



**HAL**  
open science

# Diversité et fonctions écologiques des champignons en écosystème hydrothermal marin profond

Thomas Le Calvez

► **To cite this version:**

Thomas Le Calvez. Diversité et fonctions écologiques des champignons en écosystème hydrothermal marin profond. Interactions entre organismes. Université Rennes 1, 2009. Français. NNT: . tel-00465055

**HAL Id: tel-00465055**

**<https://theses.hal.science/tel-00465055>**

Submitted on 18 Mar 2010

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

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

N° Ordre : **3899**  
de la thèse

**Thèse**

**présentée**

**DEVANT L'UNIVERSITE DE RENNES 1**

**Pour obtenir**

**Le grade de : DOCTEUR DE L'UNIVERSITE DE RENNES 1**

**Mention : BIOLOGIE**

**PAR**

**Thomas Le Calvez**

**Equipe Rôle de la Biodiversité dans les Processus Ecologiques**

**UMR 6553 Ecobio**

**Université de Rennes 1 / CNRS**

**Ecole Doctorale Vie-Agro-Santé**

**UFR Sciences de la Vie et de l'Environnement**

**Diversité et fonctions écologiques des  
champignons en écosystème hydrothermal  
marin profond**

*Le 20 avril 2009 devant la commission d'examen*

**Composition du jury :**

**Francis Martin**

**Philippe Normand**

**Peter Young**

**Ivan Couée**

**Philippe Vandenkoornhuys**

Directeur de recherche, INRA Nancy

Directeur de recherche, Université de Lyon 1

Professeur, Université d'York

Professeur, Université de Rennes 1

Professeur, Université de Rennes 1

Rapporteur

Rapporteur

Examineur

Examineur

Directeur de thèse



# Remerciements

---

Je tiens avant tout à remercier les membres du jury d'avoir accepté de lire et de juger ce travail. Je tiens donc à remercier Messieurs Philippe Normand, Francis Martin, Ivan Couée, ainsi que Peter Young, qui a fait l'effort de lire ce manuscrit en français, je vous en suis tous reconnaissant.

Un grand merci à mon directeur de thèse, Philippe Vandenkoornhuysse, pour m'avoir accueilli dans son équipe et m'avoir donné la chance de travailler sur ce sujet passionnant. Son optimisme à toute épreuve, parfois déconcertant, m'a surtout aidé à surmonter les nombreuses difficultés inhérentes à une thèse. Je le remercie également de m'avoir donné l'opportunité de travailler dans des conditions vraiment très agréables durant ces années, de m'avoir fait profiter de ses nombreuses collaborations et de m'avoir permis de voyager autant. En dehors de ses compétences professionnelles, ses qualités « humaines » doivent ici être mises en valeur ; elles ont permis une forte cohésion au sein de notre équipe.

Je tiens également à remercier chaleureusement Georges Barbier pour m'avoir accueilli dans son laboratoire de l'ESMISAB pour mon stage de Master 2, et m'avoir permis ainsi de me plonger dans l'hydrothermalisme et dans l'univers fascinant des champignons. Merci à toi Georges de m'avoir « introniser » dans le monde de la Recherche. Je tiens par ailleurs à remercier tout le personnel de l'ESMISAB, dont la sympathie m'aura permis d'oublier le climat brestois... merci à tous et spécialement à Gaétan Burgaud, pour l'heureuse collaboration que nous avons développée, ainsi qu'à Marie Hélène Thomas.

Cette thèse n'aurait sans doute pas été si agréable si je n'avais pas partagé le bureau de Mademoiselle Fanny Ramel ; sa gentillesse et son humour m'auront été d'un grand soutien. J'espère Fanny ne pas t'avoir trop martyrisé durant ces trois ans de « colocation bureaucratique » !! Notre amitié perdurera bien après nos soutenances de thèse.

Merci également à Gwenola Gouesbet Jan, Cécile Sulmon et Ivan Couée (une seconde fois), pour les nombreux fous rires au café du matin (entres autres)!... et pour leur aide.

Special thanx à l'équipe de choc : Nolwenn, Cécile, Stéphane pour l'ambiance des manips de labo, pour leurs coups de mains, les soirées...Merci à vous les filles, particulièrement à Stéphane, pour son aide lors de la seconde année de thèse !

Un grand merci également à Alexis Dufresne, dont l'aide bioinformatique, et le temps qu'il m'a accordé, m'ont été d'un secours capital pour l'avancement de ce travail.

Merci à toute l'équipe RBPE (feu IBTM) : André-Jean, Bertrand, Marie-Paule, Nathalie, Françoise, Myriam, Pierre, Daniel, Luc, et tous les autres !!!

Merci à la force vive d'ECOBIO, stagiaires, thésards, post doc, ATER, et plus particulièrement à toi Yann, forcément ! Mais également à : Emilie, Cécile, Benoît, Marion, Marie Lise, Amandine, Lisa, Denis, Fab & Nico ...

Je remercie également tout le personnel d'ECOBIO, et particulièrement Jocelyne Béven et Sandra Rigaud, pour les tracassés administratifs qu'a pu provoquer mon séjour américain, mais également à Saïd Nassur qui m'a notamment bien aidé pour l'impression de ce manuscrit, ainsi qu'Oscar Lima qui saura pourquoi !

Merci également à Laetitia Guillot pour son implication dans la mise en place de la base de données. Son sérieux et ses compétences informatiques auront permis à ce projet de se développer très rapidement.

Je tiens également à exprimer ma gratitude et mon respect à Eugene Koonin, pour m'avoir permis de séjourner plus de deux mois dans les bureaux du NCBI, Terre Sacrée de la Bioinformatique. Ces réflexions, son aide et son soutien m'ont permis d'aborder plus sereinement le flot de données auquel nous étions confrontés. Merci également à Yuri Wolf de m'avoir pris en charge au sein du NCBI ; sa patience, son calme et son accessibilité m'ont été d'un grand secours au milieu de la nébuleuse informatique. Ce séjour restera pour moi l'un des souvenirs impérissables de cette thèse.

Merci également aux financeurs, sans qui ce travail n'aurait pas été possible : le Ministère de l'enseignement supérieur et de la recherche, et la Fondation Total pour la biodiversité et la mer.

Un énorme merci également à ma famille, pour son soutien continu tout au long de ma scolarité : merci à mes parents et à ma sœur Florence.

**Enfin, merci à toi Emilie,  
pour Tout...**

# Glossaire

---

- Apomorphe** : se dit des transformations évolutives d'un caractère, de l'état ancestral vers un état dérivé.
- Attraction des longues branches** : artefact de reconstruction phylogénétique, regroupant les organismes à taux d'évolution rapides, sans liens de parenté.
- Biotrophe** : se dit d'un organisme fixé sur un organisme vivant, se développant à ses dépens, sans entraîner la mort.
- Cheminée hydrothermale** (événement hydrothermal) : larges édifices pouvant atteindre 10 mètres de hauteur, constitués de l'accrétion des métaux provenant des fluides hydrothermaux.
- Chimioautotrophie** : capacité d'un organisme à utiliser une source d'énergie minérale pour la production de carbone organique et d'énergie.
- Chimiohétérotrophie** : capacité d'un organisme à utiliser une source de carbone organique comme source de carbone et d'énergie.
- Coenocytiques** : hyphes résultant de divisions nucléaires répétées, sans divisions cellulaires concomitantes.
- Contig** : séquences nucléotidiques obtenues par assemblages des reads.
- Dorsale océanique** : chaîne de montagnes sous-marine de plus de 60.000 km de long, localisée entre deux plaques lithosphériques qui s'écartent.
- Eumycota** : champignons vrais (en opposition aux fongiformes).
- Fumeurs noirs** : fluide hydrothermal chaud résurgent de la croûte océanique, composé d'eau de mer chargée en métaux dissous. Des sulfures combinés à ces métaux vont former des particules noires.
- Homoplasie** : similarité chez une ou plusieurs espèces, de séquences (d'ADN ou de protéines), qui n'est pas due à l'héritage d'un ancêtre commun. Plusieurs cas d'homoplasies peuvent exister: l'analogie, la convergence, le parallélisme, la réversion.
- Métagénomique** : étude des génomes de l'ensemble des organismes d'un milieu.
- Monophylétique** : en phylogénie, groupe qui comprend une espèce ancestrale et tous ses descendants (en opposition avec polyphylétique). Un groupe monophylétique est également appelé clade.
- Nœud** (phylogénie) : point de rencontre de trois branches ou segments de branches dans un arbre. En phylogénie, le nœud représente un groupe comprenant les taxons frères en aval de ce nœud.
- Outgroup** : groupe externe en phylogénie.
- Read** (lecture) : séquences nucléotidiques brutes générées par pyroséquençage.
- Plésiomorphe** : se dit de l'état ancestral d'un caractère.
- Thalle** : structure dépourvue de racine, de tige ou de feuille.
- Tubicoles** : organismes vivant dans un tube, qu'ils ont construit.



## Abréviations

---

**ADNc**: ADN complémentaire  
**APS**: Adénosine 5'-Phosphosulfate  
**BAC**: Bacterial Artificial Chromosome  
**BLAST**: Basic Local Alignment Tool  
**BSA**: Bovine Serum Albumine; Serum d'albumine Bovine  
**Champignons MA** : champignons mycorhiziens à arbuscules  
**ciPCR** : Culture-independant PCR ; PCR indépendante de la culture  
**COG**: Cluster of Orthologous Group  
**Ct**: Cycle treshold  
**DMSO**: diméthylsulfoxyde  
**dNTP** : dinucléotide triphosphate  
**EDTA** : acide éthylène-diamine-tétraacétique  
**EPR**: East Pacific Rise (Ride Est-Pacifique)  
**GTR**: General Time Reversible  
**IPTG**: Isopropyl  $\beta$ -D-1-thiogalactopyranoside  
**KEGG**: Kyoto Encyclopedia of Genes and Genomes  
**K2P** : modèle de Kimura à 2 paramètres  
**LCA**: Last Common Ancestor  
**MAR**: Mid-Atlantic Rige (Ride Medio-Atlantique)  
**ML**: Maximum Likelihood (maximum de vraisemblance)  
**MP**: Maximum Parsimony (maximum de parcimonie)  
**NJ**: Neihgbor Joining  
**qPCR**: PCR quantitative  
**TBE**: Tris Borate EDTA  
**TBR**: Tree Bissection Recognition  
**Tris**: Trihydroxymethyl-amino-methane  
**WGS**: Whole Genome Sequencing  
**X-Gal**: 5-bromo-4-chloro-3-indolyl-beta-D-galactopyranoside  
**YAC**: Yeast Artificial Chromosome





# Sommaire

---

<b>Introduction générale</b>	<b>p.1</b>
<b>I-1 Présentation du Règne fongique</b>	<b>p.8</b>
1-1 Caractères généraux	p.9
1-2 Croissance	p.10
1-3 Reproduction	p.10
1-4 Modes de vie	p.11
1-4-1 Saprophytisme	p.11
1-4-2 Parasitisme et pathogénie	p.11
1-4-3 Symbiose	p.11
1-5 Organisation taxonomique	p.12
1-5-1 Phylogénie des champignons	p.13
1-5-2 Caractéristiques des organismes des 5 phyla fongiques	p.16
1-5-2-1 Chytridiomycota	p.16
1-5-2-2 Zygomycota	p.16
1-5-2-3 Glomeromycota	p.16
1-5-2-4 Dicyotes (Ascomycota et Basidiomycota)	p.17
1-6 Histoire évolutive des champignons	p.18
1-6-1 Données fossiles	p.19
1-6-2 Relations évolutives entre les groupes de champignons	p.20
1-7 Ecologie marine	p.23
1-7-1 Champignons marins	p.23
1-7-2 Champignons hydrothermaux	p.25
<b>I-2 Présentation de l'écosystème hydrothermal marin profond</b>	<b>p.27</b>
2-1 : Historique et contexte	p.28
2-2 : Principe de l'hydrothermalisme	p.31
2-3 : Localisation géographique	p.33

2-4 : Peuplements biologiques	p.33
2-4-1 Fonctionnement général	p.34
2-4-2 Communautés biologiques	p.35
2-4-2-1: Bactéries	p.35
2-4-2-2 Archaea	p.37
2-4-2-3 Eucaryotes	p.37
2-5 L'hydrothermalisme océanique : le milieu berceau de la vie ?	p.41

### **I-3 Métagénomique** p.44

3-1 Définition	p.45
3-2 Historique	p.46
3-3 La révolution des nouvelles méthodes de séquençage	p.48
3-4 Traitement des données	p.50
3-5 Applications	p.52
3-5-1 Applications de la métagénomique à la composition taxonomique d'un environnement	p.52
3-5-2 Applications de la métagénomique à la compréhension du fonctionnement des écosystèmes	p.54

## **Partie II : Matériel et méthodes** p.59

### **II-1 Sites d'études et échantillons** p.60

### **II-2-Analyse de la diversité** p.64

2-1 Extraction de l'ADN génomique total	p.64
2-2 Amplification du gène codant l'ARNr 18S	p.64
2-3 Clonage de l'ADN fongique	p.65
2-3-1 Purification	p.65
2-3-2 A-Overhanging	p.65
2-3-3 Ligation dans pGEM-T	p.66
2-3-4 Transformation par choc thermique de la souche d'E.coli DH5- $\alpha$	p.66
2-3-5 Criblage des clones bactériens	p.67
2-3-6 Séquençage	p.68
2-3-7 Détection des chimères	p.68

2-3-8 Effort d'échantillonnage	p.68
2-3-9 Soumission des séquences à Genbank	p.69
2-3-10 Culture des isolats fongiques	p.69
2-4 Analyse phylogénétiques	p.70
2-4-1 Phylogénie, définition	p.70
2-4-2 Alignement	p.70
2-4-3 Méthodes de reconstructions phylogénétiques	p.70
2-4-3-1 Méthode phénétique	p.70
2-4-3-2 Calcul des valeurs de bootstrap	p.71
2-4-3-3 Méthode de Maximum de Parcimonie	p.72
2-4-3-4 Méthode de Maximum de Vraisemblance	p.73
<b>II-3 Quantification du nombre de copies du gène codant l'ARNr18S</b>	<b>p.77</b>
<b>par PCR quantitative</b>	
3-1 Dessin des amorces	p.77
3-2 Calcul de la gamme	p.78
3-2-1 Extraction et linéarisation des plasmides	p.78
3-2-2 Calcul du nombre de copies dans les standards	p.79
3-2-3 Préparation de la gamme	p.79
3-3 Quantification	p.79
<b>II-4 Création d'une base de données de marqueurs moléculaires de</b>	<b>p.82</b>
<b>Champignons</b>	
<b>II-5 Analyse du métagénome</b>	<b>p.86</b>
5-1 Pyroséquençage	p.86
5-1-1 Préparation de la banque d'ADN	p.86
5-1-2 Amplification clonale	p.86
5-1-3 Pyroséquençage	p.86
5-2 Analyses bioinformatiques	p.89
5-2-1 Analyse des contigs assemblés	p.90

5-2-1-1 Analyse BLAST X	p.90
5-2-1-2 Effort de séquençage	p.94
5-2-1-3 Analyse des contigs fongiques	p.95
5-2-2 Analyse des séquences générées, non assemblées (reads)	p.97
5-2-2-1 Filtrage des données	p.97
5-2-2-2 Analyse BLASTX	p.98
5-2-2-3 Analyse MEGAN	p.102
5-2-2-4 Reconstruction des voies métaboliques	p.105

**Partie III : Mise en place d'une base de données experte de marqueurs moléculaires fongiques : PHYMYCO-DB**

Résumé	p.111
Publication : PHYMYCO-DB: a curated database for analyses of fungal diversity and evolution	p.113

**Partie IV : Diversité fongique dans les écosystèmes hydrothermaux marins profonds**

Résumé	p.124
Publication: Fungal Diversity in Deep Sea Hydrothermal Ecosystems	p.126

**Partie V : Analyse du métagénome extrait d'un écosystème profond**  
**Hypothèses sur le mode de vie des champignons,**  
**Reconstructions métaboliques du compartiment bactérien**

<b>V-1 Choix de l'échantillon</b>	<b>p.152</b>
1-1 Analyse de diversité complémentaire	p.153
1-2 Quantification des molécules d'ARNr 18S dans les échantillons hydrothermaux	p.153
1-3 Détermination de l'échantillon	p.155

**V-2 Présentation des résultats de l'analyse métagénomique**

Hypothèses de fonctionnement du compartiment microbien (fongique et bactérien) des écosystèmes hydrothermaux.

**Partie VI Diversité des souches fongiques filamenteuses \_\_\_\_\_ p.184**  
**isolées d'écosystèmes hydrothermaux marins profonds**

Résumé \_\_\_\_\_ p.185

Publication: Diversity of culturable marine filamentous \_\_\_\_\_ p.186  
fungi from deep-sea hydrothermal vents

**Discussion générale et perspectives de recherche \_\_\_\_\_ p.212**

I- Diversité des champignons en écosystème hydrothermal \_\_\_\_\_ p.213

II- Fonctions des champignons en écosystème hydrothermal \_\_\_\_\_ p.219

**Références bibliographiques \_\_\_\_\_ p.226**



# Liste des figures

---

- Figure 1:** **p.3**  
Extrait de l'arbre phylogénétique, établi par le regroupement d'études moléculaires et structurales, montrant les relations entre les organismes du groupe des Opisthokontes.
- Figure 2:** **p.4**  
Représentation schématique des résultats d'horloge moléculaire, effectuée par l'analyse du gène codant l'ARNr 18S, sur la base des résultats de Vandenkoornhuyse *et al.*, (2002).
- Figure 3:** **p.5**  
Echelle géologique, reprenant les principales d'hypothèses de l'émergence des champignons.
- Figure 4:** **p.14**  
Représentation de l'arbre de la vie, établi par le regroupement d'études moléculaires et structurales.
- Figure 5:** **p.15**  
Photographies de champignons des 5 phyla.
- Figure 6:** **p.22**  
Phylogénies des champignons extraites de Liu *et al.*, (2006) s'appuyant sur les séquences RBP1 et RBP2.
- Figure 7:** **p.30**  
Exemples de matériels utilisés pour les prélèvements en écosystèmes hydrothermaux marins profonds.
- Figure 8:** **p.32**  
Schéma d'une section longitudinale d'un volcan sous marin hydrothermal et des principaux mécanismes impliqués dans l'hydrothermalisme.
- Figure 9:** **p.33**  
Localisations géographiques des sites hydrothermaux. Les principaux sites de la Dorsale Médio-Atlantique (MAR) et de la dorsale Est-Pacifique (EPR) sont représentés.
- Figure 10:** **p.36**  
Différents types de métabolismes bactériens (ici les  $\epsilon$ -protéobactéries cultivées) associés à l'hydrothermalisme.



<b><u>Figure 11:</u></b>	<b>p.37</b>
Photographie de microscopie électronique à balayage d'archaea : <i>Pyrococcus abyssi</i> .	
<b><u>Figure 12:</u></b>	<b>p.40</b>
Photographies des principaux invertébrés des écosystèmes hydrothermaux.	
<b><u>Figure 13:</u></b>	<b>p.42</b>
Hypothèses de scénarii évolutifs, ayant mené à la formation des cellules de bactéries et d'archaea, à partir de la formation de compartiments inorganiques de Pyrite.	
<b><u>Figure 14:</u></b>	<b>p.47</b>
Etapas de l'analyse et de l'identification des gènes à partir d'un métagénome microbien, selon une stratégie de transformations de vecteurs plasmidiques ou viraux.	
<b><u>Figure 15:</u></b>	<b>p.50</b>
Etapas de l'analyse et de l'identification des gènes à partir d'un métagénome microbien, selon une stratégie de pyroséquençage.	
<b><u>Figure 16:</u></b>	<b>p.55</b>
Historique des études métagénomiques depuis 2003.	
<b><u>Figure 17:</u></b>	<b>p.61</b>
Cartographie des principaux sites étudiés sur la ride Medio Atlantique.	
<b><u>Figure 18:</u></b>	<b>p.61</b>
Localisation géographique des sites hydrothermaux de la ride Est Pacifique.	
<b><u>Figure 19:</u></b>	<b>p.66</b>
Carte circulaire du plasmide pGEM-T (Promega).	
<b><u>Figure 20:</u></b>	<b>p.81</b>
Exemple de courbe standard obtenu avec la gamme de standard.	
<b><u>Figure 21:</u></b>	<b>p.84</b>
Capture d'écran de la page d'accueil de PHYMYCO-DB.	
<b><u>Figure 22:</u></b>	<b>p.84</b>
Capture d'écran présentant l'arborescence taxonomique pour extraire les informations de PHYMYCO-DB.	
<b><u>Figure 23:</u></b>	<b>p.85</b>
Capture d'écran présentant les fichiers d'alignement téléchargeables à partir de PHYMYCO-DB.	
<b><u>Figure 24:</u></b>	<b>p.88</b>
Schéma des étapes de pyroséquençage.	

<b><u>Figure 25:</u></b>	<b>p.88</b>
Schéma de la réaction de pyroséquençage lors de l'incorporation d'un nucléotide par la polymérase.	
<b><u>Figure 26:</u></b>	<b>p.89</b>
Schéma global des stratégies utilisées pour analyser les données issues du pyroséquençage.	
<b><u>Figure 27:</u></b>	<b>p.96</b>
Symboles utilisés pour la classification des KOGs.	
<b><u>Figure 28:</u></b>	<b>p.106</b>
Schéma général du fonctionnement du serveur mgRAST .	
<b><u>Figure 29:</u></b>	<b>p.120</b>
Flowchart of the data in the PHYMYCO-DB database.	
<b><u>Figure 30:</u></b>	<b>p.121</b>
Visualisation of different examples of sequences deleted from manual curations.	
<b><u>Figure 31:</u></b>	<b>p.139</b>
Phylogenetic positions of deep sea hydrothermal fungi.	
<b><u>Figure 32:</u></b>	<b>p.140</b>
Estimates of fungal community diversity as a function of sampling effort for environmental samples.	
<b><u>Figure 33:</u></b>	<b>p.143</b>
Phylogenetic affinities within the Ascomycota phylum of the fungal SSU rRNA sequences from a deep sea hydrothermal ecosystem.	
<b><u>Figure 34:</u></b>	<b>p.145</b>
Phylogenetic affinities within the Basidiomycota phylum of the fungal SSU rRNA sequences from a deep sea hydrothermal ecosystem.	
<b><u>Figure 35:</u></b>	<b>p.147</b>
Phylogenetic affinities within the Zygomycota phylum of the fungal SSU rRNA sequences from a deep sea hydrothermal ecosystem.	
<b><u>Figure 36:</u></b>	<b>p.149</b>
Phylogenetic affinities within the Chytridiomycota phylum of the fungal SSU rRNA sequences from a deep sea hydrothermal ecosystem.	
<b><u>Figure 37:</u></b>	<b>p.154</b>
Analyses phylogénétiques des séquences de champignons d'échantillons isolés d'écosystèmes hydrothermaux.	

<b><u>Figure 38:</u></b>	<b>p.158</b>
Taxonomic distribution by domain of life using first BLAST hit.	
<b><u>Figure 39:</u></b>	<b>p.161</b>
KOG distribution of 178 fungal contigs.	
<b><u>Figure 40:</u></b>	<b>p.164</b>
Hypothetical fungal cell metabolism, from fungal contigs extracted from metagenome.	
<b><u>Figure 41:</u></b>	<b>p.168</b>
Taxonomic distribution of bacterial reads, according to MEGAN software.	
<b><u>Figure 42:</u></b>	<b>p.168</b>
Taxonomic distribution of bacterial reads, according to MEGAN software.	
<b><u>Figure 43:</u></b>	<b>p.170</b>
Kegg map for carbon fixation pathways analysed using ASGARD.	
<b><u>Figure 44:</u></b>	<b>p.171</b>
Map for the Acetyl coA pathway	
<b><u>Figure 45:</u></b>	<b>p.172</b>
Kegg map for reductive carboxylate cycle pathways analysed using ASGARD.	
<b><u>Figure 46:</u></b>	<b>p.174</b>
Kegg map for glycolytic and gluconeogenetic pathways analysed using ASGARD.	
<b><u>Figure 47:</u></b>	<b>p.176</b>
Kegg map for methane metabolism pathways analysed using ASGARD.	
<b><u>Figure 48:</u></b>	<b>p.178</b>
Kegg map for nitrogen metabolism pathways analysed using ASGARD.	
<b><u>Figure 49:</u></b>	<b>p.180</b>
Kegg map for sulfur metabolism pathways analysed using ASGARD.	
<b><u>Figure 50:</u></b>	<b>p.182</b>
Hypothetical interactions between microorganisms in this habitat.	
<b><u>Figure 51:</u></b>	<b>p.208</b>
Physiological analysis of the fungal collection.	
<b><u>Figure 52:</u></b>	<b>p.210</b>
SSU rRNA phylogenetic positions of deep-sea fungal isolates (colored terminals).	

# Liste des tableaux

---

<b><u>Tableau 1:</u></b>	<b>p.20</b>
Temps de divergence, en millions d'années, entre les principaux groupes.	
<b><u>Tableau 2:</u></b>	<b>p.34</b>
Différents types de métabolismes bactériens retrouvés en écosystèmes hydrothermaux.	
<b><u>Tableau 3:</u></b>	<b>p.50</b>
Comparaison des techniques de séquençage de nouvelle génération.	
<b><u>Tableau 4:</u></b>	<b>p.91</b>
Récapitulatif sur l'utilisation des programmes de recherche BLAST.	
<b><u>Tableau 5:</u></b>	<b>p.95</b>
Organismes bactériens de référence utilisés pour le test d'effort de séquençage.	
<b><u>Table 6:</u></b>	<b>p.141</b>
Summary of samples analyzed in the study. Identification, nature of samples, location (MAR: Mid-Atlantic Ridge, EPR: East Pacific Rise), and depth are indicated. Numbers of phylotypes and corresponding numbers of sequences are also reported. Results from qPCR analyses are presented in the last row.	
<b><u>Tableau 7:</u></b>	<b>p.155</b>
Caractéristiques de l'échantillon pyroséquéncé.	
<b><u>Table 8:</u></b>	<b>p.159</b>
Sequencing effort by comparisons of coverage tests, from 2 runs or 3 runs of pyrosequencing.	
<b><u>Table 9:</u></b>	<b>p.165</b>
Number of protein and nucleotide sequences in Genbank in February 2009 (whole database and RefSeq databases), for each fungal phylum.	
<b><u>Table 10:</u></b>	<b>p.206</b>
Culture collection of filamentous fungi from deep-sea hydrothermal vents.	
<b><u>Tableau 11:</u></b>	<b>p.216</b>
Les principaux habitats de l'océan profond, leurs caractéristiques et les principales applications biotechnologiques associées.	
<b><u>Tableau 12:</u></b>	<b>p.222</b>
Génomes nucléaires complets de champignons, annotés et disponibles dans Genbank, en Février 2009.	



## **Partie I : Introduction générale**

---

## Introduction générale

On estime à 1.5 millions le nombre d'espèces de champignons vivant sur Terre (Hawksworth, 2001). Cependant, seulement 14% de ces organismes ont été découverts, ce qui peut être expliqué par la difficulté de cultiver ces organismes en conditions de laboratoire mais aussi en raison d'un manque de connaissances et d'un sous-échantillonnage de ces organismes dans de nombreux biotopes de la planète (Hawksworth & Rossman, 1997 ; Hawksworth, 2001 ; Neubert *et al.*, 2006). Connus depuis l'Égypte ancienne où les processus fermentaires qu'ils empruntent étaient déjà utilisés dans l'alimentation, les champignons représentent un groupe d'organismes majeurs à la fois dans le fonctionnement des écosystèmes mais également dans des applications biotechnologiques qu'ils peuvent potentiellement diriger (Schmit & Mueller, 2007 ; Bérdy, 2005). On peut par exemple citer la découverte en 1928 de la pénicilline produite par une souche fongique *Penicillium notatum*, qui ouvrira la voie à la biosynthèse d'antibiotiques utilisés dans la pharmacopée depuis 1940. De nos jours encore, les processus fermentaires de certaines levures sont utilisés dans de nombreux domaines en agro-alimentaire, comme par exemple la « levure de boulanger » *Saccharomyces cerevisiae*, utilisée pour la panification mais également en brasserie. Cette levure est aussi un modèle d'études important en génétique et biologie moléculaire et cellulaires (par exemple Carson & Hartwell, 1985, ou plus récemment Gertz *et al.*, 2009).

La capacité de production de métabolites secondaires, telles que les molécules antibiotiques, ont favorisé l'utilisation biotechnologique des champignons. On recensait en 2006 plus de 8600 métabolites bioactifs produits par des souches fongiques, dont 4900 auraient une activité antibiotique (Bérdy, 2005). La majorité des organismes producteurs de molécules bioactives seraient filamenteux et une très faible part d'organismes levuriformes serait impliquée. Pour exemple, 900 composés bioactifs ont pu être isolés des champignons du genre *Penicillium* (Bérdy, 2005). Cependant, en considérant le grand nombre d'espèces fongiques non cultivées, de nombreuses souches productrices de molécules bioactives restent à découvrir. Mieux appréhender la diversité globale des champignons sur Terre permettrait donc potentiellement de découvrir des organismes aux applications biotechnologiques intéressantes. De plus, compte tenu de leur rôle pivot dans le fonctionnement des écosystèmes terrestres aériens, comprendre leurs rôles dans des écosystèmes où ces organismes ont été peu

ou pas étudiés permettrait également de mieux comprendre le fonctionnement global de ces écosystèmes.

Le Règne fongique constitue un groupe monophylétique au sein de l'Arbre de la vie et un groupe frère des animaux au sein des Opisthokontes (Figure 1) (Baldauf, 1999 ; Baldauf *et al.*, 2004). Une meilleure compréhension de la diversité fongique permettrait de mieux comprendre les étapes évolutives majeures ayant conduit à la diversification des Opisthokontes à la radiation des champignons. L'état des connaissances sur les relations évolutives entre champignons sera abordé au chapitre I-1 de cette thèse.

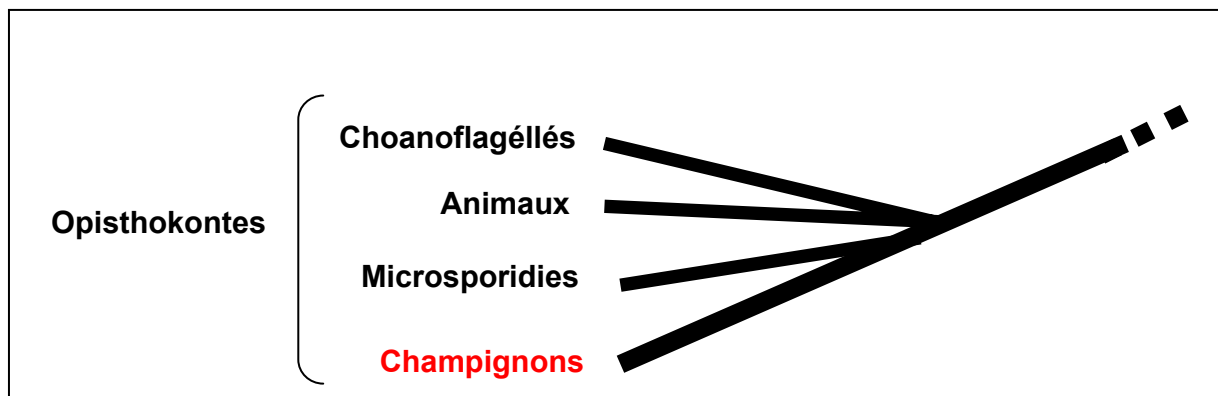


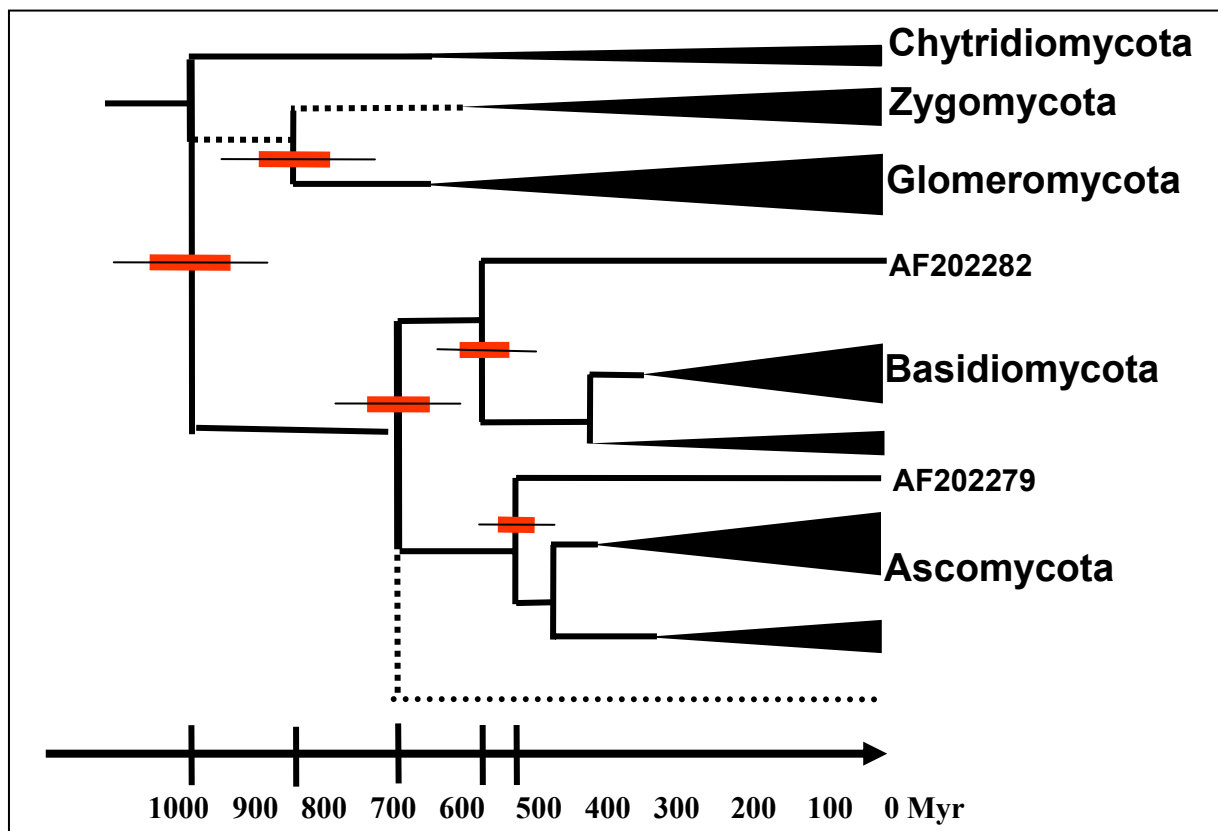
Figure 1: Extrait de l'arbre phylogénétique, établi par le regroupement d'études moléculaires et structurales, montrant les relations entre les organismes du groupe des Opisthokontes. Extrait d'après Baldauf *et al.*, 2004 in *Assembling the Tree of Life*.

## Contexte

L'étude réalisée lors de cette thèse trouve ses racines dans des travaux publiés par Vandenkoornhuyse *et al.* (2002a). Ce travail établissait la diversité globale fongique associée aux racines de plantes par une approche moléculaire. Bien que déjà étudiée, cette niche écologique a révélé une diversité de champignons insoupçonnée. Parmi les séquences obtenues lors de cette étude, des séquences représentant des branches profondes des phyla Ascomycota et Basidiomycota ont été découvertes (Vandenkoornhuyse *et al.*, 2002a). En raison de ces nouvelles branches évolutives mises en évidence, un nouveau regard pouvait être porté sur l'évolution des champignons. Ceci a stimulé la réalisation d'analyses d'horloges moléculaires à partir de ces nouvelles données dont les résultats sont présentés dans la figure 2.



Pour chaque phylum fongique, cette analyse suggère l'émergence hypothétique des champignons il y a 1 milliard d'années, une estimation antérieure de 100 à 200 millions d'années par rapport aux données généralement acceptées (Figure 2). En replaçant ces résultats sur une échelle géologique (Figure 3), l'émergence hypothétique des champignons aurait eu lieu 400 à 500 millions d'années avant les premières traces fossiles de vie en milieu aérien (estimées à 380 millions d'années). L'émergence des champignons aurait donc eu lieu avant la formation des continents (Figure 3), donc en milieu océanique. Bien que présentant certaines différences, ce travail d'analyse est cohérent avec une étude publiée par Heckman *et al.* (2001).



**Figure 2:** Représentation schématique des résultats d'horloge moléculaire, effectuée par l'analyse du gène codant l'ARNr 18S, sur la base des résultats de Vandenkoornhuysen *et al.*, (2002a). Les barres rouges horizontales indiquent l'incertitude quant à la datation compte tenu des différentes calibrations utilisées et sont accompagnées d'un intervalle de confiance (en millions d'années).

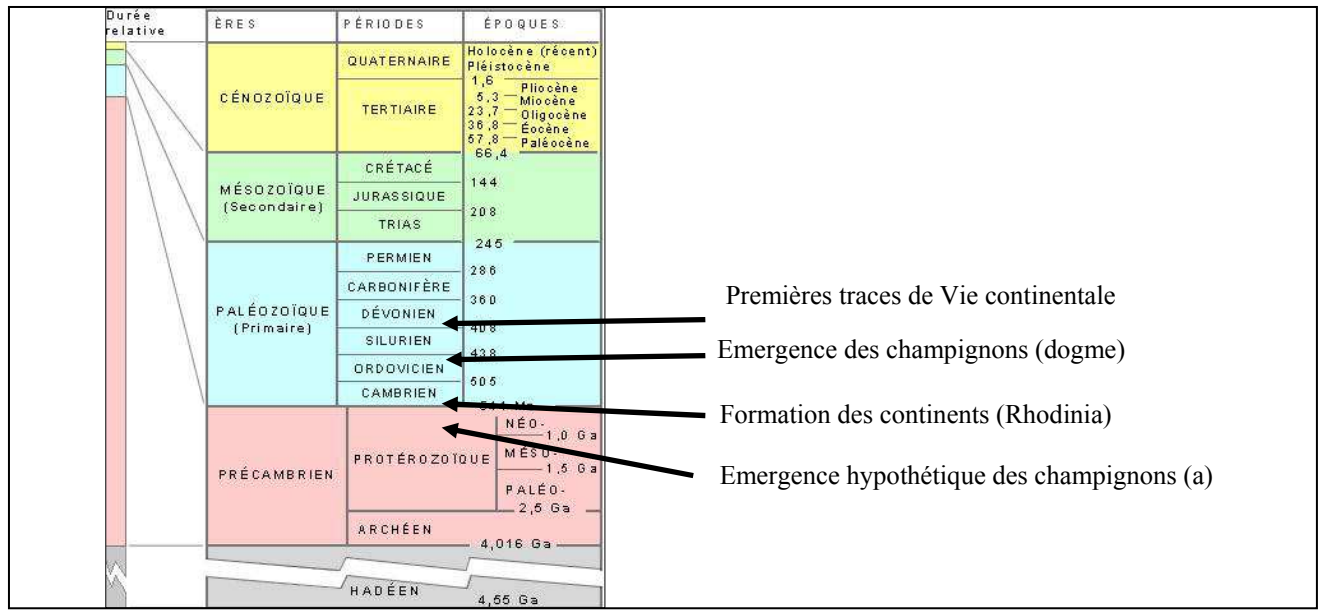


Figure 3: Echelle géologique reprenant les principales d'hypothèses de l'émergence des champignons.

a- d'après l'estimation sur la base des résultats de Vandenkoornhuyse *et al.* (2002a), présentés dans la figure 2.

En partant du postulat d'une émergence et d'une diversification des champignons en milieu marin contrairement au dogme d'une diversification des groupes majeurs de champignons en milieu aérien, nous avons choisi d'étudier la diversité fongique en écosystème hydrothermal marin profond. Cet écosystème est considéré par certains auteurs comme le milieu berceau de la vie (Wächtershäuser, 1988) et partagerait des caractéristiques chimiques avec l'Océan Primitif (Holm, 1992). Les détails de cette théorie seront expliqués dans le Chapitre I-2 de cette thèse.

## Objectifs

L'objet de ce travail de thèse a été d'analyser dans un premier temps la diversité des champignons au niveau des écosystèmes hydrothermaux marins profonds, au niveau de différentes zones géographiques. Cette analyse nous a permis d'effectuer un inventaire des organismes présents dans ces écosystèmes peu ou pas étudiés en matière de diversité fongique. Cette analyse de diversité (présentée dans la partie IV) nous permet également de

proposer des théories évolutives sur les radiations des groupes majeurs de champignons. Outre l'étude de la diversité, nous nous sommes également intéressés à l'étude du fonctionnement des champignons présents dans ces écosystèmes par une approche métagénomique, afin de mieux percevoir leur rôle au sein de ces habitats. Cette étude prédictive a été envisagée par la génération de données de séquençage de masse. Ce travail nous a également permis de mieux appréhender le fonctionnement microbien global de cet écosystème ; les résultats de l'analyse métagénomique seront présentés dans le chapitre V de cette thèse.

## **Plan de la thèse**

Cette thèse abordera dans la partie introductive les connaissances sur le Règne des champignons et plus particulièrement sur les relations évolutives entre les principaux phyla fongiques (partie I-1). L'écosystème hydrothermal marin profond sera également décrit (partie I-2). Une étude bibliographique fait le point sur l'utilisation de stratégies de séquençage de masse dans le but de caractériser les environnements microbiens par approches prédictives fonctionnelles (partie I-3).

La seconde partie de ce document, le Matériel et Méthodes (partie II), présentera d'une part les sites étudiés ainsi que les échantillons prélevés, pour ensuite s'intéresser aux techniques ayant permis de caractériser la diversité fongique de ces environnements ; seront ainsi abordées les méthodes utilisées pour construire une base de données de séquences nucléotidiques spécifiques aux champignons, ainsi que les méthodes d'inférences phylogénétiques utilisées pour déterminer les organismes présents dans les échantillons étudiés. Un rappel sur les techniques de séquençage de masse de dernière génération sera également effectué, ainsi que sur les méthodes informatiques que nous avons choisies pour traiter l'ensemble du jeu de données générées par le pyroséquençage.

Les résultats de cette étude sont présentés sous la forme de 4 articles, précédés d'un résumé.

Le premier article de ce travail, proposé dans la partie III, présentera la mise en place de la base de données de séquences fongiques PHYMYCO-DB.

Le second article, dans la partie IV, présentera les résultats obtenus sur l'étude de la diversité fongique, et les hypothèses évolutives qui en découlent.

La troisième article récapitulera les résultats obtenus lors de l'analyse du jeu de données métagénomiques qui a permis d'émettre des hypothèses prédictives du fonctionnement des organismes fongiques dans ces écosystèmes, et des interactions biotiques

vraisemblables. Cet article fait suite à une explication détaillée des raisons du choix de l'échantillon étudié. Par ailleurs cet article est complété par une analyse détaillée du métagénome global sur la base des données de séquences obtenues. L'ensemble de ce travail est présenté dans la partie V.

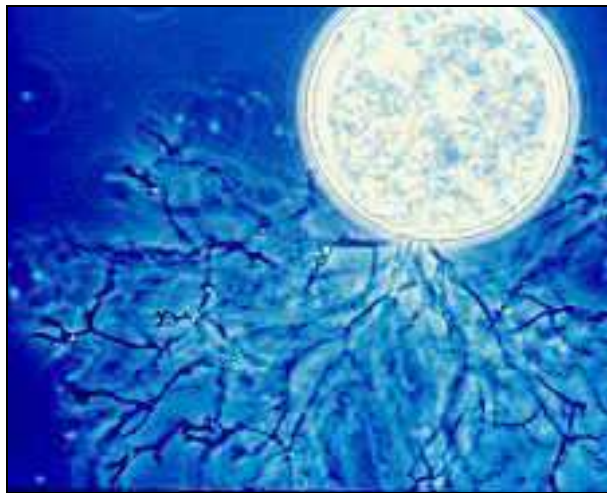
Le dernier article, en partie VI, présente les résultats obtenus en collaboration avec l'ESMISAB de Brest, sur l'étude de la fraction cultivable des organismes fongiques isolés des mêmes échantillons hydrothermaux; des notions de physiologie préliminaires de ces organismes seront abordées, notamment concernant la capacité des souches isolées à croître en milieu salin, et donc sur leur capacité à se développer en milieu marin.

Cette thèse se terminera par une synthèse des principaux résultats obtenus, qui seront discutés, ainsi que sur les perspectives de recherche de ce travail.



## Chapitre I -1 : Présentation du Règne fongique

---



# I-1 Présentation du Règne fongique

## 1-1 Caractères généraux

Les champignons représentent l'un des plus importants groupes d'organismes sur Terre et jouent un rôle clé dans un grand nombre d'écosystèmes (Mueller & Schmit, 2007). Ce sont des organismes eucaryotes à mode de reproduction sexuée ou asexuée. Les spores produites peuvent avoir un rôle dans la dispersion des champignons, mais peuvent également jouer un rôle dans la survie de l'organisme lorsque les conditions environnementales deviennent défavorables (Madelin, 1994). Leur mode de nutrition se fait par absorption en libérant dans un premier temps des enzymes hydrolytiques dans le milieu extérieur. Ces organismes sont dépourvus de chlorophylle et sont tous hétérotrophes ; le glycogène est le polysaccharide de réserve principal (Carlile & Watkinson, 1994 ; Redecker, 2002). Tous les champignons ont une paroi constituée de chitine, polysaccharide très résistant constitué de résidus N-acétylglucosamine (Carlile & Watkinson, 1994).

D'un point de vue structural, on trouve une grande variété de champignons. Ils sont classés en deux grandes catégories : la forme levure unicellulaire et la forme mycélienne pluricellulaire constituée d'hyphes (Redecker, 2002). Certaines espèces ont la capacité d'adopter les deux formes, levure et mycélienne, tandis que d'autres sont restreintes à l'une ou l'autre (Jennings & Lysek, 1996). La forme levure apporte un avantage pour la croissance dans les milieux où la pression osmotique est forte car cela diminue la surface de l'organisme. La forme mycélienne permet au champignon d'avoir une croissance radiale importante et de coloniser rapidement un milieu. Cette forme mycélienne assure donc une surface maximale de contact et permet une exploration et une recherche de nutriments dans les trois dimensions (Carlile & Watkinson, 1994 ; Jennings & Lysek, 1996).

D'un point de vue métabolique les champignons sont des chimiohétérotrophes, c'est à dire qu'ils utilisent du carbone organique comme source d'énergie (Carlile & Watkinson, 1994 ; Redecker, 2002). Ce sont des organismes aérobies pour la grande majorité, mais certaines levures peuvent être aéro-anaérobie et participer à des processus fermentaires (Carlile & Watkinson, 1994).

## 1-2 Croissance

La croissance des champignons mycéliens est assurée par des hyphes qui sont constitués de cellules hétérocaryotiques (Ascomycota et Basidiomycota) ou coenocytiques (Zygomycota et Glomeromycota). Leur extension est restreinte à l'apex. Après division, l'article apical nouvellement formé peut se séparer du reste du mycélium par une cloison (mycélium septé) ou non (mycélium siphonné) (Jennings & Lysek, 1996). Les hyphes vont se brancher en réseau, déterminant en partie la morphologie macroscopique du thalle (Carlile & Watkinson, 1994). La croissance et la nutrition vont se faire de façon concomitante ; la croissance sera réalisée par une extension de la paroi à l'apex, par un apport continu de chitine. Dans le même temps, au niveau de l'apex également, des enzymes hydrolytiques seront déversées dans le milieu extérieur (Carlile & Watkinson, 1994).

## 1-3 Reproduction

Le cycle sexuel des champignons se déroule en trois étapes : plasmogamie, caryogamie et méiose (Jennings & Lysek, 1996). La plasmogamie correspond à la fusion cellulaire entre deux cellules haploïdes. La cellule résultante est appelée dicaryon car elle possède deux types de noyaux haploïdes. Les deux noyaux vont fusionner lors de la caryogamie puis la méiose va convertir une cellule diploïde en quatre cellules haploïdes (Carlile & Watkinson, 1994). On recense également des modes de reproduction différents de celui qui précède : certains organismes garderont un mode de vie haploïde, d'autres un mode de vie uniquement diploïde, tandis que certains organismes (Deutéromycètes) n'ont pas de capacité de reproduction sexuée (Carlile & Watkinson, 1994). Les Gloméromycètes ont quant à eux un mode de reproduction très mal compris même si le mode de reproduction asexuée soit généralement accepté chez les organismes de ce phylum (Schüßer *et al.*, 2001 ; Redecker, 2002). En effet, la diversité intraspécifique élevée pour ce phylum peut être expliquée par des phénomènes de recombinaison dans les hyphes et spores coenocytiques (Vandenkoornhuyse *et al.*, 2001) ou par réassortiment de noyaux différents (Sanders, 2004).

Les spores peuvent être répandues dans le milieu de façon passive ou active par le champignon mais leur dispersion se fera toujours passivement, selon différents modes : une dispersion par le vent, par les animaux (notamment les insectes), mais également par la graine des plantes colonisées (Carlile & Watkinson, 1994). L'eau est aussi un vecteur important de dissémination des spores ; il est à noter que les zoospores (chez les organismes



du phylum Chytridiomycota) ont la faculté de nager grâce à leur flagelle (Carlile & Watkinson, 1994).

## **1-4 Modes de vie**

Les champignons ont un rôle très important dans la dégradation de la matière organique et constituent une part importante des décomposeurs sur Terre (Lutzoni *et al.*, 2004). De plus, certains champignons peuvent être phytopathogènes ou provoquer des mycoses chez les animaux (Lutzoni *et al.*, 2004). Un troisième mode de vie, symbiotique, est également très répandu.

### **1-4-1 Saprophytisme**

Les champignons ont un rôle très important dans le recyclage de la matière organique sur Terre (Galagan *et al.*, 2003). Leur capacité d'exploration via l'extension des hyphes, couplée à la capacité de largage d'enzymes hydrolytiques, ont permis une colonisation d'une grande variété de substrats. Dans le sol, les champignons participent au cycle de l'azote par la dégradation de l'humus. Ils ont la capacité de consommer la cellulose ainsi que la lignine et sont considérés comme les principaux recycleurs de matière organique à partir de matériels végétaux (Lutzoni *et al.*, 2004).

### **1-4-2 Parasitisme et pathogénie**

Environ 20% des espèces de champignons connues sont capables de parasitisme. On trouvera des parasites obligatoires, facultatifs ou opportunistes (Lutzoni *et al.*, 2004). Les opportunistes sont des organismes saprophytes qui vont s'attaquer aux organismes dont les défenses sont affaiblies (Rinaldi, 1989). Les champignons peuvent attaquer tous les groupes du vivant, comme par exemple les plantes, les insectes, les animaux mais aussi les bactéries et les autres champignons (Lutzoni *et al.*, 2004).

### **1-4-3 Symbiose**

Les associations symbiotiques entre champignons et végétaux supérieurs (mycorhize) constituent la forme de symbiose la plus répandue à l'échelle planétaire (Jennings & Lysek, 1996). On estime que 90% des végétaux contractent spontanément cette association (Smith & Read, 1997).

Les champignons vont développer un réseau de filaments mycéliens à partir de la racine et vont être impliqués dans la nutrition minérale des plantes. C'est d'ailleurs une association symbiotique qui aurait permis aux plantes de coloniser le milieu terrestre (Simon *et al.*, 1993). Outre leur capacité à augmenter l'exploration du milieu extérieur, les champignons vont également contribuer à la phytoprotection par élicitation des mécanismes de défense (Smith & Read, 1997), et produire des substances antibiotiques permettant de lutter contre d'autres microorganismes pathogènes à la plante. La contrepartie de l'association est que le champignon va recevoir de la plante des photosynthétats. Il existe plusieurs types de relations mycorhiziennes. Les plus fréquentes sont les champignons mycorhiziens à arbuscules (MA) (endosymbiose) et les ectomycorhizes (ectosymbioses) (Jennings & Lysek, 1996). Il a été démontré que la diversité des champignons MA dans le sol contrôle la diversité des plantes en un lieu (Van der Heijden *et al.*, 1998) et donc contrôle indirectement l'ensemble de l'écosystème.

Une autre forme de symbiose fortement répandue sur Terre est représentée par les lichens, qui sont constitués d'une association entre champignon (principalement du phylum Ascomycota) et une cyanobactérie. L'algue, capable de photosynthèse, va fournir les molécules organiques carbonées au champignon qui en retour fournira les éléments minéraux à l'algue (Gargas *et al.*, 1995).

## **1-5 Organisation taxonomique**

Le concept d'espèces chez les champignons est une notion alimentant les débats. Ces espèces peuvent être différenciées de différentes façons ; historiquement, les champignons étaient reconnus sur la base de leur morphologie selon le concept morphologique d'espèce (Taylor, 2000). La grande majorité des espèces décrites à ce jour reposent sur ce concept, par une analyse des traits phénotypiques (morphologiques et/ou biochimiques) (Taylor, 2000 ; Guarro *et al.*, 1999). Les champignons peuvent être également décrits selon leurs habitats, leurs localisations géographiques, selon le concept écologique de l'espèce (Taylor, 2000). L'application récente du concept phylogénétique de l'espèce, qui décrit l'espèce comme un groupe monophylétique au sein duquel le taux d'homologie entre séquences est élevé, a été favorisée par les progrès des techniques d'analyses moléculaires, et tend à être massivement

employée en mycologie (Mischler & Brandon 1987). Ce sera d'ailleurs le concept que nous adopterons dans ce travail.

### **1-5-1 Phylogénie des champignons**

L'essor des techniques moléculaires depuis les années 90, notamment le développement de la technique de PCR (Polymerase Chain Reaction), a permis l'étude des champignons à l'échelle de la spore mais aussi sur de très petites quantités d'ADN, ce qui a permis d'identifier les espèces réticentes à la culture (Taylor, 2000). Un gène ou marqueur moléculaire fongique est sélectivement amplifié à l'aide d'amorces spécifiques des champignons, puis le gène est séquencé. Le concept phylogénétique de l'espèce consiste à comparer les séquences d'ADN entre plusieurs individus afin de comparer les substitutions nucléotidiques. Ce concept va de plus fournir des informations sur l'évolution des espèces (information diachronique), le nombre de substitutions d'une séquence d'ADN étant proportionnelle au temps de divergence depuis l'ancêtre commun selon l'hypothèse de l'horloge moléculaire.

L'analyse de l'Arbre de la vie réalisé par Baldauf (2004) sur les bases de données moléculaires et structurales, montre que les champignons appartiennent au groupe des Opisthokontes (Figure 1), au sein de l'Empire Eucaryote (l'arbre complet est présenté en Figure 4). Historiquement, les champignons étaient classés dans le groupe des Plantes et étaient considérés comme proches des végétaux ; leur première classification en tant que Règne date de 1969 (Whittaker, 1969). Une étude basée sur l'analyse du gène codant l'ARNr18S a permis de suggérer un ancêtre commun aux animaux et champignons (Wainright *et al.*, 1993), mais ce sont des études phylogénétiques sur l'analyse du facteur d'élongation EF1-alpha qui ont prouvé que champignons et animaux étaient des groupes frères, au sein du groupe des Opisthokontes (Baldauf & Palmer, 1993; Baldauf, 1999) (Figures 1 et 4). Le groupe des Opisthokontes intègre aussi les Choanoflagellés qui constitueraient une branche ancienne d'un point de vue évolutif. La caractéristique principale des Opisthokontes est la présence des cellules sexuées mobiles à un flagelle polaire. Une autre caractéristique des Opisthokontes est l'utilisation du codon UGA pour la transcription du tryptophane. De plus, tous ces organismes possèdent du glycogène et de la chitine (Baldauf, 1999).

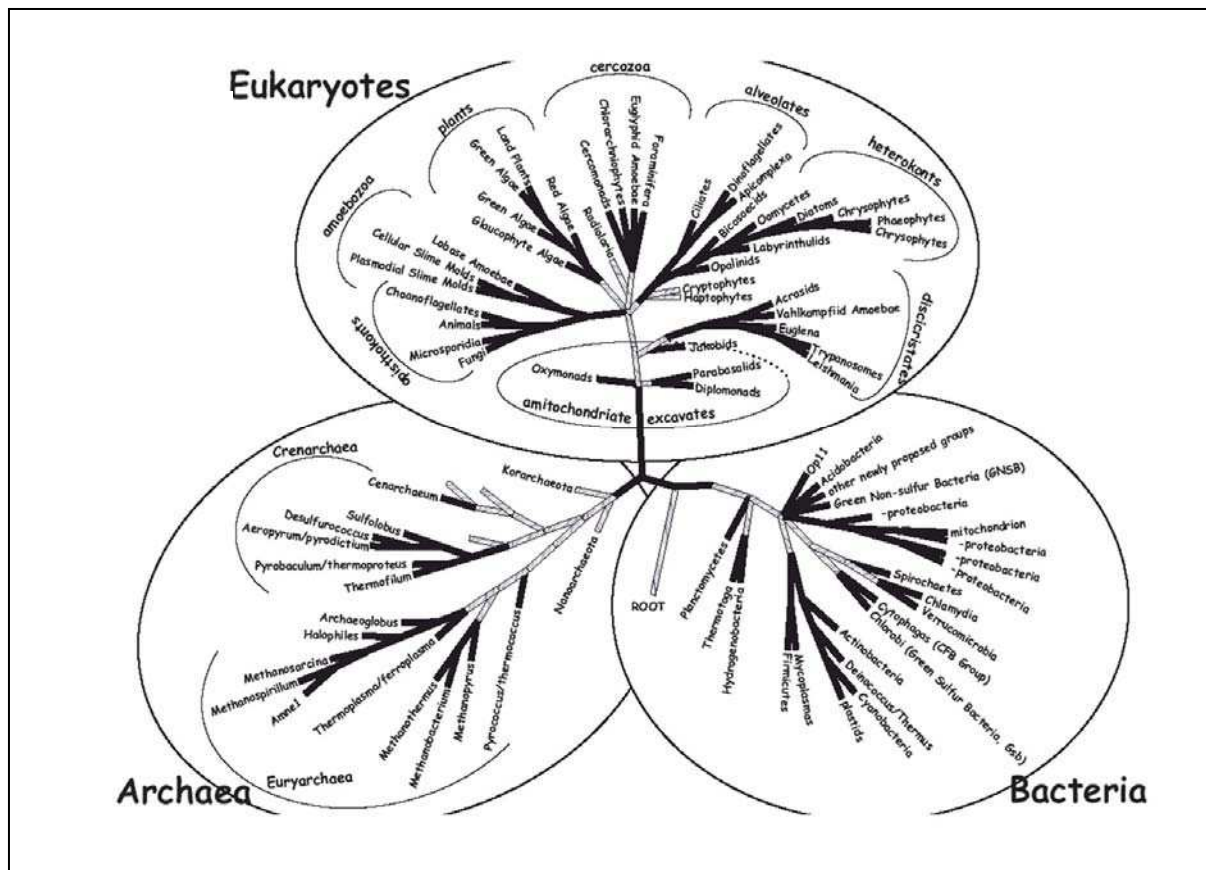


Figure 4: Représentation de l'arbre de la vie, établi par le regroupement d'études moléculaires et structurales ; extrait d'après Baldauf *et al.*, 2004 in Assembling the Tree of Life.

L'organisation des champignons en phyla est actuellement revisitée sur des bases phylogénétiques. Cinq phyla sont aujourd'hui acceptés selon une analyse phylogénétique reposant sur 6 gènes et plus de 200 organismes (James *et al.*, 2006). Si la monophylie des Ascomycota, Basidiomycota et des Glomeromycota est confirmée, les Chytridiomycota et Zygomycota sont polyphylétiques (James *et al.*, 2006 ; Bar Hen *et al.*, 2008).

Des photographies illustrant un organisme pour chacun des 5 phyla sont présentées en Figure 5.

**Basidiomycota** : *Agaricus sp.* (d'après [www.mycoweb.org](http://www.mycoweb.org))



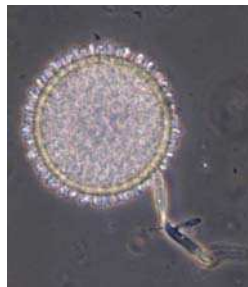
**Ascomycota** : *Galliella sp.* (d'après <http://www.clarku.edu>)



**Zygomycota** : *Pilobolus sp.* (d'après [www.mycology.com](http://www.mycology.com))



**Glomeromycota** : *Glomus sp.* (d'après <http://www.tari.gov.tw>)



**Chytridiomycota** : *Rhizophidium sp.* (d'après <http://www.mycology.com>)

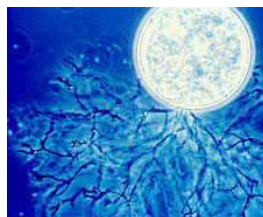


Figure 5: Photographies de champignons des 5 phyla.

## 1-5-2 Caractéristiques des organismes des 5 phyla fongiques

### 1-5-2-1 *Chytridiomycota*

Les organismes du phylum Chytridiomycota sont les seuls champignons à posséder des spores uniflagellées (zoospores) (Jennings & Lysek, 1996). C'est la lignée évolutive la plus ancienne des champignons (James *et al.*, 2000 ; James *et al.*, 2006) qui constitue un clade polyphylétique (James *et al.*, 2006 ; Bar-Hen *et al.*, 2008). Les positions basales de ce phylum n'ont jamais pu être clairement établies et supportées de façon robuste (James *et al.*, 2000 ; James *et al.*, 2006). La présence de spores flagellées semble restreindre ces organismes aux milieux aquatiques et dans les sols humides (James *et al.*, 2000). Les organismes de ce phylum sont souvent microscopiques mais peuvent aussi produire un mycélium. La plupart des Chytrides sont saprotrophes, aérobies ou anaérobies ; ils sont capables de dégrader un grand nombre de substrats (Powell, 1993 ; Shearer *et al.*, 2007). On recense également de nombreux pathogènes ; l'un d'eux a récemment été responsable d'une impressionnante mortalité de batraciens aux Etats Unis (*Batrachochytrium dendrobatidis* ; Longcore *et al.*, 1999). Des parasites d'algues ont également été reportés en milieu marin (*Chytridium polysiphoniae* ; Küpper *et al.*, 2006). Environ 1000 espèces ont été décrites au sein de ce phylum, ce qui correspond à environ 1% des espèces décrites de champignons (Taylor *et al.*, 2004).

### 1-5-2-2 *Zygomycota*

Les Zygomycota constituent un groupe ancien des champignons ayant divergé après les Chytridiomycota (James *et al.*, 2006 ; Bar-Hen *et al.*, 2008). Ce groupe formé d'organismes microscopiques hétérogènes est polyphylétique (James *et al.*, 2006 ; Tanabe *et al.*, 2000 ; Bar-Hen *et al.*, 2008). Différents modes de vie sont retrouvés, les plus communs étant le saprophytisme et le parasitisme (d'insectes principalement). Plus de 600 espèces ont été décrites à ce jour, ce qui représente moins de 1% des champignons décrits (Hawksworth 1991 ; Hawksworth & Rossman, 1997 ; Hawksworth, 2001).

### 1-5-2-3 *Glomeromycota*

Historiquement, les organismes de ce phylum étaient placés au sein des Zygomycota, dans l'ordre des Glomérales, un groupe qui regroupait les champignons mycorhiziens à arbuscules (Morton & Benny, 1990). Une analyse phylogénétique du gène codant l'ARNr 18S

a démontré la monophylie de l'ensemble des champignons mycorhiziens à arbuscules ce qui a permis d'ériger un nouveau phylum: les Glomeromycota (Schüßer *et al.*, 2001). A ce jour, 160 espèces ont été décrites en adoptant le 'traditionnel' concept d'espèce morphologique principalement basé sur la morphologie de la paroi des spores. Néanmoins, de nombreux phylotypes ont été découverts par approches moléculaires (Helgason *et al.*, 1998 ; Vandenkoornhuyse *et al.*, 2002b ; Opik *et al.*, 2008). Toutes les espèces de ce phylum sont des organismes symbiotiques biotrophes stricts de plantes (formant des mycorhizes à arbuscules).

#### ***1-5-2-4 Dicaryotes (Ascomycota et Basidiomycota)***

Les phyla Ascomycota et Basidiomycota forment le groupe des Dicaryotes et représentent la majorité des espèces de champignons décrites, en l'occurrence 67000 espèces (Taylor *et al.*, 2004) sur les 100.000 recensées.

Les organismes du phylum Ascomycota comptent 45000 espèces décrites à ce jour (Hawksworth 1991 ; Hawksworth, 2001 ; Taylor *et al.*, 2004) et constituent la quasi totalité des champignons capables de former des associations lichéniques. Des modes de vie saprophytes et parasites sont également largement répandus. On retrouve également chez ces organismes les champignons utilisés en agroalimentaire (*Saccharomyces cerevisiae*) ou en pharmacologie (*Penicillium chrysogenum*).

Les organismes du phylum Basidiomycota regroupent 22000 espèces décrites (Taylor *et al.*, 2004). Leur mode de vie est principalement saprophyte : ce sont d'ailleurs les organismes fongiques ayant les capacités de dégradation de matériels ligno-cellulolytique les plus élaborées (par exemple, Hibbet & Donoghue, 2001). On retrouve également des organismes symbiotiques de plantes ou parasites d'animaux.

## 1-6 Histoire évolutive des champignons

Les champignons représentent un groupe d'eucaryotes anciens qui est donc constitué d'organismes très diversifiés (Hawksworth, 2001). Comme nous l'avons vu dans le chapitre précédent, les champignons les plus primitifs appartenant au phylum des Chytridiomycota possèdent un caractère ancestral ayant été perdu au cours de la diversification des organismes du Royaume des champignons: un flagelle postérieur présent au niveau des zoospores, qui représenterait un avantage pour la dissémination des spores en milieu aquatique. Ces spores flagellées auraient disparu lors de la colonisation des écosystèmes terrestres lors de l'évolution (James *et al.*, 2000). Cependant, comme nous l'avons signalé, les relations phylogénétiques entre les organismes à la base des phylogénies ne sont pas clairement établies et ouvrent la voie à de nombreuses hypothèses sur l'évolution des champignons (James *et al.*, 2006, Bar-Hen *et al.*, 2008). Nous allons aborder dans ce chapitre les connaissances sur la radiation des groupes majeurs de champignons, en les mettant en rapport avec l'histoire évolutive d'autres groupes d'eucaryotes, plantes et animaux principalement.

Afin de reconstruire l'histoire évolutive des champignons, des calculs d'horloges moléculaires ont été établis. Le principe est le suivant : les gènes codants vont subir des mutations au cours du temps, et la quantité de mutations accumulées est proportionnelle au temps séparant l'organisme étudié de l'ancêtre commun. Afin de connaître le tempo évolutif, c'est à dire la fréquence de ce taux de mutation par unité de temps, des calibrations multiples doivent être réalisées par l'utilisation de différentes données fossiles. Ainsi, un nombre de mutations donné correspondra à une distance temporelle. Ces horloges moléculaires sont sensibles au marqueur moléculaire étudié, et les marqueurs à taux d'évolution lents seront privilégiés (c'est le cas des gènes codant les ARNr18S par exemple) pour situer un événement évolutif ancien. Pour calibrer l'horloge, l'idéal est d'obtenir une datation aussi précise que possible d'organismes fossiles, par l'usage d'analyses isotopiques. (Berbee & Taylor, 1993, Redecker *et al.*, 2000).



### 1-6-1 Données fossiles

Il est couramment proposé que les champignons, et plus particulièrement les mycorhizes à arbuscules (phylum Glomeromycota), ont facilité la colonisation du milieu aérien par les plantes (Simon *et al.*, 1993). Des fossiles de mycorhizes à arbuscules ont été retrouvés au Dévonien inférieur, (au niveau d'une roche sédimentaire, la « Rhynie Chert »), datant de 400 millions d'années (tableau 1). L'observation microscopique de ces fossiles végétaux primitifs a montré des interactions mycorhiziennes à arbuscules telles qu'on les connaît aujourd'hui. Le saut macroévolutif conféré par le mutualisme aurait permis des adaptations multiples requises pour la colonisation du milieu aérien il y a 400 millions d'années (Selosse & Le Tacon, 1998), accompagnant ensuite la radiation des plantes dans ces milieux (Remy *et al.*, 1994). En 2000, de nouveaux fossiles de spores et d'hyphes de champignons MA ont été retrouvés datant de l'Ordovicien (460 Ma) (tableau 1), période où les seules plantes présentes étaient des bryophytes ; les champignons MA auraient donc été présents avant la radiation des plantes vasculaires (Redecker *et al.*, 2000). Aujourd'hui, il est intéressant de noter que les hépatiques sont mycorhizées (Opik *et al.*, 2008). Cette mycorhization très ancienne tend à expliquer aujourd'hui pourquoi la plupart des plantes sont en symbiose avec ces champignons MA.

Des fossiles d'Ascomycota, et plus précisément de Sordariomycetes ont également été retrouvés dans les « Rhynie Cherts » (Taylor, 1999) (tableau 1).

Il est à noter que les premiers calculs d'horloges moléculaires ne se basaient que sur le gène codant l'ARNr18S, et qu'aucun gène codant des protéines n'avait été utilisé. Heckman *et al.* (2001) ont utilisé des séquences d'acides aminés de protéines disponibles dans les bases de données, afin d'estimer la datation hypothétique de la radiation entre les principaux groupes de champignons, ainsi que le temps de divergence entre champignons et animaux, par des calibrations multiples. Ces analyses qui sont en conflit avec d'autres travaux suggèrent la divergence des Glomeromycota postérieure à celle des Chytridiomycota, mais antérieure à la séparation Ascomycota/Basidiomycota, il y a entre 1400 et 1200 millions d'années respectivement (tableau 1). Cependant, ces résultats sont fortement dépendants des points de calibration choisis, comme par exemple la divergence entre champignons et animaux, qui pour certains auteurs se serait déroulé il y a 1576 Millions d'années (Wang *et al.*, 1999) ou 841 Millions d'années pour d'autres (Heckman *et al.*, 2001). Un récapitulatif des temps de divergence séparant les groupes majeurs de champignons est proposé dans le tableau 1 (d'après Taylor *et al.*, 2004). Il regroupe les datations d'émergence hypothétiques des

différents phyla et comprend les datations réalisées à partir des travaux de Vandenkoornhuyse (2002a), base de nos hypothèses de travail (horloge moléculaire proposée en Figure 2).

Groupes comparés (le groupe de référence entre parenthèses)	estimation rRNA (a)	estimation rRNA (b)	estimation rRNA (c)	estimation 119 protéines (d)	Age du fossile le plus ancien dans le groupe de référence
(Chytridiomycota) vs Zygomycota + Ascomycota + Basidiomycota	~ 550	~ 660	~1050 ± 100	~ 1458 ± 70	~ 400 (e)
Chytridiomycota + (Zygomycota) vs Ascomycota + Basidiomycota	~ 490	~ 590	~880 ±80	~ 1107 ± 56	~ 460 (f)
(Ascomycota) vs Basidiomycota	~ 390	~ 560	~700 ±70	~ 1208 ± 108	~ 400 (g)

Tableau 1: Temps de divergence, en millions d'années, entre les principaux groupes fongiques (modifié d'après Taylor *et al.*, 2004 in Assembling the Tree of Life)

a- Horloge moléculaire calibrée en utilisant des données fossiles (Berbee & Taylor, 1993).

b- Horloge moléculaire calibrée en utilisant des données fossiles et un temps de divergence entre animaux et champignons estimée à 965 Millions d'années (Berbee & Taylor, 2006).

c- Horloge moléculaire calculée sur la base des résultats de Vandenkoornhuyse *et al.* (2002a).

d- Horloge moléculaire calibrée en utilisant un temps de divergence entre plantes, animaux et champignons, estimée à 1576 Millions d'années (Heckman *et al.*, 2001).

e- Fossiles de Chytridiomycota isolés de Rhynie Chert (Taylor & Taylor, 1997).

g- Spores et hyphes de glomales (Glomeromycota) de l'Ordovicien (Redecker *et al.*, 2000).

g- Pyrenomycetes (Ascomycota) fossilisés, isolés de Rhynie Chert (Taylor *et al.*, 1999).

## 1-6-2 Relations évolutives entre les groupes de champignons

La classification des champignons et l'histoire évolutive des organismes de ce Règne sont actuellement en profondes révisions. L'application de techniques moléculaires pour résoudre les relations taxonomiques entre ces organismes a permis de mieux comprendre ces relations (White *et al.*, 1990, Hibbet *et al.*, 2007). A l'heure actuelle, la taxonomie communément acceptée par la communauté mycologiste est celle proposée par Hibbett (Hibbet *et al.*, 2007). Les premières études moléculaires sur l'histoire évolutive des

champignons ont dans un premier temps démontré la monophylie de ce Règne, sur la base de l'analyse du gène codant l'ARNr18S (Bowman *et al.*, 1992). Ces analyses étaient renforcées par des observations morphologiques et biochimiques : tous partagent en effet une même voie de biosynthèse de la lysine, et possèdent une paroi de nature chitineuse (Bruns *et al.*, 1992). Afin d'établir plus précisément les relations évolutives entre les différents groupes de champignons, une étude a été réalisée par James *et al.* (2006), utilisant 6 marqueurs moléculaires les plus utilisés en mycologie: les gènes codant les ARNr 18S, ARNr 5.8S, ARNr 28S, le gène codant le facteur d'élongation EF1- $\alpha$ , les gènes codant les sous unités de l'ARN polymérase II (RPB1, RPB2). Cette étude a permis de renforcer des hypothèses évolutives, notamment pour les groupes ayant divergé le plus tôt (James *et al.*, 2006). L'hypothèse communément acceptée est que les champignons possédant des cellules flagellées (Chytridiomycota) représentent les organismes les plus anciens d'un point de vue phylogénétique, et constituent un groupe frère des organismes non flagellés. La colonisation de ces milieux aériens aurait eu pour conséquence une croissance de ces organismes sous une forme filamenteuse, ainsi qu'un relargage aérien des spores (James *et al.*, 2006). L'analyse multigène de James *et al.* (2006) a de plus démontré à nouveau, la monophylie des Ascomycota, Basidiomycota et Glomeromycota. Ce travail a également permis de regrouper les Ascomycota et Basidiomycota en un sous-Règne, appelé les Dikarya, en s'appuyant sur la présence d'hyphes dicaryotes au cours du cycle sexuel de ces organismes. D'après cette étude, le groupe des microsporidies, constitués d'organismes parasites, pourrait constituer un nouveau phylum (Hibbett *et al.*, 2007).

L'analyse de la phylogénie résultant de l'étude multigènes de James *et al.* (2006) a permis d'émettre des hypothèses sur les événements de la transition « milieu marin - milieu terrestre » ; il apparaît à la lecture de la phylogénie, et notamment par le fait que Chytridiomycota et Zygomycota représentent des groupes paraphylétiques, que cette transition ne se soit pas faite par un saut macroévolutif ayant conduit à la perte du flagelle, mais que cette transition ait nécessité 6 pertes successives de flagelles durant l'évolution des champignons. Cette hypothèse est contradictoire aux observations et interprétations de Liu *et al.* (2006) se basant sur des phylogénies utilisant les gènes codant pour les sous unités de l'ARN polymérase II (RPB1 et RPB2), dont la phylogénie est proposée à la figure 6. Ce travail montre en effet une différenciation de 2 groupes : l'un contenant les organismes porteurs de flagelles (Chytridiomycota), l'autre tous les organismes ayant

perdu le flagelle au cours de l'évolution (Ascomycota, Basidiomycota, Glomeromycota et Zygomycota). Cette séparation, statistiquement soutenue, suggère qu'il n'y aurait eu qu'un seul événement évolutif ayant mené à la perte du flagelle. En conclusion, les événements ayant mené à la perte du flagelle chez les champignons, et les événements ayant mené à la colonisation des écosystèmes terrestres, restent incertains ; mieux connaître les branches profondes du Règne fongique reste une condition nécessaire pour évaluer avec précision ces scénarii évolutifs.

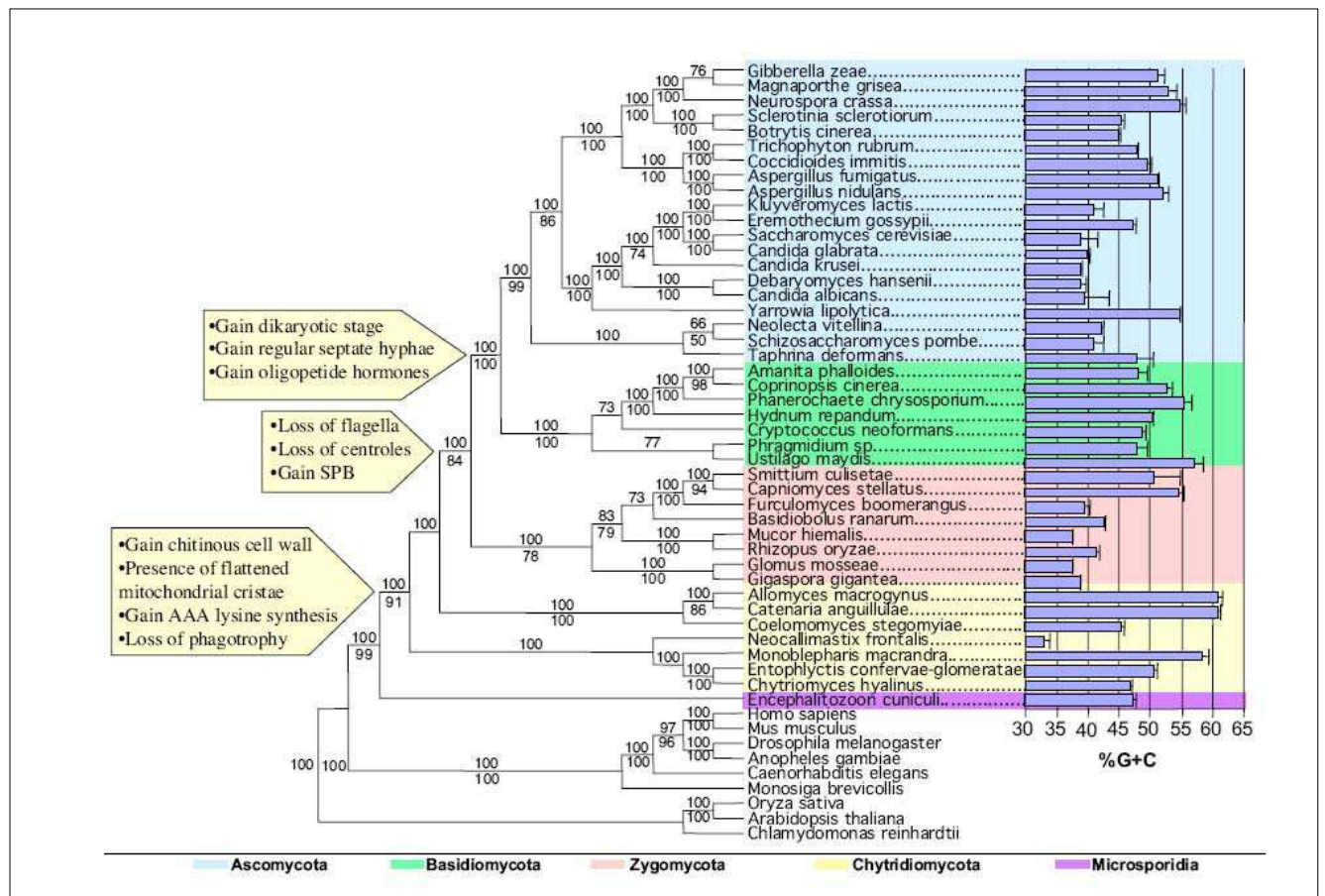


Figure 6: Phylogénies des champignons extraites de Liu *et al.*, 2006 ; *BMC Evol. Biol* 6: 74-86, s'appuyant sur les séquences RBP1 et RBP2. Les hypothèses de scénarii évolutifs sont émises sur cette phylogénie, et notamment l'hypothèse impliquant un seul événement évolutif ayant mené la perte de flagelle.

L'étude de James *et al.* (2006) a de plus permis de constater que les organismes les plus anciens d'un point de vue évolutif partageaient un mode de vie parasite (*Rozella allomycis*, Chytridiomycota), avec une croissance de type amiboïde, à l'opposé du saprophytisme et d'une croissance filamenteuse pour les organismes ayant divergé plus récemment. Les modes de vie pathogène de nombreux organismes des phyla Zygomycota et Chytridiomycota rendent difficiles leur placement au sein de la phylogénie des champignons ; les séquences de ces organismes pathogènes, à vitesse d'évolution plus élevées, vont entraîner des biais dans la reconstruction des phylogénies, notamment par le phénomène d'attraction des longues branches (Philippe & Laurent, 1998 ; Redecker, 2002 ; James *et al.*, 2006). Résoudre les relations entre les groupes à la base de la phylogénie des champignons permettrait de mieux comprendre les événements ayant mené à l'émergence et à la diversification de ces derniers.

## **1-7 Ecologie marine**

On estime à 1.5 millions le nombre d'espèces de champignons (Hawksworth, 2001). Cette estimation est principalement calculée sur la base du ratio espèces fongiques / espèces végétales dans différentes régions écologiques bien étudiées (Schmit & Mueller., 2007 ; Hawksworth & Rossman, 1997; Hawksworth, 2001). Les prédictions les plus conservatives et les plus basses considèrent un minimum de 700.000 espèces fongiques (Schmit & Mueller, 2007). Sur cette base nous pouvons supposer que seuls 1/7<sup>ème</sup> des champignons présents sur Terre (14%) ont été décrits. Ce biais de nos connaissances est à la fois dû à la difficulté de cultiver ces organismes en conditions de laboratoires, qui est nécessaire pour appliquer le 'traditionnel' concept morphologique d'espèce, mais également au manque de données concernant certains écosystèmes, comme les forêts tropicales par exemple (Hawksworth & Rossman, 1997).

### **1-7-1 Champignons marins**

A l'heure actuelle, la connaissance de la diversité de la microflore fongique océanique est embryonnaire. Cette constatation est vraisemblablement liée à l'idée très répandue que les champignons sont inféodés au milieu terrestre par leur histoire évolutive. Depuis une vingtaine d'années pourtant, des études sur les champignons marins ont visé à isoler des

molécules potentiellement utiles issues de ces organismes (Miller, 2000) : la connaissance du domaine fongique marin offrant un réservoir d'applications industrielles, pharmaceutiques notamment (Munn, 2004). De nombreux habitats marins sont colonisés par ces organismes (mangroves, littoral, océan...) (Hyde *et al.*, 1998). Kohlmeyer et Kohlmeyer (1979) ont défini dès 1979 que certains champignons semblent être inféodés exclusivement au milieu marin ou estuarien.

Historiquement, la première description d'un champignon exclusivement marin remonte à 1869 (*Phaeosphaeria typharum*) (Desmazières, 1849), mais leur étude a véritablement débuté en 1944 avec une publication de Barghoorn et Linder décrivant des champignons isolés de mangroves se développant sur le bois en décomposition (Barghoorn & Linder, 1944).

D'un point de vue écologique, ces champignons marins ont adopté trois types de modes de vie: (1) la symbiose, avec des algues par exemple (*Turgidosculum ulvae* forme une symbiose avec *Ulva vexata*); (2) le parasitisme, d'algues notamment (*Thallassoascus tregoubovii*), ou d'animaux (mollusques,...); (3) le saprophytisme (*Ophiobolus australiensis*) (Kohlmeyer & Kohlmeyer 1979). Les espèces saprophytes ont révélé majoritairement une forte activité cellulolytique (Pointing *et al.*, 1992); leur capacité à dégrader le bois couplée à la capacité à se développer à partir de matière organique animale morte leur confère un rôle prédominant dans la décomposition des matières organiques en milieu océanique (Mozouras, 1986). On estime à ce jour que moins de 1500 espèces ont été isolées des océans (Munn, 2004). Parmi le faible nombre de travaux publiés, la majorité d'entre elles se sont concentrées sur la zone littorale (Kohlmeyer & Kohlmeyer 1979 ; Hawksworth, 2001 ; Shearer *et al.*, 2007). Les études portant sur les champignons du domaine océanique ont quant à elles révélé l'existence d'Eumycota filamenteux (phylum Ascomycota principalement), et de levures (phylum Basidiomycota majoritaire) qui semblent représenter le groupe plus répandu (Kohlmeyer & Kohlmeyer 1979). Les organismes du phylum Zygomycota seraient quant à eux absents des écosystèmes marins (Shearer *et al.*, 2007). Il apparaît aux vues des travaux réalisés jusqu'alors que de nombreux substrats et milieux marins sont encore à explorer en termes de diversité fongique. L'inventaire de la flore fongique marine devrait donc permettre de découvrir de nouveaux taxons et la diversité fongique, sans doute sous-estimée pourrait être ré-évaluée.

En 1979, Kohlmeyer et Kohlmeyer avaient observé que des champignons pouvaient se développer dans des environnements marins profonds et s'adapter ainsi à de fortes

pressions mais très peu de données relatives aux champignons des grands fonds sont disponibles à l'heure actuelle (Hyde *et al.*, 1998). La plupart des publications du domaine océanique profond se sont intéressées à l'étude de levures (Nagahama *et al.*, 2001) et très peu d'espèces de champignons filamenteux ont été répertoriées. Des levures et des champignons filamenteux barophiles ont été isolées de sédiments marins profonds (Raghukumar & Raghukumar, 1988 ; Damare *et al.*, 2006). Très récemment, un grand nombre de phylotypes fongiques (des phyla Ascomycota et Basidiomycota) ont été détectés dans l'océan profond, au niveau d'un bassin anoxique hypersalin de la Méditerranée (Edgcomb *et al.*, 2009). Il a été proposé que la fréquence d'isolement des levures diminuait avec la profondeur de l'échantillonnage ; les levures du phylum Ascomycota seraient majoritaires dans les eaux peu profondes, tandis que les levures du phylum Basidiomycota seraient majoritaires dans les grandes profondeurs (Munn, 2004).

### **1-7-2 Champignons hydrothermaux**

L'étude des champignons en milieu marin étant assez limitée, les champignons en milieu hydrothermal ont en conséquence été très peu étudiés. La découverte de ces organismes dans ces milieux est très récente, car ce n'est qu'en 2002 que plusieurs études s'intéressant aux micro-eucaryotes ont détecté des séquences de natures fongiques en écosystème hydrothermal, au niveau des rides Médio-atlantique et Est-Pacifique (López-García *et al.*, 2003 ; Edgcomb *et al.*, 2002). En dépit du fait que ces études pionnières aient permis de détecter la présence de champignons dans ces environnements, aucune indication taxonomique ou écologique supplémentaires n'avaient pu être apportées. Ce n'est qu'en 2007 que des études moléculaires portant exclusivement sur l'analyse de la diversité des champignons en milieu océanique profond, ont été entreprises (Bass *et al.*, 2007). Ces études ont révélé la présence d'organismes fongiques au niveau de la ride Medio-Atlantique (sites Lucky Strike et Rainbow), classifiés dans le phylum Basidiomycota (classe des Ustilagomycetes) (Bass *et al.*, 2007). Cependant, ces auteurs ont conclu à une faible diversité de champignons dans le domaine profond, qui serait dominé par des organismes levuriformes, la forme levure offrant une meilleure adaptation aux fortes pressions hydrostatiques de ces milieux (Bass *et al.*, 2007). Une étude menée la même année semble appuyer cette constatation, où des levures noires de l'ordre des Chaetothyriales (phylum Ascomycota) ont été retrouvées au niveau de moules isolées de sites hydrothermaux du Bassin de Fidji (Van Dover *et al.*, 2007). Cette étude qui s'appuie sur des observations microscopiques, prouve le

caractère pathogène de tels organismes, pouvant provoquer des mortalités de masse dans les lits de moules (Van Dover *et al.*, 2007) et ainsi participer activement au fonctionnement de l'écosystème. Il s'agit ainsi de la première étude ayant une portée écologique sur les champignons d'écosystèmes hydrothermaux marins. La première collection de champignons (filamenteux et levuriformes) isolés d'écosystèmes hydrothermaux est publiée depuis peu (Burgaud *et al.*, 2009).

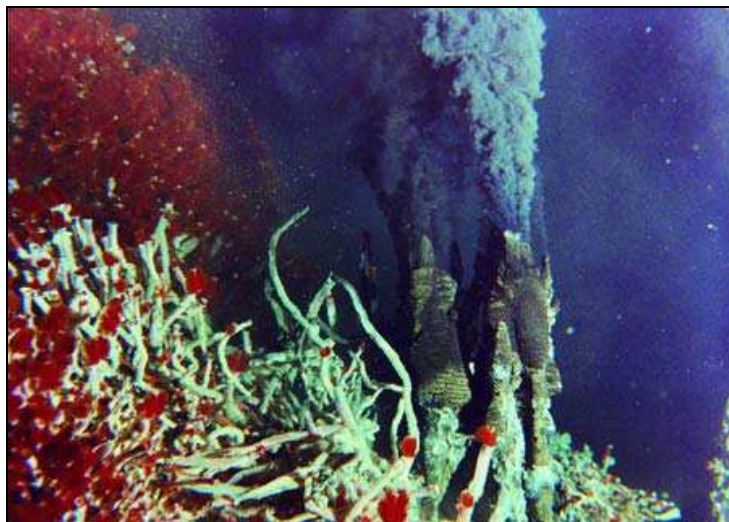
*Les objectifs de cette thèse étaient donc doubles : apporter des connaissances sur la diversité des champignons au sein des écosystèmes hydrothermaux, puis prédire les fonctions portées par ces organismes dans ces milieux. Comme nous le verrons dans le chapitre suivant, l'écosystème hydrothermal est considéré par certains auteurs comme le milieu berceau de la vie. Plus largement et moins spéculativement, nous savons aujourd'hui que la température moyenne des océans était élevée (70°C) il y a 3,5 milliards d'années et a progressivement baissée jusqu'au Précambrien (Robert & Chaussidon, 2006). Ainsi, l'étude des organismes fongiques dans ces écosystèmes hydrothermaux pourrait apporter des informations sur les relations évolutives des organismes ayant divergé le plus tôt au sein de l'arbre de la vie fongique.*





## **Chapitre I-2 : Présentation des écosystèmes hydrothermaux marins profonds**

---



## **I-2 Présentation de l'écosystème hydrothermal marin profond**

### **2-1 : Historique et contexte**

L'océan profond est caractérisé par une profondeur de plus de 1000 mètres et représente 75% de la surface totale des océans : il représente ainsi l'un des biotopes les plus importants sur Terre (Prieur *et al.*, 1995 ; Kaiser *et al.*, 2005). C'est un environnement aphotique, qui contrairement à la plupart des écosystèmes de la Terre ne repose pas sur la photosynthèse pour la production primaire d'énergie. Les organismes qui s'y développent vont dépendre essentiellement de la matière organique sédimentant des couches supérieures de l'océan vers les couches inférieures (Jannash & Taylor, 1984). D'importantes communautés animales et bactériennes se développant sur les immenses carcasses de baleines sont ainsi régulièrement retrouvées. Cependant, étant donné le faible apport de matière organique dans ces systèmes abyssaux, la productivité est très faible et on considère souvent les abysses comme de véritables déserts, avec une biomasse inférieure à  $1\text{g/m}^2$  (Desbruyeres, 2000). Ces écosystèmes sont influencés par 3 paramètres majeurs : une forte pression hydrostatique, une faible température ( $2^{\circ}\text{C}$  de moyenne) et une faible disponibilité en nutriments (Prieur, 1997).

En 1977 cependant, une équipe de géologues américains va faire une découverte qui va révolutionner la connaissance des écosystèmes marins (Corliss *et al.*, 1979). A 2500 mètres de profondeur, en étudiant le plancher océanique sur la ride Est Pacifique au niveau de la dorsale des Galapagos (Galapagos Spreading Center), les scientifiques à bord du submersible l'Alvin vont découvrir une faune exubérante composée d'organismes non identifiés se développant autour de sources tièdes ( $15^{\circ}\text{C}$ ). Cette découverte spectaculaire a ainsi mis en évidence l'existence d'une faune hydrothermale ouvrant la voie à l'étude de ces écosystèmes. Rapidement, de nombreuses expéditions vont être mises en place, explorant progressivement différentes régions du globe: l'exploration a dans un premier temps concernée le Pacifique oriental puis occidental, pour ensuite étudier la dorsale médio-atlantique (1985), puis indienne (Minic *et al.*, 2006).

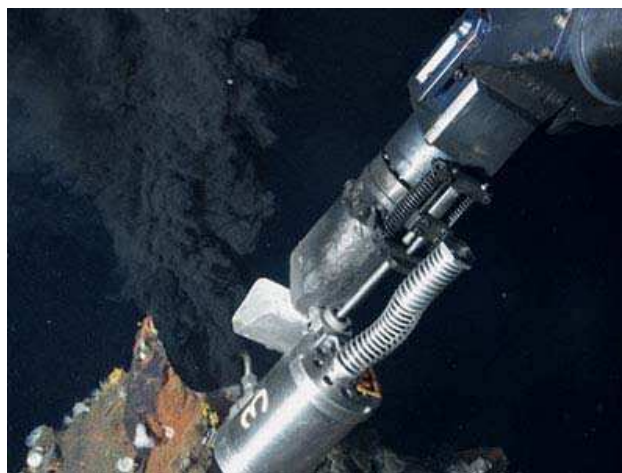
Cette découverte a donc permis de révéler un écosystème totalement inconnu. Cependant, bien que l'excitation liée à cette découverte fût grande dans la communauté scientifique, de nombreux problèmes inhérents notamment à la profondeur et aux conditions chimiques agressives du milieu ont dû être surpassés. De nouveaux équipements ont ainsi dû être mis au point afin d'atteindre le lieu d'étude et d'effectuer les prélèvements dans les meilleures conditions. On peut ainsi noter par exemple la mise au point des systèmes de prélèvements de fluide résistants aux fortes températures (Figure 7), ainsi que tout un système ingénieux permettant un prélèvement stérile des échantillons du lieu de prélèvement vers la surface (Jørgensen & Boetius, 2007). Des systèmes de submersibles radioguidés ont également été mis en place (ROV : Remote Operated Vehicle ; exemple du Nautile, Figure 7). L'exploration de ces milieux nécessite de plus une interconnexion étroite entre différents domaines scientifiques : une étude bathymétrique doit dans un premier temps être réalisée, suivie par une étude sismologique afin d'obtenir des indices de volcanisme. Ce n'est qu'une fois ces conditions réunies que biologistes, chimistes et géologues peuvent explorer les écosystèmes hydrothermaux.



Nautile, d'après <http://arn.gui.free.fr>



Exemple de préleveur de crevettes, d'après <http://www.insu.cnrs.fr>



Exemple de préleveur de fluide, d'après [Ifremer.fr](http://Ifremer.fr)

Figure 7: Exemples de matériels utilisés pour les prélèvements en écosystèmes hydrothermaux marins profonds.

## 2-2 : Principe de l'hydrothermalisme

L'hydrothermalisme océanique prend son origine autour des activités volcaniques et tectoniques se déroulant au fond des mers au niveau des dorsales océaniques. Ces dorsales longues de plus de 60.000 km et de profondeur moyenne de 2600 m, constituent le plus grand système volcanique de la Terre (Minic *et al.*, 2006). Elles sont le lieu de l'expansion du plancher océanique et sont situées entre les plaques tectoniques ; la vitesse d'ouverture des plaques peut atteindre jusqu'à 20 cm par an (pour les dorsales à expansion rapide de la ride Est Pacifique : EPR) (Minic *et al.*, 2006).

### Exemple des fumeurs noirs de la Ride Medio-Atlantique (site Menez Gwen)

Au niveau de ces dorsales, sous l'action tectonique de l'expansion lithosphérique, le plancher océanique est considérablement fissuré. L'eau de mer, froide et dense, va s'engouffrer dans ces fissures. Tout en s'enfonçant plus profondément sous la roche, l'eau de mer va être réchauffée par l'énergie géothermique provoquée par la remontée du magma sous l'effet de l'expansion océanique. La température de l'eau qui atteint alors plus de 300°C, va réagir avec la roche (basaltique dans la plupart des cas) provoquant des modifications chimiques de sa composition, variables selon le type de roche traversée. L'eau devient anoxique, le potassium, le calcium et les sulfates sont retirés; le fluide s'enrichit en particulier en sulfures mais aussi en cuivre, zinc et fer.

Sous l'action de la chaleur l'eau devient moins dense, et va ressurgir au niveau du plancher océanique. Cette résurgence va s'accompagner de modifications chimiques. Le contact du fluide hydrothermal chaud (350°C en moyenne), acide, anoxique, chargé en métaux, avec l'eau de mer froide (2°C) oxygénée, va provoquer la précipitation des métaux du fluide. Ces métaux se combinent avec les sulfures présents dans le milieu pour former des sulfures métalliques (sulfures de fer, appelés Pyrite), donnant un aspect sombre au fluide : on parle alors de « fumeurs noirs » qui seront le seul cas de mécanisme hydrothermal auquel nous nous intéresserons dans notre étude. L'accrétion de ces minéraux va permettre la formation d'édifices pouvant atteindre 15 mètres de hauteur, les cheminées hydrothermales, composées de sulfures polymétalliques (Jørgensen & Boetius, 2007). Un schéma explicatif est fourni dans la figure 8.

De tels systèmes sont très instables dans le temps compte tenu de la forte activité tectonique et volcanique. La durée de vie d'un site hydrothermal peut ainsi passer de quelques

décennies pour les systèmes à expansion rapide (type Ride Est Pacifique) à plusieurs milliers d'année pour les systèmes à expansion lente (type Ride Medio-Atlantique : MAR) (Tsurumi & Tunncliffe, 2001 ; Minic, 2006). Une autre source d'instabilité provient d'un gradient de mélange turbulent spatio-temporel entre le fluide hydrothermal et l'eau de mer. C'est au niveau de cette zone de mélange que va pouvoir se développer la vie hydrothermale en raison des forts gradients comme les gradients d'oxydo-reduction, de pH, de température (Gage & Tyler, 1991).

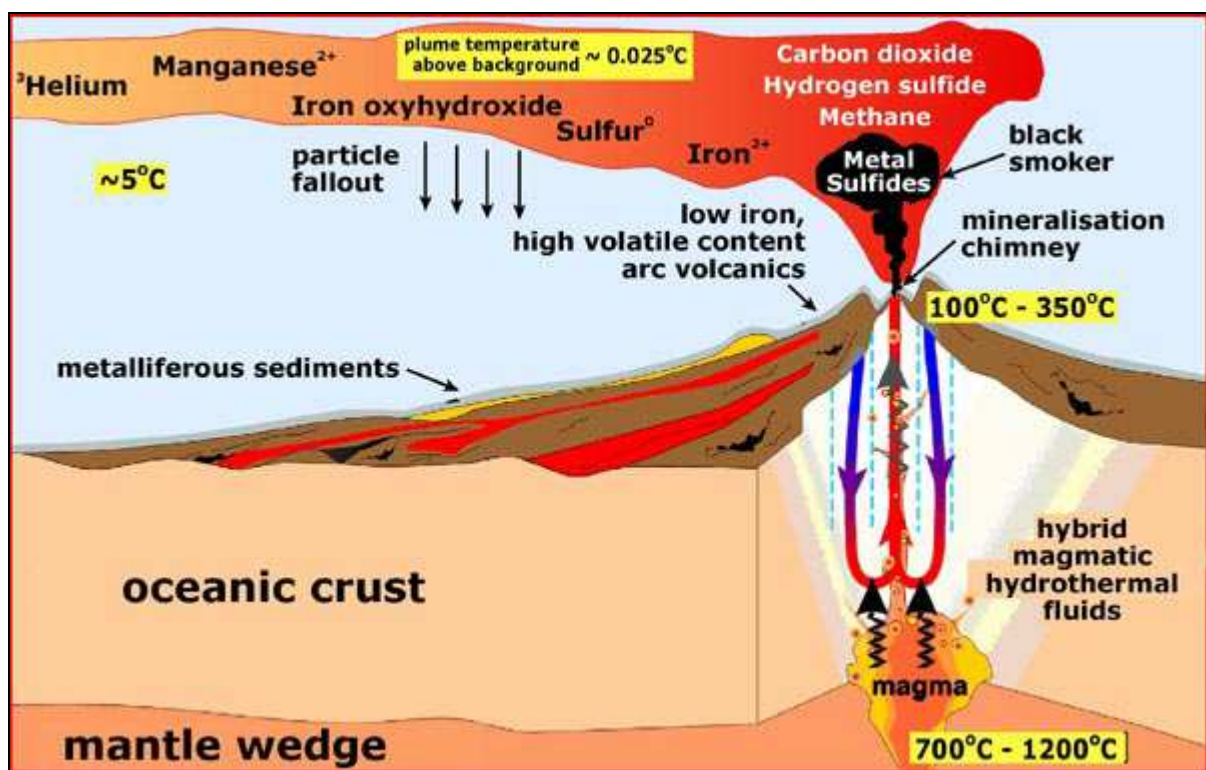


Figure 8: Schéma d'une section longitudinale d'un volcan sous marin hydrothermal et des principaux mécanismes impliqués dans l'hydrothermalisme.

D'après <http://oceanexplorer.noaa.gov>

## 2-3 : Localisation géographique

Les écosystèmes hydrothermaux se distribuent le long de la ride océanique. Compte tenu des difficultés d'études à de telles profondeurs, seule une infime partie a été découverte. Les zones les plus étudiées (Figure 9) correspondent à la ride Medio-Atlantique, à la ride Est-Pacifique, la triple jonction Juan de Fuca, ainsi que les bassins arrière arc du Sud Ouest Pacifique (Desbruyères *et al.*, 2000).

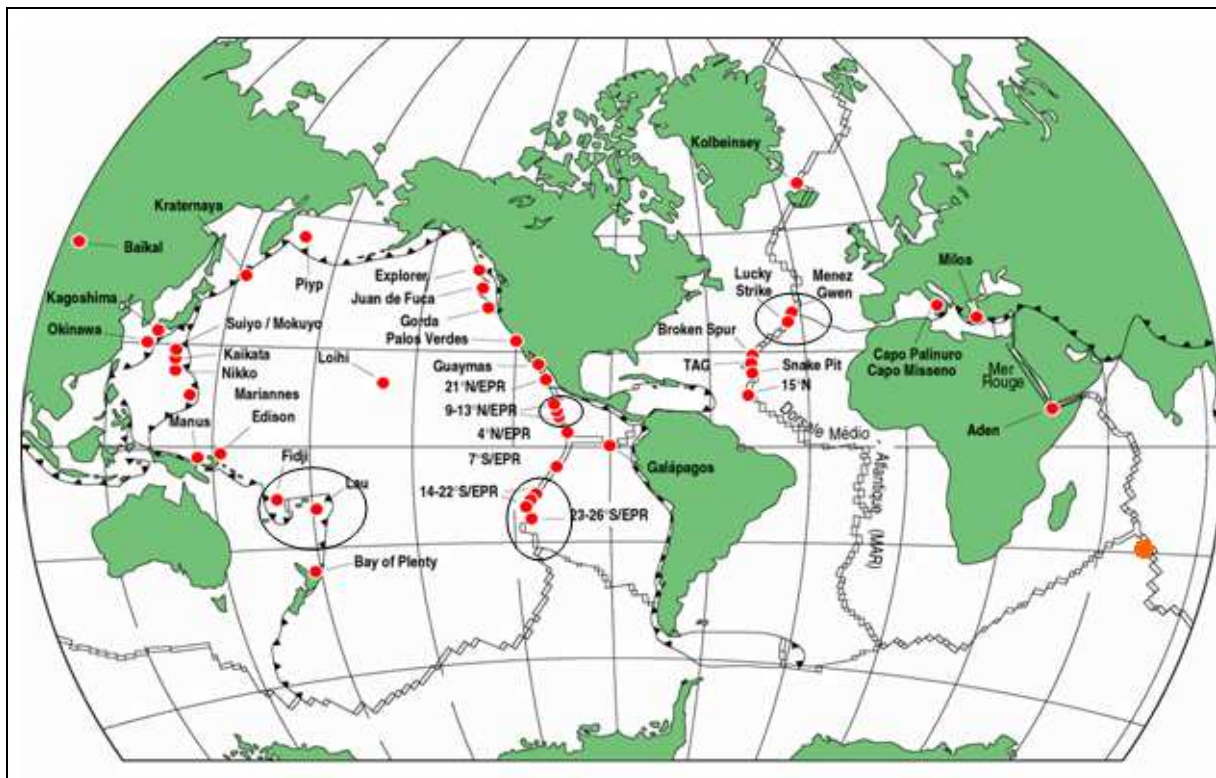


Figure 9: Localisations géographiques des sites hydrothermaux. Les principaux sites de la Dorsale Médio-Atlantique (MAR) et de la dorsale Est-Pacifique (EPR) sont représentés.

## 2-4 : Peuplements biologiques

Les écosystèmes hydrothermaux sont peuplés par une importante communauté biologique (pouvant atteindre jusqu'à  $50 \text{ Kg.m}^{-2}$ ; Prieur, 1997) très productive, distribuée autour des événements hydrothermaux. Ces communautés reposent sur une production primaire provenant de la chimioautotrophie bactérienne, entretenue par les composés réduits émis au niveau des fumeurs (Cavanaugh, 1983). Ces communautés sont donc étroitement liées à



l'émission du fluide réducteur et sont dépendantes des conditions physico-chimiques du milieu (Le Bris *et al.*, 2006).

La distribution de ces organismes va être grandement affectée par leur capacité à résister et à s'adapter aux conditions agressives du fluide émis ; on trouvera ainsi par exemple des espèces hyperthermophiles à la sortie du fluide, tandis que les espèces mésophiles se situeront plus loin du mélange fluide / eau de mer (Desbruyeres *et al.*, 2000). Malgré le fort taux d'endémisme des populations hydrothermales (environ 90% selon Minic *et al.*, 2006), des espèces dites pénétrantes sont également retrouvées, c'est à dire des espèces vivants dans le milieu abyssal mais pouvant accéder aux sites hydrothermaux (dans les parties moins toxiques) dans un but de prédation (Desbruyeres *et al.*, 2000).

Les communautés hydrothermales se répartissent de façon différente selon d'une part les types de sous système (Dorsale à expansion lente type MAR et dorsales à expansion rapide type EPR), et d'autre part selon la localisation géographique de ces sites (MAR et EPR par exemple) (Wolf, 2005). Il a par ailleurs été démontré que les organismes vivant autour des événements hydrothermaux des dorsales à expansion rapide privilégiaient le sulfure comme accepteur d'électrons, contre une utilisation préférentielle du méthane et de l'hydrogène pour les dorsales à expansion lente (Charlou *et al.*, 2002).

#### **2-4-1 Fonctionnement général**

Sans lumière et donc sans photosynthèse, les peuplements hydrothermaux se développent à partir de chimiosynthèse bactérienne, qui constitue ainsi la production primaire d'un écosystème complexe et très productif. Les bactéries (et archaea) vont tirer leur énergie de l'oxydation des composés réduits émis par le fluide, tels que les sulfures. Cette énergie servira à convertir le carbone minéral en carbone organique assimilé par l'organisme autotrophe, ou pourra servir de source de carbone à des organismes hétérotrophes. Les écosystèmes hydrothermaux caractérisés par de forts gradients physico-chimiques (oxygène dissous, pH, éléments minéraux), offrent ainsi un grand nombre d'habitats au sein d'un même système permettant la coexistence de nombreux organismes et microorganismes (Jørgensen & Boetius, 2007) (Tableau 2 et Figure 10). Il a d'ailleurs été prouvé que ce sont les conditions physico-chimiques du milieu qui vont organiser les communautés biologiques, qui se répartiront principalement selon les gradients d'oxygène dissous, de température et d'hydrogène sulfuré (H<sub>2</sub>S) (Luther III *et al.*, 2001).

## 2-4-2 Communautés biologiques

### 2-4-2-1 Bactéries

L'étude des microorganismes procaryotes a connu un essor grâce à l'avènement des techniques moléculaires ; les premières études de la flore microbienne utilisaient les techniques microbiologiques classiques telles que les observations microscopiques et les analyses biochimiques, qui nécessitaient un isolement préalable des organismes dans les conditions de laboratoire. Or, ces conditions sont difficiles à recréer artificiellement et la majorité des organismes sont réticents à ce type de culture (Jørgensen & Boetius, 2007). L'utilisation de techniques moléculaires s'est avérée indispensable pour mieux comprendre le fonctionnement biologique de ces milieux.

Il a été démontré qu'il existe au moins quatre types d'habitats hydrothermaux avec leur communauté microbienne associée (Figure 10): (1) les populations bactériennes libres associées aux fluides, se développant et se reproduisant à l'interface sous-sol / mer ; (2) les microorganismes libres regroupés en tapis se développant sur les strates de surface et qui sont exposées aux courants hydrothermaux ; (3) les associations endo ou exo symbiotiques de microorganismes avec la faune des événements ; (4) les microorganismes des panaches hydrothermaux (Jeanthon, 2000) (exemple des  $\epsilon$ -protéobactéries proposée dans la figure 10). L'abondance et la diversité des différents habitats et substrats présents dans le milieu permettent la coexistence d'un grand nombre de métabolismes bactériens différents (Tableau 2). La figure 10 représente les différents types de métabolisme et d'espèces d' $\epsilon$ -protéobactéries en fonction de leur distribution autour du fumeur.

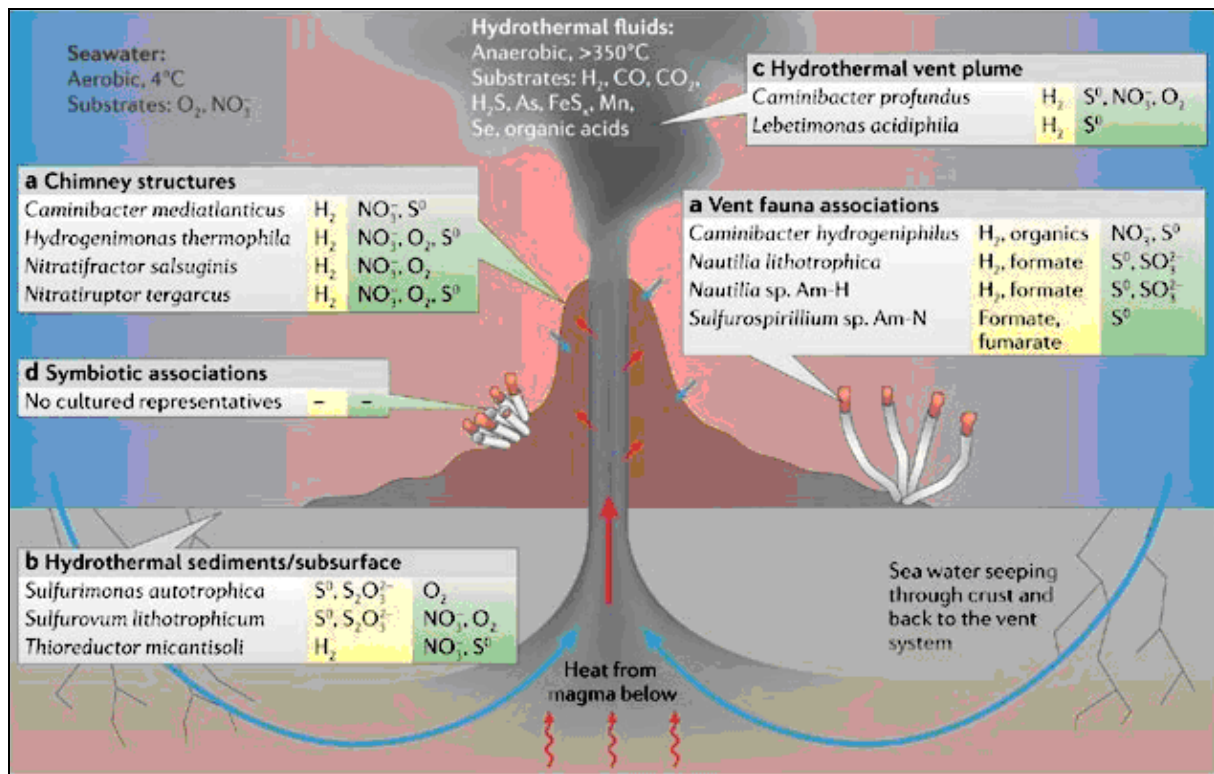


Figure 10: Différents types de métabolismes bactériens (ici les ε-protéobactéries cultivées) associés à l'hydrothermalisme, extrait d'après Campbell *et al.*, 2006. *Nature Reviews Microbiology* 4, 458-468. (S<sup>0</sup>: Soufre élémentaire; SO<sub>3</sub><sup>2-</sup>: sulphite; S<sub>2</sub>O<sub>3</sub>: thiosulphate).

Donneurs d'électrons	Accepteurs d'électrons	Source de carbone	Métabolisme	Exemple de genres bactériens
S <sup>2-</sup> , S <sup>0</sup> , S <sub>2</sub> O <sub>3</sub> <sup>2-</sup>	O <sub>2</sub>	CO <sub>2</sub>	Oxydant le soufre	<i>Beggiatoa</i>
H <sub>2</sub>	O <sub>2</sub>	CO <sub>2</sub>	Oxydant l'hydrogène	<i>Alcaligenes</i>
NH <sub>4</sub> <sup>+</sup> , NO <sub>2</sub> <sup>-</sup>	O <sub>2</sub>	CO <sub>2</sub>	Nitrifiant	<i>Nitrosococcus</i>
Fe <sup>2+</sup> , Mn <sup>2+</sup>	O <sub>2</sub>	CO <sub>2</sub>	Oxydant le fer et le manganèse	<i>Siderococcus</i>
CH <sub>4</sub> , CO	O <sub>2</sub>	CH <sub>4</sub> , CO, CO <sub>2</sub>	Méthylotrophe et oxydant le monoxyde de carbone	<i>Methylomonas</i>
Carbone organique	O <sub>2</sub>	Carbone organique	Hétérotrophe aérobie	<i>Pseudomonas</i>
H <sub>2</sub>	NO <sub>3</sub> <sup>-</sup>	CO <sub>2</sub>	Dénitrifiant	<i>Bacillus</i>
H <sub>2</sub>	S <sup>0</sup> , SO <sub>3</sub> <sup>2-</sup> , S <sub>2</sub> O <sub>3</sub> <sup>2-</sup> , SO <sub>4</sub> <sup>2-</sup>	CO <sub>2</sub>	Réduisant le soufre et le sulfate	<i>Desulfurobacterium</i>
H <sub>2</sub>	S <sup>0</sup> , SO <sub>3</sub> <sup>2-</sup> , S <sub>2</sub> O <sub>3</sub> <sup>2-</sup> , SO <sub>4</sub> <sup>2-</sup>	Carbone organique	Réduisant le soufre et le sulfate	<i>Thiobacillus</i>
H <sub>2</sub>	CO <sub>2</sub>	CO <sub>2</sub>	Méthanogène et acétogène	<i>Methanobacterium</i>
Carbone organique	NO <sub>3</sub> <sup>-</sup>	Carbone organique	Dénitrifiant	<i>Acinetobacter</i>
Carbone organique	SO <sub>4</sub> <sup>2-</sup>	Carbone organique	Réduisant le sulfate	<i>Desulfovibrio</i>
Carbone organique	Carbone organique	Carbone organique	Fermentatif	<i>Clostridium</i>

Tableau 2 : Différents types de métabolismes bactériens retrouvés en écosystèmes hydrothermaux.

### **2-4-2-2 Archaea**

Les archaea constituent un groupe du Vivant récemment découvert sur la base d'analyses moléculaires (Woese & Fox, 1977) et sont séparées en 3 phyla : les Euryarcheota, les Crenarcheota (Woese *et al.*, 1990) et le phylum récemment découvert des Nanoarcheota (Huber *et al.*, 2002). Les archaea représentent la majorité des organismes hyperthermophiles (température optimale de croissance supérieure à 80°C) dans ces environnements, et incluent des méthanogènes (*Methanococcus*), des réducteurs de sulfate (*Archaeoglobus*), et des autotrophes facultatifs (*Pyrococcus*; Figure 11) ; tous ces organismes sont anaérobies stricts (Prieur, 1995). Des organismes résistants à plus de 110°C ont été isolés de ces écosystèmes : *Pyrolobus fumarii* a une température optimale de croissance de 113°C (Blöchl *et al.*, 1997) et la souche 121 détient le record avec une croissance optimale à 121°C (Kashefi & Lovley, 2003).

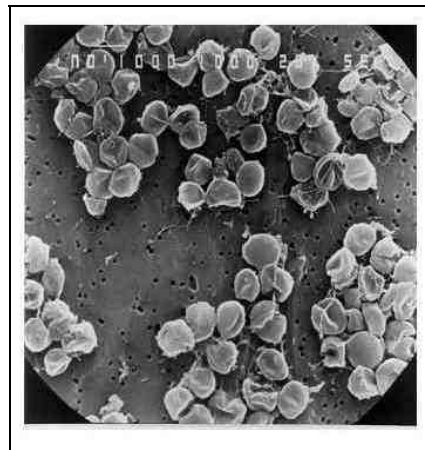


Figure 11: Photographie de microscopie électronique à balayage d'archaea : *Pyrococcus abyssi*, d'après Ifremer.fr.

### **2-4-2-3 Eucaryotes**

Les eucaryotes supérieurs découverts près des sources hydrothermales sont sans doute les organismes les plus spectaculaires inféodés à ces milieux. La communauté animale est dominée par les mollusques, les arthropodes et les polychaetes (Wolff, 2005). Leurs capacités à réguler rapidement leur métabolisme face aux conditions fluctuantes du milieu, couplées à

des pièces anatomiques protectrices (coquilles, tubes) permettent cette prédominance dans le milieu. Plus de 500 espèces endémiques au milieu hydrothermal étaient recensées en 2005 (Wolff, 2005 ; Le Bris *et al.*, 2006).

La composition des distributions faunistiques est différente entre les différents grands systèmes hydrothermaux, dépendant de l'activité tectonique des dorsales et des conditions physico-chimique associées. Voici une liste non exhaustive des organismes eucaryotes les plus présents dans ces écosystèmes et étudiés lors de cette thèse. Des photographies de ces organismes sont proposées dans la figure 12.

### *Bathymodiolus*

Les espèces du genre *Bathymodiolus* sont des modioles, de la famille des Mytilidae. Ce sont des organismes filtreurs pouvant obtenir leurs nutriments directement du milieu extérieur ; elles sont fixées à leur substrat (roche) par le byssus. L'une des caractéristiques principales de ce genre de Mytilidae est leur association obligatoire avec des bactéries symbiotiques. Les endosymbioses qu'elles possèdent sont de deux types, toutes les deux situées dans leurs branchies au niveau de cellules spécialisées, les bactériocytes : des symbiontes oxydant le soufre et des symbiontes méthylothropes. Ces symbiontes vont tirer leur énergie de l'oxydation des composés réduits émanant du fumeur, et permettre ainsi l'assimilation de carbone organique par la modiole (Fujiwara *et al.*, 2000 ; Fisher, 1990). L'occurrence de ces symbiontes dépend de l'espèce de *Bathymodiolus* considérée ; certaines espèces sont en association obligatoire avec deux symbiontes (*Bathymodiolus azoricus*), tandis que d'autres espèces possèdent seulement un type symbiotique (*Bathymodiolus thermophilus*) (Fujiwara *et al.*, 2000). Les espèces du genre *Bathymodiolus* sont les espèces dominantes de MAR, où elles forment des colonies importantes près des fumeurs (Figure 12) ; elles sont retrouvées dans des habitats mésophiles (20°C de moyenne), à pH légèrement acide à neutre, dans des environnements généralement riches en sulfure (H<sub>2</sub>S), en méthane (CH<sub>4</sub>), et en métaux lourds (Duperron *et al.*, 2006).

### *Rimicaris exoculata*

Les crevettes du genre *Rimicaris* se retrouvent dans les écosystèmes hydrothermaux atlantiques où elles abondent en véritables essaims (Figure 12). Cependant, malgré leur abondance dans ces sites, très peu d'études ont analysé le métabolisme de ces crevettes. Ainsi, leur source de carbone n'est pas connue et seules des hypothèses de fonctionnements symbiotiques ont pu être émises (Rieley *et al.*, 1999).

### *Alvinella pompejana*

Ces annélides polychètes sont uniquement retrouvés dans les sites hydrothermaux de l'océan Pacifique. Ils vivent dans des tubes fixés aux cheminées hydrothermales à des températures avoisinant 20 °C (mais pouvant atteindre 80°C). L'une des caractéristiques de ces organismes est la présence d'épibioses bactériennes (Figure 12) au niveau épidermique constituées de bactéries chimioautotrophes filamenteuses (Desbruyeres *et al.*, 1998).

### *Riftia pachyptila*

Ces vestimentifères tubicoles sont retrouvés uniquement le long de la ride Est Pacifique. Le tube, chitineux que les vers ne quittent jamais sert de protection aux conditions agressives du milieu. Ces organismes sont dépourvus d'appareil digestif et on trouvera à la place un trophosome qui est un organe spécialisé où se logent des bactéries endosymbiotiques oxydant le soufre. Par le biais d'un important panache branchial (Figure 12), le ver fournit les éléments énergétiques aux symbiotes qui en retour lui fournissent le carbone organique nécessaire à sa survie (Colaço *et al.*, 2002).

*Bathymodiolus azoricus*



*Rimicaris exoculata*



*Alvinella pompejana*



*Riftia pachyptila*



Figure 12: Photographies des principaux invertébrés des écosystèmes hydrothermaux, d'après Ifremer.fr.

## 2-5 L'hydrothermalisme océanique : le milieu berceau de la vie ?

Bien que très controversée, l'hypothèse du milieu hydrothermal comme milieu d'émergence de la vie sur Terre (Wächtershäuser, 1990) repose sur de nombreux arguments. Il constituerait ainsi pour certains auteurs le lieu de synthèse des premières molécules organiques, permettant ainsi l'établissement des premières formes de vie (Huber *et al.*, 2003 ; Huber & Wächtershäuser, 1997). Le premier argument repose sur l'observation de phylogénies moléculaires, où il est apparu que les formes les plus primitives de l'arbre de la vie étaient constituées d'organismes thermophiles chimioautotrophes (Huber & Wächtershäuser, 1997). Par leur profondeur, les écosystèmes hydrothermaux océaniques auraient de plus été les seuls environnements où une vie primitive aurait pu être protégée des impacts météoritiques et des rayons ultraviolets (Holm, 1992). La théorie d'un système de métabolisme de surface a été initiée par Wächtershäuser en 1988 (Wächtershäuser, 1988), se basant sur les théories de l'ancêtre commun des trois domaines de la vie de Woese (Woese & Fox, 1977 ; Woese, 1979 ; Woese, 1990). Cet ancêtre commun serait inorganique, acellulaire, et capable d'autotrophie par processus autocatalytiques. Les cristaux de Pyrite (FeS) formés au niveau des fumeurs par précipitation du fer constitueraient les ligands à partir desquels sont générés spontanément des peptides (Huber *et al.*, 2003). La Pyrite chargée positivement permettrait l'adhérence des composés anioniques en contact avec une eau de mer chaude. Un jeu de sélection négative par dégradation des liaisons chimiques aurait ainsi permis la génération d'une bibliothèque aléatoire de peptides au niveau des cheminées hydrothermales (Huber *et al.*, 2003). Cette théorie implique donc que les compartiments formés par les cristaux de Pyrite constitueraient les ancêtres d'une organisation cellulaire, permettant à la fois des mécanismes de concentration, nécessaires à la catalyse de réactions biochimiques, et des mécanismes de production d'énergie, via les réactions d'oxydo-reduction (provenant de la transformation des ions ferreux et de sulfure d'hydrogène en Pyrite), indispensables à la fixation de carbone et donc à l'établissement de la vie (Huber *et al.*, 2003 ; Koonin & Martin, 2005). Le milieu hydrothermal fournissant de nombreuses surfaces autocatalytiques, de nombreux gradients thermiques permettant la catalyse d'une grande variété de réactions chimiques et d'activités enzymatiques, ainsi que de nombreuses molécules indispensables aux synthèses organiques (CO<sub>2</sub>, H<sub>2</sub>S, acides aminés) constituerait donc un incubateur à l'établissement de molécules organiques. La figure 13 reprend les mécanismes hypothétiques ayant mené, à partir de substrats inorganiques (Pyrite composant les événements hydrothermaux), à l'émergence des bactéries et archaea (Koonin & Martin 2005).



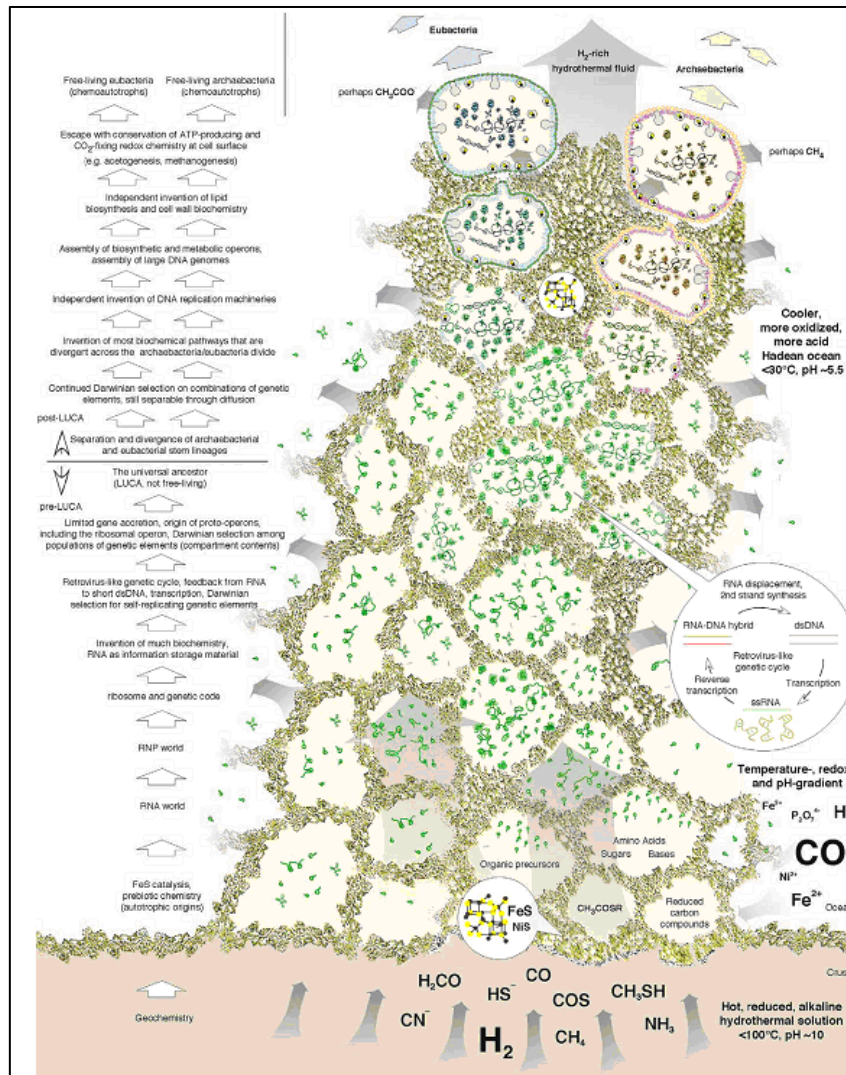


Figure 13: Hypothèses de scénarii évolutifs, ayant mené à la formation des cellules de bactéries et d'archaea, à partir de la formation de compartiments inorganiques de Pyrite. Schéma extrait d'après Koonin & Martin 2005. *Trends in Genetic*. 21: 647-654.

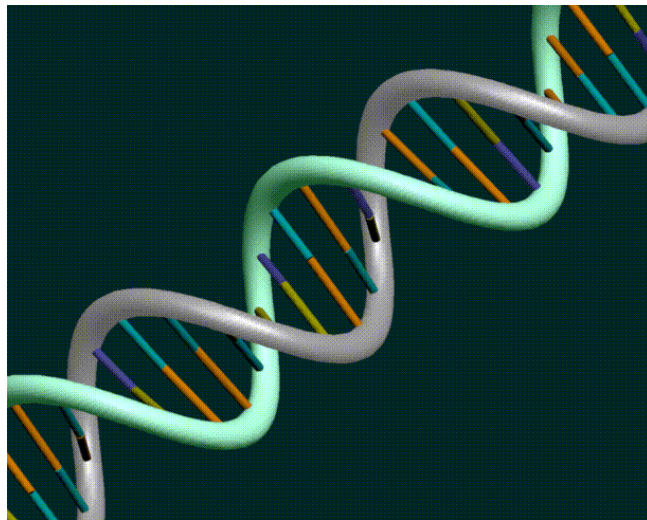
Un argument supplémentaire soutenant l'hypothèse de l'émergence de la vie en milieu hydrothermal marin est proposé par Robert et Chaussidon (2006), qui ont démontré par analyse isotopique du silicium que les conditions hydrothermales, et notamment les conditions de température, étaient communes à celles des températures de surface des océans (70°C) il y a 3.5 milliards d'années (Précambrien, Archéen). Il est ainsi proposé que les microorganismes actuels vivant en écosystème hydrothermal posséderaient des caractéristiques physiologiques et métaboliques similaires aux premiers microorganismes ayant émergé sur Terre (Martin *et al.*, 2008a).

*Dans le cadre de l'étude menée lors de cette thèse, l'analyse des communautés fongiques dans ces environnements nous permettra d'obtenir des informations sur un champ d'étude jusqu'alors quasi vierge. Outre l'aspect diversité, d'un point évolutif l'une des attentes de notre travail était de trouver des organismes ayant divergé très tôt au sein du Règne fongique, en accord avec la théorie présentée précédemment du milieu hydrothermal comme lieu d'apparition de la vie : retrouver dans ces écosystèmes des organismes à la base de la phylogénie des champignons nous permettra ainsi d'élaborer des hypothèses évolutives ayant mené à la radiation des différents groupes de champignons. Bien sûr cela ne permettrait pas de consolider l'hypothèse du milieu hydrothermal comme berceau de la vie.*



## Chapitre I-3 : Métagénomique

---



## I-3 Métagénomique

Très peu étudiés en terme de diversité spécifique, les champignons des écosystèmes hydrothermaux marins, et plus particulièrement les fonctions qu'ils portent, restent totalement inconnus. Comme nous l'avons abordé dans le chapitre 'Introduction générale', la majorité des (micro)organismes ne peuvent pas être cultivés en conditions de laboratoire et cela est d'autant plus vrai quand on étudie des écosystèmes aux caractéristiques physico-chimiques si particulières. Afin d'approcher les fonctions portées par les champignons isolés de ces écosystèmes, nous avons choisi une approche métagénomique qui nous permettra directement de nous intéresser aux gènes (et donc aux fonctions portées par ces gènes) des champignons, en s'affranchissant de toute étape de culture. Dans ce chapitre, après un rappel sur les définitions et l'historique des développements de technique de séquençage, nous aborderons leurs applications pour l'analyse des communautés microbiennes.

### 3-1 Définition

La métagénomique est l'étude des génomes de l'ensemble des organismes d'un milieu donné (Handelsman *et al.*, 1998). A l'heure actuelle, les analyses métagénomiques ont été menées uniquement sur des microorganismes (bactéries, archaea, virus, picoplancton). La majorité de ces microorganismes n'étant pas cultivables dans des conditions de laboratoire, la métagénomique offre un accès au pool génétique de la communauté en s'affranchissant de toute étape de culture et va donc permettre une compréhension prédictive du fonctionnement écosystémique (Turnbaugh & Gordon, 2008). Une étude d'Amann *et al.* (1995) a démontré que seuls 0.001 à 0.1% des microorganismes marins et 0.3 % des microorganismes du sol pouvaient être cultivés *in vitro*, ce qui réduit grandement la compréhension des écosystèmes microbiens quand seules des techniques basées sur la culture des microorganismes sont employées (Amann *et al.*, 1995 ; Singh *et al.*, 2008).

A ma connaissance, la première utilisation du terme « métagénomique » date de 1998, avec des travaux de Handelsman *et al.* (1998) s'intéressant à l'étude des génomes de tous les microorganismes présents dans l'environnement (ici le sol), et non pas aux seuls organismes cultivables. Cette vision ouvrira ainsi la voie à des travaux permettant une meilleure

compréhension du fonctionnement des écosystèmes microbiens, et sera également utilisée dans des approches biotechnologiques (recherches de gènes codant des enzymes particulières, par exemple : Bull *et al.*, 2000). L'application de ces techniques aux organismes non cultivables a également rendu possible l'analyse du génome d'organismes symbiotiques (Woyke *et al.*, 2006 ; Robidart *et al.* 2008).

### 3-2 Historique

Les premiers travaux de métagénomique imposaient une étape de clonage de l'ADN environnemental qui permettait une amplification du matériel génétique de départ; on peut noter par exemple, l'étude pionnière de Schmidt *et al.* (1991) qui s'intéressait à l'analyse d'une communauté picoplanctonique marine en clonant et sélectionnant les séquences d'ARNr 16S à partir de l'ADN génomique total extrait.

L'étape de clonage, proposée pour la première fois par Pace *et al.* (1986), qui peut s'effectuer par l'intermédiaire de BAC (Bacterial Artificial Chromosomes) ou de cosmides par exemple, permet d'isoler de larges fragments d'ADN pour former une banque génomique ; le criblage de cette banque de clones permet de rechercher des gènes entiers, voire des voies métaboliques complètes (DeLong *et al.*, 2006 ; Tyson *et al.*, 2004 ; Woyke *et al.*, 2006,...). Un schéma résumant les principales étapes de l'étude des métagénomies via la construction de banques génomiques est proposé en figure 14.

Pour les inserts courts, c'est à dire inférieurs à 10kb, l'ADN est incorporé dans un plasmide puis une souche bactérienne est transformée. Après multiplication *in vitro* des bactéries portant le plasmide et son insert, on peut alors recourir au séquençage. Dans le cas des inserts plus longs, d'autres vecteurs sont utilisés : les cosmides pour des inserts dont la taille est comprise entre 25 et 35 kb (Entcheva *et al.*, 2001), les BACs pour des inserts de 200 Kb (Pace *et al.*, 1986) et des YACs (Yeast Artificial Chromosome) au delà de 200 Kb (Murray *et al.*, 1986) et jusqu'à 400 Kb (figure 14). Dans les cas de constructions de banques métagénomiques par clonage, deux options peuvent être envisagées. La première, qui consiste à séquencer les clones portant des marqueurs phylogénétiques, a été employée notamment dans l'analyse de marqueurs de gènes codant l'ARNr16S des communautés microbiennes présentes dans le sol (Barns *et al.*, 1999). La seconde option est centrée sur la fonction portée par les fragments de gènes clonés. Les fragments d'ADN extraits sont clonés dans des vecteurs d'expression ; les clones sont ensuite criblés en fonction de leur activité *in vitro* (par exemple Beja *et al.*, 2000a ; Handelsman, 2004). Les criblages fonctionnels ont par exemple

permis de mettre en évidence la protéorhodopsine chez des bactéries marines (Beja *et al.*, 2000b; Beja *et al.*, 2000c) ou encore d'identifier des enzymes impliquées dans la dégradation de composés aromatiques (Suenaga *et al.*, 2007). Ces méthodes qui impliquent un séquençage massif de nombreux clones, sont limitées par la complexité des écosystèmes étudiés et aussi par des biais de clonage importants. Pour ce dernier point il est très connu que les régions riches en AT sont réfractaires au clonage.

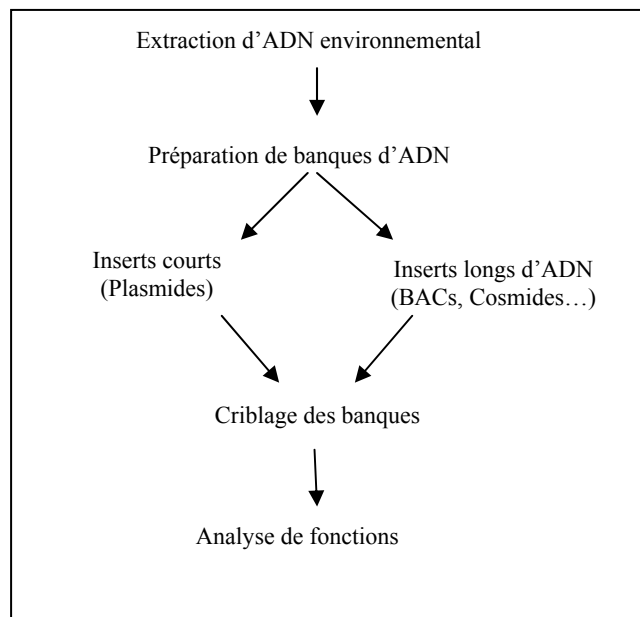


Figure 14: Etapes de l'analyse et de l'identification des gènes à partir d'un métagénome microbien, selon une stratégie de transformations de vecteurs plasmidiques ou viraux. (d'après Streit & Schmitz, 2004).

Des techniques d'enrichissement ont également été appliquées afin de cibler une fraction microbienne active d'un écosystème. La technique de Stable Isotope Probing (SIP) proposée par Radajewski *et al.* (2000) consiste en un marquage de substrat à l'aide d'un isotope stable. Les bactéries actives ayant métabolisé le substrat vont incorporer l'isotope marqué dans leur génome. Une extraction secondaire de l'ADN des bactéries composant la communauté permettra de déterminer les microorganismes actifs dans le processus écologique. L'analyse des ARN enrichis (SIP-ARN) offre l'avantage d'étudier à échelle de

temps réduite et avec un effet dilution limité de l'isotope, les microorganismes actifs (sensibilité accrue), acteurs d'une ou de plusieurs fonctions écologiques (Manefield *et al.*, 2002 ; Vandenkoornhuyse *et al.*, 2007).

Une avancée importante dans l'analyse des communautés microbiennes océanique a été initiée par l'expédition Global Ocean Sampling 2004 (Venter *et al.*, 2004). Cette expédition consistait à échantillonner des eaux de surface le long de la mer de Sargasse et de séquencer en masse l'ensemble des échantillons par la technique de Whole Genome Sequencing (WGS) par séquençage 'traditionnel'. A l'issue du WGS, 1.987.936 séquences d'ADN ont été produites. Sur l'ensemble de ces séquences, 1.214.207 séquences ont pu être identifiées et annotées comme des gènes codant des protéines; 1412 séquences de gènes codant des ARNr ont pu être obtenues. Cette étude, bien que révélant la puissance d'une telle approche pour l'étude de communautés microbiennes, a également montré des limites, notamment dans la difficulté d'assembler de larges contigs à partir des séquences obtenues par WGS, en partie due à la diversité spécifique importante échantillonnée, ainsi que dans la difficulté d'annoter la majorité des séquences codantes. Des problèmes liés aux transferts horizontaux de gènes entre groupes d'organismes différents ont également pu être soulignés à partir de cette étude pionnière (Venter *et al.*, 2004).

### **3-3 La révolution des nouvelles méthodes de séquençage**

L'un des tournants majeurs dans l'étude des communautés microbiennes environnementales a été la mise au point d'une nouvelle technologie de séquençage appelée le pyroséquençage (Margulies *et al.*, 2005). Le principe de cette technique, développée par 454 technologies puis Roche, est expliqué dans la partie Matériel et Méthodes de cette thèse. Le principal avantage de cette technique est de donner un accès à moindre coût à des centaines de milliers de séquences produites simultanément à partir d'un échantillon environnemental, sans passer par une étape préalable de clonage traditionnel de banque génomique (Figure 14). La quantité d'ADN requise au départ est très faible (de l'ordre de 5 µg ; Harkins & Jarvie, 2007). D'autres méthodes de séquençage sont actuellement en développement mais à l'heure actuelle, la technique de pyroséquençage remplace largement la stratégie de Sanger (Sanger *et al.*, 1977) dans les études métagénomiques microbiennes. Les séquences générées par pyroséquençage sont plus courtes (< 500 pb aujourd'hui) que celles produites par séquençage de Sanger (700 pb en moyenne). Il est à noter que des comparaisons entre les résultats générés



par pyroséquençage et ceux générés par les techniques de Sanger ont donné des résultats équivalents, approuvant la robustesse de cette technique (Gharizadeh *et al.*, 2006, Gill *et al.*, 2006, Frias-Lopez *et al.*, 2008).

Comme nous l'avons évoqué précédemment, d'autres méthodes de séquençage de nouvelle génération ont également été mises au point. On peut citer pour exemple la technique de séquençage dite de Solexa (Illumina Genome Analyzer ; Bennett, 2004) qui implique une amplification préalable du matériel génétique sur une surface solide, suivie par une synthèse par incorporation de nucléotides modifiés (extrémité 3'OH bloquée) et marqués par un fluorochrome associés à des colorants (dyes). L'incorporation des nucléotides se fait grâce à une ADN polymérase isothermale. Cette technologie permet d'obtenir en 4 jours de 40 à 50 millions de séquences d'une longueur comprise entre 32 et 60 pb (Mardis, 2008). Une troisième technologie de séquençage, commercialisée par Applied Biosystems sous le nom de SOLID Sequencer (Sequencing by Oligo Ligation and Detection), a été mise sur le marché en 2007. Cette technique, basée sur l'incorporation d'oligonucléotides semi-dégénérés sur une surface solide permet d'obtenir en 5 jours de 3 à 4 Gb de données de séquences, de longueur moyenne de 25-35pb. Une comparaison de ces trois méthodes est proposée dans le tableau 3. L'utilisation de ces 2 dernières techniques (Solexa et SOLID) n'est pas appropriée dans le cas de l'étude des écosystèmes microbiens, la taille des fragments générés étant largement trop faible pour obtenir une annotation précise.

	<b>Roche (454)</b>	<b>Illumina</b>	<b>SOLID</b>
Chimie	Pyroséquençage	Séquençage par synthèse	Séquençage par ligation
Amplification	PCR par émulsion	Amplification sur surface solide	PCR par émulsion
Mb/run	500 Mb	1300 Mb	3000 Mb
Temps/run	5h30	4 jours	5 jours
Longueur des séquences	350 pb	40 pb	30 pb
Coût par run	8000 \$	8950 \$	17447 \$

Tableau 3: Comparaison des techniques de séquençage de nouvelle génération. Modifié d'après Mardis, 2008.

### 3-4 Traitement des données

Face aux jeux de données de séquences générées par pyroséquençage qui se révèlent souvent d'une grande complexité, Raes *et al.* (2007) proposent un schéma général pour étudier ces résultats. La procédure, schématisée dans la figure 15 est la suivante :

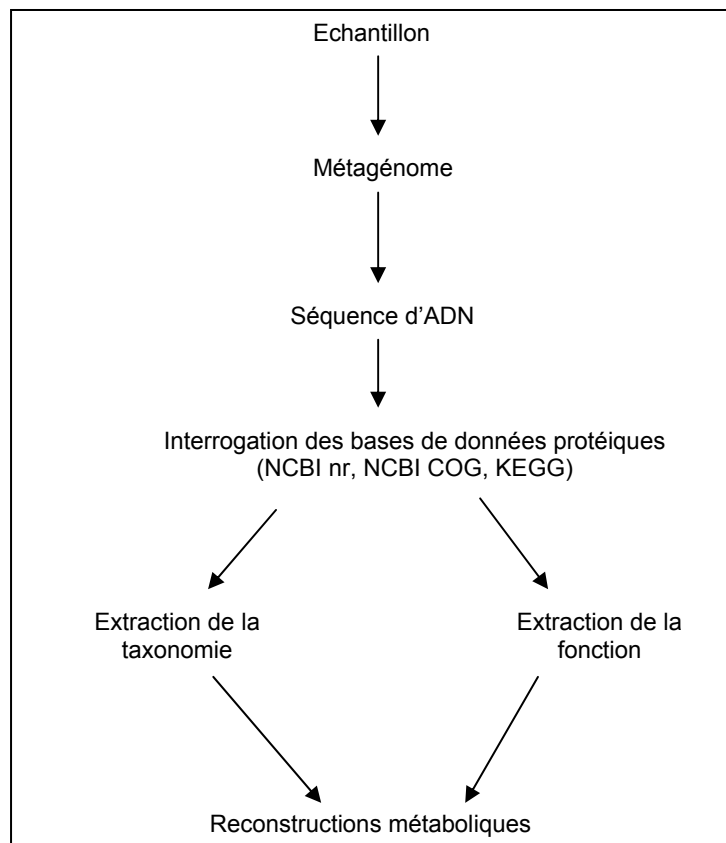


Figure 15: Etapes de l'analyse et de l'identification des gènes à partir d'un métagénome microbien, selon une stratégie de pyroséquençage (d'après Raes *et al.*, 2007. *Current Opinion in Microbiology*. 10:490-498).

A partir des séquences générées, une étape d'assemblage en contigs est effectuée afin de disposer de fragments plus longs et donc plus informatifs. Dans le cas des environnements possédant des communautés complexes, cette étape d'assemblage peut être délicate car il n'est pas possible d'utiliser un génome de référence pour guider l'assemblage des séquences en contigs. Il s'agit ici de séquençage *de novo*. L'assemblage sera bien sûr plus aisé quand la

diversité des organismes composant le métagénome sera faible (Tyson *et al.*, 2004). Une seconde stratégie d'assemblage, utilisée lorsque les génomes de référence ne sont pas disponibles, repose sur la pré-assignation des séquences en groupes phylogénétiques. Pour cela, les séquences sont clusterisées en fonction de leur composition nucléotidique (% GC des séquences, propriétés de séquences spécifiques à un clade bactérien), puis assemblées dans un second temps (Garcia *et al.*, 2006). Une fois de plus, cette étape est rendue difficile lors de l'étude de communautés complexes. Dans le cas de ces environnements complexes, un effort de séquençage supplémentaire doit être effectué afin d'obtenir une couverture des génomes dominants plus importante (Tringe *et al.*, 2005).

La composition spécifique d'un métagénome est un élément descriptif important : elle permet de connaître la contribution des espèces microbiennes dans le fonctionnement de l'écosystème étudié ; pour cela, chaque séquence/contig est comparé aux bases de données mondiales par des procédures BLAST (Basic Local Alignment Search Tool, Altschul *et al.*, 1990). Cette comparaison est rendue possible par le fait que les gènes codant les enzymes aux fonctions importantes pour la cellule seront conservés au cours de l'évolution, autorisant une approche comparative de ces séquences (Koonin & Galperin, 2002). L'approche comparative permet la prédiction des fonctions des protéines en transférant les informations obtenues à partir d'organismes modèles ou de séquences disponibles dans les bases de données vers des homologues non caractérisés (Koonin & Galperin, 2002). L'approche BLAST sera donc utilisée pour rechercher des régions homologues des séquences inconnues (nucléotidiques ou protéiques) dans les bases de données (Koonin & Galperin, 2002). La comparaison des séquences protéiques sera plus informative que celle utilisant des données nucléotidiques; ceci est dû d'une part au fait qu'il existe 20 acides aminés contre seulement 4 types de nucléotides, et d'autre part, compte tenu de la redondance du code génétique, autorisant des taux de pression sélective différentes en fonction de la position des nucléotides, ce qui entraîne un bruit de fond (le troisième nucléotide d'un triplet étant peu ou pas soumis aux pressions de sélection). Cependant dans le cas de l'étude des régions non codantes, une comparaison des séquences nucléotidiques sera privilégiée (Koonin & Galperin, 2002).

A l'issue de la requête, le premier résultat de BLAST est en général récupéré pour chaque séquence puis leur taxonomie déduite (par exemple Poinar *et al.*, 2006). Cette technique qui a l'avantage d'être rapide, possède des défauts : elle est très dépendante des annotations antérieures de séquences dans les bases de données et elle ne permet pas de caractériser des événements de transferts horizontaux de gènes qui pourraient biaiser l'analyse (Koonin & Galperin, 2002 ; Krause *et al.*, 2006 ; Mavromatis *et al.*, 2007). Par ailleurs les

recherches BLAST sont basées sur des procédures heuristiques. D'autres méthodes, utilisant des données de pyroséquençage ont également été mises au point et seront abordées plus en détail dans le chapitre I-3-5-1.

La seconde étape de l'analyse des données consiste en l'annotation fonctionnelle des séquences générées. Pour cela, les procédures BLAST sont également utilisées, contre des bases de données protéiques (NCBI non-redundant database, NCBI Cluster of Orthologous Groups,...), et le premier résultat de BLAST pour chaque séquence est récupéré. Une fois de plus cette procédure sera sensible aux compositions des bases de données ; il a été démontré que jusqu'à 50% des données pouvaient obtenir une annotation fonctionnelle par cette méthode (Raes *et al.*, 2007). Ces procédures BLAST peuvent être affinées en complétant par exemple l'analyse par la recherche de domaines protéiques (Krause *et al.*, 2006). Par ailleurs des algorithmes spécifiques de recherche de gènes, indépendants des recherches BLAST, ont été mis au point (Krause *et al.*, 2006).

### **3-5 Applications**

Les avancées technologiques des méthodes de séquençage à haut débit ont permis d'analyser les génomes d'organismes non cultivés, de révéler de nouveaux gènes ou de nouvelles voies métaboliques, ainsi que des métabolismes complets associés à des environnements. Des comparaisons de compositions génétiques entre différents environnements ont également été entreprises (DeLong, 2005). Dans le cadre de cette thèse nous aborderons uniquement les applications des méthodes de pyroséquençage pour l'étude du compartiment microbien des écosystèmes, mais ces techniques sont également très largement utilisées pour le séquençage *de novo* de génomes entiers (exemple du séquençage du génome nucléaire du mammoth ; Poinar *et al.*, 2006).

#### **3-5-1 Applications de la métagénomique à la composition taxonomique d'un environnement**

Les premières études métagénomiques s'intéressaient particulièrement à la composition taxonomique d'un environnement donné. Ces études s'appuyaient sur l'analyse de marqueurs phylogénétiques amplifiés par PCR, et notamment le gène codant pour l'ARNr 16S. Elles ont dès lors permis la découverte de nouvelles lignées taxonomiques bactériennes

(Pace, 1997). Dans un second temps, les génomes des organismes cultivables en laboratoire pouvaient être séquencés. Les techniques de pyroséquençage appliquées à ces études taxonomiques permettent d'obtenir des centaines de milliers de séquences simultanément : la taille des séquences générées (amplicons de l'ARNr 16S) allant de 100 à 300 pb pour les pyroséquenceurs de première génération (500 pb actuellement), permettent une bonne résolution de la composition phylogénétique des échantillons (Sogin *et al.*, 2006 ; Liu *et al.*, 2007). De plus, cette technique donne accès aux organismes sous-représentés dans l'environnement, mais qui peuvent tenir un rôle important dans l'écosystème (Sogin *et al.*, 2006). L'accès à ce pool d'organismes sous-représentés constituait un biais important lors de l'étude des écosystèmes bactériens par clonage / séquençage (Singh *et al.*, 2008).

La composition taxonomique d'un environnement donné peut être obtenue en effectuant un séquençage massif de courtes régions de marqueurs phylogénétiques, par une approche dite de « tag sequencing » (Sogin *et al.*, 2006). Une étude de 2008 (Huse *et al.*, 2008) a prouvé que l'amplification massive par pyroséquençage de courts fragments de gène codant les ARNr bactériens (amplification des régions hypervariables V3 et V6), était efficace pour l'assignation taxonomique des séquences produites. Cependant les procédures BLAST classiques n'apparaissent pas suffisamment robustes pour une annotation taxonomique de ces séquences et un algorithme de recherche particulier a été mis en place (GAST, Global Alignment for Sequence Taxonomy) qui permet d'utiliser les fragments complets des gènes codant les ARNr et des phylogénies précalculées avec ces fragments entiers, pour assigner une taxonomie (jusqu'au niveau spécifique) aux séquences V6 amplifiées (Huse *et al.*, 2008). Ces méthodes ont été utilisées avec succès pour évaluer la composition taxonomique bactérienne d'échantillons d'eau de mer profonde (Sogin *et al.*, 2006), ou d'échantillons hydrothermaux (Sogin *et al.*, 2006 ; Huse *et al.*, 2008).

De nouvelles méthodes non basées sur l'utilisation de marqueurs phylogénétiques ont également été mises au point et se basent sur les résultats de recherches BLAST. Le logiciel MEGAN (Metagenome Analyzer) a été développé à cet effet (Hudson *et al.*, 2007). Les séquences générées sont comparées aux bases de données protéiques en utilisant l'outil BLAST; sur la base des résultats, une assignation taxonomique se basant sur l'algorithme du « Lowest Common Ancestor » (algorithme de l'ancêtre commun le plus ancien) est effectuée. Cette méthode a été appliquée avec succès sur les échantillons pour ré-analyser des jeux de données métagénomiques de Venter *et al.* (2004) par exemple (Hudson *et al.*, 2007).

### **3-5-2 Applications de la métagénomique à la compréhension du fonctionnement des écosystèmes**

Outre l'accès à la composition taxonomique d'un échantillon et par extension d'un environnement donné, l'étude de l'ensemble des génomes d'une communauté permet de révéler les différentes voies métaboliques portées par cette communauté. Cette analyse nécessite une annotation fonctionnelle des gènes identifiés et se focalisera donc plus spécialement sur les gènes codants. Pour cela comme nous l'avons vu, les séquences générées sont comparées aux bases de données dans le but d'obtenir une annotation fonctionnelle. En général deux approches sont privilégiées : la première consiste à analyser les séquences d'ADN contre la banque de données KEGG (Kyoto Encyclopedia of Genes and Genomes) ; cette analyse va permettre de placer les gènes codant des protéines sur les voies métaboliques proposées par KEGG. Chaque voie métabolique regroupant toutes les fonctions cellulaires sera ainsi cartographiée, ce qui donnera une information permettant de prédire le fonctionnement d'un écosystème particulier mais aussi de déterminer les voies métaboliques dominantes vraisemblablement préférentielles pour cet écosystème (Gill *et al.*, 2006). La seconde approche consiste à classifier les séquences d'ADN en fonction de la nomenclature proposée par COG (Cluster of Orthologous Groups ; Tatusov *et al.*, 2003) ; cette analyse COG va permettre de relier les gènes possédant les mêmes fonctions, en étudiant les relations évolutives entre ces gènes (Gill *et al.*, 2006). La classification hiérarchique permettra également de déterminer quelles sont les voies dominantes dans le métagénome. Comme pour les études taxonomiques, ces analyses nécessitent de disposer d'une bonne couverture de séquençage, afin de pouvoir mettre en évidence les voies métaboliques sous-représentées dans le jeu de données (Gill *et al.*, 2006 ; Krause *et al.*, 2006).

Les analyses de métagénomiques ont été appliquées à de nombreux environnements et ont dès à présent permis de répondre à des questions de natures diverses. De nouveaux processus métaboliques ont ainsi pu être découverts, comme la mise en lumière d'un nouveau type de métabolisme photosynthétique bactérien dans les océans (Beja *et al.*, 2000 b; Beja *et al.*, 2000c). Un historique de ces études métagénomiques depuis 2003 est proposé dans la figure 16.

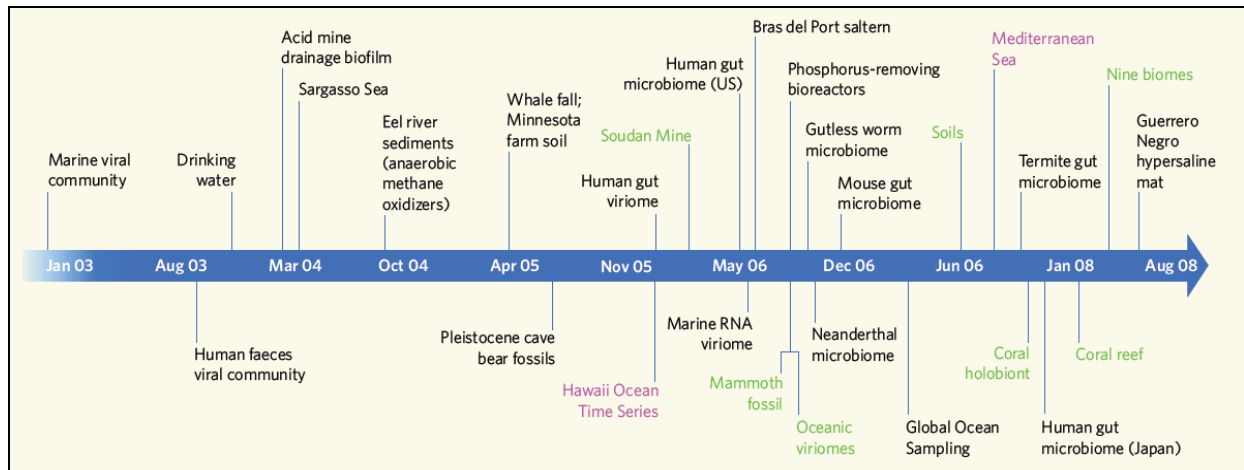


Figure 16: Historique des études métagénomiques depuis 2003. D'après Hugenholtz et Tyson, 2008. *Nature* 455: 481-483.

Une analyse métagénomique de communautés microbiennes vivant dans les eaux de drainage de mines acides a permis de déterminer les organismes dominants dans le biofilm par le séquençage de banques de gènes codant l'ARNr 16S couplé au séquençage du métagénome entier (Tyson *et al.*, 2004). Cinq organismes microbiens se sont révélés dominants. L'analyse du métagénome a permis de retrouver la quasi-totalité des gènes de ces organismes dominants, autorisant une reconstruction métabolique du fonctionnement du biofilm. A partir des séquences codantes retrouvées chez les 2 organismes bactériens dominants (*Leptospirillum sp.*, et *Ferroplasma sp.*), une reconstruction de leur métabolisme cellulaire a pu être réalisée et les interactions entre bactéries du biofilm ont pu être proposées (Tyson *et al.*, 2004).

L'utilisation de techniques de pyroséquençage permet également de se focaliser sur une voie métabolique donnée dans un environnement ou sur certains gènes présents dans l'environnement. En 2006, une étude de Leininger *et al.* sur les voies d'oxydation du méthane dans le sol, a permis de déterminer les organismes dominants impliqués dans cette voie métabolique à partir du pyroséquençage d'ADNc (ADN complémentaire aux ARN extraits, obtenus par rétrotranscription). Des clusters de gènes impliqués dans les mécanismes de dénitrification dans le sol ont également été retrouvés par pyroséquençage permettant de mieux comprendre leur organisation génétique, mais également de révéler les organismes impliqués dans ces mécanismes (Demanèche *et al.*, 2009).

Une analyse de communautés microbiennes productrices de biogaz en fermenteur a également été menée par une approche métagénomique utilisant le pyroséquençage (Schlüter *et al.*, 2008). L'analyse des séquences générées a permis de révéler des voies métaboliques impliquées dans les processus fermentaires, mais également de reconstruire les voies de méthanogénèse impliquant les organismes de cet environnement. Des hypothèses de fonctionnement des communautés microbiennes impliquées dans ces voies de méthanogénèse, couplées à l'analyse des organismes microbiens dominants, ont permis de révéler la puissance du pyroséquençage pour analyser les fonctions métaboliques portées par des consortia microbiens (Schlüter *et al.*, 2008).

Les analyses métagénomiques se sont également portées sur l'étude des communautés microbiennes ou virales des écosystèmes marins. Ainsi par exemple, Culley *et al.* (2006) ont analysé les communautés de virus à ARN dans des écosystèmes marins, très peu étudiées jusque là. Des différences entre les communautés virales de différents environnements ainsi que la reconstruction de génomes de virus inconnus (reconstruction facilitée par la faible taille des génomes viraux) ont ainsi pu être établies. Outre le fait d'avoir révélé une diversité spécifique virale importante non soupçonnée, des analyses phylogénétiques complémentaires basées sur la séquence d'acides aminés d'ARN polymérase ARN dépendante, ont prouvé que ces virus inconnus étaient distants des autres groupes de virus, renforçant l'intérêt des techniques de métagénomique pour l'étude de ces communautés (Culley *et al.*, 2006). L'essor de ces techniques a également permis de comparer les communautés virales isolées dans différentes zones géographiques océaniques (Angly *et al.*, 2006). Par ailleurs récemment, le pyroséquençage a été utilisé pour reconstruire des génomes viraux à partir de populations isolées d'hôtes humains résistants aux traitements médicaux (Eriksson *et al.*, 2008).

L'utilisation des techniques de séquençage à haut débit a également permis d'étudier les organismes symbiotiques d'un ver hydrothermal *Olavius algarvensis* (Woyke *et al.*, 2006). Les séquences d'ADN générées ont été dans un premier temps classifiées phylogénétiquement sur la base de l'analyse du gène codant l'ARNr16S, et sur la base d'algorithmes de clusterisation permettant la séparation des séquences de 4 organismes symbiotiques différents. Dans un second temps, une reconstruction physiologique des voies métaboliques de chaque symbionte a pu être réalisée sur la base de l'annotation des gènes. Un schéma métabolique regroupant les fonctions portées par chaque symbionte a ainsi pu être élaboré. Outre le fait que cette étude ait permis d'analyser les voies métaboliques portées par les génomes d'organismes jamais cultivés en laboratoire, les relations entre ces organismes et



les conditions physico-chimiques de l'environnement ont également pu être établies (Woyke *et al.*, 2006).

*Lors du travail réalisé dans cette thèse, nous avons choisi une approche métagénomique originale nous permettant d'étudier le métabolisme d'un organisme fongique inconnu, non cultivé, à partir de l'analyse de l'ensemble des séquences nucléotidiques produites par séquençage massif. Ce travail s'est accompagné d'une analyse globale de l'ensemble des métabolismes microbiens présents dans un environnement hydrothermal marin (Ride Médio-atlantique), qui s'est avérée efficace pour proposer des hypothèses de fonctionnement global du compartiment microbien dans cet environnement.*

## **Partie II : Matériel et méthodes**

---

# Matériel et méthodes

Dans la partie Matériel et méthodes de cette thèse nous aborderons dans un premier temps les sites d'échantillonnage sur lesquels nous avons travaillé (partie II-1). Dans un second temps, les différentes méthodes que nous avons utilisées pour approcher la diversité fongique seront discutées (partie II-2). Ces études de diversité ont été complétées par une analyse quantitative des gènes codant l'ARNr18S fongique, présentée dans la partie II-3 de ce chapitre. Dans la 4ème partie de ce chapitre, nous aborderons les outils ayant permis de mettre en place une base de données experte regroupant des marqueurs moléculaires fongiques (partie II-4). La dernière partie quant à elle présentera les méthodes et outils liés au pyroséquençage, ainsi que les stratégies d'analyse choisies nous ayant permis d'établir des hypothèses d'écologie fonctionnelle (partie II-5).

## II-1 Sites d'études et échantillons

La collecte d'échantillons en écosystèmes hydrothermaux profonds est soumise à de nombreuses contraintes techniques, que nous avons évoquées dans le premier chapitre de cette thèse. L'une des principales difficultés est de localiser les lieux d'échantillonnage, qui nécessite une succession de campagnes océanographiques. Les échantillons que nous avons étudiés lors de cette thèse ont tous été prélevés lors de campagnes menées entre 1991 et 2005 (figures 17 et 18) ; ces échantillons ont été cryoconservés à  $-70^{\circ}\text{C}$  en présence d'un cryoprotecteur (DMSO 5% généralement). Un matériel adapté aux fortes pressions hydrostatiques et aux conditions chimiques du milieu a ainsi dû être développé ; tous ces prélèvements ont été effectués à l'aide d'un submersible et d'équipements associés tels que des seringues en titane pour l'échantillonnage de fluides, des caisses isothermes pour les prélèvements d'échantillons d'animaux...

Lors de cette thèse, des échantillons de diverses natures ont été étudiés, avec à la fois des prélèvements minéraux (morceaux de sulfures de roches, de sédiments), ou animaux (*Bathymodiolus azoricus*, *Rimicaris exoculata*...). La liste complète des échantillons est présentée ci-après ; les localisations géographiques des sites étudiés sont présentées dans les figures 17 et 18.

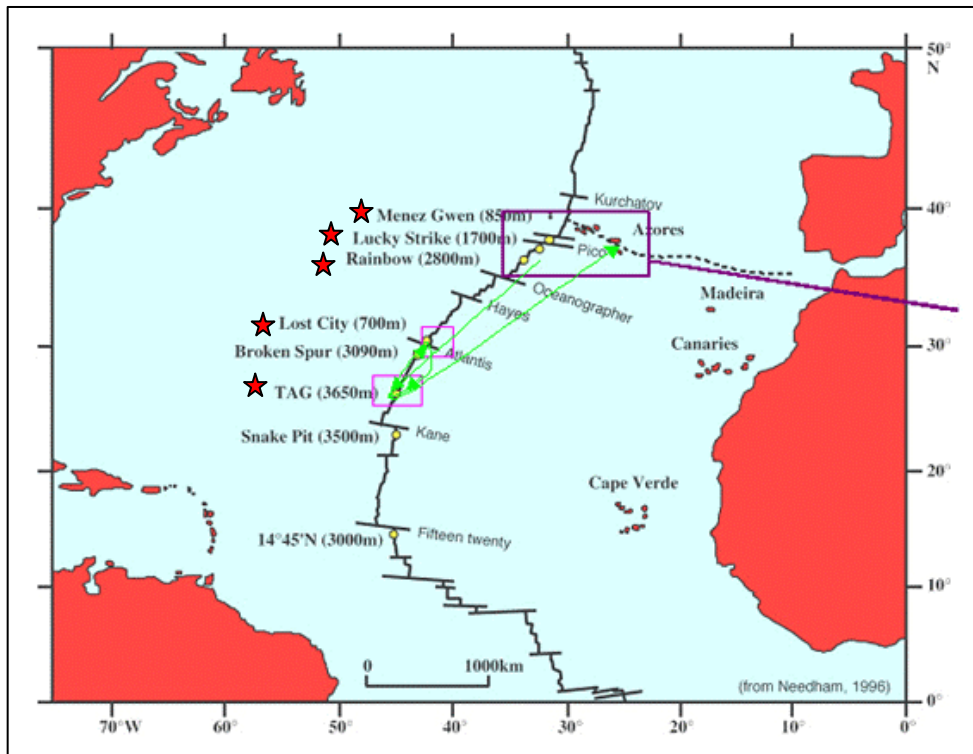


Figure 17: Cartographie des principaux sites étudiés sur la ride Medio Atlantique ; les profondeurs des sites sont indiquées entre parenthèses. *Les étoiles rouges représentent les sites étudiés lors de cette thèse.*

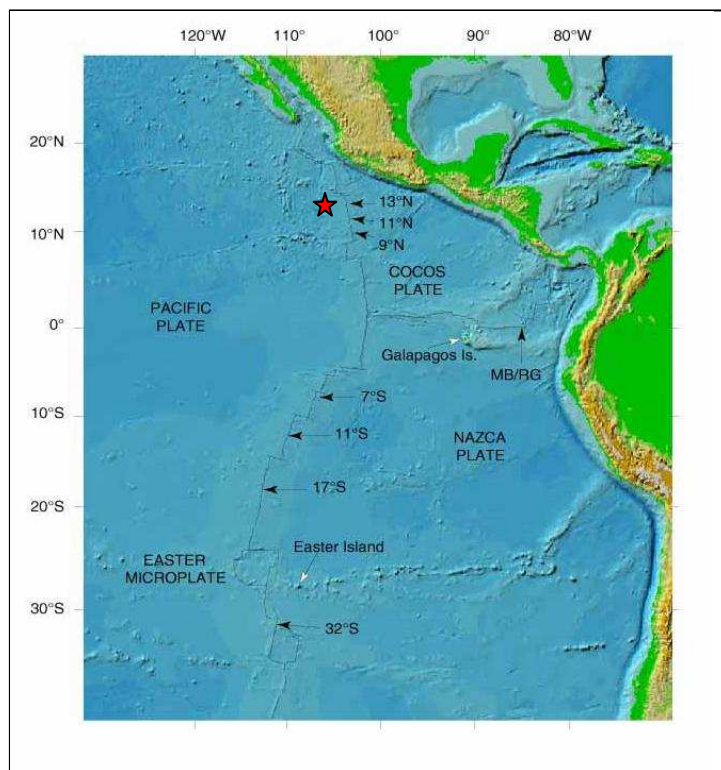


Figure 18: Localisation géographique des sites hydrothermaux de la ride Est Pacifique. *L'étoile rouge représente le site étudié lors de cette thèse ; d'après <http://www.mbari.org>.*

Echantillons prélevés sur la Ride medio-Atlantique (voir figure 17)  
(en bleu : prélèvements d'origine animale)

MARVEL 1997 (Sites Lucky Strike et Menez Gwen)	
MV2E1	Grattage de la face externe d'une coquille de <i>Bathymodiolus</i>
MV2E2	Broyat du corps de <i>Bathymodiolus</i>
MV2E3	Surface de roches ou les modioles n'étaient pas fixées
MV3E1	Matériel minéral + biofilm
MV3E2	Matériel minéral + surface verte
MV3E3	Matériel minéral noir très poreux
MV5E1	Sulfure friable de la roche portant les moules
MV5E2	Grattage de la face externe d'une coquille de <i>Bathymodiolus</i>
MV5E3	Broyat du corps de <i>Bathymodiolus</i>

EXOMAR 2005 (Sites Rainbow, Tag et Lost City)	
Ex6E01	<i>Rimicaris exoculata</i>
Ex6E02	<i>Rimicaris exoculata</i>
Ex6E03	<i>Rimicaris exoculata</i>
Ex6E04	<i>Rimicaris exoculata</i>
Ex6E05	<i>Chorocaris chacei</i>
Ex7E01	<i>Rimicaris exoculata</i>
Ex7E02	<i>Rimicaris exoculata</i>
Ex7E03	<i>Rimicaris exoculata</i>
Ex7E04	<i>Rimicaris exoculata</i>
Ex8E01	Sédiment
Ex8E02	Sédiment
Ex8E03	Morceaux de sulfures de cheminée active
Ex8E04	<i>Rimicaris exoculata</i>
Ex8E05	<i>Rimicaris exoculata</i>
Ex8E06	<i>Rimicaris exoculata</i>
Ex8E07	<i>Rimicaris exoculata</i>
Ex9E1	Fragments de revêtement externe d'un morceau de sulfures recouverts de crevettes
Ex9E2	<i>Rimicaris exoculata</i>
Ex9E3	<i>Bathymodiolus azoricus</i>
Ex9E4	<i>Bathymodiolus azoricus</i>
Ex9E5	<i>Bathymodiolus azoricus</i>
Ex9E6	<i>Bathymodiolus azoricus</i>
Ex9E7	<i>Bathymodiolus azoricus</i>
Ex9E8	<i>Bathymodiolus azoricus</i>
Ex9E9	<i>Bathymodiolus azoricus</i>
Ex9E10	<i>Bathymodiolus azoricus</i>
Ex13E2	Sédiment
Ex13E3	Sédiment
Ex13E4	2 <i>R. exoculata</i>
Ex13E5	3 <i>Chorocaris chacei</i>
Ex13E6	3 <i>Rimicaris exoculata</i> juvéniles
EX14E2	3 <i>Rimicaris exoculata</i> juvéniles
EX14E3	4 <i>R. exoculata</i>
Ex15E1	Morceaux de sulfures
Ex15E2	3 <i>R. exoculata</i>
Ex15E3	2 morceaux de tubes de polychètes

Echantillons prélevés sur la Ride Est Pacifique (voir figure 18)  
(en bleu : prélèvements d'origine animale)

<b>Campagne HERO 1991 (Ride Est Pacifique 13°N)</b>	
H18E1	Morceaux de tube d' <i>Alvinella pompejana</i> broyés
H18E3	Morceaux de tube d' <i>A. pompejana</i> broyés
H18E7	Morceaux d'épiderme dorsale postérieur d' <i>A. pompejana</i> couvert de bactéries filamenteuses
H18E9	eau environnant <i>A. pompejana</i>
H18E11	broyat de tubes d' <i>A. pompejana</i>
H18E12	Broyat d'épiderme d' <i>A. pompejana</i> couvert de bactéries
H21E4	<i>Alvinella pompejana</i>
H21E5	<i>Paralvinella</i>
H22E2	Sécrétions de <i>Paralvinella</i> fixées à la cheminée
H22E5(1)	Tentacules d' <i>Alvinella caudata</i>
H22E5(3)	Partie antérieure d' <i>Alvinella caudata</i>
H22E5(5)	Partie postérieure d' <i>Alvinella caudata</i>
H26E11(1)	Fluides prélevés dans un massif de <i>Riftia pachyptila</i>
H27E2	Fluide de <i>Riftia pachyptila</i>

## **II-2-Analyse de la diversité**

Deux types d'approches complémentaires sont envisageables pour aborder la diversité des champignons : une approche visant à mettre en place des collections d'isolats à partir d'étalements d'échantillons bruts, et une approche indépendante de la culture. L'isolement de souches fongiques présente l'avantage de pouvoir effectuer des analyses physiologiques. Néanmoins, compte tenu de la difficulté d'isoler et de maintenir en culture des souches fongiques à partir de ces environnements pour aborder la question de la diversité fongique, une approche moléculaire indépendante de la culture semble judicieuse. Les différentes étapes permettant l'analyse de cette diversité sont présentées dans ce chapitre, à partir de l'extraction de l'ADN des échantillons environnementaux jusqu'aux analyses phylogénétiques. Les études sur les isolats ont été réalisées en collaboration avec le laboratoire dirigé par Georges Barbier (ESMISAB, Brest, France).

### **2-1 Extraction de l'ADN génomique total**

Après une batterie de tests effectués en utilisant différents protocoles et différents kits disponibles, il a été décidé sur la base des résultats (rendement et amplificabilité de l'ADN extrait) que toutes les extractions d'ADN seraient réalisées à l'aide du kit d'extraction Dneasy plant mini kit (Qiagen), en respectant le protocole proposé par le fournisseur. Les extractions ont été effectuées à partir de 120 µl d'échantillons liquides ou 100 mg d'échantillons solides. Les échantillons additionnés d'un mélange de billes de verre ou d'une bille de tungstène ont été broyés par vibro-broyage (30 secondes, 30 tours/secondes) (Retsch).

### **2-2 Amplification du gène codant l'ARNr 18S**

L'amplification de la petite sous unité du gène codant l'ARNr 18S a été effectuée à l'aide de différents jeux d'amorces. Le jeu d'amorces MH2/MH4 a été privilégié car il a précédemment montré son efficacité pour amplifier un grand nombre de séquences fongiques (Vandenkoornhuyse & Leyval 1998). Cependant, quand aucune amplification n'a été obtenue par ce jeu, un second jeu d'amorces, EF3/EF4, modifié à partir de Smit *et al.* (1999) a été utilisé (les compositions sont données ci-dessous). Les réactions d'amplification du gène cible sont réalisées dans les conditions suivantes, sur un thermocycleur PTC-100 (MJ Research).

Pour le jeu d'amorces MH2 (5'TTCGATGGTAGGATAGAGG3') / MH4 (5'GTCTCACTAAGCCATTC3'), le mélange réactionnel d'amplification pour un volume final de 25 µl contient, 2 µL d'ADN environnemental extrait, 16.375 µl d'eau ultra pure ; 2.5 µl de tampon de PCR 10X contenant MgCl<sub>2</sub> 15 mM (Applied Biosystems), 2 µl de déoxyribonucléotides-triphosphate (dNTPs) (1,25 mM) (dATP, dCTP, dGTP et dTTP); 1 µL d'amorce sens (MH2, 1 pmol/µl) ; 1µl d'amorce anti-sens (MH4, 1 pmol/µl) ; 0.125 µl de *Taq*-ADN polymérase (5 U/µl) (Applied Biosystems).

Le cycle d'amplification consiste en une première étape de dénaturation de l'ADN à 94°C pendant 1 min 30, suivi d'un cycle de 30 sec à 94°C, 1 min 15 sec à 48°C (avec une diminution de 0.1°C à chaque cycle) et 1 min 30 sec à 72°C. Ce cycle est répété 37 fois. Une étape d'élongation finale de 10 min à 72°C achève l'amplification par PCR.

Les résultats de l'amplification sont révélés par électrophorèse (110V durant 20 minutes) sur un gel d'agarose (Sigma) à 1% réalisé dans du TBE 0,5 x [solution stock TBE 10X : 108 g de Tris Base, 55g d'acide borique, 20 mL EDTA à 0.5M, pH 8.3, qsp 1 L].

Le gel est ensuite immergé dans un bain de bromure d'éthidium (0,5 mg/L, Sigma) pour révéler les amplicons aux U.V.

Pour le second jeu d'amorces EF3 (5'TCCTCTAAATGACCAGTTTG3') et EF4 (5'GGAAGGGNTGTATTTATTAGAT3'), les conditions d'amplification sont les mêmes à l'exception de la température d'hybridation des amorces élevée à 53°C.

## **2-3 Clonage de l'ADN fongique**

### **2-3-1 Purification**

Après amplification, les fragments d'ADN fongiques ont été purifiés avant d'être clonés. Toutes les purifications ont été effectuées à l'aide du kit High Pure PCR product Purification kit (Roche Applied Science) selon les instructions du fournisseur.

### **2-3-2 A-Overhanging**

L'étape de « A-Overhanging » consiste à greffer à l'ADN purifié, une déoxy-Adenosine à l'extrémité 3'. Le vecteur plasmidique utilisé pour cloner le fragment dispose quant à lui d'un déoxy-Thymidine à son extrémité 5'. Cette étape permet donc d'améliorer l'étape de ligation ultérieure, en particulier lorsque le produit PCR est stocké au delà de quelques heures. Un mélange réactionnel comprenant 1 U *Taq* polymérase, 0.5 µl d'ATP



(0.125 mM), 1  $\mu$ l de Tampon Taq, 5  $\mu$ l d'ADN purifié, 3.3  $\mu$ l d'eau ultrapure, a été placé 15 minutes à 72°C.

### 2-3-3 Ligation dans pGEM-T

La ligation dans le plasmide pGEM-T (Promega) a été effectuée selon les recommandations du fournisseur (la carte du plasmide est présentée dans la figure 19). Un mélange contenant 5  $\mu$ l de tampon de ligation 2X, 1  $\mu$ l de plasmide pGEM-T linéaire, 1  $\mu$ l de T4 DNA ligase, 1.5  $\mu$ l de produits PCR (après étape de A-overhanging), et 1.5  $\mu$ l d'eau ultrapure, a été réalisé, puis incubé une nuit à 4°C.

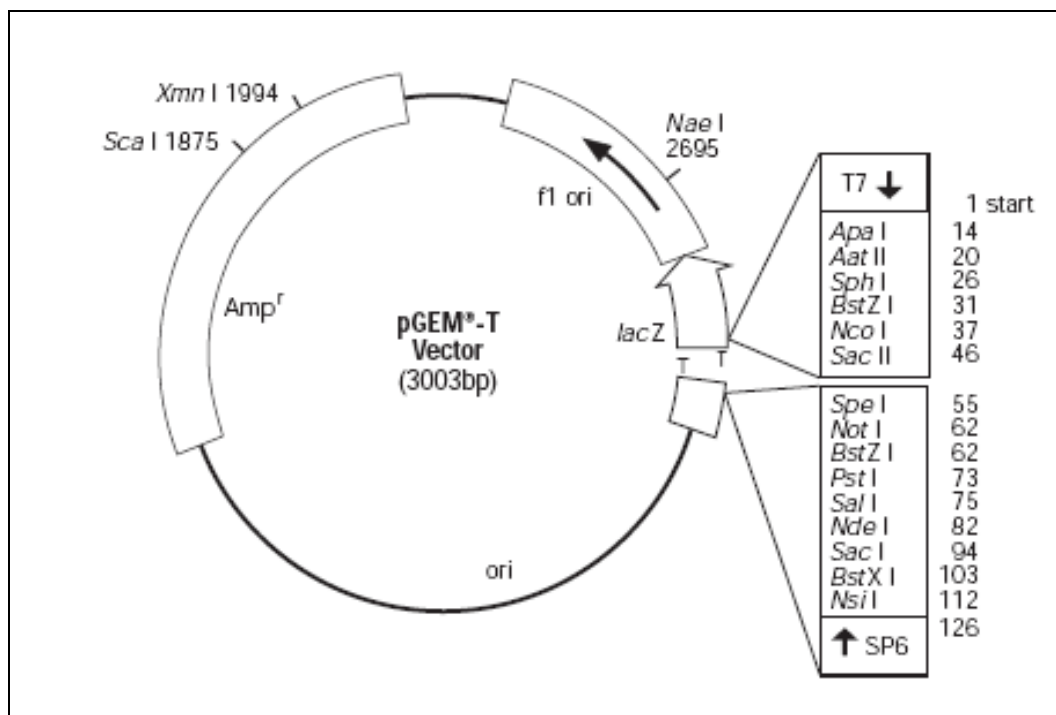


Figure 19: Carte circulaire du plasmide pGEM-T (Promega).

### 2-3-4 Transformation par choc thermique de la souche d'E.coli DH5- $\alpha$

30  $\mu$ l de suspension de cellules compétentes d'*Escherichia coli* DH5- $\alpha$  (Invitrogen) ont été additionnées à 3  $\mu$ l de produits de ligation, sur glace fondante, puis mélangées

doucement à l'embout de la pipette, avant d'être incubées 30 minutes sur glace fondante. L'intégration du vecteur plasmidique dans les bactéries compétentes a été effectuée par choc thermique : le mélange est incubé 1 minute 15 secondes à 42°C, puis est placé 2 minutes sur glace. 250 µl de milieu nutritif SOC (tryptone 2.0; extraits de levure 0.5% ; NaCl 10 mM ; KCl 2.5 mM ; MgCl<sub>2</sub> 10 mM ; MgSO<sub>4</sub> 10 mM et glucose 20 mM) additionnés d'1 µl d'ampicilline (50 mg/ml) sont ajoutés et le mélange est incubé à 37°C pendant 3 heures ; durant ces 3 heures les tubes ont délicatement été retournés à 3 reprises. Au terme des 3 heures d'incubation, 70 µl et 140 µl de suspensions bactériennes sont déposés sur gélose de milieu LB [milieu Luria bertani : tryptone 1.0%; extraits de levure 0.5% ; NaCl 1%; agar bactériologique :15 g/L , auxquels est ajoutée de l'ampicilline (50 mg.mL<sup>-1</sup>), 1ml/L de solution de X-Gal (200 mg/10 ml de Di MéthylFormamide) et 1 ml/L de la solution d'IPTG (100 mM)]. Les étalements sont incubés une nuit à 37°C. Les cultures sont alors stockées au réfrigérateur en attente du criblage des clones positifs.

### **2-3-5 Criblage des clones bactériens**

Les clones positifs (qui apparaissent blancs par le phénomène d'  $\alpha$ -complémentation) sont piqués à l'aide d'un embout stérile de pipette (0.5-10 µl), pour l'inoculation de 100 µl de LB + ampicilline (1 ml/L d'une solution de 50 mg/ml) contenus dans une cupule de plaque de culture cellulaire (IWAKI). Les plaques contenant chacune 96 clones différents sont incubées pendant 16h à 37°C, sous agitation (Eppendorf Thermomixer).

Le criblage des clones cultivés s'effectue par PCR en utilisant les amorces T7 (5'-TAATACGACTCACTATA-3') et SP6 (5'-TATTTAGGTGACACTATAG-3') (Operon), portées sur le plasmide et flanquant l'insert (figure 19).

Les PCR ont été effectuées dans un volume final de 25 µl contenant 0,125 µl de *Taq*-ADN polymérase (5 U/µl) (Applied Biosystems), 2.5 µl de tampon *Taq*-ADN polymérase 10X (Applied Biosystem), 2 µl de dNTPs (1,25 mM) , 0.6 µl d'amorces (1 pmol/µl) et 16.375 µl d'eau stérile.

Le cycle d'amplification utilisé est le suivant : 1 min 30 à 94°C, puis 35 cycles composés de : 30 sec à 94°C, 1 min 15 à 55°C, 1 min à 72°C. L'amplification s'achève par une étape d'élongation finale de 10 min à 72°C.

5 µl d'amplicons sont ensuite déposés sur gel d'agarose puis visualisés aux UV de façon à distinguer les vrais positifs (amplification de taille attendue). Les amplicons positifs sont purifiés (High Pure PCR product Purification kit ; Roche Applied Science) avant séquençage.

### **2-3-6 Séquençage**

Les fragments de taille attendue sont séquencés (plateforme de séquençage, Ouest Genopôle ; Station Biologique de Roscoff, Finistère, France).

Le séquençage se fait selon la méthode « BigDye terminator V3.1 » (Applied Biosystems) en utilisant comme amorce de séquençage d'ADN, soit T7 soit SP6 respectivement pour le brin sens et le brin anti-sens (séquençage double sens). Chaque réaction est effectuée dans un volume final de 5 µl avec 0,5 µl de mélange de séquençage dilué au 1/8<sup>ème</sup> (0,125X), 0,75 µl de tampon de séquençage (Tris HCl 400 mM pH 9 ; MgCl<sub>2</sub> 5 mM) à la concentration finale de 1X, 1 µl d'amorce à 5 µM et entre 50 et 100 ng d'insert amplifié. Les réactions sont réalisées sur un Robocycleur Gradient 96 (Stratagene) selon le programme suivant : une dénaturation de 5mn à 96°C, suivie de 50 cycles de dénaturation (96°C pendant 30s), d'hybridation des amorces (54°C pendant 30secondes) et d'élongation (60°C pendant 4mn).

Enfin, le produit de réaction de séquençage est purifié dans 20 µL de solution d'injection pour éliminer les nucléotides fluorescents restants et analysés sur un séquenceur automatique 16 capillaires ABI Prism 3100 (Applied Biosystems). De façon systématique, tous les fragments d'ADN étudiés ont été séquencés dans les 2 sens. Les 2 brins d'ADN ont dans un second temps été assemblés en contig puis vérifiés pour les nucléotides présentant des ambiguïtés (Sequencher 4.8 ; Gene Codes Corporation).

### **2-3-7 Détection des chimères**

Toutes les séquences obtenues sont comparées à des bases de données, afin d'exclure tout fragment provenant d'associations chimériques, en utilisant CHIMERA\_CHECK 2.7 de la base de données Ribosomal Database Project (<http://rdp.cme.msu.edu/html/analyses.html>).

### **2-3-8 Effort d'échantillonnage**

L'effort d'échantillonnage est calculé en produisant des courbes de raréfaction à partir des séquences d'ADN, par procédure de ré-échantillonnage. Ces courbes ont été produites par le logiciel EstimateS (Colwell, 2006). Pour cela, pour un échantillon donné, une matrice de présence/absence de phylotypes (seuil estimé à 97% de similarités) est établie.

### 2-3-9 Soumission des séquences à Genbank

Les fragments de gènes codant pour l'ARNr18S obtenus ont été intégrés dans la base de données Genbank en utilisant le logiciel Sequin, qui permet de soumettre des batchs de séquences. La séquence correspondant au plasmide a dans un premier temps été retiré en utilisant VecScreen, pour toutes les séquences. Les numéros d'accèsion attribués sont les suivants :

Echantillon	Numéro d'accèsion Genbank
MV2E1	EF638466 - EF638515
H18E12	EF638516- EF638564
MV2E2	EF638565 - EF638614
MV2E3	EF638615- EF638636
MV5E1	EF638637- EF638654
MV5E2	EF638655- EF638686
Isolats	EF638687 - EF638706

### 2-3-10 Culture des isolats fongiques

En complément de nos études de diversité moléculaires, indépendantes de la culture, des cultures d'isolats fongiques ont été établies. Ces travaux cultureux ont été réalisés au laboratoire d'Ecologie Microbienne de Georges Barbier à l'ESMISAB (Brest, France), essentiellement par Gaétan Burgaud et Georges Barbier. Les différents milieux d'enrichissement et les techniques utilisées pour isoler ces souches seront présentés dans la partie Matériel et Méthodes de la dernière publication présentée dans cette thèse (partie VI).

## **2-4 Analyse phylogénétiques**

### **2-4-1 Phylogénie, définition**

Il existe différentes méthodes phylogénétiques s'appuyant sur la comparaison de caractères spécifiques. Seules les méthodes de phylogénie moléculaire seront abordées ici. Ces méthodes moléculaires s'appuient sur la comparaison de séquences d'ADN de gènes définis, appelés marqueurs phylogénétiques. Pour pouvoir être considéré comme un bon marqueur phylogénétique, un gène doit posséder les caractéristiques suivantes :

- L'ubiquité : le marqueur doit être présent chez tous les organismes comparés.
- L'orthologie : la similitude observée entre les organismes doit être due à une descendance commune ; le marqueur ne doit pas être soumis à des pressions de transferts horizontaux.
- La conservation : le marqueur doit posséder des régions comparables, avec des taux d'évolutions proches entre organismes.

### **2-4-2 Alignement**

Dans le cas des phylogénies moléculaires, la première étape après le choix du marqueur est l'alignement des séquences nucléotidiques.

Le logiciel ClustalX 2.0 (Larkin *et al.*, 2007) a été utilisé lors de cette thèse pour aligner les séquences d'ADN de champignons selon le principe d'alignement multiple global. Les méthodes de reconstructions phylogénétiques reposent toutes sur l'alignement des séquences obtenu ; un alignement de qualité est donc absolument indispensable : ainsi, une optimisation manuelle de l'alignement en complément de l'alignement logiciel est systématiquement adoptée.

### **2-4-3 Méthodes de reconstructions phylogénétiques**

Nous avons vu que l'alignement des séquences nucléotidiques était la clé de voûte de toute analyse phylogénétique. Ces analyses reposent sur différentes méthodes qu'il convient généralement d'appliquer pour vérifier la congruence des analyses et la validité des noeuds.

#### ***2-4-3-1 Méthode phénétique***

Les méthodes phylogénétiques basées sur les distances entre les séquences d'ADN (analyses phénétiques) sont de préférence associées à un algorithme d'agglomération de

Neighbor Joining (Saitou & Nei, 1987) ayant l'avantage d'autoriser des vitesses d'évolution variables entre terminaux (hypothèse d'horloge moléculaire relaxée).

On admet dans ce cas que la distance évolutive entre 2 séquences d'ADN est égale au nombre de substitutions qui se sont produites entre 2 lignées évolutives depuis l'ancêtre commun. L'analyse se fait à partir de l'alignement multiple généré par ClustalX à partir duquel une matrice de distance est calculée pour chaque paire de séquences : on obtient donc une matrice de toutes les distances entre chaque paire de séquences. Dans un second temps, une agglomération séquentielle est effectuée en calculant les matrices en étoile intermédiaires.

Une fois les deux espèces les plus proches regroupées, une nouvelle matrice de distance est calculée, et à nouveau les terminaux les plus proches sont rapprochés et ainsi de suite jusqu'à ce que toutes les séquences aient été intégrées dans l'arbre.

Nous avons utilisé le modèle de Kimura à 2 paramètres (K2P) lors de toutes nos reconstructions phénétiques ; dans ce modèle les transitions ( $\alpha$ ) et les transversions ( $\beta$ ) ne sont pas équiprobables ( $\alpha \neq \beta$ ).

Les distances sont calculées selon la formule :

$$d_{K2P} = \frac{1}{2} \ln(1/(1-2P-Q)) + \frac{1}{4} \ln(1/(1-2Q))$$

*avec P et Q les fréquences de transition/transversion*

L'un des inconvénients majeurs de ces méthodes phénétiques est le bruit qu'induisent les phénomènes d'homoplasies (similitudes non héritées), qui seront confondues avec les homologies vraies. Or, pour étudier les relations phylogénétiques entre organismes, repérer uniquement les caractères homologues s'avère crucial afin de distinguer l'état primitif (plésiomorphe) de l'état dérivé (apomorphe). Pour cela on utilisera des méthodes basées sur les caractères, comme les méthodes de parcimonie et de vraisemblance que nous allons introduire ci-après.

#### ***2-4-3-2 Calcul des valeurs de bootstrap***

Le calcul de la valeur de bootstrap est utilisé pour valider la robustesse des nœuds d'un arbre. Un pseudo ré-échantillonnage (tirage avec remise des colonnes de la matrice) est effectué à partir de l'alignement initial pour produire une matrice artificielle. Autant de matrices artificielles que de répliquats de bootstrap sont générées ; à partir de chacune d'entre elles un arbre est calculé : la valeur de bootstrap correspondra au nombre d'itérations où un nœud est trouvé en commun entre les topologies.

### ***2-4-3-3 Méthode de Maximum de Parcimonie***

Cette méthode, comme son nom l'indique, va privilégier l'arbre qui minimise le nombre de pas évolutifs. Le nombre global de mutations est obtenu en faisant la somme des mutations le long de chaque branche ; tous les changements sont ainsi comptés entre chaque séquence et l'arbre Maximum de Parcimonie comportera le nombre minimum d'événements évolutifs. Cette construction va se faire contre un groupe externe (outgroup). Le choix de l'outgroup dans cette méthode est important : le groupe externe doit être une séquence si possible primitive au groupe d'organismes étudiés, qui ne doit pas être ni trop proche (afin d'augmenter la résolution de l'analyse), ni trop éloignée des autres séquences afin d'éviter les biais et artefacts de reconstruction (comme le phénomène d'attractions des longues branches par exemple).

La méthode de Maximum de Parcimonie repose sur une triple hypothèse de base :

- le nombre de changements évolutifs entre les branches doit être minimum.
- les sites évoluent de manière indépendante.
- la vitesse évolutive est lente et constante au cours du temps.

L'analyse suit le déroulement suivant :

- une recherche des sites informatifs est effectuée à partir de l'alignement multiple des séquences. Un site est informatif s'il y a au moins deux types de nucléotides présents dans ce site et si chacun d'eux est représenté dans au moins deux des séquences comparées.
- toutes les topologies possibles sont analysées (algorithme exact)
- le nombre minimum de substitutions est calculé pour chaque arbre en faisant la somme des changements requis pour expliquer la topologie.
- l'arbre dont la topologie implique le minimum de changements est conservé : c'est l'arbre le plus parcimonieux.

La recherche de l'arbre parcimonieux se fait en utilisant des algorithmes exacts (pour un nombre de terminaux inférieur à 30), ou des algorithmes approchés pour des jeux de données comprenant plus de 30 lignes : on parle alors d'analyse heuristique. L'analyse heuristique va permettre de rechercher un arbre s'approchant de l'optimal. Pour l'analyse heuristique, différentes stratégies d'analyse de réarrangements des branches existent. Nous avons choisi d'utiliser la procédure de réarrangement la plus profonde qui consiste à tester différentes configurations de repositionnement d'un sous-arbre d'une topologie initiale contre l'autre sous arbre et réciproquement (procédure TBR : Tree Bisection Recognition).

Afin d'optimiser cette procédure, des pondérations peuvent être ajoutées, comme la pondération des transitions par rapport aux transversions par exemple. Les caractères portant peu de changements sont également plus lourdement pondérés que des sites très variables.

La méthode de Maximum de Parcimonie a été appliquée aux jeux de données lors de cette thèse, en utilisant PAUP 4.0  $\beta$  10 (Swofford, 1991).

Les détails du script de la procédure sont les suivants.

**Script de la procédure : Maximum de parcimonie(MP)**

*#Le fichier d'alignement est ouvert dans PAUP au format nexus#*

**outgroup nom entier de la séquence choisie**

**hsearch multrees=no addseq=random nreps=100 swap=tbr**

**lscore 1-N/wts=ignore nst=2 tratio=est rates=gamma #pour N arbres sauvegardés#**

*#Fonction qui calcule les valeurs de probabilité des arbres équiprobables dans MP pour savoir quelle est réellement la meilleure solution). →La valeur absolue la plus faible correspond à la meilleure valeur de vraisemblance #*

**savetrees file=nom de la matriceMP.tre format=nexus brlens=yes**

**OPTION :**

**reweight basewt=100 index=RC fit=maximum**

*#pour recalculer la MP en pondérant les caractères. Si plusieurs topologies obtenues alors recalculer le lscore#*

**Script de la procédure pour calculer les valeurs de Bootstrap**

**hsearch multrees=no addseq=random nreps=15 swap=tbr**

**bootstrap nreps=500 treefile= nom de la matriceMPBoot.tre**

***2-4-3-4 Méthode de Maximum de Vraisemblance***

Cette méthode évalue l'hypothèse d'une histoire évolutive, en termes de probabilités pour un modèle de distance donné. L'analyse Maximum de Vraisemblance (Maximum Likelihood ; ML) produit le meilleur arbre sous l'hypothèse du modèle de distance choisi. La vraisemblance est calculée en additionnant le log des vraisemblances pour chaque site



(calculés par une méthode probabiliste), qui donne ainsi le log de vraisemblance de l'arbre entier (plus facile à manipuler).

Le modèle choisi par rapport au jeu de données initial est primordial car les probabilités calculées vont dépendre du modèle évolutif choisi. Ce modèle sera choisi sur la base des jeux de données initiaux ; le modèle le plus approprié sera choisi en utilisant ModelTest (Posada & Crandall, 1998).

L'utilisation de la méthode ML repose sur une triple hypothèse de base :

- la probabilité d'un changement est indépendante de l'histoire antérieure du site.
- les probabilités de substitutions ne changent pas avec le temps ou entre les arbres.
- les changements sont réversibles dans le temps.

Il est possible d'utiliser des modèles à plusieurs paramètres, qui autorisent ainsi une correction pour la proportion de sites pouvant changer, une correction pour la pondération des transitions/transversions, une variation intersites pouvant suivre une distribution statistique.

Lors de nos analyses, le modèle GTR +  $\Gamma$  + I (General Time Reversible) a été très largement privilégié, car correspondant le mieux à nos jeux de données selon Modeltest: c'est un modèle à 12 paramètres et 4 fréquences qui regroupe les caractéristiques suivantes :

- les 4 bases ont des fréquences différentes
- il existe 6 types de mutations réciproques (AC, AT, AG, CT, CG, TG)

La variation de substitutions entre sites est corrigée par la distribution Gamma (paramètre  $\Gamma$ ). Un paramètre  $\alpha$  est calculé à partir du jeu de données, et va correspondre au taux de variations intersites (plus  $\alpha$  est élevé, plus la variation sera faible). De plus, certains sites peuvent être considérés comme invariables (paramètre I). Plus un modèle est saturé en informations, plus le temps de calcul de maximum de vraisemblance sera long.

### **Choix du modèle :**

Le fichier d'alignement au format nexus est analysé par le logiciel Modeltest 3.7 (Posada & Crandall, 1998), qui teste plus de 60 modèles évolutifs sur le jeu de données initial ; à l'issue de cette procédure le modèle qui correspond le plus aux données sera choisi et utilisé dans la suite de la procédure.

## Script MODELTEST

### Dans PAUP :

Charger le fichier aligné

Charger le fichier modelblockPAUPb10

*#Le calcul commence aussitôt : 2 fichiers sont générés #*

« *modelfit.log* » : ce sont les résultats calculés pour chacun des 56 modèles

« *modeltest.scores* » : fichier récapitulatif des différentes valeurs obtenues qui vont servir au classement des différents modèles avec "Modeltest3.7.win.exe". Prévoir de le coller dans le même dossier que "Modeltest3.7.win.exe" pour la suite.

### Dans Modeltest3.7.win.exe :

Ouvrir Modeltest3.7.win.exe sous DOS :

Dans démarrer, faire exécuter et taper cmd

Descendre l'arborescence des fichiers jusqu'à Modeltest3.7.win.exe

Lancement des tests sur les différents modèles :

Taper : Modeltest3.7.win.exe<modeltestscores>results.txt puis Entrée

→ Le fichier "*results.txt*" est généré aussitôt dans le dossier bin

## Script de la procédure : Maximum Likelihood. (ML)

*#Le fichier d'alignement est ouvert dans PAUP au format nexus#*

**set criterion=parsimony**

**outgroup *nom entier de la séquence choisie***

**hsearch multrees=no addseq=random nreps=100 swap=tbr**

**showtree**

**describetrees 1/plot=phylogram brlens=yes**

**savetrees file=*nom de la matrice*MP2.tre replace**

**set criterion=distance**

**dset distance=gtr**                    *#Modèle GTR, le plus saturé, choisi par le ModelTest#*

**showdist**

**NJ**

**dset objective=lsfit power=2**

**hsearch**

```
gettrees file= nom de la matriceMP2.tre
lscore 1/wts=ignore nst=2 tratio=est rates=gamma
shape=estimate
set criterion=likelihood
lset tratio=previous
shape=previous
hsearch addseq=asis
savetrees file= nom de la matriceML format=nexus brlens=yes
bootstrap nreps =50
```

## **II-3 Quantification du nombre de copies du gène codant l'ARNr18S par PCR quantitative**

Nous avons abordé dans la partie précédente les différentes méthodes que nous avons utilisées pour analyser la diversité fongique par des approches moléculaires. Nous avons choisi d'effectuer une analyse par PCR quantitative (qPCR) des marqueurs fongiques, afin de connaître le nombre de copies de gènes codant l'ARNr fongique dans les échantillons ayant présenté un signal positif à l'étape précédente ; cette quantification nous permettra de plus d'aborder et d'infirmer les éventuels problèmes liés à une contamination extérieure de nos échantillons. Finalement, l'analyse de PCR quantitative sera utilisée pour choisir l'échantillon utilisé pour l'étude métagénomique : ce dernier doit posséder une faible diversité spécifique reliée à un fort nombre de copies de gènes codant l'ARNr18S.

### **3-1 Dessin des amorces**

A la différence des analyses phylogénétiques nécessitant d'amplifier des fragments longs d'ADN, la qPCR n'est efficace que sur des amplicons d'une taille inférieure ou égale à 300 pb. Dans ce but, la première étape a donc été de dessiner des amorces permettant une amplification d'un fragment de 300 pb du gène codant l'ARNr18S, dans des régions suffisamment conservées et spécifiques des champignons. Pour cela, des alignements du gène codant l'ARNr18S générés par la base de données PHYMYCO-DB ont été utilisés (voir partie 4 de Matériel et Méthodes), afin de déterminer les régions conservées. L'amorce MH2 utilisée précédemment, a été utilisée comme point d'ancrage sur l'alignement ; la seconde amorce antisens a donc été dessinée à 300 pb de bases environ de MH2. Afin d'augmenter la probabilité d'amplifier sélectivement le gène codant l'ARNr18S fongique, plusieurs jeux d'amorces ont été dessinés à l'aide du logiciel Beacon Designer, puis comparés. Cette comparaison implique une utilisation couplée des différents jeux d'amorces par PCR classique, en utilisant des gradients de températures. Cette analyse a permis de sélectionner le jeu d'amorces approprié ainsi que la température d'hybridation optimale qui ont été utilisés par la suite lors de la qPCR.

Le jeu d'amorces choisi est le suivant, et permet l'amplification de fragments de 280 pb :  
MH2 : 5'TTCGATGGTAGGATAG 3' ; FungqPCR1 : 5'TGTCGGGATTGGGTAATTT 3'.

### **3-2 Calcul de la gamme**

La seconde étape consiste à établir une gamme de standards. Cette gamme servira de référence pour la quantification du nombre de copies du gène codant l'ARNr18S dans nos échantillons, en permettant le calcul de l'efficacité de la réaction. Pour cela, nous avons étalonné la gamme à partir des plasmides réalisés contenant un fragment du gène codant l'ARNr18S.

#### **3-2-1 Extraction et linéarisation des plasmides**

Des clones bactériens transformés précédemment ont été relancés en culture à 37°C en milieu LB + Ampicilline, une nuit. Les plasmides et leurs inserts contenus dans les clones ont été séquencés précédemment : on peut ainsi en déduire avec précision la masse molaire exacte du plasmide contenant l'insert.

Les plasmides ont été purifiés par le GenELUTE five minutes Plasmid Miniprep (SIGMA), selon les recommandations du fournisseur. La concentration des plasmides a été déterminée par spectrophotométrie au Nanodrop (Thermo Scientific).

Les plasmides purifiés ont été linéarisés. L'enzyme de restriction *PstI* a été choisie car le plasmide ne possède qu'un seul site de coupure pour cette enzyme. La digestion a été effectuée dans un volume final de 20 µl, en mélangeant de l'eau ultrapure (14.3 µl), au tampon de restriction spécifique à l'enzyme choisi (ici, tampon H, 2 µl), à de la BSA acétylée (0.2 µl), puis à l'ADN plasmidique (3 µl). L'enzyme *PstI* a ensuite été ajoutée (0.5 µl) puis le mélange a été incubé 2 heures à 37°C. Le plasmide linéarisé a été précipité dans 2 volumes d'isopropanol, puis le mélange a été incubé à 2h à -20°C, puis centrifugé à 16000g pendant 20 minutes. Le culot a ensuite été lavé avec 200 µl d'éthanol 70%, puis une seconde centrifugation à 16000g pendant 10 minutes a été effectuée ; le culot séché a été resuspendu dans 10 µl d'eau ultrapure.

### 3-2-2 Calcul du nombre de copies dans les standards

Le calcul du nombre de copies dans les standards implique de connaître la concentration des plasmides linéarisés puis la longueur totale du plasmide (plasmide + insert). Le calcul se fait par la formule suivante :

$$\text{Molécules}/\mu\text{l} = a / (\text{longueur du plasmide} * 660) * 6.022.10^{23}$$

a = concentration du plasmide ( $\mu\text{g}/\mu\text{l}$ )

660 : poids molaire moyen d'un déoxynucléotide

$6.022.10^{23}$  : constante molaire

### 3-2-3 Préparation de la gamme

A partir des concentrations obtenues, la gamme de standard a pu être établie par dilutions successives de facteur 10 partant du standard le plus concentré ( $10^9$ ) vers le moins concentré ( $10^3$ ) dans un volume final de 15  $\mu\text{L}$ .

## 3-3 Quantification

La quantification a été effectuée à l'aide d'un protocole basé sur l'incorporation de SYBR Green par l'ADN double brin.

Les réactions ont été effectuées en triplicat dans des tubes optiques (Biorad) dans 10  $\mu\text{l}$  final, comme suit : 5  $\mu\text{l}$  de SYBR Green Supermix 2X (Biorad) qui contient 2.7  $\mu\text{M}$  de chaque dNTPs et 0.42 U d'iTaq DNA polymerase (Biorad) ; 0.3  $\mu\text{l}$  d'amorces sens (0.35  $\mu\text{M}$ ) ; 0.3  $\mu\text{l}$  d'amorces antisens (0.35  $\mu\text{M}$ ) et 1.5  $\mu\text{L}$  d'ADN. Les mélanges ont été effectués automatiquement, à l'aide du robot pipeteur/diluteur epMotion 5075 (Eppendorf). Les réactions de qPCR ont été effectuées dans un Chromo 4 thermocycler (MJ Research), en triplicats internes, et par 2 fois (réplicats indépendants).

Le programme qPCR suivant a été choisi :

- 1- 1 minute à 94°C
- 2- 30 secondes à 94°C

- 3- 30 secondes à 48°C
- 4- 20 secondes à 72°C
- 5- Revenir à l'étape 2, 35 fois
- 6- Lecture de la plaque et traçage de la courbe de fusion de 65°C à 95°C, lecture tous les 0.5°C.

Les résultats de la qPCR ont été analysés à l'aide du logiciel Opticon Monitor (MJ Research), qui permet de comparer les profils d'expression.

La première étape consiste à analyser les courbes de températures de fusion (Melting Curve), qui permet d'observer la présence éventuelle de plusieurs amplicons de différentes tailles, ou la dimérisation éventuelle des amorces (plusieurs pics). Il convient à cette étape de vérifier si la température de fusion observée correspond à celle de l'amplicon et non à celles des amorces. L'efficacité de la qPCR est calculée à l'aide des standards, qui permettent de tracer une droite (figure 20), « nombre de copies en fonction des Ct » (Cycle Treshold, cycle seuil qui correspond au nombre de cycles de la PCR pour que la fluorescence dépasse un seuil de détection), dont l'équation donnera la valeur de l'efficacité, selon la formule :

$$Q_n = Q_0 \cdot E^n$$

Q : quantité d'ADN

n : nombre de cycle

Q<sub>0</sub> : Quantité d'ADN initiale

E : Efficacité de la réaction, calculée à partir de l'équation de la gamme des standards.

Une quantification relative absolue et relative a été effectuée pour les échantillons étudiés.

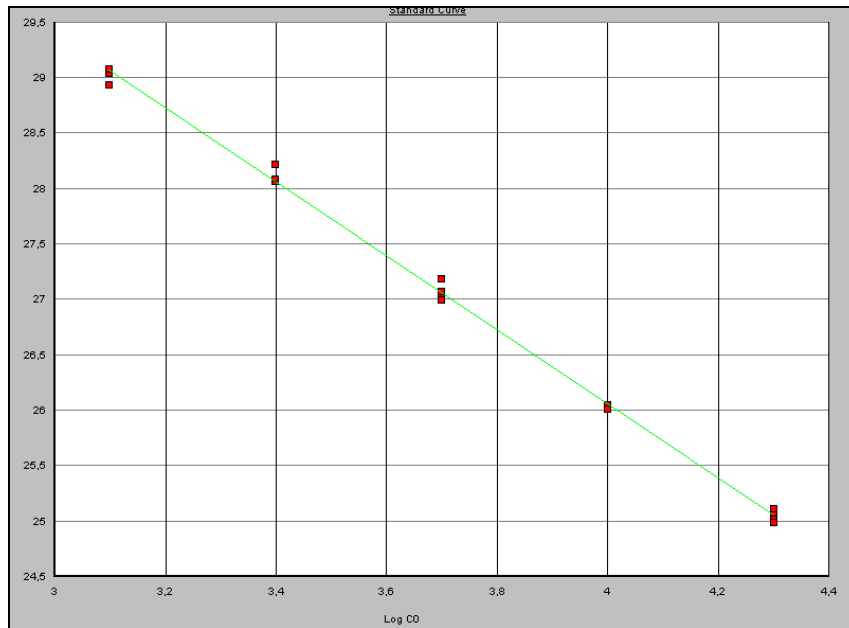


Figure 20: Exemple de courbe standard obtenue avec la gamme de standard.

en abscisse : log du nombre de copies dans le standard ;  
 en ordonnée : Ct (Cycle Treshold). L'efficacité de la qPCR est calculée en fonction de la pente de cette droite. *Les points de dilution de standards les plus concentrés ont été retirés avant l'analyse.*



## II-4 Création d'une base de données de marqueurs moléculaires de champignons

Face au nombre considérable de données moléculaires dans les banques de données mondiales de type Genbank, et le caractère artefactuel d'une fraction des séquences présentes dans ces banques, nous avons créé une base de données personnelle. Cette base de données, nommée PHYMYCO-DB (figures 21, 22, 23) permet de récupérer des séquences nucléotidiques de champignons provenant d'organismes cultivés mais également de séquences provenant d'études environnementales. Elle peut être utilisée en complément d'autres outils disponibles en ligne s'intéressant également aux séquences fongiques, comme par exemple ARB (Ludwig *et al.*, 2004), qui regroupe les séquences de gènes codant l'ARNr 18S uniquement pour les souches fongiques cultivées, et UNITE (Kõjalg *et al.*, 2005), utilisée pour l'identification d'organismes ectomycorhiziens à partir des séquences ITS (Internal Transcribed Spacer). PHYMYCO-DB contient pour l'instant les gènes codant l'ARNr18S et le facteur d'élongation EF1-alpha. La mise en place de cet outil bioinformatique s'est réalisée dans le cadre d'une collaboration avec l'IRISA (Institut de Recherche en Informatique et Systèmes Aléatoires, Rennes, France). Cette base de données se présente sous la forme d'un site web, disponible à l'adresse suivante :

<http://phymycodb.genouest.org/>

Le choix des marqueurs, notamment le gène codant l'ARNr18S, a été justifié par la masse importante de séquences disponibles dans les bases de données, permettant une comparaison entre un très grand nombre d'organismes, ainsi que par le fait que ce marqueur est jusqu'à présent utilisé dans un grand nombre d'études phylogénétiques.

PHYMYCO-DB version 1 a été programmée en PERL (*les scripts de la version 1 sont fournis en annexe*). La version 2 de PHYMYCO-DB a été construite en utilisant MySQL4, à l'aide de phpMyAdmin (<http://www.phpmyadmin.net/>). Cette seconde procédure de programmation a été choisie pour faciliter le fenêtrage de l'interface, et également pour faciliter la mise en ligne de nouveaux outils pouvant enrichir la base de données. Le site web a été créé en utilisant PHP version 4.3.9 et les langages de programmation AJAX et HTML.

Dans un premier temps, les séquences nucléotidiques fongiques (pour les marqueurs ARNr18S et EF1- $\alpha$ ) ont été extraits à partir de Genbank, en utilisant les requêtes suivantes :

“[Organism] and (ssu|SSUrRNA|SSU rRNA|18SrRNA|18S) not

(16S|mitoch\*|28S|5.8S|ITS|Internal Transcribed Spacer|internal transcribed spacer)”

“[Organism] and (EF1 alpha|EF-1 alpha|EF1-alpha|EF-1alpha|EF-1-alpha|EF1alpha)”

Ces requêtes permettent de récupérer toutes les séquences nucléotidiques des deux marqueurs, en excluant les marqueurs mitochondriaux.

Des filtres automatiques ont été ajoutés à ces requêtes afin d'exclure les séquences inférieures à 700 pb et supérieures à 2500 pb. De plus, les séquences contenant plus de 10 nucléotides indéterminés consécutifs ont été exclues.

Ces premiers filtres automatisés ne se sont pas avérés suffisants pour exclure certaines séquences clairement « fausses » ; des séquences résultant d'une mauvaise annotation, d'erreurs de séquençage, ou d'erreurs d'orientation étaient présentes dans le jeu de données. Un « nettoyage » expert a dû être entrepris. Pour cela, plusieurs alignements regroupant des séquences d'un même phylum, ont été générés. Sur la base de ces alignements, les séquences résultant d'erreurs de séquençage ou ayant une mauvaise annotation ont été retirées. Les séquences composées d'insertions d'homopolymères, possédant un fort taux de nucléotides indéterminés, ou une mauvaise orientation ont également été exclues.

Les séquences nucléotidiques peuvent être extraites à partir du site web, selon la taxonomie proposée par Genbank et sont donc disponibles sous la forme d'une arborescence. Il est ainsi possible de récupérer l'ensemble des séquences à partir d'un niveau taxonomique donné (du phylum jusqu'à l'espèce) (figure 22).

Des alignements effectués par phylum (à l'aide de ClustalW ; Larkin *et al.*, 2007) sont également téléchargeables à partir du site. Un fichier d'alignement par marqueur et par phylum sont ainsi téléchargeables, à l'exception du phylum Ascomycota, constitué de séquences divergentes, et pour lesquelles un alignement par marqueur et par sous phylum a été effectué. Ainsi, 19 fichiers de séquences alignées, optimisés à l'oeil, sont disponibles à partir du site web (figure 23).

Une dernière fonctionnalité a également été incluse dans le site web afin de faciliter l'analyse des données ; il s'agit d'alignements de séquences directement exécutables à partir du site web. Ainsi l'utilisateur, après avoir extrait les séquences à partir de notre interface peut ajouter au besoin ses séquences personnelles, et choisir éventuellement un groupe externe, puis lancer l'alignement. L'utilisateur récupèrera les données par courriel envoyé automatiquement une fois l'alignement terminé.

Une publication sur la mise en place de cette base de données a été réalisée, et sera présentée dans la partie III de cette thèse.

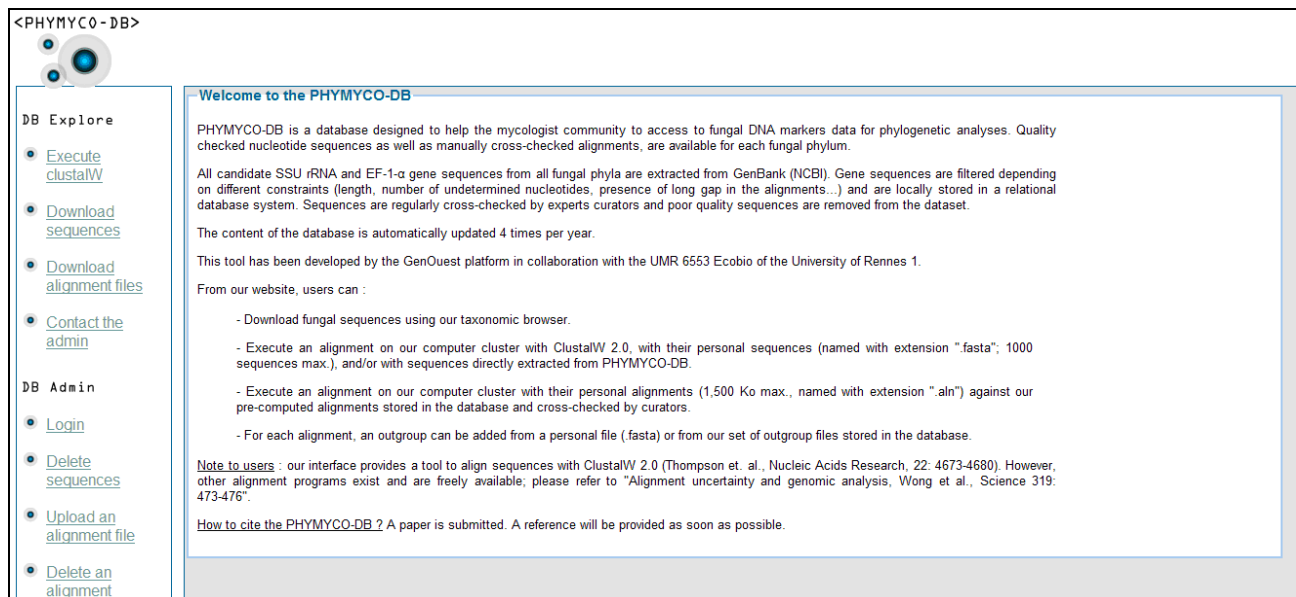


Figure 21: Capture d'écran de la page d'accueil de PHYMYCO-DB.

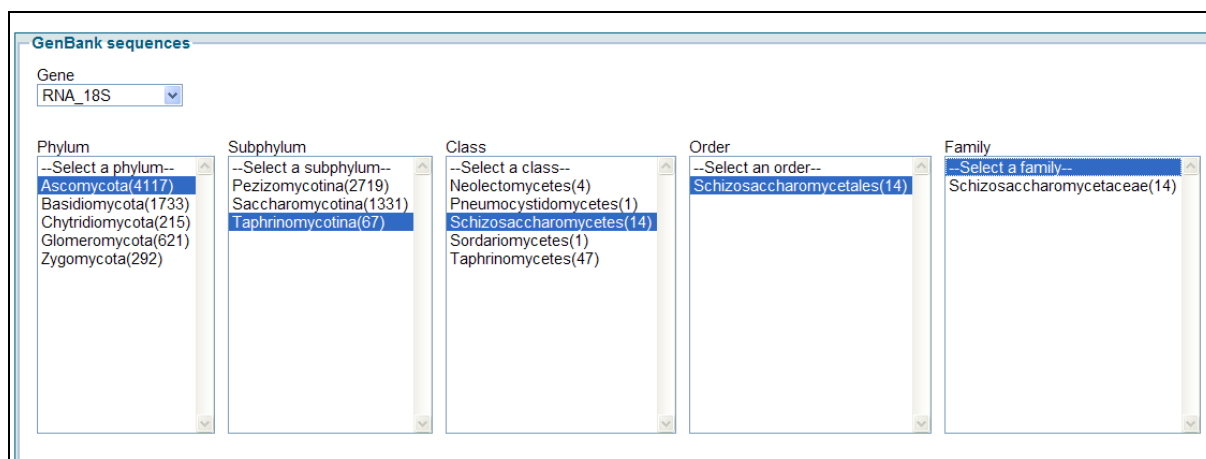


Figure 22: Capture d'écran présentant l'arborescence taxonomique pour extraire les informations de PHYMYCO-DB.

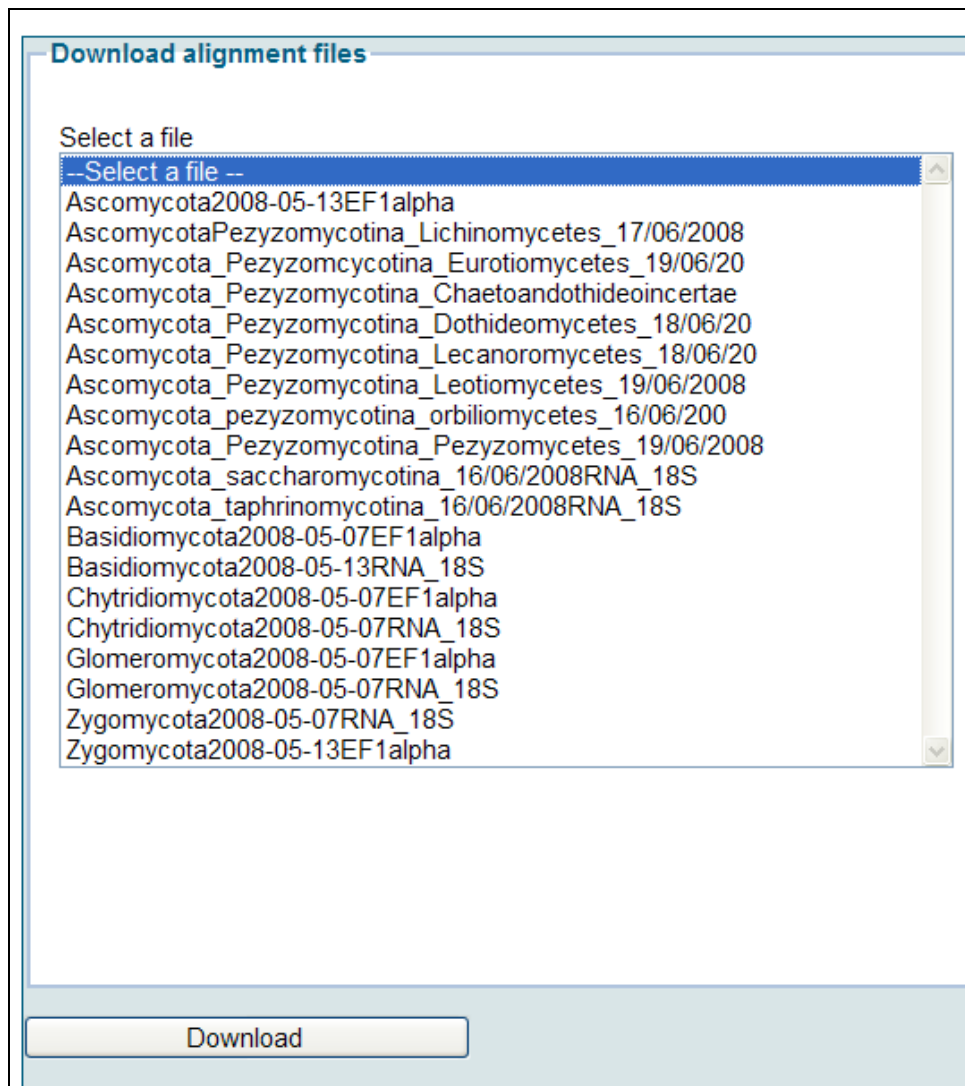


Figure 23: Capture d'écran présentant les fichiers d'alignement téléchargeables à partir de PHYMYCO-DB.

## **II-5 Analyse du métagénome**

### **5-1 Pyroséquençage**

Trois runs de pyroséquençage ont été effectués sur l'échantillon choisi après qPCR et analyse de la diversité, dans le but d'accéder au métagénome de l'échantillon et de le caractériser. Cette étape de pyroséquençage a été effectuée à l'aide du GSFIX (454 technologies, ROCHE, figure 24) au sein des services Recherche et Développement de Roche. Cette méthode de pyroséquençage génère beaucoup d'informations et ne requière pas la construction de BAC ou d'autres formes de banques génomiques de longs fragments. Les différentes étapes du pyroséquençage sont présentées dans les figures 24 et 25.

#### **5-1-1 Préparation de la banque d'ADN**

La banque métagénomique est construite par PCR en émulsion après rupture physique de l'ADN (nébulisation) (figure 24).

#### **5-1-2 Amplification clonale**

L'amplification clonale se fait en émulsion : les fragments individuels d'ADN sont fixés sur des billes. Sur chaque bille, incluse dans un nanoréacteur, se fera l'amplification (dans le nanoréacteur: ADN, amorces, dNTPs, polymérase, substrats). On obtient ainsi une banque d'ADN simple brin (à raison de millions de copies par bille).

Les billes portant l'ADN simple brin (ADN sb) sont déposées sur une lame (Pico Titer Plate, 4.6 millions de puits). Dans chaque puit de la plaque vient se loger une bille. Ainsi, chaque puit contiendra une bille portant l'ADN amplifié clonalement.

En simultanément, on aura un séquençage des fragments génomiques dans plusieurs centaines d'alvéoles de la lame.

#### **5-1-3 Pyroséquençage**

Le pyroséquençage peut se résumer en 4 grandes étapes (figure 24) :

a- Une amorce de séquençage est hybridée à un ADN sb, amplifié par PCR, et incubé avec les enzymes ADN polymérase, ATP sulfurylase, luciférase, et apyrase, ainsi que les substrats Adénosine 5' Phosphosulfate (APS) et la luciférine.

b- Chacun des 4 dNTPs est ajouté séquentiellement. La polymérase catalyse l'incorporation du dNTP complémentaire à celui porté sur le brin d'ADN cible.

Chaque incorporation est accompagnée par un relargage de pyrophosphate (PPi) en quantité équimolaire au taux de nucléotide incorporé.

c- L'ATP sulfurylase convertit quantitativement le PPi en ATP en présence de l'Adénosine 5'Phosphosulfate (APS). Cet ATP produit est utilisé par la luciférase qui transforme la luciférine en oxoluciférine et libère un photon (figure 25). La lumière produite par la réaction catalysée par la luciférine est détectée par une caméra CCD (Charge Coupled Device : Récepteurs à transfert de charge) puis intégrée sous forme de pic dans un pyrogramme à partir duquel est déduite la séquence. C'est donc une méthode de séquençage par synthèse.

d- L'apyrase, une enzyme dégradant les nucléotides, dégrade continuellement les dNTPs non incorporés et l'ATP en excès. Quand la dégradation est complète, un autre dNTP est ajouté.

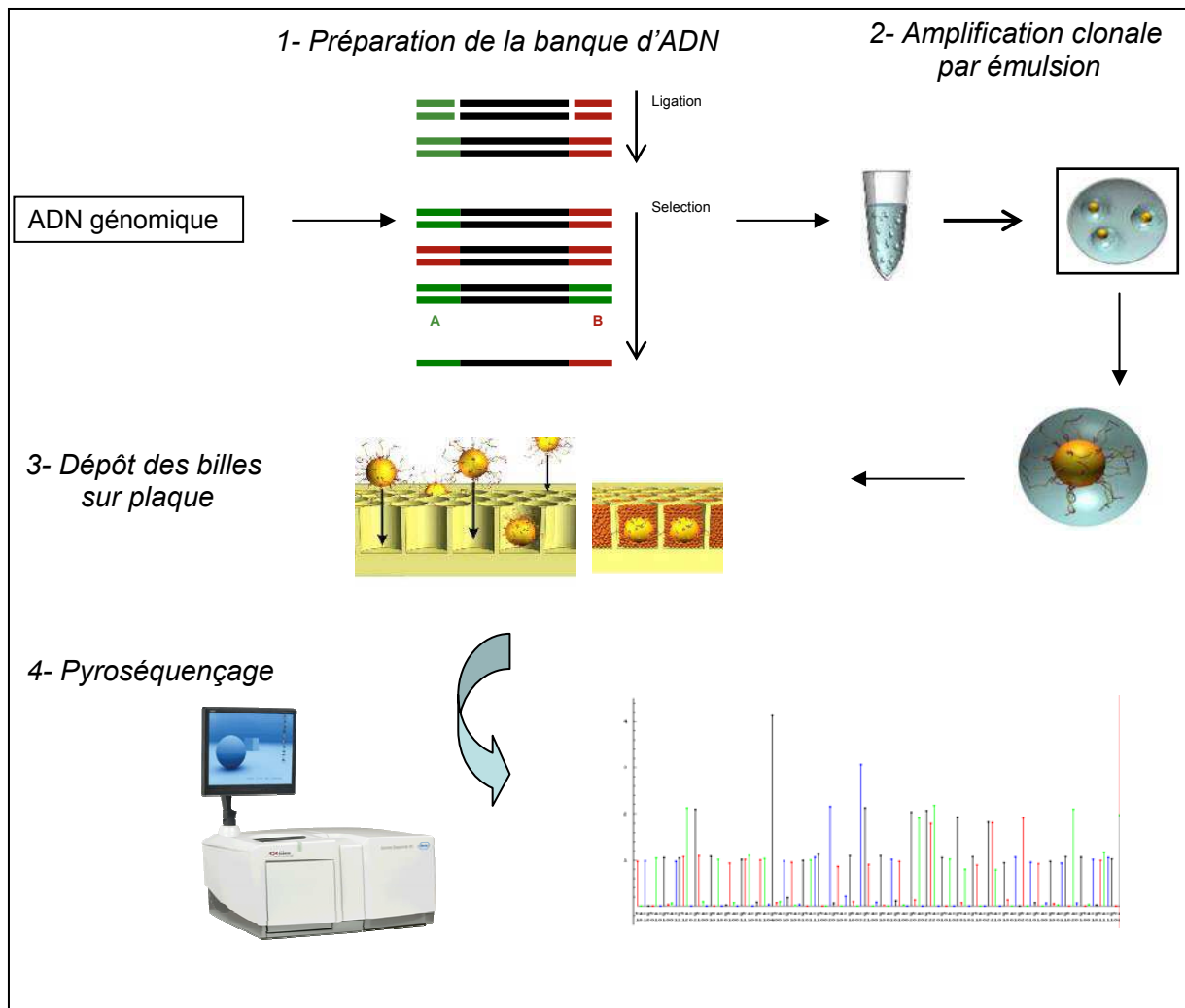


Figure 24: Schéma des étapes de pyroséquençage, modifié d'après ROCHE.

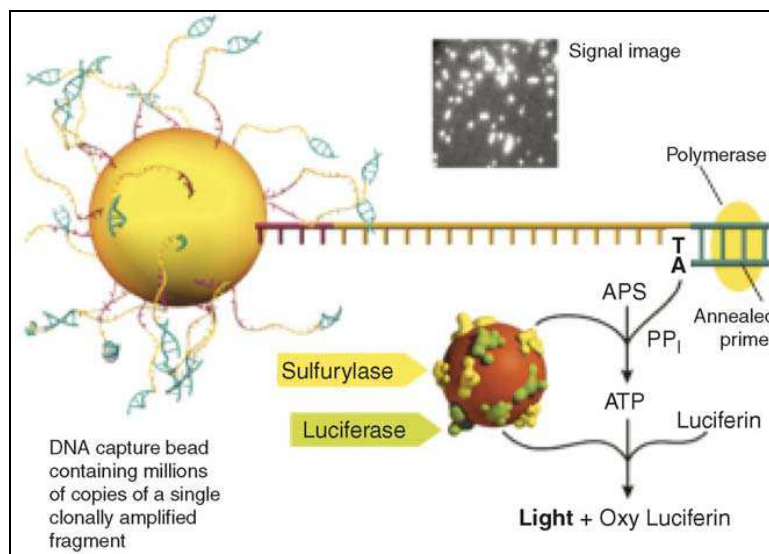


Figure 25: Schéma de la réaction de pyroséquençage lors de l'incorporation d'un nucléotide par la polymérase. D'après Mardis, 2008. *Trends in Genetics*. 24: 133-141.

## 5-2 Analyses bioinformatiques

Après l'étape de pyroséquençage, un assemblage en contigs des séquences générées, a été entrepris à l'aide du logiciel Genome Sequencer *De Novo Assembler Software* (Roche Applied Science, Mannheim, Germany) : 168.909 contigs de 352 pb de moyenne ont ainsi pu être obtenus (couverture: 2.88X) à partir des 1.441.839 séquences générées par pyroséquençage.

Nous avons choisi dans un premier temps d'analyser les contigs assemblés ; cette analyse a été effectuée au NCBI en collaboration avec Eugene V. Koonin et Yuri Wolf, pendant un séjour de 2 mois effectué là bas. Dans un second temps, nous avons choisi d'analyser toutes les séquences brutes générées par pyroséquençage (lectures), afin de maximiser les sources d'information. Toutefois, le nombre de séquences générées étant très important (1.441.839 séquences), nous présenterons dans cette thèse uniquement les résultats obtenus sur une fraction non négligeable de ce jeu de données (427.616 séquences). Les analyses sur l'ensemble du jeu de données sont actuellement en phase finale d'analyse. Le schéma général de la stratégie que nous avons utilisé pour traiter les données est présenté dans la figure 26.

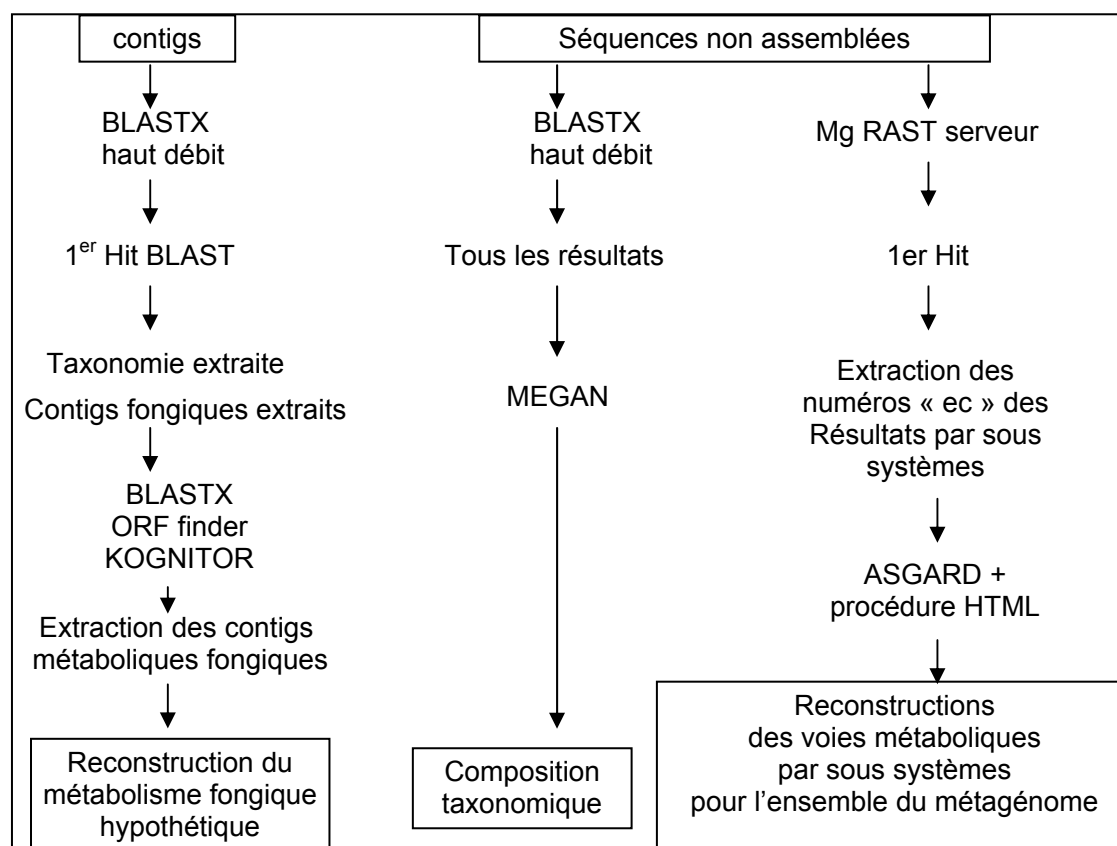


Figure 26: Schéma global des stratégies utilisées pour analyser les données issues du pyroséquençage (reads = séquences non assemblées).



## 5-2-1 Analyse des contigs assemblés

### 5-2-1-1 Analyse *BLAST X*

Une analyse BLASTX a été effectuée sur l'ensemble des contigs générés. L'analyse BLAST (Basic Local Alignment Search Tool) a pour but de comparer des séquences requêtes (de types nucléiques ou protéiques) à des séquences présentes dans les bases de données (RefSeq, nr, EMBL, DDBJ...), afin de trouver des similarités entre ces séquences. Les différents types d'analyse BLAST sont présentés dans le tableau 4.

L'un des principaux objectifs des analyses de séquences (ADN et protéiques) est d'identifier des séquences homologues et d'utiliser les conservations de structures et de séquences pour prédire des activités biochimiques, pour déduire les fonctions écologiques (Koonin et Galperin, 2002).

L'analyse BLASTX compare un jeu de données nucléotidiques, à une base de données protéiques ; pour cela, la séquence requête est traduite dans les 6 cadres de lecture, chaque cadre de lecture traduit est ensuite comparé à la base de données protéiques. A l'issue de la recherche un score (bitscore) est donné par le programme, il correspond à la similitude estimée entre la séquence requête et la séquence présente dans la base de données. Un second score, la e-value est également donné : il correspond à la probabilité que la séquence trouvée dans la base de données soit liée uniquement au hasard. Plus la e-value est faible, plus cette probabilité est faible. Il est ainsi considéré que des séquences ayant un bitscore élevé, associé à une e-value faible, correspondent à des séquences homologues.

Dans notre étude, un seuil d'e-value à  $10^{-3}$  a été choisi et seul le premier résultat du BLAST (1<sup>er</sup> hit) a été retenu ; les banques « RefSeq » (Reference Sequence database ; Pruitt *et al.*, 2007 ) et « nr » (non-redundant database) de Genbank ont été testées.

La base de données RefSeq est une collection non redondante de séquences (ADN, ARN ou protéines) annotées de façon riche et précise : elles constituent donc des séquences de référence ; une entrée RefSeq correspond à une seule molécule naturelle, appartenant à un seul organisme.

La base de données nr quant à elle est beaucoup plus volumineuse et regroupe toutes les entrées de nucléotides de Genbank, ainsi que les entrées de la base RefSeq ; elle regroupe en plus toutes les séquences déposées dans EMBL (European Molecular Biology Laboratory),

DDBJ (DNA Databank of Japan) et PDB (Protein Databank). Un rappel des différentes utilisations des recherches BLAST est fourni dans le tableau 4.

Ces analyses BLASTX, coûteuses en temps de calcul ont été réalisées sur des clusters informatiques (Ferme de calcul du NCBI, et Clusters de Ouest Genopole). Un environnement UNIX a été utilisé pour lancer ces requêtes.

<b>Programme</b>	<b>Séquence requête</b>	<b>Type de requête utilisé pour la recherche</b>	<b>Base de données utilisée pour la recherche</b>
BLASTN	ADN	ADN	ADN
BLASTP	Protéine	Protéine	Protéine
BLASTX	ADN	ADN traduit	Protéine
TBLASTN	Protéine	Protéine	ADN traduit
TBLASTX	ADN	ADN traduit	ADN traduit

Tableau 4: Récapitulatif sur l'utilisation des programmes de recherche BLAST.

Comme nous l'avons indiqué, dans un premier temps tous les contigs ont été analysés par BLASTX contre la base de données RefSeq ; les contigs possédant de cette façon une annotation ont été isolés, les redondances ont été éliminées, puis leur taxonomie a été extraite selon la taxonomie du NCBI. Les contigs de nature fongique ont ainsi été extraits de cette façon.

L'exécution d'une commande BLAST sur une banque de données installée en local, nécessite dans un premier temps le formatage du jeu de données en format FASTA, qui se fait grâce à la fonction « format db », selon :

```
Formatdb -i Requete -p F -o T -n sequencefile
```

L'option -pF indique que le jeu de données sera de type nucléotidique. Les options complètes de la fonction formatdb sont accessibles via le site web :

<http://rcc.uga.edu/applications/bioinformatics/ncbiblast2210/formatdb.html>.

L'option `-i` introduit le fichier requête

L'option `-o T` se réfère au format FASTA de la requête.

L'option `-n` est utilisée pour le nom de sortie du fichier.

La recherche BLAST est ensuite lancée grâce à la ligne :

```
blastall -p blastx -d refseq_protein -e 0.001 -i QUERY -o out.QUERY
```

L'option `-p` est utilisée pour le choix du programme BLAST (BLASTn, BLASTX,...)

L'option `-d` est reliée au choix de la base de données.

L'option `-i` est reliée à la requête

L'option `-o` renvoie au fichier de sortie.

L'option `-e` renvoie au seuil d'e-value choisi.

La liste complète des options de la commande blastall est disponible sur le site web

[http://blast.genome.jp/sit-bin/show\\_man?blast2](http://blast.genome.jp/sit-bin/show_man?blast2).

Les scripts exécutés au NCBI, avec l'aide de Yuri Wolf et d'Eugene Koonin sont présentés ci-après. Ils ne sont applicables que localement sur les fermes de calcul dédiées du NCBI.

### Scripts utilisés au NCBI pour l'analyse BLASTX des contigs générés par pyroséquencage

**dos2unix 454AllContigs.fna**

*#commande qui précise l'environnement UNIX#*

**fake\_gi 454AllContigs.fna -b=1000001 > All.fna**

*#le fichier est renommé en All.fna ; la première séquence aura pour numéro « 1000001 »#*

**formatdb -i All.fna -p F -o T -n All**

*#commande qui formate le fichier du jeu de données pour une analyse BLAST ultérieure#*

**nemp -r="^\>gi\|(d+)" All.fna | random\_subset > AllRand.gi**

*#commande qui permet de réarranger le jeu de données aléatoirement#*

```
head -500 ../data/AllRand.gi > test.gi
```

```
head -5 ../data/AllRand.gi > tmp.gi
```

```
fastacmd -d ../data/All -i tmp.gi | blastall -p blastx -d refseq_protein -e 10 -m 0 -I T -z  
20000000 -b 10 -v 10 -F F -C F | less
```

```
fastacmd -d ../data/All -i tmp.gi | blastall -p blastx -d refseq_protein -e 10 -m 9 -I T -z  
20000000 -b 10 -v 10 -F F -C F -o tmp.tab
```

*#ces 4 lignes de commande sont utilisées pour tester la procédure sur un sous échantillon du jeu de données (respectivement 500 séquences, puis 5 séquences). #*

```
batch_farmblast ../data/AllRand.gi -prog=blastall -p=blastx -dq=  
/net/frosty/vol/export1/lecalvet/meta/data/All -ds=refseq_protein -od=blast -e=1 -b=10 -  
v=10 -C=F -t=@ -clean -log=test.log -Fq=unified50limit -hp=  
/net/frosty/vol/export1/lecalvet/meta/protein -bt=100 -mail=lecalvet@ncbi.nlm.nih.gov -  
work
```

*#cette commande exécute la procédure BLASTX vs RefSeq sur les fermes de calcul.#*

```
btabs2tas blast/AllRand.*.tab -ecut=0.001 | nemp -r="^\d+\t\d+" | sort -n > AllTopHit.lst
```

*#cette commande permet d'appliquer un filtrage des hits de BLASTX selon une e-value de 10<sup>-3</sup>, puis d'extraire les annotations des premiers hits de BLAST.#*

```
nemp -r="^\t(\d+)" AllTopHit.lst | sort -n | uniq > AllTopHit.gi
```

*#les redondances sont éliminées par cette ligne de commande#*

```
cat AllTopHit.gi | gi2taxdat > AllTopHit.g2t
```

```
nemp -r="^\t(*)" AllTopHit.g2t | sort | uniq > tmp.txt
```

```
cat AllTopHit.lst | subs_keys AllTopHit.g2t > AllTopHit.tax
```

```
fastacmd -i AllTopHit.gi -d refseq_protein -t T | grep ">" | nemp -  
r="^\>gi\|(\d+)\S*\s+(.*)" > AllTopHit.g2d
```

*#Ces 4 lignes de commande permettent d'extraire la taxonomie (AllTopHit.tax) et les annotations (AllTopHit.g2d) de tous les 1ers hits de Blast#*

### ***5-2-1-2 Effort de séquençage***

Afin d'estimer l'effort de séquençage, nous avons dans un premier temps réalisé une analyse BLASTX (comme décrite précédemment) sur la base des résultats de 2 runs de pyroséquençage (soit environ 900 000 séquences). La taxonomie des contigs a été extraite toujours sur la base du premier résultat de BLAST.

Les contigs ont ensuite été classifiés, puis séparés selon leur taxonomie (extraction par phylum et sous phylum). A l'issue de cette classification, nous avons pu déterminer les organismes bactériens les plus représentés dans notre jeu de données. Nous avons alors choisi les organismes bactériens dans les phyla dominants (voir le tableau 5), et pour lesquels un génome complet était disponible dans RefSeq. De plus, le choix de ces génomes de référence a été orienté par le nombre de protéines qu'il regroupait. Toutes les protéines de ces organismes ont alors été extraites à partir de Genbank (sous la forme de liste Gi : Gene identification).

Le principe du test est le suivant : pour chaque génome complet dont nous disposons, les contigs provenant d'un même phylum, préalablement extraits, ont été analysés par BLASTX contre les protéines du génome complet (utilisées ici comme base de données). La procédure BLASTX décrite précédemment a été utilisée, en respectant les mêmes options de filtrage. A l'issue de cette analyse on peut donc connaître le taux de recouvrement du nombre de protéines du génome de référence par les contigs (c'est à dire le nombre de protéines du génome complet recouvert par au moins un contig de notre jeu de données) ; nous avons établi qu'un taux de couverture de 70% serait idéal pour poursuivre notre analyse de métagénome. Les résultats de ce test sont présentés dans la partie V de cette thèse. A l'issue de ce test, nous avons choisi d'effectuer un run supplémentaire de pyroséquençage afin d'augmenter le taux de couverture. Nous avons choisi de comparer les résultats de couverture de génome obtenus à l'issue de 2 runs de séquençage avec ceux obtenus après 3 runs. Si le gain de couverture s'avère minime, nous considérerons que l'effort de séquençage a été suffisant.

<b>Organismes testés</b>	<b>Rang taxonomique</b>	<b>Nombre de protéines présentes dans la base de données (Gi List)</b>
<i>Gramella forsetii</i> KT0803	Bacteroidetes	3585
<i>Cytophaga hutchinsonii</i> ATCC33406	Bacteroidetes	3785
<i>Parvibaculum lavamentivorans</i>	$\alpha$ -protéobactérie	3636
<i>Silicibacter sp.</i> TM1040	$\alpha$ -protéobactérie	4254
<i>Colwellia psychrerytraea</i> 34H	$\gamma$ -protéobactérie	4912
<i>Methylococcus capsulatus str. Bath.</i>	$\gamma$ -protéobactérie	2956
<i>Nitrosococcus oceani</i> ATCC 19707	$\gamma$ -protéobactérie	3023

Tableau 5: Organismes bactériens de référence utilisés pour le test d'effort de séquençage.

### ***5-2-1-3 Analyse des contigs fongiques***

Les annotations fonctionnelles des contigs ont été déduites par BLASTX, selon la procédure décrites précédemment. De plus, la taxonomie de chaque premier hit de BLAST a été extraite sur la base de ces résultats, selon la taxonomie du NCBI. Ainsi, les contigs annotés comme « fongiques » ont pu être extraits puis analysés plus en détail.

Dans le but de mieux caractériser les annotations assignées aux contigs fongiques, une seconde analyse BLASTX a été effectuée. De plus, une analyse ORF finder (NCBI ; <http://www.ncbi.nlm.nih.gov/gorf/gorf.html>) a été réalisée. L'outil ORF finder (Open Reading Frame Finder) permet de trouver tous les cadres ouverts de lecture d'une séquence requête.

Ainsi, par ces procédures, des annotations fonctionnelles claires ont pu être établies (y compris pour certains contigs classifiés comme « hypothétiques » par BLASTX). Afin d'affiner la classification des contigs, les séquences fongiques ont ensuite été analysées par KOGNITOR (NCBI ; <http://www.ncbi.nlm.nih.gov/COG/>) qui permet d'effectuer une comparaison contre une banque d'orthologues eucaryotes. De cette manière, on obtient une classification de tous les contigs fongiques, selon la nomenclature KOG (figure 27).

Les contigs liés au métabolisme (nomenclature : Q, C, G, E, F, H, I, P) ont ensuite été extraits, dans le but de reconstruire les voies métaboliques portées par les séquences

fongiques. Chaque contig annoté a dans ce but été comparé aux cartes métaboliques KEGG (Kyoto Encyclopedia of Genes and Genomes ; <http://www.genome.jp/kegg/>), qui permet de situer les enzymes fongiques prédites sur des cartes de voies métaboliques.

#### **INFORMATION STORAGE AND PROCESSING**

[J] Translation, ribosomal structure and biogenesis

[A] RNA processing and modification

[K] Transcription

[L] Replication, recombination and repair

[B] Chromatin structure and dynamics

#### **CELLULAR PROCESSES AND SIGNALING**

[D] Cell cycle control, cell division, chromosome partitioning

[Y] Nuclear structure

[V] Defense mechanisms

[T] Signal transduction mechanisms

[M] Cell wall/membrane/envelope biogenesis

[N] Cell motility

[Z] Cytoskeleton

[W] Extracellular structures

[U] Intracellular trafficking, secretion, and vesicular transport

[O] Posttranslational modification, protein turnover, chaperones

#### **METABOLISM**

[C] Energy production and conversion

[G] Carbohydrate transport and metabolism

[E] Amino acid transport and metabolism

[F] Nucleotide transport and metabolism

[H] Coenzyme transport and metabolism

[I] Lipid transport and metabolism

[P] Inorganic ion transport and metabolism

[Q] Secondary metabolites biosynthesis, transport and catabolism

#### **POORLY CHARACTERIZED**

[R] General function prediction only

[S] Function unknown

**Figure 27:** Symboles utilisés pour la classification des KOGs.

## 5-2-2 Analyse des séquences générées, non assemblées (reads)

### 5-2-2-1 Filtrage des données

Comme nous l'avons expliqué précédemment, nous avons choisi d'analyser l'ensemble des séquences non assemblées issues du pyroséquençage, en complément de l'analyse des contigs assemblés. Cette analyse complémentaire a pour but d'extraire le maximum d'informations du jeu de données initial, en évitant les biais éventuels potentiellement apportés par l'assemblage automatique.

Dans un premier temps les séquences de faible qualité ont été filtrées, selon les recommandations de Huse *et al.* (2007). Les séquences présentant un ou plusieurs nucléotides indéterminés (N) ou celles présentant une faible complexité (biais dans la composition nucléotidiques) ont été retirées du jeu de données initial, à l'aide du script PERL suivant :

*Tous les scripts PERL présentés ci-après ont été écrits en collaboration avec Alexis Dufresne.*

#### **Script PERL pour supprimer les séquences ayant une composition biaisée et celles contenant des nucléotides indéterminés.**

```
#!/usr/bin/perl
#Ce script identifie les séquences qui n'ont pas de composition biaisée
#(qui contiennent les 4 bases) et qui ne contiennent pas de N

use strict;
use warnings;
use Bio::Seq;
use Bio::SeqIO;

open (OUTFILE1, ">high_quality_sequences");
open (OUTFILE2, ">sequences_with_N");
open (OUTFILE3, ">sequence_low_complexity");

my $seqin = Bio::SeqIO->new( -'format' => 'fasta' , -file => $ARGV[0]);

while((my $seqobj = $seqin->next_seq())) {

my $id = $seqobj->display_id;

#Recupère l'ID de la seq

my $str = $seqobj->seq();

#Recupère la séquence (sans l'ID)
```



```

if ($str =~ /A/ and $str =~ /C/ and $str =~ /G/ and $str =~ /T/) {

    if ($str =~ /^[^ACGT]/) {
#pour identifier séquences contenant des N

        print OUTFILE2 ">$id\n";
        print OUTFILE2 "$str\n";
    }

    else {

#séquences sans N et contenant les quatre nucléotides

        print OUTFILE1 ">$id\n";
        print OUTFILE1 "$str\n";
    }
}

else {

print OUTFILE3 ">$id\n", "$str\n";

}
}

```

### 5-2-2-2 Analyse BLASTX

Une analyse BLASTX (contre la base de données nr) a été effectuée sur 427.616 séquences générées par pyroséquençage (après filtrage du jeu de données par le script PERL retirant les séquences de composition nucléotidique biaisée). Un cluster informatique de Ouest Genopole (Genocluster, Rennes, France) a été utilisé à cet effet. Les scripts permettant la soumission au cluster sont présentés ci-après. Afin de ne pas saturer les nœuds du cluster, le fichier de 427.616 séquences a dans un premier temps été fractionné en plusieurs fichiers de 100 séquences ; le script programmé en PERL, pour le fractionnement du fichier est également présenté.

#### **Script PERL utilisé pour fragmenter le fichier contenant toutes les séquences issues du pyroséquençage, en plusieurs fichiers de 100 séquences**

```

#! /usr/bin/perl -w

#On parcourt le fichier principal contenant toutes les séquences
#On sépare les séquences en groupes de 100 séquences au maximum en les
#comptant et en ouvrant un nouveau fichier toutes les 100 séquences

use strict;
my $compteur_fichier = 1;
my $compteur_seq = 0;
my $ligne = "";

```

```

my @all_seq = ();
my $nom_fichier = "";

open (FILE, $ARGV[0]);

@all_seq = <FILE>;

close FILE;

open (TEMP, '>454_fichier_1.fna');

foreach $ligne (@all_seq) {

    if ($ligne =~ /^>EX/) {

#Si la ligne correspond au début d'une séquence fasta (>identifiant)

        $compteur_seq ++;

        if ($compteur_seq <= 100) {

#si le compteur de séquence est inférieur ou égal à 100, on écrit la
#ligne dans le fichier 454_fichier_XXX.fna

            print TEMP $ligne;

        }

        else {

#Si le compteur de séquence est supérieur à 100, on ferme le fichier
#454_fichier_XXX.fna, on remet le compteur de séquence à 1, on
#incrémente le compteur de fichier

#on ouvre un nouveau fichier (454_fichier_YYY.fna où YYY = XXX+1) et on
écrit la ligne dedans

            $compteur_seq = 1;
            $compteur_fichier ++;
            $nom_fichier = "454_fichier_".$compteur_fichier.".fna";
            close TEMP;
            open (TEMP, ">$nom_fichier");
            print TEMP $ligne;

        }

    }

    else {

        print TEMP $ligne;

#il s'agit d'une ligne de la séquence au format fasta. On l'écrit dans
# le fichier ouvert à ce moment là

    }

}

```

## **Script PERL de soumission au cluster informatique de la plateforme Ouest Genopole (GenoCluster)**

```
#!/bin/bash

#Soumission des 4277 fichiers de 100 séquences nucléotidiques (métagénome
#sources hydrothermales) pour une analyse BLASTX contre NCBI_nr dans
#le cadre d'une analyse de la composition phylogénétique avec Megan

#compteur de job total

i=0

#compteur de job lancé. On attend 30s tous les deux jobs lancés
#pour permettre aux autres utilisateurs de lancer leurs travaux

j=0

#Tant que le nombre total de jobs est inférieur au nombre de fichiers à
analyser

while test $i != 4277
do

#On incrémente le nombre total de job

i=$((i+1))

#On incrémente le compteur de jobs lancés et on soumet le job en cours

j=$((j+1))s

nom_fichier_requete="454_fichier_${i}.fna"

nom_fichier_resultats="454_fichier_${i}.blastx"

nom_job="Reads_454_metagenome_${i}"

#les calculs sur genocluster sont envoyés dans une file d'attente
# (Sun Grid Engine, commande qsub)
#les informations pour lancer le calcul avec qsub doivent être lues
#dans un script shell (fichier temp.sh) généré comme suit :

echo "##! /bin/bash" > temp.sh

#spécification du shell d'execution

echo "##$ -S /bin/bash" >> temp.sh

#adresse mail de l'utilisateur lançant les travaux; -m : envoie d'un mail
au début, à la fin et si job s'arrête

echo "##$ -M adresse@mail" >> temp.sh
echo "##$ -m n" >> temp.sh

#initialisation de l'environnement pour BLAST sur genocluster

echo ". /local/env/envncbi.sh" >> temp.sh
```

```
#commande BLAST

echo "blastall -p blastx -d /db/nr/current/blast/All/nr -i
./fichier_fasta/$nom_fichier_requete -e 0.001 -o $nom_fichier_resultats" >>
temp.sh

#Tant que le nombre de job lancés est inférieur ou égale à 2, on soumet
#le job en cours

if [$j -le 2]

then

#commande pour lancer le job sur genocluster avec le script généré
/usr/local/sge/bin/lx24-amd64/qsub -cwd -N $nom_job temp.sh

rm temp.sh

else

#Sinon, on attend 30 secondes, on remet le compteur de jobs lancés à
#zéro et on lance la job en cours

sleep 30s

j=0

/usr/local/sge/bin/lx24-amd64/qsub -cwd -N $nom_job temp.sh

rm temp.sh

fin

done
```

### 5-2-2-3 Analyse MEGAN

Afin d'obtenir une vision de la composition taxonomique de notre échantillon plus affinée que celle fournie par les résultats du « Premier hit » de BLAST, le logiciel MEGAN (Metagenome Analyzer ; Hudson *et al.* 2007) a été utilisé. Ce logiciel repose sur l'algorithme du Dernier Ancêtre Commun (Last Common Ancestor, LCA). Chaque fichier de résultat généré par BLASTX (contre la base de données nr) sur les séquences filtrées a été intégré dans MEGAN, pour cette reconstruction taxonomique. Dans un premier temps, les annotations générées par BLASTX, ont été extraites. Tous les résultats de BLASTX, sont ainsi récupérés, après filtrage par e-value de  $10^{-3}$ . Le script utilisé pour extraire cette information est présenté ci-après. Dans un second temps chaque fichier généré est directement intégré dans MEGAN.

#### **Script PERL pour extraire les annotations des séquences obtenues après BLASTX.**

```
#!/usr/bin/perl

use strict;
use warnings;

my @big_genbank;
my @small_genbank;
my @annotation;
my $ligne_big_genbank;
my $ligne;
my $locus;
my $size;
my @alias_cds;
my @product;
my $i;
my @coord_cds;
my @strand_cds;
my @translation;
my @seq_nucleotides_genbank;
my $sequence_read;
my $seq;
my $CDS = 0;

open (GBK_FILE, $ARGV[0]);

@big_genbank = <GBK_FILE>;
close GBK_FILE;

foreach $ligne_big_genbank (@big_genbank) {

    push (@small_genbank, $ligne_big_genbank);

    if ($ligne_big_genbank =~ /\s{6,8}\d{1,4}\s[acgt]{10}.*/) {
#lignes contenant les séquences nucléotidiques des reads dans le
#fichier Genbank

        push (@seq_nucleotides_genbank, $ligne_big_genbank);
    }
}
```



```

        else {
            print $locus, "\t", $size, "\n";
        }

        open (OUTFILE, ">>liste_reads_MG_RAST.fas");
        print OUTFILE ">", $locus, "\n", $sequence_read,
"\n";

        close OUTFILE;
    }

    }

    $locus = "";
    $size = "";
    @alias_cds = ();
    @product = ();
    @coord_cds = ();
    @strand_cds = ();
    @translation = ();
    $seq = "";
    $CDS = 0;
}
}

#####
#Subroutines
#####

#subroutine pour que la valeur de chaque feature
#soit sur une seule ligne

sub nice_genbank_file {
my (@genbank) = @_ ;
my $line;
my @annotation;
my $feature;

foreach $line (@genbank) {

if ($line =~ /\s{21}\/\./.*="/) {
push (@annotation, $line);

#on place la ligne n dans @annotation
}
elsif ($line =~ /\s{21}(\w*.*\n)/) {
$feature = pop(@annotation);

#on extrait la ligne n-1 et on la place dans la variable $feature

chomp ($feature);
$feature = $feature." ".$1;

#on concatène ligne n-1 avec ligne n (sans partie avec espace)

push (@annotation, $feature);

#on place dans le tableau @annotation la nouvelle ligne formée a partir de
n et n-1

```

```

}
else {
push (@annotation, $line);
}
}

return @annotation;

#on renvoie @annotation vers le tableau @annotation de la partie
#principale du script
}

#####
#####

sub nice_sequence_file {

#subroutine pour que la séquence nucléotidique des reads soit sur une
#seule ligne

my (@seq_nucleotides_genbank) = @_;
my $line;
my @annotation;
my $feature;

$line = join('',@seq_nucleotides_genbank);

chomp ($line);
$line =~ s/\d//g;
$line =~ s/\s//g;

return $line;
}

```

#### ***5-2-2-4 Reconstruction des voies métaboliques***

L'information des séquences filtrées a été utilisée pour reconstruire l'ensemble des voies métaboliques présentes dans notre échantillon. Ainsi, après filtrage des séquences (script PERL supprimant les séquences ayant une composition biaisée), 412.343 séquences (de longueur moyenne 238 pb) ont directement été soumises au serveur mgRAST (Metagenomic Genome Rapid Annotation Using Subsystem Technology; <http://metagenomics.nmpdr.org/>; Meyer, *et al.*, 2008). Ce serveur permet une annotation des fragments d'ADN, propose leur classification phylogénétique, ainsi qu'une reconstruction des voies métaboliques présentes dans l'échantillon. Les étapes de reconstructions utilisées par le serveur mgRAST sont présentées dans la figure 28.



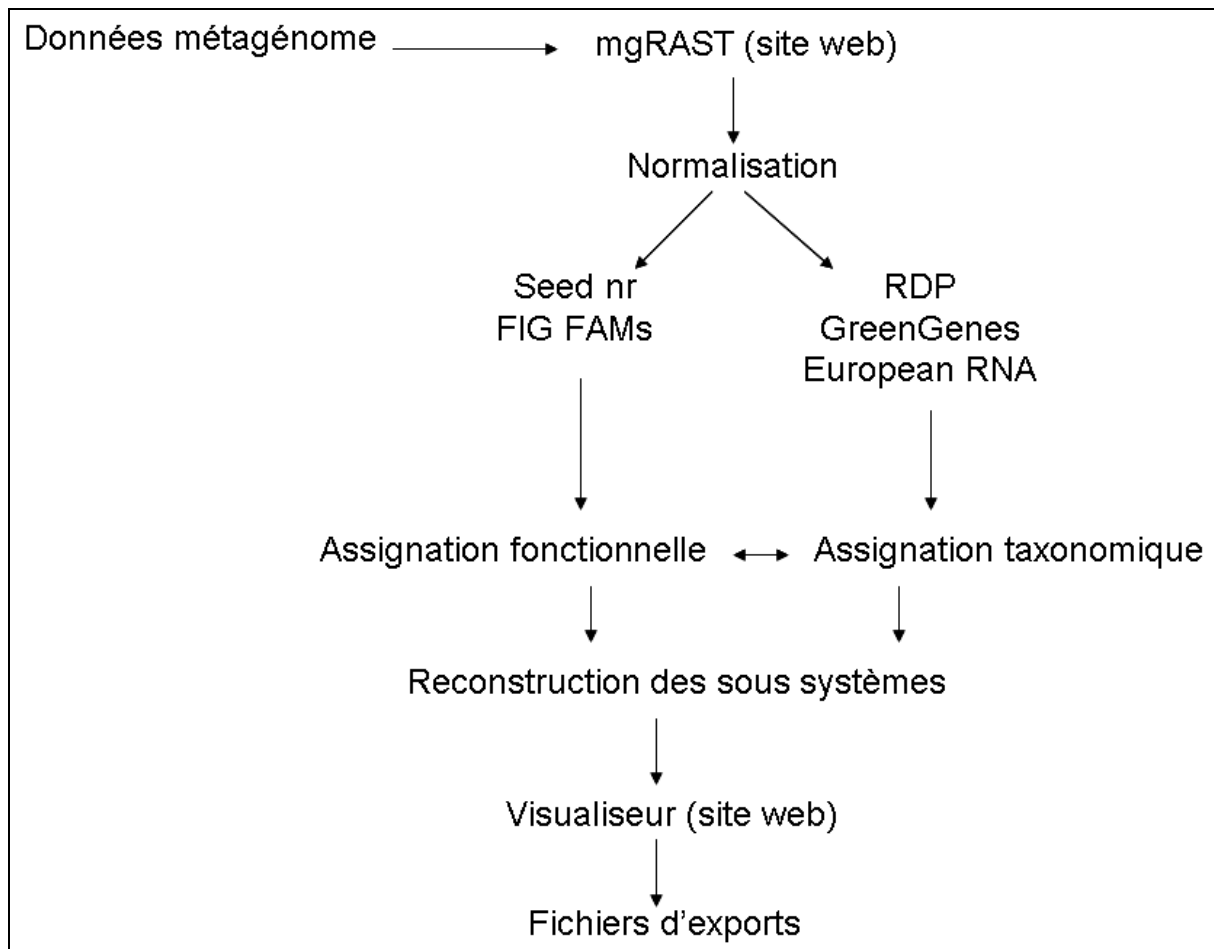


Figure 28: Schéma général du fonctionnement du serveur mgRAST ; d'après Meyer *et al.*, 2008 ; *BMC Bioinformatics*. 9: 386.

Bien que cette reconstruction métabolique soit directement possible via le serveur, nous avons choisi une solution alternative afin de mieux contrôler les résultats générés. Les fichiers d'annotation générés par le serveur mgRAST ont ainsi été ré-analysés par nos soins ; toutes les annotations contenant des numéros de classification d'enzymes (ec number : enzymes classifications ; convention KEGG de nomenclature des enzymes) ont été extraites. Ces numéros « ec » ont ensuite été intégrés dans le logiciel ASGARD-1.0.1 ; ce logiciel s'appuie sur la présence ou l'absence de ces numéros « ec » pour reconstruire les voies métaboliques, selon les cartes KEGG. Une procédure supplémentaire de modification de pages HTML a été effectuée afin de relier la présence des « ec » à leur nombre dans notre jeu de données : on peut ainsi reconstruire les voies métaboliques KEGG présentes dans notre

échantillon. Il est à noter que pour cette étape de reconstruction métabolique, aucune assignation taxonomique n'a été effectuée. On aura donc une reconstruction globale prédictive des fonctions métaboliques du métagénome, qui ne seront donc pas organismes spécifiques.

### Script PERL utilisé pour produire le fichier nécessaire au fonctionnement du logiciel ASGARD-1.0.1 (fonction color\_map)

```
#!/usr/bin/perl -w
use strict;

my @Kegg_pathways;

#Fichier contenant la liste de toutes les voies métaboliques et de
#toutes les enzymes KEGG. Nom : Kegg_pathways

my @annotation; #Fichier d'annotation du métagénome (4441177_3.tab =
fichier d'annotation MG_RAST reformaté)
my $ligne;
my $pathway;
my $i = 0;
my $EC_number;
my $motif;

open (EC_FILE, $ARGV[0]);
open (TAB, $ARGV[1]);

@Kegg_pathways = <EC_FILE>;
@annotation = <TAB>;

close EC_FILE;
close TAB;

foreach $ligne (@Kegg_pathways) {

    if ($ligne =~ />.*/) {

        $pathway = $ligne;
        $i = 0;

    }

    elsif ($ligne =~ /(\d*\.\d*\.\d*|\-)\.(\d*|\-)\t.*/) {
#Si l'enzyme est présente dans le fichier d'annotation, alors on écrit
#son numéro et son nom dans le fichier de sortie ainsi que le nom de
#la voie métabolique correspondante si ce n'est pas déjà fait

        $EC_number = $1;

        if (grep(/\(EC\s*$EC_number\)/, @annotation)) {

#On recherche un motif exact
#Le motif est recherché est sous cette forme : (EC 1.1.1.11)
#On ne fait pas la recherche uniquement sur le numero EC
```

#Ex: on si on cherche 1.1.1.11 grep va trouver aussi 1.1.1.1 même si ce  
#dernier n'est pas dans le fichier d'annotations

```
        if ($i == 0) {  
            $i = 1;  
#permet de savoir si le nom de la voie métabolique a déjà été écrit  
#dans le fichier de sortie. $i= 0 => nom pas encore écrit; $i = 1 nom déjà  
#écrit  
                print $pathway;  
                print $ligne;  
            }  
            else {  
                print $ligne;  
            }  
        }  
    }  
}
```

## **Partie III : Mise en place d'une base de données experte de marqueurs moléculaires fongiques : PHYMYCO-DB**

---

## Résumé

Le premier objectif de cette thèse était d'étudier la diversité fongique isolée d'échantillons hydrothermaux. Pour cela nous avons choisi d'analyser le gène codant l'ARNr18S des champignons présents dans l'environnement. Cette étude repose sur des analyses phylogénétiques effectuées sur la base des séquences obtenues. Pour mener à bien ces analyses, nous avons choisi de collaborer avec l'IRISA (Institut de recherche en informatique et systèmes aléatoires ; Rennes, France) pour mettre en place une base de données experte regroupant des marqueurs génétiques fongiques.

Le premier marqueur, le gène codant pour la petite sous unité de l'ARNr, est présent en multiples copies chez tous les organismes vivants. Du fait de son implication centrale dans la traduction protéique, sa séquence est très conservée et peu sujette aux phénomènes de transferts de gènes horizontaux. Ces caractéristiques en font un marqueur modèle en phylogénie et il a été très largement utilisé à cet effet. Cette utilisation massive a résulté en sa très large représentation dans les bases de données de séquences mondiales. Cependant, cette représentation massive possède des inconvénients, dont le principal est lié à la présence de séquences de mauvaise qualité (Nilsson *et al.*, 2006). L'utilisation de ces séquences « erronées » va alors conduire à des résultats biaisés de phylogénies.

Le second marqueur fongique que nous avons choisi d'extraire est le gène codant le facteur d'élongation EF1-  $\alpha$ . Ce gène, présent en simple copie chez la plupart des organismes est impliqué dans la synthèse des protéines. Ce marqueur a notamment été utilisé pour prouver que champignons et animaux étaient des groupes frères (Baldauf, 1999 ; Baldauf *et al.*, 2000).

Les ressources disponibles en ligne pour analyser les séquences fongiques (Ribosomal Database Project, SILVA...) ne nous semblaient pas appropriées pour obtenir les séquences tout au long de l'Arbre de la Vie fongique et nous avons donc entrepris de construire une base de données pour récupérer ces marqueurs phylogénétiques et les analyser en toute confiance. Nous avons choisi d'extraire toutes les séquences de ces 2 marqueurs moléculaires à partir de Genbank selon des critères sélectifs expliqués dans la partie Matériel et méthodes de cette thèse (partie II-4). La mise au point de cette base de données a donc permis de récupérer les séquences fongiques en suivant l'arborescence taxonomique de Genbank, et d'effectuer des alignements de haute qualité à partir des jeux de données extraits. Cette base de données a dans un second temps été utilisée pour résoudre les futures analyses phylogénétiques réalisées lors de cette thèse.

La base de données PHYMYCO-DB est accessible à l'adresse suivante :

<http://phymycodb.genouest.org/>. Au total, la base de données regroupe 6978 séquences d'ARNr 18S et 431 séquences d'EF1-  $\alpha$  fongiques, ainsi que 19 fichiers d'alignements précalculés.

# PHYMYCO-DB: a curated database for analyses of fungal diversity and evolution

Le Calvez Thomas<sup>1,3</sup>, Guillot Laetitia<sup>2\*</sup>, Alexis Dufresne<sup>1,3</sup> and Vandenkoornhuyse Philippe<sup>1,3</sup>

<sup>1</sup> UMR 6553 CNRS ECOBIO, FR/90 IFR 2116 CAREN, Université de Rennes 1, Campus de Beaulieu, 35042 Rennes cedex, France

<sup>2</sup> Plateforme OUEST-Genopole® Equipe Symbiose, IRISA-INRIA, Campus de Beaulieu, 35042 Rennes cedex, France

<sup>3</sup> Université Européenne de Bretagne, UEB (European University of Brittany)

First and second authors have contributed equally

---

## Summary

Unculturable microorganisms form the highest pool of diversity on earth and are identified by analysing molecular taxonomic targets and applying the phylogenetic species concept. The analysis of phylogenetic affinities is based on known sequences. However, the reliability of this information also needs to be analysed to ensure correct phylogenetic interpretation.

We propose a fungal sequence database, PHYMYCO-DB for use in facilitating the analysis of fungal diversity in the environment and addressing evolutionary questions.

We focus on two widely used molecular taxonomic markers: the *SSU rRNA* and *EF1- $\alpha$*  gene sequences. All available fungal sequences from GenBank were filtered to satisfy sequence quality criteria. 6978 and 431 high quality fungal sequences of the *SSU rRNA* gene and *EF1- $\alpha$*  gene respectively, are provided.

PHYMYCO-DB offers different tools and can be used (i) to extract high quality fungal sequences from our browser for each of the 5 fungal phyla and at all taxonomic levels; (ii) to extract specific alignments that we have already performed

Users may also use our website to launch alignments with their personal sequences and stored data. PHYMYCO-DB is accessible through the URL <http://phymycodb.genouest.org/>

## Introduction

In recent decades there has been an exponential increase in the number of gene sequences available in public databases. The use of molecular phylogenies and molecular taxonomic markers has revealed the considerable extent of genetic diversity in living organisms especially those not easily distinguished by morphological characters, such as prokaryotes and unicellular eukaryotes. As the species is the unit of diversity measurement, systematics and evolutionary analyses (Purvis & Hector 2000), an efficient means of identifying boundaries and number of species is required. Molecular methods and implicit adoption of the phylogenetic species concept offer a standardized approach to delimit species (Vandenkoornhuysen *et al.*, 2002a). Thanks to progress in sequencing technologies and bioinformatic methods, gene sequencing and comparison of sequences to detect orthologous counterparts with those in databases, is relatively straightforward. This approach, along with adoption of the phylogenetic species concept (Taylor *et al.*, 2000), can also be successfully applied to non-cultivable organisms which represent up to 99% of the prokaryotic world (Amann *et al.*, 1995). The choice of molecular marker is determinant since the data must have a limited proportion of homoplasy (non-inherited similarities) and enough phylogenetic information (i.e. length and number of terminals) to allow accurate species demarcation.

In this context, the gene coding for the small subunit of the ribosomal RNA (*SSU rRNA*) is often seen as the “ultimate” molecular marker (Woese *et al.*, 1990; Woese, 2000). The *SSU rRNA* gene is present in all living organisms. Its sequence is highly conserved between taxa, reflecting strong functional constraints on the translational machinery. Indeed, most mutations in the *SSU rRNA* gene sequence reduce the stability of the secondary structure of the *SSU rRNA* molecule and thus the efficiency of protein synthesis. Furthermore, this gene, like other informational genes, appears to be less subject to horizontal gene transfers and is believed to provide better inferences of “true” organismal phylogenies (Choi & Kim, 2007).

Another advantage of using the *SSU rRNA* gene sequence is its huge representation in international public databases (Genbank, Benson *et al.*, 2008; EMBL, Kulikova *et al.*, 2004; DDBJ, Sugawara *et al.*, 2008), which allows comparisons between a wide variety of organisms (e.g. for review see Avise, 2004).

However, as many of the sequences deposited and stored in international public databases (GenBank/EMBL/DDBJ) could be (i) of artefactual origin (i.e. chimeric sequences), (ii) of poor quality and even (iii) erroneously annotated (Nilsson *et al.*, 2006), subsequent phylogenetic reconstructions may be highly biased (Bridge *et al.*, 2003) and jeopardize



interpretation. Assessing the reliability of sequences is an important prerequisite to analyses. However, because of its time-consuming nature and due to lack of awareness, this key step is not always taken into account.

Although several tools, such as the Ribosomal Database Project (Cole *et al.*, 2007), and SILVA (Pruesse *et al.*, 2007), exist online for the analysis of *SSU rRNA* gene sequences, nothing to date has allowed researchers to obtain relevant access to the entire range of fungal *SSU rRNA* gene sequences available.

Other genes, such as those coding for the elongation factor *EF1- $\alpha$*  (*tef1*), for  $\beta$ -tubulin (*tub1*, *tub2*), actin (*act1*...), or for RNA polymerase II subunits (*RPB1* and *RPB2*), can be used as alternative markers. The robustness of phylogenies could be improved by combined use of these markers to produce consensus trees or supertrees (Baldauf *et al.*, 2000; Gontcharov *et al.*, 2004; James *et al.*, 2006).

Our aims were therefore (i) to develop an easy-to-use fungal-dedicated database with stored sequences known to be of high quality and for which the selected molecular markers are widely acknowledged, (ii) to produce a tool, based on anchor sequences covering the fungal tree, that can be automatically updated, (iii) to produce high quality multiple alignments for use in testing environmental sequences or evolutionary hypotheses.

## **Material and methods**

We developed a curated database to extract fungal phylogenetic markers, with filters of sequence quality, from all available sequences (Genbank/EMBL/DDBJ), to permit highly reliable phylogenetic analyses. In PHYMYCO-DB version 1 (Figure 29), we provide all the sequence data for nuclear *SSU rRNA* genes and *EF1- $\alpha$*  genes that satisfied our filters. In this way, users can extract fungal datasets, and perform online sequence alignments using ClustalW 2.0 (Larkin *et al.*, 2007) based on extracted or personal data (FASTA format or aln format). This database is clearly extensive and, as the Genbank data increases, other markers will be included, and complete fungal genomes inserted.

### **Database structure**

PHYMYCO-DB is a relational database built on top of a MySQL 4 database using phpMyAdmin (<http://www.phpmyadmin.net/>), a tool written in PHP intended to handle the administration of MySQL over the Web. The PHYMYCO-DB database is managed by the

administrators using the web interfaces developed with PHP version 4.3.9, and the HTML and Ajax programming languages.

### **GenBank data**

The nuclear *SSU rRNA* and *EF1- $\alpha$*  genes were extracted from the Genbank database, using the following requests: “[Organism] and (ssu|SSUrRNA|SSU rRNA|18SrRNA|18S) not (16S|mitoch\*|28S|5.8S|ITS|Internal Transcribed Spacer|internal transcribed spacer)” and “[Organism] and (EF1 alpha|EF-1 alpha|EF1-alpha|EF-1alpha|EF-1-alpha|EF1alpha)”, and different filter parameters: *SSU rRNA* gene sequences shorter than 1000 nucleotides and longer than 2500 nucleotides were rejected, *EF1- $\alpha$*  genes sequences shorter than 700 nucleotides and longer than 2500 nucleotides were rejected, as well as sequences containing more than 10 consecutive undetermined nucleotides.

The datasets were then cross-checked by expert curators. After alignment, sequences with (i) erroneous sequencing or (ii) erroneous annotation (iii) homopolymers insertion, (iv) high levels of undetermined nucleotides, (v) erroneous alignment, (vi) reverse complementary sequences, were deleted in a manual cleaning process (Figure 30). The database is automatically updated 4 times per year.

Currently, 6978 *SSU rRNA* gene sequences are stored in the PHYMYCO-DB database (4117 Ascomycota, 1733 Basidiomycota, 215 Chytridiomycota, 621 Glomeromycota, and 292 Zygomycota). PHYMYCO-DB also contains 431 *EF1- $\alpha$*  gene sequences (237 Ascomycota, 58 Basidiomycota, 10 Chytridiomycota, 4 Glomeromycota, and 122 Zygomycota).

Nineteen different alignment files are now available, with a specific alignment for the *SSU rRNA* and *EF1- $\alpha$*  gene sequences for each phylum. However, due to the high number of *SSU rRNA* gene sequences available for the Ascomycota phylum, alignment matrices at a lower systematic rank (i.e. order level) are provided.

### **Downloading GenBank data**

Users can easily select sequences by navigating from our interface through Genbank taxonomy and download them in a FASTA format file. The numbers of sequences stored in the database for each taxonomic level are given in brackets.

### **Aligning a dataset with ClustalW**

Users can launch a ClustalW 2.0 alignment on our back-end computer cluster, by selecting GenBank sequences through the GenBank Taxonomy Browser, by uploading their own personal sequences in a FASTA file or by combining both.

Before alignment, an outgroup file from our list or from a personal file can be added as an option.

Another way to align SSU rRNA genes and EF1- $\alpha$  genes is proposed, which makes use of the alignment files already computed, cleaned and saved in the database, for each phylum and each phylogenetic marker. Users will be able to add their own ALN format alignment file then execute a new ClustalW alignment. This tool should be very efficient for analysing the sequence diversity of fungal amplicons from an environmental sample, for example. It should also be very useful to highlight new fungal lineages.

### **Discussion**

The use of molecular markers to understand phylogenetic affinities between fungi has highlighted that only 3 of the 5 fungal phyla are monophyletic: Glomeromycota, Ascomycota and Basidiomycota. The last two phyla form a sister clade named Dikarya (James *et al.*, 2006, Liu *et al.*, 2006). The Chytridiomycota and Zygomycota phyla which formed the earlier emerging lineages are polyphyletic (Lutzoni *et al.*, 2004; Bar-Hen *et al.*, 2008). As demonstrated by Bar-Hen *et al.* (2008), refined methods of sequence analysis based on identification of the outlier nucleotides within the multiple alignment matrix should help to resolve uncertainties about the phylogenetic position of a given sub-tree or branch. Furthermore, sequence artefacts may occur due to the huge number of fungal sequences in the databases with erroneous annotations resulting in biased phylogenetic positioning or interpretation. The removal of such data is tedious and not always executed with sufficient care.

To our knowledge, PHYMYCO-DB is the first database to provide exhaustive access to fungal sequences for two phylogenetic markers (*SSU rRNA* and *EF1- $\alpha$*  genes) obtained from cultivated isolates as well as environmental samples. The aligned sequences available in PHYMYCO-DB are of high quality, as a result of efficient sequence cleaning, even though they have not yet been manually optimised (Figure 29). PHYMYCO-DB can be used with other tools commonly employed in ecological and phylogenetic analyses. For example, the

SILVA RNA database project (Pruesse *et al.*, 2007) provides different sets of aligned rRNA sequences (small and large-subunit rRNA sequences) for the three domains of life. It is used with the ARB software (Ludwig *et al.*, 2004) to infer phylogenies, generate primers and probes (Amann & Fuchs 2008).

Other databases such as the “Assembling the Fungal Tree Of Life” project (AFTOL; <http://aftol.org/index.php>), and UNITE (Kõljalg *et al.*, 2005) are restricted to fungal sequences. AFTOL allows users to extract information from cultivated fungal isolates (morphological and molecular data) as well as alignments for different fungal genes that have already been published. However, environmental sequences from non-cultured fungi are missing from the database. The UNITE system assembles ITS (Internal Transcribed Spacers) sequences from Basidiomycota and Ascomycota. This molecular marker is believed to be more taxonomically informative than *SSU rRNA* gene sequences for analysing groups of organisms that have emerged ‘recently’ and are closely related (Anderson & Parkin 2007) which is the case for Ascomycota and Basidiomycota. They are often used to resolve phylogenetic relationships at the species level or at the infraspecific level (Xu *et al.*, 2007). However, as ITS display high sequence variability, obtaining reliable alignments with this marker can be difficult (D’Auria *et al.*, 2006) since accurate comparisons are hindered by the accumulated homoplasy and the high frequency of insertion/deletion events. The use of the *SSU rRNA* sequences is interesting since new groups within all the fungal phyla including Ascomycota and Basidiomycota can be detected has shown in different papers published recently (Vandenkoornhuyse *et al.*, 2002a; Bass *et al.*, 2007).

The gene encoding the elongation factor *EF1- $\alpha$*  (*tef1* gene), generally present as a single copy gene, is involved in protein synthesis and displays a low mutation rate. This molecular target has been extensively used to resolve phylogenetic affinities between eukaryotic organisms (Baldauf *et al.*, 2000; Baldauf 1999; Helgason *et al.*, 2003), and particularly the sisterhood of animals and fungi (Baldauf & Palmer, 1993). This phylogenetic marker could be helpful to resolve phylogenetic relationships between closely related fungi (Helgason *et al.*, 2003; Moon *et al.*, 2002; Tanabe *et al.*, 2004). Studies of *SSU rRNA* genes and *EF1- $\alpha$*  genes could be combined to improve the resolution of fungal phylogenetic affinities.

### **Future changes and direction**

In this first version of PHYMYCO-DB, we have focused on the *SSU rRNA* and *EF1- $\alpha$*  genes, i.e. the targets most commonly used to analyse fungal diversity and evolution. Future developments of the database will include (i) the provision of alignment matrixes in which

outlier nucleotides identified from influence functions (Bar-Hen *et al.*, 2008) will be highlighted so that users can then delete these sites; (ii) scripts and cleaning methods to allow the user to recover sequences from other phylogenetic markers, such as genes coding for  $\beta$ -tubulin (*tub1*, *tub2*), actin (*act1*...), RNA polymerase II subunits (*RPB1* and *RPB2*); (iii) incorporation of all the finished fungal genomes available. Due to the great flexibility of our database, collaborative projects with users could also be developed, on request to the corresponding author.

## Conclusions

PHYMYCO-DB is a useful tool for scientists who want to explore fungal diversity within environmental samples or cultivars. Our database contains *SSU rRNA* and *EFI- $\alpha$*  sequences from all fungal phyla, and high quality expert curated alignments which can be used with confidence for phylogenetic analyses.

## Availability and requirements

The PHYMYCO-DB database is available via a web-based interface at <http://phymycodb.genouest.org/>, on the bioinformatics GenOuest platform web site. The web interface is divided into 2 parts. The first part, entitled “*DB admin*”, is restricted to the administrators and is used for cleaning and optimising the database. The second part, entitled “*DB explore*”, is publicly accessible to all users. This is where the downloading of sequences of interest is made possible. We also offer the possibilities to download alignment files prepared from subsets of the PHYMYCO-DB database and also to perform sequence alignments using ClustalW 2.0.

## Acknowledgements

We thank the ‘Ouest Génopole’ Bioinformatics service. This work is supported by GIS Génomique Marine and a grant from the ‘Total Corporate Foundation for biodiversity and the sea’. TLC has a PhD grant from the French ministry of research.

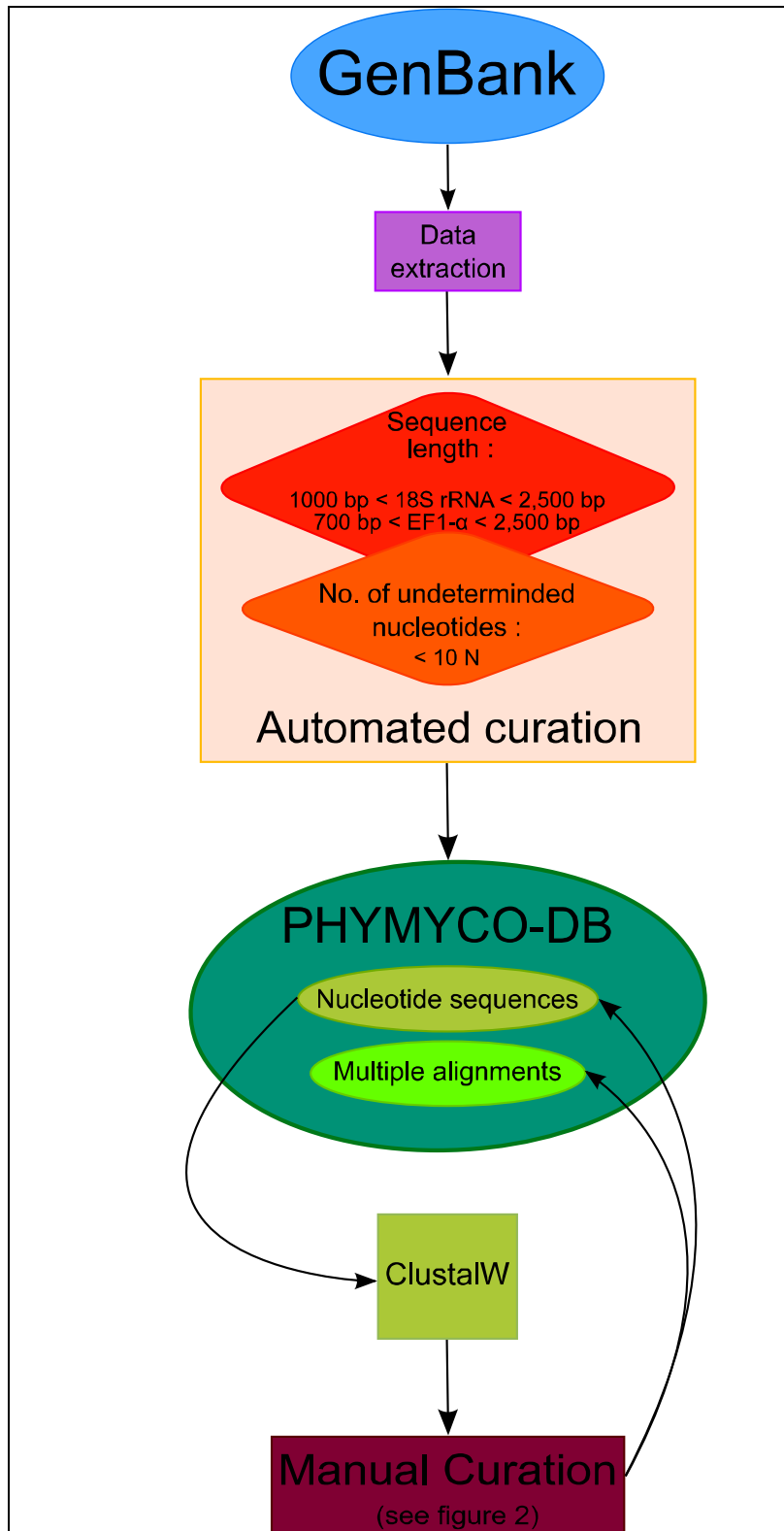


Figure 29: Flowchart of the data in the PHYMYCO-DB database. The arrows indicate the flow of gene sequences extracted from the Genbank database, through the automated and manual curation steps.

AY761069.1 Togninia	TCGACCCCGGAGGAGGAGCCTGAG-RRACG-----GCTA
AY761068.1 Togninia	TCGACCCCGGAGGAGGAGCCTGAG-RRACG-----GCTA
AY761089.1 Lentomitella	TCGACCCCGGAGGAGGAGCCTGAG-RRACG-----GCTA
AF242267.1 Rhamphoria	TCGACCCCGGAGGAGGAGCCTGAG-AGACGTAAAGTGTCCCTCTATAAACCTGATATTTTTCCGTTGAATCTCTGCTAACATATTCTACAGGGCTA
DQ471003.1 Ophiostoma	TCGACCCCGGAGGAGGAGCCTGAG-RRACG-----GCTA
AJ243295.1 Ophiostoma	TCGACCCCGGAGGAGGAGCCTGAG-RRACG-----GCTA
DQ789980.1 Melanotaenium	GATRACGAA-CGAGACC-----TTGACCTGC-
AJ495238.1 Hyphoxyma	CCCGRACCTC-TTGCGAGTTTTGGAGGTATGCGCAGCCCACTCCTAGCTCTTAGCTAGGAGG-
D31649.1 Bullera	GATRAC-GRACGAGACC-----TTRACCTGC-
AB032635.1 Cryptococcus	GATRAC-GRACGAGACC-----TTRACCTGC-
AY083224.1 Filobasidiella	GATRAC-GRACGAGACC-----TTRACCTGC-

(i) long insertions

AF164337.1 Monoblepharella	AAATCC-CTTRACGAGGACCAATTGGAGG---GCAAGTCT-
AF164336.1 Monoblepharella	AAATCC-CTTRACGAGGACCAATTGGAGG---GCAAGTCT-
AF164335.1 Monoblepharella	AAATCC-CTTRACGAGGACCAATTGGAGG---GCAAGTCT-
AY546682.1 Monoblepharella	AAATCC-CTTRACGAGGACCAATTGGAGG---GCAAGTCT-
AY854021.1 Synchytrium	-----GCAAGTCT-
Y17504.1 Hyaloraphidium	AAATCC-CTTRACGAGGACCAATTGGAGG---GCAAGTCT-
AY218377.1 Sparassis	CTGGGGGGGA-----AGRATTT---GGGAT
AY218378.1 Sparassis	CTGGGGGGGATATGCTCGCAGGGCTGAACTTAAAGGATTT
AY218380.1 Sparassis	CTGGGGGGGATATGCTCGCAGGGCTGAACTTAAAGGATTT
AY705950.1 Grifola	CTGGGGGGGATATGCTCGCAGGGCTGAACTTAAAGGATTT
AY218381.1 Sparassis	CTGGGGGGGATATGCTCGCAGGGCTGAACTTAAAGGATTT
AY218375.1 Sparassis	CTGGGGGGGATATGCTCGCAGGGCTGAACTTAAAGGATTT
AY218373.1 Sparassis	CTGGGGGGGATATGCTCGCAGGGCTGAACTTAAAGGATTT

(ii) deletions

AF516542.1 Polyporus	GGTTGAGCTGGCCTTAC-----GAGGCATGTGCTCGCCCC-----TGC'
AF516518.1 Lentinus	GGTTGAGCTGGCCTTCC-----GAGGCATGTGCAAGCCCC-----TGC'
AF516539.1 Polyporus	GGTTGAGCTGGCCTTCC-----GAGGCATGTGCAAGCCCCCTGTTWWWNCNCTTT'
AF516524.1 Polyporus	GGTTGAGCTGGCCTTCC-----GAGGCATGTGCAAGCCCC-----TGC'
AF516571.1 Polyporus	GGTTGAGCTGGCCTCAT-----GAGGCATGTGCAAGCCTCTG-----CTT
EF550481.1 Pichia	GCTGAACTTAAAGGATTTGACCGGAGGGCACCACCAGGAGTGGAGCCTGCGGCTTAAATTTG
EF550477.1 Pichia	GCTGAACTTAAAGGATTTGACCGGAGGGCACCACCAGGAGTGGAGCCTGCGGCTTAAATTTG
AJ508275.1 Pichia	GCTGAACTTAAAGGATTTGACCGGAGGGCACCACCAGGAGTGGAGCCTGCGGCTTAAATTTG
AJ508271.1 Candida	GCTNNNNCTTAAWNNANNTNNNNNANNNNGCACCACCAGGAGTNNNCRNCTSCGCCTTAAATTTRE

(iii) high number of undetermined nucleotides

AF541989.1 Uncultured	----CTGGACCTGGTGAATTTCCCGTGTGAGTCBAATTAAGCCGC--AGTCCACT
AY452676.1 Uncultured	----CTCACCTC---TATGACGCTCCATTCCAGGAGACTTGTGCAC--GGTCCATT
DQ150132.1 Sebacinaceae	----CTCACCTC---TATGACGCTCCATTCCAGGAGACTTGTGCAC--GGTCCGTC
AB211931.1 Ustilago	----CCBACCCAGCTTTGGAAATTCGATAGAGTGGGTCAAGTGGGC--ACTGAGT-
AB211932.1 Ustilago	----CCBACCCAGCTTTGGAAATTCGATAGAGTGGGTCAAGTGGGC--ACTGAGT-

(iv) wrong annotation or chimeric sequences

Figure 30: Visualisation of different examples of sequences deleted from manual curations (ClustalX 2.0 had been used for the alignment); On the left side, name of sequences highlighted in blue are sequences removed from PHYMYCO-DB: (i) presence of long insertions; (ii) presence of deletions; (iii) sequences containing high numbers of undetermined nucleotides; (iv) wrong annotation or reverse complementary sequence.

## **Partie IV : Diversité fongique dans les écosystèmes hydrothermaux marins profonds**

---



## Résumé

Depuis leur découverte en 1977, les écosystèmes hydrothermaux ont révélé de nombreuses espèces aux caractéristiques originales. Très rapidement, de nombreuses études se sont portées sur le compartiment microbien. Cependant les champignons dans ces milieux ont été très peu étudiés. Des études moléculaires s'intéressant aux microeucaryotes dans ces environnements ont prouvé leur présence (López-García *et al.*, 2003). Des levures ont également pu être isolées de ces environnements et récemment des champignons responsables de maladies de modioles ont été caractérisés (Bass *et al.*, 2007 ; Van Dover *et al.*, 2007).

L'étude que nous avons menée s'est intéressée à la diversité fongique de différents écosystèmes hydrothermaux marins (isolés de la Ride Est-Pacifique et de la Ride Médio-Atlantique), afin d'une part de mieux caractériser les organismes présents et d'autre part dans l'espoir de mieux comprendre les processus évolutifs ayant mené à la diversification des champignons. Cette dernière hypothèse s'appuie sur la théorie considérant les écosystèmes hydrothermaux comme milieux berceaux de la Vie.

Pour résoudre ces questions nous avons donc choisi d'analyser la diversité fongique sur la base de l'amplification de l'ARNr 18S, à partir de l'extraction de l'ADN total des différents échantillons dont nous disposons. Les échantillons étaient de différente nature, minérale (fragments de roches) ou animale (*Bathymodiolus azoricus* et *Alvinella pompejana*). Des analyses phylogénétiques (Neighbor Joining, Maximum Parsimony, Maximum Likelihood) ont été effectuées pour approcher cette diversité. La base de données de séquences fongiques PHYMYCO-DB établies lors de cette thèse (présentée dans le chapitre III) a été utilisée pour mener ces analyses. Une étude de PCR quantitative a également été menée sur ces mêmes échantillons afin de quantifier la fréquence des champignons dans ces échantillons. En parallèle de cette étude moléculaire, et en collaboration avec l'ESMISAB (Brest) nous avons tenté d'isoler des souches fongiques en conditions de laboratoires, à partir des mêmes échantillons ; les séquences de ces souches cultivées ont également été intégrées dans nos analyses phylogénétiques.

La culture des souches, en conditions de laboratoire (milieu salé 3% NaCl) à température ambiante, à l'obscurité, a permis d'isoler 21 souches, toutes appartenant au phylum des Ascomycota. Parmi ces 21 souches, 12 phylotypes ont été retrouvés. L'analyse directe du gène codant l'ARNr18S par ciPCR (PCR indépendante de la culture) a révélé la présence de 9 phylotypes distincts, appartenant aux phyla Basidiomycota et Chytridiomycota. Il apparaît ainsi que les conditions de culture n'aient pas permis d'isoler ces organismes,

renforçant l'intérêt de coupler ces deux approches. La PCR quantitative réalisée sur différents échantillons a permis d'estimer entre  $1.91.10^5$  et  $1.35.10^7$  le nombre de copies de gènes codant l'ARNr18S fongique, par microgramme d'ADN génomique, nous permettant de rejeter une origine contaminante des phylotypes détectés.

L'un des résultats majeurs de cette étude est la découverte d'un phylotype inconnu de Chytridiomycota, par ciPCR, qui représente une branche profonde de ce phylum, et donc une branche profonde du Règne fongique. Ce résultat, couplé aux analyses d'horloges moléculaires effectuées précédemment et en accord avec la théorie d'émergence de la Vie en milieu hydrothermal marin, permet de supposer une émergence et surtout une diversification des Champignons en milieu marin.

Une connaissance plus complète du compartiment fongique au sein de ces écosystèmes permettrait ainsi de mieux comprendre leur rôle au sein de ces environnements. Une valorisation biotechnologique de cette diversité pourrait également être envisagée. Ce travail est présenté sous forme d'une publication, soumise à *Applied and Environmental Microbiology*.

# Fungal Diversity in Deep Sea Hydrothermal Ecosystems

Thomas Le Calvez<sup>1</sup>, Gaëtan Burgaud<sup>2</sup>, Stéphane Mahé<sup>1</sup>, Georges Barbier<sup>1</sup>, Philippe Vandenkoornhuys<sup>2</sup>

1 UMR 6553 CNRS ECOBIO, FR/90 IFR 2116 CAREN, Université de Rennes 1, Campus de Beaulieu, 35042 Rennes cedex, France

2 Laboratoire de Biodiversité et Ecologie Microbienne (EA 3882), Université de Bretagne Occidentale, ESMISAB, Technopôle Brest Iroise, 29280 Plouzané, France

---

## Abstract :

Deep marine hydrothermal ecosystems are considered oases of life in oceans. Since their discovery in the late 1970's many endemic species of bacteria, archaea and other organisms such as annelids and crabs have been described. Considerable knowledge has been acquired about the diversity of (micro)organisms in these ecosystems but the diversity of fungi has not been studied to date. However these are regarded as key organisms in terrestrial ecosystems because of their ecological functions and especially their ability to degrade organic matter. This lack of knowledge reflects the widely held belief that fungi are terrestrial organisms. The first inventory of such organisms in deep hydrothermal marine environments was established in this study. Fungal diversity was investigated by analysing the small subunit ribosomal RNA gene sequence amplified by culture-independent PCR in DNA extracts from hydrothermal samples and from a culture collection that had been successfully established. Our work revealed an unsuspected diversity of species within 3 of the 5 fungal phyla. We found a new branch of Chytridiomycota forming an ancient evolutionary lineage. Many of the identified species are unknown, even at higher taxonomic levels within the Chytridiomycota, Ascomycota and Basidiomycota. This work opens the way to new studies on the diversity, ecology and physiology of fungi in oceans and might stimulate new prospecting for biomolecules. From an evolutionary point of view, the emergence and diversification of fungi in ocean can no longer be excluded.

## Introduction

Since the discovery of hydrothermal vent ecosystems thirty years ago, unexpected species diversity has been revealed and shed light on the functional coupling between geosphere and biosphere. When submersibles dive to the seafloor, they bring numerous organisms back to the surface, and have allowed the description of nearly two new species per month (Fisher *et al.*, 2007). Deep-sea hydrothermal ecosystems are considered as hotspots of microbial diversity on the seafloor. Indeed, they are biomass-productive ecosystems due to the large range of chemical compounds released by the polymetallic sulfite chimneys or “black smokers” which represent a huge quantity of chemical energy that is available for liberation (Ramirez-Llodra *et al.*, 2007). The vent fluid, having been heated close to the magma chamber, can attain 400°C when emitted. It is also characterized by a lack of dissolved oxygen, strong acidity (pH 2-3), a high concentration of electron donors, i.e., reduced compounds such as methane and hydrogen sulfide, and the presence of heavy metals (Tivey, 2007). Continual mixing with the cold ocean waters (2-4°C), rich in electron acceptors, then creates a dynamic chemical disequilibrium and a source of energy for the microorganisms that control the rates of redox reactions (Jørgensen & Boetius, 2007).

Each ridge displays varied geochemistry and the vent fluid differs, even at scales as small as the fractures, pipes and porosities in the black smokers, thereby creating a diversity of microhabitats for biota (Fisher *et al.*, 2007; Ramirez-Llodra *et al.*, 2007; Jørgensen & Boetius, 2007). Microbes have colonized these different microsites and an impressive diversity of bacteria and archaea species has been described and even more has been observed. Some are chemolithoautotrophic primary producers and at the base of the trophic web in these habitats. In addition to these key microorganisms, other primary producers, like the photosynthetic green sulfur bacteria, are reported to be able to harvest energy from geothermal radiation (Beatty *et al.*, 2007). The organic compounds produced by the primary producers are consumed by heterotrophic microbes and animals through symbioses, filtering of free-living microorganisms or grazing of biofilms.

Considering hydrothermal vent cycling of carbon and comparing it to the carbon cycling in terrestrial ecosystems, key organisms, the fungi, have not yet been studied. Contemporary organisms living in land ecosystems are well adapted. However, the first colonizers of land faced a harsh physical environment (Selosse & Le Tacon, 1998). Thus it is suggested that early colonization of land was only possible by a macroevolutionary jump (i.e. acquisition of

multiple competences during a single evolutionary step) (Selosse & Le Tacon, 1998). It has been suggested that the establishment of eukaryotes on land may only have been possible as a consequence of the association of a fungus and a phototroph (Selosse & Le Tacon, 1998). The earliest known fossil fungus (460 My) was supposed to form arbuscular mycorrhiza (Redecker *et al.*, 2000). The idea that the colonization of land was made possible by this symbiotic association is supported elsewhere (Simon *et al.*, 1993). Based on molecular clock estimates, some authors suggest that fungi appeared a long time before Devonian land colonization. The major radiations within fungi may even have appeared during the Precambrian (1460 My to 960 My) (Heckmann *et al.*, 2001), 300 My or more before land colonization. These analyses are currently under discussion, due to the absence of fossil evidence to support them and because of possible bias in the molecular clock estimations (Berbee & Taylor, 2006). However, assuming that the proteins estimations are correct, it could be hypothesized that fungi emerged and diversified in water before their colonization of land. An ancestral character, the unwalled flagellated gamete, is found in old lineages of fungi. This is generally regarded as an advantage for dispersal and reproduction in water. The loss of flagella in fungi appears to have been a single evolutionary step (Liu *et al.*, 2006), and has been used as an argument to sustain the dogma that higher fungi diversified on land, before a secondary colonization of water ecosystems. It remains a common belief among ecologists, microbiologists and mycologists that fungi are found (almost) exclusively on land. A recent study has shown that this loss of flagella may have occurred at least four times during fungal evolution (James *et al.*, 2006) suggesting a more complicated evolutionary paradigm than previously thought. Despite the view that fungi diversified in terrestrial ecosystems, cultural approaches have resulted in the identification of a small, ecologically defined group of marine higher fungi that includes filamentous ascomycetes, their anamorphs, and yeasts (Kohlmeyer & Kohlmeyer, 1979). We suggest an alternative hypothesis to the loss of flagella. It can be argued that the loss of motile gametes in fungi has been compensated by the resistance and long-range dispersal of both mitotic and meiotic spores. We suggest that this evolutionary innovation in eukaryotes should have led to colonization and long term persistence in many new environments, including land. From the molecular clock predictions (Heckman *et al.*, 2001) and partly in agreement with our computations (P. Vandenkoornhuyse, unpublished), we suggest that this evolutionary event might have appeared approximately 1,200 to 1,000 My ago, and therefore in oceans. If true, we would expect to find a large variety of fungi, including those producing non-flagellated spores, in oceanic samples even within isolated habitats.

Consequently, the main aims of the present work were to test for the presence of fungi in deep hydrothermal ecosystems, by establishing a culture collection and by a culture-independent approach involving direct amplifications of the fungal SSU rRNA gene from environmental samples, and then to analyze and evaluate the diversity of such fungi in these particular environments. The deep marine hydrothermal ecosystem was an interesting target to test our hypotheses for various reasons. Firstly, the hydrothermal ecosystem is suggested to be the ‘cradle of life’ (Wächtershäuser, 1990; Huber & Wächtershäuser, 1998; Koonin & Martin, 2005) even if more recently, it has been shown that life emerged first in a mesophilic and diversified in thermophilic conditions (Boussau *et al.*, 2008). Secondly, the hydrothermal conditions were very mundane during the Precambrian (Robert & Chaussidon, 2006). Thirdly, use of general eukaryotic primers to perform ciPCRs, has already led to the detection of fungi close to hydrothermal vents (López-García *et al.*, 2003). Fungal sequences have also been detected in deep sea anoxic sediments (Dawson & Pace 2002; Edgcomb *et al.*, 2002; Bass *et al.*, 2007) and barophilic fungal strains have been isolated from this environment (Damare *et al.*, 2006).

## **Material and methods**

### **Studied sites and sampling**

Samples were collected during 2 oceanographic cruises: i) HERO (30/09/1991-04/11/1991) on the East Pacific Rise at Elsa site (12°48’N, 103°57’W; -2630 m) and ii) MARVEL (29/08/1997-13/09/1997) on the Mid-Atlantic Ridge at the Menez Gwen (37°51’N, 31°31’W; -860 m) and Lucky Strike sites (37°17’N, 32°16’W; -1700 m). Deep sea sampling was performed using the DSV Nautilie. Support research vessels were “Nadir” for the HERO cruise and “Atalante” for MARVEL ([www.ifremer.fr/fleet//index.php](http://www.ifremer.fr/fleet//index.php)). Animals and rocks were collected using sampling boxes, washed on board, disinfected with ethanol and filled with sterile seawater. All the boxes were waterproof to prevent contamination from the water column during ascent. They were opened for sampling during dives, then closed and placed in the insulated crate of the submersible to maintain a low temperature until their subsequent use on board.

## **Samples preparation & fungal cultures**

On board, in the lab, sterility was obtained with a Bunsen burner and a vertical laminary flow hood. Solid samples were taken from their container with sterile strips, laid on a sterile Petri dish, rinsed well with sterile seawater, then crushed with a sterile pestle and mortar. After this, the samples were aliquoted and preserved by deep-freeze storage (-80°C) until subsequent use.

During the HERO cruise, Sabouraud (Sab) chloramphenicol solid medium (AES Laboratory) was used for aerobic enrichment cultures at 30°C and atmospheric pressure. During the MARVEL cruise, 5 solid culture media were used. Ac medium contained, per liter: potato starch (Sigma) 5 g, yeast extract 0.5 g, peptone 1 g, sea salts 30 g and PIPES buffer (Sigma) 6.05 g. Starch was replaced by cellobiose (Sigma) 5 g in Cc medium, by glucose (Sigma) 5 g in Gc medium, by arabic gum 5 g and olive oil 5 g in Lc medium, by xylan oat spelts 5 g in Xc medium. Pc medium contained, per liter: brain heart infusion (Difco) 9 g, NaCl 23g and PIPES buffer 6.05 g. The pH was adjusted to 7.5 with NaOH 4N. The 5 media were supplemented, per liter, with agar 15 g and chloramphenicol 500 mg. Cultures were grown aerobically at 25°C (ambient temperature) and atmospheric pressure. Finally, all strains were able to grow on GYPS medium containing, per liter: glucose 1 g, yeast extract 1 g, peptone 1 g, starch 1 g, sea salts 30 g. Pure cultures were obtained and kept in culture collection by continuous culture and then freeze-dried at -80°C.

**Extraction of DNA** Each sample was homogenized with sterile glass beads and ground for 28s at 30 rpm in a Beat Beater (MM 301, Retsch). Different methods of DNA extraction were tested to obtain amplifiable fungal DNA. The method chosen was the DNeasy Plant mini kit (QIAGEN) following the manufacturer's recommendations. DNA from cultivated strains was extracted with Fast DNA Spin Kit (MP Biomedicals).

## **Cloning and sequencing**

The fungal small subunit ribosomal RNA (SSU rRNA) genes were amplified by PCR using the set of primers MH2 (5'TTCGATGGTAGGATAGAGG3') and MH4 (5'GTCTCACTAAGCCATTC3') (Vandenkoornhuysse & Leyval, 1998) for environmental sequences and with NS1 (5'GTAGTCATATGCTTGTCTC3') and ITS5R (5'CCTTGTTACGACTTTTACTTCC3') primers for cultivated strains (White et al., 1990). The AU2/AU4 primers set (Vandenkoornhuysse et al., 2002a) did not amplify any of the

samples analyzed. DNA was amplified by touch-down PCR with an initial step at 94°C for 1.30 min then 37 cycles of 94°C for 30s, 48°C for 1 minute 15 sec (with a decrease of 0.1°C at each cycle) and 72°C for 1 min 30 sec . The PCR amplification ended with a final elongation step at 72°C for 10 min. For those samples which did not yield any product we used another primer set, modified from Smit et al. (1999): EF3 (5'TCCTCTAAATGACCAGTTTG3') and EF4 (5'GGAAGGGNTGTATTTATTAGAT3') with the same PCR conditions except for the annealing temperature (53°C instead of 48°C). The amplification fragments were purified with High Pure PCR product purification kit (ROCHE) and were cloned in pGEM-T vector (Promega) and DH5 $\alpha$  competent cells (Gibco BRL Life Technologies). From each clone library, 16 to 120 randomly selected positive clones were sequenced (ABI-PRISM Dye Terminator Cycle Sequencing Ready Reaction Kit, Perkin Elmer) for both strands using primers T7 or SP6 for environmental sequences and NS1, NS3 (5'GCA AGTCTGGTGCCAGCAGCC3') and ITS5R (White et al., 1990) for the culture sequences. The contig of the two sequences strands was calculated using Sequencher 4.6 (GeneCodes) and alignment uncertainties were checked manually. Chimeric artifacts were deleted from the dataset using CHIMERA\_CHECK 2.7 (Ribosomal Database Project, <http://rdp.cme.msu.edu/html/analyses.html>).

The sequences reported in this paper have been deposited in the GenBank database (accession nos. EF638466-EF638515, sample MV2E1; EF638516-EF638564, sample H18E12; EF638565-EF638614, sample MV2E2; EF638615-EF638636, sample MV2E3; EF638637-EF638654, sample MV5E1; EF638655-EF638686, sample MV5E2; EF638687-EF638706, cultivable strains). The matrix is available upon request to the corresponding author.

### **rRNA fungal database**

A fungal SSU rRNA gene database was created so that the detected sequences could be analyzed (<http://phymycodb.genouest.org/>). Algorithms were used to recover SSU rRNA gene sequences for each fungal phylum from the GenBank/EMBL/DDBJ databases. Filters were used to eliminate those sequences that were too short (< 1000 bp), too long (> 2500 bp), or of poor quality, or, having included at least 10 consecutive undetermined nucleotides (N) (Le Calvez *et al.*, submitted). A multiple alignment was performed for each phylum, using Clustal X 1.81, in a matrix containing all sequences from our databases and our hydrothermal sequences.



After using this protocol and completing the various analyses, we were able to recover 1733 sequences from Basidiomycota, 215 from Chytridiomycota, 4117 from Ascomycota, 292 from Zygomycota and 621 from Glomeromycota, representing all the different branches of fungal phylogeny. Each phylum was then subjected to a multiple alignment step followed by neighbor-joining analyses. Phylogenetic trees were visualized by Treeview 1.6.6. The phylogenetic neighbors closest to the analyzed environmental sequences were selected and subsequent phylogenetic analyses were performed (see below).

### **Diversity and phylogenetic analyses**

Rarefaction curves were computed for each sample. The number of species was quantified for 100 random combinations of 1 to N sequences and also by performing 100 bootstrap pseudoreplicates (Colwell 2006) implemented in EstimateS. A multiple alignment was performed using CLUSTALX 1.81 (Thompson et al., 1997) and refined by eye. This included our environmental sequences, sequences from the fungal cultures isolated, and the most relevant representative fungal SSU rRNA gene sequences of the Chytridiomycota, Basidiomycota and Ascomycota phyla (93 sequences including 18 representative environmental sequences and 26 sequences from fungal cultures). The alignment was performed on 1145 nucleotides as detailed above. Phylogenetic analyses were done as follows. CLUSTALX 1.81 was used for neighbor-joining (NJ) phylogenies with distance correction, gaps omission, and 1000 bootstrap pseudoreplicates. PAUP 4.0β10 was used for maximum parsimony (MP) using a heuristic tree search with 500 replicates of random addition, tree bisection and reconnection (TBR) as branching algorithm, each of the 500 bootstrap iterations utilizing 10 random-addition replicates and TBR branch swapping. In addition, the data were analyzed using the ML-GTR + I + G procedure, with 50 iterations. Modeltest 3.7 software (Posada & Crandall, 1998) was used to select the model.

After these analyses, phylotypes were specified using a cutoff of 98% (pairwise distance computed using PAUP 4.0β10).

### **qPCR assays**

Reactions were performed in a final volume of 10μl using iQ SYBR Green Supermix (Biorad) which contained SYBR Green PCR buffer, 2.7 μM each of dATP, dTTP, dGTP and dCTP, 0.42 U of iTaq DNA polymerase (Biorad). 0.35 μM of each primer (MH2:

5'TTCGATGGTAGGATAG 3' and FungqPCR1 : 5'TGTCGGGATTGGGTAATTT 3') were then added to the mix.

All reactions were performed in optical tubes (Biorad), with 8.5 µl of the master mix added first, followed by 1.5 µl of template. The tubes were sealed with microseal films (Biorad). All reactions were performed with the Chromo 4 thermocycler (MJ Research), using an initial denaturation at 94°C for 3 minutes to activate the enzyme, followed by 35 or 40 denaturing steps at 94°C for 30 seconds, annealing-extension at 48°C for 45 seconds, followed by plate reading. The dissociation curve from 65°C to 95°C was measured after the last qPCR cycle. All data were analyzed using Opticon Monitor 3 (MJ Research).

This approach was used to study six samples in triplicate, in two independent runs. The three samples which exhibited the strongest signals in these two preliminary runs were compared in a final run. Seven different plasmid concentrations were used to construct a standard curve ( $10^9$  to  $10^3$  copies in 1.5 µl) for absolute quantification. The number of copies in the standards was calculated using the following formula:

$\text{Molecules}/\mu\text{l} = a / (\text{plasmid length} * 660) * 6.022 \cdot 10^{23}$ ; where  $a$  is the plasmid concentration ( $\mu\text{g}/\mu\text{l}$ ), 660 is the average molecular weight of one base pair and  $6.022 \cdot 10^{23}$  the molar constant.

## Results

### Diversity of the culture collection

The cultivation and isolation of microbes (bacteria and archaea) from deep hydrothermal samples is a generally difficult task that includes many constraining manipulations. Nevertheless, we were able to establish a fungal collection with different samples from several origins without using complex culturing procedures such as pressurized fermentors (see Material and Methods). These fungi were isolated using a classic medium supplemented with sea salt, at one atmosphere and 25°C. Physiological analyses demonstrated that the strains were able to grow in deep-sea salinity (3% weight/volume of NaCl) and at low temperatures (data not shown). The SSU rRNA gene sequence analyses revealed that all 21 isolates in the collection belonged to the phylum Ascomycota, as did previously isolated marine strains (Kohlmeyer & Kohlmeyer, 1979). A subset of the cultivated Ascomycota belonged to the Dothideomycetes (Figure 31, phylotypes 14-16) and Sordariomycetes (Figure

31, phylotypes 17, 18, 20) classes, and at least one new species from each of these groups was discovered. Black yeasts were also detected, with two phylotypes of *Exophiala* (order Chaetothyriales) (Figure 31, phylotypes 10-13), a group that includes pathogens and saprophytes. Finally, one phylotype was obtained from cultures clustered with the mitosporic Ascomycota (Figure 31, phylotype 19), which have been described as ‘common’ pathogenic fungi in tidal seawater (Berbee *et al.*, 1992; Shearer *et al.*, 2007). As expected from our knowledge, the spores produced by the cultivated isolates were not flagellated, even if some isolates appeared to be specific to the oceanic habitat (Burgaud *et al.*, 2009).

### **Fungal diversity by culture-independent approach**

We investigated the fungal SSU rRNA genes from environmental samples (Table 6) directly amplified by ciPCRs using fungal-specific primers. A customized database was constructed to analyze in depth the detected SSU rRNA gene sequences. This database contained all available fungal SSU rRNA sequences that passed filters for length and quality (144 Chytridiomycota; 266 Zygomycota; 418 Glomeromycota; 2,327 Basidiomycota; and 4,270 Ascomycota), with at least one representative of each known fungal genus in Genbank. All these 7,425 sequences, which represent the whole fungal tree of life were used to determine, with high confidence, the phylogenetic affinities of the new sequences. A secondary matrix of sequences was then designed from these computations along with BLASTn analyzes (Altschul *et al.*, 1990), that contained the closest relatives of the analyzed environmental SSU rRNA sequences (for each phylum, Neighbor Joining phylogenetic trees using the entire datasets are provided as Supplementary Material Figure 33, 34, 35 and 36). A total of 20 distinct phylotypes from the ciPCR and the fungal culture collection was detected (Figure 31). The ciPCR results revealed a previously unsuspected diversity of fungal phylotypes in the vent ecosystems, with 9 distinct phylotypes of which 5 were new at the genus or higher taxonomic level within Chytridiomycota (Figure 1, phylotypes 1, 2) and Basidiomycota (Figure 31, phylotypes 3, 4, 9).

On the one hand, the Basidiomycota phylotypes detected belonged to the Agaricomycotina (Figure 31, phylotypes 3-8), mostly to the genera *Cryptococcus*, a phylotype already detected in another hydrothermal area (Takishita *et al.*, 2006) and their anamorphs *Filobasidium* (EPR) which are of economic, agricultural and medical importance. Phylotype 9 (Figure 31), close to uncultured fungi, had already been detected in deep sea and hydrothermal sediments (Bass *et al.*, 2007) and fluids (López-García *et al.*, 2007). Moreover, Homobasidiomycete yeasts from the order Auriculariales (Figure 31, phylotype 7) were also

found. On the other hand, the retrieved Chytridiomycota phylotypes belonged to species that have not been described previously (Figure 31, phylotypes 1, 2) and, in particular, to an apparent ancient evolutionary lineage in the order Chytridiales. These sequences shared only 95.7 % identity with their closest relative, *Chytridium polysiphoniae* (Figure 31). The phylogenetic position of this *C. polysiphoniae* is uncertain because it is a pathogenic fungus (Küpper *et al.*, 2006), and such fungi tend to show a long-branch attraction bias (Figure 31). When *C. polysiphoniae* was removed, along with the other ambiguously placed pathogenic Chytridiomycota, the new group of species formed one of the most ancient branches of the fungal kingdom (see Figure 36). It is most probable that the phylogenetic position of these newly discovered fungi was determined correctly because the branches within this group were short; a result not expected if the terminals were pathogens.

Different fungal communities were discovered when the fungal diversity in MAR vs. EPR (East Pacific Rise) was compared, these two ridges being known to possess distinct endemic animal communities (Wolff, 2005) (i.e. *B. azoricus* and *A. pompejana* samples respectively). However, Agaricomycotina (Figure 31, phylotypes 3-8) and Dothideomycetes (Figure 31, phylotypes 14-16) sequences were found in both ecosystems. Agaromycotina fungi seem to be widely distributed because they were also revealed in deep sea sediments (Gadanhó & Sampaio, 2005; Nagahama *et al.*, 2001) and methane seeps (Takishita *et al.*, 2006).

### **Estimation of fungal SSU rRNA gene copy number in environmental samples**

Quantitative PCR analyses were performed to indirectly assess the fungal SSU rRNA gene copy number in environmental samples taken from sites on the Mid Atlantic Ridge (MAR) and East Pacific Rise (EPR) to test the possibility of contamination of the environmental samples by exogenous fungi. In this case the number of fungal SSU rRNA gene copies would be expected to be low. A set of specific fungal primers was therefore designed on the basis of the diversity analyses, as described in Material and Methods. Strikingly, the number of SSU rRNA gene copies ranged from  $1.91 \cdot 10^5$  to  $1.35 \cdot 10^7$  per microgram of genomic DNA. Thus, this analysis clearly allowed us to reject the hypothesis of an exogenous origin of all these fungi (i.e. contamination of the samples by the water column or during conditioning). As the highest occurrence of fungi was not observed in the bodies of animals (*Bathymodiolus azoricus* or *Alvinella pompejana*) but on the exterior of the shell (*B. azoricus*) and the surrounding rock sample (Table 6), we hypothesized that most of these fungi were not biotrophic.

## Discussion

One of the main results of the study presented herein is the evidence of an old Chytridiomycota lineage, unknown either at the genus level or at a higher taxonomic rank (Figure 31 and Figure 36). As protein clock analyses (Heckman *et al.*, 2001) had suggested that fungi emerged in oceans approximately a billion years ago during the Proterozoic era of the Precambrian, deep branches, such as this unknown phylotype, were expected from our working hypotheses. It is thus possible that the emergence and initial diversification of fungi occurred in a marine environment. This study provides an argument in favor of this hypothesis but do not provide the demonstration. The flagellum loss of motile gametes in fungi, which likely occurred more than once, (James *et al.*, 2006) does not invalidate our conclusion since this loss can be compensated in terms of fitness by the resistance and long-range dispersal of both mitotic and meiotic spores in aquatic environments. If the molecular clock estimates are correct, this hypothesis implies that fungi diversified in oceans before colonizing the land during the late Silurian or early Devonian period (i.e. approximately 400 million years ago). In this study we found a variety of fungi belonging to 3 of the 5 fungal phyla, which is one argument supporting this assumption even if secondary colonizations from land to marine ecosystems cannot be exclude. This conclusion contradicts the widely accepted hypothesis that the diversification of fungi occurred on land. The possible early colonization of hydrothermal habitats and the ecological function of the deep branching Chytridiomycota are being addressed in an ongoing metagenomic analysis of the fungi living in deep marine hydrothermal ecosystem.

### Fungal diversity in marine hydrothermal ecosystems

Despite the effectiveness of the primer set used here for amplification of Ascomycota (Vandenkoornhuyse & Leyval, 1998), we did not find any of these Ascomycota by ciPCR. Rarefaction curves from resampling and pseudo-resampling (bootstrap) procedures indicated that the number of sequences analyzed might be insufficient to represent the entire fungal diversity within all samples. One or two additional phlotypes might be discovered (Figure 32) by doubling the sequencing effort. We suggest that the abundance of Ascomycota in the environmental sample was extremely low and that cultivation resulted in amplification of these rare organisms. This could explain why ciPCR and subsequent sequencing failed to

detect them. Should this be the case, it is likely that the fungal diversity described here has been underestimated.

The observation that a specific fungal community was found for each sample might be explained by the diversity of the surrounding physico-chemical conditions (i.e. gradients of temperature, pH, etc...). A low redundancy of phylotypes was observed in all the samples analyzed. This allows us to argue that the diversity of fungi in hydrothermal ecosystems is high and that the recently reported numbers of living fungal species (Hawksworth, 2001) are substantial underestimates. Given the high fungal SSU rRNA gene copy number found for a given sample (up to  $1.91 \cdot 10^5$  copies /  $\mu\text{g}$ ), it can also be suggested that these fungi might be involved in important ecological functions. Moreover, 3 of the phylotypes found in this work had already been detected by ciPCR in previous studies (López-García *et al.*, 2007; Bass *et al.*, 2007). This also argues for rejection of the hypothesis that these sequences originated from contamination.

As we have no data on how the identified fungi live, harvest energy, and interact, their roles in these hydrothermal ecosystems remain unclear. Since fungi are considered key organisms in land ecosystems, they could be as important in oceanic ecosystems especially at the bottom of oceans. We need better knowledge of the fungal diversity and functions in deep hydrothermal habitats and, more generally, in oceans. These numerous, uncharacterized fungi might become a new source of drugs and biotechnological discoveries.

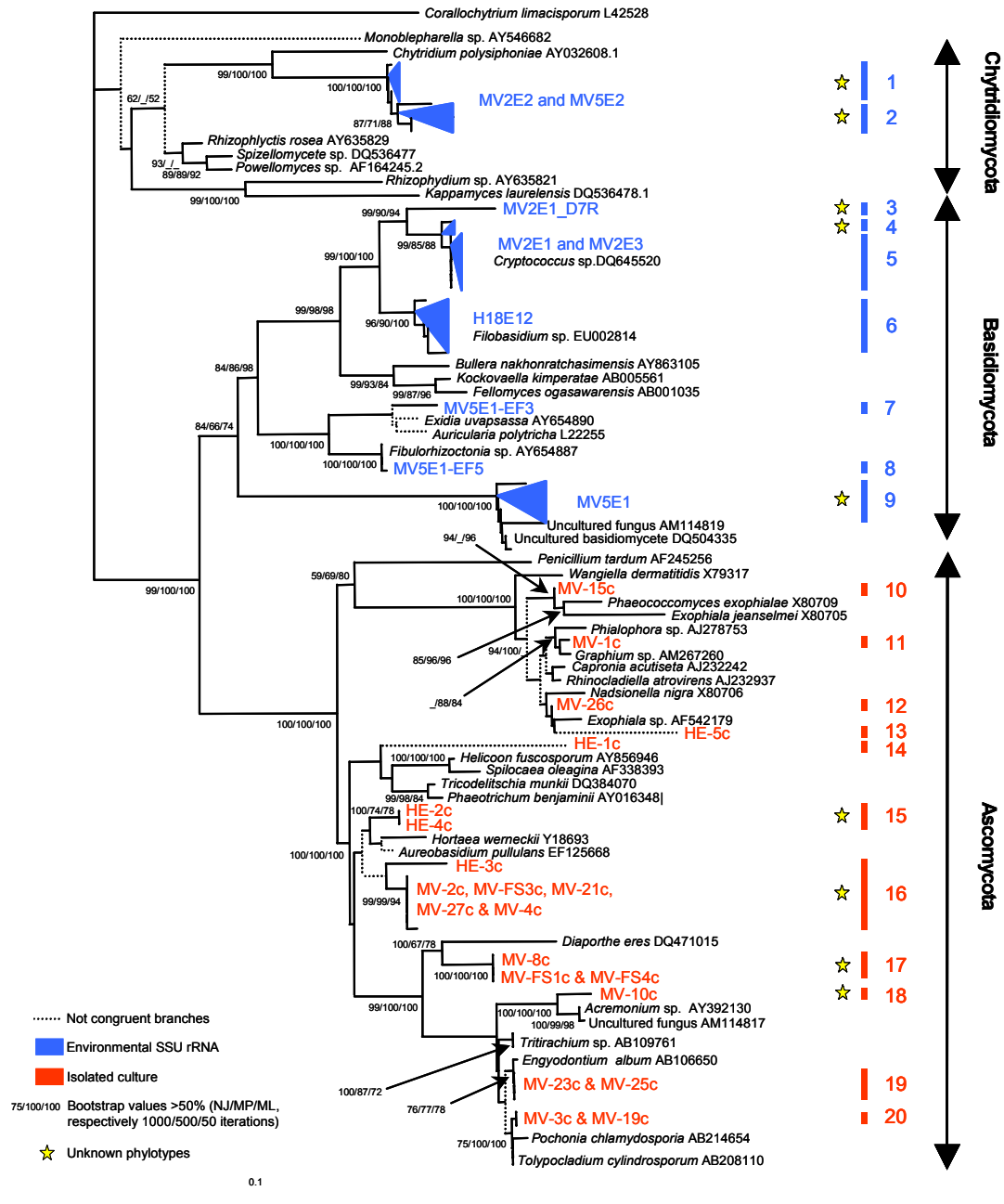
## **Acknowledgments**

We greatly acknowledge Eugene Koonin J. Peter W. Young and Myriam Bormans for valuable comments on the manuscript, for discussions and suggestions. We thank Laetitia Guillot ('Ouest Génomique' Bioinformatics) for her efficiency in constructing the fungal database. We are also grateful to the 'Ouest Génomique' sequencing service. This work was supported by GIS Génomique Marine and a grant from the 'Total Corporate Foundation for biodiversity and the sea'.

Supplementary information made available.

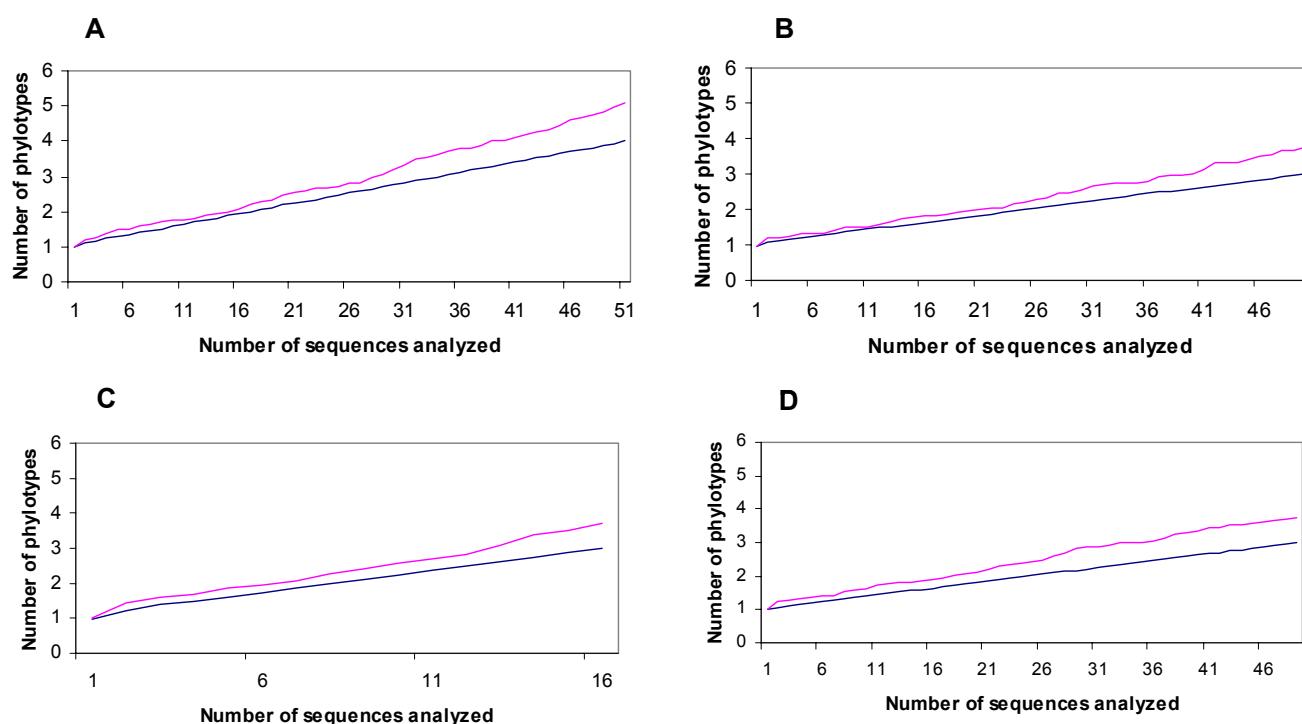
**Figure 31: Phylogenetic positions of deep sea hydrothermal fungi.** This consensus tree includes environmental SSU rRNA isolated from environmental ecosystems (blue), and isolated cultures (orange) from the same samples, with just the closest known related SSU rRNA fungal sequences. Tree was constructed using the NJ algorithm. Bootstrap values >50% are indicated at the nodes (respectively NJ/MP/ML estimated from 1000/500/50 iterations). Yellow stars mark the phylotypes considered to be new ones. The scale represents a distance of 0.1 changes per position computed from the NJ-K2P model. Dotted lines on the consensus tree represent branches not recovered in all three analyses. Names in orange refer to the identification number in the culture collection: MV isolates from the MARVEL samples (MAR) and HE from “HERO” samples (EPR) (Table 6). The following cultures correspond to sample MV2E1: MV1c-MV4c, MV21c, MV23c, MV-FS1c and MV-FS3c; to sample MV2E2: MV8c, MV10c, MV25c and MV-FS4c; to sample MV2E3: MV15c, MV19c, MV26c and MV27c; to sample H18E9: HE5c; to sample H18E11: HE1c-3c and finally to sample H18E12: HE4c. Within Ascomycota, HE-1c and HE-3c terminals were not defined as new phylotypes since their phylogenetic affinities remained unclear.

Fig. 31





**Figure 32: Estimates of fungal community diversity as a function of sampling effort for environmental samples.** Rarefaction curves were computed from 100 bootstrap replicates (pink) and the expected richness function (blue) is the number of phylotypes estimated (random sampling without replacement) from our sequences dataset generated from the ciPCRs. A : MV2E2; B : H18E12; C : MV2E1; D : MV2E3. Plots for MV5E1 and MV5E2 were not drawn since only one phylotype was found in each sample by sequence analyses of the ciPCR products using the primers MH2/MH4.



**Table 6:** Summary of samples analyzed in the study. Identification, nature of samples, location (MAR: Mid-Atlantic Ridge, EPR: East Pacific Rise), and depth are indicated. Numbers of phylotypes and corresponding numbers of sequences are also reported. Results from qPCR analyses are presented in the last row.

<b>Samples</b>	<b>Nature</b>	<b>Location</b>	<b>Depth (m)</b>	<b>Phylotypes</b>	<b>Number of SSU rRNA sequences</b>	<b>Copies of rRNA per <math>\mu\text{g}</math> of genomic DNA</b>
MV2E1	Exterior of shell of <i>Bathymodiolus azoricus</i>	Menez Gwen, MAR	860	8	59	$1.91 \cdot 10^5$
MV2E2	<i>B. azoricus</i> (body)	Menez Gwen, MAR	860	5	47	$1.483 \cdot 10^6$
MV2E3	Surrounding friable rocks	Menez Gwen, MAR	860	6	25	$3.01 \cdot 10^6$
MV5E1	Sulphide of surrounding rocks	Lucky Strike, MAR	1700	3	18	N/A
MV5E2	Exterior of shell of <i>B. azoricus</i>	Lucky Strike, MAR	1700	2	32	$1.35 \cdot 10^7$
H18E12	<i>A. pompejana</i> (epiderm)	Elsa, EPR	2630	2	49	N/A

### **Supplementary informations**

**Figure 33: Phylogenetic affinities within the Ascomycota phylum of the fungal SSU rRNA sequences from a deep sea hydrothermal ecosystem.** The NJ tree calculated includes SSU rRNA gene sequences of the fungal isolates (red). On the left, the tree shown was trimmed from a phylogeny using all the Ascomycota sequences available yet which satisfied the filters of quality and length (4270 sequences); on the right, parts of the tree are shown to highlight the environmental phylotypes affinities. Bootstrap values >50% are indicated at the nodes (for 100 iterations). The scale represents a distance of 0.1 changes per position computed from the NJ-K2P model. For nomenclature, please refer to figure 31 in the manuscript.

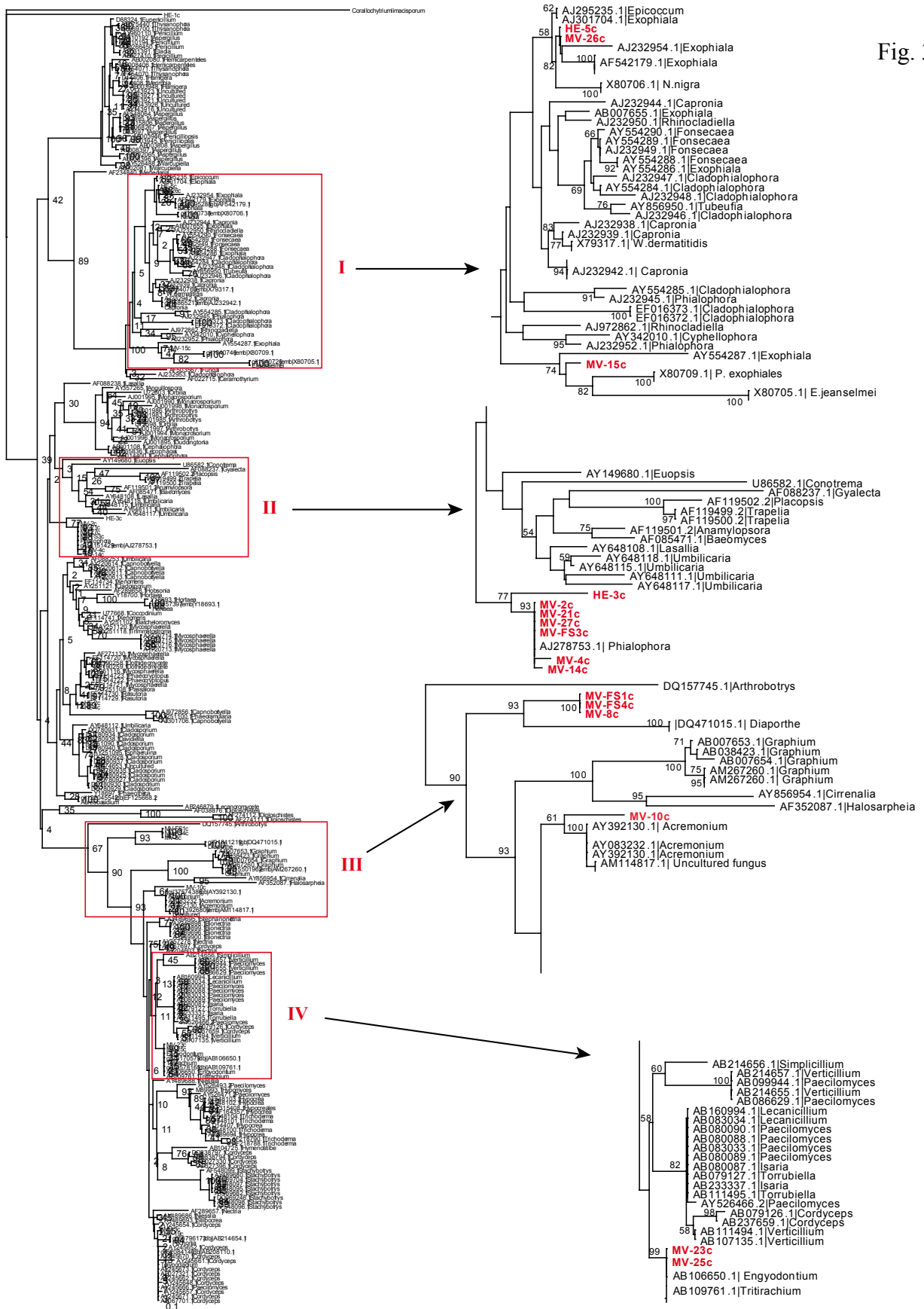


Fig. 33

**Figure 34: Phylogenetic affinities within the Basidiomycota phylum of the fungal SSU rRNA sequences from a deep sea hydrothermal ecosystem.** The NJ tree calculated includes SSU rRNA gene sequences revealed after ciPCR from the samples, H18E12, MV5E1, MV2E1 and MV2E3. On the left, the tree computed from all Basidiomycota sequences available yet which satisfied the filters of quality and length (2327 sequences); on the right, parts of this tree to highlight the environmental phylotypes affinities. Bootstrap values >50% are indicated at the nodes (for 100 iterations). The scale represents a distance of 0.1 changes per position computed from the NJ-K2P model. For nomenclature, please refer to figure 31 in the manuscript.



**Figure 35: Phylogenetic affinities within the Zygomycota phylum of the fungal SSU rRNA sequences from a deep sea hydrothermal ecosystem.** The NJ tree calculated includes SSU rRNA gene sequences revealed after ciPCR from the samples MV5E2 and MV2E2 (green). All the Zygomycota sequences available yet which satisfied the filters of quality and length were included (266 sequences). Bootstrap values >50% are indicated at the nodes (for 100 iterations). The scale represents a distance of 0.1 changes per position computed from the NJ-K2P model. For nomenclature, please refer to figure 31 in the manuscript.





**Figure 36: Phylogenetic affinities within the Chytridiomycota phylum of the fungal SSU rRNA sequences from a deep sea hydrothermal ecosystem.** The NJ tree calculated includes SSU rRNA gene sequences revealed after ciPCR from the samples MV5E2 and MV2E2 (green). All the Chytridiomycota sequences available yet which satisfied the filters of quality and length were used to compute this tree (163 sequences). The tree was constructed using the NJ algorithm. Bootstrap values >50% are indicated at the nodes (respectively NJ/MP estimated from 1000/250 iterations). The scale represents a distance of 0.1 changes per position computed from the NJ-K2P model. For nomenclature, please refer to figure 31 in the manuscript.





**Partie V : Analyse du métagénome extrait d'un écosystème profond. Hypothèses sur le mode de vie des champignons, et fonctionnement global de l'écosystème**

---

## **V- Analyse métagénomique**

Les premières parties de cette thèse ont présenté les différentes étapes ayant permis d'accéder à l'analyse de la diversité fongique associée aux écosystèmes hydrothermaux. Nous avons ainsi pu établir la présence de nombreux phylotypes fongiques, en combinant l'analyse de diversité directement à partir des échantillons environnementaux, et l'analyse des séquences nucléotides extraites à partir de souches fongiques isolées à partir des mêmes échantillons. Cette étude de diversité, basée sur des analyses phylogénétiques, était renforcée par la création d'une base de données experte, regroupant des marqueurs moléculaires, extraits à partir de la base de données Genbank. Cependant, aucune donnée sur le rôle écologique et le fonctionnement de ces organismes au sein de ces écosystèmes n'avait été obtenue par ces analyses. Pour résoudre ces questions, nous avons alors choisi une approche métagénomique, qui nous a permis d'accéder au pool entier de gènes de tous les organismes vivant dans ces environnements.

Cette cinquième partie, qui présente les résultats de cette analyse métagénomique, est articulée en 3 axes : le premier axe montrera les étapes ayant mené au choix de l'échantillon hydrothermal utilisé pour cette étude. Le second axe, présentera les résultats obtenus sur l'analyse des contigs assemblés ayant permis de reconstruire un fonctionnement hypothétique du métabolisme fongique dans ces environnements. Le dernier axe s'intéressera quant à lui à l'étude des séquences non assemblées, ayant permis d'établir des hypothèses sur le fonctionnement général du compartiment microbien au sein de l'écosystème. Ces deux derniers axes seront présentés sous la forme d'une ébauche de publication, sur la base de résultats de pyroséquençage (voir chapitre Matériel et méthodes).

### **V-1 Choix de l'échantillon**

La première étape de l'analyse métagénomique a consisté à choisir l'échantillon hydrothermal présentant les caractéristiques optimales pour répondre à nos hypothèses de travail.

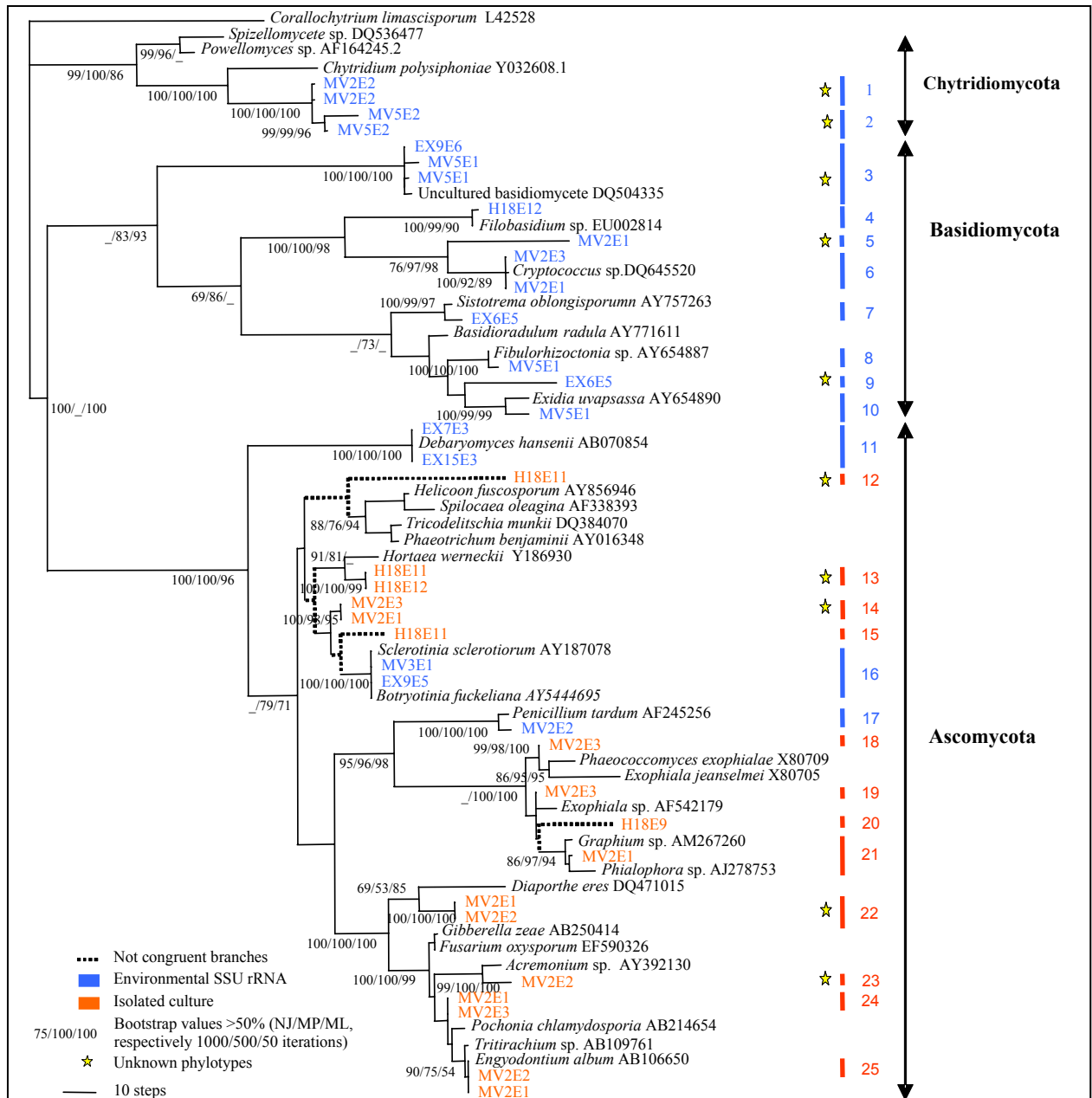
## 1-1 Analyse de diversité complémentaire

En complément des analyses phylogénétiques présentées dans la 4<sup>ème</sup> partie de cette thèse, nous avons choisi d'étendre l'analyse de diversité sur des échantillons supplémentaires, obtenus lors de la campagne EXOMAR (voir le chapitre Matériel et Méthodes).

Cette analyse a permis de mettre en évidence 25 phylotypes appartenant aux phyla Ascomycota, Basidiomycota et Chytridiomycota (figure 37); 10 de ces phylotypes représentent des phylotypes inconnus (figure 37). En accord avec la phylogénie présentée dans la 4<sup>ème</sup> partie de cette thèse, aucune séquence d'organismes appartenant aux phyla Glomeromycota et Zygomycota n'ont été retrouvés. De façon intéressante, certaines séquences extraites d'échantillons de la campagne EXOMAR, ont également été retrouvées lors des campagnes précédentes (échantillons EX9E6 : phylotype 3, et EX9E5 : phylotype 1). Cette analyse de diversité complémentaire a également permis de retrouver des séquences appartenant au phylum Ascomycota, directement amplifiées à partir d'échantillons environnementaux (phylotypes 11, 16 et 17, figure 37). L'une de ces phylotypes (phylotype 17) semble appartenir au genre *Penicillium*; bien que ces organismes soient ubiquistes, une contamination extérieure ne peut pas être totalement exclue pour ce phylotype. Des levures du genre *Debaryomyces*, ont également été retrouvées (phylotype 11). Ces levures ont déjà été rencontrées en milieu marin (Kutty & Philip, 2008).

## 1-2 Quantification des molécules d'ARNr 18S dans les échantillons hydrothermaux

Afin de quantifier le nombre de copies de gènes codant l'ARNr 18S dans nos échantillons, nous avons choisi une approche par PCR quantitative, dont le protocole est présenté dans la partie Matériel et méthodes de cette thèse. Cette analyse, nous a permis d'obtenir une quantification absolue pour différents échantillons hydrothermaux candidats pour l'analyse métagénomique. Ainsi, le nombre de copies de gènes codant l'ARNr18S variait de 1.91 copies par  $\mu\text{g}$  d'ADN génomique à  $1.35 \cdot 10^7$  copies par  $\mu\text{g}$  d'ADN génomique.



**Figure 37** : Analyses phylogénétiques des séquences de champignons d'échantillons isolés d'écosystèmes hydrothermaux. L'arbre consensus présenté comprend les séquences des gènes codant les ARNr18S, directement amplifiés à partir de l'environnement (bleu), ou à partir des isolats fongiques isolés des mêmes échantillons (rouge). L'arbre a été construit en utilisant un algorithme MP, et les valeurs de bootstrap > 50% sont indiquées aux nœuds (respectivement pour NJ/MP/ML : 1000/500/50 itérations). L'échelle représente 10 pas évolutifs. Les phylotypes inconnus sont symbolisés par une étoile. Les échantillons étudiés ont été isolés lors des campagnes MARVEL (MV), HERO (H18E) et EXOMAR (EX).

### 1-3 Détermination de l'échantillon

Les résultats couplés des analyses phylogénétiques et des quantifications des molécules d'ARNr 18S, ont été utilisés pour déterminer l'échantillon semblant le plus approprié pour l'analyse métagénomique : ce dernier devait présenter une faible diversité couplée à une forte fréquence de gènes codant l'ARNr18S. Ainsi, nous avons choisi l'échantillon présentant une diversité faible (un phylotype unique), et originale (phylotype inconnu formant une lignée évolutive ancienne), combiné à un grand nombre de molécules d'ARNr 18S, permettant de s'assurer d'une forte fréquence de champignons dans l'échantillon. L'échantillon MV5E2 a ainsi été choisi pour être pyroséquencé ; les caractéristiques de cet échantillon sont présentées dans le tableau 7. L'analyse du métagénome contenu dans cet échantillon ayant pour objectif de nous permettre de résoudre les voies métaboliques portées par les gènes de cet organisme inconnu. De plus, le caractère ancestral de cette lignée de Chytride nous donne accès au matériel génétique d'un champignon ayant potentiellement conservé les caractères d'une lignée primitive des champignons.

Echantillon	Site d'échantillonnage	Nombre de copies d'ARNr 18S / $\mu$ g d'ADN génomique	ADN extrait utilisé pour le pyroséquencage
MV5E2: Grattage de la surface externe de la coquille de <i>Bathymodiolus azoricus</i>	MAR: Lucky Strike	$1.35.10^7$	6 ng

Tableau 7 : Caractéristiques de l'échantillon pyroséquencé.

*Les résultats des travaux préliminaires obtenus sur la base de l'analyse de l'ensemble des contigs produits par pyroséquencage, et de près de 450 000 séquences non assemblées, sont présentés sous forme d'une publication. Les résultats obtenus à partir des contigs, qui nous ont permis d'établir des hypothèses sur les rôles des champignons dans leur environnement seront discutés dans un premier temps; la seconde partie de la publication portera sur l'étude des séquences non assemblées nous ayant permis d'établir des hypothèses sur le fonctionnement écosystémique global. Toutes les informations sur les techniques utilisées sont présentées dans la partie Matériel et méthodes de cette thèse.*



## **V-2 Présentation des résultats de l'analyse métagénomique**

### **Hypothèses sur le mode de vie des champignons, Reconstructions métaboliques du compartiment bactérien**

#### **Introduction**

Life in the deep sea relies solely on organic matter sinking from the upper layers. As no light penetrates these ecosystems, no photosynthesis can occur and productivity is therefore very low. Abyssal ecosystems are thus considered as deserts as their biomass does not exceed  $1\text{g.m}^{-2}$ . In contrast, deep sea hydrothermal ecosystems are considered as hot spots of productivity on Earth, with biomasses which can exceed  $50\text{kg.m}^{-2}$  (Desbruyeres *et al.*, 2000). Hydrothermalism is dependent on the volcanic activities occurring in different parts of the oceanic ridges. The hydrothermal fluids emitted by black smokers result from the heating of seawater, in contact with the magma mantle, and are enriched with reduced compounds, such as hydrogen sulfide ( $\text{H}_2\text{S}$ ) and methane ( $\text{CH}_4$ ) or carbon dioxide ( $\text{CO}_2$ ) (Jannash 1989; Karl, 1995). The primary production on which the biological communities in these ecosystems depend is finely tuned by chemoautotrophic bacteria, which constitute the basis of the food web (Jannash, 1989; Fisher, 1990). These bacteria (free-living or symbionts) harvest their energy by oxidizing the reduced compounds emitted by the fluids (Van Dover & Fry, 1994; Childress, 1991). The energy released can then be used to convert mineral carbon to organic carbon which may subsequently be used by heterotrophic organisms. Inhabitants of these environments live at the mixing of hydrothermal fluid with seawater interface (Desbruyeres *et al.*, 2000).

Due to the strong gradients in physico-chemical conditions (dissolved oxygen, temperature, pH...), species inhabiting deep sea hydrothermal ecosystems are mostly endemic (90%) and many new taxa have been identified since the discovery of these ecosystems in 1977 (Ramirez-Llodra, 2007; Wolff, 2005). The finding of Eukaryotic invertebrate organisms such

as tubeworms (*Alvinella pompejana*), mussels (*Bathymodiolus sp.*), and shrimps (*Rimicaris sp.*) has been widely documented. In addition to the eukaryotes and bacteria, archaea (sulfur reducers and methanogens) have been found. These archaea have also been observed in the hotter parts of this ecosystem (Raguenees *et al.*, 1995).

For more than a decade, microbiological surveys in these ecosystems have focused on bacteria and archaea, and fungal diversity has been analysed only very recently (Bass *et al.*, 2007; Burgaud *et al.*, 2009; Le Calvez *et al.*, submitted). As fungi are fundamental to the functioning of terrestrial ecosystems through their recycling of organic compounds (C, N, S, P...), and numerous symbiotic or pathogenic associations, it is hypothesized that, similarly, they may exert important ecological functions in marine hydrothermal ecosystems. However nothing is known to date about how these fungi live and what ecological functions they perform. It is now possible to predict metabolic pathways from genes and thereby deduce lifestyles by a metagenomic approach, the metagenome being defined as the sum of the genomes of all organisms living in a given environmental sample (Handelsman *et al.*, 1998).

New sequencing technologies, especially pyrosequencing (Margulies *et al.*, 2005), provide access to thousands of nucleotidic sequences at reasonable cost, compared to traditional Sanger sequencing technologies. Pyrosequencing does not require the construction of BAC libraries and has been shown to be very effective for metagenomic analyses (e.g. Gharizadeh *et al.*, 2006, Gill *et al.*, 2006; Liu *et al.*, 2007; Frias-Lopez *et al.*, 2008; Schlüter *et al.*, 2008).

We report here, an original metagenomic analysis based on deep DNA sequencing technologies, and designed to better understand both the functioning of unknown fungal organisms and overall functioning of the ecosystem. This approach gave us direct access to a battery of fungal gene fragments in our sample, mixed with genes from other organisms sharing the habitat. One particular sample was chosen on the basis of its low fungal diversity (a single phylotype), the originality of this diversity (an old fungal lineage within the Chytridiomycota) and its estimated frequency within the sample. In this paper we report the results of the metagenomic analysis. In the first part we focus on the assembled sequences (hereafter named contigs) to predict the fungal functions and metabolisms. In the second part, we investigate the whole metagenome to access the taxonomic composition of our sample and reconstruct the global functioning of the microbial assemblage in this environment by direct sequence analysis.

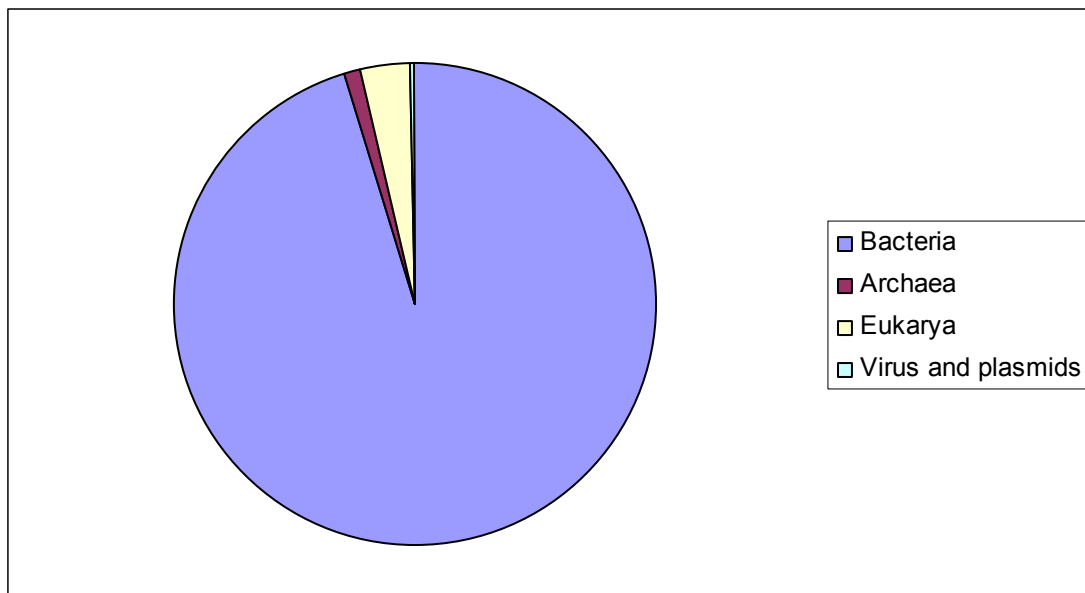
## Results and discussion

### Pyrosequencing results

6 ng of DNA were extracted from the selected sample (MV5E2), and used as template to perform 3 pyrosequencing runs with the GS FLX sequencer (Roche Applied Science), without any whole genome amplification step. The *de novo* assembly of sequences was performed with Newbler Assembler Software (Roche Applied Science). In total, 168.909 contigs (average length: 352 bp; coverage: 2.88x) were assembled from 1.441.839 single reads (average length: 227.5 bp) and 328.493.329 nucleotides were sequenced for this study. Each contig was assembled by means of 3.66 sequences.

### Sequencing effort

The first step was to check that the sequencing effort was sufficient to properly analyse the metabolic properties of our sample. As indicated in the Material and Methods section, we chose to extract taxonomy from the BLASTX results (BLASTX against Refseq, using an e-value cut-off of  $10^{-3}$ ), using contigs as a query. An overview of Superphylum distribution, based on the results of the first BLAST hit, is provided in figure 38.



**Figure 38:** Taxonomic distribution by domain of life using first BLAST hit. Taxonomy was extracted after BLASTX analyses against metagenomic contigs (see Material and Methods section for details).

We highlighted a large majority of bacterial contigs in the dataset. We chose those microorganisms which were dominant (see Material and Methods), and for which genomic information was available. We then did coverage tests using these microorganisms as reference (table 8).

Organisms tested	proteins in database	Top hits	% proteins in database with 1 hit ( 2 runs = 110996 contigs)	% proteins in database with 1 hit (3 runs = 168.909 contigs)
<i>Gramella forsetii</i> KT0803	3585	1996	55.67	59.35
<i>Cytophaga hutchinsonii</i> ATCC 33406	3785	1444	38.15	45.23
<i>Parvibaculum lavamentivorans</i> DS-1	3636	1686	44.39	46.3
<i>Silicibacter</i> sp. TM1040	4254	2083	48.96	54.16
<i>Colwellia psychrerythraea</i> 34H	4912	2367	44.76	48.2
<i>Methylococcus capsulatus</i> str. Bath	2956	1839	56.52	62.2
<i>Nitrosococcus oceani</i> ATCC 19707	3023	1855	56.96	61.4

**Table 8:** Sequencing effort by comparisons of coverage tests, from 2 runs or 3 runs of pyrosequencing (details provided in Material and Methods)

The results of this test indicated that a supplementary pyrosequencing run, which resulted in a gain of 35 % of the assembled contigs, also increased the genome coverage for the dominant microorganisms in our sample. However, the third sequencing run only provided a gain in coverage of the mean data of 4.35 % for 7 microorganisms tested (table 8). It was therefore concluded that the sequencing effort was sufficient to study the metabolic properties of our metagenomic dataset. We refined our understanding of the taxonomic composition by doing a deep analysis of the unassembled reads; the results of this analysis are provided in the second part of this manuscript.

## **Contigs analyses and prediction of fungal functions in hydrothermal ecosystems**

Fungi occur in diverse marine habitats, where they contribute to the cycling of organic matter, mostly through their lignocellulolytic and pathogenetic abilities (Hyde *et al.*, 1998). The fungal functions in deep marine hydrothermal ecosystems are unknown but can be predicted from metagenomic analysis of the extracted DNA. The sample chosen was obtained from the external surface of a mussel shell isolated from the Mid-Atlantic Ridge (Lucky Strike site, 1600 m depth). This mussel, *Bathymodiolus azoricus*, is a deep-sea hydrothermal organism known to possess bacterial endosymbionts with proven methanotrophic and thioautotrophic metabolisms (Fujiwara *et al.*, 2000).

We chose to determine the ecological functions of the fungi in these habitats by focussing on the contigs. This choice was governed by the fact that these DNA sequences are longer than the unassembled reads and should be more informative and more reliable for the BLAST analyses. However, the dark side of this strategy is that the unassembled fungal DNA sequences are wasted.

Extraction of the fungal contigs was based on the first BLASTX hit results using the curated RefSeq database (see Material and Methods for details). 335 fungal sequences were extracted from these 168.909 contigs, of which 84 could directly be annotated on the basis of the BLASTX results. An ORF finder (NCBI) was used to obtain complementary annotations from contigs tagged as hypothetical (251 contigs). Finally, 178 non-redundant annotated contigs were obtained with high confidence. All contigs were then compared with the KOG database. The KOG distribution is shown in figure 39.

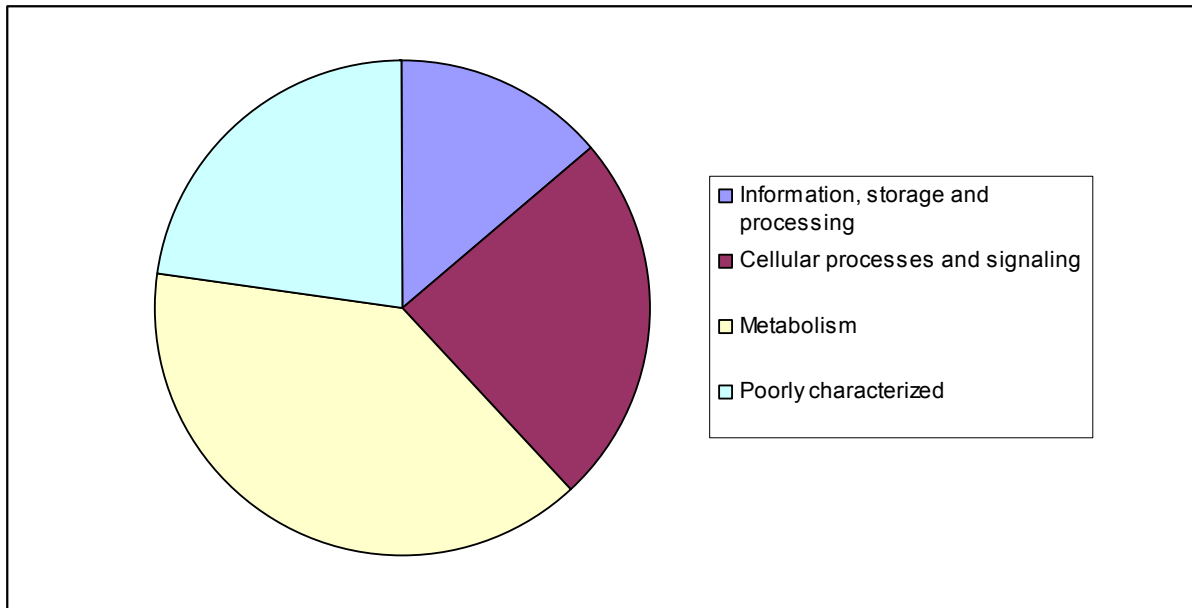


Figure 39: KOG distribution of 178 fungal contigs.

It was predicted from this classification that 79 contigs harboured metabolic functions. This expert annotation was then used to predict fungal metabolism based on KEGG mapping. We here present the presumed metabolic abilities of the unknown fungus, for aminoacids, carbohydrates, and energy metabolisms, and for specific abilities such as antibiotics production. The results obtained with the other contigs classified in the 3 other KOG categories (i.e. “information, storage and processing”, “cellular processes and signalling”, and “poorly characterized”) are not shown as they were mostly ‘house keeping genes’: for example, contigs in “Cellular Processes and signalling” were mainly annotated as protein kinases involved in signal transduction mechanisms. These contigs are widely shared by living organisms and could not be used to produce hypotheses about how fungi live in this ecosystem or what ecological functions they exert. Contigs from the KOG category “Information, storage and processing” were not really informative. Some, for example, were annotated as “DNA polymerase” or “DNA topoisomerase”, and we could not take unfair advantage of this information to resolve our questions. The “poorly characterized” fungal contigs were not taken into account.

The hypothetical fungal cell derived from all this information is presented in figure 40, and is based solely on the contigs annotated as “metabolic”.

### Amino acid metabolism :

We found protein-coding genes involved in the synthesis and degradation of 15 different amino acids from 23 different fungal enzymes we extracted from our dataset. Enzymes were found for cysteine, lysine, histidine, glutamate, tyrosine, glycine, serine, threonine, leucine, valine, isoleucine, alanine, aspartate, and phenylalanine metabolisms. Enzymes involved in the synthesis and degradation of selenocysteine were also found. As fungi are able to synthesize almost all aminoacids a pathogenic lifestyle could be rejected. Enzymes associated with nitrogen metabolism were also predicted. These enzymes are involved in the transformation of nitroalkanes to nitrites (2-nitropropane dioxygenase ec 1.13.11.32), and ammonia to carbamoyl-P (Carbamoyl-phosphate synthase 6.3.4.16). Nitroalkanes could then be integrated into nitrogen metabolism and used in preference to nitrates available in the environment. After the transformation of nitroalkanes to nitrites (ec 1.13.11.32), and possibly to ammonia, we were able to highlight the use of nitrogen compounds to synthesize aminoacids (6.3.4.16).

### Carbohydrates and Energy metabolism :

Protein-encoding genes involved in the pentose phosphate pathways (Glucose-6-P isomerase ec 5.3.1.9 ; 6-phosphogluconate dehydrogenase ec 1.1.1.44 ; Phosphoglucomutase ec 5.4.2.2 ; Phosphoketolase ec 4.1.2.9 ; 2-Deoxy-D-ribose-5P ec 4.1.2.4), the citrate cycle (Phosphoenolpyruvate carboxykinase ec 4.1.1.49 ; Aconitate hydratase ec 4.2.1.3 ; Succinate dehydrogenase ec 1.3.5.1 ; 2-oxoglutarate ferredoxin oxidoreductase ec 1.2.7.3 ; Glutamate decarboxylase ec 4.1.1.15) and glycolysis, were predicted from the annotation. For glycolysis, we found enzymes for the aerobic carboxylation of glucose (Glucose-6-P isomerase ec 5.3.1.9; Glyceraldehyde-3-P dehydrogenase ec 1.2.1.12), as well as enzymes involved in the anaerobic decarboxylation of pyruvate (Alanine dehydrogenase ec 1.4.1.1; Acetyl coA synthetase ec 6.2.1.1), a process characteristic of yeast-like metabolism. We hypothesized that the phylogenetic species found (an old chytrid lineage) might be able to ensure both yeast-like and filamentous-like metabolism, a feature which could have been lost during subsequent Chytridiomycota evolution. Different energy sources would be used. We found enzymes involved in oxidative phosphorylation (Succinate dehydrogenase ec 1.3.5.1; NADH dehydrogenase ec 1.6.5.3) that would constitute an additional sign of aerobic metabolism. Signatures of genes involved in glycogen catabolism (Glycogen Phosphorylase ec 2.4.1.1),

and the use of formate, pyruvate, butyn-1-ol, propyn-1-ol were also predicted. An ability to degrade other complex carbohydrates and hydrocarbons was hypothesized. Surprisingly, no heterotrophic gene signatures were found, such as genes encoding for chitinases or glucosidases (typically found in fungal organisms). This might stem from the limited number of fungal contigs detected in our dataset: the main carbon source of these organisms remains to be elucidated. However, an unusual use of carbon can be suggested, as explained below.

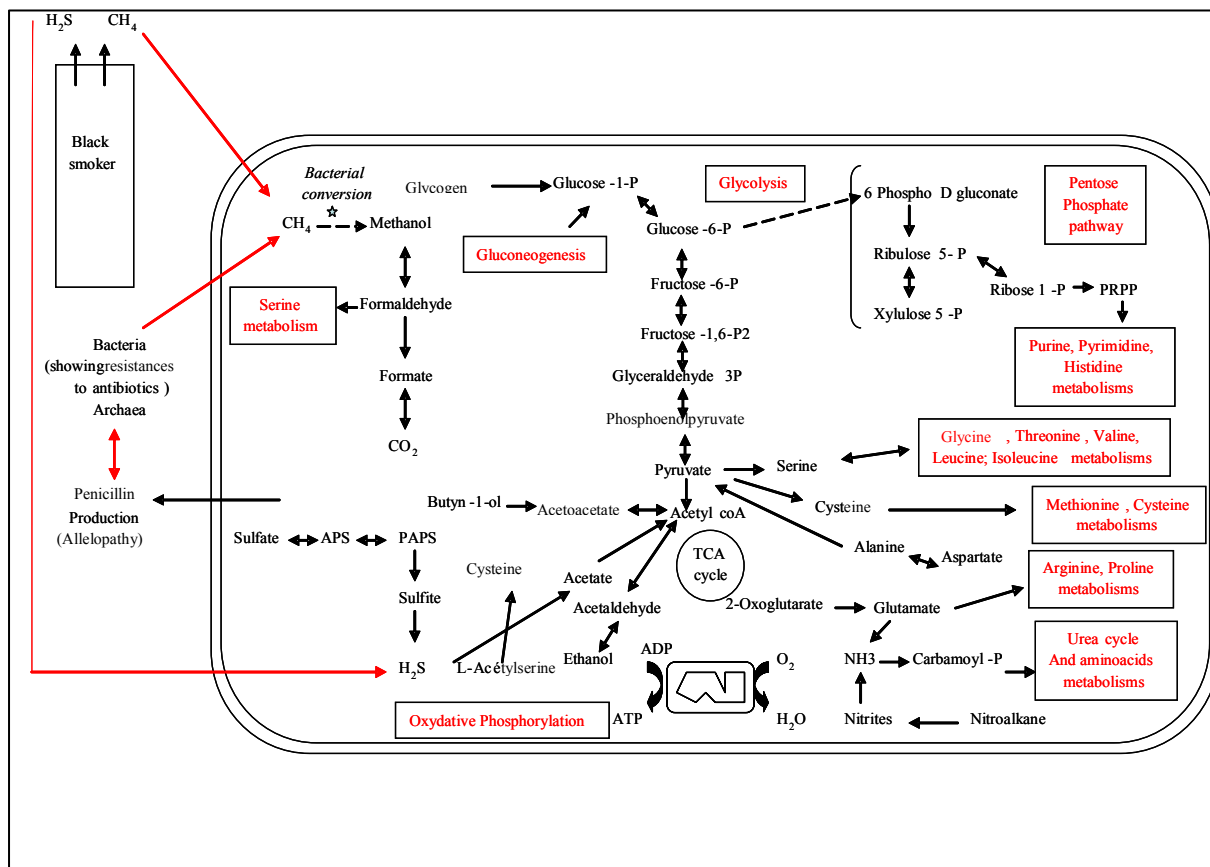
#### Interactions between fungi and their environment.

Enzymes involved in methane metabolism were found, particularly alcohol dehydrogenase (ec 1.1.99.8) which drives the transformation of methanol into formaldehyde. Although no gene coding the enzyme involved in methane transformation to methanol (Methane monooxygenase ec 1.14.13.25) was found, a hypothetical methanotrophic metabolism could not be rejected. Our results suggest an alternative hypothesis: bacteria could convert the CH<sub>4</sub> naturally available in the environment to methanol which in turn could be used by fungi (see figure 40).

We also found fungal contigs involved in sulfur metabolism (3 of the enzymes in this cycle detected), especially the anabolic reduction of sulfate pathway (Sulfate adenylyltransferase ec 2.7.7.4), and the use of H<sub>2</sub>S, naturally abundant in the environment (Cysteine synthase ec 2.5.1.47). In addition, a hypothetical pathogenic lifestyle could again be rejected as these organisms are not dependent on an organic sulfur contribution. The fungal DNA annotation suggests a biotic interaction signature i.e. the presence of genes involved in penicillin biosynthesis (Alpha amino adipate reductase ec 6.3.2.26; Isopenicillin N synthase ec 1.21.3.1). This fungal antibiotic biosynthesis was coupled with the finding of bacterial contigs involved in penicillin degradation (figure 40), and bacterial penicillin binding proteins. We suggest that the fungus produces and emits antibiotics into the environment (allelopathy mechanism) to improve its ability to compete in colonising the ecological niche or habitat.

A hypothetical scheme, based on all these Best BLASTX hits analyses of metagenomic contigs, is proposed. This scheme represents the principal mechanisms that could be highlighted from the fungal contigs (figure 40).





**Figure 40:** Hypothetical fungal cell metabolism, from fungal contigs extracted from the metagenome. Red arrows represent interactions between the fungi and the environment.

### Limits of the analysis

The results presented here are important in terms of knowledge based on predictions. They also provided a range of testable hypotheses. However, they were limited by the low number of fungal contigs extracted without ambiguity from the complete dataset. There are is fact has different explanations for this. First, the annotation is dependent on the sequences already annotated. The fungal phylotype present in the analysed sample is far from being a model organism. It is a highly original fungus belonging to the Chytridiomycota phylum which has not been studied extensively. The fungal sequences are therefore under-represented. Table 9 shows the numbers of proteins and nucleotide sequences in Genbank (February 2009), in the whole database, and in RefSeq: only 412 chytrid proteins are stored in Genbank, and would result in the low coverage of fungal contigs in our dataset. At the same time, this study provided more than 300 non redundant contigs. Moreover, no whole genome sequence for Chytrids with complete annotation is available as yet, although ongoing projects exist. Although the analysis of contigs makes sense, the fungal genome, like other eukarya, contains introns and repeated elements which makes it difficult to produce contigs for these organisms. Presumably a portion of the fungal data was not taken into account by the contig sequence analyses. It would be interesting to test another hydrothermal sample, exhibiting a high frequency of gene coding SSU rRNA and a phylotype belonging to well studied fungal phyla, such as organisms from the Ascomycota and Basidiomycota, for example. In this way, we could compare the efficiencies of BLASTX analyses to discriminate contigs annotated as fungal.

	Proteins		Nucleotides	
	All	RefSeq	All	RefSeq
Chytridiomycota	412	141	1796	8
Ascomycota	770127	303266	1108144	303689
Basidiomycota	149834	55946	194609	58062
Zygomycota	1583	80	4510	3
Glomeromycota	384	20	11543	1

Table 9: Number of protein and nucleotide sequences in Genbank in February 2009 (whole database and RefSeq databases), for each fungal phylum.

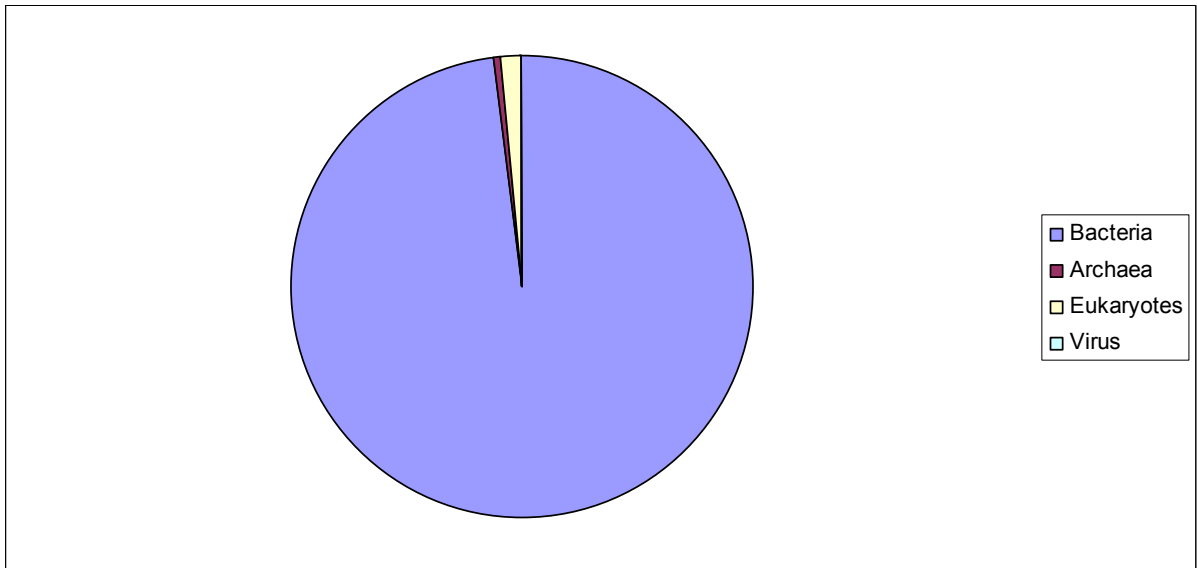
The analyses presented herein enabled us to predict the different components of the metabolic abilities of an old fungal lineage inhabiting a deep-sea hydrothermal ecosystem. They provide information about a previously unknown organism which is therefore under-represented in the public database. This study underlines the necessity to improve efforts in whole genome sequencing, both to better understand the metabolic abilities of such organisms and also to highlight their roles in the environment.

## **Direct analyses of sequences and prediction of global functioning of the metagenome**

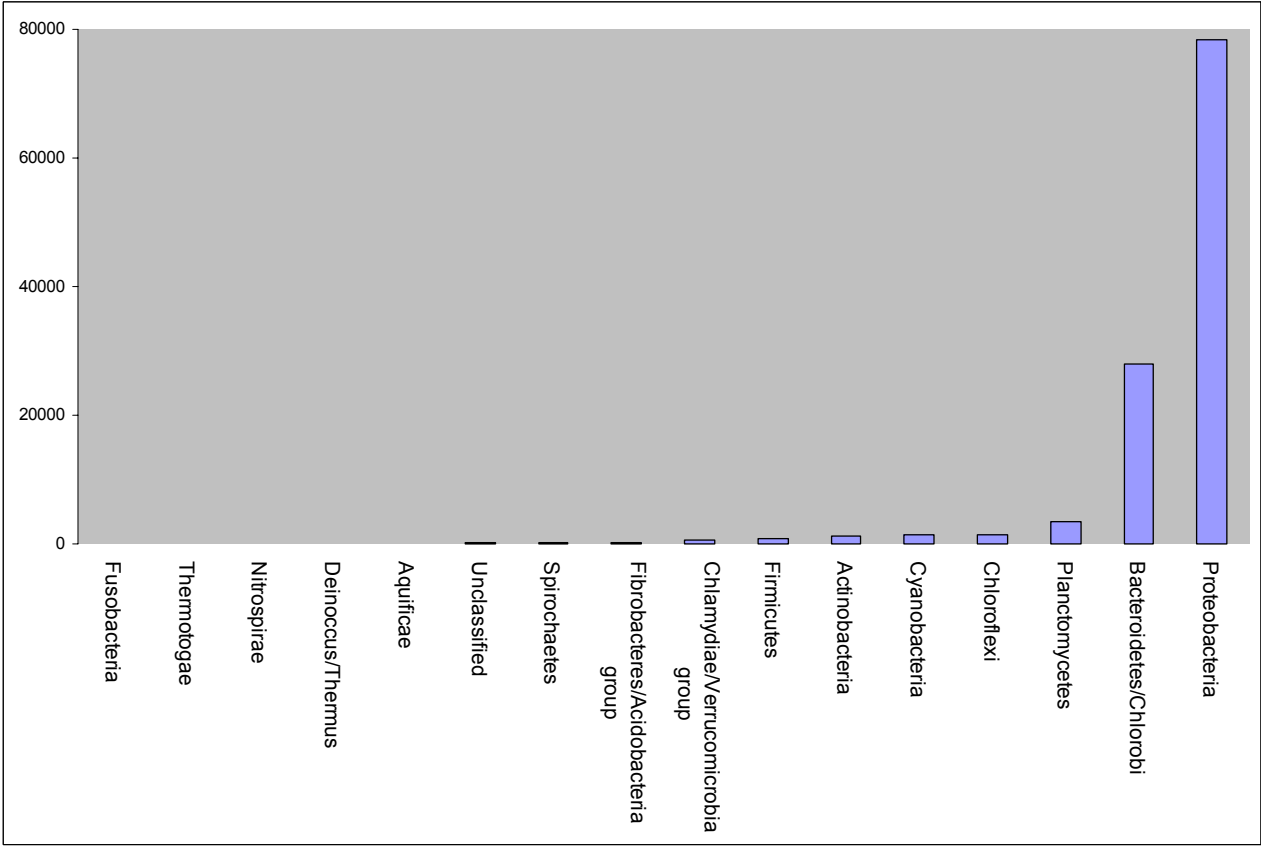
### **Taxonomic distribution of the metagenomic dataset**

To access the taxonomic diversity of our sample, we used the MEGAN software, based on the Last Common Ancestor Algorithm. We used 427.616 reads. These were first analysed with BLASTX (against the nr database), using an e-value threshold of  $10^{-3}$ . The BLASTX results files were then integrated into MEGAN: In this way, a taxonomic assignation could be obtained from 208.113 reads. Analysis of the taxonomic distribution at the Superphylum level of our dataset revealed a large majority of bacterial reads, with more than 98% of reads assigned as bacterial (see figure 41 for the taxonomic distribution at Superphylum level, in agreement with figure 38). Proteobacteria and Bacteroidetes contribute to 68% and 24% of the bacterial contigs respectively (figure 42). Surprisingly, the majority of reads from the Proteobacteria phylum belonged to the Gamma Proteobacteria, and only 5% were assigned to the Epsilon Proteobacteria class which seems to be dominant in hydrothermal environments (figure 32) (Campbell *et al.*, 2003; Corre *et al.*, 2001). The under-representation of these organisms in our dataset could be due to the sensitivity of MEGAN analyses, which is dependent on protein coding genes stored in the Genbank nr database, but could also be due to the sample we chose, which had been isolated from a mesophilic area of the Lucky Strike vent. Finally, the Superphyla distribution of the MEGAN results is similar to the Best BLAST hit strategy indicated in figure 38.

The MEGAN results provided an overview of community distribution in our sample, and revealed a very complex community. Because of this complexity, most reads could not be assembled into contigs (Tringe *et al.*, 2005).



**Figure 41:** Taxonomic distribution by domain of life using MEGAN software. Taxonomy was extracted from the Best BLASTX results using metagenomic contigs as a query (see Material and Methods section for details).



**Figure 42:** Taxonomic distribution of bacterial reads, according to MEGAN software.

## **Analysis of bacterial metabolisms present in the metagenomic sample**

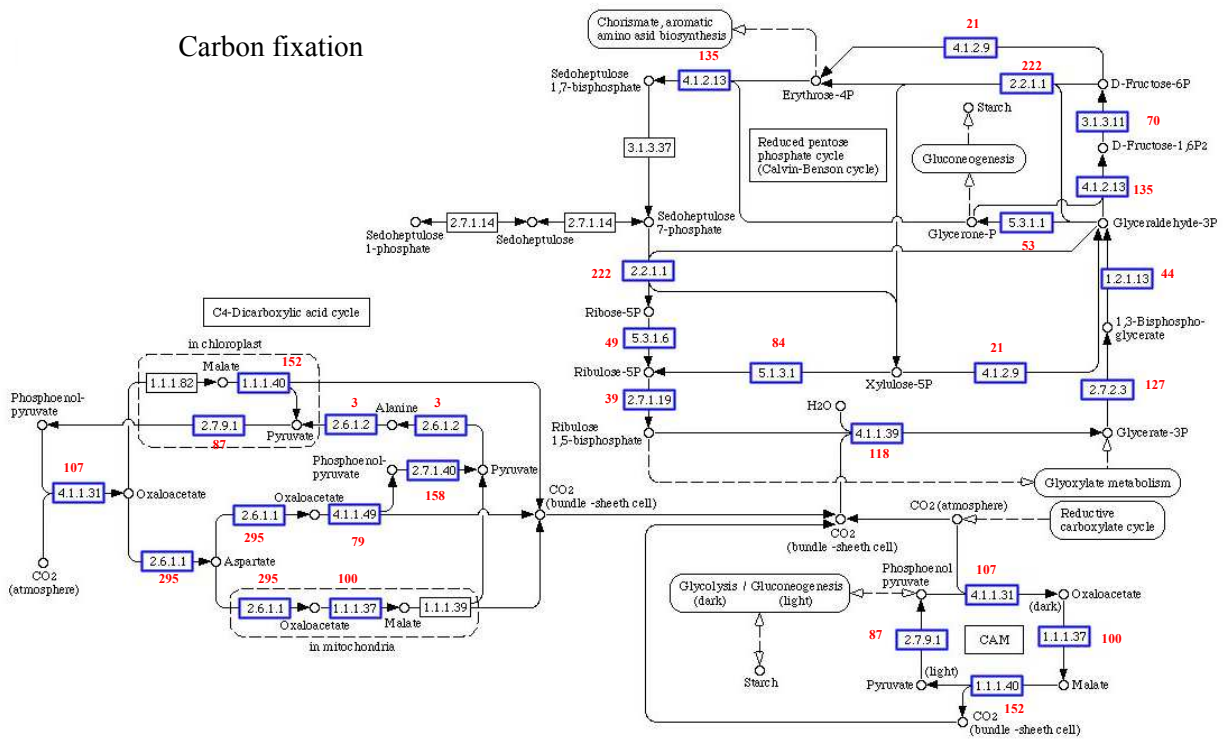
To complete our metagenomic study and obtain insights into the global functioning of the ecosystem, we submitted our reads dataset to the automated RAST server. Each read was then compared with the RAST subsystems and integrated into the KEGG metabolic pathways, using ASGARD software as a complement (see Material & Methods for more details). No preliminary data binning was applied but as the vast majority of reads with MEGAN were assigned to Bacteria, we reasoned that the RAST analysis would principally focus on bacterial reads. We present hypotheses about the functioning of the major metabolic pathways within the ecosystem. Our approach is based on gene function prediction, although the informative MEGAN results (e.g. dominant organisms), were also taken into account.

### **Carbon Assimilation**

Carbon assimilation in deep hydrothermal ecosystems relies on chemolithoautotrophy (Pimenov *et al.*, 2000). As expected in this environment, no sequence associated with photosynthetic mechanisms was found, which also confirmed that no contamination occurred during sampling.

Many reads were assigned to Ribulose Bi phosphate Carboxylase (RuBisCo) (ec 4.1.1.39): which can fix CO<sub>2</sub>, with H<sub>2</sub>O, to form 2 molecules of Phosphoglycerate from the substrate (figure 43). Most autotrophic microorganisms, including hydrothermal-living organisms, use the Calvin Benson Cycle to assimilate CO<sub>2</sub> (Childress & Fisher, 1992). We found all the known enzymes in this cycle (figures 43, 45). Another CO<sub>2</sub> fixation pathway could also be predicted from our dataset (figure 44). Acetyl CoA is formed from CO<sub>2</sub> (or CO) and is used as a source of carbon and energy for growth. This autotrophic fixation pathway is used by methanogenic or sulphate-reducing bacteria (Pezacka & Wood, 1988). We can assume from these data that autotrophic mechanisms are operating in this habitat. The different types of autotrophic metabolism apparent from the metagenome analysis confirm the complexity of community composition indicated by the MEGAN analyses.

## Carbon fixation



**Figure 43:** Kegg map for carbon fixation pathways analysed using ASgard. Blue rectangles represent enzymes predicted from the analysis of 427.616 sequences from the metagenome dataset. Number of occurrences of enzymes found is written in red.

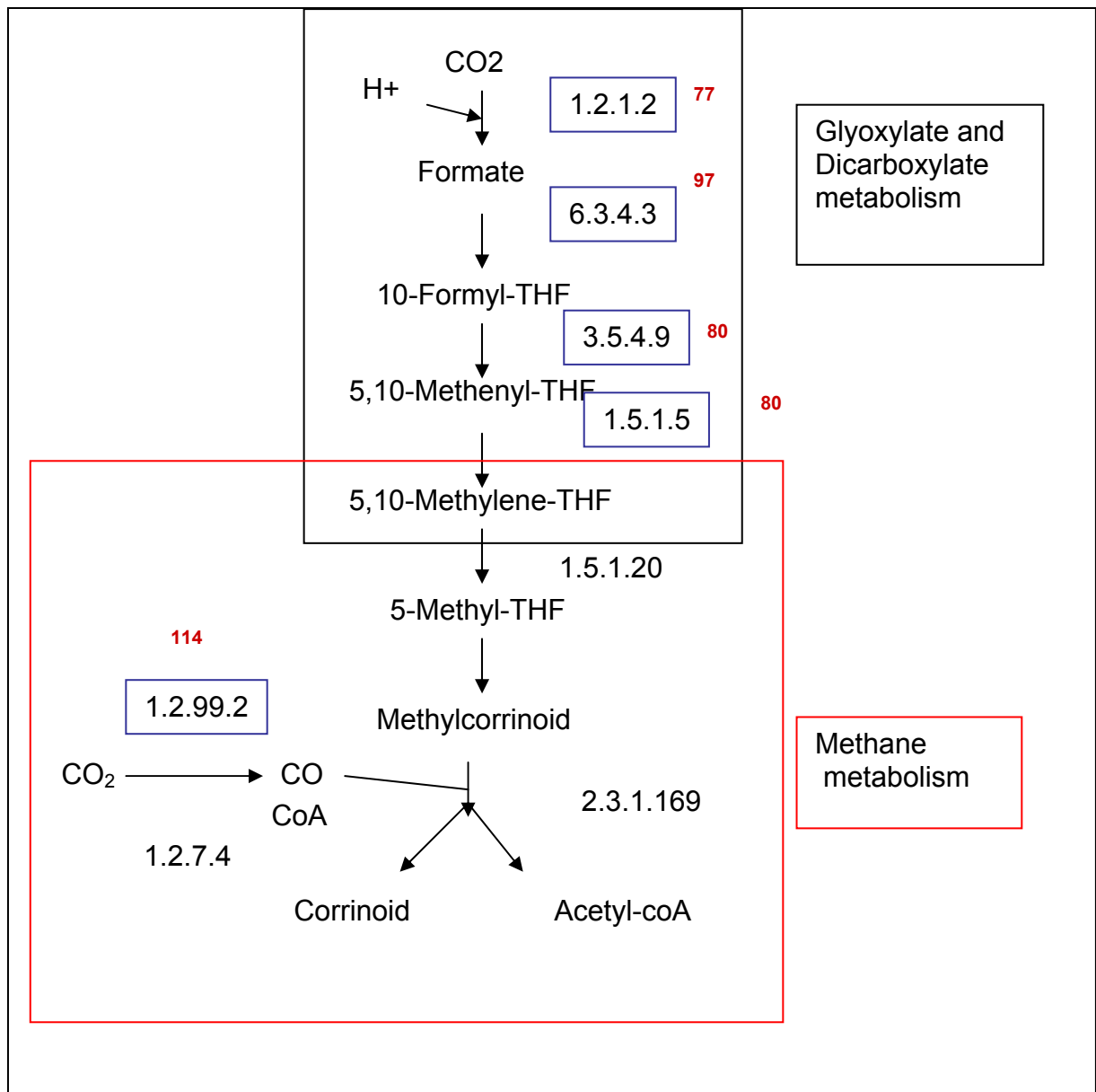
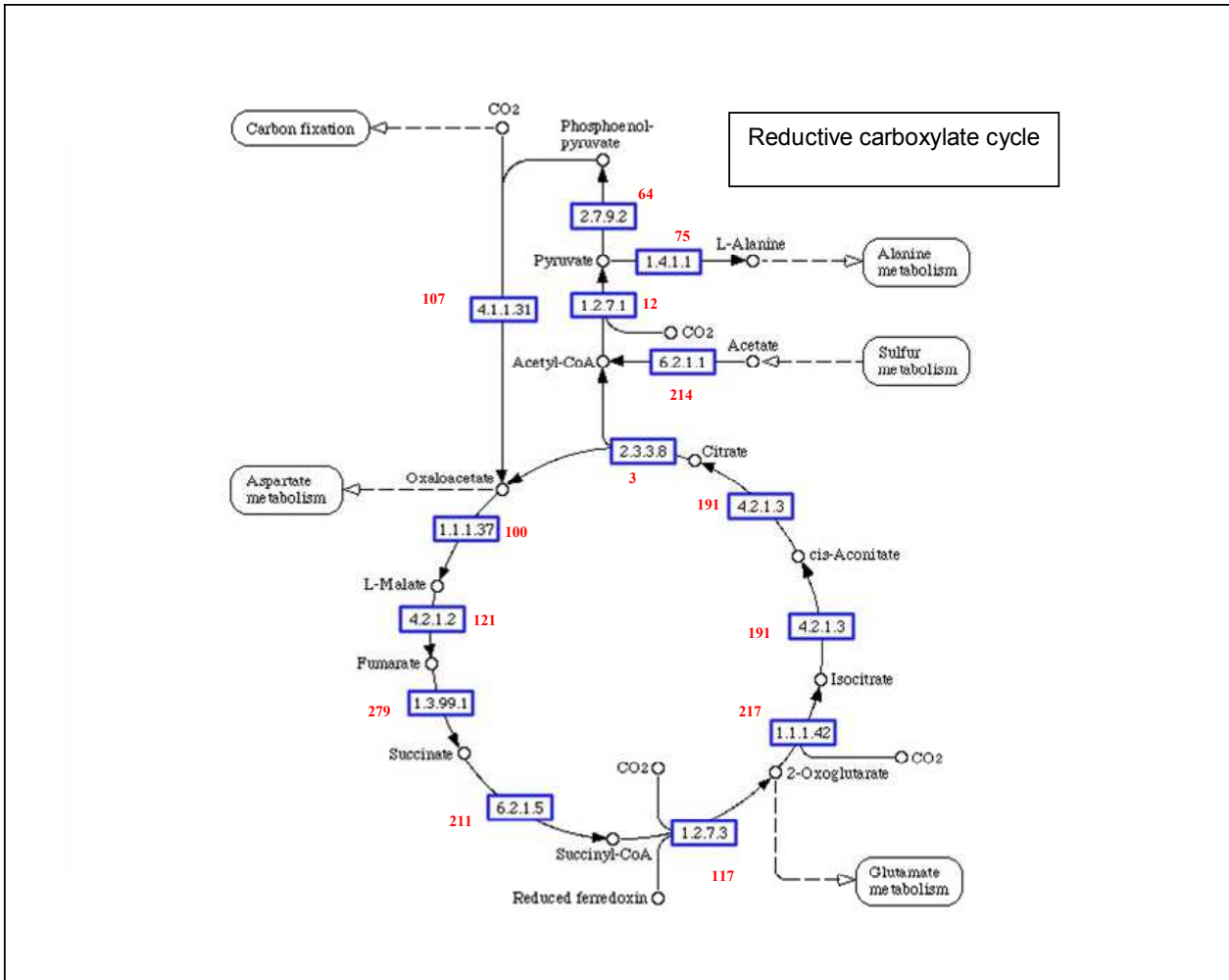


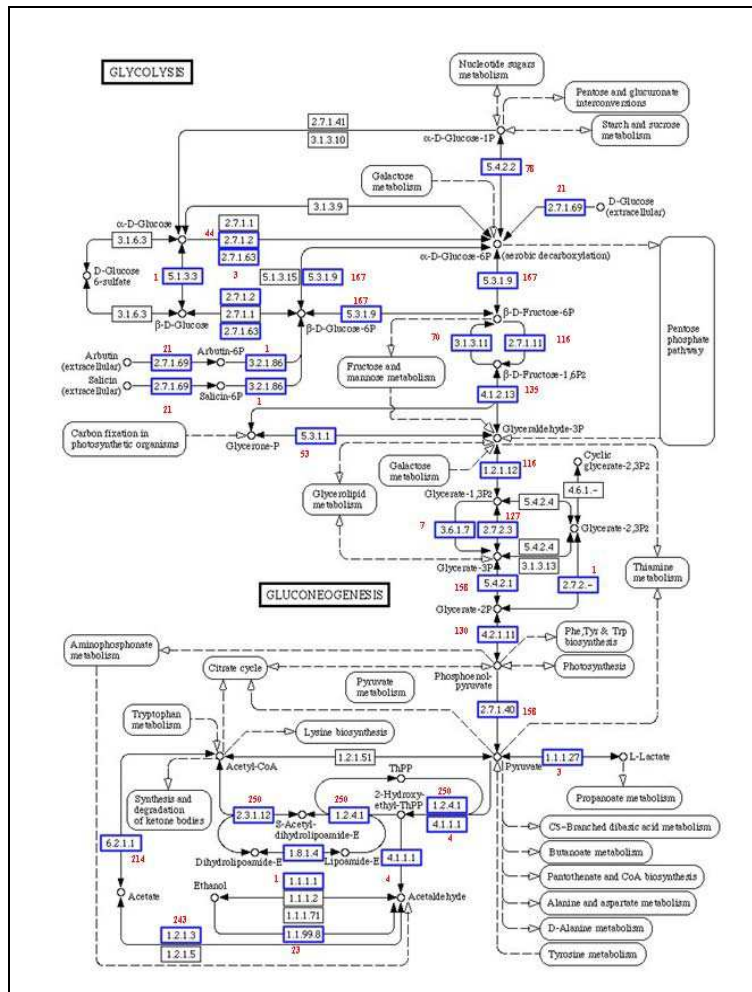
Figure 44: Map for the Acetyl coA pathway (not available from KEGG). Blue rectangles represent enzymes predicted from the analysis of 427.616 sequences of the metagenome dataset. Number of occurrences of enzymes found is written in red.





**Figure 45:** Kegg map for reductive carboxylate cycle pathways analysed using ASGARD. Blue rectangles represent enzymes predicted from the analysis of 427.616 sequences from the metagenome dataset. Number of occurrences of enzymes found is written in red.

We also found many enzymes involved in the heterotrophic fixation of carbon. Organic carbon produced by autotrophs can be used by heterotrophic bacteria, archaea and eukarya. We found all enzymes of the fructose, mannose, galactose, glycogen, sucrose, and aminosugars cycles (not shown). Classical glycolytic and gluconeogenic pathways were also recovered from our metagenome dataset (figure 46). Other kinds of heterotrophy can be predicted since enzymes involved in the catabolism of complex aromatic compounds like dichlorobenzoate, benzoate, fluorobenzoate, and other molecules such as dichloroethane were found. These compounds are supposed to be abundant in this kind of ecosystem and could be important sources of organic carbon assimilation and energy, independently of autotrophic mechanisms. Anaerobic organisms in hydrothermal sediments have been shown to be able to oxidize aromatic compounds (Tor & Lovley, 2001), and benzoate-like molecules (Tor *et al.*, 2003), which indicates that microbial consortia able to completely oxidize complex organic compounds are present (Tor *et al.*, 2003). Such oxidations are coupled with the reduction of molecules such as sulfites or nitrates (Tor *et al.*, 2003). These abilities to degrade complex organic compounds would also indicate detoxification and adaptation to the environment, a common feature of symbiotic hydrothermal organisms (Alayse-Danet *et al.*, 1987).



**Figure 46:** Kegg map for glycolytic and gluconeogenic pathways analysed using ASgard. Blue rectangles represent enzymes predicted from the analysis of 427.616 sequences from the metagenome dataset. Numbers indicated in red are the numbers of occurrences of enzymes found.

### Methane assimilation

At the hydrothermal site studied, Lucky Strike, methane is emitted by the smokers within fluids (Pimenov *et al.*, 2000), and 115.2  $\mu\text{M}$  has been detected (in the fluid) at this particular site (Desbruyères *et al.*, 2001). Methanotrophy is therefore very common. Different genera of methanotrophic bacteria have been identified in the hydrothermal vent plumes, sediments, and also in symbiotic associations, with *Bathymodiolus* for example (Nercessian *et al.*, 2005; Robinson & Cavanaugh, 1995).

As expected, the analyses of the metagenome dataset revealed enzymes involved in methanotrophy (figure 47). We found enzymes transforming methane to methanol (Methane monooxygenase ec 1.14.13.25; figure 47), methanol to formaldehyde (ec 1.11.1.6/1.11.1.7/1.11.99.8, figure 47), formaldehyde to formate (Alcohol dehydrogenase; ec 1.1.99.8) and formate to  $\text{CO}_2$  (Formate dehydrogenase ec 1.2.1.2). Formaldehyde could be used to form aminoacids (serine, lysine, or threonine), as a source of organic carbon formation by the Pentose Phosphate pathways, or to provide energy by glycolysis (figure 46). The resulting  $\text{CO}_2$  could be used to produce Acetyl CoA, which in turn could be an entry point in amino acid production, or in many other cycles, via the formation of pyruvate molecules.  $\text{CO}_2$  could also be released and then used by autotrophic organisms. An additional indication of methanotrophic mechanisms is the presence of enzymes involved in methane transport (not shown).

As very few reads were assigned to archaea, we hypothesized that in this particular habitat, methanotrophy would mainly be mediated by bacteria. From MEGAN analyses and metabolic predictions, we found signatures of the two types of methanotrophic mechanisms (type I and type II) where a *Methylococcus* is likely the major microorganisms involved. It has recently been shown that *Methylococcus* species display great metabolic flexibility, including an ability to oxidize hydrogen and sulfur, which provides an advantage in terms of adaptation in deep marine hydrothermal ecosystems (Warn *et al.*, 2004). However, it is apparent from figure 47 that only 4 reads were assigned to methane monooxygenase, the enzyme driving the transformation of methane to methanol. There is thus discordance between the taxonomic analyses, which reveal a strong presence of methanogenic bacteria, and the functional analyses. A possible explanation could be the structure of this enzyme in three sub-units, located in the membranes, the genes of which display considerable variation. For example, it has been shown that *Methylococcus capsulatus* bacteria, isolated from the deep-sea, share an original methane monooxygenase, specific to these species (Hiep-Hoa *et al.*, 1998).

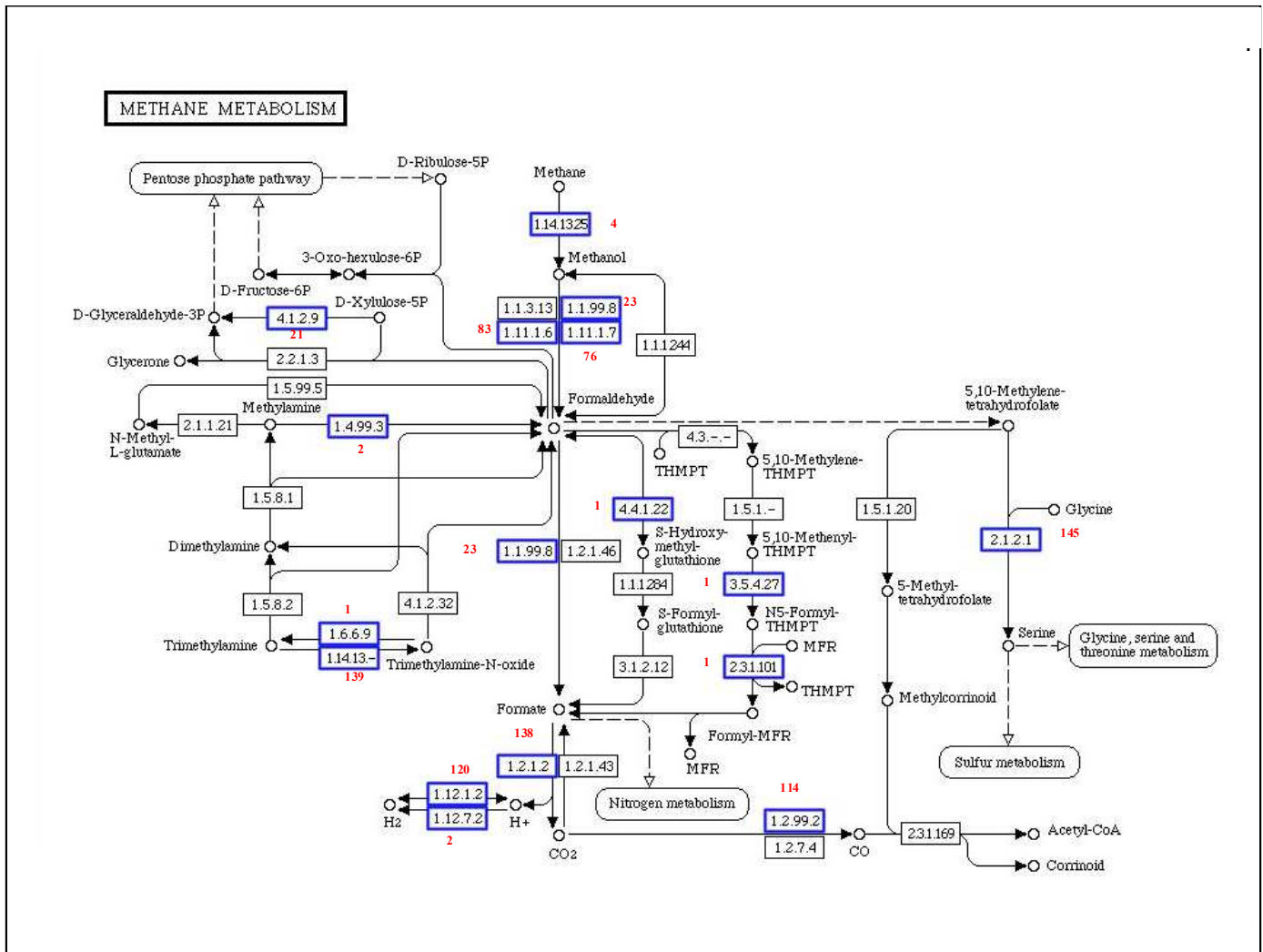


Figure 47: Kegg map for methane metabolism pathways analysed using ASGAR. Blue rectangles represent enzymes predicted from the analysis of 427.616 sequences from the metagenome dataset. Number of occurrences of enzymes found is written in red.

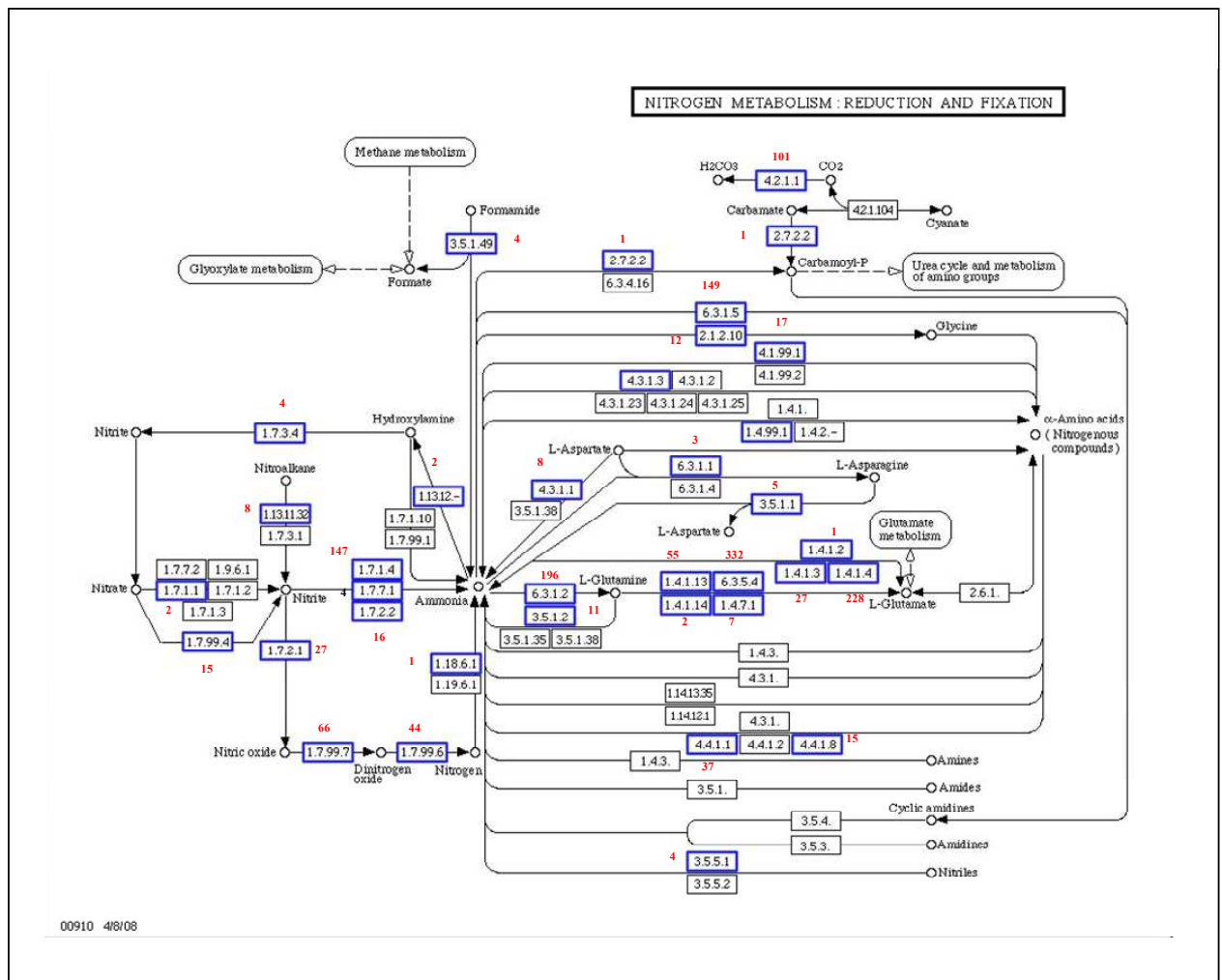
## Nitrogen uptake

Nitrogen assimilation is a complementary reaction of associated autotrophic mechanisms, and is generally essential for chemoautotrophic symbionts (Lee *et al.*, 1999). As reported by Lam *et al.* (2003), ammonia can be removed from seawater by both aerobic and anaerobic processes. In hydrothermal ecosystems, aerobic chemolithoautotrophic ammonia-oxidizing bacteria are mainly responsible for ammonia removal (Lam *et al.*, 2003).

We investigated the mechanisms involved in nitrogen uptake (figure 48), and highlighted two major metabolic pathways. The entry point into the nitrogen cycle seems to be nitrate. Nitrate concentrations ranged from 0.5  $\mu\text{M}$  to 20.03  $\mu\text{M}$  at the studied site (Desbruyeres *et al.*, 2001). This can be metabolized into nitrites via nitrate reductase (ec 1.7.1.1 and ec 1.7.99.4; figure 48). The nitrite produced can be converted either to ammonium (3 enzymes figure 48) or to  $\text{N}_2\text{O}$  (ec 1.7.2.1, figure 48), then to NO then  $\text{N}_2$  (figure 48). A denitrification process may occur in this habitat given that all the required enzymes were predicted from the sequence analysis. However, the number of times that the gene encoding the first enzyme was found (conversion of nitrate to nitrite) was small, although micro-organisms such as *Nitrosomonas* and *Nitrosococcus*, *Bacillus*, *Pseudomonas*, were detected by the MEGAN analyses.

It can also be concluded from the metagenome analysis that ammonium is crucial to nitrogen cycling. This can either be produced from nitrate or from  $\text{N}_2$  (only one gene signature for the latter, figure 48). It can be taken up from the environment and is a key molecule in amino acid anabolism (cellular utilisation) and catabolism (degradation). However, no data is available about the ammonium concentration in this environment.

An alternative hypothesis could be proposed: nitrate and ammonium would be produced from  $\text{N}_2$  in a very localized area and widely distributed into the environment from this source.



**Figure 48:** Kegg map for nitrogen metabolism pathways analysed using ASGARD. Blue rectangles represent enzymes predicted from the analysis of 427.616 sequences from the metagenome dataset. Number of occurrences of enzymes found is written in red.

## Assimilation of Sulfur

At Lucky Strike, a high quantity of hydrogen sulfide is released into the environment within the fluids of this black smoker (Cavanaugh, 1983). Part of this H<sub>2</sub>S is precipitated as polymetallic sulfide to form hydrothermal chimneys. Some of the emitted sulfur is not precipitated and can be used as an energy source by autotrophic microorganisms, and by sulfur-oxidising micro-organisms (Goffredi *et al.*, 1997). Primary production in these ecosystems involves sulfur-oxidising bacteria. The MEGAN analyses suggest that bacteria from the *Beggiatoa* genus are the main micro-organisms involved in this process (Nelson *et al.*, 1989). Interestingly, sulfur-oxidizing symbionts are also present and, according to the MEGAN results, an equal proportion of free-living and symbiotic organisms would be involved in sulfur oxidation. H<sub>2</sub>S would be used as an electron donor and oxidised to sulfate molecules (figure 49). Sulfate can be used in cellular metabolism: sulfurous chemolithoautotrophic bacteria may assimilate CO<sub>2</sub> (via acetate molecules) by oxidizing H<sub>2</sub>S (Cysteine synthase ec 2.5.1.57; figure 49). Many symbiotic organisms in deep sea hydrothermal ecosystems, such as endosymbionts of the tubeworm *Riftia pachyptila*, are able to fix CO<sub>2</sub> in this way (Robinson *et al.*, 1998). From the gene prediction and MEGAN analyses, we can suggest that *Beggiatoa*-like bacteria would convert H<sub>2</sub>S to elemental sulfur (S) and sulfate, while the *Thiobacillus*-like bacteria in our metagenome dataset, would oxidize sulfur and thiosulphate to fuel the cycle (Sulfur dioxygenase ec 1.13.11.84/ Thiosulphate reductase ec 2.8.1.5; figure 49) as shown previously in hydrothermal environments (Durand *et al.*, 1993).

The metagenome analysis also revealed a possible anaerobic reduction of sulfates. In this case, sulfates are first activated to APS (Sulfate adenylyltransferase ec 2.7.7.4), which are mainly transformed to sulfites by adenylyl reductase (ec 1.8.99.2; figure 49). Sulfites are then transformed to H<sub>2</sub>S by sulfite reductase (ec 1.8.3.1; figure 49). However, only one read was associated with sulfite reductase, and could be due to BLASTX sensitivity. The anaerobic reduction of sulfates involved bacteria in the *Desulfovibrio* and *Geobacter* genera. H<sub>2</sub>S produced in this way, would mainly be metabolised to acetate, which, in turn, would be incorporated into the TCA cycle to fix CO<sub>2</sub>. We also found enzymes involved in the catabolic reduction of sulfate, in an anaerobic process, which implied the presence of bacteria from the Desulfuromonadale order.





## **Hypotheses of mechanisms of metagenome functioning**

Analysis of the whole metagenome revealed the presence of different metabolic processes governed by different actors. The possible contributions of the different organisms were studied to predict ecosystem functioning. Possible interactions between the organisms and the mechanisms that they drive were also proposed. A hypothetical model is provided in Figure 50. The main metabolisms detected were: catabolic reduction of sulfates, anabolic reduction of sulfates, sulfur oxidation, methane assimilation, anaerobic reduction of nitrates and denitrification (figure 50). All these mechanisms could be linked to the chemical properties of the environment studied, and our metagenomic approach proved to be a pertinent means of accessing the entire metabolic diversity within an environment. The figure 51 presented herein try to draw all the different bacterial metabolisms in the ecosystem; we also suggest in this scheme, the speculative interactions between organisms in the bacterial consortia.

We also draw from this scheme, the hypothetical interactions between fungi and their environment. The main interactions detected would result in allelopathic mechanisms, for ecological niche competition, but bacteria could also be useful to fungi in this environment by metabolising the CH<sub>4</sub> emitted by smokers, in a form which could then be used by fungal organisms.

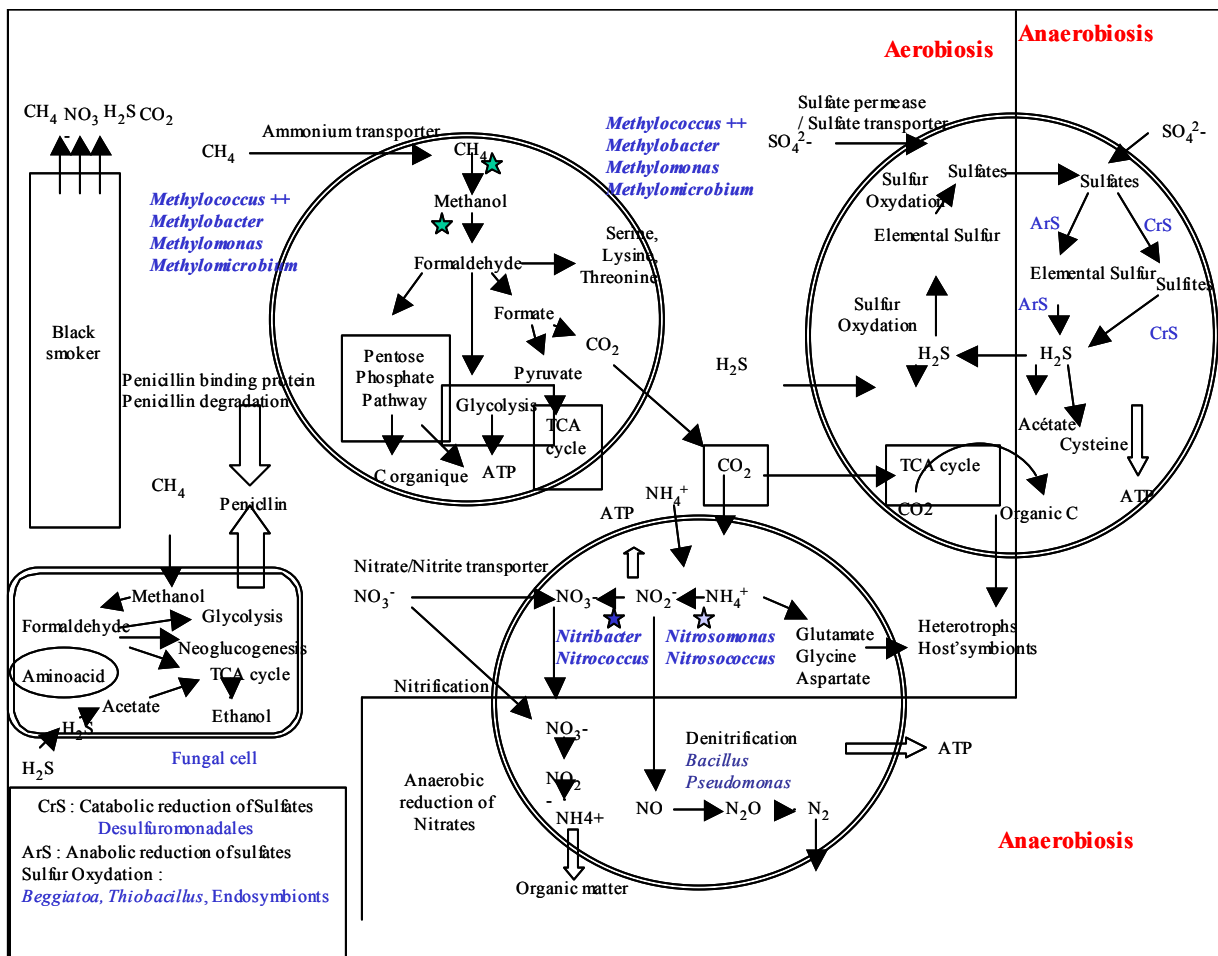


Figure 50: Hypothetical interactions between microorganisms in this habitat. Circles represent hypothetical bacterial cells and rectangles the hypothetical fungal cell.

The proposed major organisms driving the pathways are indicated in blue. Aerobic and anaerobic compartments are represented.

## **Concluding remarks**

The analysis of our metagenomic dataset revealed a complex bacterial community. Exploring the nucleotide sequences of protein-coding genes for the subsystems (KEGG maps), combined with taxonomic information obtained with MEGAN software, allowed us to rely, at least in part, on a function to the each micro-organism involved. “Who’s doing what?” is a central question in modern ecology and crucial to our understanding of the different hierarchical levels, whether molecules, individuals, communities, or ecosystems.

Our study could be reproduced in other habitats, but the complexity of the sample(s) analysed clearly limits its comprehension.

Also, as demonstrated in previous metagenomic studies, our results are dependent on the composition of the databases, and if little information is available about a given group in the international databases this results in under-represented sequence information even if the organism occurs frequently in the analysis. Testing the predictions by complementary studies, such as characterization of the expressed genes in the sample (metagenomic and metatranscriptomic analyses), nutrients acquisition (for an example by isotope probing strategy) should provide a more comprehensive story. However, RNAs in themselves have other limits related to the non constitutive expression of genes, maturation of RNAs, etc...

Exploring fungal metabolism in this environment was not easy. This could be directly related to the choice of sample (unknown species from the old Chytridiomycota lineage). Since only limited data were available for the Chytrids (Table 9), only a limited number of fungal contigs could be extracted. This is a pioneer and cutting edge study in terms of methods and data analyses. It can therefore be speculated that re-analysis of the dataset in future years will provide more details about the fungal function in this sample due to the rapid increase in size of the sequence data bank. Although our tests indicated that the sequencing effort was satisfactory, additional pyrosequencing with new kits which provide reads of <500 bp might be of interest.



**Partie VI : Diversité des souches fongiques filamenteuses  
isolées d'écosystèmes hydrothermaux marins profonds**

---

## Résumé

L'étude de la diversité fongique réalisée lors de cette thèse s'est principalement focalisée sur l'analyse du matériel génétique fongique isolé de ces écosystèmes ; ce travail était complété par l'analyse taxonomique de souches isolées des mêmes échantillons. Ce travail, réalisé en collaboration avec l'ESMISAB (Brest, France) a permis d'isoler des souches représentant 12 phylotypes appartenant tous au phylum Ascomycota.

**Les résultats présentés dans cette partie sous la forme d'une publication acceptée dans *Environmental microbiology*, s'intéressent à l'étude réalisée par l'ESMISAB, notamment par Gaétan Burgaud et Georges Barbier, sur l'étude d'une collection de souches isolées de différents échantillons hydrothermaux prélevés dans différentes zones géographiques.** L'isolation des souches a été effectuée sur milieu salé (NaCl 3%) à température et pression ambiantes. D'autres milieux de culture ont également été testés. Le travail effectué à l'ESMISAB a permis d'isoler des souches filamenteuses (62 souches) et des souches levuriformes (35 souches) mais l'étude présentée dans cette publication ne s'est intéressée qu'aux formes filamenteuses.

Des études physiologiques, principalement basées sur la culture des souches filamenteuses dans différentes conditions de température et de salinité a permis de discriminer les souches halophiles / non halophiles, ainsi que les souches psychrophiles / mésophiles. Des hypothèses d'endémisme ont été discutées sur la base de ces résultats.

L'analyse taxonomique de la collection, dans un premier temps menée par analyse microscopique puis par analyse phylogénétique (amplification du gène codant l'ARNr18S couplée à l'amplification des régions ITS de l'ARNr dans certains cas) dans un second temps. A l'exception d'une souche, tous les isolats fongiques appartiennent au phylum des Ascomycota, retrouvés dans au moins 13 ordres différents. Il apparaît que la majorité des souches aient été isolées d'invertébrés, principalement des modioles (*Bathymodiolus azoricus*) et des crevettes (*Rimicaris exoculata*).

Des hypothèses sur les modes de vie de ces organismes sont également discutées dans cette publication, notamment sur le parasitisme supposé des souches appartenant à l'ordre des Chaetothyriales (phylum Ascomycota), groupe de levures noires. Ces levures ont été identifiées lors de précédentes études, comme responsables d'épizooties de modioles en écosystème hydrothermal (Van Dover *et al.*, 2007).

La publication, sous presse à l'heure de l'écriture de cette thèse, est disponible à l'adresse suivante : <http://www3.interscience.wiley.com/cgi-bin/fulltext/122212184/PDFSTART>

# Diversity of culturable marine filamentous fungi from deep-sea hydrothermal vents

Gaëtan Burgaud<sup>1</sup>, Thomas Le Calvez<sup>2</sup>, Danielle Arzur<sup>1</sup>, Philippe Vandenkoornhuyse<sup>2</sup> and Georges Barbier<sup>1</sup>

<sup>1</sup>Laboratoire Universitaire de Biodiversité et Ecologie Microbienne (EA3882), IFR 148 Université Européenne de Bretagne, Université de Brest, ESMISAB, Technopole Brest-Iroise, 29280 Plouzané, France.

<sup>2</sup>UMR CNRS ECOBIO, FR/IFR CAREN, Université Européenne de Bretagne, Université de Rennes 1, Campus de Beaulieu, 35042 Rennes cedex, France.

---

## Summary

Since now very few studies have been carried out on deep-sea marine fungi, thus this field remains relatively unknown. However, their presence inside benthic microbial eukaryotes at deep-sea vents was recently pointed out from molecular microbial ecology studies. We report here an attempt to describe the culturable part of mycological communities in deep-sea vent ecosystems that is an important step in understanding their diversity, abundance, and function. Physiological characterization revealed strains that are more or less adapted to deep-sea conditions. Those results suggest the presence of true marine organisms and other more ubiquitous. Phylogenetical characterization highly correlated to physiological data revealed the presence of fungi that have been previously described and unknown ones until now, belonging to new taxonomic groups. This survey encourages for further work in order to complete descriptions and also to describe the ecological role of these organisms in such extreme environments.



## Introduction

Fungi belong to microbial eukaryotes and colonize diverse habitats, terrestrial, aquatic and marine. Marine fungi comprise a small and ecologically defined group of filamentous ascomycetes, their anamorphs and yeasts (Kohlmeyer & Kohlmeyer, 1979) but they have been little studied until now. Described higher marine fungi are listed as 467 species belonging to 244 genera (Kis-Papo, 2005). Fungi from aquatic and marine habitats have received much attention in recent years to assess their diversity and to elucidate their ecological role (Kohlmeyer & Kohlmeyer, 1979; Hyde *et al.*, 2000; Shearer *et al.*, 2007). Production of bioactive metabolites was also screened and reported (Cuomo *et al.*, 1995; Christophersen *et al.*, 1998; Cueto *et al.*, 2001; Rowley *et al.*, 2003; Lin *et al.*, 2005; Badhury *et al.*, 2006). Most of the reported studies on marine fungal diversity have focused on tropical mangroves, salt marshes and open oceanic regions (Jones & Alias, 1997; Buchan *et al.*, 2002; Kohlmeyer & Kohlmeyer, 1979). In tidal zones, in brackish and estuarine habitats, fungi are involved in the degradation of organic matter; their ecological importance is established as saprobes, pathogens and symbionts in algae, higher plants and animals (Kohlmeyer & Kohlmeyer, 1979; Hyde *et al.*, 1998). Concerning yeasts, pink basidiomycetes seems to be the most common group of yeasts found in the marine environment (Lorenz & Molitoris, 1997) which presence was correlated with organic matter resulting from polluted waters, water boundaries, decaying plankton blooms and macroalgae. The role of fungi and yeasts in marine habitats seems underestimated (Hyde *et al.*, 1998), particularly in extreme environments such as deep-sea. Presence of decaying algae, higher plants and wood in the deep-sea suggest that fungi may be transported to the deep-sea attached to such substratum and/or to particulate organic matters, they could also sediment as spores or mycelial filaments (Lorenz & Molitoris, 1997). Over the past several years, there has been an increasing interest in diversity of microbial eukaryotes in extreme environments. Studies employed culture-independent methods combined with sequence analysis of the small subunit ribosomal RNA genes (SSU rRNA) of microeukaryotes. This approach was applied to survey the microeukaryotic diversity in extreme environments such as acidic and iron rivers (Zettler *et al.*, 2001; Zettler *et al.*, 2003; Gadanho & Sampaio, 2006), anoxic shallow sediments of marine water and freshwater (Dawson & Pace, 2002; Stoeck *et al.*, 2006), suboxic waters and anoxic sediments in salt marsh (Stoeck & Epstein, 2003), anoxic deep-sea waters (Takishita *et al.*, 2005; Zuendorf *et al.*, 2006) and deep-sea hydrothermal vents (Edgcomb *et al.*, 2002;

López-García *et al.*, 2003; Gadanho & Sampaio, 2005; López-García *et al.*, 2007; Bass *et al.*, 2007). These studies revealed a majority of sequences belonging to *Alveolata* and *Stramenopiles*. Fungi were scarce or even absent but some fungal sequences harvested at deep-sea vents were novel. Scientists studying microeukaryotic diversity in an acidic iron river revealed new fungal phylotypes with fungal-specific primers that were not detected when universal eukaryotic primers were used (Gadanho & Sampaio, 2006). These results suggested that all the previous studies could have underestimated fungal diversity in extreme environments. Considering the revealed existence of original fungal diversity at deep-sea vents, a new evident following step was isolation and description of corresponding organisms. Yeast diversity was assessed at Mid-Atlantic Ridge and revealed 12 phylotypes belonging to Ascomycota and 7 to Basidiomycota (Gadanho & Sampaio, 2005). 33% of the isolated strains belonged to new phylotypes. In this study, we have decided to assess the presence of culturable fungi with an emphasis on filamentous fungi at deep-sea vent ecosystems. To our best knowledge, no previous report has dealt with culturable filamentous fungi from deep-sea vent ecosystems. Fresh and frozen samples collected at Mid-Atlantic Ridge (MAR), South West Pacific-Lau Basin and East Pacific Rise (EPR) during six oceanographic cruises: BIOLAU (1989), DIVANAUT2 (1989), HERO (1991), MARVEL (1997), EXOMAR (2005) and MoMARDREAM-Naut (2007), were used to isolate culturable fungi. We report here a global description of the resulting culture collection of filamentous fungi.

## **Material and methods**

### **Environmental sampling**

*Hydrothermal vents.* 210 hydrothermal samples were collected during 6 oceanographic cruises:

- i) BIOLAU in the Lau Basin, South West Pacific (12/05/1989-27/05/1989; 20°3.0' S, 176°7.8' W; -2620m);
- ii) DIVANAUT2 (19/06/1994-01/07/1994; on the Mid-Atlantic Ridge at Menez Gwen (37°51'N, 31°31'W; -900 m) and Lucky Strike (37°17'N, 32°16'W; -1650 m) sites;
- iii) HERO on the East Pacific Rise at Elsa site (30/09/1991-04/11/1991; 12°48'N, 103°57'W; -2630 m);
- iv) MARVEL (29/08/1997-13/09/1997) on the Mid-Atlantic Ridge at Menez Gwen (37°51'N, 31°31'W; -900 m) and Lucky Strike (37°17'N, 32°16'W; -1650 m) sites;
- v) EXOMAR (25/07/2005-28/08/2005) on the Mid-Atlantic Ridge at Rainbow

(36°08'N, 34°00'W, -2300 m), TAG (26°02'N, 44°54'W, -3630 m) and Lost City (30°04'N, 42°12'W, -900 m) sites; vi) MoMARDREAM-Naut (08/07/2007-19/07/2007) on the Mid-Atlantic Ridge at Rainbow site (36°08'N, 34°00'W, -2300 m).

*Surface.* Samples of *Sargassum fluitans* were found floating at ocean surface in the TAG area during EXOMAR. Individuals were sampled with surrounding seawater for comparison with deep-sea hydrothermal samples using sterile strip and tubes.

Deep-sea sampling was performed using the Deep Submergence Vehicle Nautilie for BIOLAU, HERO, DIVANAUT2 MARVEL and MoMARDREAM-Naut

(<http://www.interridge.org/en/node/4897>) cruises and the ROV “Victor 6000” for EXOMAR cruise. Support research vessels were “Nadir” for HERO, BIOLAU and DIVANAUT2 cruises, “Atalante” for MARVEL and EXOMAR cruises and “Pourquoi Pas ?” for MoMARDREAM-Naut cruise ([www.ifremer.fr/fleet//index.php](http://www.ifremer.fr/fleet//index.php)). Deep-sea waters were collected with titanium syringes, sediments with a sediment sampler, shrimps with a fauna aspirator, alvinellids, mussels, sponges and minerals using insulated boxes. After washing, those equipments were sterilized (sediment sampler) or ethanol disinfected (titanium syringes, aspirator containers and pipes, sampling boxes) and filled with sterile seawater.

All samples were processed directly after the Nautilie or ROV recovery. On board, in the lab, sterility was obtained with Bunsen burner or using a vertical laminary flow hood. Solid samples were taken in their container with sterile strips, laid on a sterile Petri dish and abundantly rinsed with sterile seawater before crushing in a mortar with a pestle, both autoclaved previously. Sterile seawater was added for crushing when necessary. Sterile seawater contained (per liter) 30 g of Sea salts (Sigma) diluted in 1 liter distilled water filtered on 0.22 $\mu$  and autoclaved to obtain sterility. Collected seawaters, sediments and crushed samples were used to inoculate culture media and frozen in cryotubes at -70°C with DMSO as cryoprotectant (5% v/v) for further use. During MoMARDREAM-Naut cruise, dissections were realised on animal samples in order to analyze precisely fungal location.

### **Enrichment conditions**

During HERO cruise, Sabouraud (Sab) chloramphenicol solid medium (AES Laboratoire) was used for aerobic enrichment cultures at 30°C and atmospheric pressure. During MARVEL cruise, 5 solid culture media were used. Ac medium contained per liter: potato starch (Sigma) 5 g, yeast extract 0,5 g, peptone 1 g, sea salts 30 g and PIPES buffer (Sigma)

6,05 g. Starch was replaced by cellobiose (Sigma) 5 g in Cc medium, by glucose (Sigma) 5 g in Gc medium, by arabic gum 5 g and olive oil 5 g in Lc medium, by xylan oat spelts 5 g in Xc medium. Pc medium contained per liter: brain heart infusion (Difco) 9 g, NaCl 23g and PIPES buffer 6,05 g. pH was adjusted to 7.5 with NaOH 4N. The 5 media were supplemented per liter with agar 15 g and chloramphenicol 500 mg. Cultures were done aerobically at 25°C (ambient temperature) and atmospheric pressure. During EXOMAR and MoMARDREAM-Naut cruises, GYPS medium was used and contained per liter: glucose 1 g, yeast extract 1 g, peptone 1 g, starch 1 g, sea salts 30 g. This medium was supplemented per liter with agar 15 g and chloramphenicol 500 mg. pH was also adjusted to 7.5. Cultures were done aerobically at 4°C, 15°C, 25°C (ambient temperature), 45°C (only during EXOMAR) at atmospheric pressure. Pure cultures were obtained by streaking and central picking on their respective enrichment media and kept in culture collection by continuous culture on GYPS medium supplemented per liter with agar 15 g.

Nutrient plates exposed under laminary flow hood during each isolation were processed as control. No fungal colonies yielded on control plates indicating no occurrence of aerial contaminations during isolation. Exception made for the laminary flow hood during EXOMAR cruise that led to the isolation of *Penicillium* on control plates. All *Penicillium* strains with identical morphology isolated during this cruise were eliminated.

Each purified strain from our collection has been integrated to the “Souchothèque de Bretagne”

(<http://www.ifremer.fr/souchotheque/internet/htdocs/generique.php?pagebody=catalogue.php>) and are available with an LMSA number associated to their GenBank number.

### **Physiological characterization**

All experiments were done in triplicate. Fungal strains were grown on solid and yeast in liquid GYPS medium. The effect of temperature was determined at 5, 15, 25 and 35°C. The effect of salinity was determined in modifying sea salts concentrations in GYPS medium: 0, 15, 30, 45 and 60 g per liter.

### **Growth measurement and Statistical analyses**

After central picking on Petri dishes, mycelium diameter was measured each 7 days for a maximum of 5 weeks to characterize fungal growth. Principal Component Analysis used to analyze physiological results was done with “Statbox 6.6” software (Grimmersoft).

### **DNA Extraction and Purification**

DNA Extraction was performed by FastDNA Spin Kit (MP Biomedicals) that is specific for fungi and yeasts following the manufacturer’s recommendations.

### **DNA Amplification**

18S rRNA gene sequences were amplified with NS1 (5'-GTAGTCATATGCTTGTCTC-3') and SR6R (5'-AAGTAGAAGTCGTAACAAGG-3') primers (White *et al.*, 1990) for genetic identification of fungi (Vandenkoornhuyse *et al.*, 2002a). When SR6R was unefficient, 1020r (5'-AAACATCCTTGGRAATGCTTT-3') or 1055r (5'-AAGAACGGCCATGCACCAC) (Atkins, 2000) primers were used. Internal Transcribed Spacer and rRNA 28S (partial) were amplified using ITS5 (5'-GGAAGTAAAAGTCGTAACAAGG-3') and LR6 (5'-CGCCAGTTCTGCTTACC-3') primers (respectively White *et al.*, 1990; Vilgalys and Hester, 1990). All PCR reactions were performed in 20µL reaction volumes containing 1X Taq Titanium Buffer (Ozyme), 2mM of each of the four dNTPs (Promega), 0.08µM of each primer (Prologo), 1U of Taq Titanium DNA polymerase (Ozyme) and 1µL of genomic DNA. Reaction mixtures were incubated in a PTC-200 thermal cycler (MJ Research). PCR temperature profile consisted of an initial denaturation step at 94°C for 2 min, followed by 30 cycles of 30 sec at 94°C, 30 sec at 54°C and 2 min at 72°C and a final extension step of 2 min at 72°C before a conservation at 4°C. A negative control in which DNA was replaced by sterile water was included. PCR products were controlled by electrophoresis in 0,8% (w/v) agarose gel (Promega) in 0.5X Tris-Borate EDTA (TBE) buffer at 90V for 90 min and stained with ethidium bromide. Molecular size marker was Lambda DNA/EcoR1 + Hind III Markers (Promega). DNA banding patterns were visualized under UV transillumination and picture files were generated using Gel-Doc 2000 (Biorad).

## Phylogenetic analyses

Fungal amplicons were sequenced with NS1, NS3 (5'-GCAAGTCTGGTGCCAGCAGCC-3'), NS5 (5'-AACTTAAAGGAATTGACGGAA-3') (White *et al.*, 1990) and SR6R primers for SSU rRNA gene sequences. For ITS and LSU rRNA, ITS5 and LR6 primers were used. Sequences were obtained by “Big Dye Terminator” technology (Applied Biosystems). This work was done at “Ouest Génopole” sequencing facility in the “Station Biologique de Roscoff” ([www.sb-roscoff.fr](http://www.sb-roscoff.fr)).

Chromatograms obtained were translated in nucleotidic sequences with DNA Baser v.2.10 (<http://www.dnabaser.com>). After cleaning, sequences were imported to MEGA 4.0 software (Tamura *et al.*, 2007). Each sequence was analyzed in order to find GenBank sequences with close BLAST-N hits (Altschul *et al.*, 1990). Additional 18S rRNA sequences from clones harvested during environmental deep-sea studies were included in our analysis (Edgcomb *et al.*, 2002; López-García *et al.*, 2003; Takishita *et al.*, 2005; López-García *et al.*, 2007; Bass *et al.*, 2007). Similarities between sequences were assessed using pairwise distance calculation with MEGA 4.0. Finally, sequences from the Assembling the Fungal Tree Of Life (AFTOL) project have been added to our phylogenetic tree (James *et al.*, 2006). Sequences were trimmed to ensure that all sequences had the same start and end point.

All SSU rRNA sequences were aligned using ClustalW v.1.83 (Thompson *et al.*, 1994), one of the most widely used packages, and other packages described as faster and more accurate: MAFFT v.5 (Kato *et al.*, 2005) and MUSCLE v.3.6 (Edgar, 2004). After visual checking and manual curation, each alignment was analyzed for the Bayesian estimation of phylogeny using MrBayes v.3.1.2 software (Ronquist and Huelsenbeck, 2003). Three million generations option has been set to run the Metropolis-coupled Monte Carlo Markov Chain method (*mcmc*). After generation 3.000.000, the average standard deviations of split frequencies were in increasing values:  $p=0.01373$  (ClustalW v.1.83),  $p=0.02190$  (MAFFT v.5) and  $p=0.02199$  (MUSCLE v.3.6). For all alignment softwares,  $p<0.05$  showing that convergence occurred. Software generating the smallest frequency (ClustalW v.1.83) has been chosen in order to produce an evolutionary tree that would be accurate and robust. Chosen alignment was analyzed using MODELTEST v.3.7 (Posada & Crandall, 1998) in order to obtain the more realistic evolutionary model used for phylogenetic analyses (TrN+I+G model; Gamma distribution shape parameter = 0.6470). Phylogeny was then evaluated using two different methods: i) Bayesian inference (BI) with MrBayes v.3.1.2 analysis using 3.000.000 generations and the *mcmc* method. The tree search included two *mcmc* searches with four chains (setting default temperature for heating the chains) and a sampling frequency of 100

generations. A “burnin” of 7500 (25% of the 3.000.000 generations / 100 sample frequency) was set in order to exclude the first 7500 trees generated; ii) Maximum Likelihood (ML) with 100 bootstrap iterations using PHYML (Guindon *et al.*, 2005) and the parameters obtained with MODELTEST v.3.7. Final phylogenetic tree topology was realized using MrBayes v.3.1.2 analysis results. Nodes in the tree are showing Bayesian posterior probabilities and ML bootstraps respectively.

## Results

### Culturable fungi isolated

A total of 210 samples were processed for fungal growth of which 42 (20%) yielded to isolation of fungi. Consequently, all other samples were free of culturable fungi in our culture conditions (80%). Hydrothermal samples consisted mostly in shrimps (71) and mussels (48) and for lower numbers in smoker rock scrapings (25), alvinellids (20), tubeworms (12), sediments (8), other animals (8), colonization modules (4), seawater (3), corals (3) and gastropods (2). Ocean surface seaweeds and seawater were sampled as control (respectively 4 and 2 samples). The final number of fungi obtained on isolation plates was 97. Morphological characteristics examined under optical microscope permitted to separate filamentous fungi and yeasts. Observations revealed branching hyphae and spores for filamentous fungi (62 isolates) and budding cells for yeasts (35 isolates). Only filamentous fungi have been exploited in this study. Samples used to build our fungal culture collection and their origins are shown in table 10. Filamentous fungi were mostly isolated from *Bathymodiolus azoricus* mussels (36 isolates) and deep-sea hydrothermal shrimps (14 isolates) when considering absolute frequencies. Those results indicate that fungi were much more associated with animals rather than mineral substrate. Fungal strains were mainly isolated from fresh samples directly on board during oceanographic cruises. However, cryoconserved samples of past cruises enabled isolation of a few strains (12 isolates/55 cryoconserved samples processed). Several strains (MV16, MV18, MV20, MV22) did not grow during the purification step and were finally lost.

Fungi were not found at the same rate in all the studied sites. The greatest part of filamentous fungi was found from MAR at Menez Gwen and Rainbow sites. Isolates from Menez Gwen dominated the filamentous fungi culture collection (27) compared with Rainbow (21) even if

the number of samples processed was lower for Menez Gwen (16) than Rainbow (97). This corresponds in fact with the presence of one *Bathymodiolus azoricus* individual that gave a high number of fungal isolates in Menez Gwen. We did not succeed to isolate any fungal strain at Lucky Strike site. Statistical distribution tests have been performed in order to know the distribution type of fungi in hydrothermal sites. The variance to mean ratio ( $\sigma^2/m$ ) was calculated for each site (Cancela da Fonseca, 1966). Significant values corresponds with  $(\sigma^2/m)-1 > 2 \cdot ((2n/(n-1))^2)^{1/2}$ . In all sites studied with sufficient number of isolates, an aggregate distribution ( $\sigma^2/m > 1$ ) was observed: Rainbow (1.75), Elsa (1.71), Menez Gwen (8.12) and Lost City (2). This indicates that fungi were located in specific niches inside the ecosystem (mainly mussels and shrimps). For TAG site, this value was not significant. No fungi were isolated from Lucky Strike and Lau Basin.

We used a chi-square analysis to test independence of fungal isolation from locations and kind of samples collected. The result was that fungal isolation was very highly dependent ( $p < 0,001$ ) on hydrothermal sites ( $p = 2,32 \cdot 10^{-14}$ ) and kind of samples collected ( $p = 2,02 \cdot 10^{-30}$ ). Thus, hydrothermal sites and kind of samples had a very highly significant influence on isolation. This very high dependence strongly supports the idea that our isolates were not contaminants. Moreover, when pieces of samples were deposited on Petri dishes, fungal colonies accounting for our collection always grew from those pieces indicating clearly that fungi were physically associated with samples. Concerning positive enrichments, most frequently, we observed only one fungal colony per isolation Petri dish. However, in some cases, mussel samples (3) yielded to isolation of several colonies (Table 10).

### **Physiological characterisation**

To discriminate individuals (filamentous fungi) for their ability to live in deep-sea hydrothermal ecosystems, principal component analysis (PCA) was carried out on growth data (colony diameters after 14 days of growth) at different temperatures and salinities (See Figure 51). Axis 1 explained 59% of the total variance and has a significance of growth axis. Axis 2 appeared as a temperature behaviour axis and gathers 20% of the total variance. 79% of the whole variance is visible on the first plan (Axis 1-Axis 2) that permitted a reliable discussion of the individual positions. On this first plan, we can observe individuals with fast growth on the right side of the figure and other with low growth on the left side distributed depending on their temperature preferendum.



When considering colony diameter at 14 days of growth, a majority of strains (35 located in the upper part of the figure 51) were psychrotrophs, defined as *those that grow at low temperatures but also above 20°C* (Deacon, 2006). Remaining strains cannot grow at the lower temperature tested in this study (5°C) and thus cannot be defined as psychrotrophs but rather as mesophiles defined as *those that grow between 10°C and 40°C with an optimal growth temperature of 20-35°C* (Hawksworth *et al.*, 1995; Deacon, 2006). No psychrophiles, defined as *those having an optimum growth at no more than 16°C and a maximum growth of about 20°C* (Deacon, 2006), were isolated in our study.

A third axis explained 8% of the total variance and is a salinity behaviour axis independent of the two first axes. Strain Ex2 was strictly halophile (no growth without sea salts) and though strongly correlated with this third axis. Four strains (Mo9, MV8, MVFS1 and MVFS4) were inversely correlated with the third axis and defined as non-halophiles (difficulty to grow increasing with sea salts present in the media). Considering colony diameters at 14 days of growth depending on the salinity parameter, 4 groups were defined:

- *Strictly halophile*: Ex2
- *Halophile*: Ex1, Ex7, Ex8, Ex9, Mo4, Mo6, Mo7, Mo12, Mo14, Mo15, Mo16, Mo17, Mo18, Mo19, MV2, MV4, MV10, MV14, MV21, MV23, MV25, MV26, MV27, HE3,
- *Halotolerant*: HE2, HE4, HE7, Mo3, Mo5, Mo8, MV13, DIVA1,
- *Non-halophile*: Ex3, Ex4, Ex5, Ex6, Mo1, Mo2, Mo9, Mo10, Mo11, Mo13, MV1, MV3, MV6, MV7, MV8, MV9, MV11, MV12, MV15, MV17, MV19, MV24, MVFS1, MVFS2, MVFS3, MVFS4, HE1, HE5, HE6.

### **Phylogenetic diversity**

The first goal of this study was to determine whether culturable fungal strains could be isolated from deep-sea hydrothermal ecosystems that were never harvested for such organisms. Based on the gene encoding SSU ribosomal RNA, our purpose was first to identify each fungal strain and, if not, to assess their taxonomic originality based on sequence similarities with closest neighbors in phylogenetic trees. SSU rRNA sequences from culture-independent studies in deep-sea ecosystems were integrated to this phylogenetical analysis. When sequences from our culture collection could not be identified by 18S rRNA, complementary genetic analyses were performed using ITS and 28S rRNA sequences.

Our 18S rRNA sequences indicated the presence of Ascomycota and Basidiomycota in our culture collection (See Figure 52). The distribution of isolates was unequal between these two phyla as majority were ascomycetes represented by Pezizomycotina subphylum. Only one isolate (HE6) was affiliated to Basidiomycota and was close to *Tilletiopsis pallescens* (16 mismatches on 1651bp).

Within Ascomycota, in order of decreasing cluster size, Helotiales were dominant (14 strains), twelve strains of which formed cluster Hel2 (MV2, MV4, MV14, MV21, MV27, MVFS3, Mo12, Mo14, Mo15, Mo16, Mo17, Mo18; See Figure 52). Eleven of those strains were isolated from *Bathymodiolus azoricus* “micro-environment” (internal and external faces of mussels) and one from *Rimicaris exoculata* (Mo12; See table 10). Their ITS rRNA gene sequences were similar to *Cadophora malorum* (100% similarity). Hel1 corresponds with two strains (HE3 and DIVA1) closely similar (4 mismatches on 1431bp for 18S rRNA) isolated respectively from *Alvinella pompejana* and *Bathymodiolus azoricus*. Their taxonomic originality was confirmed with 18S, ITS and 28S rRNA gene sequence analyses: respectively 23 mismatches on 1431bp (98% similarity) with *Helicodendron paradoxum*, 18 mismatches on 470bp (96% similarity) with *Phialophora* sp. and 42 mismatches on 1055bp (95% similarity) with *Hyphodiscus hymeniophilus*.

Chaetothyriales order pooled eleven strains. Three of which (MV1, MV6 and Mo13) were clearly identified as *Exophiala* sp. and the eight others were close and still belonging to the *Exophiala* clade. Based on their 18S rRNA genes, the closest neighbor for members of Hypocreales were *Paecilomyces lilacinus* (MV3, MV19 and MV24) and *Lecanicillium lecanii* (MV9 and MV13). However, five strains (MV10, Ex7, Ex8, MV23 and MV25) were separated from available reference sequences of known species. Eight strains belonged to Coniochaetales and were close to *Lecythophora* genus. Interestingly, four of those strains (MV8, MVFS1, MVFS4 and Mo9) were isolated from the deep-sea mussel *Bathymodiolus azoricus* and the deep-sea shrimp *Rimicaris exoculata* and the four others (Ex3 to Ex6) from an ocean surface alga (*Sargassum fluitans*). Six strains were affiliated to the Eurotiales order. Five were close to *Aspergillus* sp. (Ex1, Ex9, Mo4, Mo6 and Mo7). ITS and 28S rRNA gene sequence analyses has permitted to clearly identify *Aspergillus sydowii* (100% similarity). Strain Ex2 was clearly identified as *Eurotium herbariorum* (100% similarity). Four strains (Mo1, Mo2, Mo10 et Mo11) were close to *Aureobasidium pullulans* (99,8% similarity). Those strains were isolated during MoMARDREAM-Naut cruise at Rainbow site from rocks scrapping and *Rimicaris exoculata*. Two strains isolated from *Rimicaris exoculata* (Mo3 and Mo5) were clearly identified as *Geomyces pannorum* (100% similarity). Representatives of

Dothideomycetes (HE2 and HE4), Xylariales (HE7), Orbiliales (HE1), Capnodiales (Mo8) were also observed in our study.

Based on SSU rRNA genes, isolates affiliated to Capnodiales (Mo8), Dothideomycetes et Chaetothyriomycetes incertae sedis (Mo1, Mo2, Mo10, Mo11) and Eurotiales (Ex2, Ex1, Ex9, Mo4, Mo6, Mo7) clusters were grouped close to environmental clones obtained from deep-sea hydrothermal ecosystems with culture-independent methods. Mo8 grouped closely (99,15% - 8 different nucleotides on 946bp) to one clone detected in the sulfide and hydrocarbon-rich hydrothermal sediment of the Guaymas Basin at 2000 meter deep in the Gulf of California (AY046698). Mo1, Mo2, Mo10 and Mo11 close to *Aureobasidium pullulans* were phylogenetically associated (100% similarity) to clones harvested in the fluid-seawater mix samples processed on Mid-Atlantic Ridge in Lost-City site at 750-900 meters deep and phylogenetically close (99,74% - 3 different nucleotides on 1160bp) to clones obtained near Bismarck wreck at 4790 meter deep (EU154991). Ex2 identified as *Eurotium herbariorum* grouped close (99,60% - 5 different nucleotides on 1257bp) to one clone isolated from 3 different deep-sea samples: near Bismarck wrecks and in the drake passage water column at 2000-3000 meter and 200-500 meter deep (EU154986) and also phylogenetically close (99,43% - 6 nucleotide differences on 1054bp) to one clone harvested in sediment of the Guyamas Basin (AY046710). Our *Aspergillus* strains (Ex1, Ex9, Mo4, Mo6, Mo7) were also close (98,6% - 16 differences on 1147bp) to the later (AY046710).

18S rRNA sequences indicated a high heterogeneity of our isolates among Ascomycota that grouped within thirteen different orders. Several isolates have no similar sequences in the GenBank database. This suggests that novel species/genera were isolated even if we have to consider that fungal 18S rRNA sequence databases could be incomplete.

## **Discussion**

### **Risk of contamination**

When scientists are trying to describe fungal diversity in an environment, they must consider that fungal contaminants could be isolated. But first, definition of a fungal contaminant has to be discussed. Two types of contaminant could occur in deep-sea microbiology depending on contamination origin: i) anthropogenic contaminant appearing during ROV or submarine sampling or isolation on oceanographic vessel (fungal airborne spores for example); ii) terrigenous contaminant occurring in deep-sea vents due to natural phenomena as oceanic

currents or sedimentation. These last contaminants are circulators that could survive in deep-sea vents or adapt to hydrothermal conditions.

Deep-sea hydrothermal vents are complex ecosystems, not only on a biological point of view but also for sampling which is expensive and consequently scarce. This situation limits the use of complex experimental designs to study the risk of contamination. Contamination can occur and have been visualized only once during EXOMAR cruise due to a default of the laminary flow-hood. Control plates led to the obtention of *Penicillium* strains and consequently *Penicillium* strains isolated during EXOMAR with the same morphology have been eliminated. Strains considered as marine are those that have grown directly from hydrothermal substrates. Considering our results, we can observe that 80% of the samples processed were sterile for fungi on the high amount of culture media processed. No fungi were isolated at one hydrothermal site (Lucky Strike) although same methods were employed for sampling. Our chi-square results have clearly indicated that our isolates were very highly dependent on hydrothermal sites and on type of samples processed. Finally, for 5 clusters, encompassing 38 filamentous fungi isolates (on our 62 strains), samples that led to isolation were processed at different dates during 2 or even 3 different oceanographic cruises. Same strains isolated in the same place after several years strongly support that those fungi were not anthropogenic contaminants or circulators and had an ecological role at deep-sea ecosystems. All our results suggest that sampling methods and materials used were not a root of anthropogenic contaminations.

### **Occurrence in hydrothermal vents**

To understand the importance of fungi at hydrothermal vents, we needed first to assess their occurrence in these ecosystems. Considering that fungal DNA signatures were previously reported in hydrothermal studies, our goal was to know if fungal strains from such locations were culturable. We succeed in isolating fungi from a relatively low part of the samples collected (20%). Such a scarce degree of presence may indicate that fungi were not widely distributed in deep-sea hydrothermal vents. This statement appears in agreement with previous culture-independent studies that harvested only 18 deep-sea fungal sequences from 239 clones (Bass *et al.*, 2007) and only one fungus type from several deep-sea hydrothermal mussels of South-Pacific sites (Van Dover *et al.*, 2007). Statistical distribution tests led to visualization of an aggregate distribution in all sites studied with sufficient numbers of isolates. This is concordant with previous study (Van Dover *et al.*, 2007) and could explain

this moderate fungal presence. Such a degree of presence may also reflect a slow fungal colonization process due to deep-sea conditions. This hypothesis is in agreement with a previous study that did not recover any fungal sequences from a 15 days colonization module deployment when significant numbers of sequences were obtained from Titanic and Bismarck wreck areas (Bass *et al.*, 2007). Another hypothesis to explain this relatively low isolation rate is the use of culture media rich in organic compounds which may be selected fungal strains associated with animal tissues, an evident source of organic matter. Oligotrophic strains were likely ignored because of the rich culture media chosen. Looking for oligotrophic fungal strains in the deep-sea should be done in the future using nutrient-free, low-nutrient silica gels or other solidifying agents commonly used for soil samples (Wainwright, 2005).

We have isolated filamentous fungi from all sites sampled excepted Lau Basin (frozen samples) and Lucky Strike (fresh samples). A previous study reported prevalence higher than 58% of fungal epizootic diseases on mussels of Mussel Hill hydrothermal site in the North Fiji Basin (Van Dover *et al.*, 2007). Our negative results are in agreement with this work that simultaneously did not recover any fungal infected samples from Lau Basin. Consequently, our absence of isolation for Lau Basin do not seem abnormal even if we have to consider that frozen samples may have had a negative impact on culturability. Absence of fungal isolates at Lucky Strike appears more surprising because clones were harvested in a previous study (López-García *et al.*, 2007). It could be the result of a too low sampling effort on this site (only 6 samples) and/or because of the presence of uncultivated strains.

Only one fungal strain was isolated from sediment indicating a low abundance of culturable fungi in this kind of substrate. Other sediment samples from Guaymas Basin were studied for fungi (data not shown). Using 35 frozen sediment samples from 0 to 40 cm depth for cultivation, no fungi were isolated. Those data are concordant with previous studies on deep-sea sediments that led to obtention of only eight fungal sequences from 515 clones (Edgcomb *et al.*, 2002; Bass *et al.*, 2007).

Our first results appear encouraging since more than sixty filamentous fungal strains were isolated. Filamentous fungi appear associated to animal samples as mussels and shrimps (when considering absolute frequencies). Filamentous fungi were associated to mussels especially inside animal (from flesh) and outside (from shell and byssi). Further studies should be useful to assess their role as symbionts, parasites or saprotrophs.

## Endemic vs Ubiquist

Successful isolation of culturable fungi led to the outcoming question whether strains were endemic (autochthonous) or ubiquitous (allochthonous). Considering growth depending on temperature and salinity, our culture collection was divided in two groups. Firstly, psychrotrophs that could be common deep-sea marine fungi, adapted to low temperatures conditions in such environments. We can hypothesize that those psychrotrophs are real deep-sea endemic strains. Secondly, mesophiles that did not show specific adaptation to live in dominant cold deep-sea conditions. But, considering their physiology, those strains were able to survive and maybe to use specific niches at deep-sea vents. So, we can here hypothesize that those mesophiles could be ubiquitous fungi present at hydrothermal vents due to sedimentation or currents. Those hypotheses are in agreement with the cold *in situ* temperatures measured in shrimp swarms at Rainbow site:  $13.2 \pm 5.5$  °C (Desbruyères *et al.*, 2001) and 3.8–14.7 °C (Zbinden *et al.*, 2004) and in *Bathymodiolus azoricus* mussel beds:  $6.0 \pm 3.9$  °C at Rainbow and  $10.1 \pm 0.5$  °C at Menez Gwen (Desbruyères *et al.*, 2001).

Some strains have shown difficulties to grow under our cultivation conditions that were not necessarily optimal. Hydrostatic pressure which effect was not studied in this study could be an important parameter to control for their growth. Indeed, although previous works have dealt with the effective growth of marine fungi (Raghukumar & Raghukumar, 1998; Damare *et al.*, 2006) and marine yeasts (Lorenz & Molitoris, 1997) under elevated hydrostatic pressure, no true piezophilic fungi have been reported to date.

Adaptation or natural selection of fungal strains to deep-sea vent conditions may occur. Eight filamentous fungi belonging to *Lecythophora* genus were isolated from different substrates: 4 strains were isolated from *Sargassum fluitans* from the ocean surface used as control and 4 from *Bathymodiolus azoricus* shell and flesh and *Rimicaris exoculata* (Table 10). PCA and crude values from physiological analyses indicated that the 4 strains from searface were mesophiles (not able to grow at 5°C) and the 4 strains from vents were psychrotrophs and therefore able to grow at 5°C (Figure 51). Considering the Raghukumar hypothesis (see next part), sinking organic materials might bring fungal strains in deep-sea that could adapt to cold deep-sea conditions.

## Diversity and potential role in deep-sea ecosystems

Our cultures showed the presence of five strains (Ex1, Ex9, Mo4, Mo6 and Mo7) belonging to *Aspergillus* sp (Eurotiales order) isolated during two different oceanographic cruises (EXOMAR and MoMARDREAM-Naut) from Rainbow and Lost City sites. Complete analyses based on rRNA Internal Transcribed Spacer (ITS1-5.8S-ITS2) sequencing revealed the species *A. sydowii*. This species is a common terrestrial fungus isolated from diverse habitats and causing an epizooty among sea fan corals (Alker *et al.*, 2001). Recent study reported *A. sydowii* in deep-sea sediments (Raghukumar *et al.*, 2004). Our five strains were described as mesophiles in physiological analysis (Figure 51) and so not specifically adapted to usual deep-sea marine conditions. A first possibility could be that these ubiquitous strains are contaminants incoming from ocean surface or transported by currents. Their presence in marine ecosystems is congruent with the fact that all *Aspergillus* isolated were halophiles. But, in our case, these five strains were isolated at 2 different dates, persons, oceanographic cruises and sites. Moreover, our strains grouped close to environmental sequences from deep-sea environments: AY046710 (Edgcomb *et al.*, 2002) and EU154986 (Bass *et al.*, 2007). These data suggested that our *Aspergillus* strains were naturally existent at deep-sea. *Aspergillus* may have a role in deep-sea environments even if they are not indigenous and even if they are not in optimal growth conditions. Occurrence of fungi in deep-sea sediments from the Chagos Trench (Indian Ocean) at 5000 m depth was reported (Raghukumar *et al.*, 2004; Damare *et al.*, 2006; Damare & Raghukumar, 2007). Their study yielded almost exclusively the mitosporic fungus *Aspergillus sydowii* present at core depths of 160 to 370 cm. Spores of *A. sydowii* from their core samples germinated and grew at elevated hydrostatic pressures and low temperatures. They hypothesized that “*such spores might eventually sink to the deep-sea surficial sediments, undergo natural selection mechanisms with time and acquire capabilities to grow and multiply in the presence of suitable nutrient sources*” (Raghukumar *et al.*, 2004). Our paper reveals the first presence of *A. sydowii* at deep-sea hydrothermal vents where such microorganism may grow and reproduce thanks to temperature gradients and to high and diverse nutrient sources in this particular ecosystem.

Other mesophilic strains were isolated (MV19 and MV24) and clearly identified as *Paecilomyces lilacinus*. They were isolated from *Bathymodiolus azoricus* samples. Their real development in deep-sea is less evident because those strains were clearly non-halophiles.

One strain in our collection (Ex2) was strictly halophile and characterized as *Eurotium herbariorum*. Our results from cultures were well correlated with those from culture-independent methods: AY046710 (Edgcomb *et al.*, 2002) and EU154986 (Bass *et al.*, 2007).

Sequences close to our organism have been detected in several deep-sea environments and also in the water column. Moreover, *E. herbariorum* have been isolated in other extreme environments: hypersaline waters of solar salterns (Butinar *et al.*, 2005) and the Dead Sea waters (Jin *et al.*, 2005). In these environments, *E. herbariorum* was described as the most common species isolated and thus characterized as indigenous. In our case, Ex2 was isolated from a sediment core sample where this strain may have a role in decomposition of organic matter in deep-sea floor.

Helotiales order gathered 14 strains in our collection that were isolated during different oceanographic cruises. Twelve strains were identified as *Cadophora malorum* and described physiologically as psychrotrophs and also mainly halophiles (except MVFS3 that was non-halophile). These last twelve strains corresponding to a single fungal species (identical 18S, ITS and 28S rRNA sequences) were isolated from two different hydrothermal sites: Menez-Gwen (6) and Rainbow (6) and thus support the ecological significance of this species in the deep-sea vents. *Cadophora malorum* has been reported largely in Antarctica on mosses (Azmi & Seppelt, 1998; Tosi *et al.*, 2002) that demonstrated their wide distribution on the Antarctic continent. A recent study has physiologically demonstrated that *Cadophora* sp strains isolated in Antarctic environments were psychrotrophs. Scientists have speculated that the *Cadophora* sp studied (containing *C. malorum*) were endemic in Antarctica (Duncan, 2007). *Geomyces pannorum* was also frequently isolated as psychrotroph in Antarctic soils and speculated as indigenous in Antarctica (Duncan, 2007). Moreover, *G. pannorum* was isolated from Arctic cryopegs and marine deposits (Kochkina *et al.*, 2007). Two of our strains were clearly identified as *G. pannorum* and were characterized as psychrotrophs and also halotolerant. Such psychrotrophic strains from our collection harvested in cold deep-sea vents could be the result of exchanges with polar environments. This hypothesis is concordant with previous study on deep-sea and Antarctic isolates of *Psychrobacter* reporting that these deep-sea microorganisms probably originate in polar regions and disseminate with the global deep ocean circulation (Maruyama *et al.*, 2000). Eleven *C. malorum* from our collection were isolated from *Bathymodiolus azoricus*: 10 from mussels and 1 from a *Polynoidae* scale-worm associated to one mussel sampled. Some scale-worms seem to be able to leave their host mollusks to infest new ones (Britayev *et al.*, 2007). Thus, fungi may be disseminated by scale-worms in mussel communities.



Occurrence of fungi inducing diseases in deep-sea vents animals has been reported only once. An emergent epizootic event was reported in the deep-sea mussels *Bathymodiolus brevior* at the Mussel Hill hydrothermal vent in Fiji Basin in which fungi were the causal agent (Van Dover *et al.*, 2007). A *Capronia*-like fungus (order Chaetothyriales) elicited a host immune response in mussels and was associated with tissue deterioration (using histological observations, genetic analyses, Fluorescent *In Situ* Hybridization technique but not culture-dependent methods). Eleven isolates of our collection were also affiliated to the Chaetothyriales order. Ten strains were characterized as low growing strains and non-halophiles and only one was psychrotroph and halophile (MV26). They were mostly isolated from *Bathymodiolus azoricus* mussels (>80%). This set of results suggests that fungi of the Chaetothyriales order were predominant in deep-sea mussels of different species (*B. brevior* and *B. azoricus*) and of different locations (South West Pacific and Mid-Atlantic Ridge sites) and can sometimes be associated with pathologic situation.

Our results proved that those fungi isolated from deep-sea mussels are not symbionts because a large amount of living samples were processed and only a few led to the visualization of a fungal presence. Our mussel samples looked apparently healthy with a normal creamy coloration without necrosis or lesions and confirmed the Van Dover hypothesis that “*identification of fungi in otherwise healthy individuals makes it clear that the fungus is not a strict saprophyte*” (Van Dover *et al.*, 2007). It suggests that fungi may be facultative parasites or opportunistic pathogens of deep-sea animals. This hypothesis is consistent with the fact that we only processed living animal samples. That could support the hypothesis that hydrothermal vents host a unique array of animal parasites (Moreira & López-Garcia, 2003). Four strains isolated during MoMARDREAM cruise at Rainbow site and identified as *Aureobasidium pullulans* were observed for the first time in hydrothermal vents. This yeast-like fungus was previously isolated from deep-sea sediments of the Pacific Ocean (Li *et al.*, 2007). These strains were all psychrotrophic. Three isolates (Mo2, Mo10 and Mo11) were obtained from *Rimicaris exoculata* and one from a chimney rock scrapping (Mo1). Food intake of *R. exoculata* is composed of rock particles (Segonzac *et al.*, 1993). Our study may indicate that a transfer could occur between fungi present on chimney rock and organisms scrapping it. Those culture-dependent results are concordant with previous culture-independent works. Indeed, our strains sequences match with environmental sequences obtained from deep-sea environments: EU154991 and DQ504331 (See Figure 52). Even if those strains were characterized as non-halophiles, these results indicated that *A. pullulans* is

frequently observed in Atlantic deep-sea environments at different depths, from wrecks (4790 meters) to hydrothermal sites (750-900 and 2300 meters).

Occurrence of fungi in healthy and diseased corals has been widely reported in shallow waters. Fungi are described as skeletal-components of healthy, partially dead and diseased corals (Ravindran *et al.*, 2001). One fungal parasite was found to cause necrotic patches on five different corals (*Porites lutea*, *Porites lichen*, *Montipora tuberculosa*, *Goniopora* sp and *Goniastra* sp). Infections were due to *Scolecobasidium* sp (Raghukumar & Raghukumar, 1991). The same genus was found in deep-sea corals (Mo19), near Rainbow hydrothermal vent site (Figure 52), indicating that this halophilic fungus also occurred in corals at several thousand meters depth and could be implicated in coral diseases. This is the first description of such an association in the whole deep-sea.

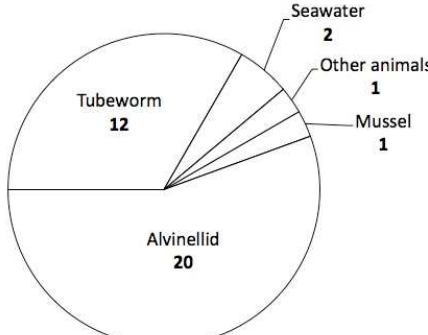
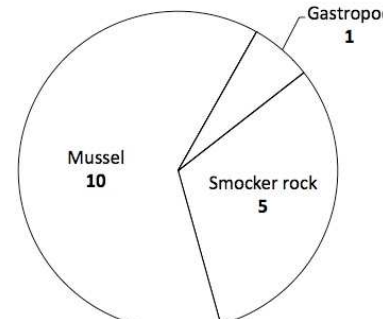
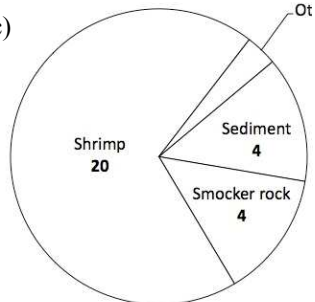
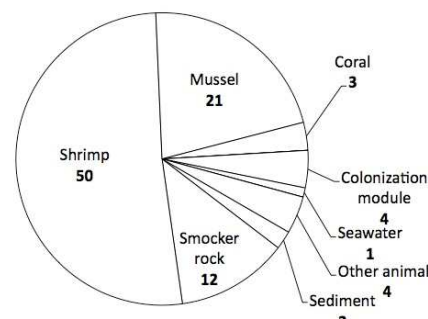
HE7 was characterised as an halotolerant fast growing fungus phylogenetically close to *Eutypa lata*, a genus mostly isolated from mangrove woods (Alias *et al.*, 1995; Sarma *et al.*, 2001). HE2 and HE4 were also halotolerant and close to *Dothideomycete* sp., belonging to an ascomycetous class mainly composed of plant pathogens. HE6 was non-halophile and close to *Tilletiopsis pallescens*, also described as plant pathogen. Wood fungi recovered in deep-sea hydrothermal vents may be explained by sunken wood presence in deep-sea (Palacios *et al.*, 2006).

**Our study revealed a non-exhaustive culturable fungal diversity until now never harvested in deep-sea environments. The preventive measures taken (sterile sampling equipments, laminary flow-hood on board...) permit to control contamination. Moreover, physiological analyses confirmed that fungi isolated from several deep-sea vents mostly gathered likely indigenous species. Further works concerning the effect of hydrostatic pressure on growth (use of pressure vessels), metabolic activities and the use of diverse media are needed to assess more their adaptation to deep-sea conditions. *In situ* observations could also permit to approach the ecological importance of such organisms and to know whether they form dense or scarce communities.**

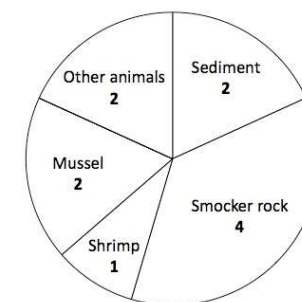
## **Acknowledgements**

We thank the chief scientists of the BIOLAU, DIVANAUT2, HERO, MARVEL, EXOMAR and MoMARDREAM-Naut cruises, pilots and support crews of oceanographic vessels and Deep Submergence Vehicles of Ifremer. We greatly acknowledge Dr. Jean-luc Jany for valuable help and comments on phylogenetical analyses. We thank all members of the GDR Ecchis for discussions and suggestions. We also thank ANR Deep-Oases and French Research Ministry for financial support. We finally thank Françoise Gail and the CHEMECO / European Science Foundation EURODEEP for discussions and financial support.

**Table 10: Culture collection of filamentous fungi from deep-sea hydrothermal vents.**

LOCATION (Depth)	STRAIN (Culture medium)	SAMPLE PROCESSED (Type and Number)	
<u>EPR; 13°N; Elsa site (-2630m)</u>			
H18E11: <i>Alvinella pompejana</i> tube	HE1, HE2 and HE3 (Sab)		
H18E12: <i>A. pompejana</i> with bacterial mat	HE4 (Sab)		
H18E09: Seawater surrounding <i>A. pompejana</i>	HE5 (Sab)		
H13E04: <i>Riftia pachyptila</i> branchial plume	HE6 (GYPS)		
H18E01: <i>A. pompejana</i> tube	HE7 (Sab)		
<u>MAR; Menez-Gwen (-860m)</u>			
DV14E01: <i>Bathymodiolus azoricus</i> flesh	DIVA1 (GYPS)		
MV2E01: External face of a shell of <i>B. azoricus</i>	MV1 (Ac), MV2 (Lc), MV3 (Xc), MV4 (Xc), MV21 (Cc), MV22 (Pc), MV23 (Pc), MVFS1 (GYPS), MVFS2 (GYPS) and MVFS3 (GYPS)		
MV2E02: <i>B. azoricus</i> tissue	MV6 (Ac), MV7 to MV12 (Cc), MV13 (Pc), MV14 (Xc), MV24 (Pc), MV25 (Pc) and MVFS4 (GYPS)		
MV2E03: <i>B. azoricus</i> mineral substrate	MV15 (Ac), MV16 (Cc), MV17 (Lc), MV18 (Pc), MV19 (Xc), MV26 (Cc) and MV27 (Xc).		
<u>MAR; TAG (-3650m)</u>			
Ex13E04: <i>Rimicaris exoculata</i>	Ex7 (GYPS)		
<u>MAR; Rainbow (-2300m)</u>			
Ex6E01: <i>Rimicaris exoculata</i>	Ex1 (GYPS)		
Ex8E02: Sediment	Ex2 (GYPS)		

MoPR1: Smoker rocks	Mo1 (GYPS)
MoPR2 to MoPR5: <i>Rimicaris exoculata</i>	Mo2 to Mo13 (GYPS)
MoPR6: <i>Bathymodiolus azoricus</i> byssus	Mo14, Mo15 and Mo18 (GYPS)
MoPR6: <i>Bathymodiolus azoricus</i> tissue	Mo16 (GYPS)
MoPR6: Commensal worm in <i>Bathymodiolus azoricus</i>	Mo17 (GYPS)
MoPR9: Coral	Mo19 (GYPS)
<u>MAR: Lost City (-700m)</u>	
MV5E03: <i>Bathymodiolus azoricus</i> tissue	MV20 (Lc)
Ex18E01: Siliceous sponge	Ex8 and Ex9 (GYPS)

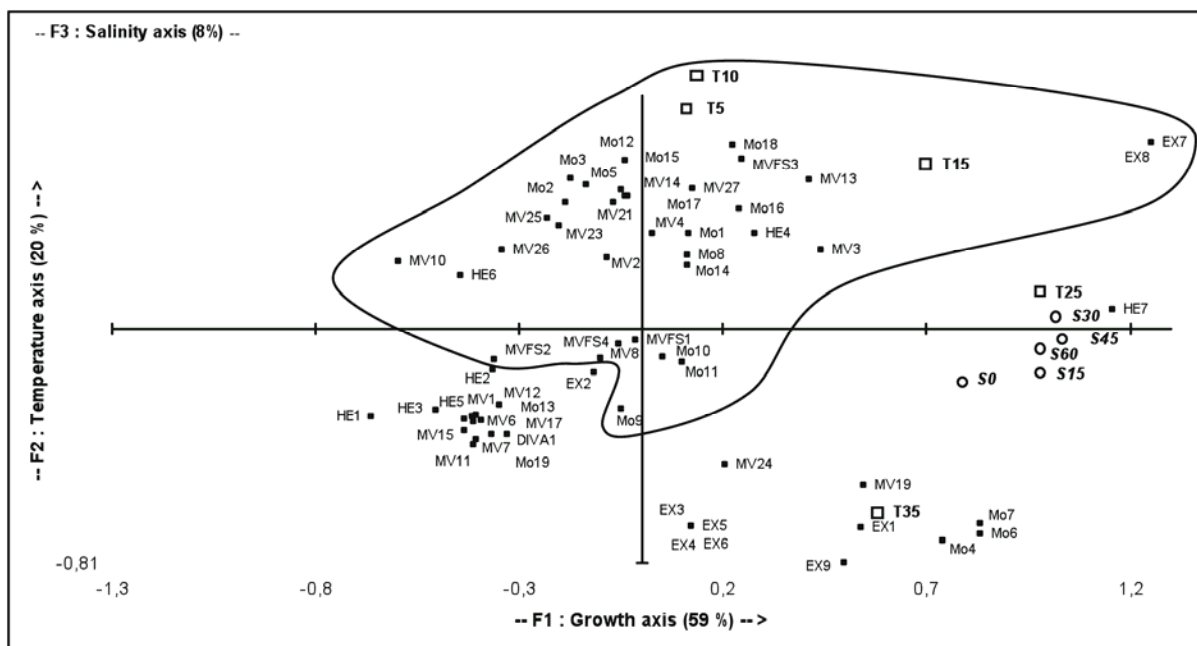



---

MAR (Mid-Atlantic Ridge); EPR (East Pacific Rise); H (HERO); DV (DIVANAUT2); MV (MARVEL); Ex (EXOMAR) and Mo (MoMARDREAM-Naut). Ac (Potato starch); Cc (Cellobiose); Gc (Glucose); Lc (Arabic gum); Xc (Xylan); Pc (Brain heart); GYPS (Glucose – Yeast extract – Peptone – Starch) and Sab (Sabouraud).

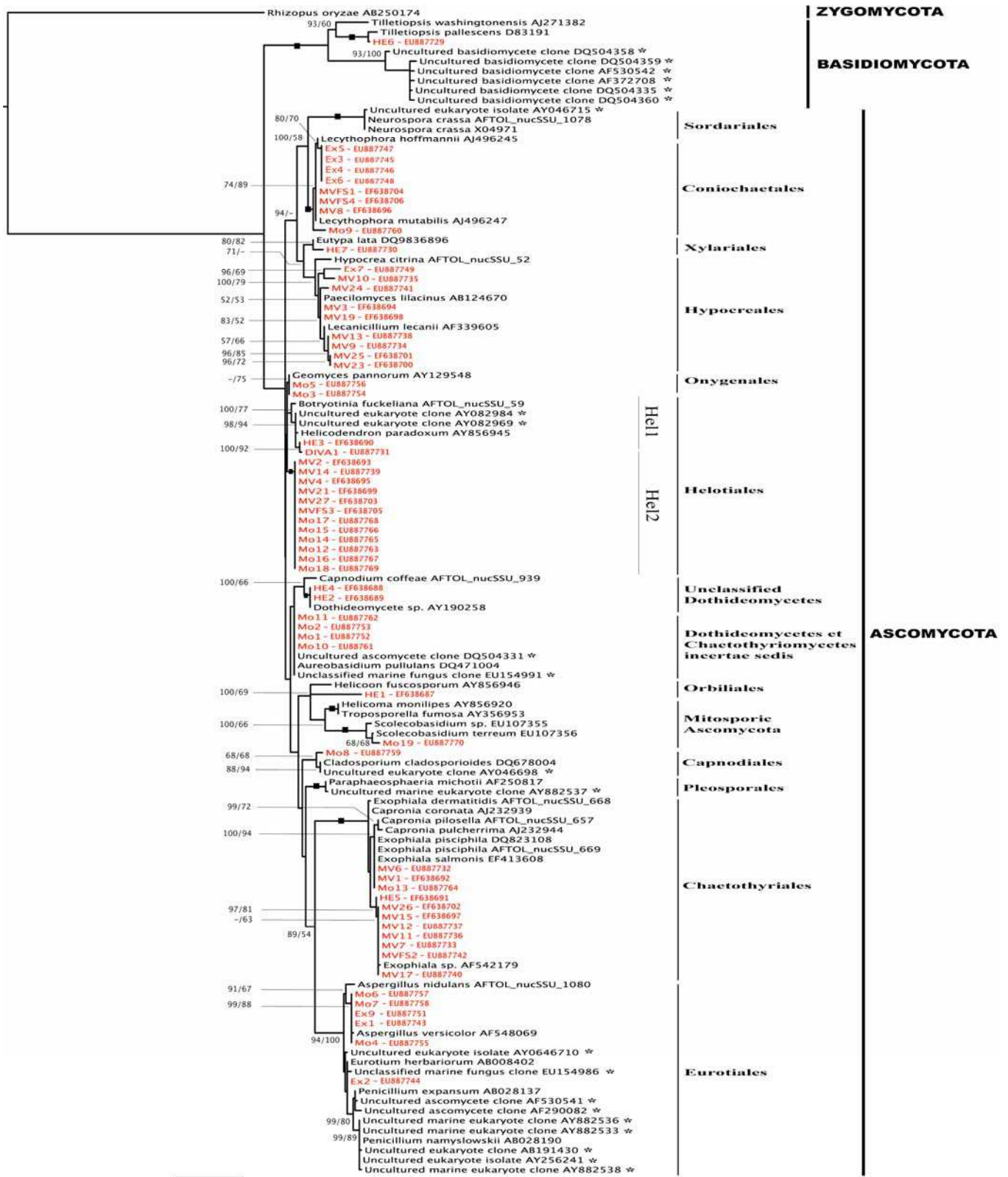
Six samples of seaweed (*Sargassum fluitans*) collected at the ocean surface of the TAG site were processed as control and yielded to the isolation of filamentous fungi (Ex3 to Ex6). Twelve strains (HE1, HE2, HE3, HE4, HE5, HE6, HE7, MVFS1, MVFS2, MVFS3, MVFS4 and DIVA1) were isolated from frozen samples.

**Figure 51: Physiological analysis of the fungal collection.** This figure was obtained from Principal Component Analysis of fungal strains colony diameters (individuals) at different temperatures and salinities (variables). Mycelial colony diameters were measured after 14 days of incubation on GYPS medium depending on growth conditions (4 temperatures at 3% sea salts and 5 salinities at strain optimal temperatures). Strains encircled were psychrotrophs (able to grow at 5°C) and others were mesophiles. Filled squares, empty squares and empty circles correspond respectively to the strains, the temperature and the salinity parameters.



**Figure 52**: SSU rRNA phylogenetic positions of deep-sea fungal isolates (colored terminals). Topology was built using MrBayes v.3.1.2 from a ClustalW 1.83 alignment. Node support values are given in the following order: MrBayes posterior probabilities / PHYML 100 bootstraps. Black squares represent nodes supported by an excess of 0.95 posterior probabilities and 95% bootstraps. Included environmental sequences from deep-sea clones are marked with an asterisk. *Rhizopus oryzae* (AB250174) belonging to the Zygomycota phylum was used as outgroup. All sequences are listed with their GenBank accession numbers.

Fig.52







## **Discussion générale et perspectives de recherche**

---

Le premier objectif de ce travail de thèse a été d'apporter des connaissances supplémentaires à un sujet très peu exploité jusque là, à savoir la diversité des champignons en écosystème hydrothermal marin profond. Le second objectif était de comprendre les fonctions écologiques de tels organismes.

La découverte de ces écosystèmes est relativement récente (Corliss *et al.*, 1979), et de nombreuses espèces jamais décrites (d'eucaryotes supérieurs, de bactéries et d'archaea principalement) sont découvertes très régulièrement. L'intérêt porté à ces espèces relève surtout du fait que celles-ci doivent développer des adaptations pour vivre dans des conditions physico-chimiques agressives.

Ce n'est que très récemment que des premières traces de vie fongique ont été associées à ces écosystèmes (Edgcomb *et al.*, 2002 ; López-García *et al.*, 2003). Dès lors, l'analyse ciblée des champignons a permis d'en isoler une fraction (Bass *et al.*, 2007 ; Van Dover *et al.*, 2007 ; Burgaud *et al.*, 2009). Cependant, les connaissances sur la fraction fongique de ces habitats restent limitées, ceci étant renforcé par le fait que les champignons sont considérés pour beaucoup comme des organismes inféodés aux écosystèmes terrestres.

Cette partie discutera des principaux résultats que nous avons obtenus lors de cette thèse, puis des perspectives de recherche pouvant être menées.

## **I- Diversité des champignons en écosystème hydrothermal**

### **Approche indépendante de la culture**

Nous avons choisi une approche moléculaire indépendante de la culture pour améliorer la connaissance de la diversité spécifique des champignons dans ces écosystèmes. Cette analyse a été effectuée en amplifiant directement à partir des échantillons environnementaux les gènes codant l'ARNr 18S spécifiques aux champignons, puis en réalisant des analyses phylogénétiques nous permettant de déterminer les organismes présents.

Nous avons créé dans ce but une base de données dédiée à ce type de travail, PHYMYCO-DB, regroupant des marqueurs phylogénétiques fongiques extraits de Genbank. Ce travail a été effectué en collaboration avec l'IRISA (Rennes, France). Après récupération automatique des données, puis d'un « nettoyage » automatique et manuel des séquences de mauvaise qualité pouvant biaiser les analyses ultérieures, la base de données contenait 6978 séquences de gènes codant l'ARNr18S et 431 séquences de gènes codant EF1- $\alpha$ .

Une fois la base de données mise au point, nous avons alors effectué les analyses de diversité fongique sur un large panel d'échantillons hydrothermaux. Au terme de ces analyses, nous avons pu révéler la présence de 13 phylotypes fongiques, appartenant aux phyla Ascomycota, Basidiomycota et Chytridiomycota. De façon intéressante, certains phylotypes ont été détectés dans des environnements hydrothermaux lors de précédentes études (Takishita *et al.*, 2006 ; López-García *et al.*, 2007 ; Bass *et al.*, 2007), ou plus largement dans l'océan profond pour d'autres phylotypes (Gadanhó & Sampaio 2005; Nagahama *et al.*, 2001). Ceci tend à prouver que ces habitats sont assez largement colonisés par les champignons et constitue aussi un argument pour rejeter une origine contaminante des organismes étudiés.

### **Approche culturelle**

En parallèle de cette approche indépendante de la culture, nous avons choisi d'isoler les souches fongiques à partir des mêmes échantillons par isolement à pression atmosphérique. Ce travail a été réalisé en collaboration avec l'ESMISAB (Brest, France). Cette approche a permis d'isoler 13 phylotypes supplémentaires, exclusivement dans le phylum Ascomycota. Étonnamment, aucun phylotype commun n'a pu être retrouvé par ces deux approches. Des études précédentes utilisant les mêmes jeux d'amorces que ceux que nous avons utilisé (Vandenkoornhuyse & Leyval, 1998) ont permis d'amplifier très largement des séquences du phylum Ascomycota, et les différences observées entre les études culturelles et environnementales ne doivent pas être liées à des biais d'amplification dus à ces amorces. L'une des explications possibles peut provenir du fait que les organismes du phylum Ascomycota isolés lors de cette étude s'adaptent plus facilement aux conditions de culture de laboratoire. L'adaptation à la pression atmosphérique pourrait ainsi constituer un biais important lors de la mise en culture. Cependant, il existe très peu d'études sur l'effet de la pression hydrostatique sur le développement et la croissance des champignons. Les mécanismes d'adaptation des bactéries aux pressions hydrostatiques sont d'ailleurs également mal connus et il a été démontré que des cultures de souches bactériennes isolées dans l'océan profond, dans des conditions de pression atmosphériques, entraînaient des biais dans les compositions des communautés bactériennes dominantes (Yanagibayashi *et al.*, 1998). On peut noter que des équipements (dérivés de ceux utilisés en bactériologie) permettant de cultiver des souches fongiques dans les conditions de pression hydrostatiques similaires à celles du domaine profond ont été mis au point et ont permis de maintenir en culture des levures à métabolisme fermentaire (Lorenz & Molitoris, 1997). D'autre part, compte tenu des

conditions chimiques fluctuantes du milieu hydrothermal, il peut s'avérer difficile de réunir les conditions optimales de culture en laboratoire. Les analyses complémentaires réalisées par l'ESMISAB ont permis au total d'isoler 98 souches fongiques à partir de nombreux échantillons (62 souches d'organismes filamenteux et 36 souches de levure). Seules les souches d'organismes filamenteux ont été étudiées plus en détail, notamment par des analyses physiologiques (cultures sous différentes conditions de températures et de salinité).

La majorité des souches isolées dans les conditions de laboratoire ne semblent pas être des halophiles stricts ; la question de l'endémisme de ces souches, discutée dans la partie VI de cette thèse, reste toutefois à élucider. Il apparaît difficile cependant par des approches culturelles d'aborder ces questions, les organismes isolés seront vraisemblablement ceux possédant les capacités adaptatives les plus développées : de tels organismes pourraient donc être ubiquistes, capables de s'adapter et de se développer dans différentes conditions physico-chimiques du milieu, à la fois dans des écosystèmes terrestres et marins. On peut supposer que des champignons isolés à plusieurs reprises, retrouvés dans différents sites hydrothermaux seraient les organismes effectivement inféodés à ces milieux. L'avantage certain de ces mises en culture est l'accès à des opportunités biotechnologiques.

### **Diversité globale**

En combinant les approches culturelles et environnementales, nous avons mis en évidence la présence de champignons répartis dans 15 ordres différents d'Ascomycota, 6 ordres de Basidiomycota, et un ordre de Chytridiomycota. Parmi ces organismes, les analyses phylogénétiques ont permis de révéler des champignons jamais décrits jusqu'à ce jour. Par ailleurs, aucun champignon appartenant aux Zygomycota, et Glomeromycota n'ont été retrouvés par nos analyses. Ce travail a donc permis de mettre en lumière une forte diversité de champignons en écosystème hydrothermal marin profond, contrairement aux précédentes observations concluant à une faible diversité fongique dans ces environnements (Bass *et al.*, 2007). Etant donné que de nombreux sites hydrothermaux restent à analyser en terme de diversité fongique, notre étude pionnière peut ouvrir la voie à une investigation plus approfondie de ce domaine peu étudié jusqu'à présent. Les écosystèmes hydrothermaux pourraient ainsi constituer un réservoir de biodiversité fongique, offrant l'espoir de découvrir de nombreux champignons originaux. Ceci nous permettrait de renforcer les connaissances sur la diversité globale des champignons sur Terre et potentiellement de valoriser cette diversité à des termes biotechnologiques.

### Explorer la diversité fongique : approche biotechnologique.

L'océan, et plus particulièrement l'océan profond, sont reconnus comme des sources de découverte de nouveaux organismes aux capacités biotechnologiques importantes (Bull *et al.*, 2000). Cependant, l'un des obstacles à contourner vient des difficultés à maintenir en culture ces organismes. Malgré cela, l'océan a d'ores et déjà permis d'isoler un nombre de molécules important, principalement exploitées par l'industrie pharmaceutique, mais également par d'autres industries comme l'agroalimentaire (Bull *et al.*, 2000 ; Marris, 2006). En 2000 par exemple, les produits de la biotechnologie marine ont engendré des ventes estimées à 100 milliards \$ US (Convention sur la Diversité Biologique, 2005). Les microorganismes de l'océan profond se révèlent ainsi être une source de molécules naturelles très importantes (Marris, 2006). Le tableau 11 présente les principaux habitats de l'océan profond, leurs caractéristiques principales et les applications biotechnologiques majeures pouvant être exploitées (Bull *et al.*, 2000). Ces différents milieux se révèlent particulièrement intéressants car les organismes doivent y développer des adaptations spécifiques aux conditions extrêmes liées à l'habitat. Les écosystèmes hydrothermaux marins offrent donc de grands espoirs dans ce domaine et les propriétés des microorganismes (bactéries et archaea) résidants ont déjà été valorisées par de nombreuses applications industrielles (Bull *et al.*, 2000).

Habitat	Caractéristiques principales	Opportunités biotechnologiques principales
Fosses océaniques	Forte pression hydrostatique	Catalyseurs biologiques, nouveaux ou améliorés
Abysses, suintements froids	Faible température	Catalyseurs biologiques, bioremédiations, surfactants
Eau de mer	Faible concentration en nutriments	Catalyseurs et ligands à forte affinité
Events hydrothermaux	Forte température, métaux	Catalyseurs biologiques thermostables
Sédiments, épibioses, symbioses	Forte concentration en nutriments Mécanismes de défense	Composés bioactifs nouveaux impliqués dans la signalisation cellulaire, la défense

Tableau 11: Les principaux habitats de l'océan profond, leurs caractéristiques et les principales applications biotechnologiques associées. D'après Bull *et al.*, 2000. *Microbiology and Molecular Biology Reviews*. 64: 573-606.

Les champignons en écosystème hydrothermal marin se révéleraient donc être une source potentielle majeure de molécules originales. A l'heure actuelle, des recherches de composés bioactifs isolées à partir de champignons marins ont été entreprises avec succès : les champignons endophytes (d'algues), ou en associations symbiotiques avec des animaux (mollusques, éponges ou coraux) semblent les cibles privilégiées pour ces études (Raghukumar, 2008). Dès à présent, des molécules antibiotiques et antitumorales par exemple, ont ainsi pu être isolées (Bugni & Ireland, 2004). Les environnements extrêmes, et notamment l'océan profond, ont également été prospectés depuis peu dans un but de valorisation biotechnologique fongique (ou mycotechnologie) (Raghukumar, 2008). L'océan profond se révèle particulièrement intéressant car les champignons doivent s'adapter à la fois à une forte concentration en ions sodium, à un milieu oligotrophe, ainsi qu'à une forte pression hydrostatique (Raghukumar, 2008). L'espoir de valoriser la diversité fongique en écosystème hydrothermal est donc réel. Pour cela cependant, le maintien en culture des souches isolées est déterminant mais comme nous l'avons vu précédemment dans notre étude, cette étape culturale semble avoir apporté un biais dans la composition de la communauté fongique. Des efforts supplémentaires doivent donc être apportés pour améliorer cette étape. Une prospection préalable plus large par approche moléculaire pourrait donc permettre de cibler les organismes présents et ainsi de mieux contrôler les conditions de mises en culture futures. Des recherches d'activités antibiotiques sont d'ores et déjà en cours à l'ESMISAB sur la collection de champignons hydrothermaux.

### **Hypothèses sur l'évolution des champignons**

L'un des résultats majeurs de ce travail a été la découverte de deux phylotypes de Chytridiomycota, groupés dans une branche profonde du Règne fongique. Ces organismes appartiendraient ainsi à une lignée évolutive ancienne, sinon la plus ancienne, des champignons. Ces résultats sont en accord avec les analyses d'horloge moléculaire réalisées sur le jeu de données de Vandenkoornhuyse *et al.* (2002a) présentées dans la figure 2, qui ont été la source de nos hypothèses de travail et qui nous ont mené à nous intéresser à ces environnements. Cette analyse avait proposé une émergence des champignons il y a 1 milliard d'années. A cette époque, l'océan primitif partageait certaines caractéristiques communes avec le milieu hydrothermal contemporain (Robert & Chaussidon, 2006), considéré par certains auteurs comme le lieu d'émergence de la vie (Huber & Wächtershäuser, 1997 ; Huber *et al.*, 2003 ; Koonin & Martin 2005). Nous espérons ainsi trouver dans ces environnements des organismes aux caractères ancestraux. Certains auteurs considèrent d'ailleurs ces

écosystèmes comme des environnements contenant des espèces très anciennes d'un point de vue évolutif, qui constitueraient ainsi des formes de fossiles vivants (Newman, 1985). Néanmoins, le fait d'avoir rencontré des champignons, même anciens, ne saurait soutenir l'hypothèse berceau de la vie du milieu hydrothermal. Ces hypothèses sont cependant contredites par des études de fossiles d'eucaryotes supérieurs isolés de sédiments hydrothermaux et d'études d'horloges moléculaires qui tendent à montrer une radiation plus récente des groupes majeurs des invertébrés hydrothermaux (émergence au Cénozoïque ou Mésozoïque, durant les 100 dernières millions d'années) (Little & Vrijenhoek, 2003) : cette question reste toutefois à l'heure actuelle controversée.

Notre hypothèse de travail de départ impliquait une diversification des champignons dans l'océan, avant une colonisation de l'écosystème terrestre aérien. Lors de notre étude, nous avons trouvé une large diversité de champignons, répartie dans 3 phyla. Aucun organisme du phylum Zygomycota n'a été retrouvé. Ceci peut en partie s'expliquer d'une part par la faible occurrence de ces organismes dans les milieux aquatiques plus généralement, où seules quelques espèces parasites de rotifères ou d'algues ont pu être isolées (Benny *et al.*, 2001). D'autre part, les Zygomycota constituent l'un des groupes de champignons les moins étudiés (Benny *et al.*, 2001). Il serait alors intéressant d'étudier plus précisément la diversité de ces organismes dans les écosystèmes hydrothermaux (et plus largement en milieu marin), par la mise au point d'amorces spécifiques par exemple, dans l'espoir de mettre en lumière ou d'infirmer leur présence.

Pour l'instant, nous avons pu apporter des éléments de réponse en faveur de la diversification des champignons en écosystèmes marins, même si nous n'avons pas prouvé cette théorie. D'autres études complémentaires devront être réalisées et pourraient s'orienter selon 3 axes majeurs : le premier d'entre eux serait d'analyser plus précisément la fraction de Chytridiomycota, issue de l'échantillonnage d'un grand nombre de sites hydrothermaux. Etant donné la difficulté d'isoler de tels organismes en conditions de laboratoire, une approche moléculaire sera privilégiée, en amplifiant sélectivement les organismes de ce phylum. Des tentatives d'isolement et de mise en culture pourront également être réalisées afin de comprendre la diversité de ces organismes de façon plus exhaustive, mais également de mieux comprendre leurs modes de vie. Ceci nous permettra également d'apporter des éléments de réponse sur les théories évolutives que nous avons développées. Le second axe qui pourrait s'avérer particulièrement intéressant serait de tenter de détecter des séquences d'organismes du phylum Zygomycota, comme nous l'avons indiqué



précédemment, dans le but d'appuyer les théories d'une émergence mais surtout d'une diversification des groupes majeurs de champignons en milieu marin. Ceci est particulièrement vrai car ces organismes constituent un groupe ancien d'un point de vue évolutif, ayant vraisemblablement émergé après les organismes à la base du phylum Chytridiomycota (James *et al.*, 2006 ; Bar-Hen *et al.*, 2008). Finalement, le dernier axe de recherche devra s'orienter sur l'étude étendue de la composante fongique totale de ces environnements, pour avoir une vision globale de cette diversité.

## **II- Fonctions des champignons en écosystème hydrothermal**

### **Hypothèses sur les fonctions écologiques des champignons en écosystèmes hydrothermaux**

Malgré l'importance apportée à l'étude des écosystèmes marins profonds, et plus particulièrement à l'étude des écosystèmes hydrothermaux, de nombreuses questions sur le fonctionnement global de ces écosystèmes restent en suspens. Pour l'océan profond, récemment, le rôle des virus dans le cycle du recyclage de la matière organique, par une induction massive de mortalité procaryotique, a très récemment été prouvé (Danovaro *et al.*, 2008), confirmant la nécessité d'étendre les analyses à des organismes jusque là peu étudiés. Comme nous l'avons indiqué tout au long de cette thèse, les champignons hydrothermaux ont été très peu étudiés et à l'heure actuelle très peu de données sur leurs rôles écologiques dans ces écosystèmes sont disponibles ; pour l'instant seules des modes de vie parasite ont pu être prouvés (Van Dover *et al.*, 2007 ; Burgaud *et al.*, 2009) : les implications des levures noires de l'ordre des Chaetothyriales, ont été décrites dans des phénomènes de mortalité de modioles hydrothermales (Van Dover *et al.*, 2007). De telles levures ont également été isolées dans notre étude, à la fois par les approches environnementales et culturelles, associées aux modioles, ce qui pourrait confirmer le rôle pathogène de ces organismes. A l'heure actuelle, seules les modioles semblent être affectées par ces pathogènes, des études supplémentaires pourraient permettre de découvrir des champignons associés à des pathogénies d'autres invertébrés hydrothermaux. Les analyses moléculaires ont permis de détecter des séquences fongiques associées à d'autres animaux (crevettes et vers essentiellement), ce qui nous permet de supposer que le spectre d'hôte des champignons dans ces milieux est large, des analyses complémentaires pourraient confirmer cette hypothèse. Les pathogènes pourraient donc jouer un rôle important dans le cycle de la matière organique dans ces environnements. Notre étude

a également permis de découvrir des espèces au mode de vie parasite supposé: il s'agirait principalement d'espèces d'Ascomycota « mitosporiques ». Ceci est cohérent avec des observations suggérant que les écosystèmes hydrothermaux renferment une large diversité d'organismes parasites (Moreira & López-Garcia, 2003). Une approche complémentaire, afin de localiser les organismes fongiques dans leur environnement pourrait également être réalisée, par des techniques d'hybridation *in situ* (méthode Fish : Fluorescent *In Situ* Hybridization). Les données moléculaires que nous avons engendré lors de ce travail de thèse pourraient ainsi se révéler des bases solides à l'élaboration de sondes fluorescentes spécifiques aux champignons, afin d'apporter des connaissances supplémentaires sur l'écologie de ces organismes.

L'analyse métagénomique quant à elle, nous a permis de révéler des voies métaboliques d'un organisme divergent au sein des Chytridiomycota. L'analyse des séquences fongiques extraites du métagénome, ne nous a pas permis clairement de définir des fonctions écologiques, mais des hypothèses sur des interactions entre les champignons et leur environnement ont pu être établies. Un mode de vie parasite de l'organisme présent dans l'échantillon environnemental a pu être rejeté. Aucune signature d'hétérotrophie n'a pu être retrouvée, notamment aucune enzyme codant des hydrolases (chitinases par exemple) n'a été détectée. Ceci ne nous a donc pas permis d'établir un mode de vie saprotrophe de ces organismes. Étonnamment, des enzymes impliquées dans la carboxylation aérobie du glucose, ainsi que des enzymes impliquées dans la décarboxylation anaérobie du pyruvate, ont été retrouvées. Cette double capacité métabolique, pourrait donc constituer une caractéristique originale de ces organismes, qui pourrait représenter un métabolisme ancien d'un point de vue évolutif, perdu au cours de l'évolution par les organismes du phylum Chytridiomycota. Par ailleurs, l'analyse du métagénome nous a permis de détecter des interactions entre champignons et environnement : (1) des interactions biotiques entre champignons et bactéries ont été révélées par des enzymes fongiques impliquées dans la production de molécules antibiotiques (pénicilline) ; ces observations ont été renforcées par la présence dans le jeu de données d'enzymes bactériennes impliquées dans la dégradation de ces molécules antibiotiques. Ainsi, nous avons pu détecter un phénomène d'allopédie, qui permettrait aux champignons de coloniser la niche écologique. Par ailleurs, nous avons pu supposer une interaction entre champignons et bactéries dans le cycle du méthane : les champignons pouvant prendre en charge le méthanol métabolisé dans un premier temps par les bactéries à partir des molécules de méthane présentes dans l'environnement. (2) des interactions abiotiques ont également pu être révélées, par la prise en charge par les champignons de molécules

d'hydrogène sulfuré émises par les fumeurs noirs par exemple, ou encore la possibilité de métaboliser des molécules complexes (nitroalkanes...). L'analyse métagénomique s'est donc avérée intéressante pour approcher les fonctions métaboliques portées par ces organismes inconnus, malgré le faible nombre de séquences fongiques extraites du jeu de données. Les limites de la technique métagénomique que nous avons employée sont discutées dans la partie suivante.

### **Analyse métagénomique : résultats, limites et perspectives**

L'analyse métagénomique que nous avons envisagée pour comprendre les rôles des champignons dans ces écosystèmes souligne aussi certaines limites. La principale difficulté que nous avons rencontrée lors de cette étude est due au nombre limité de contigs fongiques que nous avons pu extraire de l'ensemble du jeu de données, largement dominé par des fragments de gènes bactériens. Ceci peut s'expliquer par la méthodologie que nous avons employée, et l'échantillon que nous avons choisi d'étudier. Notre stratégie, classiquement utilisée dans d'autres analyses métagénomiques s'intéressant à la fraction microbienne des environnements (par exemple Schlüter *et al.*, 2008), s'appuyait sur la comparaison des séquences du métagénome aux banques de données protéiques mondiales. Le principal inconvénient de cette stratégie tient donc au fait que celle-ci repose avant tout sur la composition des bases de données, qui peut être biaisée soit par la présence de séquences mal annotées (dans le cas de bases de données non « administrées »), et par la représentativité des organismes dans ces mêmes bases de données. L'approche originale que nous avons choisie se proposait d'étudier un organisme fongique inconnu, rendant difficile la comparaison dans les bases de données, sans disposer d'organisme de référence. Plus généralement, très peu de séquences de champignons du phylum Chytridiomycota sont disponibles, rendant sans doute difficile une comparaison basée sur des homologies. Aucun génome complet nucléaire de ces organismes n'est actuellement disponible dans Genbank, mais des projets de séquençage sont en cours (notamment pour *Batrachochytridium dendrobatidis*) et devraient être prochainement intégrés dans cette base de données. Des efforts supplémentaires pour combler les manques de données génétiques relatives à ces organismes devront donc être entrepris. Pour comparaison, à l'heure actuelle, 14 génomes nucléaires complets d'Ascomycota sont disponibles dans Genbank, contre seulement 2 génomes de Basidiomycota ; par ailleurs aucun génome nucléaire complet n'est disponible pour les organismes des phyla Glomeromycota, Zygomycota et Chytridiomycota. Les génomes complets nucléaires, présents dans Genbank sont présentés dans le tableau 12. Cette différence dans les efforts de séquençage est

étroitement corrélée aux nombres de protéines dans ces mêmes bases de données (présenté dans le tableau 9 de cette thèse ; partie V). Il faut noter qu'il existe bien plus de génomes de champignons, complètement séquencés (par exemple *Laccaria bicolor* ; Martin *et al.*, 2008b), mais ce tableau présente uniquement les organismes présents dans la base de données que nous avons utilisée, dont les séquences nucléotidiques et les annotations associées sont disponibles.

<b>ASCOMYCOTA</b>
<i>Aspergillus niger</i> CBS 513.88
<i>Aspergillus nidulans</i> FGSC A4
<i>Aspergillus terreus</i> NIH2624
<i>Aspergillus fumigatus</i> Af293
<i>Gibberella zeae</i> PH-1
<i>Chaetomium globosum</i> CBS 148.51
<i>Yarrowia lipolytica</i> CLIB122
<i>Schizosaccharomyces pombe</i> 972h
<i>Debaryomyces hansenii</i> CBS767
<i>Saccharomyces cerevisiae</i>
<i>Ashbya gossypii</i> ATCC 10895
<i>Pichia stipitis</i> CBS 6054
<i>Candida glabrata</i> CBS 138
<i>Kluyveromyces lactis</i> NRRL Y-1140
<b>BASIDIOMYCOTA</b>
<i>Cryptococcus neoformans</i> var. <i>neoformans</i> B-3501A
<i>Cryptococcus neoformans</i> var. <i>neoformans</i> JEC21

**Tableau 12 :** Génomes nucléaires complets de champignons, annotés disponibles dans Genbank, en Février 2009.

Afin de vérifier les résultats de notre analyse métagénomique, il serait intéressant de ré-analyser le même jeu de données que nous avons généré lorsque les bases de données se seront enrichies en protéines de Chytridiomycota, ce qui nous permettrait vraisemblablement d'optimiser l'extraction des contigs fongiques de notre métagénome. Il pourrait également être intéressant de comparer les résultats de cette analyse à ceux résultant du pyroséquençage d'un second échantillon hydrothermal dont nous disposons : cet échantillon pourrait être

choisi sur la base de la quantification du gène codant l'ARNr 18S selon la méthodologie employée dans notre étude. La comparaison tiendrait cette fois à l'analyse d'un échantillon ayant révélé une diversité toujours faible, mais relative à un organisme proche de ceux déjà décrits, comme par exemple un organisme du phylum Ascomycota. Il serait alors intéressant de comparer le nombre de contigs extraits de cette façon.

Compte tenu de la forte diversité spécifique que nous avons mis en lumière dans notre échantillon, de la faible quantité d'ADN de départ (principalement liée à la difficulté d'isoler des échantillons hydrothermaux et donc de disposer de matériel en grande quantité), notre étude a tout de même permis de mettre en évidence les métabolismes bactériens majeurs présents dans le site étudié. Cette approche peut donc être appliquée à un grand nombre d'écosystèmes, afin d'approcher les voies métaboliques bactériennes dominantes, et sans doute de mieux comprendre le cyclage de la matière organique (C, N, O). Disposer d'échantillons environnementaux (hydrothermaux, ou de tout autre écosystème sur Terre) de façon à permettre une extraction d'ADN total en quantité plus importante que celle dont nous disposons, permettrait de mettre en évidence les organismes sous-représentés et les voies métaboliques qu'ils portent, pouvant jouer un rôle majeur dans le fonctionnement de l'écosystème.

Des analyses complémentaires pourraient être réalisées sur la base de nos résultats, offrant des perspectives de recherche multiples. L'analyse métagénomique pourrait d'une part être renforcée par l'étude de la fraction microbienne active dans l'environnement, par des approches de métatranscriptomiques. Cette approche consiste à extraire les ARNm dans l'échantillon ; après rétrotranscription des ARNm, les gènes exprimés dans l'environnement sont analysés (recherches d'homologies dans les bases de données). Cette approche permet d'éclairer les processus actifs dans l'environnement. De précédentes études ont utilisé cette technique avec succès, notamment pour l'analyse des gènes exprimés par des communautés microbiennes marines, impliqués dans les cycles des nutriments (Poretsky *et al.*, 2005). A cet effet, le pyroséquençage se révélerait également un outil puissant. Nous pourrions ainsi envisager une telle approche dans le but de valider les prédictions de fonctionnement métabolique établies sur la base de l'étude du métagénome.

### **Le défi bioinformatique**

L'analyse de données issues du pyroséquençage nécessite une puissance informatique considérable. Ceci est principalement lié à la quantité de séquences générées, à la taille de ces

séquences, et au temps de calcul nécessaire pour analyser le métagénome. Des algorithmes de comparaison spécifiques efficaces doivent être mis en place, au même titre que des algorithmes puissants pour l'assemblage des séquences générées en contigs (notamment dans les cas de communautés complexes). On peut noter que la taille des séquences produites augmente à chaque nouvelle de génération de kits de pyroséquençage (les pyroséquenceurs 454 de génération actuelle produisent des séquences de 400-500 pb en moyenne), rendant plus aisée les recherches d'homologie. D'autre part, compte tenu de la nature même de ces méthodes de recherche d'homologie, des serveurs informatiques dédiés (ou des clusters informatiques) doivent être utilisés ; par exemple, pour les analyses BLASTX, chaque séquence est traduite dans les 6 cadres de lecture, et chacune de ces combinaisons est comparée à l'ensemble de la base de données. Pour des jeux de données constitués de plusieurs milliers, voir de millions de séquences, le temps de calcul est donc très important. A l'heure actuelle il existe des outils informatiques, accessibles en ligne, pour faciliter l'analyse des données métagénomiques : c'est le cas de mgRAST (Meyer *et al.*, 2008) que nous avons utilisé dans ce travail par exemple, mais on peut également citer par exemple IMG/M (Integrated Microbial genomes with Microbial Samples ; Markowitz *et al.*, 2008), élaboré pour la comparaison des données de métagénome, ou encore CAMERA (Community Cyberinfrastructure for Advanced Marine Microbial Ecology Research and Analysis ; Seshadri *et al.*, 2007), qui regroupe les données de séquences métagénomiques, ainsi que les informations environnementales associées. Des améliorations peuvent encore être apportés à ces outils : nous avons souligné dans cette thèse les difficultés à contrôler les résultats générés par mgRAST, que nous avons choisi de compléter par des analyses BLASTX personnelles ; d'autre part, les annotations présentes dans les bases de données telles que CAMERA, peuvent contenir des erreurs, qu'il sera difficile de contrôler. Ainsi, l'analyse informatique des jeux de données métagénomique constitue un véritable défi pour la communauté microbiologiste, dans l'espoir de mieux comprendre le fonctionnement de fraction microbienne des écosystèmes.

*Comme nous l'avons évoqué tout au long de cette thèse, l'étude des champignons marins en écosystème hydrothermal est un axe de recherche très récent, offrant des perspectives de travail nombreuses. Le travail réalisé lors de cette thèse a particulièrement appuyé l'intérêt d'étudier la microflore fongique dans ces environnements extrêmes et soutient donc la nécessité de renforcer les connaissances liés à ces organismes.*



## Références bibliographiques

---



## A

- Alayse-Danet A.M., Desbruyeres D., Gaill F. 1987.** The possible nutritional detoxification role of the epibiotic bacteria of Alvinellid polychaetes : review of the current data. *Symbiosis*. 4: 51-62.
- Alias S., Kuthubutheen A., Jones E. 1995.** Frequency of occurrence of fungi on wood in Malaysian mangroves. *Hydrobiologia*. 295: 97-106.
- Alker A., Smith G., Kim K. 2001.** Characterization of *Aspergillus sydowii* (Thom et Church), a fungal pathogen of Caribbean sea fan corals. *Hydrobiologia*. 460: 105-111.
- Altschul S.F., Gish W., Miller W., Myers EW., Lipman D.J. 1990.** Basic Local Alignment Search Tool. *Journal of Molecular Biology*. 215: 403-410.
- Amann R.I., Ludwig W., Schleifer K.H. 1995.** Phylogenetic identification and in situ detection of individual microbial cells without cultivation. *Microbiological Reviews*. 59: 143-69.
- Amann R., Fuchs B.M. 2008.** Single-cell identification in microbial communities by improved fluorescence in situ hybridization techniques. *Nature Reviews Microbiology*. 6: 339-348.
- Anderson I.C., Parkin P.I. 2007.** Detection of active soil fungi by RT-PCR amplification of precursor rRNA molecules. *Journal of Microbiological Methods*. 68: 248-253.
- Angly F.E., Felts B., Breitbart M., Salamon P., Edwards R.A., Carlson C., Chan A.M., Haynes M., Kelley S., Liu H, Mahaffy J.M., Mueller J.E., Nulton J., Olson R., Parsons R., Rayhawk S., Curtis A., Suttle C.A, Rohwer F. 2006.** The marine viromes of four oceanic regions. *PloS Computational Biology*. 4: e368.
- Atkins M.S. 2000.** Ancyromonadida: A new phylogenetic lineage among the Protozoa closely rRelated to the common ancestor of Metazoans, Fungi, and Choanoflagellates (Opisthokonta). *Journal of Molecular Evolution*. 51: 278-285
- Avise JC, 2004.** Molecular Markers, Natural History, and Evolution. Sunderland, Massachusets. (Sinauer Associates eds).
- Azmi O., Seppelt R. 1998.** The broad-scale distribution of microfungi in the Windmill Islands region, continental Antarctica. *Polar Biology*. 19: 92-100.

## B

- Baldauf S. L., Palmer J. D. 1993.** Animals and fungi are each other's closest relatives: Congruent evidence from multiple proteins. *Proceedings of the National Academy of Sciences, USA*. 90: 11558-11562.
- Baldauf S. L. 1999.** A search for the origins of animals and fungi: Comparing and combining molecular data. *American Naturalist*. 154: S178-S188.
- Baldauf S.L., Roger A.J., Wenk-Siefert I., Doolittle W.F. 2000.** A kingdom-level phylogeny of eukaryotes based on combined protein data. *Science*. 290: 972-7.

- Baldauf S. L.,** Bhattacharya D., Cockrill J., Hugenholtz P., Pawlowski J., Simpson A.G.B. **2004.** The Tree of Life: An Overview, Chapter 4 in *Assembling the Tree of Life*. Eds. Cracraft and Donoghue, Oxford University Press, USA.
- Barghoorn E.S., Linder D.H. 1944.** Marine fungi: their taxonomy and biology. *Farlowia*. 1: 395-467.
- Bar-Hen A.,** Mariadassou M., Poursat M.A., Vandenkoornhuysen P. **2008.** Influence function for robust phylogenetic reconstructions. *Molecular Biology and Evolution*. 25:869–873.
- Barns S. M.,** Takala S. L., Kuske C. R. **1999.** Wide distribution and diversity of members of the bacterial kingdom Acidobacterium in the environment. *Applied and Environmental Microbiology*. 65: 1731-1737.
- Bass D.,** Howe A., Brown N., Barton H., Demidova M., Michelle H., Li L., Sanders H., Watkinson S.C.C., Willcock S., Richards T.A.A. **2007.** Yeast forms dominate fungal diversity in the deep oceans. *Proceedings of the Royal Society B*. 274: 3069-3077.
- Beja O.,** Suzuki M.T., Koonin E.V., Aravind L., Hadd A., Nguyen L.P., Villacorta L., Amjadi M., Garrigues C., Jovanovich S.B. Feldman R.A., DeLong E.F. **2000a.** Construction and analysis of bacterial artificial chromosome libraries from a marine microbial assemblage. *Environmental Microbiology*. 2: 516-529.
- Beja O.,** Aravind L., Koonin E.V., Suzuki M.T., Hadd A., Nguyen L.P., Jovanovich S.B. Gates C.M., Feldman R.A., Spudich J.L., Spudich E.N., DeLong E.F. **2000b.** Bacterial rhodopsin: evidence for a new type of phototrophy in the sea. *Science*. 289: 1902-1906.
- Beja O.,** Spudich E.N., Spudich J.L., Leclerc M., DeLong E.F. **2000c.** Proteorhodopsin phototrophy in the ocean. *Nature*. 411: 786-789.
- Bennett S. 2004.** Solexa Ltd. *Pharmacogenomics*. 5 : 433-438.
- Benny G.L.,** Humber R.A., Morton J.B. **2001.** Zygomycota: Zygomycetes. In *The Mycota VII. Systematic and Evolution part A*. (Esser K., Lemke P.A. eds).
- Benson D.A.,** Karsch-Mizrachi I., Lipman D.J., Ostell J., Wheeler D.L. **2008.** GenBank. *Nucleic Acids Research*. 36: D25-D30.
- Berbee M.L.,** Kloockener B., Taylor J.W. **1992.** Yeast biology. *Science* 257: 1610-1611.
- Berbee M.L., Taylor J.W. 1993.** Dating the evolutionary radiations of the true fungi. *Canadian Journal of Botany*. 71: 114-1127.
- Berbee M.L., Taylor J.W. 2006.** Dating divergences in the fungal tree of life: review and new analyses. *Mycologia*. 98: 838-849.
- Bérdy J. 2005.** Bioactive microbial metabolites. A personal view. *The Journal of Antibiotics*. 58: 1-26.
- Bhadury P.,** Mohammad B.T., Wright P.C. **2006.** The current status of natural products from marine fungi and their potential as anti-infective agents. *Journal of Industrial Microbiology and Biotechnology*. 33: 325-337.
- Blöchl E.,** Rachel R., Burggraf S., Hafenbradl D., Jannasch H.W., Stetter K.O. **1997.** *Pyrolobus fumarii* gen. and sp. nov. represents a novel group of archaea extending the upper temperature limit for life to 113°C. *Extremophiles*. 1: 14 -21.
- Boussau B.,** Blanquart S., Neacsulea A., Lartillot N., Gouy M. **2008.** Parallel adaptations to high temperatures in the Archaean eon. *Nature*. 456: 942-946.
- Bowman B.H,** Taylor J.W., Brownlee A.G., Lee J., Lu S.D., White T.J. **1992.** Molecular evolution of fungi: relationships of the basidiomycetes, ascomycetes, and chytridiomycetes. *Molecular Biology and Evolution*. 9: 285-296.

- Bridge P.D., Roberts P.J., Spooner B.M., Panchal G. 2003.** On the unreliability of published DNA sequences. *New phytologist*. 160: 43-48.
- Britayev T.A., Martin D., Krylova E.M., Von Cosel R., Aksiuk T.S. 2007.** Life-history traits of the symbiotic scale-worm *Branchiopolynoe seepensis* and its relationships with host mussels of the genus *Bathymodiolus* from hydrothermal vents. *Marine Ecology*. 28: 36-48.
- Bruns T.D., Vilgalys R., Barns S., Gonzales D., Hibbet D.S., Lane D.J., Simon L., Stickel S., Szaro T.M., Weisburg W.G., Sogin M. 1992.** Evolutionary relationships within the fungi: analyses of nuclear small subunit rRNA sequences. *Molecular Phylogenetics and Evolution*. 1: 231-241.
- Buchan A., Newell S.Y., Moreta J.I.L., Moran M.A. 2002.** Analysis of Internal Transcribed Spacer (ITS) regions of rRNA genes in fungal communities in a Southeastern US Salt Marsh. *Microbial Ecology*. 43: 329-340.
- Bugni T.S., Ireland C.M. 2004.** Marine-derived fungi: a chemically and biologically diverse group of microorganisms. *Natural Products Report*. 21:143-163.
- Bull A.T., Ward A.C., Goodfellow M. 2000.** Search and discovery strategies for biotechnology: the paradigm shift. *Microbiology and Molecular Biology Reviews*. 64: 573-606.
- Burgaud G., Le Calvez T., Arzur D., Vandenkoornhuysse P., Barbier G. 2009.** Diversity of culturable marine filamentous fungi from deep-sea hydrothermal vents. *Environmental Microbiology*. In Press.
- Butinar L., Zalar P., Frisvad J.C., Gunde-Cimerman N. 2005.** The genus *Eurotium*-members of indigenous fungal community in hypersaline waters of salterns. *FEMS Microbiology Ecology*. 51: 155-166.

## C

- Campbell B.J., Stein J.L., Cary S.C. 2003.** Evidence of chemolithoautotrophy in the bacterial community associated with *Alvinella pompejana*, a hydrothermal vent polychaetes. *Applied and Environmental Microbiology*. 69: 5070-5078.
- Campbell B.J., Engel A.S., Porter M.L., Takai K. 2006.** The versatile  $\epsilon$ -proteobacteria: key players in sulphidic habitats. *Nature Reviews Microbiology*. 4: 458-468.
- Cancela da Fonseca J.P. 1966.** L'outil statistique en biologie du sol. III. Indices d'intérêt écologique. *Revue d'Ecologie et de Biologie du Sol*. 3: 381-407.
- Carlile M.J., Watkinson S.C. The Fungi. 1994.** (Academic Press eds).
- Carson M. J., Hartwell L. 1985.** CDC17: An essential gene that prevents telomere elongation in yeast. *Cell*. 42: 249-257.
- Cavanaugh C.M. 1983.** Symbiotic chemoautotrophic bacteria in marine invertebrates from sulphide-rich habitats. *Nature*. 302: 58 - 61.
- Charlou J.L., Donval J.P., Fouquet Y., Jean-Baptiste P., Holm N. 2002.** Geochemistry of high H<sub>2</sub> and CH<sub>4</sub> vent fluids issuing from ultramafic rocks at the Rainbow hydrothermal field (36°14'N, MAR). *Chemical Geology*. 191: 345-359.

- Chenna R.**, Sugawara H., Koike T., Lopez R., Gibson T.J., Higgins D.G., Thompson J.D. **2003**. Multiple sequence alignment with the Clustal series of programs. *Nucleic Acids Research*. 31: 3497.
- Childress J.J.**, Fisher C.R., Favuzzi J.A., Sanders N.K. **1991**. Sulfide and carbon dioxide uptake by the hydrothermal vent clam *Calymene magnifica* and its chemoautotrophic symbionts. *Physiological Zoology*. 64: 1444-1470.
- Childress J.**, Fisher C. **1992**. Biology of hydrothermal vent animals: physiology, biochemistry, and autotrophic symbioses. *Oceanography and Marine Biology*. 30: 337-441.
- Choi I.G.**, Kim S.H. **2007**. Global extent of horizontal gene transfer. *Proceedings of the National Academy of Sciences, USA*. 104: 4489-4494.
- Christophersen C.**, Crescente O., Frisvad J.C., Gram L., Nielsen J., Nielsen P.H., Rahbaek L. **1998**. Antibacterial activity of marine-derived fungi. *Mycopathologia*. 143: 135-138.
- Colaço A.F.**, Dehairs F., Desbruyères D. **2002**. Nutritional relations of deep-sea hydrothermal fields at the Mid-Atlantic Ridge: a stable isotope approach. *Deep-Sea Research I*. 49: 395-412.
- Cole J.R.**, Chai B., Farris R.J., Wang Q., Kulam-Syed-Mohideen A.S., McGarell D.M., Bandela A.M., Cardenas E., Garrity G.M., Tiedje J.M. **2007**. The ribosomal database project (RDP-II): introducing myRDP space and quality controlled public data. *Nucleic Acids Research*. 35: D169–D172.
- Colwell RK.** **2006**. EstimateS, version 8.0: statistical estimation of species richness and shared species from samples (software and user's guide). Freeware for Windows and Mac OS available: <http://viceroy.eeb.uconn.edu/EstimateS>.
- Convention sur la diversité biologique :** diversité biologique marine et côtière. **2005**. UNEP.
- Corliss J.B.**, Dymond J., Gordon L.I., Edmond J.M., von Herzen R.P., Ballard R.D., Green K., Williams D., Bainbridge A., Crane K., van Andel T.H. **1979**. Submarine Thermal Springs on the Galápagos Rift. *Science*. 203: 1073-1083.
- Corre E.**, Reysenbach A.L., Prieur D. **2001**. Epsilon Proteobacterial diversity from a deep-sea hydrothermal vent on the Mid-Atlantic Ridge. *FEMS Microbiology Letters*. 205: 329-335.
- Cueto M.**, Jensen P.R., Kauffman C., Fenical W., Lobkovsky E., Clardy J. **2001**. Pestalone, a New Antibiotic Produced by a Marine Fungus in Response to Bacterial Challenge. *Journal of Natural Products*. 64: 1444-1446.
- Culley A.I.**, Lang A.S., Suttle C.A. **2006**. Metagenomic analysis of coastal RNA virus communities. *Science*. 312: 1795-1798.
- Cuomo V.**, Palomba I., Perretti A., Guerriero A., D'Ambrosio M., Pietra F. **1995**. Antimicrobial activities from marine fungi. *Journal of Marine Biotechnology*. 2: 199-204.

## D

- Danovaro R.**, Dell'Anno A., Corinaldesi C., Magagnini M., Noble R., Tamburini C., Weinbauer M. **2008**. Major viral impact on the functioning of benthic deep-sea ecosystems. *Nature*. 454: 1084-1087.
- Damare S.**, Raghukumar C., Raghukumar S. **2006**. Fungi in deep-sea sediments of the Central Indian Basin. *Deep-Sea Research pt I*. 53: 14-27.

- Damare S., Raghukumar C. 2007.** Fungi and Macroaggregation in Deep-Sea Sediments. *Microbial Ecology*. 56:168-177.
- D'Auria G., Pushker R., Rodriguez-Valera F. 2006.** IWoCS: analyzing ribosomal intergenic transcribed spacers configuration and taxonomic relationships. *Bioinformatics*. 22: 527–531.
- Dawson R.C., Pace N.R. 2002.** Novel kingdom-level eukaryotic diversity in anoxic environments. *Proceedings of the National Academy of Sciences, USA*. 99: 8324-8329.
- Deacon J. 2006.** Environmental conditions for growth and tolerance of extremes. *Fungal Biology*. 142-157.
- DeLong E.F. 2005.** Microbial community genomics in the ocean. *Nature Reviews of Microbiology*. 3: 459-469.
- DeLong E.F., Preston C.M., Mincer T., Rich V., Hallam S.J., Frigaard N.U., Martinez A., Sullivan M.B., Edwards R., Brito B.R., Chisholm S.W., Karl D.M. 2006.** Community genomics among stratified microbial assemblages in the ocean's interior. *Science*. 311: 496-503.
- Demanèche S., Philippot L., David M.M., Navarro E., Vogel T.M., Simonet P. 2009.** Characterization of denitrification gene clusters of soil bacteria via a metagenomic approach. *Applied and Environmental Microbiology*. 75: 534-537.
- Desbruyères D., Chevaldonne P., Alayse A.M., Jollivet D. Lallier F.H., Jouin-Toulmond C., Zal F., Sarradin P.M., Cosson R., Caprais J.C., Arndt C., O'Brien J., Guezennec J., Hourdez S., Riso R., Gaill F., Laubier L., Toulmond A. 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 research Part 2*. 45: 383-422.
- Desbruyères D., Almeida A., Biscoito M., Comtet T., Khripounoff A., Le Bris N., Sarradin P.M., Segonzac M. 2000.** A review of the distribution of hydrothermal vent communities along the northern Mid-Atlantic Ridge: dispersal vs environmental controls. *Hydrobiologia*. 440: 201-216.
- Desbruyères D., Biscoito M., Caprais J.C., Colaço A., Comtet T., Crassous P., Fouquet Y., Khripounoff A., Le Bris N., Olu K., Riso R., Sarradin P.M., Segonzac M., Vangriesheim A. 2001.** Variations in deep-sea hydrothermal vent communities on the Mid-Atlantic Ridge near the Azores plateau. *Deep-Sea Research part I*. 48: 1325-1346.
- Desmazières J.B., 1849.** Plantes cryptogames de France, 2<sup>nd</sup> ed., No. 1778. Lille.
- Duperron S., Sibuet M., Dubilier N., Gaill F. 2006.** Symbiosis in mytilids from chemosynthetic ecosystems: relationship between symbiont diversity and environmental conditions. *Geophysical Research Abstracts*. 8: 10593.
- Duncan S.M. 2007.** Fungal diversity and cellulolytic activity in the historic huts, Ross Island, Antarctica (PhD).
- Durand P., Reysenbach A.L., Pace N. 1993.** Isolation and characterization of *Thiobacillus hydrothermalis* sp. Nov, a mesophilic obligately chemolithotrophic bacterium isolated from a deep-sea hydrothermal vent in Fiki Basin. *Archives of Microbiology*. 159: 39-44.

## E

- Edgcomb V.P., Kysela D.T., Teske A., de Vera Gomez A., Sogin M.L. 2002.** Benthic eukaryotic diversity in the Guaymas Basin hydrothermal vent environment. *Proceedings of the National Academy of Sciences, USA*. 99: 7658-7662.
- Edgcomb V., Orsi W., Leslin C., Epstein S.S., Bunge J., Jeon S., Yakimov M.M., Behnke A., Stoeck T. 2009.** Protistan community patterns within the brine and halocline of deep hypersaline anoxic basins in the eastern Mediterranean Sea. *Extremophiles*. 13:151-167.
- Entcheva P., Liebl W., Johann A., Hartsch T., Streit W.R. 2001.** Direct cloning for enrichment cultures, a reliable strategy for isolation of complete operons and genes from microbial consortia. *Applied and Environmental Microbiology*. 67: 89-99.
- Eriksson N., Pachter L., Mitsuya Y., Rhee S.Y., Wang C., Gharizadeh B., Ronaghi M., Shafer R.W., Beerenwinkel N. 2008.** Viral population estimation using pyrosequencing. *PLoS Computational Biology*. 4: e1000074.

## F

- Fisher C.R. 1990.** Chemoautotrophic and methanotrophic symbioses in marine invertebrates. *Review of Aquatic Sciences*. 2: 399-613.
- Fisher C.R., Takai K., Le Bris N. 2007.** Hydrothermal vent ecosystems. *Oceanography* 20: 14-23.
- Frias-Lopez J., Shi Y., Tyson G.W., Coleman M.L., Schuster S.C., Chisholm S.W., DeLong E.F. 2008.** Microbial community gene expression in ocean surface waters. *Proceedings of the National Academy of Sciences, USA*. 105: 3805-3810.
- Fujiwara Y., Takai K., Uematsu K., Tsuchida S., Hunt J.C., Hashimoto J. 2000.** Phylogenetic characterization of endosymbionts in three hydrothermal vent mussels: influence on host distributions. *Marine Ecology*. 208: 147-155.

## G

- Gadanhó M., Sampaio J.P. 2005.** Occurrence and diversity of yeasts in the Mid-Atlantic ridge hydrothermal fields near the Azores archipelago. *Microbial Ecology*. 50: 408-417.
- Gadanhó M., Sampaio J.P. 2006.** Microeukaryotic diversity in the extreme environments of the Iberian Pyrite Belt: a comparison between universal and fungi-specific primer sets, temperature gradient gel electrophoresis and cloning. *FEMS Microbiology Ecology*. 57: 139-148.
- Gage J.D., Tyler P.A. Deep-sea biology. 1991.** A natural history of organisms at the deep-sea floor. Cambridge University Press.
- Galagan J.E., Calvo S.E., Borkovich K.A., Selker E.U., Read N.D., Jaffe D., FitzHugh W., Ma L.J., Smirnov S., Purcell S., Rehman B., Elkins T., Engels R., Wang S., Nielsen C.B., Butler J., Endrizzi M., Qui D., Ianakiev P., Bell-Pedersen D., Nelson M.A., Werner-Washburne M., Selitrennikoff C.P., Kinsey J.A., Braun E.L.,**

- Zelter A., Schulte U., Kothe G.O., Jedd G., Mewes W., Staben C., Marcotte E., Greenberg D., Roy A., Foley K., Naylor J., Stange-Thomann N., Barrett R., Gnerre S., Kamal M., Kamvysselis M., Mauceli E., Bielke C., Rudd S., Frishman D., Krystofova S., Rasmussen C., Metzberg R.L., Perkins D.D., Kroken S., Cogoni C., Macino G., Catcheside D., Li W., Pratt R.J., Osmani S.A., DeSouza C.P., Glass L., Orbach M.J., Berglund J.A., Voelker R., Yarden O., Plamann M., Seiler S., Dunlap J., Radford A., Aramayo R., Natvig D.O., Alex L.A., Mannhaupt G., Ebbole D.J., Freitag M., Paulsen I., Sachs M.S., Lander E.S., Nusbaum C., Birren B. **2003**. The genome sequence of the filamentous fungus *Neurospora crassa*. *Nature*. 422: 859-868.
- García M.H.**, Ivanova N., Kunin V., Warnecke F., Barry K.W., McHardy A.C., Yeates C., He S., Salamov A.A., Szeto E., Dalin E., Putnam N.H., Shapiro H.J., Pangilinan J.L., Rigoutsos I., Kyrpidis C., Blackall L.L., Mc Mahon K.D., Hugenholtz P. **2006**. Metagenomic analysis of two enhanced biological phosphorus removal (EBPR) sludge communities. *Nature Biotechnology*. 24: 1263-1269.
- Gargas A.**, DePriest P.T., Grube M., Tehler A. **1995**. Multiple origins of lichen symbioses in fungi suggested by SSU rDNA phylogeny. *Science*. 268: 1492-1495.
- Gertz J.**, Siggia E.D., Cohen B.A. **2009**. Analysis of combinatorial cis-regulation in synthetic and genomic promoters. *Nature*. 457: 215-218.
- Gharizadeh B.**, Herman Z.S., Eason R.G., Jejelowo O., Pourmand N. **2006**. Large-scale pyrosequencing of synthetic DNA: a comparison with results from Sanger dideoxy sequencing. *Electrophoresis*. 27: 3042-3047.
- Gill S.R.**, Pop M., DeBoy R.T., Eckburg P.B., Turnbaugh P.J., Samuel B.S., Gordon J.I., Relman D.A., Fraser-Liggett C.M., Nelson K.E. **2006**. Metagenomic analysis of the human distal gut microbiome. *Science*. 312:1355-1359.
- Goffredi S.K.**, Childress J.J., Desaulniers N.T., Lallier F.H. **1997**. Sulfide Acquisition by the vent worm Riftia Pachyptila appears to be via uptake of HS<sup>-</sup> rather than H<sub>2</sub>S. *The Journal of Experimental Biology*. 200: 2609-2616.
- Gontcharov A.A.**, Marin B., Melkonian M. **2004**. Are combined analyses better than single gene phylogenies? A case study using SSU rDNA and rbcL sequence comparisons in the Zygnematophyceae (Streptophyta). *Molecular Biology and Evolution*. 21: 612-624.
- Guarro J.**, Gené J., Stchigel A.M. **1999**. Developments in Fungal Taxonomy. *Clinical Microbiology Reviews*. 12: 454-500.
- Guindon S.**, Lethiec F., Duroux P., Gascuel O. **2005**. PHYML online-a web server for fast maximum likelihood-based phylogenetic inference. *Nucleic Acids Research*. 33: 557-559.

## H

- Handelsman J.**, Rondon M.R., Brady S.F., Clardy J., Goodman R.M. **1998**. Molecular biological access to the chemistry of unknown soil microbes: a new frontier for natural products. *Chemistry and Biology*. 5:R245-R249.
- Handelsman J.** **2004**. Metagenomics: Application of Genomics to Uncultured Microorganisms. *Microbiology and Molecular Biology Reviews*. 68: 669-685.

- Harkins T., Jarvie T. 2007.** Metagenomics analysis using the Genome Sequencer™ FLX system. *Nature Methods*. iii-iv.
- Hawksworth D.L. 1991.** The fungal dimension of biodiversity: magnitude, significance and conservation. *Mycological Research*. 95: 641-655.
- Hawksworth D.L., Kirk P.M., Sutton B.C., Pegler D.N. 1995.** In Ainsworth and Bisby's Dictionary of Fungi. (8th ed.) *CAB International, Wallingford*. 1-616.
- Hawksworth D.L., Rossman A.Y. 1997.** Where are all the undescribed fungi? *Phytopathology*. 87: 888-891.
- Hawksworth D.L. 2001.** The magnitude of fungal diversity: the 1.5 million species estimate revisited. *Mycological Research*. 105: 1422-1432.
- Heckman D.S., Geiser D.M., Eidell B.R., Stauffer R.L., Kardos N.L., Hedges S.B. 2001.** Molecular evidence for the early colonization of land by fungi and plants. *Science*. 293: 1129-1133.
- Helgason T., Daniell T.J., Husband R., Fitter A.H., Young J.P.W. 1998.** Ploughing up the wood-wide web? *Nature*. 394: 431.
- Helgason T., Watson I.J., Young J.P.W., 2003.** Phylogeny of the Glomerales and Diversisporales (Fungi: Glomeromycota) from actin and elongation factor 1-alpha sequences. *FEMS Microbiology Letters*. 229: 127-132.
- Hibbet D.S., Donoghue M.J. 2001.** Analysis of character correlations among wood decay mechanisms, mating systems, and substrate ranges in homobasidiomycetes. *Systematic Biology*. 50: 215-242.
- Hibbett D.S., Binder M., Bischoff J.F., Blackwell M., Cannon P.F., Eriksson O.E., Huhndorf S., James T., Kirk P.M., Lücking R., Lumbsch H.T., Lutzoni F., Matheny P.B., McLaughlin D.J., Powell M.J., Redhead S., Schoch C.L., Spatafora J.W., Stalpers J.A., Vilgalys R., Aime M.C., Aptroot A., Bauer R., Begerow D., Benny G.L., Castlebury L.A., Crous P.W., Dai Y.C., Gams W., Geiser D.M., Griffith G.W., Gueidan C., Hawksworth D.L., Hestmark G., Hosaka K., Humber R.A., Hyde K.D., Ironside J.E., Kõljalg U., Kurtzman C.P., Larsson K.H., Lichtwardt R., Longcore J., Miadlikowska J., Miller A., Moncalvo J.M., Mozley-Standridge S., Oberwinkler F., Parmasto E., Reeb V., Rogers J.D., Roux C., Ryvarden L., Sampaio J.P., Schüßler A., Sugiyama J., Thorn R.G., Tibell L., Untereiner W.A., Walker C., Wang Z., Weir A., Weiss M., White M.M., Winka K., Yao Y.J., Zhang N. 2007.** A higher-level phylogenetic classification of the Fungi. *Mycological Research*. 111: 509-547.
- Hiep-Hoa T.N., Sean J. E., Hon-Kay Y., Sunney I. C. 1998.** The Particulate Methane Monooxygenase from *Methylococcus capsulatus* (Bath) Is a Novel Copper-containing Three-subunit Enzyme. Isolation and characterization. *Journal of biological chemistry*. 273: 7957-7966.
- Holm N.G. 1992.** Why are hydrothermal systems proposed as plausible environments for the origin of life? *Origin of Life and Evolution of Biosphere*. 22: 5-14.
- Huber C., Wächtershäuser G. 1997.** Activated Acetic Acid by Carbon Fixation on (Fe,Ni)S Under Primordial Conditions. *Science*. 276: 245-247.
- Huber C., Wächtershäuser G. 1998.** Peptides by activation of amino acids with CO on (Ni,Fe)S surfaces: implications for the origin of life. *Science*. 281: 670-672.
- Huber H., Hohn M.J., Reinhard R., Fuchs T., Verena C.W., Stetter K.O. 2002.** A new phylum of Archaea represented by a nanosized hyperthermophilic symbiont. *Nature*. 417: 63-67.



- Huber C.**, Eisenreich W., Hecht S., Wächtershäuser G. **2003**. A Possible Primordial Peptide Cycle. *Science*. 301: 938-940.
- Hudson D.**, Auch A.F., Qi J., Schuster S.C. **2007**. MEGAN analysis of metagenomic data. *Genome Research*. 17:377-386.
- Hugenholtz P.**, Tyson G.W. **2008**. Metagenomics. *Nature*. 455: 481-483.
- Huse S.M.**, Huber J.A., Morrison H.G., Sogin M.L., Welch D.M. **2007**. Accuracy and quality of massively parallel DNA pyrosequencing. *Genome Biology*. 8: R143.
- Huse S.M.**, Dethlefsen L., Huber J.A., Welch D.M., Relman D.A, Mitchell L., Sogin M.L. **2008**. Exploring Microbial Diversity and Taxonomy Using SSU rRNA Hypervariable Tag Sequencing. *PloS Genetics*. 4 : e1000255
- Hyde K.D.**, Jones E.B.G., Leñaño E., Pointing S.B., Vrijmoed L.L.P. **1998**. Role of fungi in marine ecosystems, *Biodiversity and Conservation*. 7: 1147-1161.
- Hyde K.D.**, Sarma V.V., Jones E.B.G. **2000**. Morphology and taxonomy of higher marine fungi. In *Marine Mycology: A Practical Approach* (Hyde K.D, Pointing S.B. eds). 172-204.

## J

- James T.Y.**, Porter D., Leander C.A., Vilgalys R., Longcore J.E. **2000**. Molecular phylogenetics of the Chytridiomycota supports the utility of ultrastructural data in chytrid systematics. *Canadian Journal of Botany*. 78: 336-350.
- James T.Y.**, Kauff F., Schoch C.L., Matheny P.B., Hofstetter V., Cox C.J., Celio G., Gueidan C., Fraker E., Miadlikowska J., Lumbsch H.T., Rauhut A., Reeb V., Arnold A.E., Amtoft A., Stajich J.E., Hosaka K., Sung G.H., Johnson D., O'Rourke B., Crockett M., Binder M., Curtis J.M., Slot J.C., Wang Z., Wilson A.W., Schübler A., Longcore J.E., O'Donnell K., Mozley-Standridge S., Porter D., Letcher P.M., Powell M.J., Taylor J.W., White M.M., Griffith G., Davies D.R., Humber R.A., Morton J.B., Sugiyama J., Rossman A.Y., Rogers J.D., Pfister D.H., Hewitt D., Hansen K., Hambleton S., Shoemaker R.A., Kohlmeyer J., Volkmann-Kohlmeyer B., Spotts R.A., Serdani M., Crous P.W., Hughes K.W., Matsuura K., Langer E., Langer G., Untereiner W.A., Lücking R., Bädler B., Geiser D.M., Aptroot A., Diederich P., Schmitt I., Schultz M., Yahr R., Hibbett D.S., Lutzoni F., McLaughlin D.J., Spatafora J.W., Vilgalys R. **2006**. Reconstructing the early evolution of Fungi using a six-gene phylogeny. *Nature*. 443: 818-822.
- Jannash H.W.**, Taylor C.D. **1984**. Deep sea microbiology. *Annual Reviews of Microbiology*. 38: 487-514.
- Jannasch, H.W.** **1989**. Chemosynthetically sustained ecosystems in the deep sea. In, *Autotrophic Bacteria*. Springer Verlag. (Schlegel H.G., Bowien, B. eds.), 147–166.
- Jeanthon C.** **2000**. Molecular ecology of microbial vent communities. *Antonie van Leeuwenhoek*. 77: 117-133.
- Jennings D.H.**, Lysek G. **1996**. Fungal biology: understanding the fungal lifestyle. (Bios Scientific publishers eds).
- Jin Y.**, Weining S., Nevo E. **2005**. A MAPK gene from Dead Sea fungus confers stress tolerance to lithium salt and freezing-thawing: Prospects for saline agriculture. *Proceedings of the National Academy of Sciences*,

USA. 102: 18992-18997.

**Jones E.B.G, Alias S.A. 1997.** Biodiversity of mangrove fungi. In *Biodiversity of tropical microfungi* (Hyde K.D. ed.) 71-92.

**Jørgensen B.B., Boetius A. 2007.** Feast and famine – microbial life in the deep-sea bed. *Nature Reviews of Microbiology*. 5: 770-781.

## K

**Kaiser M.J., Attrill M.J., Jennings S., Thomas D.N., Barnes D.K.A., Brierly A.S., Polunin N.V., Raffaelli D.G., Williams P.J. 2005.** Marine ecology. Processes, systems and impacts. Oxford University Press.

**Karl D.M. 1995.** Ecology of free-living, hydrothermal vent microbial communities. In *The Microbiology of Deep-Sea Hydrothermal Vents*. (Karl, D.M. ed.). CRC Press, 35–124.

**Kashefi K., Lovley D. R. 2003.** Extending the upper temperature limit for life. *Science*. 301: 934.

**Katoh K., Kuma K., Toh H., Miyata T. 2005.** MAFFT version 5: improvement in accuracy of multiple sequence alignment. *Nucleic Acids Research*. 33: 511-518.

**Kis-Papo T. 2005.** Marine Fungal Communities In *The Fungal Community: Its Organization And Role In The Ecosystem* (Dighton J., James F White Jr J.F., Peter Oudemans P. eds). 23: 61-91.

**Kochkina G.A., Ivanushkina N.E., Akimov V.N., Gilichinskii D.A., Ozerskaya S.M. 2007.** Halo- and psychrotolerant Geomyces fungi from arctic cryopegs and marine deposits. *Microbiology*. 76: 31-38.

**Kohlmeyer J., Kohlmeyer E. 1979.** Marine mycology: the higher fungi. London: Academic Press.

**Kõljalg U., Larsson K.H., Abarenkov K., Nilsson R.H., Alexander I.J., Eberhardt U., Erland S., Høiland K., Kjoller K., Larsson E., Pennanen T., Sen R., Taylor A.F.S., Tedersoo L., Vrålstad T., Björn M., Ursing B.M. 2005.** Unite: a database providing web-based methods for the molecular identification of ectomycorrhizal fungi. *New Phytologist*. 166: 1063-1068.

**Koonin E.V., Galperin M.Y. 2002.** Sequence - Evolution - Function: computational approaches in comparative genomics. Kluwer Academic Publishers.

**Koonin E.V., Martin W. 2005.** On the origin of genomes and cells within inorganic compartments. *Trends in Genetic*. 21: 647-654.

**Krause L., Diaz N.N., Bartels D., Edwards R.A., Puhler A., Rohwer F., Meyer F., Stoye J. 2006.** Finding novel genes in bacterial communities isolated from the environment. *Bioinformatics*. 22: 281-289.

**Kulikova T., Aldebert P., Althorpe N., Baker W., Bates K., Browne P., van den Broek A., Cochrane G., Duggan K., Eberhardt R., Faruque N., Garcia-Pastor M., Harte N., Kanz C., Leinonen R., Lin Q., Lombard V., Lopez R., Mancuso R., McHale M., Nardone F., Silventoinen V., Stoehr P., Stoesser G., Tuli M.A., Tzouvara K., Vaughan R., Wu D., Zhu W., Apweiler R. 2004.** The EMBL nucleotide sequence database. *Nucleic Acids Research*. 32: D27-D30.

**Küpper** F.C., Maier I., Müller D.T., Loiseaux-De Goer S., Guillou L. **2006**. Phylogenetic affinities of two eukaryotic pathogens of marine macroalgae, *Eurychasma dicksonii* (Wright) Magnus and *Chytridium polysiphoniae* Cohn. *Cryptogamie Algologie*. 27: 165-184.

## L

**Lam** P., Cowen J.P., Jones R.D. **2004**. Autotrophic ammonia oxidation in a deep-sea hydrothermal plume. *FEMS Microbiology Ecology*. 47:191-206.

**Larkin** M.A., Blackshields G., Brown N.P., Chenna R., McGettigan P.A., McWilliam H., Valentin F., Wallace I.M., Wilm A., Lopez R., Thompson J.D., Gibson T.J., Higgins D.G. **2007**. Clustal W and Clustal X version 2.0. *Bioinformatics*. 23:2947-2948.

**Lee** R.L., Robinson J.J., Cavanaugh C.M. **1999**. Pathways of inorganic nitrogen assimilation in chemoautotrophic bacteria-marine invertebrate symbioses: expression of host and symbionts glutamine synthetase. *The Journal of Experimental Biology*. 202: 289-300.

**Le Calvez** T., Burgaud G., Mahé S., Barbier G., Vandenkoornhuysen V. **Submitted**. Fungal Diversity in Deep Sea Hydrothermal Ecosystems.

**Leininger** S., Urich T., Schloter M., Schwark L., Huber J.A., Welch D.M., Huse S.M., Neal P.R., Arrieta J.M., Herndl G.J. **2006**. Archaea predominate among ammonia-oxidising prokaryotes in soils. *Nature*. 442: 806-809.

**Le Bris** N., Govenar B., Le Gall C., Fisher C.R. **2006**. Variability of physico-chemical conditions in 9° 5'N EPR diffuse flow vent habitats. *Marine chemistry*. 98: 167-182.

**Little** C.T.S., **Vrijenhoek** R.C. **2003**. Are hydrothermal vent animals living fossils? *Trends in ecology and evolution*. 18: 582-588.

**Li** H., Chi Z., Wang X., Duan X., Ma L., Gao L. **2007**. Purification and characterization of extracellular amylase from the marine yeast *Aureobasidium pullulans* N13d and its raw potato starch digestion. *Enzyme and Microbial Technology*. 40: 1006-1012.

**Lin** X., Huang Y., Fang M., Wang J., Zheng Z., Su W. **2005**. Cytotoxic and antimicrobial metabolites from marine lignicolous fungi, *Diaporthe* sp. *FEMS Microbiology Letters*. 251: 53-58.

**Liu** Y.J., Hodson M.C., Hall B.D. **2006**. Loss of the flagellum happened only once in the fungal lineage: phylogenetic structure of Kingdom Fungi inferred from RNA polymerase II subunit genes. *BMC Evolutionary Biology*. 6: 74-87.

**Liu** Z., Lozupone C., Hamady M., Bushman F.D., Knight R. **2007**. Short pyrosequencing reads suffice for accurate microbial community analysis. *Nucleic Acids Research*. 35: e120-1-10.

**Longcore** J.E., Pessier A.P., Nichols D.K. **1999**. *Batrachochytrium dendrobatidis* gen. et sp. nov., a chytrid pathogenic to amphibians. *Mycologia*. 91: 219-227.

**López-García** P., Philippe H., Gail F., Moreira D. **2003**. Autochthonous eukaryotic diversity in hydrothermal sediment and experimental microcolonizers at the Mid-Atlantic-ridge. *Proceedings of the National Academy of Sciences, USA*. 100: 697-702.

- López-García P.**, Vereshchaka A., Moreira D. **2007**. Eukaryotic diversity associated with carbonates and fluid–seawater interface in Lost City hydrothermal field. *Environmental Microbiology*. 9: 546-554.
- Lorenz R.**, **Molitoris H.P.** **1997**. Cultivation of fungi under simulated deep-sea conditions. *Mycological Research*. 101: 1355-1365.
- Ludwig W.**, Strunk O., Westram R., Richter L., Meier H., Yadhukumar, Buchner A., Lai T., Steppi S., Jobb G., Förster W., Brettske I., Gerber S., Ginhart A.W., Gross O., Grumann S., Hermann S., Jost R., König A., Liss T., Lüßmann R., May M., Nonhoffl B., Reichel B., Strehlow R., Stamatakis A., Stuckmann N., Vilbig A., Lenke M., Ludwig T., Bode A., Schleifer K.H. **2004**. ARB: a software environment for sequence data. *Nucleic Acids Research*. 32: 1363-1371.
- Luther III G.W.**, Rozan T.F, Taillefertk M., Nuzzio D.B., Di Meo C., Shank T. M., Lutz R.A., Craig C. S. **2001**. Chemical speciation drives hydrothermal vent ecology. *Nature*. 410: 813-816.
- Lutzoni F.**, Kauff F., Cox C.J., McLaughlin D., Celio G., Dentinger B., Padamsee M., Hibbett D., James T.Y., Baloch E., Grube M., Reeb V., Hofstetter V., Schoch C., Arnold A.E., Miadlikowska J., Spatafora J., Johnson D., Hambleton S., Crockett M., Shoemaker R., Sung G.H., Lücking R., Lumbsch T., O'Donnell K., Binder M., Diederich P., Ertz D., Gueidan C., Hansen K., Harris R.C., Hosaka K., Lim Y.W., Matheny B., Nishida H., Pfister D., Rogers J., Rossman A., Schmitt I., Sipman H., Stone J., Sugiyama J., Yahr R., Vilgalys R. Assembling the fungal tree of life: progress, classification, and evolution of subcellular traits. **2004**. *American Journal of Botany*. 91: 1446–1480.

## M

- Madelin T.M.** **1994**. Fungal aerosols: a review. *Journal of aerosol science*. 25: 1405-1412.
- Manefield M.**, Whiteley A.S., Griffiths R.I., Bailey M.J. **2002**. RNA stable isotope probing, a novel means of linking microbial community function to phylogeny. *Applied and Environmental Microbiology*. 68: 5367-5373.
- Mardis E.R.** **2008**. The impact of next-generation sequencing technology on genetics. *Trends in Genetics*. 24: 133-141.
- Margulies M.**, Egholm M., Altman W.E., Attiya S., Bader J.S., Bemben L.A., Berka J., Braverman M.S., Chen Y.J., Chen Z., Dewell S.B., Du L., Fierro J.M., Gomes X.V., Godwin B.C., He W., Helgesen S., Ho C.H., Irzyk G.P., Jando S.C., Alenquer M.L.I., Jarvie T.P., Kshama B., Jirage K.B., Kim J.B., Knight J.R., Lanza J.R., Leamon J.H., Lefkowitz S.M., Lei M., Li J., Lohman K.L., Lu H., Makhijani V.B., McDade K.E., McKenna M.P., Myers E.W., Nickerson E., Nobile J.R., Plant R., Puc B.P., Ronan M.T., Roth G.T., Sarkis G.J., Simons J.F., Simpson J.W., Srinivasan M., Tartaro K.R., Tomasz A., Vogt K.A., Volkmer G.A., Wang S.H., Wang Y., Weiner M.P., Yu P., Begley R.F., Rothberg J.M. **2005**. Genome sequencing in microfabricated high-density picolitre reactors. *Nature*. 437: 376-380.
- Marris E.** **2006**. Drugs from the deep. *Nature*. 443: 904-905.
- Martin W.**, Baross J., Kelley D., Russell M.J. **2008**. Hydrothermal vents and the origin of life. *Nature Reviews Microbiology*. 6: 805-814.
- Maruyama A.**, Honda D., Yamamoto H., Kitamura K. **2000**. Phylogenetic analysis of psychrophilic bacteria isolated from the Japan Trench, including a description of the deep-sea species *Psychrobacter pacificensis* sp. nov. *International Journal of Systematic and Evolutionary Microbiology*. 50: 835-846.

- Mavromatis K.**, Ivanova N., Barry K., Shapiro H., Goltsman E., McHardy A.C., Rigoutsos I., Salamov A., Korzeniewski F., Land M., Lapidus A., Grigoriev I., Richardson P., Hugenholtz P., Kyrpides N.C. **2007**. Use of simulated data sets to evaluate the fidelity of metagenomic processing methods. *Nature Methods*. 4: 495-500.
- Meyer F.**, Paarman D., D'Souza M., Olson R., Glass E.M., Kubal M., Paczian T., Rodriguez A., Stevens R., Wilke A., Wilkening J., Edwards R.A. **2008**. The metagenomics RAST server: a public resource for the automatic phylogenetic and functional analysis of metagenomes. *BMC Bioinformatics*. 9: 386.
- Miller J.D.** **2000**. Screening for secondary metabolites. In Marine mycology. A practical approach. (Hyde K.D. et Pointing S.B. eds). Fungal Diversity Press. University of Hong Kong.
- Minic Z.**, Serre V. Hervé G. **2006**. Adaptation des organismes aux conditions extrêmes des sources hydrothermales marines profondes. *Compte Rendus. Biologies*. 329: 527-540.
- Mischler B.D.**, **Brandon**, R.N. **1987**. Individuality, pluralism, and the phylogenetic species concept. *Biological Philosophy*. 2, 397-414.
- Moreira D.**, **López-García P.** **2003**. Are hydrothermal vents oases for parasitic protists? *Trends in Parasitology*. 19: 556-558.
- Morton J.B.**, **Benny G.L.** **1990**. Revised classification of arbuscular mycorrhizal fungi (Zygomycetes): a new order, Glomales, two new suborders, Glomineae and Gigasporineae, and two new families, Acaulosporaceae and Gigasporaceae, with an emendation of Glomaceae. *Mycotaxon*. 37: 471-491.
- Moon C.D.**, Miles C.O., Jarlfors U., Schardl C.L. **2002**. The evolutionary origins of three new *Neotyphodium* endophyte species from grasses indigenous to the Southern Hemisphere. *Mycologia*. 94: 694-711.
- Mozouras R.** **1986**. Decay of wood by micro-organisms in aquatic habitats. *Record of the Annual Convention of the British Wood Preserving Association*. 8: 1-18.
- Mueller G.M.**, **Schmit J.P.** **2007**. Fungal biodiversity: what do we know ? What can we predict? *Biodiversity and Conservation*. 16: 1-5.
- Munn C.B.** **2004**. Marine microbiology. Ecology and applications. Garland Sciences/ BIOS Scientific publishers.
- Murray A.W.**, Schultes N.P., Szostak J.W. **1986**. Chromosome length controls mitotic chromosome segregation in yeast. *Cell*. 45: 529-536.

## N

- Nagahama T.**, Hamamoto M., Nakase T., Horikoshi K. **2001**. Distribution and identification of red yeasts in deep-sea environments around the northwest Pacific ocean. *Antonie Van Leeuwenhoek International Journal of general and molecular microbiology*. 80: 101-110.
- Nelson D.C.**, Wirsén C.O., Holger W., Jannasch H.W. **1989**. Characterization of Large, Autotrophic *Beggiatoa* spp. abundant at hydrothermal vents of the Guaymas Basin. *Applied and Environmental Microbiology*. 55: 2909-2917.
- Nercessian 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. *Environmental Microbiology*. 7: 118-132.

**Neubert K.**, Mendgen K., Brinkmann H., Wirsal S.G.R. **2006**. Only a Few Fungal Species Dominate Highly Diverse Mycofloras Associated with the Common Reed. *Applied and Environmental Microbiology*. 72: 1118-1128.

**Newman W.A.** **1985**. The abyssal hydrothermal vent invertebrate fauna. A glimpse of antiquity? *Bulletin of the biological society of Washington*. 6: 231-242.

**Nilsson R.H.**, Ryberg M., Kristiansson E., Abarenkov K., Larsson K.H., Kõljalg U. **2006**. Taxonomic reliability of DNA sequences in public sequence databases: A fungal perspective. *PloS One*. 1: e59.

## O

**Opik M.**, Moora M., Zobel M., Saks U. Wheatley R., Wright F., Daniell T. **2008**. High diversity of arbuscular mycorrhizal fungi in a boreal herb-rich coniferous forest. *New Phytologist*. 179: 867-876.

## P

**Pace N.R.**, Stahl D.A., Lane D.J., Olsen G.J. **1986**. Analyzing natural microbial populations by ribosomal rRNA sequences. *Advances in microbial ecology*. 9: 1-55.

**Pace N.R.** **1997**. A molecular view of microbial diversity and the biosphere. *Science*. 276: 734-740.

**Palacios C.**, Zbinden M., Baco A.R., Treude T., Smith C.R., Gaill F., Lebaron P., Boetius A. **2006**. Microbial ecology of deep-sea sunken wood : quantitative measurements of bacterial biomass and cellulolytic activities. *Cahier de Biologie Marine*. 47: 415-420.

**Pezacka E.**, **Wood H.G.** **1988**. Acetyl-coA Pathway of Autotrophic Growth. Identification of the methyl-binding site of the CO dehydrogenase. *The Journal of biological chemistry*. 263: 16000-16006.

**Philippe H.**, **Laurent J.** **1998**. How good are deep phylogenetic trees? *Current opinion in genetics and development*. 8: 616-623.

**Pimenov N.V.**, Lein A.Y., Sagalevich A.M., and Ivanov M.V. **2000**. Carbon Dioxide Assimilation and Methane Oxidation in Various Zones of the Rainbow Hydrothermal Field. *Microbiology*. 69: 689-697.

**Poinar H.N.**, Schwarz C., Qi J., Shapiro B., MacPhee R.D.E., Buigues B., Tikhonov A., Huson D., Tomsho L.P., Auch A., Rampp M., Miller W., Schuster S.C. **2006**. Metagenomics to paleogenomics: Large scale sequencing of mammoth DNA. *Science*. 331: 392-394.

**Pointing S.**, Vrijmoed L.L.P., Jones E.B.G. **1992**. A qualitative assessment of lignocellulose degrading enzyme activity in marine fungi. *Botanica marina*. 41: 293-298.

**Poretsky R.S.**, Bano N., Buchan A., LeCleir G., Kleikemper J., Pickering M., Pate W.M., Moran M.A., Hollibaugh J.T. **2005**. Analysis of microbial gene transcripts in environmental samples. *Applied and environmental microbiology*. 71: 4121-4126.

**Posada D.**, **Crandall K.** **1998**. Modeltest: testing the model of DNA substitution. *Bioinformatics*. 14: 817-818.

**Powell M.J.** **1993**. Looking at mycology with a Janus face: a glimpse at Chytridiomycetes active in the environment. *Mycologia*. 85: 1-20.

**Prieur D.**, Erauso G., Jeanthon C. **1995**. Hyperthermophilic life at deep sea hydrothermal vents. *Planetary and Space Science*. 43: 115-122.

- Prieur D. 1997.** Microbiology of deep sea hydrothermal vents. *Marine Biotechnology*. 15: 242-244.
- Pruesse E., Quast C., Knittel K., Fuchs B., Ludwig W., Peplies J., Glöckner F.O. 2007.** SILVA: a comprehensive online resource for quality checked and aligned ribosomal RNA sequence data compatible with ARB. *Nucleic Acids Research*. 35: 7188-7196.
- Pruitt K.D., Tatusova T., Maglott D.R. 2007.** NCBI Reference Sequence (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins. *Nucleic Acids Research*. 35: D61-D65.
- Purvis A., Hector A., 2000.** Getting the measure of biodiversity. *Nature*. 405: 212-219.

## R

- Radajewski S., Ineson P., Parekh N.R., and Murrell J.C. 2000.** Stable-isotope probing as a tool in microbial ecology. *Nature*. 403: 646-649.
- Raes J., Foerstner K.U., Bork P. 2007.** Get the most out of your metagenome: computational analysis of environmental sequence data. *Current Opinion in Microbiology*. 10: 490-498.
- Raghukumar C., Raghukumar S. 1991.** Fungal invasion of massive corals. *Marine Ecology*. 12: 251-260.
- Raghukumar C., Raghukumar S. 1998.** Barotolerance of fungi isolated from deep-sea sediments of the Indian Ocean. *Aquatic Microbial Ecology*. 15: 153-163.
- Raghukumar C., Raghukumar S., Sheelu G., Gupta S.M., Nagender N.B., Rao B.R. 2004.** Buried in time: culturable fungi in a deep-sea sediment core from the Chagos Trench, Indian Ocean. *Deep-Sea Research Part I*. 51: 1759-1768.
- Raguenes G., Meunier J.R., Antoine E., Godfroy A., Caprais J.C., Lesongeur F., Guezennec J., Barbier G. 1995.** Biodiversity of hyperthermophilic Archaea from hydrothermal vents of the East Pacific Rise. *Comptes rendus de l'Académie des sciences*. 318: 395-402.
- Ramirez-Llodra E., Shank T.M., German C.R. 2007.** Biodiversity and biogeography of hydrothermal vent species: thirty years of discovery and investigations. *Oceanography*. 20: 30-41.
- Ravindran J., Raghukumar C., Raghukumar S. 2001.** Fungi in *Porites lutea*: association with healthy and diseased corals. *Diseases of aquatic organisms*. 47: 219-228.
- Redecker D., Kodner R., Graham L.E. 2000.** Glomalean fungi from Ordovician. *Science*. 289: 1920-1921.
- Redecker D. 2002.** New views on fungal evolution based on DNA markers and the fossil. *Research in Microbiology*. 153: 125-130.
- Remy W., Taylor T.N., Hass H., Kerp H. 1994.** Four hundred-million-year-old vesicular arbuscular mycorrhizae. *Proceedings of the National Academy of Sciences, USA*. 91: 11841-11843.
- Rieley G., Van Dover C.L., Hedrick D.B., Eglinton G. 1999.** Trophic ecology of *Rimicaris exoculata*: a combined lipid abundance/stable isotope approach. *Marine Biology*. 133:495-499.
- Rinaldi M. G. 1989.** Emerging opportunists. *Infectious Disease Clinics of North America*. 3: 65-76.
- Robert F., Chaussidon M. 2006.** A paleotemperature curve for the Precambrian oceans based on silicon isotopes in cherts. *Nature*. 443: 969-972.

- Robidart J.C.**, Bench S.R., Feldman R.A., Novoradovsky A., Podell S.B., Gaasterland T., Allen E.E., Felbeck H. **2008**. Metabolic versatility of the *Riftia pachyptila* endosymbiont revealed through metagenomics. *Environmental Microbiology*. 10: 727-737.
- Robinson J.J.**, **Cavanaugh C.M.** **1995**. Expression of form I and form II Rubisco in chemoautotrophic symbioses: Implications for the interpretation of stable carbon isotope values. *Limnology and Oceanography*. 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*. *Journal of Bacteriology*. 180: 1596-1599.
- Ronquist F.**, **Huelsenbeck J.P.** **2003**. MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics*. 19: 1572-1574.
- Rowley D.C.**, Kelly S., Kauffman C.A., Jensen P.R., Fenical W. **2003**. Halovirs A-E, new antiviral agents from a marine-Derived fungus of the genus *Scytalidium*. *Bioorganic & Medicinal Chemistry*. 11: 4263-4274.

## S

- Sanger F.**, Nicklen S., Coulson A.R. **1977**. DNA sequencing with chain-terminating inhibitors. Proceedings of the National Academy of Sciences, USA. 74, 5463-5467.
- Schlüter A.**, Bekel T., Diaz N.N., Dondrup M., Eichenlaub R., Gartemann K.H., Krahn I., Krause L., Krömeke H., Kruse O., Mussnug J.H., Neuweger H., Niehaus K., Pühler A., Runte K.J., Szczepanowski R., Tauch A., Tilker A., Viehöver P., Goesmann A. **2008**. The metagenome of a biogas-producing microbial community of a production-scale biogas plant fermenter analysed by the 454-pyrosequencing technology. *Journal of Biotechnology*. 136: 77-90.
- Saitou N.**, **Nei M.** **1987**. The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Molecular biology and evolution*. 4: 406-25.
- Sanders I.R.**, **2004**. Intraspecific genetic variation in arbuscular mycorrhizal fungi and its consequences for molecular biology, ecology, and development of inoculum. *Canadian Journal of Botany*. 82: 1057-1062.
- Sarma V.V.**, Hyde K.D., Vittal B. **2001**. Frequency of occurrence of mangrove fungi from the east coast of India. *Hydrobiologia*. 455: 41-53.
- Schüßer A.**, Schwarzott D., Walker C. **2001**. A new fungal phylum, the Glomeromycota: phylogeny and evolution. *Mycological Research*. 105: 1413-1421.
- Schmidt T.M.**, DeLong E.F., Pace N.R. **1991**. Analysis of a marine picoplankton community by 16S rRNA gene cloning and sequencing. *Journal of Bacteriology*. 173: 4371-8.
- Schmit J.P.**, **Mueller G.M.** **2007**. An estimate of the lower limit of global fungal diversity. *Biodiversity and Conservation*. 16: 99-111.
- Segonzac M.**, De Saint-Laurent M., Casanova B. **1993**. L'énigme du comportement trophique des crevettes Alvinocarididae des sites hydrothermaux de la dorsale médio-atlantique. *Cahiers de biologie marine*. 34:



535-571.

- Shearer C.A.**, Descals E., Kohlmeyer B., Kohlmeyer J., Marvanova L., Padgett D., Porter D., Raja H.A., Schmit J.P., Thornton H.A., Voglymayr H. **2007**. Fungal biodiversity in aquatic habitats. *Biodiversity and Conservation*. 16: 49-67.
- Selosse M.A.**, **Le Tacon F.** **1998**. The land flora: a phototroph-fungus partnership? *Trends in Ecology and Evolution*. 13: 15-20.
- Simon L.**, Bousquet J., Levesque R.C., Lalonde M. **1993**. Origin and diversification of endomycorrhizal fungi and coincidence with vascular land plants. *Nature*. 363: 67-69.
- Singh B.**, Gautam S.K., Verma V., Kumar M., Singh B. **2008**. Metagenomics in animal gastrointestinal ecosystem : Potential biotechnological prospects. *Anaerobe*. 14: 138-144.
- Smit E.**, Leeflang P., Glandorf B., van Elsas J.D., Wernars K. **1999**. Analysis of fungal diversity in the wheat rhizosphere by sequencing of cloned PCR-amplified genes encoding 18S rRNA and temperature gradient gel electrophoresis. *Applied and Environmental Microbiology*. 65: 2614-2621.
- Smith S.E.**, **Read D.J.** **1997**. Mycorrhizal Symbiosis. Academic Press, Cambridge.
- Sogin M.L.**, Morrison H.G., Huber J.A., Welch D.M., Susan M., Huse S.M., Neal P.R., Arrieta J.M., Herndl G.J. **2006**. Microbial diversity in the deep sea and the underexplored "rare biosphere". *Proceedings of the National Academy of Sciences, USA*. 32: 12115-12120.
- Stoeck T.**, **Epstein S.** **2003**. Novel eukaryotic lineages inferred from Small-Subunit rRNA analyses of oxygen-depleted marine environments. *Applied and Environmental Microbiology*. 69: 2657.
- Stoeck T.**, Hayward B., Taylor G.T., Varela R., Epstein S.S. **2006**. A multiple PCR-primer approach to access the microeukaryotic diversity in environmental samples. *Protist*. 157: 31-43.
- Streit W.R.**, **Schmitz R.A.** **2004**. Metagenomics-The key to the uncultured microbes. *Current opinion in microbiology*. 7: 492-498.
- Suenaga H.**, Ohnuk T.I., Miyazaki K. **2007**. Functional screening of a metagenomic library for genes involved in microbial degradation of aromatic compounds. *Environmental Microbiology*. 9: 2289-2297.
- Sugawara H.**, Ogasawara O., Okubo K., Gojobori T., Tateno Y. **2008**. DDBJ with new system and face. *Nucleic Acids Research*. **36**: D22-D24.
- Swofford D.L.** **1991**. PAUP: Phylogenetic Analysis Using Parsimony, Version 3.1 Computer program distributed by the Illinois Natural History Survey, Champaign, Illinois.

## T

- Takishita K.**, Miyake H., Kawato M., Maruyama T. **2005**. Genetic diversity of microbial eukaryotes in anoxic sediment around fumaroles on a submarine caldera floor based on the small-subunit rDNA phylogeny. *Extremophiles*. 9: 185-196.
- Takishita K.**, Tsuchiya M., Reimer J.D., Maruyama T. **2006**. Molecular evidence demonstrating the basidiomycetous fungus *Cryptococcus curvatus* is the dominant microbial eukaryote in sediment at the Kuroshima Knoll methane seep. *Extremophiles*. 10: 165-169.

- Tamura K.**, Dudley J., Nei M., Kumar S. **2007**. MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) Software Version 4.0. *Molecular Biology and Evolution*. 24:1596-1599.
- Tanabe Y.**, O'Donnell K., Saikawa M., Sugiyama J. **2000**. Molecular phylogeny of parasitic Zygomycota (Dimargaritales, Zoopagales) based on nuclear small subunit ribosomal DNA sequences. *Molecular Phylogenetics and Evolution*. 16: 253-262.
- Tanabe Y.**, Saikawa M., Watanabe M.M., Sugiyama J. **2004**. Molecular phylogeny of Zygomycota based on EF1-alpha and RPB1 sequences: limitations and utility of alternative markers to rDNA. *Molecular Phylogenetics and Evolution*. 30: 438-449.
- Tatusov R.L.**, Fedorova N.D., Jackson J.D., Jacobs A.R., Kiryutin B., Koonin E.V., Krylov D.M., Mazumder R., Mekhedov S.L., Nikolskaya A.N., Rao B.S., Smirnov S., Sverdlov A.V., Vasudevan S., Wolf Y.I., Yin J.J., Natale D.A. **2003**. The COG database: an updated version includes eukaryotes. *BMC Bioinformatics*. 4:41.
- Taylor T.N.**, Hass T., Remy W., Kerp H. **1999**. The oldest fossil ascomycetes. *Nature*. 399: 648.
- Taylor J.W.**, Jacobson D.J., Kroken S., Kasuga T., Geiser D.M., Hibbet D.S., Fisher M.C. **2000**. Phylogenetic species recognition and species concept in fungi. *Fungal Genetics and Biology*. 31: 21-32.
- Taylor J.W.**, Spatafora J., O'Donnell K., Lutzoni F., James T., Hibbett D.S., Geiser D., Bruns T.D., Blackwell M. **2004**. The Fungi. In *Assembling the Tree of Life* (Joel Cracraft, Michael J. Donoghue eds). Oxford University Press.
- Taylor T.N.**, Hass H., Kerp H., Krings M., Hanlin R.T. **2005**. Perithecial ascomycetes from the 400 million year old Rhynie chert: an example of ancestral polymorphism. *Mycologia*. 97: 269-285.
- Thompson J.D.**, Higgins D.G, Gibson T.J. **1994**. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Research*. 22: 4673-4680.
- Thompson J.D.**, Gibson T.J., Plewniak F., Jeanmougin F., Higgins D.G. **1997**. The ClustalX windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tool. *Nucleic Acids Research*. 24: 4876-4882.
- Tivey M.K.** **2007**. Generation of seafloor hydrothermal vent fluids and associated mineral deposits. *Oceanography*. 20: 50-65.
- Tor J.M.**, **Lovley D.R.** **2001**. Anaerobic degradation of aromatic compounds coupled to Fe (III) reduction by *Ferroglobus placidus*. *Environmental Microbiology*. 3: 281-287.
- Tor M.J.**, Amend J.P., Lovley D.R. **2003**. Metabolism of organic compounds in anaerobic,hydrothermal sulphate-reducing marine sediments. *Environmental Microbiology*. 5: 583-591.
- Tosi S.**, Casado B., Gerdol R., Caretta G. **2002**. Fungi isolated from Antarctic mosses. *Polar Biology*. 25: 262-268.
- Tringe S.G.**, Von Mering C., Kobayashi A., Salamov A.A., Chen K., Chang H.W., Podar M., Short J.M., Mathur E.J., Detter J.C., Bork P., Hugenholtz P., Rubin E.M. **2005**. Comparative metagenomics of microbial communities. *Science*. 308: 554-557.
- Tsurumi M.**, **Tunncliffe V.** **2003**. Tubeworm-associated communities at hydrothermal vents on the Juan de Fuca Ridge, northeast Pacific. *Deep Sea Research part I*. 50: 611-629.

**Turnbaugh P.J., Gordon J.I. 2008.** An invitation to the marriage of metagenomics and metabolomics. *Cell*. 134: 709-713.

**Tyson G.W., Chapman J., Hugenholtz P., Allen E.E., Ram R.J., Richardson P.M., Solovyev V.V., Rubin E.M., Rokhsar D.S., Banfield J.F. 2004.** Community structure and metabolism through reconstruction of microbial genomes from the environment. *Nature*. 428: 37-43.

## V

**Vandenkoornhuysen P., Leyval C. 1998.** SSU rDNA sequencing and PCR-fingerprinting reveal genetic variation within *Glomus mossae*. *Mycologia* 90: 791-797.

**Vandenkoornhuysen P., Leyval C., Bonnin I. 2001.** High genetic diversity in AM fungi: evidence for recombination events. *Heredity*. 87: 243-253.

**Vandenkoornhuysen P., Baldauf S.L., Straczek J., Leyval C., Young J.P.W. 2002a.** Extensive fungal diversity in plant roots. *Science*. 295: 2051.

**Vandenkoornhuysen P., Husband R., Daniell T.J., Watson I.J., Duck J.M., Fitter A.H., Young J.P.W. 2002b.** Arbuscular mycorrhizal community composition associated with two plant species in a grassland ecosystem. *Molecular Ecology*. 11: 1555-64.

**Vandenkoornhuysen P., Mahé S., Ineson P., Staddon P., Ostle N., Fitter A.H., Young J.P.W. 2007.** Active root-inhabiting microbes identified by rapid incorporation of plant-derived carbon into RNA. *Proceedings of the National Academy of Sciences USA*. 104: 16970-16975.

**Van der Heijden M.A.G., Klironomos J.N., Ursic M., Moutoglou P., Streitwolf-Engel R., Boller T., Wiemken A., Sanders I.R. 1998.** Mycorrhizal fungi diversity determines plant biodiversity, ecosystem variability and productivity. *Nature*. 396: 69-72.

**Van Dover C.L., Fry B. 1994.** Microorganisms as food resources at deep-sea hydrothermal vents. *Limnology and Oceanography*. 39: 51-57.

**Van Dover C.L., Ward M.E., Scott J.L., Underdown J., Anderson A., Gustafson C., Whalen M., Carnegie R.B. 2007.** A fungal epizootic in mussels at a deep-sea hydrothermal vent. *Marine ecology*. 28: 54-52.

**Venter J.C., Remington K., Heidelberg J.F., Halpern A.L., Rusch D., Eisen J.A., Wu D., Paulsen I., Nelson K.E., Nelson W., Fouts D.E., Levy S., Knap A.H., Lomas M.W., Neelson K., White O., Peterson J., Hoffman J., Parsons R., Baden-Tillson H., Pfannkoch C., Rogers Y.H., Smith H.O. 2004.** Environmental genome shotgun sequencing of the Sargasso Sea. *Science*. 304, 66-74.

**Vilgalys R., Hester M. 1990.** Rapid genetic identification and mapping of enzymatically amplified ribosomal DNA from several *Cryptococcus* species. *Journal of Bacteriology*. 172: 4238-4246.

## W

**Wächtershäuser G. 1988.** Before enzymes and templates: theory of surface metabolism. *Microbiological Reviews*. 52: 452-484.

- Wächtershäuser G. 1990.** Evolution of the first metabolic cycle. *Proceedings of the National Academy of Sciences, USA.* 87: 200-204.
- Wainright P.O., Hinkle G., Sogin M.L., Stickel S.K. 1993.** Monophyletic origins of the Metazoa: an evolutionary link with fungi. *Science.* 260: 340-342.
- Wainwright M. 2005.** Oligotrophic Growth of Fungi. In *The Fungal Community: Its Organization And Role In The Ecosystem.* (Dighton J., White Jr J.F., Oudemans P. eds). 23: 643-658.
- Wang D.Y.C., Kumar S., Hedges S.B. 1999.** Divergence time estimates for the early history of animal phyla and the origin of plants, animals and fungi. *Proceedings of the Royal Society of London, Series B- Biological Sciences.* 266: 163-171.
- Whittaker R. H. 1969.** New concepts of kingdoms of organisms. *Science.* 163: 150-160.
- White T.J., Bruns T., Lee S., Taylor J.W. 1990.** Amplification and direct sequencing of fungal ribosomal RNA genes for phylogenetics. In: *PCR protocols: a guide to methods and applications* (Innis M., Gelfand D., Sninsky J., White T. eds). Academic Press, New York. 315-322.
- Woese C., Fox G. 1977.** Phylogenetic structure of the prokaryotic domain: the primary kingdoms. *Proceedings of the National Academy of Sciences USA.* 74: 5088-5090.
- Woese C.R. 1979.** A proposal concerning the origin of life on the planet Earth. *Journal of Molecular Evolution.* 13: 95-101.
- Woese C.R., Kandler O., Wheelis M.L. 1990.** Towards a natural system of organisms: proposal for the domains Archaea, Bacteria, Eukarya. *Proceedings of the National Academy of Sciences, USA.* 91: 6721- 6728.
- Woese C.R. 2000.** Interpreting the universal phylogenetic tree. *Proceedings of the National Academy of Sciences, USA.* 97: 8392-8396.
- Wolff T. 2005.** Composition and endemism of deep sea hydrothermal fauna. *Cahiers de Biologies Marines.* 46: 97-104.
- Woyke T., Teeling H., Ivanova N.N., Huntemann M., Richter M., Gloeckner F.O., Boffelli D., Anderson I.J., Barry K.W., Shapiro H.J., Szeto E., Kyrpides N.C., Musmann M., Amann R., Bergin C., Ruehland C., Rubin E.M., Dubilier N. 2006.** Symbiosis insights through metagenomic analysis of a microbial consortium. *Nature.* 443: 950-955.

## X

- Xu P., Han Y., Wu J., Lv H., Qiu L., Chang R., Jin L., Wang J., Yu A., Chen C., Nan H., Xu X., Wang P., Zhang D., Zhang S., Li W., Chen W. 2007.** Phylogenetic Analysis of the Sequences of rDNA Internal Transcribed Spacer (ITS) of *Phytophthora soja*. *Journal of Genetics and Genomics.* 34: 180-188.

## Y

- Yanagibayashi M., Kato C., Li L., Nogi Y., Inada T., Taira K., Kimura T., Suzuki K., Horikoshi K. 1998.** Changes in bacterial community in Japan Trench sediment during cultivation without decompression. *Deep Sea Research I.* 14: 553-560.

## Z

- Zbinden M.**, Le Bris N., Gaill F., Compere P. **2004**. Distribution of bacteria and associated minerals in the gill chamber of the vent shrimp *Rimicaris exoculata* and related biogeochemical processes. *Marine Ecology Progress Series*. 284: 237-251.
- Zettler L.A.A.**, Gomez F., Zettler E., Keenan B.G., Amils R., Sogin M.L. **2001**. Eukaryotic diversity in Spain's River of Fire. *Journal of Eukaryotic Microbiology*. 48: 293-297.
- Zettler L.A.A.**, Messerli M.A., Laatsch A.D., Smith P.J.S., Sogin M.L. **2003**. From genes to genomes: beyond biodiversity in Spain's Rio Tinto. *Biological Bulletins*. 204: 205-209.
- Zuendorf A.**, Bunge J., Behnke A., Barger K.J.A., Stoeck T. **2006**. Diversity estimates of microeukaryotes below the chemocline of the anoxic Mariager Fjord, Denmark. *FEMS Microbiology Ecology*. 58: 476-491.

## Résumé

Les champignons sont des microorganismes essentiellement connus en écosystèmes aériens, s'étant diversifiés en milieu continental. A partir d'analyses d'horloges moléculaires, nous avons émis l'hypothèse d'une diversification des champignons dans les océans. Par ailleurs, de nombreux auteurs considèrent les écosystèmes hydrothermaux marins profonds comme le Berceau de la Vie : nous avons donc choisi d'étudier ces milieux, très peu exploités en terme de diversité fongique, dans l'espoir de retrouver des champignons ayant conservé des caractères ancestraux de ce Règne. Des analyses de diversité (établies après la création d'une base de données moléculaires dédiée), par PCR indépendantes de la culture (amplification du gène codant l'ARNr18S), ont révélé des organismes se branchant à la base de la phylogénie des champignons (phylum Chytridiomycota), en accord avec notre hypothèse de travail, couplée à une forte diversité spécifique fongique (phyla Chytridiomycota, Basidiomycota, et Ascomycota). Des souches isolées de ces écosystèmes en laboratoire ont également permis de détecter de nombreuses nouvelles espèces de champignons (phylum Ascomycota). Ces 2 approches combinées nous ont ainsi permis de révéler une diversité insoupçonnée dans ces milieux. Des hypothèses évolutives sur la diversification des champignons en milieu marin ont ainsi pu être proposées.

Le second objectif de notre travail était de connaître les rôles de ces organismes au sein de ces écosystèmes. Pour cela, nous avons choisi une approche métagénomique originale : l'échantillon choisi sur la base de la fréquence du gène codant l'ARNr18S et des études de diversité, a été pyroséquencé (GS FLX ; 454 Life Sciences, ROCHE). L'échantillon testé contenait un seul phylotype fongique, constituant une branche profonde des champignons (phylum Chytridiomycota). Six ng d'ADN extraits ont été pyroséquencés en 3 runs indépendants : 168.909 contigs (longueur moyenne : 352 pb) ont été assemblés à partir de 1.441.839 séquences individuelles.

Deux approches ont alors été choisies. La première consistait à analyser les contigs générés, afin d'extraire les fragments de gènes fongiques et de reconstruire leur métabolisme hypothétique. Des hypothèses de fonctionnement métabolique des champignons dans ces milieux, établies sur la base de recherche d'homologies dans les bases de données (BLASTX, ORF finder, KOG,...) ont pu être établies. La seconde approche, basée cette fois sur l'étude des séquences non assemblées, consistait à reconstruire les voies métaboliques du métagénome (BLASTX, mgRAST, ASGARD), et reconstituer la composition taxonomique de notre échantillon (MEGAN). Les résultats de l'analyse métagénomique présentés dans cette thèse, nous ont permis d'émettre des hypothèses sur les modes de vie des champignons en écosystème hydrothermal et nous ont également permis de reconstruire avec succès les métabolismes bactériens dominants dans l'écosystème. Les limites de cette analyse, ainsi que les nombreuses perspectives qu'offrent ce travail sont également développées.

**Mots clés :** Diversité fongique, ARNr18S, PHYMYCO-DB, Métagénomique, Pyroséquencage, Annotations fonctionnelle, Reconstructions métaboliques.

## Summary

Fungi are known as terrestrial microorganisms which have diversified in land. From molecular clock estimates we hypothesized their diversification in oceans. A current theory of emergence of life on Earth suggests that the deep marine hydrothermal environment might be the Cradle of Life. Thus, we have tested and analyzed the diversity of fungi in this particular ecosystem, in the aim to detect fungi displaying ancestral traits within the fungal Kingdom. Diversity analyses (established after the creation of a personal dedicated molecular database), were performed from ciPCR (amplification of a fragment of the gene encoding 18SrRNA), allowed to highlight the presence of a new branch forming one of the most ancient evolutionary lineage of fungi (phylum Chytridiomycota), in agreement with our working hypothesis, along with a large fungal diversity (phyla Chytridiomycota, Basidiomycota, and Ascomycota). Fungal strains isolated in lab from these environments allowed us to detect new fungal species (phylum Ascomycota). Both strategies revealed an unsuspected fungal diversity. Evolutionary hypotheses about fungal diversification in marine ecosystems have been proposed.

The second purpose of this work was to understand the ecological functions of these fungi in these ecosystems. In this aim, we chose an original metagenomic approach: one particular sample, selected on the basis of both fungal diversity and 18SrRNA gene frequency (qPCR assays), was pyrosequenced (GS FLX, 454 Life Sciences, ROCHE). Sample we analysed contained one phylotype forming an old fungal lineage (phylum Chytridiomycota). Six ng of DNA were pyrosequenced in 3 independent runs : 168.909 contigs (length mean : 352 bp) were assembled from 1.441.839 individual reads.

Two different strategies were chosen to analyze the metagenome. The first one was to analyze contigs in depth, to extract and annotate fungal genes in the aim to reconstruct the hypothetical fungal metabolism. Genes functions predictions relied on homologies searches in public database (BLASTX, ORF finder, KOG...).

The second strategy focused on unassembled reads, to reconstruct the different metabolisms of the metagenome (BLASTX, mgRAST server, and ASGARD software), and to better understand the taxonomic composition of our sample (using MEGAN). The results from these analyses, shown in this thesis, have allowed us to suggest hypotheses about fungal lifestyle in its environment, and allowed to successfully reconstruct dominant bacterial metabolisms in the ecosystem. Limits of this analysis, as well as the numerous perspectives of this work were also developed.

**Key words** : Fungal diversity, 18SrRNA, PHYMYCO-DB, Metagenomic, Pyrosequencing, Functional annotations, Metabolic reconstructions.