



**HAL**  
open science

# Diversité des branches évolutives basales du règne des champignons dans les écosystèmes hydrothermaux marins profonds

Stéphane Mahé

► **To cite this version:**

Stéphane Mahé. Diversité des branches évolutives basales du règne des champignons dans les écosystèmes hydrothermaux marins profonds. Biodiversité et Ecologie. Université Rennes 1, 2012. Français. NNT: . tel-00797898

**HAL Id: tel-00797898**

**<https://theses.hal.science/tel-00797898>**

Submitted on 7 Mar 2013

**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.



**THÈSE / UNIVERSITÉ DE RENNES 1**  
*sous le sceau de l'Université Européenne de Bretagne*

pour le grade de  
**DOCTEUR DE L'UNIVERSITÉ DE RENNES 1**

*Mention : Biologie*

**Ecole doctorale Vie-Agro-Santé**

présentée par

**Stéphane Mahé**

Préparée à l'unité de recherche UMR 6553  
EcoBio  
UFR Sciences de la vie et de l'environnement

---

**Diversité des  
branches évolutives  
basales du règne  
des champignons  
dans les écosystèmes  
hydrothermaux  
marins profonds**

**Thèse soutenue à Rennes  
le 30 Mai 2012**

devant le jury composé de :

**Laure Guillou**

Directeur de recherche, Roscoff / *rapporteur*

**Sébastien Duperron**

Maître de conférence, UPMC CNRS MNHN / *rapporteur*

**Marc-André Sélosse**

Professeur, Université Montpellier II / *examineur*

**Théodore Bouchez**

Chercheur à Irstea-Antony / *examineur*

**Georges Barbier**

Professeur, ESMISAB Brest / *examineur*

**Philippe Vandenkoornhuyse**

Professeur, Université Rennes1 / *directeur de thèse*



## Résumé

---

Les champignons sont des organismes hétérotrophes ubiquistes jouant des rôles pivots dans de nombreux écosystèmes (*e.g.* décomposeurs, symbiontes) et formant une lignée eucaryote majeure. Les Chytridiomycota constituent les branches basales du règne des Champignons, soit une position critique pour la compréhension de la radiation évolutive fongique. Or, ce groupe a été peu étudié, ce qui ne permet qu'une résolution partielle de ces relations évolutives. Ici, nous décrivons la diversité fongique, en ciblant principalement les Chytridiomycota via des analyses de génomique environnementale. Des échantillons de diverses natures ont été collectés au niveau de sources hydrothermales marines profondes qui sont connues pour être des hotspots de diversité, avec un fort taux d'endémisme. Pour l'analyse des séquences moléculaires obtenues, une base de données (PHYMYCO-DB) portant sur des marqueurs moléculaires fiables et dédiés à la phylogénie fongique, a été créée puis mise en ligne. Des outils moléculaires développés au cours de cette thèse ont permis de récupérer six phylotypes Chytridiomycota dont quatre produisent des branches non-décrites à ce jour. De plus, la diversité obtenue n'est pas limitée au clade des Opisthokonta (*i.e.* principalement les règnes des Animaux et des Champignons) puisqu'il a également été observé un phylotype Apusozoa et deux phylotypes ayant une position indéfinie à la base des Unikonta. Ces résultats offrent des perspectives pour la description de nouveaux organismes via le séquençage de génomes ou l'imagerie. Ces organismes sont également prometteurs pour la résolution des relations évolutives chez les Opisthokonta, les Unikonta, voire les Bikonta.

## Abstract

---

Fungi are ubiquitous heterotrophs that play pivotal roles in numerous ecosystems (*e.g.* decomposers, symbionts) and form one of the major eukaryotic lineages. Among them the Chytridiomycota phylum constitutes the early diverging branches of the fungal kingdom and so can give critical informations for the understanding of the fungal evolutionary radiation. However, this group has been overlooked and consequently the roots of the fungal kingdom remain poorly resolved. Here, we describe the fungal diversity with a special interest in Chytridiomycota, via environmental genomic analyses. Different kinds of samples were collected at deep-sea hydrothermal vents which are generally considered as hotspots of diversity harbouring endemic organisms and also encompassing novel fungal lineages. To process the data collected, we developed an online user-friendly database, called PHMYCO-DB, containing quality-checked molecular markers dedicated to fungal phylogenetic studies. The molecular experimental tools designed here allowed the description of six phlotypes belonging to Chytridiomycota, four of them producing branches never described before. Also, unexpectedly, the diversity retrieved goes beyond the Opisthokonta clade boundaries (*i.e.* Animals, Fungi ...) as one phylotype of Apusozoa and two phlotypes of very diverging eukaryotes with undefined positions anterior to the known Unikonta, were obtained. These results offer perspectives for the description of new organisms via genome sequencing or imaging, but also they are promising for the resolution of the complex deep evolutionary relationships in Opisthokonta, Unikonta and Bikonta.

## Table des matières

---

Avant-Propos	p.9
Revue Bibliographique	p.11
Le règne des champignons	p.13
I – Caractères généraux des champignons	p.14
A – Structure	p.14
B – Morphologie	p.14
C – Ecologie	p.16
II – Contexte évolutif	p.17
A – Le règne des champignons dans le groupe des Opisthokonta	p.17
B – Les Apusozoa : un phylum au positionnement ambiguë	p.20
C – Origine précambrienne des champignons	p.21
III – Diversité taxonomique	p.25
A – Description de nouvelles espèces	p.25
B – Estimations de la diversité	p.27
C – Diversité par des approches moléculaires	p.27
IV – Arbre de la vie du Règne des Champignons	p.29
A – Glomeromycota, Dikarya : des clades monophylétiques	p.29
B – Les lignées basales : des clades peu résolus	p.31
C – Un nouveau phylum : les Cryptomycota	p.32
D – Les lignées basales caractérisées par la présence d'un flagelle	p.35
E – Hypothèses évolutives sur la diversification des champignons	p.37

V – Les Chytridiomycota	p.37
A – Biologie	p.37
B – Profil écologique	p.41
1) Les chytrides en eau douce	p.43
2) Les chytrides en milieu marin	p.44
VI – Les champignons marins	p.46
A – Profil écologique	p.46
B – Fonctions écologiques	p.47
C – Applications biotechnologiques	p.49
D – Les champignons dans les inventaires eucaryotes	p.49
E – Les champignons dans les inventaires fongiques	p.54
1) Diversité par des approches culturelles	p.54
2) Diversité par des approches moléculaires	p.56
3) Diversité des champignons supérieurs	p.58
4) Diversité des lignées basales	p.58
Présentation des objectifs	p.62
Matériel et Méthodes	p.65
I – Création d'une base de données dédiée aux analyses phylogénétiques fongiques	p.67
II – Design d'amorces moléculaires	p.71
III – Analyse de la diversité des Chytridiomycota en milieu hydrothermal	p.76
A – Echantillons	p.76
B – Génomique environnementale	p.79
1) Extraction de l'ADN génomique total	p.79

2) Amplifications du gène de l'ARNr 18S avec MH2/SR8R	p.82
3) Clonage des amplicons	p.85
IV – Analyses Phylogénétiques	p.85
A – Alignement	p.86
B – Méthodes de reconstructions phylogénétiques	p.86
1) Neighbor Joining	p.86
2) Méthode de maximum de vraisemblance	p.87
3) Robustesse de l'arbre inféré	p.87
4) Méthode Bayésienne	p.87
Chapitre 1	p.89
PHYMYCO-DB : a curated database for analyses of fungal diversity and evolution	
Chapitre 2	p.117
Fungal diversity in deep-sea hydrothermal ecosystems	
Chapitre 3	p.129
Early diverging branches of the fungal kingdom and opisthokonta clade in deep-sea hydrothermal vents	
Discussion – Perspectives	p.155
Références bibliographiques	p.167



Annexes

p.189

- Active root-inhabiting microbes identified by rapid incorporation of plant-derived carbon into RNA
- Integration of molecular functions at the ecosystemic level: breakthroughs and future goals of environmental genomics and post-genomics

## Avant-propos

---

La représentation des relations évolutives entre les organismes vivants sous forme d'arbre est communément utilisée de nos jours. Les premières phylogénies étaient basées par exemple, sur des caractères embryologiques, morphologiques ou des modes de nutrition. Cependant, tous les organismes ne présentent pas tous ces caractères, à commencer par les prokaryotes. Il a donc été envisagé d'utiliser des marqueurs moléculaires. Ainsi, des représentations de l'arbre de la vie ont pu être obtenues. L'une d'elle présente les organismes vivants regroupés en trois domaines : Bacteria, Archaea et Eukaryota (Baldauf *et al.*, 2004). Parmi les Eukaryota, le groupe des Opisthokonta comprend principalement le règne des animaux et le règne des champignons. Il a d'ailleurs été montré que ces deux clades sont frères (Baldauf & Palmer 1993). Au sein des Opisthokonta des zones d'ombres existent dans les lignées basales des champignons et donc au niveau de la divergence des champignons et des animaux. Ces incertitudes nécessitent l'obtention de séquences moléculaires d'organismes se positionnant au niveau de ces lignées basales pour pouvoir être résolues. En effet, pour les lignées basales des champignons, les méthodes d'observation et d'isolation sont limitées pour la description de nouvelles espèces. Et la diversité de ces lignées basales est maintenant plus accessible par les approches moléculaires mais son étude reste peu avancée par rapport aux lignées supérieures du règne des champignons. Ainsi, la proportion d'espèces décrites pour les lignées basales (*i.e.* Chytridiomycota et Zygomycota) est largement inférieure à celle des champignons supérieurs (*i.e.* Basidiomycota, Ascomycota et Glomeromycota).

Les champignons ont surtout été étudiés dans les écosystèmes continentaux et il a été montré qu'ils y jouaient un rôle clé dans le cyclage des éléments par leur caractère hétérotrophe et dans la composition écosystémique par leurs relations symbiotiques de diverses natures. En milieu marin, ces études sont beaucoup moins avancées et pour le moment, elles s'attachent principalement à décrire leur diversité. Plusieurs méthodes ont été employées pour rendre compte

de cette diversité, mais des caractéristiques inhérentes à ces méthodes ainsi qu'aux champignons rendent difficile l'obtention d'un inventaire exhaustif des groupes fongiques présents. De plus, la majorité des études ont porté sur les champignons supérieurs, laissant la diversité des lignées basales peu décrite. Récemment une étude sur la diversité fongique dans les milieux hydrothermaux marins profonds a mis en évidence l'existence de séquences distantes de la diversité connue des Chytridiomycota, la lignée ayant divergé en première du règne des champignons.

Les objectifs de ce travail de thèse sont :

- (i) de décrire la diversité des Chytridiomycota des milieux hydrothermaux marins profonds par l'approche de la génomique environnementale,
- (ii) d'analyser les relations évolutives des lignées basales des champignons par une approche phylogénétique basée sur les séquences environnementales obtenues.

Nous faisons l'hypothèse que la diversité des Chytridiomycota est beaucoup plus grande que celle décrite précédemment au niveau des sites hydrothermaux et que l'analyse de ce phylum basal du règne des champignons peut aider à la compréhension de la diversification des champignons voire des Opisthokonta.

REVUE

BIBLIOGRAPHIQUE



La synthèse des connaissances dans ce domaine, proposée ici, porte également un regard critique sur les travaux effectués pour mettre en lumière la frontière des connaissances. Dans ce chapitre, après une brève description des champignons, j'aborde la matrice des relations évolutives dans laquelle ils se trouvent et les replace dans un contexte temporel. La suite fera un état des lieux de la description des nouvelles espèces et de l'ampleur de la diversité encore inexplorée, avec une présentation de la diversité fongique telle qu'elle est connue à ce jour, avec un intérêt particulier pour les lignées basales. S'ensuivra une présentation de l'état des connaissances sur le phylum des Chytridiomycota, le premier à avoir divergé du règne des champignons.

## Le règne des champignons

---

### I – Caractères généraux des champignons

Les champignons constituent une des lignées les plus diverses des Eukaryota (Cannon, 1997 ; Blackwell, 2011) et sont perçus comme un des groupes les plus importants en raison de leurs fonctions vitales aux écosystèmes (Mueller & Schmit, 2007). De plus, ils ont un rôle économique important en tant que producteurs de biomolécules utilisées en pharmacologie et en recherche industrielle (Schmit & Mueller, 2007 ; Stajich *et al.*, 2009). Toutefois, leur étude a surtout été dirigée vers les écosystèmes continentaux (*e.g.* Richards *et al.*, 2012) et les groupes à intérêts économiques (Powell, 1993), laissant une partie de leur diversité encore insondée.

#### A – Structure

Une remarquable variété de morphologies (*Cf.* Figure 1) et de cycles de vie est observée chez les champignons (Redecker, 2002). Ils présentent principalement deux types de structures : la forme levure unicellulaire et la forme mycélienne pluricellulaire constituée d'hyphes. Les champignons supérieurs Ascomycota et Basidiomycota ont des hyphes septés, c'est-à-dire qu'ils comportent des cloisons tandis que ceux de la plupart des autres champignons sont coenocytiques (*i.e.* sans septum) (Stajich *et al.*, 2009). Ces champignons supérieurs se distinguent également par la production de structures de fructification macroscopiques (Redecker, 2002).

#### B – Reproduction

Les champignons peuvent se reproduire de manière sexuée ou asexuée. La forme sexuée est

appelée téléomorphe et la forme asexuée est appelée anamorphe. D'un point de vue taxonomique, le *Code International de Nomenclature Botanique* (ICBN ; McNeill *et al.*, 2006) accepte l'utilisation de noms différents pour chacune des deux phases, toutefois lorsque les deux phases sont connues, le nom du téléomorphe est retenu. Un des moyens de reproduction repose sur la production de spores sexués ou asexués qui peuvent prendre des formes différentes, comme les zygospores et sporangiospores chez les Zygomycota (White *et al.*, 2006) et les zoospores chez les Chytridiomycota (Gleason & Lilje, 2009). La partie V consacrée aux Chytridiomycota développe plus largement les caractéristiques propres aux zoospores. Les différents groupes de champignons présentent de multiples mécanismes de dispersion de leurs spores résultant d'une adaptation à leur environnement (Lutzoni *et al.*, 2004).

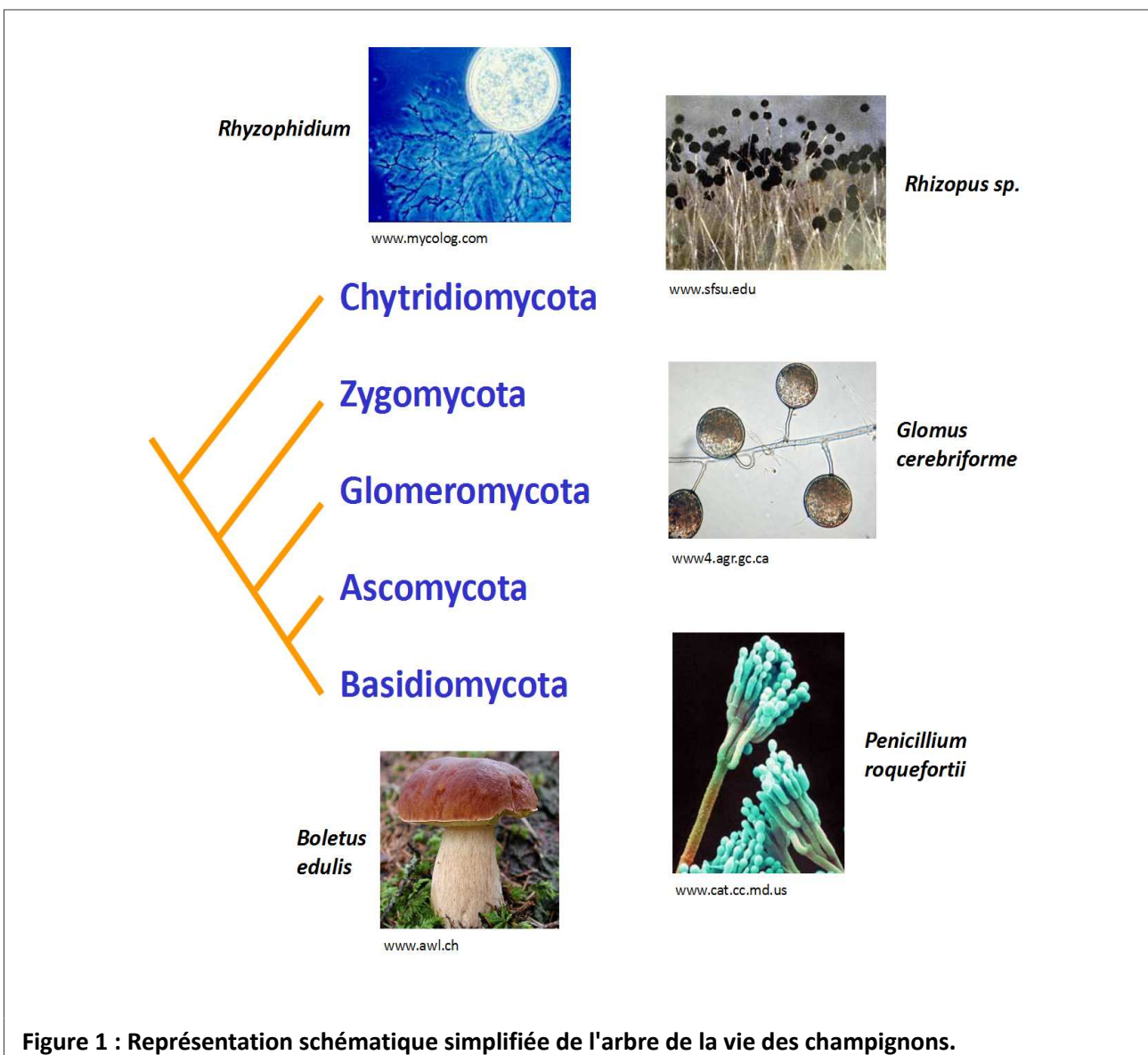


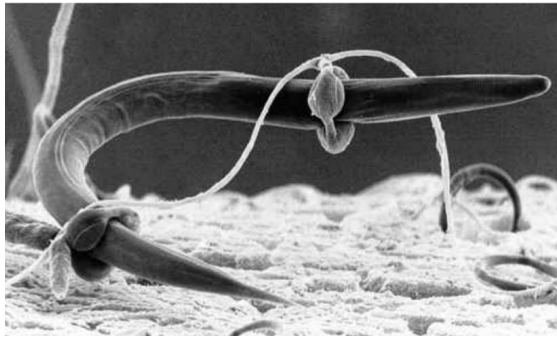
Figure 1 : Représentation schématique simplifiée de l'arbre de la vie des champignons.



## C – Ecologie

D'un point de vue écologique, les champignons sont ubiquistes, ils ont été observés dans de nombreuses régions du globe que ce soit en milieux continental ou océanique (Cf. VI – Les champignons marins). Par exemple dans les sols, les hyphes comptent pour la plus grande fraction de la biomasse (Wardle & Van der Putten, 2002). Ce sont des organismes hétérotrophes, *i.e.* qui se nourrissent de composés organiques, et ils ont principalement été décrits comme étant saprophytes. Ils se nourrissent par osmotrophie (Richards *et al.*, 2012), *i.e.* par absorption, au travers de leur paroi riche en cellulose et en chitine. En effet, ils sécrètent des enzymes pouvant cliver des polymères organiques complexes en des formes plus simples, qu'ils peuvent ensuite assimiler (Peay *et al.*, 2008 ; Richards *et al.*, 2012). De plus, leur fonction de décomposition de débris d'origine végétale (*e.g.* cellulose, lignine) ou d'origine animale (*e.g.* kératine, chitine) permet de relarguer dans l'écosystème des molécules utilisables par un plus grand nombre d'organismes (Hyde *et al.*, 2007). Ils jouent ainsi un rôle central dans le cycle du carbone et le recyclage des nutriments (Peay *et al.*, 2008) et se trouvent donc impliqués dans de nombreux réseaux trophiques (Wardle & Van der Putten, 2002). Une autre fonction incontournable des champignons est celle provenant de leur association avec les plantes. Par exemple, la plupart des plantes terrestres sont en relation mutualiste avec des Glomeromycota pour former les mycorhizes à arbuscules (*e.g.* Parniske, 2008). Le réseau d'hyphes des champignons qui prospecte le sol à partir des racines permet entre autre, la nutrition minérale et hydrique des plantes. En retour, les champignons reçoivent du carbone organique (Smith *et al.*, 2010). D'autres types de symbioses avec des organismes végétaux et animaux sont également observés. Par exemple, les champignons ont un rôle non négligeable en tant que parasites ou pathogènes. Ces fonctions sont les plus répandues et certaines autres, moins connues, ne sont pas pour autant moins importantes. Par exemple, celle du prédateur *Arthrobotrys* (Cf. Figure 2), un champignon nématophage qui pourrait être utilisé comme agent de biocontrôle d'un nématode parasite de plantes (Kerry, 2000).

Les champignons montrent donc un large panel de fonctions écosystémiques, surtout étudiées dans les milieux continentaux.



**Figure 2 : Nématode capturé par le champignon prédateur *Arthrobotrys anthonia*.** Photo de Edwig Triantaphyllou extraite de Peay *et al.* (2008).

Au cours de leur évolution, ils ont colonisé de nombreux habitats, et ont subis plusieurs phases importantes de radiation. Avant de poursuivre avec leur diversité taxonomique, nous allons nous intéresser à la place qu'ils occupent au sein de l'arbre de la vie, c'est-à-dire observer quels organismes vivants sont leurs plus proches parents.

## II – Contexte évolutif

### A – Le règne des champignons dans le groupe des Opisthokonta

La lignée des champignons et la lignée des animaux (Metazoa) partagent une histoire évolutive commune avant leur diversification (Cf. Figure 3). De nombreuses études se sont attachées à résoudre la relation qu'elles ont l'une vis-à-vis de l'autre et leurs positionnements au sein des Eukaryota, en utilisant diverses combinaisons de marqueurs moléculaires. Ainsi, Wainright *et al.* (1993) ont montré une plus grande proximité du règne des champignons avec les Metazoa plutôt qu'avec le règne des plantes, en utilisant le gène de la petite sous-unité ribosomique (gène de l'ARNr 18S). Ce résultat est également obtenu la même année par Baldauf & Palmer (1993), à l'aide de séquences protéiques dont celles du facteur d'élongation 1 $\alpha$  (EF-1 $\alpha$ ). Plus récemment, l'utilisation de grands jeux de données a confirmé ce regroupement (*e.g.* Burki *et al.*, 2007 ; Hampl *et al.*, 2009). En association avec d'autres groupes, les règnes des champignons et des animaux

forment les Opisthokonta (Cavalier-Smith & Chao, 1995 ; Adl *et al.*, 2005). Ceux-ci ne présentent pas de caractères qui leur sont propres mais la présence d'un flagelle inséré postérieurement et la forme aplatie des crêtes mitochondriales sont observées chez de nombreuses espèces et pourraient être des caractères ancestraux du groupe (Carr & Baldauf, 2011). Les Opisthokonta se divisent en Holozoa et Holomycota. Les Holozoa regroupent les Metazoa, son clade frère les Choanoflagellida, les Ichthyosporea et quatre protistes *incertae sedis* (*i.e.* qui ne sont pas assignés à un niveau taxonomique supérieur) comme *Corallochytrium limacisporum*, *Capsaspora owczarzaki*, *Ministeria* et les Aphelidea, pour ces derniers aucune séquence nucléotidique n'est encore disponible (Lang *et al.*, 2002 ; Ruiz-Trillo *et al.*, 2004 ; Steenkamp *et al.*, 2006; Carr & Baldauf, 2011) (*Cf.* Figure 4).

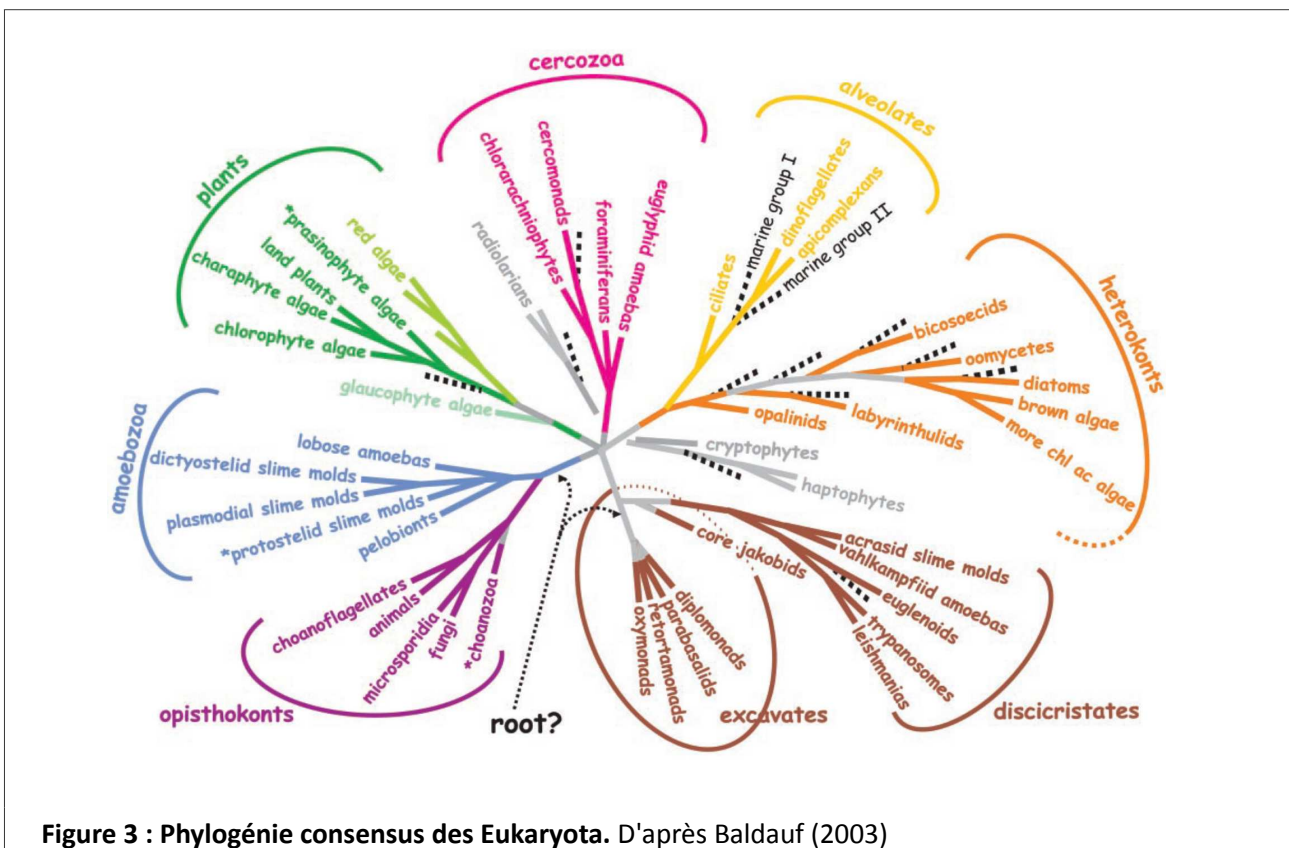
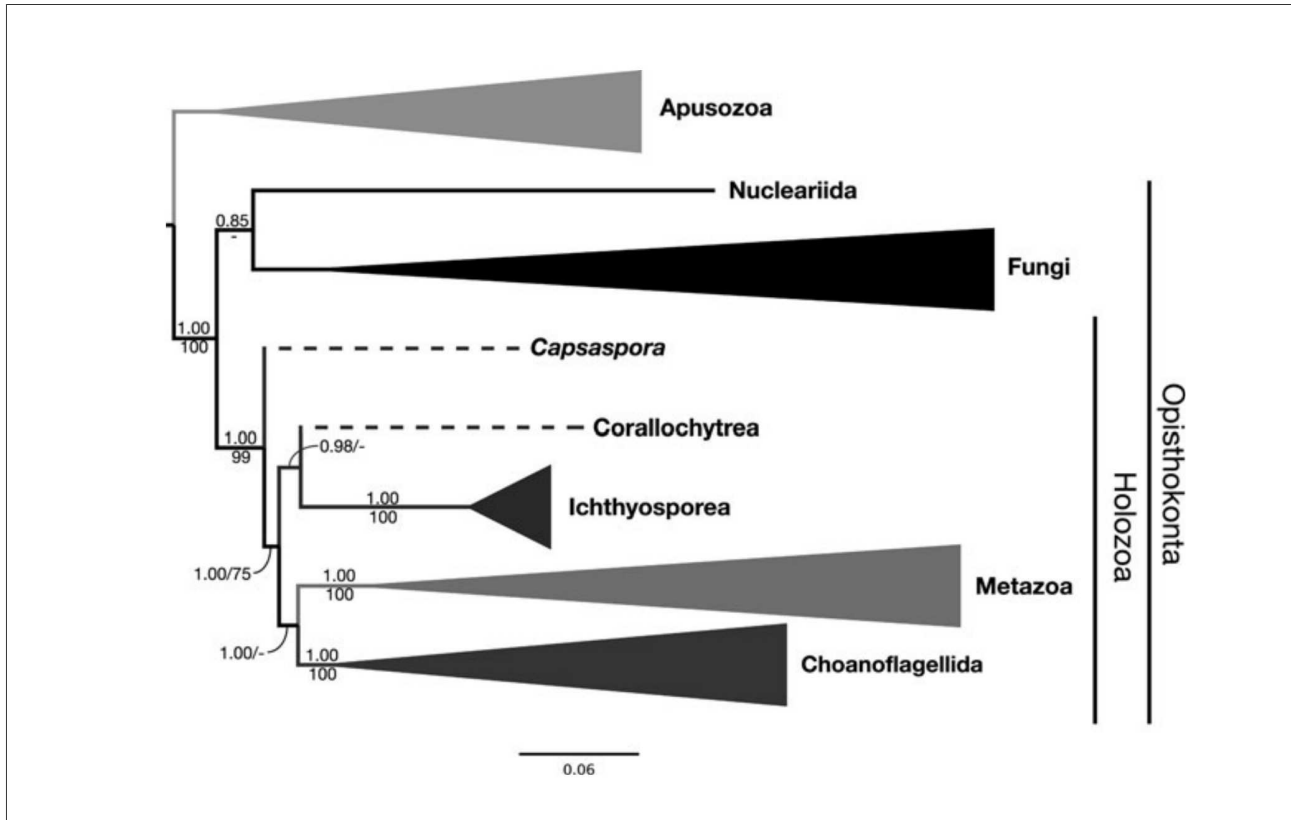


Figure 3 : Phylogénie consensus des Eukaryota. D'après Baldauf (2003)



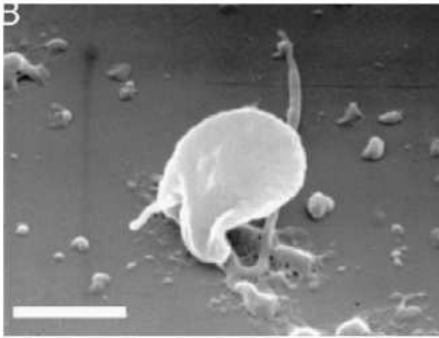
**Figure 4 : Figure simplifiée d'une phylogénie des Opisthokonta.** Quatre gènes ont été utilisés : la grande et la petite sous-unités ribosomiques (18S et 28S),  $\alpha$ -tubulin (*tubA*) et la protéine de choc thermique 90-kDa (*hsp90*). Les valeurs des probabilités postérieures d'inférence Bayésienne et les pourcentages de bootstrap du maximum de vraisemblance sont données au-dessus et en-dessous des branches respectivement. La taille des branches est proportionnelle au nombre de substitutions nucléotidiques par site selon l'échelle indiquée (0.06). Les lignes en pointillées correspondent aux taxa dont la position phylogénétique est incertaine. Figure extraite de Carr & Baldauf (2011).

Quant aux Holomycota, ils sont composés du règne des champignons et des Nucleariida (Steenkamp *et al.*, 2006 ; Lara *et al.*, 2010). Les Nucleariida ont été déplacés des Amoebozoa (Cf. Figure 3) aux Opisthokonta suite à des analyses phylogénétiques multi-gènes robustes (*e.g.* Moreira *et al.*, 2007 ; Steenkamp *et al.*, 2006 ; Liu *et al.*, 2009) alors que leur positionnement dans les Opisthokonta basé seulement sur le gène de l'ARNr 18S est faiblement supporté (Amaral-Zettler *et al.*, 2001). Ce sont des protistes amoeboïdes, non flagellés, émettant des filopodes (Adl *et al.*, 2005). Seules les lignées basales des champignons étant flagellées, la perte du flagelle semble s'être produite de manière indépendante chez les Nucleariida et les champignons (Carr & Baldauf, 2011). Par contre, des caractères potentiellement hérités d'un ancêtre commun sont

présents chez les Nucleariida et chez les champignons des lignées basales comme les déplacements amoéboïdes de certains zoospores (Gleason & Lilje, 2009). Récemment, *Fonticula alba*, anciennement *incertae sedis* Eumycetozoa (Adl *et al.*, 2005), a été repositionné comme taxon frère des Nucleariida d'après une phylogénie robuste basée sur cinq gènes : 18S, actine,  $\beta$ -tubuline, facteur d'élongation 1 $\alpha$  et protéine de choc thermique 70-kDa (Brown *et al.*, 2009). C'est la seule espèce du genre *Fonticula* et elle n'a été isolée qu'une seule fois. Elle est caractérisée par la formation d'une agrégation multicellulaire de forme conique composée de cellules amoéboïdes et qui produit des spores (Brown *et al.*, 2009). Enfin, les Opisthokonta constituent le groupe frère des Amoebozoa, et cet ensemble forme les Unikonta. Tous les autres eucaryotes sont regroupés dans les Bikonta (Cavalier-Smith, 2009). Toutefois, un certain nombre de protistes eucaryotes ne peuvent pas, pour le moment, être placés sans ambiguïté dans l'une ou l'autre de ces lignées majeures des Eukaryota et sont considérés comme des *incertae sedis* Eukaryota (Adl *et al.*, 2005).

## B – Les Apusozoa : un phylum au positionnement ambiguë

Le phylum des Apusozoa est une de ces lignées au positionnement ambiguë. En effet, son placement au sein des Bikonta ou des Unikonta n'a pu être déterminé de manière robuste avec la protéine de choc thermique 90-kDa (*hsp90*) (Stechmann & Cavalier-Smith, 2003) et le gène de l'ARNr 18S combiné au gène de la grande sous-unité ribosomique (ARNr 28S) (Moreira *et al.*, 2007). Une seule des études avec le gène de l'ARNr 18S (Cavalier-Smith & Chao, 1995) et une étude multi-gène (ARNr 18S, ARNr 28S, actine,  $\beta$ -tubuline,  $\alpha$ -tubuline et *hsp90*) supportent de manière robuste l'hypothèse des Apusozoa en clade frère des Opisthokonta (Kim *et al.*, 2006). Le phylum des Apusozoa est constitué de deux ordres : les Planomonadida (Cavalier-Smith *et al.*, 2008) et les Apusomonadida (Karpov & Mylnikov, 1989 dans Cavalier-Smith & Chao, 2010). Cavalier-Smith & Chao (2010) montrent par une analyse basée sur le gène de l'ARNr 18S un clade Opisthokonta/Apusomonadida avec un positionnement paraphylétique des Planomonadida faiblement soutenu.



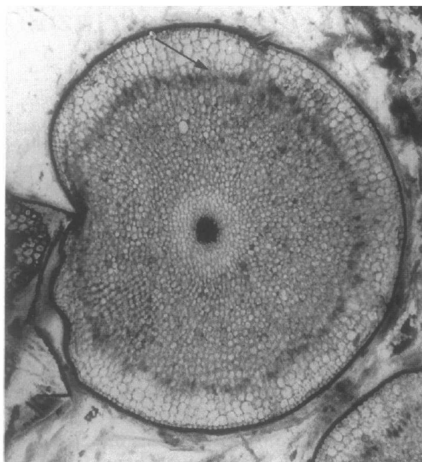
**Figure 5 : Photo de *Planomonas mylnikovi*.** Le flagelle postérieur est épais tandis que le flagelle antérieur plus fin est plus court. Echelle 1 $\mu$ m. Photo au microscope électronique à balayage de A.P. Mylnikov extraite de Cavalier-Smith *et al.* (2008).

Les Apusozoa (*Cf.* Figure 5) sont des organismes hétérotrophes libres possédant deux flagelles : un dirigé vers l'avant et l'autre orienté vers l'arrière (Adl *et al.*, 2005). Ils sont habituellement rencontrés en eau douce, en milieu marin et dans les sols. Les Apusozoa ont un statut essentiel pour la compréhension de la diversification des Eukaryota ainsi que la bifurcation entre les Unikonta et les Bikonta (Moreira *et al.*, 2007). De manière générale, une vision plus exhaustive de la diversité de ces microorganismes basaux accompagnée de l'acquisition de données moléculaires est nécessaire et permettrait de mieux résoudre les relations évolutives existantes (Carr & Baldauf, 2011).

### C – Origine précambrienne des champignons

La première datation de la radiation évolutive d'une lignée de champignons a été calculée pour comparer la diversification des champignons mycorhiziens à arbuscules (Glomeromycota) avec la colonisation des milieux terrestres par les plantes (Simon *et al.*, 1993). Cette datation repose sur l'horloge moléculaire, c'est-à-dire un taux, constant à long terme, de substitutions d'acides aminés ou de nucléotides entre deux taxa ou clades. Il s'agit donc de calibrer des arbres phylogénétiques à l'aide de données paléontologiques (fossiles) afin d'estimer un âge absolu de la divergence entre des taxa ou clades. Simon *et al.* (1993) ont utilisé 12 séquences du gène de l'ARNr 18S, mettant ainsi en évidence une diversification des Glomeromycota il y a 400 millions d'années. Un fossile de champignon mycorhizien à arbuscules associé à une plante (*Cf.* Figure 6) et retrouvé

dans un dépôt sédimentaire siliceux du début de Dévonien à Rhynie en Ecosse, (Remy *et al.*, 1994) confirme leur présence à cette période et indique que le mutualisme pourrait avoir joué un rôle important lors de la conquête du milieu terrestre par les plantes (*e.g.* Selosse & Le Tacon, 1998). En effet, cette association unit des capacités 'pré-adaptées' à la vie terrestre, couplant ainsi l'utilisation de substances minérales non-solubles par les champignons, à la photosynthèse réalisée par les plantes (Selosse & Le Tacon, 1998). Cette symbiose mutualiste pourrait être beaucoup plus ancienne puisque des fossiles de Glomales datant du milieu de l'Ordovicien, soit 55 à 60 millions d'années auparavant, ont été mis en évidence dans des dépôts sédimentaires d'origine marine peu profonds (Redecker *et al.*, 2000).



**Figure 6 : Coupe transversale d'*Aglaophyton major*, en association avec un *Glomeromycota*.** Cette plante fossile mycorhizée, âgée de 400 millions d'années a été mise en évidence dans le dépôt sédimentaire siliceux de Rhynie, en Ecosse. La flèche indique des cellules contenant les arbuscules du champignons. Photo extraite de Remy *et al.* (1994).

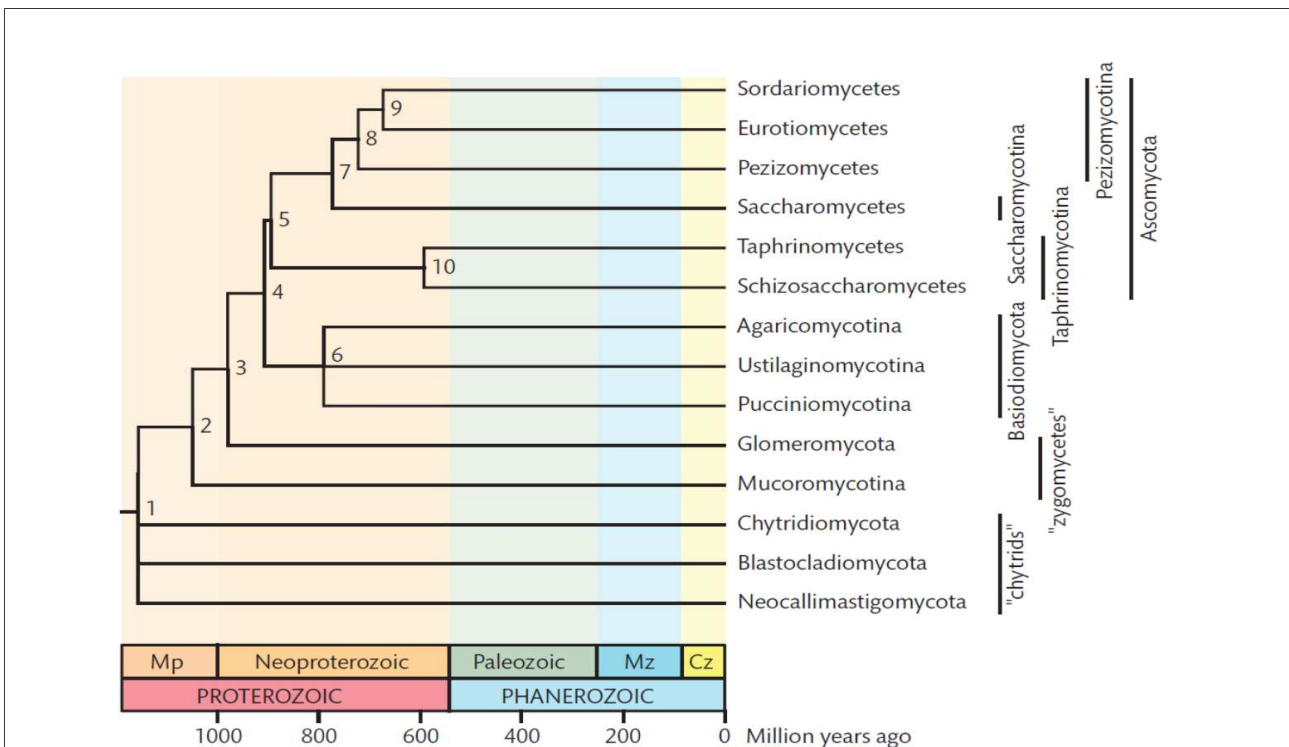
La calibration paléontologique représente une des difficultés de la datation par les calculs d'horloge moléculaire. La dégradation des structures liée à une mauvaise conservation, la taille microscopique de la plupart des champignons et le manque de paléontologues formés en mycologie rendent difficile la description de fossiles fongiques qui peuvent être confondus avec des cyanobactéries, des algues et des Oomycota (Redecker *et al.*, 2000 ; Taylor *et al.*, 2004 dans Blair, 2009). De plus, certaines formations géologiques semblent être plus favorables que d'autres à la conservation de fossiles comme le dépôt sédimentaire siliceux de Rhynie qui comporte de nombreux fossiles fongiques, par exemple un Chytridiomycota (Taylor *et al.*, 1992) ainsi que le plus

ancien fossile d'Ascomycota, âgés de 400 millions d'années (Taylor *et al.*, 1999). Enfin, il est à noter qu'une des propriétés intrinsèques de la datation d'un fossile est de donner un âge minimum à une divergence entre deux taxa, soit un âge minimum à un ancêtre commun (Benton & Ayala, 2003). Les données paléontologiques ont donc tendance à fournir des datations récentes en lien avec les collections de fossiles incomplètes, contrairement aux données moléculaires qui en général dévient les estimations vers des âges plus anciens.

Au point de vue moléculaire, des différences de taux de substitutions entre des lignées ou entre des marqueurs phylogénétiques peuvent biaiser l'estimation si elles ne sont pas prises en compte. Dans le premier cas, Drummond *et al.*, (2006) propose par exemple d'utiliser une méthode bayésienne pour relâcher l'horloge moléculaire : les taux de substitution sont inférés à partir d'une distribution statistique basée sur les points de calibration, ils peuvent donc être différents selon les branches de la phylogénie. Dans le second cas, Heckman *et al.* (2001) ont réduit les biais liés à l'utilisation de plusieurs gènes selon deux approches : en faisant une moyenne des temps de divergence estimés par les différents gènes et en pondérant par leur taille les gènes concaténés, avant de faire la moyenne des temps obtenus. L'étude de Heckman *et al.* (2001) est la première analyse à estimer les âges des divergences fongiques en utilisant des séquences protéiques de plusieurs gènes. Les résultats obtenus sont controversés, entre autres car ils proposent des divergences très anciennes, bien plus anciennes que les fossiles retrouvés (*e.g.* Taylor & Berbee, 2006 ; Sanderson, 2003). Les divergences fongiques majeures sont estimées s'être produites il y a 966 à 1 458 millions d'années, et l'origine des plantes vasculaires, il y a 700 millions d'années, toutes pendant le Précambrien (Heckman *et al.*, 2001). Taylor & Berbee (2006) leur reprochent d'avoir utilisé des points de calibration externes au groupe des champignons, impliquant des taux de substitution différents. Et également, qu'Heckman *et al.* (2001) aient utilisé une estimation controversée de la divergence animal-champignon de 1 576 millions d'années qui se base sur des fossiles qui ne peuvent être identifiés de manière indiscutable (Taylor & Berbee, 2006). Cependant, plusieurs études ont calculé des estimations similaires, montrant que la radiation des principaux groupes de champignons s'est produite au Précambrien (*Cf.* Figure 7), il y a plus de 550 millions d'années : Padovan *et al.* (2005), avec des séquences du gène de l'ARNr 18S et deux calibrations différentes pour la divergence animal-champignon, et Hedges *et al.* (2004)



avec des séquences protéiques. D'après toutes ces estimations, la divergence entre les Chytridiomycota et les autres phyla de champignons s'étend entre 893 et 1423 ou 1458 millions d'années auparavant, respectivement Padovan *et al.* (2005) et Heckman *et al.* (2001). Globalement, cette divergence se positionne bien de manière antérieure à la colonisation du milieu terrestre par les plantes.



**Figure 7 : Datation de la radiation des différentes lignées fongiques.** Les chiffres font référence aux temps de divergences utilisés ici et calculés par plusieurs études (Cf. Blair, 2009). Les noms des taxons sont issus de Hibbett *et al.* (2007). Liste des abréviations : Cz : Cénozoïque, Mp : Mésoproterozoïque, Mz : Mésozoïque et Np : Néoproterozoïque. Le Proterozoïque est une ère géologique englobée dans le Précambrien. Figure extraite de Blair (2009).

Les grands groupes actuels des champignons sont donc tous apparus au Précambrien, pendant le Protérozoïque. De nos jours, les champignons représentent l'un des groupes les plus diversifiés des Eukaryota. Nous allons voir dans la prochaine partie, la dimension de cette diversité.

### III – Diversité taxonomique

#### A – Description de nouvelles espèces

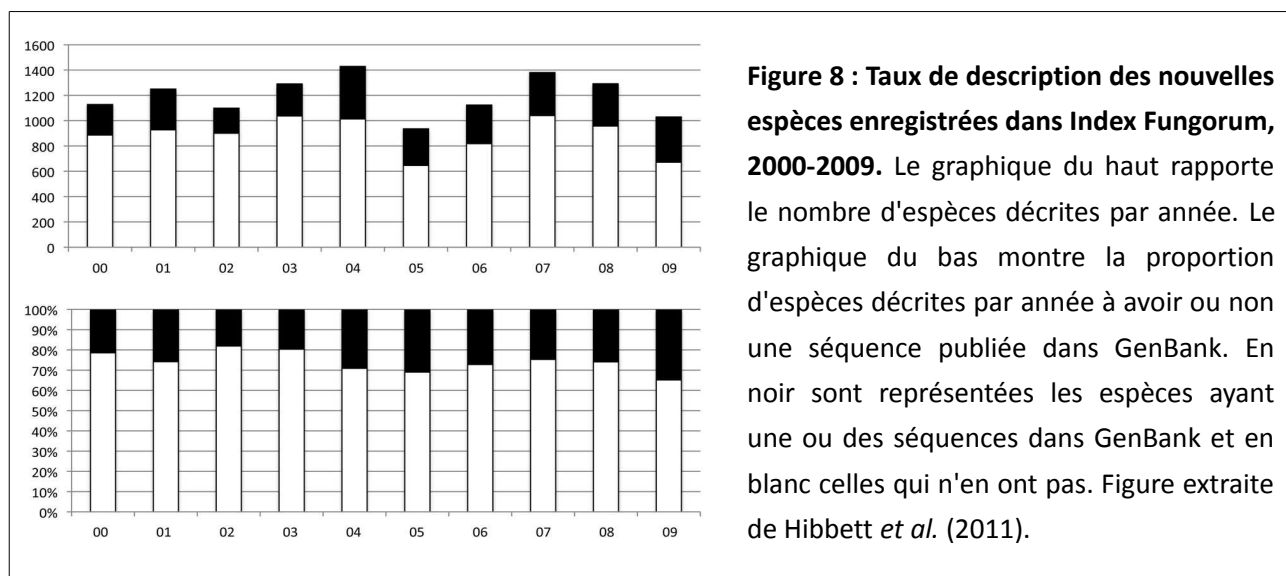
« The proportion of new species discovered in previously unstudied habitats or in revisionary studies on a regional or global level is an incontrovertible indicator of the extent of our ignorance » Hawksworth 2001

Le nombre de nouvelles espèces fongiques acceptées est difficilement dénombrable dû au large panel de médias scientifiques permettant leurs publications, contrairement aux nouvelles bactéries dont la publication est restreinte à deux revues scientifiques (Crous *et al.*, 2004). La description de nouvelles espèces fongiques est principalement basée sur le concept morphologique de l'espèce, qui souligne les divergences (Giraud *et al.*, 2008 ; Voigt & Kirk, 2011). L'originalité de la souche est validée par la publication des caractères morphologiques, physiologiques, écologiques ou phylogénétiques rendant compte de son unicité taxonomique. Les consignes demandent en général de rédiger une diagnose en latin complétée d'une description des autres caractères en anglais. L'attribution d'un nom d'espèce, plus formelle, est réglementée par le *Code International de Nomenclature Botanique* (ICBN ; McNeill *et al.*, 2006) entre autre afin de vérifier l'adéquation de l'appellation et d'éviter les redondances. Enfin, un holotype (specimen) doit être déposé dans plusieurs collections de souches internationales. Cette procédure est donc assez standardisée mais ne permet pas la différenciation entre deux noms d'espèces attribués à un même organisme (synonymie hétérotypique) ou d'une souche dont le nom d'espèce a été conservée, mais qui a été réassignée à un autre genre suite à de nouvelles études (synonymie homotypique) (Patterson *et al.*, 2010). Cela a donc une conséquence directe sur l'évaluation du nombre d'espèces fongiques acceptées.

Selon l'édition la plus récente du 'Dictionary of the Fungi' (Kirk *et al.*, 2008), 97 861 espèces sont recensées, soit 60 000 de plus depuis la première édition en 1943 (Blackwell, 2011). Durant les années 2000-2009, une moyenne annuelle de 1196 nouvelles espèces (Ascomycota, Basidiomycota et Glomeromycota) ont été décrites chaque année (Hibbett *et al.*, 2009 ; Hibbett *et al.*, 2011) (Cf. Figure 8), ce qui reste assez constant au regard des précédentes décennies, 1229

et 1097 nouvelles espèces pour les décennies 1980-1989 et 1990-1999 respectivement (Hawksworth, 2001).

Cependant, il n'est pas toujours indiqué clairement si ces chiffres concernent seulement les champignons 'vrais' puisqu'ils peuvent prendre en compte d'autres organismes étudiés par les mycologues, tels que les Eumycetozoa (1000-1100 espèces ; Fiore-Donno *et al.*, 2010) et les Oomycota (500-800 espèces ; Krings *et al.*, 2011) (Kirk *et al.*, 2008) qui pourtant ne sont plus considérés comme faisant partis du Règne des champignons depuis de nombreuses années. La connaissance de la diversité taxonomique des champignons 'vrais' est relative et souvent déséquilibrée vers certains groupes comme les macromycètes et les pathogènes de plantes (Voigt & Kirk, 2011). Cela a des répercussions directes sur la biologie évolutive (hypothèses phylogénétiques, coévolution...) mais également sur la biologie de la conservation via l'estimation de la diversité fongique sur Terre (Mueller & Schmit, 2007). Depuis une vingtaine d'année, cette estimation est un challenge pour de nombreux mycologues et écologues qui l'ont évaluée entre 500 000 et 9 900 000 espèces (Mueller & Schmit, 2007).



## B – Estimations de la diversité

L'estimation de Hawksworth (1991) d'un minimum de 1 500 000 espèces fongiques sur Terre est la plus communément admise, quand bien même elle fait l'objet de nombreuses critiques. L'étude se base sur le nombre de champignons associés à des plantes sur des îles en Europe. En moyenne, un ratio de 6:1 a été observé. La diversité globale des champignons a ensuite été extrapolée en utilisant une estimation de 270 000 plantes sur Terre. Dix ans plus tard, Hawksworth réexamine son estimation et la maintient malgré des arguments qui justifieraient une révision à la hausse (Hawksworth, 2001) : (1) le ratio entre champignons et plantes pourrait être plus élevé en milieu tropical qu'en milieu tempéré ; (2) d'autres habitats colonisés par les champignons devraient être pris en compte : insectes, sol, milieu aquatique... ; (3) les espèces cryptiques – espèces identiques morphologiquement mais qui sont des espèces distinctes – semblent courantes chez les macromycètes et (4) les approches moléculaires devraient permettre d'accéder aux espèces non-cultivables. Par exemple, en liaison avec ce dernier point, une étude propose une estimation minimum de la richesse spécifique globale des champignons du sol allant de 3 500 000 à 5 100 000 espèces (O'Brien *et al.*, 2005). Elle est extrapolée à partir de l'estimation de la richesse spécifique des champignons d'un sol de forêt, obtenue par le séquençage de masse de *Internal Transcribed Spacers* (notés ici ITS). Des critères plus ancrés dans un contexte écologique ont permis à Schmit & Mueller (2007) de proposer une estimation d'une limite basse de cette diversité. Ils l'évaluent à 712 000 espèces fongiques dans le monde. Pour cela, ils ont combiné des données collectées par des experts, qui prennent en compte la biogéographie, le taux d'endémisme et la spécificité d'hôtes des différentes espèces.

## C – Diversité par des approches moléculaires

La taxonomie traditionnelle est limitée pour la description de certaines espèces, *e.g.* les caractéristiques morphologiques utilisables pour la classification sont peu nombreuses et souvent convergentes (Redecker, 2002 ; Blackwell, 2011). De plus, une partie des champignons ne sont pas cultivables (*e.g.* Blackwell, 2011). La détection des espèces à partir d'échantillons

environnementaux en utilisant une approche moléculaire a déjà mis en évidence de nouveaux groupes fongiques à des rangs taxonomiques élevés. Ainsi, Vandenkoornhuyse *et al.* (2002) ont amplifié à partir de racines de plantes, deux séquences qui s'ancrent profondément dans les Ascomycota et les Basidiomycota. Schadt *et al.* (2003) ont révélé l'existence d'une grande diversité de champignons dans un sol de toundra recouvert de neige, parmi laquelle trois nouveaux clades au niveau du sous-phylum ou de la classe. Parfois la découverte de ces nouveaux groupes aboutit à la description formelle de nouveaux clades comme la nouvelle classe des Archaeorhizomycetes qui a pu être cultivée et décrite (Rosling *et al.*, 2011). Des séquences environnementales de ce groupe, qui était appelé 'Soil Clone Group 1', ont été obtenues pour la première fois par Porter *et al.* (2008). L'acquisition de toutes ces séquences environnementales nécessite de mettre en place un système pour les classer et les mettre en relation avec la diversité fongique cultivée et décrite.

L'approche moléculaire pour étudier la diversité des champignons promet un rendement de découverte d'espèces beaucoup plus élevé que les approches traditionnelles (Peay *et al.*, 2008 ; Hibbett *et al.*, 2011). L'identification des espèces repose alors entièrement sur des données génétiques moléculaires que l'on appelle *Molecular Operational Taxonomic Units* (MOTUs) (Hibbett *et al.*, 2011 ; Krüger *et al.*, 2012). Cela va de paire avec une évolution du *Code International de Nomenclature Botanique* (ICBN ; McNeill *et al.*, 2006) qui se traduirait par la mise à disposition de MOTUs pour la description d'un nouveau spécimen (Hibbett *et al.*, 2011). En effet, l'analyse des nouvelles espèces de champignons *sensu lato* décrites entre 2000 et 2009 montre que 74,4% d'entre elles n'ont pas de séquences mises en ligne dans GenBank (Hibbett *et al.*, 2011) (Cf. Figure 8). La génération de séquences de référence pour les espèces déjà décrites permettrait d'avoir une meilleure connaissance de l'avancement de la description de la diversité fongique globale et aussi de la variabilité et de l'applicabilité des marqueurs génétiques utilisés.

Récemment, au cours du 9<sup>ème</sup> Congrès International de Mycologie, les mycologues ont montré l'envie de moderniser le *Code International de Nomenclature Botanique* (Norvell *et al.*, 2010). Plusieurs propositions ont été faites ces dernières années, comme le dépôt des nouveaux noms dans une base de données publique, accessible en ligne (Hawksworth, 2005). Deux bases de données existent déjà, *Index Fugorum* (Kirk, 2000) et Mycobank (Crous *et al.*, 2004). Elles

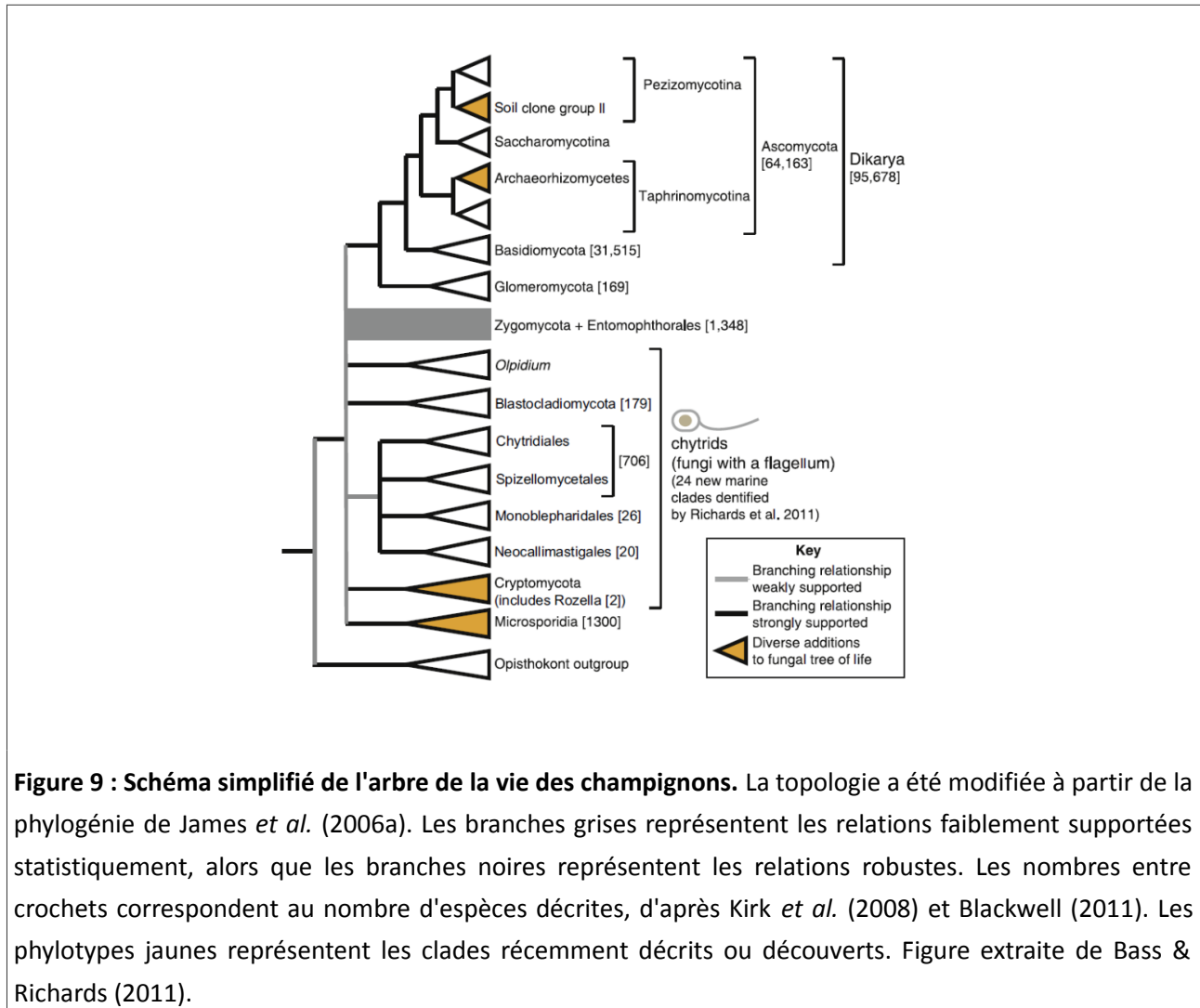
répertorient les noms donnés aux espèces fongiques *sensu lato* et mettent à disposition la description des caractères associés. Pour le moment, elles fonctionnent de manière participative et se propose d'aider les taxonomistes dans le choix du nom d'une nouvelle espèce. Egalement, la publication électronique des descriptions d'espèce est plébiscitée (Norvell *et al.*, 2010). Cela faciliterait la récupération des informations via des liens, *e.g.* les séquences nucléotidiques sur GenBank (Hibbett *et al.*, 2011). Une réelle volonté de communiquer et de rendre accessible les données est en train de s'affirmer dans la communauté des mycologues, en s'appuyant sur le développement des outils de partage en ligne.

## VI – Arbre de la vie du règne des Champignons

### A – Glomeromycota, Dikarya : des groupes monophylétiques

Les contours de l'arbre de la vie du règne des champignons ont été confirmés par l'utilisation de marqueurs moléculaires pour les analyses phylogénétiques. Ainsi les eucaryotes hétérotrophes Eumycetozoa et Oomycota, ont été déplacés hors du règne des champignons grâce à l'analyse des séquences protéiques du facteur d'élongation EF1- $\alpha$  (Baldauf & Doolittle, 1997) et des séquences du gène de l'ARNr 18S (Gunderson *et al.*, 1987) respectivement. Actuellement cinq phyla majeurs sont acceptés : Basidiomycota, Ascomycota, Glomeromycota, Zygomycota et Chytridiomycota (Cf. Figure 9). Le phylum Glomeromycota est le dernier à avoir été décrit (Schüßler *et al.*, 2001). Auparavant ce groupe correspondait à l'ordre des Glomerales et était placé parmi les Zygomycota. L'analyse des séquences du gène de l'ARNr 18S a démontré la monophylie des Glomeromycota et propose également qu'ils partagent un ancêtre commun avec les Basidiomycota et les Ascomycota. C'est la plus importante restructuration de l'arbre de la vie des champignons qui ait été faite ces dernières années. L'arbre de la vie des champignons tel que perçu aujourd'hui pourrait être une perception biaisée de la réalité de la diversité des champignons. Différentes études récentes ont permis de mettre en évidence l'existence de branches profondes inconnues dans les Ascomycota (*e.g.* Vandenkoornhuysen *et al.*, 2002 ; Schadt *et al.*, 2003) et les Basidiomycota (*e.g.*

Vandenkoornhuyse *et al.*, 2002) qui pourraient constituer de nouveaux phyla.



Les reconstructions phylogénétiques des hauts rangs taxonomiques permettent de formuler des hypothèses sur la constitution des clades fongiques majeurs et les relations existant entre eux. James *et al.* (2006a) ont combiné les données de six marqueurs moléculaires échantillonnés sur 199 taxons fongiques : les séquences nucléotidiques des gènes de l'ARNr 18S, 28S et 5.8S ainsi qu'EF1- $\alpha$  et des sous-unités de l'ARN polymérase II (*RPB1* et *RPB2*). Ils montrent que la monophylie des Ascomycota et des Basidiomycota est très bien supportée statistiquement, ce qui corrobore les résultats de Liu *et al.* (2006) obtenus par des analyses sur un seul gène (18S ou *RPB1*) ou deux

gènes combinés (*RPB1* avec *RPB2*). Ainsi, il a été proposé de les regrouper dans un sous-règne appelé Dikarya en référence à une partie de leur cycle de vie pendant laquelle les cellules des hyphes possèdent deux noyaux (Hibbett *et al.*, 2007). La phylogénie de James *et al.* (2006a) confirme également le regroupement des Glomeromycota en tant que clade frère des Dikarya. Si les lignées apparues le plus récemment semblent assez bien définies, les lignées basales de l'arbre de la vie des champignons sont assez controversées.

## B – Les lignées basales : des clades peu résolus

Les lignées fongiques basales sont décrites comme un assemblage de multiples clades de Zygomycota et Chytridiomycota (McLaughlin *et al.*, 2009). Dans leur classification, Hibbett *et al.* (2007) ont déstructuré les Zygomycota en tant que phylum pour les reclasser en quatre sous-phyla *incertae sedis*, compte tenu du caractère polyphylétique de ce clade quels que soient les marqueurs moléculaires et les stratégies mono- ou multi-loci employées (*i.e.* Tanabe *et al.*, 2004 ; James *et al.*, 2006a ; Liu *et al.*, 2006 ; White *et al.*, 2006). Les Zygomycota sont caractérisés par un taux de mutations élevé, ce qui crée de longues branches dans les phylogénies (Redecker, 2002). Seules des relations intraordinales sont obtenues de manière soutenue mais pas avec tous les marqueurs (Tanabe *et al.*, 2004). Pour les Chytridiomycota, la classification est beaucoup mieux connue. Hibbett *et al.* (2007) ont érigé l'ordre des Blastocladales au rang de phylum Blastocladiomycota et celui des Neocallimastigales devient le phylum Neocallimastigomycota, en se basant sur les phylogénies de James *et al.* (2006a) et (2006b). Toutefois, ce dernier phylum n'est pas toujours reconnu (Bass & Richards, 2011) (*Cf.* Figure 9).

Deux clades sont en cours d'analyse afin d'établir les relations qui existent entre eux et le règne des champignons (*Cf.* Figure 9). Le premier, les microsporidies, a une position phylogénétique incertaine qui est beaucoup discutée depuis plus de vingt ans (James *et al.*, 2006a). Ce sont des parasites obligatoires unicellulaires, hétérotrophes et non flagellés (Voigt & Kirk, 2011). La difficulté pour leur trouver des liens phylogénétiques interprétables par rapport à des clades existants, réside dans leur tempo évolutif (*i.e.* taux de mutations accumulées par unité de temps) très élevé. Tout d'abord placées parmi les Eukaryota comme le clade ayant divergé en



premier puis positionnées comme une lignée émergeant des Zygomycota, comme clade frère des Dikarya ou encore clade frère des champignons, leur positionnement requiert d'être reconsidéré (Cf. synthèse bibliographique de Voigt & Kirk, 2011 ; Tanabe *et al.*, 2005). Le second est celui des Cryptomycota qui a été décrit comme nouveau phylum en novembre 2011 (Jones *et al.*, 2011b). Le placement phylogénétique des lignées basales est assez délicat à soutenir car un biais d'attraction de longues branches ne peut être écarté (Lara *et al.*, 2010). Cet effet d'attraction des longues branches tend à faire se regrouper des taxons phylogénétiquement éloignés (Philippe & Laurent, 1998) et induit une position profondément enracinée erronée. Ce phénomène est souvent observé chez les parasites de manière générale et il est encore plus marqué pour les parasites intracellulaires qui accumulent plus de mutations par unité de temps que le reste des organismes (Itoh *et al.*, 2002).

### C – Un nouveau phylum : les Cryptomycota

Le genre *Rozella* s'inscrit comme une lignée basale du règne des champignons ou comme son clade frère le plus proche (James *et al.* 2006a ; 2006b ; Lara *et al.*, 2010 ; Jones *et al.*, 2011a). Les *Rozella* sont des parasites intracellulaires obligatoires de champignons zoosporiques (Held, 1981). Ils sont caractérisés par leurs zoospores uniflagellés, similaires à ceux des Chytridiomycota par un stade intracellulaire pendant lequel ils n'ont pas de paroi et se nourrissent du cytoplasme de leur hôte par phagocytose (Held, 1981). Récemment des études ont montré que des séquences environnementales issues de sols ou de milieux aquatiques formaient un nouveau clade incluant les *Rozella* (Lara *et al.*, 2010 ; Jones *et al.*, 2011a). Ces séquences environnementales formant le clade LKM11 (du nom de la première séquence de ce groupe), proviennent d'une étude sur le fonctionnement des communautés microbiennes d'un lac (van Hannen *et al.*, 1999). Elles ont par la suite été retrouvées dans des sols (*e.g.* Lesaulnier *et al.*, 2008), en eau douce (*e.g.* Lefèvre *et al.*, 2007 ; Lefèvre *et al.*, 2008 ; Lepère *et al.*, 2010), dans des sédiments marins profonds (*e.g.* Nagano *et al.*, 2010 ; Nagahama *et al.*, 2011) et des environnements anoxiques (*e.g.* Takishita *et al.*, 2010). Ce clade LKM11 regroupé avec les *Rozella* a été érigé en phylum des Cryptomycota dans le règne des champignons, pour partie en raison de la richesse de sa diversité phylogénétique (Jones *et al.*,

2011b). Jones *et al.* (2011a) ont étudié les propriétés structurales des Cryptomycota en utilisant différentes techniques moléculaires sur des échantillons environnementaux, ce qui leur a permis d'examiner plusieurs stades du cycle de vie. Seulement trois morphologies ont pu être observées : des zoospores ovoïdes uniflagellés, des kystes et des cellules sans flagelle attachées à des diatomées ou à d'autres cellules. Pour ce dernier cas, il n'a pas été possible de déterminer si leur mode de vie est parasitaire ou saprophyte. L'un des résultats majeurs est qu'aucun stade observé ne présentait de paroi contenant de la chitine (*Cf.* Encadré 1 : La chitine) ou de la cellulose, à l'inverse de ce qui est couramment retrouvé chez les champignons, du moins au niveau des spores (Adl *et al.*, 2005).

#### Encadré 1 : La chitine

La chitine assure l'intégrité de la cellule fongique et confère à la paroi une forte résistance à l'étirement, ce qui fait qu'elle est impliquée dans de nombreux processus comme par exemple la croissance, la communication, la nutrition et l'adhésion à d'autres cellules (Bowman & Free, 2006). Elle représente 1 à 2% du poids sec de la paroi des levures et entre 10 et 20% de celle des formes filamenteuses (Bowman & Free, 2006). Les parois des champignons sont considérées comme riches en chitine en comparaison à d'autres organismes. En effet, elle n'est pas exclusive aux champignons puisque des chitines synthases sont retrouvées dans les règnes formant la couronne eucaryote, à l'exclusion des plantes (Ruiz-Herrera *et al.*, 2002). Toutefois, les champignons sont les seuls à posséder des chitine synthases de la division 1 et de la division 2 (*Cf.* Figure 10), alors que les autres organismes ne possèdent qu'un petit nombre d'enzymes homologues ou paralogues (James & Berbee, 2012). La duplication des gènes codant pour les chitine synthases semble s'être produite lors d'une transition évolutive ancienne des champignons puisque quasiment tous possèdent des enzymes de ces deux domaines (Ruiz-Herrera *et al.*, 2002). L'acquisition d'une paroi riche en chitine leur a permis de s'adapter à de nombreux milieux et de se nourrir par osmotrophie (Richards *et al.*, 2012). Cependant certains champignons parasites, lorsqu'ils sont à

l'intérieur d'une cellule hôte, sont généralement sans paroi et absorbent des nutriments par phagocytose (James & Berbee, 2012). Ce pourrait être un caractère ancestral qu'ils partagent avec les *Nucleariida*, taxon frère le plus proche des champignons, (Lara *et al.*, 2010) ou une adaptation par convergence au parasitisme intracellulaire (James & Berbee, 2012) en réponse à la capacité de certains hôtes à détecter la présence de chitine via leur système immunitaire (Leonardon *et al.*, 2010).

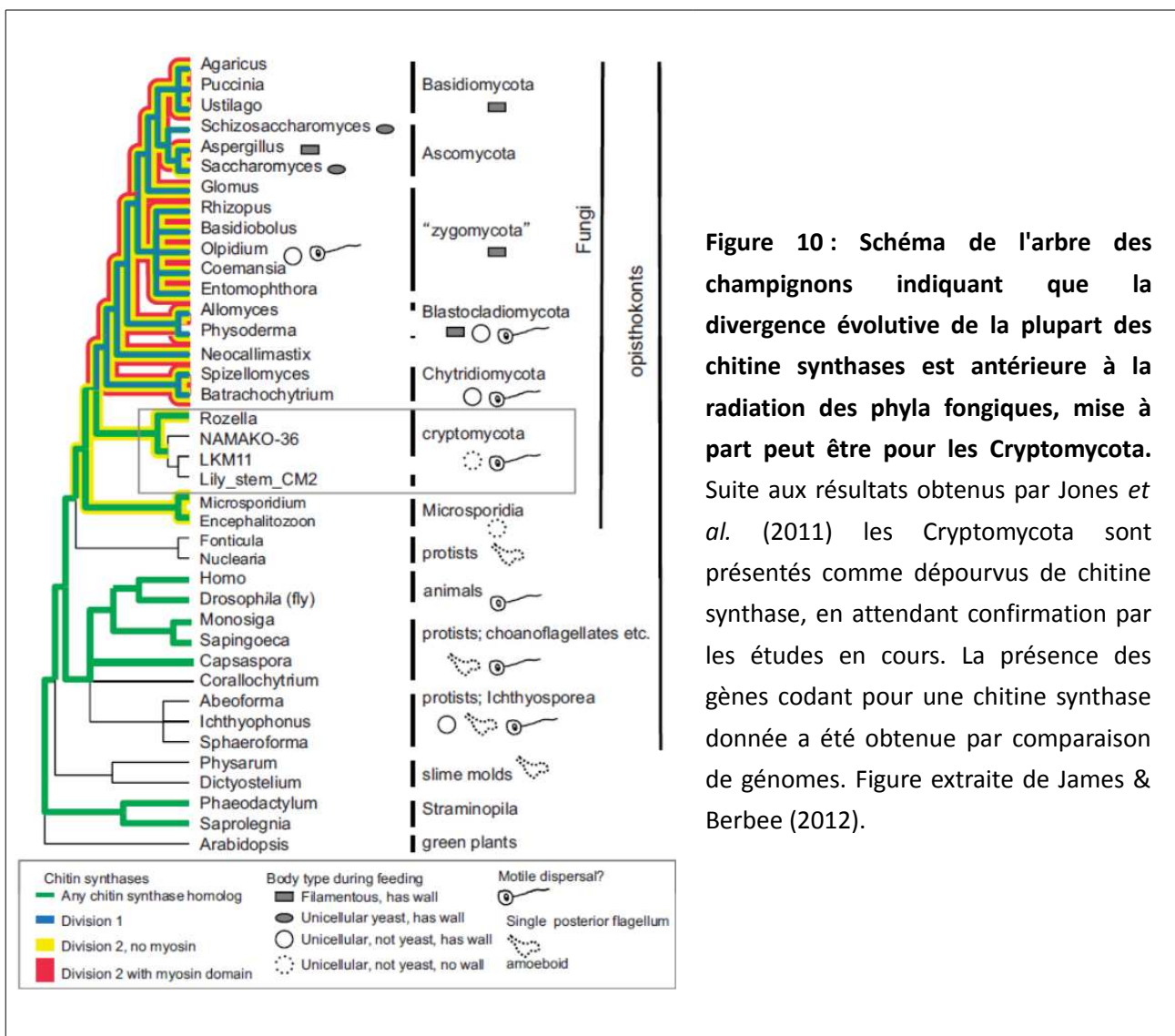


Figure 10 : Schéma de l'arbre des champignons indiquant que la divergence évolutive de la plupart des chitine synthases est antérieure à la radiation des phyla fongiques, mise à part peut être pour les Cryptomycota. Suite aux résultats obtenus par Jones *et al.* (2011) les Cryptomycota sont présentés comme dépourvus de chitine synthase, en attendant confirmation par les études en cours. La présence des gènes codant pour une chitine synthase donnée a été obtenue par comparaison de génomes. Figure extraite de James & Berbee (2012).

James & Berbee (2012) ont examiné des *Rozella* en culture pour pouvoir observer un plus grand nombre de stades du cycle de vie. Ainsi, des sporanges de résistance immatures comportant de la chitine dans leur paroi ont été mis en évidence. Pour les sporanges de résistance matures (Cf. Figure 11), la chitine n'est plus observable, en raison de la couche externe pigmentée de la paroi masquant sa présence (James & Berbee, 2012). L'analyse du génome d'une espèce de *Rozella* a révélé la présence d'au moins une chitine synthase de la division 2 (Cf. Figure 10), ce qui est en faveur de leur position basale (James & Berbee, 2012) De plus, ce positionnement peut être appuyé par le fait que les cellules de *Rozella* présentent des flagelles à une certaine période de leur cycle de vie (Held, 1981).



**Figure 11 : Photo de *Rozella allomycis* parasitant les hyphes septés d' *Allomyces* (Blastocladiomycota).** Les sporanges de dormance matures ont une paroi épaisse pigmentée tandis que les sporanges de dormances immatures sont de couleur claire. Echelle 20  $\mu$ m. Photo extraite de James & Berbee (2012).

#### D – Les lignées basales caractérisées par la présence d'un flagelle

Les lignées basales ou les taxons frères les plus proches des champignons sont en effet caractérisées par un stade 'zoospore' pendant une partie de leur cycle de vie (Cf. Figure 9). Chez les Cryptomycota, *Rozella*, Chytridiomycota et Blastocladiomycota, les spores mobiles sont généralement pourvues d'un flagelle implanté à l'arrière de la cellule, qui leur permet de se déplacer dans un milieu aquatique (Held, 1981 ; James *et al.*, 2006b ; Gleason & Lilje, 2009). De rares exceptions sont toutefois observées : plusieurs flagelles (*e.g.* certaines espèces de *Neocallimastigomycota*) ou pas de flagelles du tout (*e.g.* *Hyaloraphidium*) (Gleason & Lilje, 2009). Les autres phyla fongiques ne présentant pas de flagelle, plusieurs hypothèses ont été proposées

sur l'occurrence de la perte du flagelle au cours de l'évolution des champignons. Ces hypothèses diffèrent selon les marqueurs génétiques et les taxons utilisés pour construire les phylogénies sur lesquelles ont été placé l'événement de perte de flagelle. Ainsi, Liu *et al.* (2006) suggèrent que cette perte s'est produite une seule fois. Leur phylogénie basée sur les marqueurs *RPB1* et *RPB2* regroupe les Zygomycota avec les Glomeromycota dans un clade monophylétique, lui même en position paraphylétique avec les Ascomycota et les Basidiomycota. De ce fait, il n'y a eu qu'une occasion pour la perte du flagelle : entre les Chytridiomycota et les autres phyla fongiques. Ce travail a été discuté et est controversé notamment en raison du fait que les Zygomycota sont généralement considérés comme polyphylétiques en utilisant d'autres cibles génétiques et ne sont d'ailleurs plus considérés comme un phylum à part entière (Hibbett *et al.*, 2007). James *et al.* (2006a), avec moins de parcimonie, ont quant à eux estimé que cette perte de flagelle s'est produite entre 4 à 6 fois. Ils se sont basés sur une phylogénie réalisée à partir de six gènes qui indiquent que les Chytridiomycota et les Zygomycota sont polyphylétiques. *Olpidium* un Chytridiomycota parasite, au placement phylogénétique difficile à résoudre, se place dans cette étude dans un clade composé exclusivement de Zygomycota, ce qui engendre deux à quatre événements de perte de flagelle. Egalement, cette phylogénie inclut les microsporidies, non-flagellées, qui se branchent avec les *Rozella*, dans les lignées basales flagellées (James *et al.*, 2006a). Ceci est à considérer avec prudence puisque le positionnement des microsporidies n'est pas clairement résolu (James *et al.*, 2006a). Enfin, une dernière hypothèse propose une perte de flagelle qui se produit au sein des Chytridiomycota, car *Hyaloraphidium* n'a pas de flagelle. Ce résultat est plutôt dû à l'échantillonnage des taxons et correspond un événement isolé. D'ailleurs, dans la Figure 9 extraite de Bass & Richards (2011), les auteurs ont modifié la topologie de James *et al.* (2006a) et ainsi elle montre que la perte de flagelle est un événement unique. Le positionnement de caractères sur des phylogénies est donc une pratique à utiliser avec précaution, plusieurs sources d'erreurs existant. Pour le positionnement de la perte du flagelle, une des faiblesses vient du fait que les lignées basales sont assez mal résolues en raison notamment d'un manque de séquences disponibles pour les marqueurs généralement utilisés pour retracer l'histoire évolutive des champignons.

## E – Hypothèses évolutives sur la diversification des champignons

En accord avec la datation de la radiation des différentes lignées de champignons, la présence de zoospores chez les organismes des lignées basales des champignons est en faveur d'une diversification des champignons en milieu aquatique (James *et al.*, 2006a ; Le Calvez *et al.*, 2009). D'une part, l'absence de structure pour une dispersion aérienne et la présence d'un flagelle, implique que la dispersion des spores et la reproduction aient lieu dans un environnement aquatique (James *et al.*, 2006a). D'autres part, des fossiles de Glomeromycota en association avec des cyanobactéries, retrouvés dans les sédiments siliceux de Rhynie renforcent cette hypothèse, en révélant une association ne pouvant se produire qu'en milieu humide ou aquatique (Taylor *et al.*, 1997). James *et al.* (2006a) ont remarqué que la perte du flagelle coïncide avec l'apparition de nouveaux modes de production et de dissémination des spores. Ainsi, il a été fait l'hypothèse que la perte des gamètes flagellés a été compensée par le développement de spores de résistance capables de disperser sur de longues distances (Le Calvez *et al.*, 2009). Cet événement pourrait avoir facilité la colonisation de nouveaux écosystèmes comme le milieu terrestre. Richards *et al.* (2012) proposent quant à eux, qu'une diversification continentale se soit produite en premier lieu et qu'elle fut suivie de plusieurs transitions marine-continentale dans les deux sens. Ils s'appuient sur le fait que les lignées marines se positionnent phylogénétiquement parmi les lignées continentales en formant des branches assez courtes. Et également, parce que des séquences appartenant aux champignons des lignées basales sont retrouvées en plus grande abondance au niveau des systèmes lenticulaires par rapport au milieu marin.

## V – Les Chytridiomycota

### A – Biologie

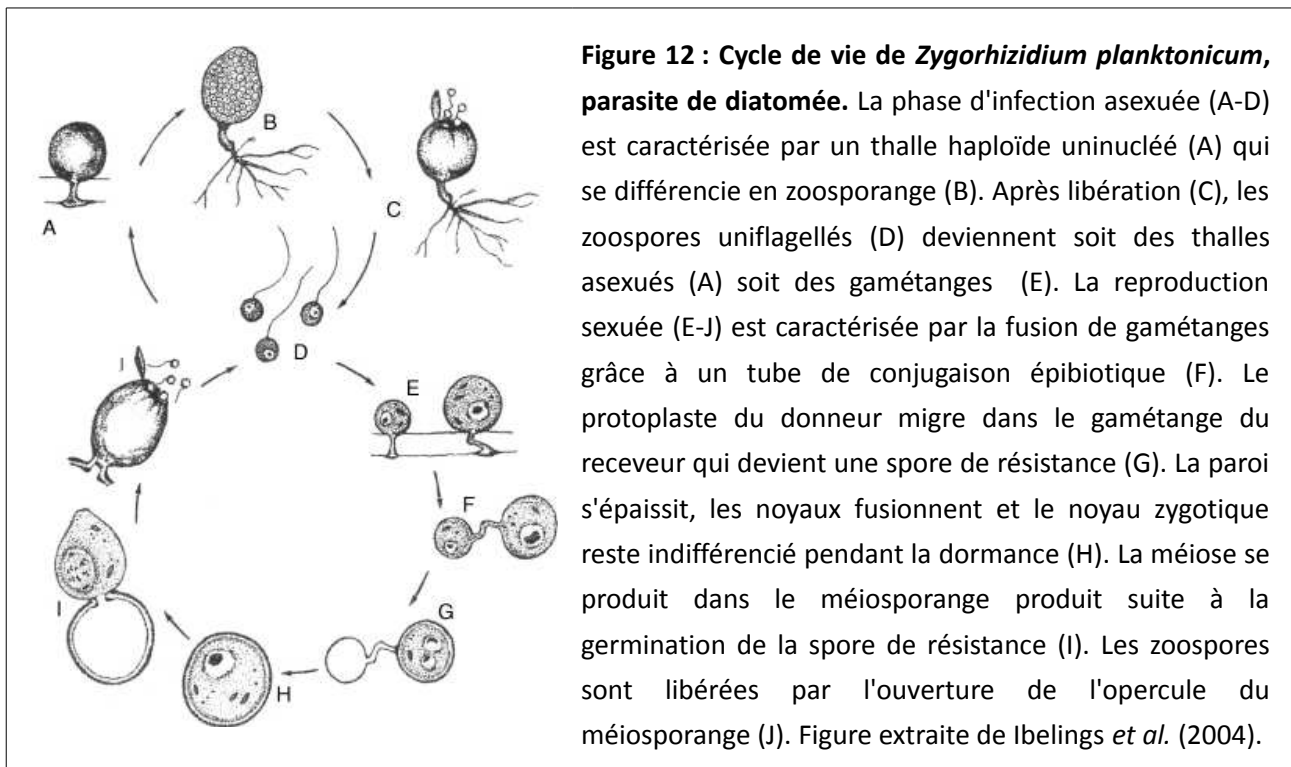
Dans cette section, seuls les Chytridiomycota *sensu stricto* seront décrits. C'est-à-dire que les caractères propres aux Blastocladiomycota et aux Neocalligomastimycota, qui ont récemment

été érigés en phyla (Hibbett *et al.*, 2007), ne seront pas abordés. Brièvement, les premiers sont des saprophytes ou parasites de champignons, d'algues, de plantes ou d'invertébrés et peuvent être anaérobiques facultatifs (James *et al.*, 2006b). Les seconds sont anaérobiques et le plus souvent retrouvés dans le rumen, le système digestif ou les fèces de mammifères et de reptiles herbivores (Liggenstoffer *et al.*, 2010). Ils ont un fort potentiel d'utilisation en biotechnologie dans les procédés de fermentation et la production de biocarburant à partir d'une biomasse lignocellulosique.

Les Chytridiomycota *sensus stricto* sont très divers phylogénétiquement (James *et al.*, 2006b). Actuellement, les Chytridiomycota sont répartis en deux groupes, la classe des Chytridiomycetes, qui comprend six ordres : Chytridiales, Spizellomycetales (Barr, 1980), Rhizophydiales (Letcher *et al.*, 2006), Cladochytridiales (Mozley-Standridge *et al.*, 2009), Rhizophlyctidiales (Letcher *et al.*, 2008), Lobulomycetales (Simmons *et al.*, 2009) et la classe des Monoblepharidomycetes (Hibbett *et al.*, 2007) qui ne contient qu'un seul ordre, les Monoblepharidales. Traditionnellement, les groupes taxonomiques des Chytridiomycetes étaient basés sur la morphologie et le développement du thalle, mais les phylogénies moléculaires ont montré que ces caractères ne possédaient pas d'information phylogénétique de bonne qualité contrairement aux caractères ultrastructuraux des zoospores qui sont plus conservés (James *et al.*, 2000). La restructuration du phylum des Chytridiomycota est donc récente. L'ordre des Chytridiales, qui contenait les cinq autres ordres des Chytridiomycetes, ne comprend plus que la famille des Chytridiaceae (Letcher *et al.*, 2005) et quelques autres clades qui nécessitent une révision de leur systématique (Mozley-Standridge *et al.*, 2009).

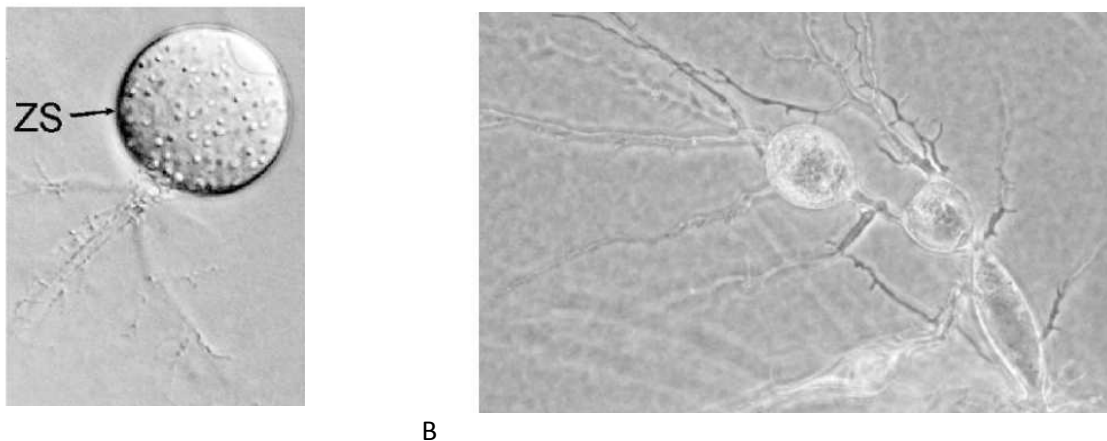
Les Chytridiomycota *sensus stricto*, que l'on appelle également chytrides (le terme 'chytride' est utilisé parfois de manière galvaudée puisqu'il inclut en plus des Chytridiomycota les Oomycota et les Eumycetozoa) comprennent environ 1000 espèces décrites (James *et al.*, 2006b). Ils sont presque tous microscopiques et peu d'espèces ont une importance économique (Powell, 1993) ce qui pourrait expliquer le manque d'intérêt qui leur est porté (James *et al.*, 2000). Comme les autres champignons, ils possèdent de la chitine dans leur paroi, des crêtes mitochondriales aplaties, ils stockent les glucides sous forme de glycogène et ils synthétisent la lysine par la voie des acides

aminoadiques (AAA). Pour le moment, deux modes de vie ont été observés, saprophytes ou parasites. Le cycle de vie d'un parasite de diatomée, *Zygorhizidium planktonicum*, est illustré dans la figure 12.



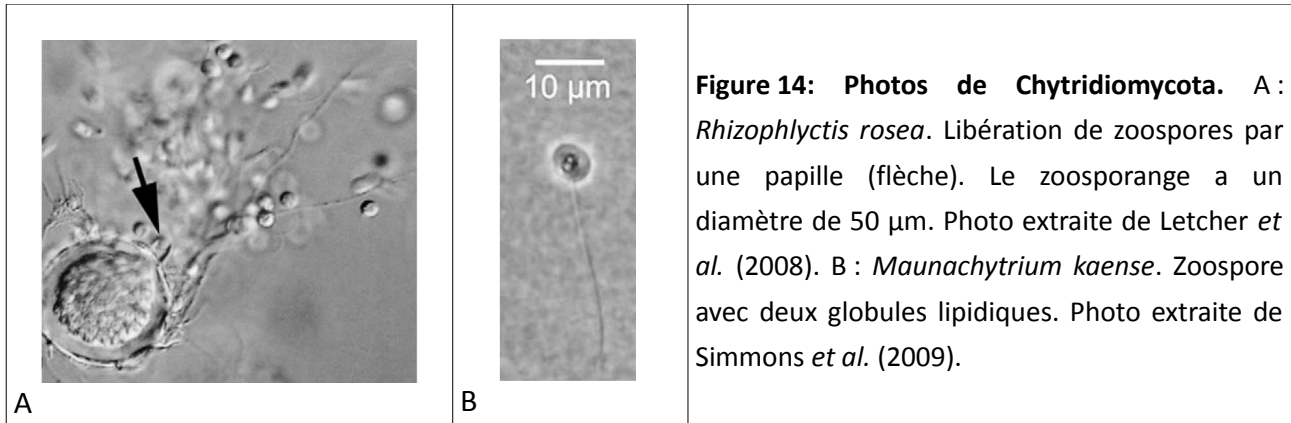
Concernant leur morphologie, on distingue les chytrides eucarpiques (*Cf.* Figure 13A) qui consiste en un sporange reproducteur, duquel partent des rhizoïdes filamenteux (hyphe anucléé) qui leur servent à s'ancrer à leur hôte ainsi qu'à se nourrir par absorption (Letcher & Powell, 2002). Et les chytrides holocarpiques dont le thalle est entièrement converti en sporange pendant la reproduction (James *et al.*, 2006b). Ensuite le thalle peut être monocentrique (*Cf.* Figure 13A), c'est-à-dire qu'un individu produit un seul sporange, ou polycentrique (*Cf.* Figure 13B), un individu est composé de plusieurs sporanges dans un réseau de rhizoïdes. Enfin, le thalle peut être végétatif ou reproducteur et on l'appelle respectivement zoosporange ou gamétange. Le zoosporange est une structure en forme de sac, à paroi mince, dont le protoplasme subit des divisions produisant des zoospores (James *et al.*, 2006b).





**Figure 13 : Photo de Chytridiomycota. A :** *Chytriumyces hyalinus*, eucarpique monocentrique avec un zoosporange (ZS) operculé d'où partent les rhizoïdes. **B :** *Catenomyces persicinus* polycentrique avec plusieurs sporanges dans les rhizoïdes. Photos extraites de James *et al.* (2006b).

Les zoospores sont des formes de dispersion active sur une courte distance (Gleason *et al.*, 2008). Ce sont des cellules sans paroi portant un unique flagelle implanté postérieurement (Cf. Figure 14). Ils s'en servent pour se propulser et pour cela, il est nécessaire qu'ils se trouvent en présence d'eau. Toutefois, certains chytrides produisent des zoospores amoéboïdes sans flagelle qui se déplacent en rampant sur les surfaces (Gleason & Lilje, 2009). Pendant la dispersion les zoospores dépendent de leur réserve énergétique jusqu'à ce qu'ils s'enkystent à un substrat (Gleason *et al.*, 2008). Les spores de résistance sont quant à eux produits de manière sexuée ou asexuée et possèdent une paroi épaisse qui leur permet de se maintenir dans des conditions qui leur sont défavorables. Les zoospores semblent avoir un rôle important dans les écosystèmes mais leur physiologie et leur écologie restent moins explorées que leurs relations phylogénétiques (Gleason & Lilje, 2009).

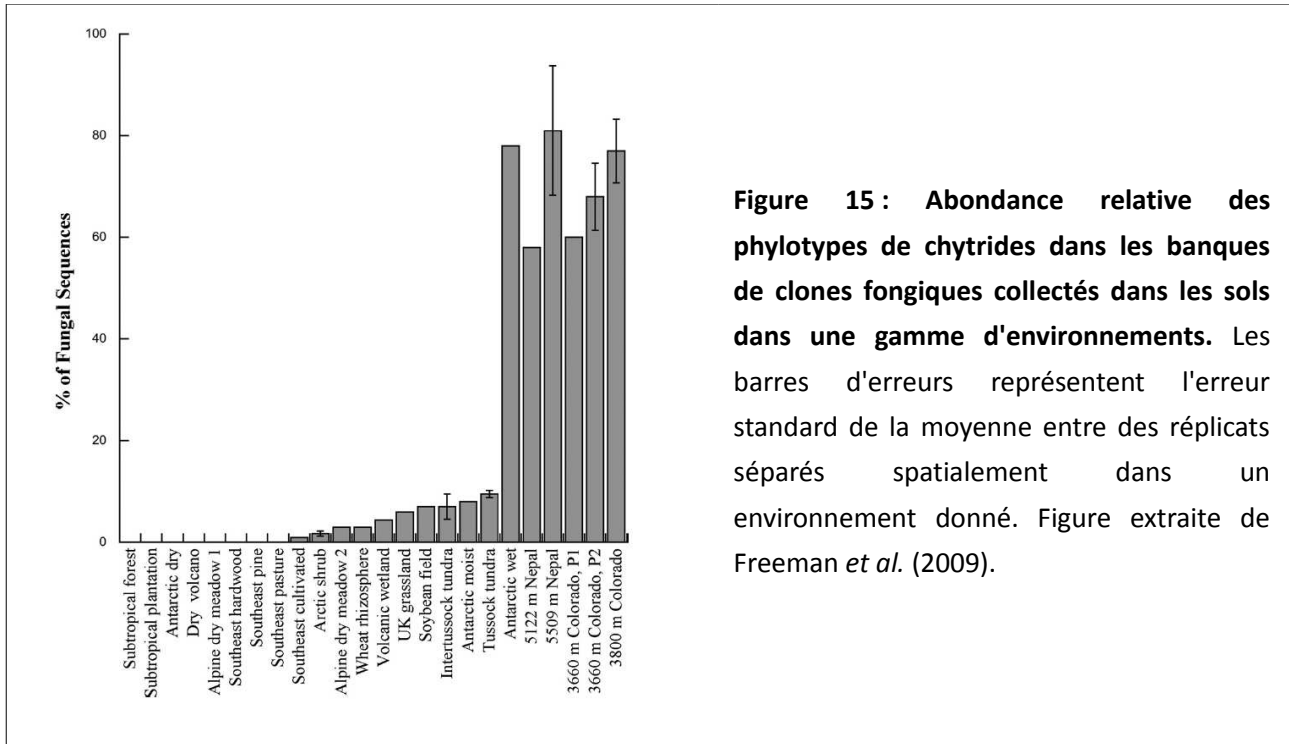


## B – Profil écologique

L'étude des chytrides dans l'environnement a longtemps été négligée probablement parce qu'ils sont passés inaperçus ou parce qu'ils ont été confondus avec d'autres organismes (Powell, 1993). Par exemple, dans les systèmes lenticulaires, ils ont été inclus parmi les protistes eucaryotes avant d'être identifiés comme des champignons (Lefèvre *et al.*, 2007). Dans la littérature, il en découle l'utilisation de nombreuses appellations pour les décrire (*i.e.* champignons zoosporiques, hétérotrophes flagellés...), ce qui porte à confusion car elles peuvent englober des organismes très différents. Quelques espèces ont toutefois reçu plus d'attention que d'autres parce qu'elles ont un impact économique ou écologique plus visible. La plus étudiée, *Batrachochytrium dendrobatidis*, décrite en 1998, est responsable de la chytridiomycose des amphibiens dans de nombreuses régions du monde (Cf. synthèse bibliographique Kilpatrick *et al.*, 2010). Elle représente la plus grande menace infectieuse pour les amphibiens et pourrait être impliquée dans leur déclin. Mais les mécanismes menant à la mort de l'hôte ne sont pas bien connus et donc la pathogénie nécessite d'être étudiée de manière plus approfondie. Egalement les plantes supérieures sont parasitées par des chytrides comme *Synchytrium endobioticum*, qui est très difficile à éradiquer et dont les spores peuvent rester viables jusqu'à 40 ans dans le sol. *S. endobioticum* provoque la galle verruqueuse de la pomme de terre, ce qui les rend impropres à la consommation et à la vente (Cf. synthèse bibliographique Powell, 1993). Dans un contexte moins anthropocentrique, les chytrides sont parasites de phytoplancton, d'algues et d'autres champignons, voire d'autres chytrides comme *Septosperma rhizophydii* qui infecte *Rhizophyidium macrosporum* (Cf. synthèse

bibliographique de Powell, 1993). Une autre fonction écosystémique des chytrides est le saprophytisme : ils dégradent les restes d'animaux et de végétaux. Ils ont la capacité de décomposer des macromolécules résistant à d'autres organismes, telles que la chitine, la kératine et la cellulose (*e.g.* pollen) (Gleason *et al.*, 2008). On retrouve d'ailleurs ces fonctions avec des impacts plus ou moins conséquents dans différents écosystèmes.

Les chytrides sont communément trouvés en eau douce, dans le milieu marin mais également dans les sols présentant périodiquement un taux d'humidité assez élevé favorable à la dissémination des zoospores (Shearer *et al.*, 2007). Ils sont ubiquistes et ont colonisé des habitats sous toutes les latitudes (Powell, 1993). Des études portant sur l'abondance des champignons dans les sols montrent qu'en général les chytrides ne sont pas détectés ou alors qu'ils représentent une petite fraction (<10%) des séquences obtenues (Figure 15) (Freeman *et al.*, 2009). Les seules exceptions font état d'une domination des communautés fongiques par les chytrides dans des sols bénéficiant d'un apport d'eau saisonnier lié à la fonte des neiges. Ainsi, ils constituent 78% des séquences fongiques en Antarctique maritime sud (Bridge & Newsham, 2009), 70% dans les Rocheuses (Colorado) et 75% dans l'Himalaya népalais (Freeman *et al.*, 2009). Les sites examinés par Freeman *et al.* sont dépourvus de plantes et la production primaire résulte de l'activité photosynthétique des algues et de photoautotrophes bactériens. Ils ont ainsi proposé que les chytrides s'y maintenaient grâce au parasitisme et/ou saprophytisme sur les algues et les cyanobactéries, ou encore grâce à la dégradation de grains de pollen provenant d'altitudes plus faibles.

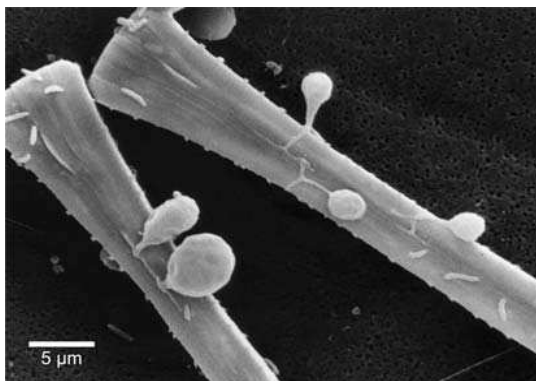


**Figure 15 : Abondance relative des phylotypes de chytrides dans les banques de clones fongiques collectés dans les sols dans une gamme d'environnements.** Les barres d'erreurs représentent l'erreur standard de la moyenne entre des répliqués séparés spatialement dans un environnement donné. Figure extraite de Freeman *et al.* (2009).

### 1) Les chytrides en eau douce

En eau douce, les chytrides ont surtout été étudiés dans les écosystèmes lenticques où ils semblent tenir un rôle trophique clé. Une étude ciblant le gène de l'ARNr 18S de picoflagellés 'non identifiés' a révélé une diversité de chytrides très élevée (Lefèvre *et al.*, 2008). 'Non identifiés' fait référence aux cellules de taille inférieure à 5µm dont les caractères ultrastructuraux ne permettent pas une identification au microscope à épifluorescence. Sur le nombre de séquences totales obtenues, 19% sont affiliées aux champignons et 75% des séquences fongiques sont identifiées comme appartenant aux chytrides. Le groupe des 'flagellés hétérotrophes', qui était traditionnellement restreint aux protistes bacterivores en milieu aquatique, est donc élargi aux champignons saprophytes et parasites (Sime-Ngando *et al.*, 2011). En lac, les chytrides peuvent avoir un impact considérable sur la dynamique du phytoplancton via le parasitisme : ils sont souvent spécifiques d'hôte, hautement infectieux et extrêmement virulents (Ibelings *et al.*, 2004 ; Kagami *et al.*, 2007). Des prévalences d'infection supérieures à 90%, ont été observées lors de blooms d'algues ou de cyanobactéries et cela a pour conséquence un effondrement relativement

rapide des populations (Kagami *et al.*, 2007). Parfois des prévalences intermédiaires sont observées et s'expliqueraient par la prédation du zooplancton sur les chytrides (Kagami *et al.*, 2009). Par exemple, *Zygorhizidium planktonicum* (chytrides) parasite des algues de grandes tailles comme *Asterionella formosa* (Figure 16), que les daphnies (zooplancton) ne peuvent consommer. *Z. planktonicum* se nourrit du protoplasme de leur hôte et ses zoospores sont ensuite broutés par les daphnies. Il y a donc un transfert d'énergie et de nutriments du phytoplancton au zooplancton (Kagami *et al.*, 2009). D'autant plus, de bonne qualité nutritionnelle, puisqu'ils sont riches en acides gras polyinsaturés, stéroïdes et glycogène (Sime-Ngando *et al.*, 2011). La matière organique peut ainsi être recyclée dans la colonne d'eau sans passer par la boucle microbienne (Lefèvre *et al.*, 2008). Les chytrides ont également un rôle dans la dégradation de la matière organique particulaire, qu'elle soit allochtone (feuilles, pollen...) ou autochtone (exuvies d'arthropodes, zooplancton mort...) (Cf. synthèses bibliographiques de Gleason *et al.*, 2008 ; Jobard *et al.*, 2010). Leur activité saprophytique pourrait être tout aussi vitale aux écosystèmes pélagiques qu'elle l'est aux systèmes du sol (Watling, 2005 & Treseder, 2005 dans Jobard *et al.*, 2010).



**Figure 16: Zoospores de *Zygorhizidium planktonicum* attachés à la surface de cellules hôtes d'*Asterionella formosa*, diatomée.** Microscopie électronique à balayage (échelle : 5  $\mu\text{m}$ ). Photo de Van Donk, extraite de Kagami *et al.* (2007).

## 2) Les chytrides en milieu marin

En milieu marin, peu d'espèces de Chytridiomycota ont été identifiées et caractérisées (Gleason *et al.*, 2011) : *Thallassochytrium gracilariopsis* (Nyvall *et al.*, 1999) qui est parasite ou mutualiste d'algues rouges, *Chytridium polysiphoniae* (Müller *et al.*, 1999) qui est parasite

obligatoire d'algues brunes et *Rhizophydium littoreum* (Shields *et al.*, 1990 dans Gleason *et al.*, 2011) qui est parasite facultatif d'algues vertes ou saprophyte. D'autres espèce du genre *Rhizophydium* ont également été décrites comme parasites de diatomées marines (Gleason *et al.*, 2011). Lors d'inventaires moléculaires portant sur les Eukaryota ou les champignons, des séquences de Chytridiomycota ont été mises en évidence. Les études correspondantes sont présentées plus en détail dans les tableaux 1 et 3 de la partie VI – Les champignons marins. Ainsi, la présence de Chytridiomycota a été observée dans des sédiments intertidaux (Dawson & Pace, 2002 ; Stoeck *et al.*, non publié), dans de l'eau de mer anaérobie à différentes profondeurs (Stoeck et Epstein, 2003 ; Zuendorf *et al.*, 2006 ; Stoeck *et al.*, 2006 ; Edgcomb *et al.*, 2011a), dans la couche supérieure de la colonne d'eau (Massana *et al.*, 2004), dans un biofilm bactérien à 1575 m de profondeur (Bass *et al.*, 2007), associée à un animal au niveau de sources hydrothermales à 860 et 1 700 m de profondeur (Le Calvez *et al.*, 2009) et dans les sédiments d'un suintement froid de méthane (2 400 m) entre 0 et 30 cm sous la surface (Thaler *et al.*, 2012). Richards *et al.* (2012) ont répertorié douze nouveaux phylotypes de champignons marins n'ayant jamais été décrits auparavant, même en milieu continental. Ces études n'ont pas toutes utilisé les mêmes couples d'amorces, certaines ciblant des Eukaryota et d'autres spécifiquement des champignons (Cf. VI – Les champignons marins), cela suggère que la diversité des Chytridiomycota en milieu marin pourrait être plus importante (Nagano *et al.*, 2010 ; Richards *et al.*, 2012).

Les chytrides ont donc été très peu étudiés en milieu naturel, et sont la plupart du temps intégrés dans des groupes fonctionnels, morphologiques ou taxonomiques. Cependant, il a été montré qu'ils peuvent jouer un rôle clé dans certains systèmes comme les écosystèmes lenticques (Sime-Ngando *et al.*, 2011) et des sols où les producteurs primaires ne sont pas des plantes (Freeman *et al.*, 2009). L'utilisation de méthodes moléculaires les ciblant de manière spécifique pourrait être un moyen de les étudier plus précisément pour surmonter les difficultés liées à leur rareté dans certains écosystèmes.

## VI – Les champignons marins

## A – Profil écologique

Le milieu marin est un écosystème très peu étudié par les mycologues comparativement au milieu continental. Les premières observations, réalisées au XIX siècle, rendaient compte de champignons associés à des algues dérivant avec le courant ou en voie de décomposition (Jones, 2011). L'étude de Barghoorn & Linder (1944) est probablement celle qui a créé un nouvel intérêt pour la mycologie marine (Kohlmeyer & Kohlmeyer, 1979 ; Jones, 2011). En examinant des échantillons de bois immergé continuellement dans l'eau de mer ou exposé aux marées dans des mangroves, ils ont observé des champignons analogues aux formes terrestres mais qui présentaient des adaptations aux conditions marines (Barghoorn & Linder, 1944). Suite à ces travaux, le nombre de champignons marins décrits par année a augmenté (Kohlmeyer & Kohlmeyer, 1979). Dans l'objectif d'étudier le rôle et l'impact des champignons en milieu marin, Kohlmeyer et Kohlmeyer (1979) ont proposé une définition écologique de ces formes fongiques observées en milieu marin, qui est de nos jours communément admise : "*Les champignons marins obligatoires se développent et se reproduisent exclusivement en milieu marin ou estuarien ; les champignons marins facultatifs sont observés en milieu continental (i.e. milieu terrestre et eau douce) et sont également capables de se développer (et probablement de se reproduire) en milieu marin.*" Le nombre actuel d'espèces marines obligatoires de champignons supérieurs est de 549 dont 54 seulement décrites après l'an 2000 (Jones, 2011). Près de 450 d'entre elles sont des Ascomycota, les autres étant des anamorphes ou des Basidiomycota. La plus grande partie de ces espèces sont lignicoles (*i.e.* associées à un substrat en bois) dans les zones tempérées et dans les mangroves tropicales (Jones, 2011). Les algues représentent le second réservoir d'espèces de champignons marins (Bugni & Ireland, 2004). On trouve également des espèces fongiques saprophytes ou endophytes, sur des plantes marines (*e.g.* *Spartina*, *Posidonia*) en milieux tropical ou tempéré. D'autres sont isolées à partir de sable, de coraux, d'algues calcaires, de coquilles de mollusques, d'exosquelettes d'hydrozoaires ou de tubes d'annélides. Enfin, des animaux marins comme les éponges forment également des associations avec les champignons (Gao *et al.*, 2008).

La plupart de ces espèces font partie de la diversité étudiée en zone littorale et sublittorale, en examinant au microscope les structures de fructification sur des substrats marins ou en les isolant sur des milieux de culture. Les méthodes moléculaires, de plus en plus utilisées pour l'identification, ont quant à elles permis de mettre en évidence des souches endophytes (Jones, 2011).

En milieu marin profond, les premières études se basaient sur l'observation des structures de fructification sur des échantillons de bois et des panneaux couverts de polyuréthane (Jones, 2011) ou sur l'incubation de sédiments ou d'eau de mer sous pression ambiante (Kohlmeyer & Kohlmeyer, 1979). Ces techniques ne permettent pas de discerner si l'organisme est un champignon indigène du milieu marin profond ou si c'est un champignon contaminant, *i.e.* provenant d'un autre habitat (Kohlmeyer & Kohlmeyer, 1979). Les approches culturales et moléculaires ont permis d'approfondir la description de la diversité fongique marine dans divers systèmes marins à différentes profondeurs (*Cf.* ci-dessous). De plus, des capacités physiologiques qui pourraient permettre aux champignons de coloniser des milieux marins profonds, ont été mis en évidence. Par exemple, Fernandes *et al.* (2004) ont montré que l'expression des gènes de *Saccharomyces cerevisiae* était modifiée par l'application d'une pression de 200 MPa pendant 30mn à des cellules. La réponse de stress spécifique à l'augmentation de la pression hydrostatique se manifeste physiologiquement par une modification de la composition de la membrane, suggérant l'existence de mécanismes d'adaptation à des environnements à haute pression comme les milieux marins profonds (*Cf.* synthèse bibliographique de Simonato *et al.*, 2006).

## B – Fonctions écologiques

Les champignons jouent un rôle important de décomposeur de substrats ligneux (*e.g.* bois) flottant dans les écosystèmes marins comme les mangroves ou d'autres zones côtières, relatif à leur fonction lignocellulolytique (Hyde *et al.*, 1998). Ainsi, ils ont un rôle clé puisqu'ils fractionnent un polymère complexe en matière organique particulaire qui devient utilisable pour d'autres organismes du réseau trophique. On l'observe également dans des systèmes riches en carbone



organique dissous produit par des organismes autotrophes (*e.g.* phytoplancton). Dans les eaux côtières d'Hawaï, denses en producteurs primaires, Gao *et al.* (2010) ont décrit une grande diversité d'Ascomycota et de Basidiomycota jusqu'à 100 m de profondeur. Un phénomène similaire a été observé au niveau des écosystèmes d'upwelling où des eaux profondes et froides remontent à la surface chargées en éléments minéraux induisant une augmentation de la biomasse autotrophe (Gutiérrez *et al.*, 2011). Au niveau du courant de Humboldt (Chili), Gutiérrez *et al.* (2011) ont mis en évidence une biomasse fongique importante sous forme d'hyphes et de spores, associée à la présence d'enzymes extracellulaires hydrolytiques. Ils ont observé que les champignons réalisaient jusqu'à 90% de l'hydrolyse extracellulaire du carbone photosynthétique pendant la saison haute de production. Des champignons dans ce type de système océanique auraient donc une forte implication dans le cycle du carbone comme c'est le cas en milieu continental. Les champignons agissent sur un autre cycle biogéochimique, celui de l'azote. Il a été montré que les champignons utilisent du nitrate pour leur respiration dans des sédiments suboxiques, participant ainsi à la dénitrification anaérobie de l'environnement marin (Jebaraj & Raghukumar, 2009). On retrouve également des champignons en symbiose avec des algues microscopiques ou des cyanobactéries sous forme de lichens, avec des algues macroscopiques formant ainsi de mycophycobioses, mais ces symbioses sont très peu documentées (Hyde *et al.*, 1998). Des communautés fongiques sont associées avec des éponges (Gao *et al.*, 2008) mais il n'a pas encore été déterminé de quel type d'interaction biologique il s'agit. Enfin, certains champignons marins sont des pathogènes d'animaux ou d'algues pouvant causer des infections sévères (Hyde *et al.*, 1998). Van Dover *et al.* (2007) ont identifié une levure pathogène comme étant l'agent responsable de la détérioration des tissus de modioles, pouvant entraîner la décimation de la population dans le bassin des Fidji. De même, les champignons endolithiques sont considérés comme des parasites nuisibles aux récifs de corail ou aux gorgones (*Cf.* Figure 17) puisqu'ils peuvent provoquer des épizooties pouvant mener à la mort de la colonie si celle-ci est en plus exposée à un stress environnemental (Alker *et al.*, 2001). Ce sont des champignons qui creusent des substrats carbonatés (*e.g.* squelettes de coraux, gorgones mais aussi formations calcaires, coquilles de mollusques...), pour se nourrir de matière organique minéralisée (Golubic *et al.*, 2005).



**Figure 17 : Photo de gorgone (*Annela sp.*) parasitée par un champignon du genre *Aspergillus*.** Photo de Sakanan Plathong, extraite de Jones *et al.* (2009).

### C – Applications biotechnologiques

Les champignons sont des sources importantes de métabolites actifs et de nouvelles structures chimiques, qui peuvent être utilisés en pharmacologie ou dans les processus industriels (Cf. synthèse bibliographique de Imhoff *et al.*, 2011). Et les champignons marins en particulier sont d'un grand intérêt puisqu'ils ont le potentiel d'apporter une nouvelle diversité biosynthétique (Bugni & Ireland, 2004). Il a également été montré que certains champignons pourraient être utilisés en biorémédiation pour leur activité enzymatique liée à la dégradation de la lignine (Hyde *et al.*, 1998 ; Raghukumar, 2000). Des études sont en cours sur leur capacité à casser les hydrocarbures, ce qui pourrait être exploité pour dégrader les nappes de pétroles d'origine accidentelle (Jones, 2011).

### D – Les champignons dans les inventaires eucaryotes

La description de la diversité des champignons a été approfondie par l'application des techniques moléculaires (*e.g.* Richards & Bass, 2005). Concernant le milieu marin, la présence de champignons a dans un premier temps été révélée à l'aide d'amorces nucléotidiques spécifiques à la séquence du gène de l'ARNr 18S du domaine des Eukaryota. En pleine mer, les champignons

semblent constituer une fraction minime de la communauté eucaryote (0,8%) dans la partie supérieure de la colonne d'eau (Massana & Pedrós-Alió, 2008). Par exemple, une seule séquence (Ascomycota) a été mise en évidence en Atlantique Nord alors que l'étude portait également sur la mer d'Alboran en Méditerranée et les mers de Weddel et de Scotia en Antarctique (Díez *et al.*, 2001). Quant à Massana *et al.* (2004), ils ont retrouvé quatre phylotypes dans des échantillons prélevés à 800 m au large de la baie de Blanes (Espagne) en Méditerranée : un Basidiomycota, un Chytridiomycota et deux séquences nouvelles assignées récemment aux Cryptomycota par Richards *et al.*, 2012. La colonne d'eau supérieure est souvent pauvre en nutriments et les eucaryotes présents sont majoritairement représentés par des unicellulaires autotrophes (les producteurs primaires) ou phagotrophes (Richards *et al.*, 2012). Elle est donc moins favorable aux champignons qui se nourrissent par osmotrophie, entre autre car les enzymes sécrétées, pour la digestion précédant l'absorption, auraient tendance à se dissiper dans le milieu liquide par diffusion (Richards *et al.*, 2012).

Dans d'autres systèmes marins, qui sont plus éloignés de la surface et des côtes ou qui présentent d'autres caractéristiques (*e.g.* milieux anoxiques, sites hydrothermaux), les champignons ont également été détectés en utilisant la même technique (*i.e.* amorces universelles eucaryotes). Le tableau 1 présente les affiliations phylogénétiques qui ont été identifiées lors de ces analyses.

**Tableau 1 : Comparaison des affiliations phylogénétiques fongiques des séquences de petites sous-unités ribosomiques obtenues lors d'études ciblant le domaine eucaryote dans différents systèmes marins.** Les prélèvements présentés sont tous de nature liquide sauf un seul, les modules de colonisation expérimentaux qui sont des substrats de différentes natures et qui dans cette étude (López-García *et al.*, 2003), sont exposés au fluide hydrothermal pendant 15 jours. L'affiliation phylogénétique indiquée correspond à l'espèce ou au genre frère des séquences environnementales déterminé par les analyses phylogénétiques de chaque étude. Ou alors, dans certains cas, l'analyse phylogénétique menée par Richards *et al.* (2012) a permis de mettre à jour ces affiliations à l'aide de séquences déposées dans GenBank postérieurement aux études originales. Enfin, si seul un nom de phylum est indiqué, les séquences environnementales constituent un nouveau clade fongique (Richards *et al.*, 2012).

Référence	Localisations	Natures des échantillons	Affiliations phylogénétiques
López-García <i>et al.</i> , 2001	Passage de Drake, Antarctique	Eau de mer froide oligotrophe à 3 000 m	Ascomycota : <i>Eupenicillium javanicum</i>
López-García <i>et al.</i> , 2003	Sources hydrothermales Lucky Strike Ride Médio-Atlantique	Modules de colonisation expérimentaux à 1 695 m Interface fluide / eau de mer	Ascomycota : <i>Eupenicillium javanicum</i> Basidiomycota : <i>Malassezia</i> (Richards <i>et al.</i> , 2012)
Stoeck & Epstein, 2003	Etang tidal dans le marais salé de Sippewissett, Cape Cod, Etats-Unis	Eau de mer microaérobie < à 1,5 m	1 Chytridiomycota 2 Cryptomycota (Richards <i>et al.</i> , 2012)
Stoeck <i>et al.</i> , 2003	Bassin de Cariaco, Vénézuéla	Eau de mer anoxique à 270, 340 et 900 m	Ascomycota : <i>Penicillium namyslowski</i>
Zuendorf <i>et al.</i> , 2006	Mariager fjord, Danemark	Eau de mer anoxique sous l'halocline à 18 m	1 Chytridiomycota (Richards <i>et al.</i> , 2012)
Stoeck <i>et al.</i> , 2006	Bassin de Cariaco, Vénézuéla	Eau de mer anoxique à 340 m	5 Ascomycota : <i>Arthrinium</i> , <i>Penicillium namyslowski</i> , <i>Bipolaris</i> 1 Chytridiomycota : idem Stoeck & Epstein, 2003 (Richards <i>et al.</i> , 2012)
López-García <i>et al.</i> , 2007	Sources hydrothermales Lost City, Ride Médio-Atlantique	Interface fluide / eau de mer entre 750 et 900 m (40 - 75°C)	Basidiomycota : <i>Malassezia</i> Ascomycota : <i>Aureobasidium pullulans</i>

Sur ces sites, entre un et six OTUs ont été retrouvés et ils appartiennent aux phyla des Basidiomycota, Ascomycota, Chytridiomycota et Cryptomycota. Stoeck *et al.* (2006) ont répliqué l'étude de Stoeck *et al.* (2003) estimant que seule une fraction de la diversité de microeucaryotes avait été décrite. Lors de cette deuxième analyse, une seule profondeur a été échantillonnée (*versus* 18 dans Stoeck *et al.*, 2003) mais par contre trois couples d'amorces (*versus* un dans Stoeck *et al.*, 2003) ont été utilisés, assortis d'un effort de séquençage plus conséquent. A partir des résultats obtenus, ils estiment qu'une partie de la diversité de micro-Eukaryota a encore échappé à cette nouvelle analyse car le redondance des OTUs retrouvés par les trois couples d'amorces est seulement de 4%. A noter que concernant les champignons cinq OTUs supplémentaires ont été amplifiés par rapport à la première analyse (*Cf.* Tableau 1). De plus, ces études ciblant l'ensemble des Eukaryota sont rarement exhaustives en partie parce que l'échantillonnage n'est pas mené jusqu'à saturation (Edgcomb *et al.*, 2002 ; Stoeck *et al.*, 2006 ; Richards & Bass, 2005).

Des études ciblant les Eukaryota dans des sédiments ont montré des profils extrêmement variables de la diversité des champignons variant de leur absence sur certains sites (Edgcomb *et al.*, 2002) à la présence exclusive de leurs séquences (Takishita *et al.*, 2006). Dawson & Pace (2002), ont utilisé deux couples d'amorces eucaryotes pour prospecter des sédiments superficiels récupérés à marée basse. Les sédiments saumâtres du Parc Aquatique de Berkeley (Etats-Unis) arborent une communauté eucaryote dominée par les champignons tandis que ces derniers sont beaucoup moins représentés dans la communauté eucaryote des sédiments du lagon de Bolinas (Etats-Unis). A de plus grandes profondeurs, Edgcomb *et al.* (2002) ont prélevé deux carottes de sédiments anoxiques superficiels (1-3cm) ainsi que de l'eau recouvrant l'un des deux points de carottage au niveau des sources hydrothermales de Guaymas (2000m). La stratégie employée a consisté à amplifier le gène de l'ARNr 18S avec un couple d'amorces universelles puis à effectuer une amplification plus spécifique avec un couple d'amorces eucaryotes (*i.e.* PCR nichée). Dans cette analyse, une seule carotte de sédiments comportait des champignons : trois Ascomycota, *Aspergillus*, *Aureobasidium*, *Neurospora* et deux Basidiomycota, *Ustilago* et *Xerocomus*. Ils représentent environ 6% des séquences eucaryotes obtenus. Cependant le problème de contamination exogène se pose puisque Edgcomb *et al.* (2002) ont également amplifié des séquences d'organismes phototrophes, qui nécessitent un rayonnement solaire pour la

photosynthèse. Ces organismes pourraient provenir de contaminations lors de l'échantillonnage. Alternativement, ces séquences pourraient provenir d'individus qui ne sont plus actifs ou enkystés et qui auraient coulé dans la colonne d'eau jusqu'à se déposer au niveau des sédiments (Edgcomb *et al.*, 2002). Afin de décrire la diversité active de la communauté eucaryote, Edgcomb *et al.* (2011b) ont alors extrait l'ARN en plus de l'ADN et comparer les profils du gène de l'ARNr 18S obtenus, et ce pour des sédiments peu profonds (1,75 - 3,9 m sous la surface) et des sédiments plus profonds (35,1 - 37,4 m sous la surface) échantillonnés au niveau de la marge et de la fosse du Pérou. Si l'on regroupe tous les échantillons, ce sont des séquences de Basidiomycota qui ont été retrouvées majoritairement. Elles sont même les seules séquences eucaryotes obtenues au niveau de certains des sites les plus profonds. De plus, les profils de diversité ARN sont différents des profils ADN, les Basidiomycota restant toutefois majoritaires. La diversité ARN ne comportant que peu ou pas d'autotrophes, cela montre que des organismes inactifs et probablement exogènes peuvent biaiser la description de la diversité ADN eucaryote dans les sédiments (Edgcomb *et al.*, 2011b). A l'inverse, Takishita *et al.* (2006) n'ont noté aucune contamination de ce genre puisqu'ils ont amplifiés exclusivement des séquences de *Cryptococcus curvatus* (levures Basidiomycota) lors d'une étude ciblant tous les Eukaryota, dans les sédiments d'un site hydrothermal riche en méthane, au Kuroshima Knoll (640 m). Cependant, lors d'observations au microscope, ils ont noté la présence de rotifères. Il s'avèrent que les séquences du gène de l'ARNr 18S de ces organismes divergent trop des amorces utilisées pour être amplifiées par celles-ci (Takishita *et al.*, 2006). En 2007, Takishita *et al.* ont utilisé les mêmes couples d'amorces sur un site similaire de la baie de Sagami. Ils ont montré la prédominance à 83,9% (site avec mollusques) et 44,6% (site sans mollusque) des séquences de *Cryptococcus curvatus* dans la communauté eucaryote. Le profil d'une diversité eucaryote dans un site donné dépend donc non-seulement de la technique employée, moléculaire *versus* culturale, mais également des molécules ciblées (*i.e.* couples d'amorces, ARN, ADN).

La présence des champignons a donc été détectée lors d'études portant sur les Eukaryota dans divers systèmes marins. Mais ces études ciblant pour la plupart un large spectre phylogénétique d'organismes parmi les protistes et non les champignons, la description de la diversité des champignons est biaisée. Durant cette dernière décennie, des études ciblant

spécifiquement les champignons ont été effectuées par des approches culturelles et moléculaires pour examiner plus finement leur diversité dans divers systèmes marins à différentes profondeurs.

## E – Les champignons dans les inventaires fongiques

### 1) Diversité par des approches culturelles

La description de la diversité fongique avec des approches culturelle et moléculaire spécifiques aux champignons fait état d'une diversité plus conséquente et étendue à un plus grand nombre de groupes fongiques que celle décrite par les études sur le domaine des Eukaryota, dans le milieu marin. Le tableau 2 présente des études récentes de mise en culture de la diversité fongique sur des sites peu profonds (entre 14 et 200 m) et des sites de plus grandes profondeurs (entre 700 et 11 000 m). Plusieurs types d'échantillons marins ont ainsi été examinés : sédiments anoxiques ou non, animaux, eau, échantillons minéraux et des modules de colonisation expérimentaux. Dans l'ensemble, ces études ont révélé la présence de formes filamenteuses et de levures des phyla Ascomycota et Basidiomycota. En revanche, aucun représentant des lignées basales fongiques n'a été obtenu en culture.

**Tableau 2 : Phyla fongiques des souches obtenues par la mise en culture d'échantillons marins de différentes natures prélevés à des profondeurs variables.** Liste des abréviations : EPR : Ride Est-Pacifique, MAR : Ride Médio-Atlantique.

Référence	Localisations & Profondeurs	Natures des échantillons	Affiliations phylogénétiques
Nagahama <i>et al.</i> , 2001	Baies de Sagami et Suruga, Japon (1 000 à 11 000 m)	Sédiments superficiels Animaux	Ascomycota Basidiomycota
Biddle <i>et al.</i> , 2005	Marge du Pérou (150 m)	Sédiments entre 30 et 157 m sous la surface	Ascomycota
Raghukumar <i>et al.</i> , 2004	Fosse des Chagos Ocean indien (5 904 m)	Sédiments entre 0 et 370 cm sous la surface	Ascomycota Anamorphes
Damare <i>et al.</i> , 2006	Bassin central indien (4 900 à 5 390 m)	Sédiments superficiels	Ascomycota
Singh <i>et al.</i> , 2010	Bassin central indien (4 000 à 5 700 m)	Sédiments superficiels	Ascomycota Basidiomycota
Gadanho & Sampaio, 2005	Sites hydrothermaux, MAR (800 à 3 150 m)	Eau : 3 à 5 m au-dessus du plancher océanique	Ascomycota Basidiomycota
Burgaud <i>et al.</i> , 2009	Sites hydrothermaux, MAR et EPR (700 à 3 650 m)	Eau, sédiments, échantillons minéraux, animaux	Formes filamenteuses : Ascomycota Basidiomycota
Le Calvez <i>et al.</i> , 2009	Sites hydrothermaux, MAR (860 & 1 700 m) et EPR (2 630 m)	Animaux Echantillons minéraux	Ascomycota
Burgaud <i>et al.</i> , 2010	Sites hydrothermaux, MAR et Bassin de Lau : Pacifique (900 à 3 630 m)	Eau, animaux, modules de colonisation expérimentaux, mues de crevette	Levures : Ascomycota Basidiomycota
Jebaraj & Raghukumar, 2009	Mer d'Arabie (14 & 26 m)	Sédiments superficiels anaérobies	Ascomycota
Jebaraj <i>et al.</i> , 2010	Mer d'Arabie (3, 25 & 200 m)	Sédiments superficiels anaérobies, eau	Ascomycota Basidiomycota



## 2) Diversité par des approches moléculaires

La méthode moléculaire a mis en évidence une diversité fongique marine étendue à un plus grand nombre de phyla, *i.e.* Ascomycota, Basidiomycota, Chytridiomycota et Cryptomycota. Les phyla retrouvés dans les six premières études portant exclusivement sur la diversité fongique marine en milieu profond et utilisant l'approche moléculaire sont présentés dans le tableau 3. Les marqueurs moléculaires choisis sont principalement le gène de l'ARNr 18S (Bass *et al.*, 2007 ; Le Calvez *et al.*, 2009 ; Nagahama *et al.*, 2011) et les ITS (Lai *et al.*, 2007 ; Nagano *et al.*, 2010). Nagano *et al.* (2010) qui ont également utilisé le gène codant pour l'ARNr 5.8S, ont remarqué que les nombreuses séquences amplifiées n'avaient pas de proches représentants dans la base de données et portaient peu ou pas d'information phylogénétique exploitable. Ils ont donc proposé d'utiliser les gènes codant pour le gène de l'ARNr 18S et 28S, ce qu'ont fait Nagahama *et al.* en 2011.

**Tableau 3 : Phyla fongiques détectés par l'amplification de marqueurs moléculaires spécifiques aux champignons à partir d'échantillons marins de différentes natures prélevés à des profondeurs variables.**

Liste des abréviations : EPR : Ride Est-Pacifique, MAR : Ride Médio-Atlantique.

Référence	Localisations & Profondeurs	Nature des échantillons	Affiliations phylogénétiques
Bass <i>et al.</i> , 2007	Epave du Bismarck (3 000 & 4 000 m)	Eau de mer	Ascomycota Basidiomycota Chytridiomycota
	Epave du Titanic (3 000 & 3 700 m)	Eau de mer	
	Site hydrothermal, MAR (2 264 m)	Sédiments, modules de colonisation expérimentaux	
	Passage de Drake (250, 500, 2 000 & 3 000 m)	Eau de mer	
	Golfe de Californie (1 575 m)	Biofilm bactérien anaérobie	
Lai <i>et al.</i> , 2007	Mer de Chine (350, 884, 1 123, 2 965 & 3 011 m)	Sédiments superficiels avec des hydrates de méthane	Ascomycota Basidiomycota
Le Calvez <i>et al.</i> , 2009	Sites hydrothermaux, MAR et EPR (860 à 2 630 m)	Animaux, échantillons minéraux,	Basidiomycota Chytridiomycota
Nagano <i>et al.</i> , 2010	Fosse d'Izu-Ogasawara (7 111 & 9 760 m)	Sédiments entre 0 et 110 cm sous la surface	Ascomycota Basidiomycota Cryptomycota
	Fosse des Mariannes (10 131 m)	Sédiments	
	Baie de Sagami, suintement froid de méthane (1 174 m)	Biofilm bactérien	
Nagahama <i>et al.</i> , 2011	Baie de Sagami, suintement froid de méthane (830 à 1 200 m)	Sédiments	Ascomycota Basidiomycota Blastocladiomycota Cryptomycota
Thaler <i>et al.</i> , 2012	Golfe de Mexico, suintement froid de méthane (2 400 m)	Sédiments entre 0 et 30 cm sous la surface	Ascomycota Basidiomycota Chytridiomycota

### 3) Diversité des champignons supérieurs

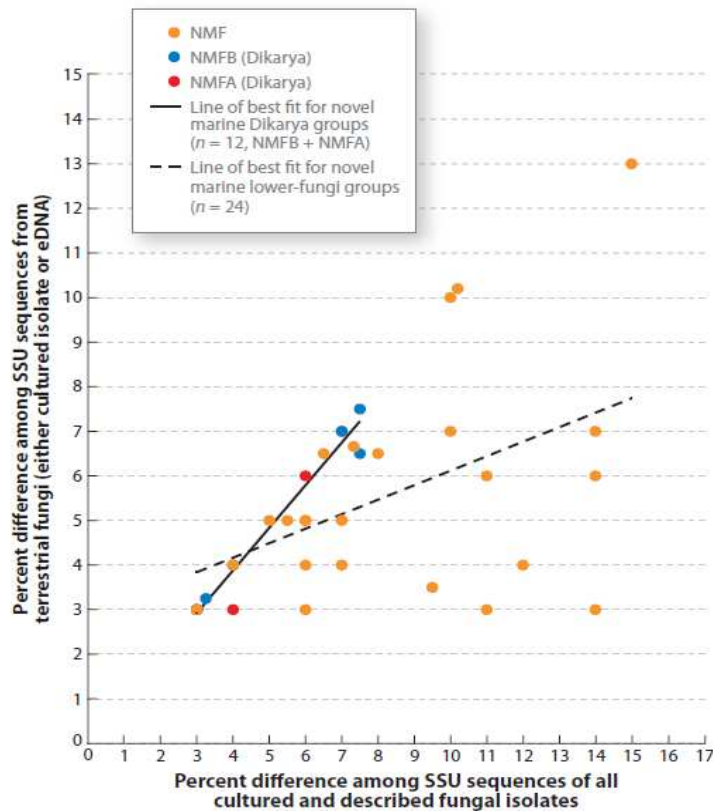
Les communautés fongiques décrites quelque soit l'approche, sont principalement composées de champignons supérieurs, *i.e.* Ascomycota et Basidiomycota, dont les séquences sont souvent proches de celles de champignons terrestres (Richards *et al.*, 2012). Les levures ont été décrites comme les plus représentées, *e.g.* : *Cryptococcus*, *Rhodotorula*, *Candida*, *Debaryomyces* (*e.g.* Gadanho & Sampaio, 2005 ; Bass *et al.*, 2007 ; Le Calvez *et al.*, 2009 ; Burgaud *et al.*, 2010). Des séquences environnementales affiliées aux levures *Malassezia* ont souvent été mises en évidence dans des études fongiques ou eucaryotes portant sur plusieurs types de systèmes marins. Elles constituent un groupe appelé de manière formelle 'groupe des levures hydrothermales et/ou anoxiques' (*e.g.* Dawson & Pace, 2002 ; López-García *et al.*, 2007 ; Bass *et al.*, 2007 ; Le Calvez *et al.*, 2009 ; Jebaraj *et al.*, 2010 ; Edgcomb *et al.*, 2011b). Les formes filamenteuses, *e.g.* *Fusarium*, *Aspergillus*, *Exophiala* et *Tilletiopsis* sont assez bien représentées dans les cultures (*e.g.* Damare *et al.*, 2006 ; Burgaud *et al.*, 2009). On les retrouve également dans la plupart des études moléculaires (*e.g.* López-García *et al.*, 2001 ; Stoeck *et al.*, 2006 ; Bass *et al.*, 2007). Enfin, des champignons produisant des structures multicellulaires en milieu continental sont également présents en milieu marin, *e.g.* *Exidia*, *Coprinus* et *Antrodia* (Bass *et al.*, 2007 ; Le Calvez *et al.*, 2009 ; Jebaraj *et al.*, 2010).

### 4) Diversité des lignées basales

Les lignées basales ont uniquement été détectées par l'approche moléculaire dans les milieux marins profonds (Cf. tableau 3) où elles peuvent constituer une grande partie de la communauté décrite. Dans les sédiments d'un suintement froid de méthane de la baie de Sagami (Nagahama *et al.*, 2011) la communauté fongique est dominée par des organismes des lignées basales. Cinq OTUs forment un clade nouveau, le 'groupe basal de clones I' qui a récemment été déterminé comme phylogénétiquement plus proche des Blastocladiomycota (Richards *et al.*, 2012), cependant il en est assez éloigné puisque l'identité est seulement de 91%. L'un de ces OTUs représente 56,2% du total des quatre banques de clones constituées (Nagahama *et al.*, 2011). Un

OTU appartenant au 'groupe basal de clones I' a également été retrouvé dans un biofilm bactérien anaérobie dans le Golfe de Californie (Bass *et al.*, 2007). Richards *et al.* (2012) dans leur analyse de la diversité des champignons marins montrent que le plus grand nombre d'OTUs est affilié aux Ascomycota et Basidiomycota mais que la majorité des nouveaux OTUs se placent parmi les champignons des lignées basales. Ils considèrent qu'un OTU correspond à un nouveau champignon marin s'il présente un minimum de 3% de divergence avec les séquences contenues dans la base de données moléculaires GenBank (Benson *et al.*, 2011). Ils ont trouvé que 24 des 36 nouveaux OTUs marins s'intègrent aux lignées basales. Toutefois il faut soustraire un OTU (GQ995437), qui provient de l'étude de la diversité des champignons de Freeman *et al.* (2009) sur des sites terrestres de hautes altitudes. De plus, ces nouveaux OTUs marins des lignées basales sont plus divergents des séquences de champignons continentaux de GenBank que les nouveaux OTUs marins des Dikarya et ils semblent être très éloignés des espèces cultivées et décrites de champignons (Cf. Figure 18).

La méthode culturale et la méthode moléculaire semblent donner accès à des diversités différentes, ce qui suggère que la communauté fongique en milieu marin serait encore plus complexe que celle décrite à ce jour (Le Calvez *et al.*, 2009 ; Richards *et al.*, 2012). Dans notre étude de la communauté fongique à des sites hydrothermaux de la ride Médio-Atlantique (Le Calvez *et al.*, 2009), nous avons obtenu neuf phylotypes appartenant aux Basidiomycota et Chytridiomycota via l'outil moléculaire tandis que les onze phylotypes cultivés sont tous des Ascomycota. Jebaraj *et al.* (2010) ont également détecté des séquences d'Ascomycota et de Basidiomycota différentes selon la méthode employée. La différence est toutefois moins marquée puisque certains phylotypes comportent des séquences issues de chacune des deux méthodes. L'hypothèse d'une diversité plus étendue que celle perçue jusqu'à présent est également renforcée par le fait que des couples d'amorces différents permettent d'accéder à des profils de diversité différents (*e.g.* Edgcomb *et al.*, 2002 ; Stoeck *et al.*, 2006 ; Jebaraj *et al.*, 2010).



**Figure 18 : Graphique comparant la différence de pourcentage d'identité entre les nouveaux OTUs marins fongiques et les séquences du gène de l'ARNr 18S des champignons continentaux issues d'échantillons environnementaux *versus* la différence de pourcentage d'identité entre les nouveaux OTUs marins fongiques et les séquences du gène de l'ARNr 18S d'espèces de champignons cultivées et décrites.** Les points oranges (Novel Marine Fungi : NMF) correspondent aux lignées basales. Les NMF présentent des niveaux de différences plus élevés avec les espèces cultivées (en abscisse) qu'avec l'ensemble des séquences de champignons continentaux (en ordonnée). Cela signifie que les affiliations des NMF sont plus proches de séquences provenant d'échantillons environnementaux plutôt que des collections de cultures. Liste des abréviations : eDNA, ADN environnemental ; n, nombre de phylotypes ; NMFA, nouveaux Ascomycota marins ; NMFB, nouveaux Basidiomycota marins. Figure extraite de Richards *et al.* (2012)

Depuis ces cinq dernières années, un renouveau de la connaissance de la diversité fongique des lignées basales vient de la publication de travaux essentiellement moléculaires décrivant une diversité originale jusqu'alors non suspectée, dans différents milieux océaniques. Ces travaux offrent des opportunités pour l'approfondissement des connaissances à la fois sur un plan systématique et sur un plan fonctionnel. Ils soulèvent également la question de l'évolution des champignons en remettant en cause l'idée assez dogmatique selon laquelle les champignons sont des organismes continentaux qui auraient coloniser secondairement le milieu marin. Et au-delà du règne fongique, ils interrogent sur les grands phénomènes de radiation chez les Eukaryota notamment celle entre les animaux et les champignons ainsi que leur trajectoire évolutive respective.

## Présentation des objectifs

---

Comme il a été vu précédemment, les Chytridiomycota ont fait l'objet d'un nombre limité d'études. De plus, les cas sur lesquels une attention particulière a été portée, ont tous en commun le fait d'être lié à l'homme pour des raisons économique ou éthique. Toutefois, l'étude des Chytridiomycota s'est développée récemment dans certains écosystèmes grâce à l'utilisation de nouvelles méthodes d'analyse de leur diversité, ce qui permet de contourner les difficultés limitant leur description par l'observation ou l'isolation. Ainsi, il a été montré par des analyses de génomiques environnementales que les communautés fongiques étaient dominées par les Chytridiomycota dans les lacs (Lefèvre *et al.*, 2007 ; 2008) et dans les sols de hautes montagnes (Freeman *et al.*, 2009).

Ce travail s'attache à décrire la diversité du phylum des Chytridiomycota en milieu hydrothermal marin profond en utilisant des techniques de génomique environnementale. Au-delà du phylum en lui-même, l'objectif est d'accéder à la diversité des organismes des lignées basales fongiques. En effet, les Chytridiomycota constituent le phylum qui a divergé en premier du règne des champignons et son positionnement est stratégique pour l'acquisition de données permettant la compréhension de l'histoire évolutive des champignons. Une étude récente de la diversité fongique en milieu hydrothermal profond a mis en évidence des séquences de phylotypes de Chytridiomycota qui n'avaient pas encore été observés (Le Calvez *et al.*, 2009). C'est pourquoi ce type d'écosystème semble pertinent pour explorer la diversité des Chytridiomycota. Afin d'analyser cette diversité, il a été mis en place une base de données qui contient des marqueurs phylogénétiques pour le règne des champignons. L'intérêt principal réside dans l'application automatique de filtres de qualité, complétée par un tri manuel qui permet d'écarter les séquences erronées, non valides pour les analyses phylogénétiques.

Le travail de thèse est présenté de la manière suivante :

## Matériel et Méthodes

Cette partie retrace de manière succincte le cheminement suivi pour pouvoir répondre à la problématique. Ces analyses ayant été entravées par des problèmes méthodologiques, ceux-ci y sont exposés avec les solutions employées pour les résoudre.

### Chapitre 1

Ce chapitre présente une publication portant sur PHYMYCO-DB, une base de données contenant des séquences moléculaires directement utilisables pour des analyses phylogénétiques, des dessins d'amorces etc... Elle est actuellement en cours de soumission à PLoS ONE.

### Chapitre 2

La publication de Le Calvez *et al.*, 2009 est présentée dans ce manuscrit d'une part parce qu'elle est à l'origine de ce travail de thèse mais également parce que j'ai contribué à sa production avec les autres auteurs, à plusieurs niveaux : obtention des séquences moléculaires, analyses phylogénétiques et rédaction du manuscrit publié. L'objectif de cette étude était de décrire la diversité du règne des champignons en milieu hydrothermal, au niveau de différentes zones géographiques. Parmi les séquences obtenues, il a été mis en évidence des séquences nouvelles des lignées fongiques basales.

### Chapitre 3

Ce chapitre correspond à la publication produite à partir des résultats obtenus sur la diversité des Chytridiomycota en milieu hydrothermal marin profond. Cette publication est en préparation pour la soumettre à Proceedings of the Royal Society B.





# MATERIEL ET METHODES



## Matériel et Méthodes

---

Dans la partie Matériel et Méthodes sera présenté l'ensemble des moyens mis en œuvre pour pouvoir répondre à la problématique abordée dans cette thèse. Toutefois, les protocoles seront décrits de manière succincte puisqu'ils sont exposés plus en détail dans chacune des publications. L'objectif ici est avant tout d'expliquer les choix méthodologiques qui ont été faits. Dans un premier temps, les démarches effectuées pour mettre en place une base de données de séquences moléculaires fongiques seront détaillées. Puis, l'amplification des séquences de Chytridiomycota ayant nécessité le design d'amorces spécifiques, les différentes étapes allant du choix du marqueur moléculaire à la validation des amorces seront décrites. Enfin, le cheminement suivi pour acquérir les données moléculaires et les analyser sera abordé en détails en exposant les problèmes techniques survenus et les ajustements effectués. Egalement, les échantillons seront présentés selon leur provenance.

### I – Création d'une base de données dédiée aux analyses phylogénétiques fongiques

L'un des objectifs de cette thèse étant l'étude de la diversité moléculaire des Chytridiomycota par le biais d'inférences phylogénétiques, cela a nécessité la confrontation des séquences environnementales obtenues avec des séquences appartenant à des organismes cultivés et décrits ainsi qu'avec d'autres séquences environnementales. Les séquences produites lors d'une étude doivent être déposées dans une base de données publique (*e.g.* GenBank) pour qu'il leur soit attribué un numéro d'accession, facilitant ainsi le suivi des séquences entre les différentes études. Cependant, cette procédure ne garantit pas la qualité des séquences et l'accumulation de séquences erronées est de plus en plus constatée (Bridge *et al.*, 2003; Vilgalys,

2003; Bridge *et al.*, 2004; Hawksworth, 2004, 2009; Holst-Jensen *et al.*, 2004; Nilsson *et al.*, 2006; Bidartondo, 2008). Cette propagation étant difficilement maîtrisable, notre solution fut de mettre en place une base de données de séquences utilisables en phylogénie, triées selon plusieurs critères de qualité. D'autres bases de données existent déjà, cependant elles sont soit universelles, *i.e.* dédiées aux séquences de tous les organismes vivants (*e.g.* Silva, Pruesse *et al.*, 2007) soit exclusives d'un groupe taxonomique (*e.g.* MaarjAM, Öpik *et al.*, 2010) ou encore dédiées à l'identification taxonomique à l'aide d'*ITS* (*e.g.* UNITE, Abarenkov *et al.* 2010a). Silva aurait pu répondre à nos attentes, cependant son utilisation s'avère complexe de par son étendue taxonomique. Ainsi, la base de données PHYMYCO-DB a été créée pour l'analyse des relations phylogénétiques du règne des champignons. Cet outil bioinformatique a été élaboré en collaboration avec l'IRISA (Institut de Recherche en Informatique et Systèmes Aléatoires, Rennes, France). Cette base de données est accessible sur le site internet suivant : <http://phymycodb.genouest.org/>. PHYMYCO-DB faisant l'objet d'une publication présentée dans le chapitre 1, seules les étapes de sa mise en place et ses fonctionnalités seront présentées ici.

Dans un premier temps, deux marqueurs moléculaires ont été choisis. Le premier est le gène de l'ARNr 18S dont la séquence est très conservée entre les différents taxons, ce qui le rend porteur d'informations évolutives de qualité. C'est aussi le gène le plus représenté dans les bases de données, couvrant ainsi un très grand nombre d'organismes. Le second est le gène du facteur d'élongation 1 $\alpha$  (EF1- $\alpha$ ), qui a été beaucoup moins séquencé mais qui du fait de son taux de mutations plus élevé que celui du gène de l'ARNr 18S, permet l'étude des relations phylogénétiques entre les niveaux taxonomiques supérieurs.

Les étapes suivantes sont réalisées quatre fois par an afin que PHYMYCO-DB soit à jour vis-à-vis des séquences régulièrement déposées sur GenBank. Tout d'abord, l'équipe de l'IRISA se charge d'extraire les séquences d'intérêt à partir de la base de données GenBank. Ensuite, ces séquences sont filtrées de manière automatique selon leur taille et leur composition nucléotidique, *i.e.* si une séquence comporte plus de dix nucléotides indéterminés consécutifs, elle n'est pas conservée. Ces filtres automatiques ne permettant pas de détecter certains types d'erreur, nous effectuons une étape manuelle de tri des séquences. Les erreurs régulièrement rencontrées sont illustrées dans la

figure 2 du Chapitre 1. Avant la mise en place en routine de la récupération des séquences pertinentes pour les analyses phylogénétiques, de nombreux échanges avec l'équipe de l'IRISA ont eu lieu afin d'ajuster l'outil bioinformatique à la réalité taxonomique. Par exemple, GenBank propose une taxonomie à partir de laquelle une arborescence est constituée sur PHYMYCO-DB pour la récupération des séquences. Or des problèmes inhérents au site internet de GenBank ont fait que des séquences n'ont plus d'attribution taxonomique et du coup, elles se retrouvent classées dans un mauvais groupe taxonomique sur PHYMYCO-DB. Ce type d'erreur doit être vérifié à chaque mise à jour. Egalement, d'un point de vue strictement biologique, le tri manuel a nécessité une connaissance experte des séquences fongiques qui a été acquise au fur et à mesure de la découverte du profil des séquences. Par exemple, la présence de longues insertions (*i.e.* 50 à 500 nucléotides) a été remarquée dans plusieurs séquences lors du premier tri. Après une recherche bibliographique, il s'est avéré que ces insertions correspondaient à des introns, qui sont d'ailleurs très répandus chez les Ascomycota associés aux algues dans les lichens, (*Cf.* Chapitre 1). L'approfondissement des connaissances sur le profil des séquences a donc impliqué un traitement différentiel des séquences.

Ainsi les utilisateurs de PHYMYCO-DB peuvent télécharger des séquences du gène de leur choix, qui respectent les critères de qualité que nous avons définis (*Cf.* Chapitre 1). Cela se fait à partir d'une arborescence dans laquelle apparaissent tous les niveaux taxonomiques, du phylum à l'espèce et à partir de chacun desquels, on peut obtenir un fichier .fasta contenant les séquences correspondantes (*Cf.* Figure 1).

Il est également possible pour l'utilisateur de télécharger des alignements effectués avec CLUSTAL W (Larkin *et al.* 2007). Ces alignements sont réalisés séparément pour chacun des gènes et pour un groupe taxonomique donné, dans la limite de 400 séquences (*Cf.* Figure 2). Enfin, l'utilisateur peut utiliser PHYMYCO-DB pour aligner, avec CLUSTAL W, ses propres séquences avec des séquences ou un alignement extrait de PHYMYCO-DB.

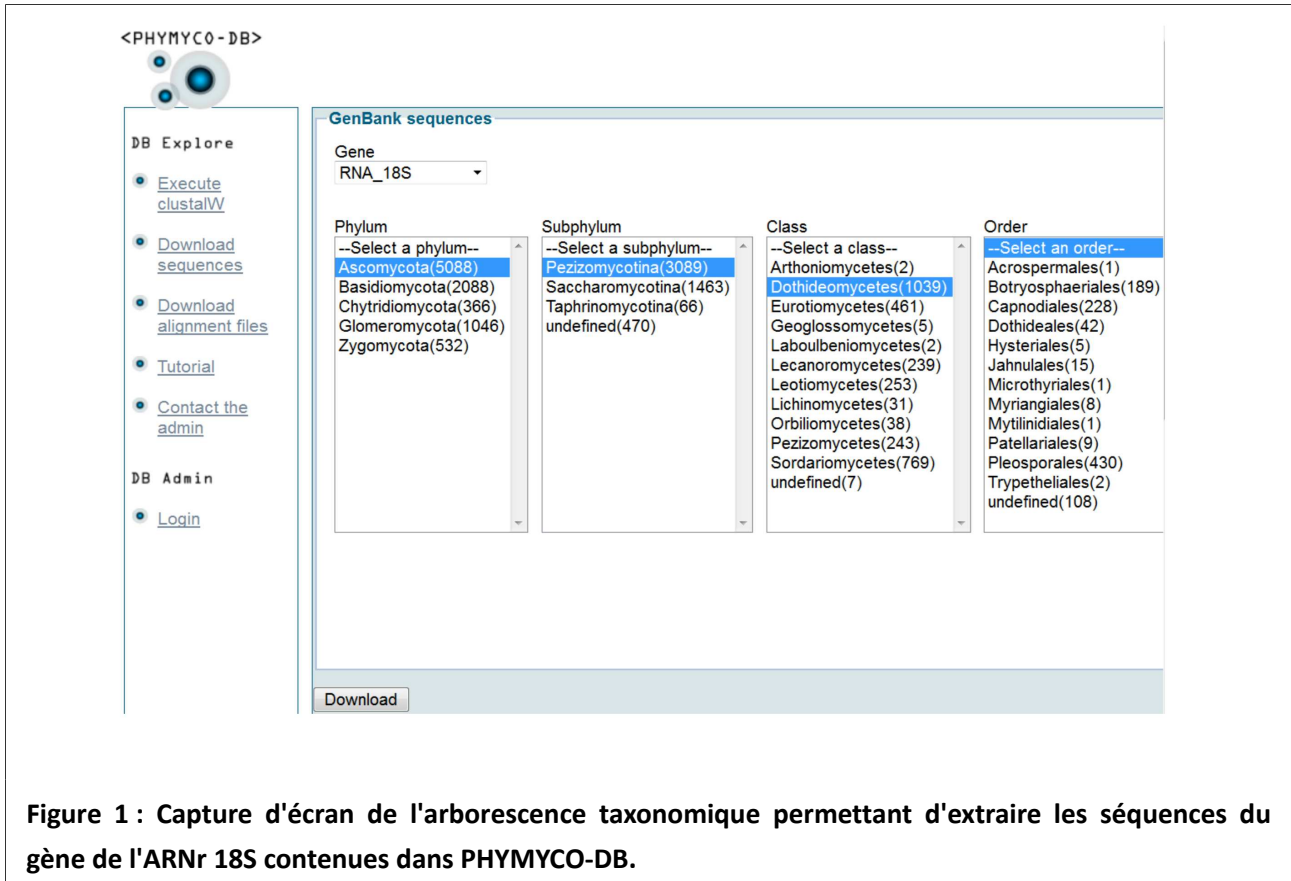


Figure 1 : Capture d'écran de l'arborescence taxonomique permettant d'extraire les séquences du gène de l'ARNr 18S contenues dans PHYMYCO-DB.

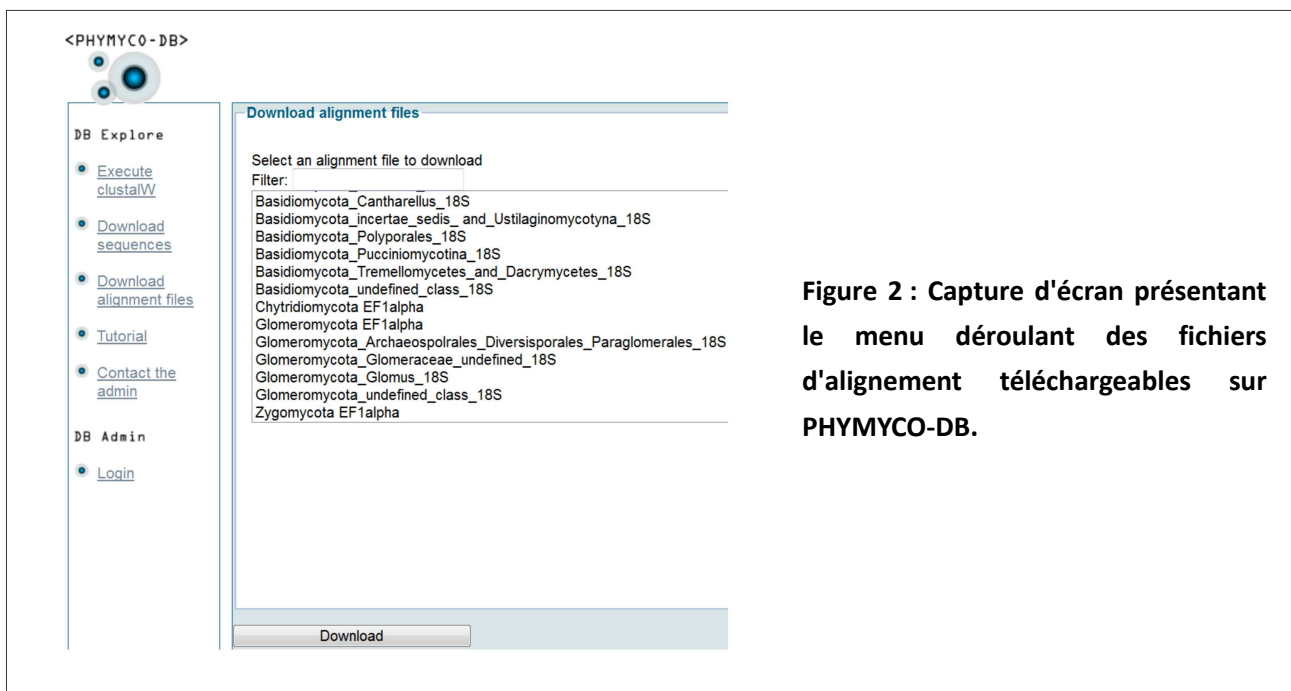


Figure 2 : Capture d'écran présentant le menu déroulant des fichiers d'alignement téléchargeables sur PHYMYCO-DB.

PHYMYCO-DB est donc une base de données fiable dont la principale fonction est l'analyse phylogénétique du règne des champignons. De plus, elle s'avère très utile et pratique pour le design d'amorces moléculaires et l'alignement rapide de séquences environnementales dans un groupe taxonomique donné. Son utilisation apporte une complémentarité non-négligeable avec les autres bases de données existantes et les autres méthodes d'analyses de séquences (*e.g.* l'alignement local proposé par BLASTn). C'est un outils voué à évoluer par l'ajout d'autres fonctionnalités (*e.g.* d'autres marqueurs moléculaires) et par l'intégration d'autres projets collaboratifs. Outre les avantages pour l'analyse phylogénétique que PHYMYCO-DB procurent (*Cf.* Chapitre 1), l'interface réalisée par l'équipe de l'IRISA est très conviviale.

## II – Design d'amorces moléculaires

Ce travail constituant la première analyse de génomique environnementale portant sur la diversité des Chytridiomycota, des outils spécifiques ont été mis au point, à commencer par les amorces moléculaires qui vont permettre d'amplifier des fragments d'ADN de ce groupe. Concernant le choix du marqueur, le gène de l'ARNr 18S s'est positionné comme un bon candidat pour les raisons indiquées ci-dessus et dans le Chapitre 1, mais également parce qu'ils ont été peu étudiés et donc les bases de données comportent très peu de séquences pour les autres marqueurs.

Une première sélection d'amorces a été effectuée à partir de publications sur les études de la diversité fongique et de la page web du laboratoire de R. Vilgalys (Duke University) : <http://www.biology.duke.edu/fungi/mycolab/primers.htm>. Ensuite, il a été vérifié que ces amorces étaient compatibles avec les Chytridiomycota. Pour ce faire, la matrice d'alignement correspondant au phylum des Chytridiomycota a été téléchargée à partir de PHYMYCO-DB et les différentes amorces ont été alignées contre cette matrice.



**Tableau1 : Liste des amorces s'alignant avec la matrice des Chytridiomycota.** Les amorces BMB-AR à NS8 sont extraites de la page web du Laboratoire de R. Vilgalys avec les amorces 'NS' provenant de White *et al.* (1990) tel qu'il est indiqué sur la page web. L'amorce EF3R a été dessinée par Smith *et al.* (1999) et MH2/MH4 par Vandenkoornhuyse *et al.* (2002). L'amorce C130 a été dessinée à la main.

Amorce	Séquence 5'-3' (brin sens)	Résultat de BLASTn
BMB-AR	CAGCAGCCGCGGTAATTCCAGC	Bactéries, Basidiomycota, Ascomycota, Chytridiomycota, Bryozoa, Metazoa, Choanoflagellida, Alveolata
BMB-CR	CAGCAGCCGCGGTAATTCCAGC	Bactéries, Basidiomycota, Ascomycota, Chytridiomycota, Bryozoa, Metazoa, Algues, Amoebozoa
SR1	TACCTGGTTGAT[C/T]CTGCCAGT	Basidiomycota, Ascomycota, Chytridiomycota, Bryozoa, Plantes, Champignons <i>incertae sedis</i>
SR2R	TGGTGGTGCATGGCCG	Bactéries, Basidiomycota, Ascomycota, Chytridiomycota, Bryozoa, Oomycota, Metazoa
SR3	TGCCCTATCAACTTTCG	Bactéries, Basidiomycota, Ascomycota, Chytridiomycota, Metazoa, Choanoflagellida, Algues, Plantes
SR6R	AAGTAAAAGTCGTAACAAGG	Bactéries, Basidiomycota, Ascomycota, Chytridiomycota, Metazoa, Bryozoa, Plantes
SR7R	AGTTAAAAGCTCGTAGTTG	Bactéries, Basidiomycota, Ascomycota, Chytridiomycota, Bryozoa, Metazoa, Choanoflagellida, Algues, Ciliés
SR8R	GATCCAGGACTTTTACTTT	Ascomycota, Chytridiomycota, champignons <i>incertae sedis</i> , Eukaryota marins non cultivés
SR9	CAGAGGTGAAATTCT	Bactéries, Basidiomycota, Ascomycota, Chytridiomycota, Bryozoa
SR10R	TTTGACTCAACACGGG	Bactéries, Basidiomycota, Ascomycota, Chytridiomycota, Bryozoa, Eukaryota marins non cultivés
NS3	GCAAGTCTGGTGCCAGCAGCC	Bactéries, Basidiomycota, Ascomycota, Chytridiomycota, Bryozoa, Metazoa, Alveolata, Algues
NS5	AACTTAAAGGAATTGACGGAAG	Bactéries, Basidiomycota, Ascomycota, Chytridiomycota, Bryozoa, Eukaryota marins non cultivés
NS7	GAGGCAATAACAGGTCTGTGATGC	Bactéries, Basidiomycota, Ascomycota, Chytridiomycota, Amoebozoa, Eukaryota marins non cultivés
NS8	TCCGTAGGTGAACCTGCGGA	Bactéries, Basidiomycota, Ascomycota, Chytridiomycota, Bryozoa, Metazoa, Algues, Plantes, Champignons <i>incertae sedis</i> , Oomycota, Ciliés
C130	TACCTTACTACTTGGATAACCGTG	Bactéries, Basidiomycota, Ascomycota, Chytridiomycota, Haptophyceae, Champignons <i>incertae sedis</i>
EF3R	CAAACCTGGTCATTTAGAGGAA	Bactéries, Basidiomycota, Ascomycota, Chytridiomycota, Bryozoa, Champignons <i>incertae sedis</i>
SR8Rmod	ATCCAGGACTTTTACTTTG	Ascomycota, Chytridiomycota, Alveolata
MH2	TTCGATGGTAGGATAGAGG	Basidiomycota, Ascomycota, Chytridiomycota, Glomeromycota, Algues, Haptophyceae
MH4	GAATGGCTTAGTGAC	Bactéries, Basidiomycota, Ascomycota, Metazoa, Algues, Plantes, Ciliés

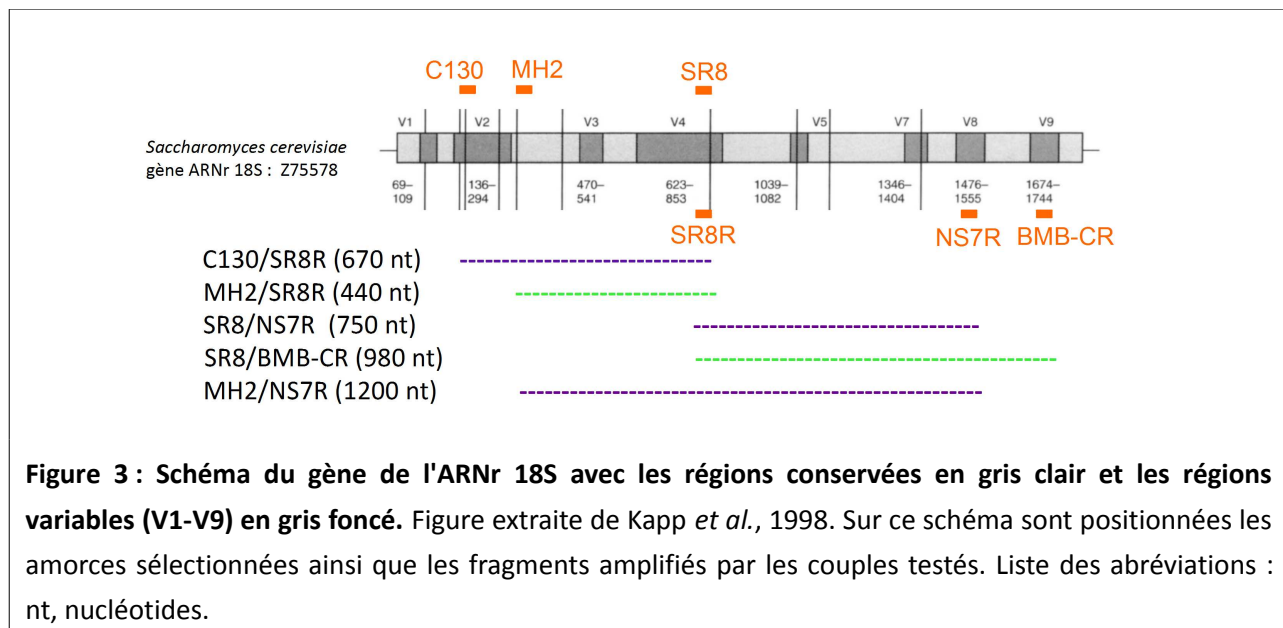
L'étape suivante a consisté à adapter ces amorces pour l'analyse de la diversité des Chytridiomycota. C'est-à-dire que la séquence des amorces a été modifiée pour qu'elle soit identique à la zone ciblée. Elles ont ensuite été testées *in silico* en les alignant localement contre GenBank avec BLASTn. Ainsi les amorces qui comportaient des identités assez fortes avec des séquences de Bacteria, d'Archeae et d'Eukaryota autres que les champignons ont été écartées (Cf. Tableau 1). Des logiciels de design d'amorces ont également été utilisés, tels que Primer-BLAST (NCBI), Beacon Designer™ (Premier Biosoft), et Fast PCR (Primer Digital). Mais aucune des amorces produites n'a été retenue car elles amplifiaient des bactéries. Et l'amorce C130 a été exclusivement dessinée à partir de l'alignement des Chytridiomycota issu de PHYMYCO-DB, en ciblant une zone conservée propre aux Chytridiomycota.

Un des principaux critères étaient d'obtenir des fragments assez courts pour être séquencés à l'aide du pyroséquenceur 454 (Roche). Les règles suivantes ont également été suivies dans la mesure du possible :

- Longueur : entre 17 et 28 nucléotides.
- Composition nucléotidique entre 50 et 60% de G/C.
- Une extrémité 3' plutôt en G ou C ou GC ou CG pour augmenter l'efficacité de l'amorçage.
- Une extrémité 3' qui ne comporte pas 3 G/C ou plus à cause de la stabilité de cette liaison.
- Température d'hybridation entre 55 et 80°C.
- Des extrémités 3' non complémentaires entre les deux amorces du couple.
- Chaque amorce ne doit pas être auto-complémentaire.

Après avoir vérifié la compatibilité des amorces pour les couples constitués (Cf. Figure 3), ceux-ci ont été testés (Cf. Tableau 2) à partir de nouvelles extractions d'ADN sur des échantillons hydrothermaux qui comportent des Chytridiomycota (Le Calvez *et al.* 2009).

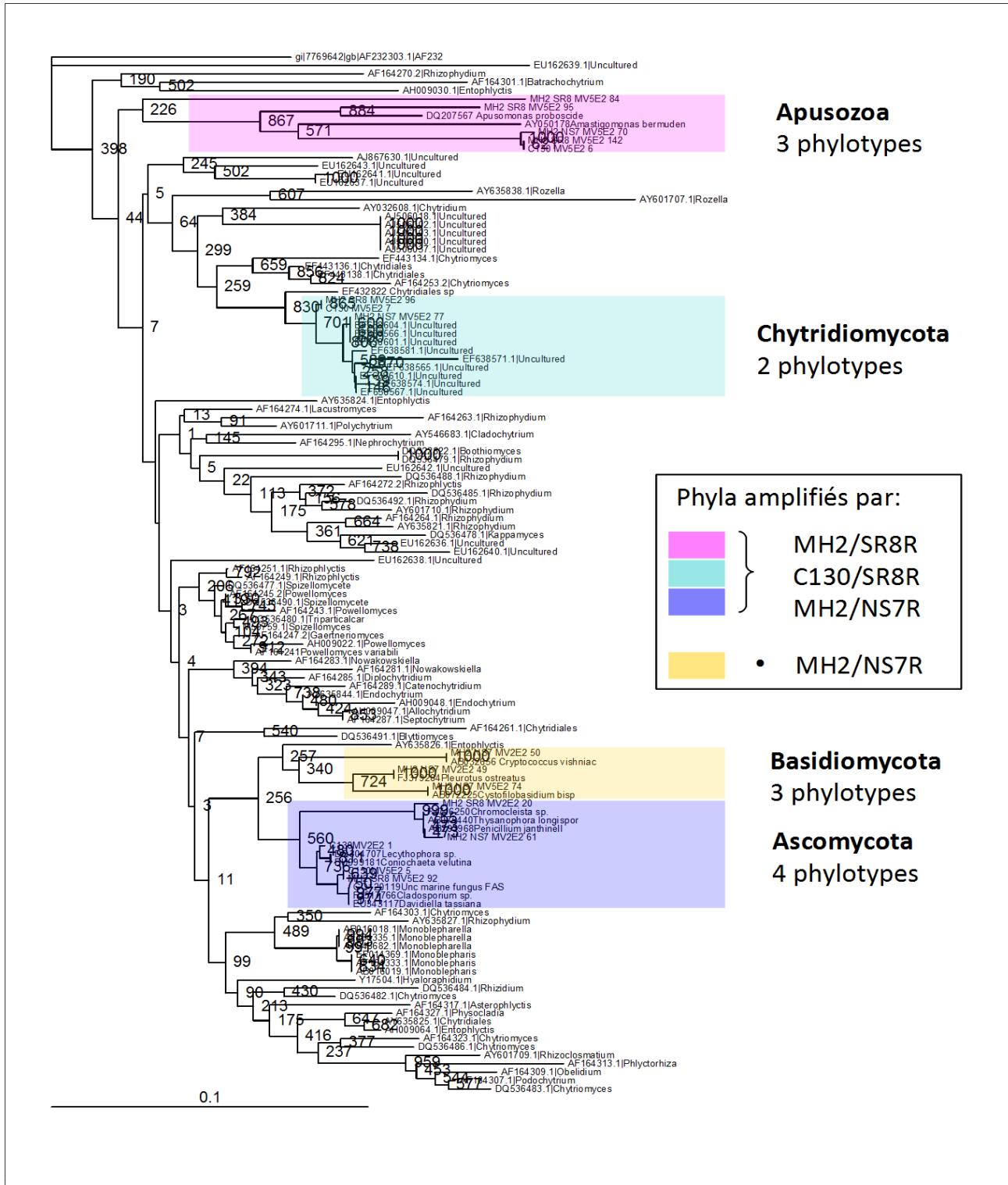
Les résultats (Cf. Figure 4) obtenus avec le couple MH2/SR8R montrent que ce couple a le potentiel pour répondre à la problématique de cette thèse. D'un point de vue méthodologique, la taille du fragment amplifié (environ 440 paires de base) est une taille convenable pour un séquençage de masse par le 454. Ce couple a permis d'amplifier les phylotypes que nous avons obtenus précédemment (Le Calvez *et al.*, 2009) et deux phylotypes d'Ascomycota.



**Tableau 2 : Séquences des amorces testées.**

Amorce	Séquence telle qu'utilisée pour l'amplification	Source
C130	TAC CTT ACT ACT TGG ATA ACC G	Design manuel
MH2	TTC GAT GGT AGG ATA GAG G	Le Calvez <i>et al.</i> 2009
SR8R modifiée	TCA AAG TAA AAG TCC TGG ATC	Vilgalys, page web
NS7R modifiée	ATC ACA GAC CTG TTA TTG CC	Vilgalys, page web
SR8 modifiée	GAT CCA GGA CTT TTA CTT TGA	Vilgalys, page web
BMB-CR	GAC GGG CGG TGT GTA C	Vilgalys, page web

Après vérification avec BLASTn, la proportion d'Ascomycota présentant une forte identité avec SR8R s'est avérée assez faible. De plus, le couple MH2/SR8R permet d'accéder à une diversité externe au clade des Opisthokonta puisque trois phylotypes d'Apusozoa ont été récupérés. Enfin, il couvre une partie de la zone variable V4 qui est une zone informative d'intérêt du point de vue phylogénétique pour des espèces relativement proches et dont les structures secondaires sont assez conservées comme chez les champignons (Wuyts *et al.*, 2000). La zone V4 est aussi plus documentée au niveau de l'échantillonnage taxonomique des bases de données que les zones situées dans l'autre moitié du gène de l'ARNr 18S (Stoeck *et al.*, 2010).



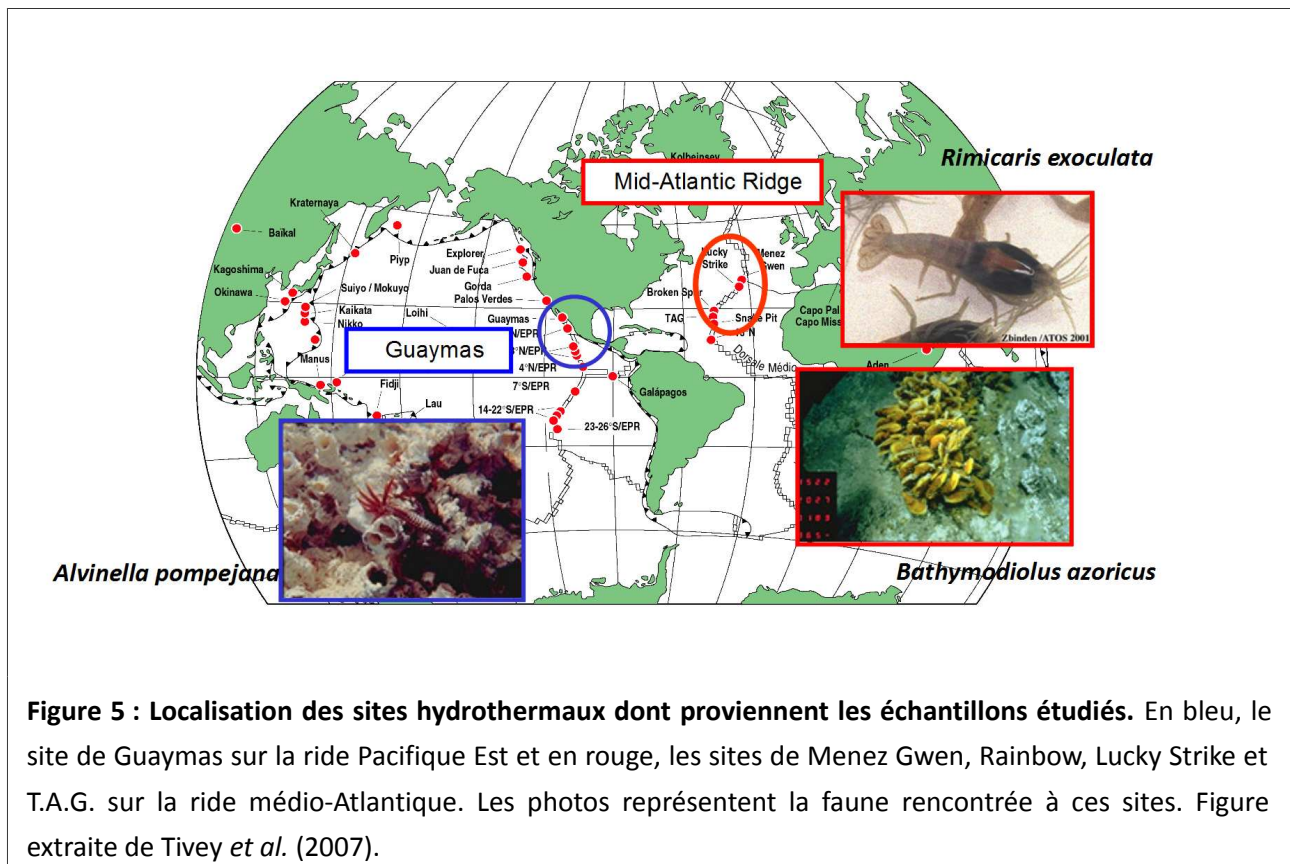
**Figure 4 :** Phylogénie présentant les différentes séquences environnementales obtenues avec les couples d'amorces MH2/SR8R, C130/SR8R et MH2/NS7R. En vert, rose et bleu, sont surlignés les phylotypes retrouvés avec les trois couples d'amorces et en jaune, ceux obtenus avec le couple MH2/NS7R seulement. La topologie a été obtenue par Neighbor-joining (K2P) avec 1000 réplicats de bootstraps.

Afin de cibler une plus grande partie de Chytridiomycota, l'amorce SR8R a été dégénérée en deux versions supplémentaires : SR82R (TCA AAG TAA AAG TCC TGG TTC) et SR83R (TCA AAG TAA CAG TCC TAG ATC). En rouge apparaissent les nucléotides modifiés.

### III – Analyse de la diversité des Chytridiomycota en milieu hydrothermal

#### A – Echantillons

Les échantillons ont été récupérés auprès de Georges Barbier qui les a étudiés avec son équipe notamment par des approches culturelles à l'ESMISAB, Brest, France. Ils ont été prélevés lors de différentes campagnes : GUAYNAUT (1991), MARVEL (1997) et EXOMAR (2005). Depuis, ils ont été conservés en présence d'un cryoprotecteur (DMSO 5% généralement) à  $-80^{\circ}\text{C}$ . Ils proviennent de différents sites hydrothermaux (Cf. Encadré 1) distribués sur la ride Pacifique Est et la ride Médio-Atlantique (Cf. Figure 5).



## Encadré 1 : Le milieu hydrothermal marin profond

Les écosystèmes hydrothermaux sont distribués le long de la ride océanique qui correspond à la zone d'expansion du plancher océanique entre les plaques tectoniques (Langmuir & Forsyth, 2007). Cette ride constitue une chaîne volcanique qui parcourt la Terre de manière continue sur plus de 60 000 km et à des profondeurs variant en moyenne entre 2000 et 5000 m (Ramirez-Llodra *et al.*, 2007). En conséquence, les conditions sont caractérisées par une obscurité totale, une forte pression et des températures avoisinant deux degrés.

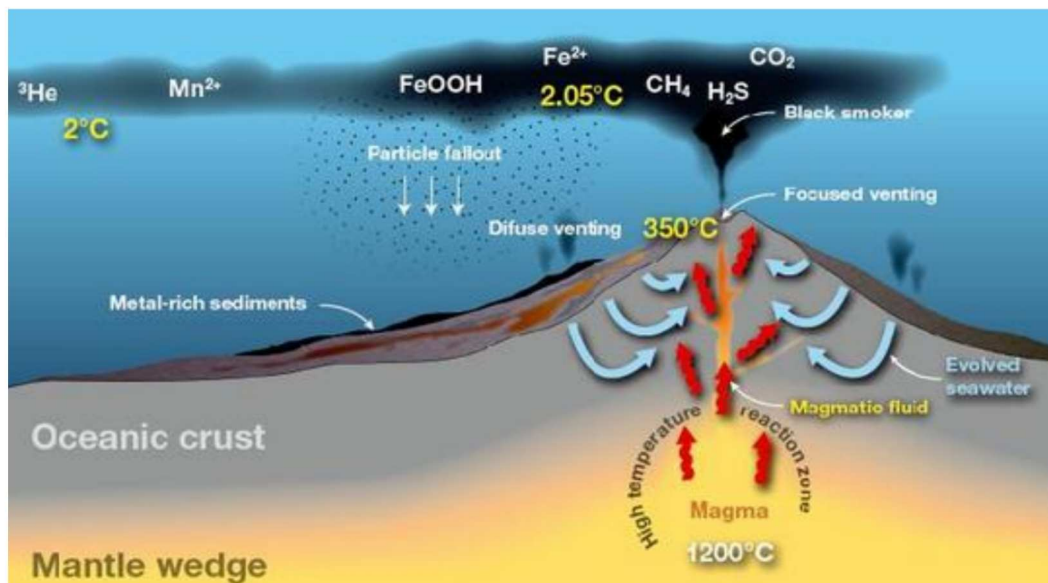
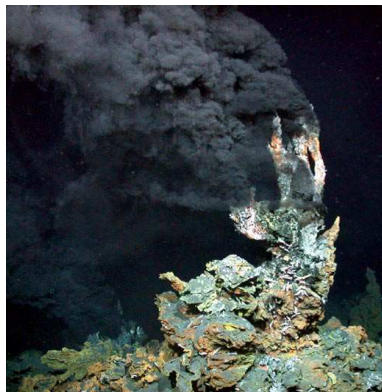


Figure 6 :Schéma d'une section longitudinale d'un volcan marin hydrothermal montrant le brassage de l'eau de mer avec le fluide hydrothermal. D'après <http://www.gns.cri.nz>.

Au niveau de ces dorsales, de l'eau de mer froide et dense s'engouffre par les fissures du plancher océanique (Cf. Figure 6) (Tivey, 2007). En profondeur, elle rentre en contact avec le magma qui remonte : sa température atteint alors plus de 300°C et sa composition chimique est modifiée. Devenue moins dense sous l'action de la chaleur, elle resurgit au niveau du plancher océanique. Le fluide émis est très chaud (*e.g.* 400°C), anoxique, acide (pH 2-3) et contient des métaux lourds et des composés réduits (*i.e.* donneurs d'électrons), *e.g.* méthane, sulfure d'hydrogène (Tivey, 2007). En entrant en contact avec l'eau de mer froide et oxygénée, les métaux du fluide précipitent et se combinent avec les sulfures pour former des sulfures

métalliques qui donnent cette couleur sombre au fluide, d'où le nom de fumeurs noirs (Cf. Figure 7). Ainsi, ce brassage continu avec l'eau de mer riche en accepteurs d'électrons, crée un déséquilibre chimique dynamique dont les microorganismes tirent leur énergie. En effet, ces producteurs primaires utilisent les composés réduits pour fixer le carbone inorganique de l'eau de mer (Ramirez-Llodra *et al.*, 2007).



A



B

**Figure 7 : A : Fumeur noir.**  
Photo extraite de  
Ramirez-Llodra *et al.* (2007).  
**B : Riftia pachyptila.**  
[www.astrobio.net](http://www.astrobio.net)

La communauté biologique repose donc sur la chimiosynthèse bactérienne et les forts gradients physico-chimiques créent un grand nombre d'habitats (Jørgensen & Boetius, 2007). En ayant seulement exploré 10% de la ride, des patrons de diversité ont été identifiés (Ramirez-Llodra *et al.*, 2007). Par exemple, la ride Médio-Atlantique est caractérisée par des communautés de moules *Bathymodiolus azoricus* et de crevettes *Rimicaris exoculata*, alors que la ride Pacific Est est dominée par le ver tubicole géant *Riftia pachyptila*. Egalement, il y a été observé un fort degré d'endémisme des espèces (*i.e.* près de 85%). Concernant la diversité, les bactéries, les archaebactéries et les protistes ont été bien étudiés depuis la découverte dans les années 70 des sources hydrothermales marines profondes. Toutefois, leur étude a d'abord consisté à mettre en place des systèmes de prélèvement stérile pour éviter toute contamination lors de la remontée des échantillons. Des études sont actuellement menées pour la prédiction de leurs fonctions.



Au cours de cette thèse, environ 90 échantillons de natures diverses ont été étudiés. Ils peuvent être regroupés sous deux catégories : les prélèvements minéraux (*e.g.* morceaux de sulfure des cheminées, sédiments de différentes profondeurs) et les animaux (*e.g.* *Riftia*, *Rimicaris*). La liste des échantillons est présentée dans les tableaux 3, 4 et 5.

## B – Génomique environnementale

La diversité des Chytridiomycota a donc été abordée par des méthodes de génomique environnementale, c'est-à-dire que les champignons n'ont pas été mis en culture pour pouvoir les identifier mais que cette diversité a été abordée au travers de l'ADN contenu dans ces échantillons environnementaux.

### 1) Extraction de l'ADN génomique total

L'ADN total a été extrait avec les kits d'extraction DNeasy plant mini kit (Qiagen) d'après la mise au point réalisée précédemment pour l'étude du règne des champignons sur le même type d'échantillons hydrothermaux (Le Calvez *et al.*, 2009). Toutefois, des tests ont été réalisés sur l'efficacité du broyage au mortier-pilon par rapport au vibro-broyage. Les deux méthodes donnant des résultats comparables, d'un point de vue pratique *i.e.* plus rapide et moins sujet aux contaminations, les échantillons ont été extraits au vibro-broyeur. Au final, 100mg ou 120µl ont été prélevés pour chacun des échantillons et placés dans des tubes de 2ml avec une bille de tungstène (3mm de diamètre). Ces tubes ont ensuite été plongés brièvement dans l'azote liquide. Ils ont ensuite été placés dans le vibro-broyeur pour deux sessions de 2mn à 30Hz séparées par une brève immersion dans l'azote liquide. Les échantillons étant de natures très différentes, le rendement des extractions est très variable. En effet, en plus de l'ADN des bactéries et probablement des archaeobactéries et des protistes, l'ADN des animaux était également extrait.



Tableau 3 : Liste des échantillons étudiés, récoltés lors de la campagne GUAYNAUT.

Campagne Océanographique GAYNAUT 1991 (Ride Pacifique Est)		
Site	Echantillon	Nature
Bassin de Guaymas	GY05Rift1	<i>Riftia</i> (Partie supérieure)
	GY05Rift4	<i>Riftia</i> (Branchies)
	GY03A3	Eau filtrée au dessus d' <i>Alvinella</i>
	GY06T6(2)	Carotte de sédiments dans la mate bactérienne (5 à 10 cm)
	GY06T7(2)	Carotte de sédiments dans la mate bactérienne (5 à 10 cm)
	GY06T7(6)	Carotte de sédiments dans la mate bactérienne (25 à 30 cm)
	GY06T7(5)	Carotte de sédiments dans la mate bactérienne (20 à 25 cm)
	GY06T8(3)	Carotte de sédiments dans la mate bactérienne (10 à 15 cm)
	GY06T6(3)	Carotte de sédiments dans la mate bactérienne (10 à 15 cm)
	GY12T2 (2)	Carotte de sédiments (5 à 10 cm)
	GY12T4 (7)	Carotte de sédiments (30 à 35 cm)
	GY12T4(5)	Carotte de sédiments (20 à 25 cm)
	GY12T1(6)	Carotte de sédiments (20 à 25 cm)
	GY12T3(4)	Carotte de sédiments (15 à 20 cm)
	GY12T1 (4)	Carotte de sédiments (15 à 20 cm)
	GY12T2 (3)	Carotte de sédiments (10 à 15 cm)
	GY12T3(3)	Carotte de sédiments (10 à 15 cm)
	GY12T4(3)	Carotte de sédiments (10 à 15 cm)
	GY12T4(1)	Carotte de sédiments (0 à 5 cm)
	GY12T3(1)	Carotte de sédiments (0 à 5 cm)

Tableau 4 : Liste des échantillons étudiés, récoltés lors de la campagne MARVEL.

Campagne Océanographique MARVEL 1997 (Ride Médio-Atlantique)		
Site	Echantillon	Nature
Lucky Strike	MV5-E1	Sulfure friable de la roche portant des moules
	MV5-E2	Grattage de la face externe d'une coquille de moule
Menez Gwen	MV2-E2	Broyat du corps de la moule

Tableau 5 : Liste des échantillons étudiés, récoltés lors de la campagne EXOMAR.

Campagne Océanographique EXOMAR 2005 1991 (Ride Médio-Atlantique)		
Site	Echantillon	Nature
Rainbow	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>
	EX9E11	<i>Bathymodiolus azoricus</i> (Byssus)
	EX9E12	<i>Bathymodiolus azoricus</i> (Byssus)
	EX9E13	<i>Bathymodiolus azoricus</i> (Byssus)
T.A.G. Trans-Atlantic Geotraverse	EX11E1(1)	<i>Rimicaris exoculata</i>
	EX11E1(2)	<i>Rimicaris exoculata</i>
	EX11E2(1)	<i>Rimicaris exoculata</i>
	EX11E2(2)	<i>Rimicaris exoculata</i>
	EX11E3	<i>Rimicaris exoculata</i>
	EX11E4	<i>Rimicaris exoculata</i> (Thorax)
	EX11E5	<i>Rimicaris exoculata</i>
	EX11E6	<i>Rimicaris exoculata</i>
	EX11E7	<i>Rimicaris exoculata</i> (Mue)
	EX11E8	<i>Rimicaris exoculata</i> (Mue)
	EX11E9	<i>Rimicaris exoculata</i>
	EX11E10	<i>Chorocaris chacei</i>
	EX11E11	<i>Rimicaris exoculata</i>
	EX13E2	Sédiment ocre
	EX13E3	<i>Rimicaris exoculata</i>
	EX13E4	<i>Rimicaris exoculata</i>
	EX13E5	<i>Chorocaris chacei</i>
	EX13E6	<i>Rimicaris exoculata</i>
	EX13E7	<i>Chorocaris chacei</i>
	EX14E2	<i>Rimicaris exoculata</i>
EX14E3	<i>Rimicaris exoculata</i>	
EX15E1	Morceaux de sulfures de Kremlin	
EX15E2	<i>Rimicaris exoculata</i>	
EX15E3	<i>Rimicaris exoculata</i>	
Lost City	EX17E1	Morceaux de carbonates porteurs de bactéries filamenteuses
	EX18E1	Morceaux d'éponges siliceuses
	EX18E4	Sédiment
	EX18E5	Sédiment
	EX19E1	Fragments de carbonates avec filaments bactériens
Lucky Strike	EX20E1	Byssus et grattage de moules sans bactéries filamenteuses

## 2) Amplifications du gène de l'ARNr 18S avec MH2/SR8R

La préparation la plus directe des échantillons pour un séquençage de masse par le 454 nécessite le design d'amorces fusionnées. En effet, ce type de séquençage nécessite l'adjonction à nos amorces d'un adaptateur, d'une clé et d'un MID (*i.e.* une dizaine de nucléotides qui permettent d'identifier l'échantillon). L'adaptateur consiste en une séquence nucléotidique qui va permettre à l'amplicon de se lier aux microbilles qui portent la séquence complémentaire (*Cf.* Figure 6). En conséquence, selon le type de pyroséquençage envisagé il n'existe qu'un couple d'adaptateur.

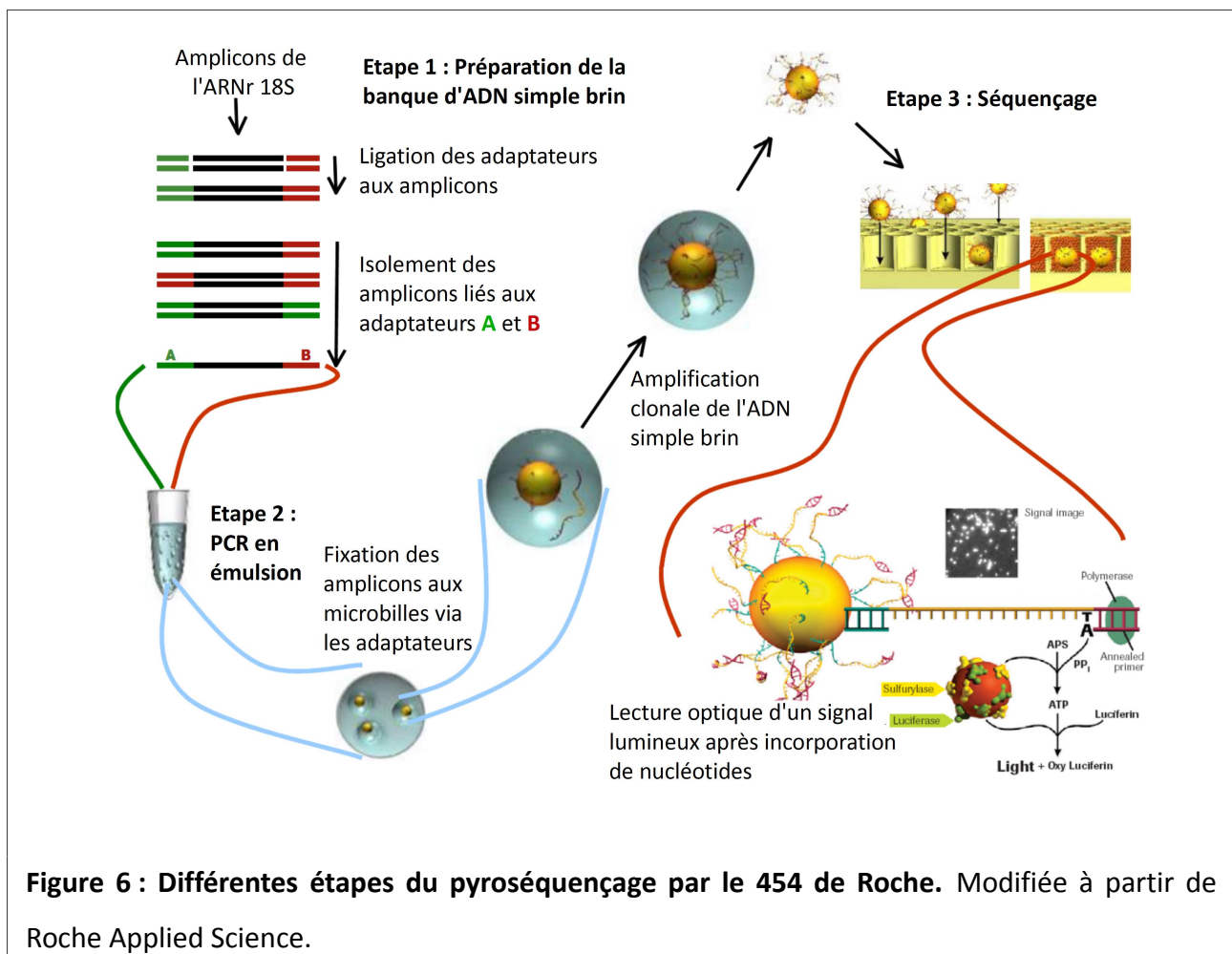


Figure 6 : Différentes étapes du pyroséquençage par le 454 de Roche. Modifiée à partir de Roche Applied Science.

Des prédictions de formation de structures secondaires ont donc été effectuée avec Mfold (Zuker, 2003) sur les différentes combinaisons possibles entre MH2 ou SR8R et l'adaptateur

A/clé/MID ou l'adaptateur B/clé/MID. La stabilité de la structure secondaire dépend des liaisons formées quand la molécule se replie sur elle-même. Pour cela, un  $\Delta G$  est calculé. Il s'agit de l'énergie libre qui correspond à l'énergie requise pour casser la structure secondaire (Zuker, 2003). Plus elle est négative, plus la structure secondaire est stable, ce qui risque d'empêcher l'hybridation de l'amorce à l'ADN génomique. Quelque soit l'adaptateur, l'amorce MH2 forme des structures secondaires très stables *i.e.* les valeurs indiquant la plus faible stabilité étaient  $\Delta G = -19$  Kcal/mol et  $\Delta G = -9$  Kcal/mol pour l'adaptateur A et B respectivement. Deux couples d'amorces fusionnées testés n'ont pas permis d'obtenir le fragment attendu. Sachant que les valeurs généralement tolérées ne descendent pas au-dessous de  $-6$  Kcal/mol, cette méthode de préparation des échantillons pour le pyroséquençage n'est donc pas envisageable avec MH2/SR8R. L'autre solution consiste à accrocher les adaptateurs/clé/MID par ligation après l'amplification des échantillons (Cf. Figure 6).

Ainsi dans un premier temps les échantillons ont été criblés pour savoir si ils comportent le fragment d'intérêt. Une mise au point de la réaction d'amplification a été réalisée pour optimiser le rendement d'amplicons produits. En effet, lors des tests des différents couples d'amorces, les séquences ont été obtenues par du clonage/séquençage classique. La quantité minimale d'ADN requise pour ce protocole n'est pas très élevée. Par contre, pour le séquençage de masse envisagé, la quantité minimale d'ADN requise pour la constitution de la banque d'ADN simple brin (Cf. Figure 6) est de 500 ng.

L'amplification du gène de l'ARN 18S par PCR (*i.e.* Polymerase Chain Reaction) avec les amorces MH2/SR8R-SR82R-SR83R selon un protocole classique des cycles de dénaturation, hybridation et élongation n'a pas permis d'obtenir des résultats satisfaisants. En effet, très peu d'amplifications ont été obtenues et lorsqu'il y avait amplification, elle était très faible et sa répétabilité n'était pas constante. Une rampe de températures a donc été effectuée pour ajuster la température d'hybridation de l'amorce. Une fois déterminée et appliquée aux cycles d'amplification classiques, le rendement n'était pas amélioré. Pour y remédier, des Touchdown PCR ont été réalisées. Ce type de protocole est utilisé, entre autre, lorsque le fragment cible contenu dans l'ADN génomique total extrait est très peu représenté. Cela consiste à appliquer une

température d'hybridation proche de la température de demi-dénaturation ( $T_m$ ) propre à l'amorce, durant les premiers cycles, pour augmenter la stringence et donc augmenter la spécificité de l'amorce. Cela a pour conséquence d'augmenter le nombre de copies du fragment attendu. Ensuite, la température d'hybridation est abaissée pour les cycles suivants afin d'augmenter l'efficacité de la PCR. Mais cela n'a pas permis d'obtenir un meilleur rendement.

Par la suite, plusieurs modifications ont été testées une à une : (i) une dilution des ADN extraits, (ii) des concentrations d'amorces différentes, (iii) une augmentation du nombre de cycles. Aucune de ces nouvelles tentatives n'a montré une amélioration du rendement de l'amplification. L'augmentation du nombre de cycles par exemple a dû être limitée à 38 cycles puisque au-delà de ce nombre apparaissaient des fragments de taille non souhaitée. Pour tester la présence d'inhibiteurs, des PCR ciblant le gène de l'ARNr 16S des bactéries ont été réalisées. Des amplifications de bonne intensité ayant été obtenues, il a été conclu que les ADN extraits ne comportaient pas d'inhibiteurs susceptibles de gêner la réaction PCR. Enfin, des amplifications avec différents Pre-Mix Failsafe™ (Epicentre Biotechnologies) ont permis de sélectionner un des Pre-Mix comme potentiel facilitateur de la PCR. Toutefois, cela a permis d'obtenir seulement un faible nombre d'amplifications supplémentaires.

Au total, sept échantillons présentant des fragments de la taille attendue ont été obtenus en double et deux échantillons n'ont été obtenus qu'une seule fois. Par contre, lors de la visualisation aux rayons ultra-violet de ces fragments sur gel d'agarose après un traitement au BET (Bromure d'éthydiu), des intensités différentes étaient perçues pour chacun d'eux, indiquant des rendements d'amplification différentiels. D'autres méthodes d'évaluation de la quantité d'ADN ont alors été appliquées, telles que les puces à ADN haute sensibilité (Agilent) et un dosage de l'ADN par picogreen. Globalement les résultats ont indiqué que tous les amplicons étaient à une concentration inférieure à  $4\text{ng}/\mu\text{l}$ . En prenant en compte les difficultés à obtenir une amplification, même très faible, le regroupement des amplifications pour constituer une banque d'amplicons pour le pyroséquençage ne permet pas d'obtenir une quantité d'ADN suffisante. Il a donc été décidé de réaliser un clonage/séquençage à partir des échantillons obtenus (*Cf.* Tableau 6).

Tableau 6 : Tableau récapitulatif des échantillons.

Echantillon	Nature	Localisation	Profondeur (m)	Dilution	Mix PCR
GY12T4(5)	Carotte de sédiment (20-25cm)	Guaymas	2000	1	Classique
				1/10	Failsafe
MV5E2	Grattage de la face externe d'une coquille de <i>B. azoricus</i>	Lucky Strike	1700	1/10	Classique
				1/10	Classique
EX6E5	Crevette : <i>Chorocaris chacei</i>	Rainbow	2300	1	Failsafe
EX8E6	Crevette : <i>Rimicaris exoculata</i>	Rainbow	2300	1	Classique
				1/10	Failsafe
EX9E1	Fragment de cheminée (sulfure)	Rainbow	2300	1/10	Failsafe
EX9E5	Modiole : <i>Bathymodiolus azoricus</i>	Rainbow	2300	1	Classique
				1/10	Failsafe
EX9E8	Modiole : <i>Bathymodiolus azoricus</i>	Rainbow	2300	1	Classique
				1	Classique
EX9E9	Modiole : <i>Bathymodiolus azoricus</i>	Rainbow	2300	1	Classique
				1	Classique
EX15E3	Crevette : <i>Rimicaris exoculata</i>	T.A.G	3645	1	Classique
				1/10	Failsafe

### 3) Clonage des amplicons

Les amplicons ont tout d'abord été purifiés avec des billes magnétiques (Agencourt AMPure) selon le protocole du fournisseur. Puis, ils ont été clonés avec le vecteur pGEM-T (Promega) et les bactéries compétentes DH5 $\alpha$  (Gibco BRL Life Technologies). Enfin, les clones positifs ont été envoyés à séquencer (Genoscreen, Lille, France).

## IV – Analyses Phylogénétiques

Les méthodes phylogénétiques reposent sur la comparaison de caractères spécifiques. Dans

cette étude, le choix a porté sur les phylogénies moléculaires qui permettent de comparer des séquences d'ADN. Un marqueur phylogénétique doit présenter des caractéristiques lui assurant des propriétés informatives sur les relations de parenté entre les taxons, il doit être (i) présent chez tous les taxons étudiés, (ii) orthologue *i.e.* les similitudes observées sont héritées d'un ancêtre commun, (iii) conservé *i.e.* il doit présenter des régions comparables entre les taxons.

## A – Alignement

L'alignement des séquences est crucial puisque c'est à partir de la matrice alignée que sont réalisées les reconstructions phylogénétiques. Pour cela, Clustal X 2.1 (Larkin *et al.*, 2007) a été utilisé pour effectuer un alignement multiple globale, qui a ensuite été optimisé manuellement.

## B – Méthodes de reconstructions phylogénétiques

### 1) Neighbor Joining

Le Neighbor Joining (NJ ; Saitou et Nei, 1987) est un algorithme d'agglomération basé sur les distances entre les séquences d'ADN (*i.e.* une analyse phénétique), qui autorise un taux de mutations différent pour chacune des branches pour relaxer l'hypothèse d'horloge moléculaire. Le modèle de Kimura à deux paramètres (K2P) a été utilisé pour calculer les distances entre les paires de terminaux. C'est un modèle qui pondère les substitutions selon qu'il s'agisse de transitions ou de transversions. Cette étape a été réalisée avec Clustal X 2.1 (Larkin *et al.*, 2007).

Cette méthode est limitée car elle ne fait pas de distinction entre les homologies vraies et l'homoplasie et donc elle n'est pas adaptée à des séquences éloignées. C'est pourquoi, des méthodes basées sur les caractères ont également été appliquées.

## 2) Méthode de maximum de vraisemblance

La méthode de maximum de vraisemblance évalue les arbres possibles, donc différentes hypothèses évolutives, en terme de probabilité pour un modèle de distance évolutive donné. L'arbre obtenu est donc celui qui a la meilleure probabilité d'expliquer les données. Le modèle décrivant au mieux les données a été choisi avec le logiciel jModelTest (Posada & Crandall 1998). Il s'agit du modèle GTR (General Time Reversible), qui consiste à pondérer les différents types de mutations selon leur fréquence au sein de la matrice, pour calculer les distances entre les paires de taxons. Ce modèle est souvent utilisé pour les méthodes de maximum de vraisemblance car il est saturé en terme de nombre de paramètres. On peut également corriger la vitesse d'évolution des sites par la distribution gamma (paramètre G) et certains sites peuvent être considérés comme invariables (paramètre I).

La reconstruction phylogénétique par le méthode de maximum de vraisemblance a été calculée avec le logiciel PHYML (Guindon *et al.* 2010) selon le modèle GTR+I+G.

## 3) Robustesse de l'arbre inféré

Les arbres obtenus par ces deux méthodes doivent être analysés pour mesurer la robustesse de l'hypothèse évolutive inférée. Dans cette étude, la pertinence des nœuds internes a été déterminée par une méthode de rééchantillonnage : le bootstrap. Le jeu de données initial est rééchantillonné pour constituer un jeu de données différent pour chaque réplicat de bootstrap. La valeur de bootstrap correspond au pourcentage d'arbres créés où le nœud existe. Cela permet de quantifier l'incertitude de chacun des nœuds de l'arbre inféré.

## 4) Méthode Bayésienne

La méthode Bayésienne est basée sur la distribution des probabilités postérieures (*i.e.* calculées *a posteriori*) des arbres phylogénétiques obtenus suivant un modèle choisi, à partir de probabilités définies *a priori*. C'est à dire que la probabilité du modèle est calculée en fonction des données, à l'inverse de la méthode de maximum de vraisemblance. Pour cette étude, le modèle



utilisé est le même que celui utilisé en maximum de vraisemblance, à savoir GTR+I+G. Les calculs ont été effectués avec Mr Bayes 3.2.1 (Ronquist & Huelsenbeck 2003).

Une fois ces trois arbres phylogénétiques obtenus, les hypothèses évolutives sont comparées pour examiner leur congruence. C'est-à-dire que l'existence des nœuds déterminés par les trois méthodes est vérifiée tout en considérant leur robustesse.

# CHAPITRE 1

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



## Chapitre 1

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

Le premier objectif de cette thèse est de décrire la diversité des lignées basales des champignons en milieu hydrothermal marin profond par une approche de génomique environnementale et des analyses phylogénétiques. Le choix du marqueur moléculaire a porté sur le gène de l'ARNr 18S pour ses qualités informatives sur les relations évolutives entre les organismes et sa bonne représentativité dans les bases de données moléculaires publiques. Cependant, la fiabilité des séquences stockées dans ces bases de données est très discutée par la communauté scientifique (*Cf.* article de ce chapitre). Nous avons donc créé une base de données pour les marqueurs phylogénétiques (ARNr 18S et EF1- $\alpha$ ) fongiques, appelés PHYMYCO-DB. Elle a été réalisée en collaboration avec une équipe de l'IRISA (Institut de recherche en informatique et systèmes aléatoires, Rennes).

L'étape majeure consiste à trier manuellement les séquences à partir d'alignements multiples. Cette étape a été réalisée avec Marie Duhamel. Près de 20% des séquences récupérées après le filtrage automatique ont été écartées par ce tri manuel. Au total, 8757 séquences du gène de l'ARNr et 648 séquences du gène de EF1- $\alpha$  de très bonne qualité sont maintenant disponibles.

PHYMYCO-DB offre les outils nécessaires pour: (i) extraire des séquences de très bonne qualité pour chacun des cinq phylums et pour tous les niveaux taxonomiques, (ii) extraire des alignements considérés comme alignements de référence, (iii) lancer des alignements de séquences personnelles ajoutées aux séquences extraites de PHYMYCO-DB. Elle est accessible via l'URL <http://phymycodb.genouest.org/>.

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

Stéphane Mahé<sup>1,2</sup>, Marie Duhamel<sup>1,2,3</sup>, Thomas Le Calvez<sup>1,2,5</sup>, Laëtitia Guillot<sup>2,4</sup>, Ludmila Sarbu<sup>2,4</sup>, Anthony Bretaudeau<sup>2,4</sup>, Olivier Collin<sup>2,4</sup>, Alexis Dufresne<sup>1,2</sup>, E. Toby Kiers<sup>3</sup> and Philippe Vandenkoornhuyse<sup>1,2\*</sup>

<sup>1</sup> Université de Rennes I, CNRS, UMR 6553 ECOBIO, Campus de Beaulieu, 35042 Rennes cedex, France

<sup>2</sup> Université Européenne de Bretagne, UEB (European University of Brittany), France

<sup>3</sup> Department of Ecological Science, Vrije Universiteit, Amsterdam, 1081 HV Amsterdam, the Netherlands

<sup>4</sup> Université de Rennes I, CNRS, UMR 6074 IRISA, Campus de Beaulieu, 35042 Rennes cedex, France

<sup>5</sup> Université de Poitiers, CNRS, UMR6008, 40 avenue du Recteur Pineau, 86022 Poitiers Cedex, France

**\*Corresponding author:** E-mail: philippe.vandenkoornhuyse@univ-rennes1.fr

Short title: A fungal database for phylogenetic studies

## **Abstract**

### *Background*

In environmental genomics studies, fungi can be identified based on nucleic acid sequences using either highly variable sequences as species barcodes or conserved sequences containing a high quality phylogenetic signal. For the latter, identification at the species level or higher taxonomic levels relies on phylogenetic analyses and the adoption of the phylogenetic species concept.

Such analysis requires that sequences are well identified and deposited in public-access databases. However, many errors exist within these publicly available sequences and these data require screening to ensure correct phylogenetic interpretation.

### *Methods & principal findings*

To facilitate phylogenetic inferences and phylogenetic assignment, we introduce a fungal sequence database. The database PHYMYCO-DB comprises fungal sequences from GenBank that have been filtered to satisfy stringent sequence quality criteria. For the first release, two widely used molecular taxonomic markers were chosen: the SSU rRNA and EF1- $\alpha$  gene sequences. Following the automatic extraction and filtration, a manual curation is performed to remove problematic sequences while preserving relevant sequences usable for phylogenetic studies. As a result of curation, ~20% of the automatically filtered sequences have been removed from the database. To demonstrate how PHYMYCO-DB can be employed, we test a set of environmental Chytridiomycota sequences obtained from deep sea samples. This exercise reiterates the importance of phylogenetic assignment of sequences rather than a simple BLASTn search.

### *Conclusion*

PHYMYCO-DB offers the tools necessary to: (i) extract high quality fungal sequences for each of the 5 fungal phyla, at all taxonomic levels, (ii) extract already performed alignments, to act as 'reference alignments', (iii) launch alignments of personal sequences along with stored data. A total of 8757 SSU rRNA and 648 EF1- $\alpha$  gene high quality fungal sequences are now available.

The PHYMYCO-DB is accessible through the URL <http://phymycodb.genouest.org/>.

**Key words:** Fungal diversity, SSU rRNA, EF1- $\alpha$ , sequence database, Chytridiomycota, environmental sequences, curation.

### **Introduction**

In recent years there has been an exponential increase in the number of gene sequences available in public-access databases. This is the result of new DNA extraction techniques and new generation sequencers that allow the collection of data at great speed. In recent years, the use of molecular taxonomic markers associated with phylogenetic analyses has revealed the considerable extent of genetic diversity in fungi, especially those that are cryptic, unculturable or not easily distinguishable by morphological characters [e.g. 1]. As the species concept is employed for diversity measurements, systematics and evolutionary analyses [2], an efficient means of identifying boundaries, and thus number of species, is required. Molecular methods and the implicit adoption of the phylogenetic species concept [3] offer a standardized approach to delimit groups of organisms [e.g. 4-6]. Thanks to progress in sequencing technologies and bioinformatic methods, the detection of orthologous sequences using databases is relatively efficient. This approach can also be successfully applied to organisms that are not available in culture, allowing us to prospect their diversity in various fungal habitats [7,8]. Of course, this approach requires choosing a relevant



molecular marker which: (i) targets a nucleic acid sequence with a limited proportion of homoplasy (i.e. non-inherited similarities), (ii) contains enough phylogenetic information which is not sensitive to paralogy (i.e. single copy genes or highly conserved genes). This allows for accurate characterization at species or higher taxonomic levels.

In this context, the nuclear gene coding for the small subunit of the ribosomal RNA (SSU rRNA) is often seen as the 'ultimate' molecular marker [9, for review 10]. 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' phylogenies [11]. Although the SSU rRNA gene can have a multicopy status within a single fungal genome, sequence variations have been shown to be extremely low or null. For example *Saccharomyces cerevisiae* has 2 SSU rRNA copies both on its chromosome XII; *Encephalitozoon cuniculi*, a Microsporidia, has 2 SSU rRNA genes copies one on its chromosome I, the other on chromosome IV. In these 2 cases, the copies display 100% identity. Thus the SSU rRNA gene is lowly sensitive to paralogy compared to LSU rRNA gene and ITS where variations among copies have been clearly shown [e.g. 12-14].

A second advantage of using the SSU rRNA gene sequence is its huge representation in international public databases - GenBank [15], EMBL/ENA [16], DDBJ [17] - which allows comparisons between a wide variety of organisms [for review 18]. One disadvantage is that because the SSU rRNA gene is highly conserved, the resolution of the phylogenetic analyses is poor for youngest fungal groups within Ascomycota. Other genes, such as

those encoding 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. Among these ones, EF1- $\alpha$  sequence data are the most abundant but only represent a small fraction of the amount of SSU rRNA yet available (i.e. less than 7% of the total number of sequences contained in PHYMYCO-DB). Generally present as a single copy gene, the EF1- $\alpha$  gene is involved in protein synthesis and displays a higher mutation rate than SSU rRNA gene. Because of these attributes, EF1- $\alpha$  protein sequences have been used to resolve phylogenetic affinities between eukaryotic organisms [19-21], and particularly the sisterhood of animals and fungi [22]. The gene sequences also have the potential to help resolve phylogenetic relationships between closely related fungi [21,23-24], but they contain a higher proportion of homoplasious positions compared to SSU rRNA gene sequences. Synergistic studies of SSU rRNA genes and EF1- $\alpha$  genes could greatly improve the resolution of fungal phylogenetic affinities. An online database incorporating data from both these sequences is a key step to achieving improved phylogenetic resolution for fungi.

#### **Pollution of public sequence database and the aim of PHYMYCO-DB**

One major obstacle for international public databases is constant pollution by non-negligible proportion of mistake-containing sequences (GenBank/EMBL/DDBJ). This problem, discussed in several articles and journal forums [e.g. 25-32], is becoming more and more obvious, but solutions remain elusive. Mistakes have different origins, including: (i) erroneous specimen identification [27], (ii) the use of separate names for different sexual stages [30], (iii) differences in taxonomy among specialists [27] and/or advances in knowledge since the time the sequence was deposited leading to wrong designations [29], (iv) the lack of precision in the description of the deposited sequences making them obsolete [33], (v) sequences resulting from artefactual origin (i.e. chimeric sequences), and (vi) sequences of poor quality with undefined positions. Even more

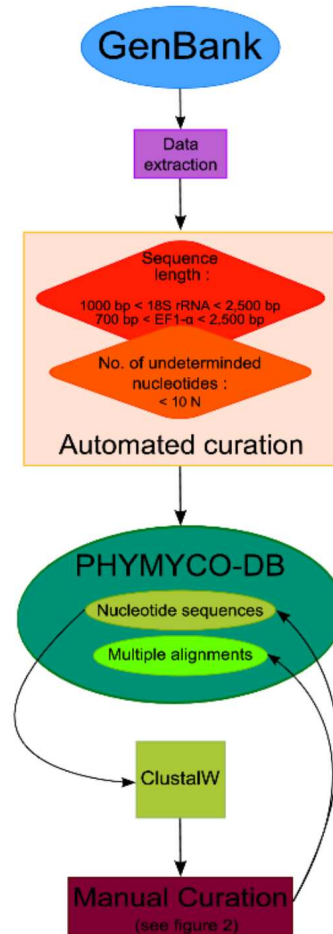
problematic is the erroneous annotated sequences that propagate within open access databases because of biased phylogenetic interpretation. Unfortunately, more and more sequence assignments are based solely on identity searches using heuristic local alignment (i.e. BLASTn searches). All these mistakes have the potential to jeopardize interpretations. Therefore, assessing the reliability of sequences is an increasingly important prerequisite to analyses.

Many of these errors can be limited via expert curation. Expert curation is critical for the continued advancement of the field because it allows for the production of sequence databases, containing accurate and reliable sequences. To date, most curated databases specialize in particular taxonomic groups [e.g. 34], collect data associated to each nucleic acid sequence, and work with specimens validated by experts and deposited in public reference collections [e.g. 33]. Several important tools, such as the Ribosomal Database Project [35], SILVA [36], Greengenes database [37] exist online for the analysis of SSU rRNA gene sequences. Apart from SILVA, these databases use automated filters to remove part of the polluting sequences. However, manual curation is an essential component of these projects and should aim to be even more stringent.

Based on lessons learned from other curated databases, our aims at PHYMCO-DB are to: (i) develop an easy-to-use fungal-dedicated database with stored sequences of high quality, (ii) use selected molecular markers that are widely acknowledged, namely SSU rRNA and EF1- $\alpha$ , (iii) produce a tool, based on anchor sequences covering the fungal tree, that can be automatically updated, along with an expert curation of the new sequences, (iv) produce high quality multiple alignments for use in testing environmental sequences or evolutionary hypotheses.

### **Database structure : design and implementation**

The sequences constituting PHYMYCO-DB version 1 (Fig. 1) were retrieved in October 2011 from the release 185 of GenBank (NCBI). The SSU rRNA and EF1- $\alpha$  genes sequences are extracted from the GenBank database, using the following queries: “fungi[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|EF1a)”. After this extraction step, automatic quality filter parameters are applied. For SSU rRNA, nucleic acid sequences that are shorter than 1000 nucleotides and longer than 2500 nucleotides are rejected. Likewise for EF1- $\alpha$  genes, sequences shorter than 700 nucleotides and longer than 2500 nucleotides are discarded. Also sequences containing more than 10 consecutive undetermined nucleotides are excluded. According to the automatic quality criteria, all accepted sequences are then stored in a MySQL 5 relational database. PHYMYCO-DB is automatically updated 4 times a year and is managed by administrators using the web interfaces developed with PHP version 4 programming language.



**Figure 1** : Flowchart of the data in the PHYMYCO-DB. The arrows indicate the flow of gene sequences extracted from the GenBank database, through the automated and manual curation steps. All the sequences made available to users has passed the 2 curation processes. After each upgrade of the database (i.e. 4 times per year), expert manual curation is performed.

Following automatic filtering, datasets are then cross-checked by expert curators (hereafter 'expert curation'). Multiple alignments are performed using Clustal X 2.1 [38] on small sequence groups (< 400 sequences), which are closely related to obtain a good quality alignment and to make the expert curation as accurate as possible. Sequences are deleted from the alignment and from the database in a manual cleaning process if they contain: errors of sequencing (i.e. containing several substitutions that are not found anywhere else, Fig. 2), errors in the annotation (i.e. a sequence with a naming inside a different group, Fig. 2), homopolymers insertions (Fig 2), many undetermined nucleotides (Fig 2), erroneous alignment or reverse complementary sequences (Fig. 2). This expert curation is time consuming but essential to obtain reliable sequences and high-quality alignments. By adopting strict rules of expert curation, subjectivity and mistakes become minimal. Following expert curation, species redundancy (i.e. identical sequences) is retained in the database to keep information about sequence origin and ecology. The detection of dubious sequences from the alignments does not result in correction of the sequence in international databases. They are, however, all removed from PHYMYCO-DB. When corrections are made for a given sequence, a new registration number is provided by GenBank for example. In this case, the corrected sequence will be automatically extracted (i.e. 4 updates per year) and will be examined by one of the expert curators.

During the development process, it became clear that our automatic filters were not stringent enough to retrieve only trustworthy sequences. For example, SSU rRNA can present intron-like regions which could also be chimeric insertions. Introns are abundant in particular lineages of fungi, especially within lichen-forming fungi (Ascomycota). These fungi can display up to eight introns in the SSU rRNA gene, as for example found in the taxon *Physconia* [39]. At the expert curation stage, we noticed that the position of introns was not consistently given in the deposited sequence description, yet they were

detectable after the alignment only. When a sequence containing introns was the only sequence of a particular genus, this sequence was kept and the intron insertion was removed from the alignment. Otherwise the sequence was discarded from the database. Employing our curation principles, we discarded 2090 additional unreliable sequences, i.e. 18% of the sequences extracted from GenBank.

Following the curation steps, 8757 SSU rRNA gene sequences have been stored in the PHYMYCO-DB (4858 Ascomycota, 2056 Basidiomycota, 366 Chytridiomycota, 1025 Glomeromycota, and 452 Zygomycota). PHYMYCO-DB also contains 648 EF1- $\alpha$  gene sequences (291 Ascomycota, 173 Basidiomycota, 10 Chytridiomycota, 25 Glomeromycota, and 149 Zygomycota). Because of the heterogeneity among the number of sequences per taxonomic rank, and because we wanted a limited number of sequences for each alignment, the taxonomic level within these alignments is variable (family to phylum level). We therefore produced a total of about 50 'reference' alignment files. These online alignments contain mainly full length sequences, even if rare, very long sequences were cut at the same length as the others. This was done to keep maximum information available. This is especially useful for designing primers, and to give a greater freedom for manipulation by online users.

**Figure 2** : Visualisation of sequences deleted by the manual curation after alignment (ClustalX 2.1). The sequences highlighted in blue illustrate examples of sequences removed from PHYMYCO-DB. Irrelevance can come from erroneous sequencing (e.g. repeated gaps), wrong annotation (e.g. sequence corresponding to another clade), high numbers of undetermined nucleotides, presence of long insertions and introns, presence of deletions and sequence inversion.

```

Cladosporium TGTERTTGGARTGAGTACRATTTTAAATCCCTTACGAGGGRACARTTGGAGGGCRACTCTGGTGCCAGCAG
Cladosporium TGTERTTGGARTGAGTACRATTTTAAATCCCTTACGAGGGRACARTTGGAGGGCRACTCTGGTGCCAGCAG
Piedraia TGTERTTGGARTGAGTACRATTTTAAATCCCTTACGAGGGRACARTTGGAGGGCRACTCTGGTGCCAGCAG
Capnodium GTT--TTGGARTGAGTACRATTTTAAATCCCTTACGAGGGRACARTTGGAGGGCRACTCTGGTGCCAGCAG
Capnodium TGTERTTGGARTGAGTACRATTTTAAATCCCTTACGAGGGRACARTTGGAGGGCRACTCTGGTGCCAGCAG
Microxyphium TGTERTTGGARTGAGTACRATTTTAAATCCCTTACGAGGGRACARTTGGAGGGCRACTCTGGTGCCAGCAG
    
```

(a) Errors of sequencings

```

Pilophorus CTCCGGGGCTCCTTGGTGATTCACTACTCAGCGATCGCATGGCCCTTGGCCGGCGATGGTTCAATTC
Pilophorus CTCCGGGGCTCCTTGGTGATTCACTACTCAGCGATCGCATGGCCCTTGGCCGGCGATGGTTCAATTC
Cladonia CTCCGGGGCTCCTTGGTGATTCACTACTCAGCGATCGCATGGCCCTTGGCCGGCGATGGTTCAATTC
Cladonia CTCCGGGGCTCCTTGGTGATTCACTACTCAGCGATCGCATGGCCCTTGGCCGGCGATGGTTCAATTC
Cladonia CTCCGGGGCTCCTTGGTGATTCACTACTCAGCGATCGCATGGCCCTTGGCCGGCGATGGTTCAATTC
Cladia CTCCGGGGCTCCTTGGTGATTCACTACTCAGCGATCGCATGGCCCTTGGCCGGCGATGGTTCAATTC
    
```

(b) Errors in the annotation

```

Saccharomyces CT-ATGCCGACTAGGGATCGGGTGGTGTTTTTT-RTTGR-CCCRCTCGGCACCTTACGAG-RAATCRAA
Saccharomyces CT-ATGCCGACTAGGGATCGGGTGGTGTTTTTT-RTTGR-CCCRCTCGGCACCTTACGAG-RAATCRAA
Saccharomyces CT-ATGCCGACTAGGGATCGGGTGGTGTTTTTT-RTTGR-CCCRCTCGGCACCTTACGAG-RAATCRAA
Saccharomyces CTNNMGCCGACTAGGGATCGGGTGGTGTTTTTT-RTTGNACCCCTCGGCACCTTACGAGRAATCRAA
Saccharomyces CTNNMGCCGACTAGGGATCGGGTGGTGTTTTTT-RTTGNACCCCTCGGCACCTTACGAGRAATCRAA
Saccharomyces CTNNMGCCGACTAGGGATCGGGTGGTGTTTTTT-RTTGNACCCCTCGGCACCTTACGAGRAATCRAA
    
```

(c) Many undetermined nucleotides

```

Blumeria CATARRCTATGCCGACTAGGGATCGGGCGATGTTATTTTTT---GACTCGCTCGGCACCTTACGAGRAAT
Leveillula CATARRCTATGCCGACTAGGGATCGGGCGATGTTATTTTTT---GACTCGCTCGGCACCTTACGAGRAAT
Pleochaeta CATARRCTATGCCGACTAGGGATCGGGCGATGTTATTTTTT---GACTCGCTCGGCACCTTACGAGRAAT
Oidium CATARRCTATGCCGACTAGGGATCGGGCGATGTTATTTTTT---GACTCGCTCGGCACCTTACGAGRAAT
Blumeria CATARRCTATGCCGACTAGGGATCGGGCGATGTTATTTTTT---GACTCGCTCGGCACCTTACGAGRAAT
Cystotheca CATARRCTATGCCGACTAGGGATCGGGCGATGTTATTTTTT---GACTCGCTCGGCACCTTACGAGRAAT
    
```

(d) Homopolymers

```

Uncultured AAGTAAAAGTCTTGGTTCCCGACACGCCAGTGAAGGGCATGAGGTTCCCCAGAGGAAAGGCGCCGGCCGG
Uncultured AAGTAAAAGTCTTGGTTCCCGACACGCCAGTGAAGGGCATGAGGTTCCCCAGAGGAAAGGCGCCGGCCGG
Uncultured AAGTAAAAGTCTTGGTTCCCGACACGCCAGTGAAGGGCATGAGGTTCCCCAGAGGAAAGGCGCCGGCCGG
Ascomycete AAGCAGAAAGTTAGGGG-ATCGAAGACGATCAGATACCGTCTGAGTCTTAAACCAATATATGCCGACTAGGG
Capnobotryella AAGCAGAAAGTTAGGGG-ATCGAAGACGATCAGATACCGTCTGAGTCTTAAACCAATATATGCCGACTAGGG
Ascomycota AAGCAGAAAGTTAGGGG-ATCGAAGACGATCAGATACCGTCTGAGTCTTAAACCAATATATGCCGACTAGGG
    
```

(e) Reverse complementary sequences

```

Taphrina AAGGAATTGACGGGAGGGCCACCAC-GGAGT-----
Taphrina AAGGAATTGACGGGAGGGCCACCAC-GGAGT-----
Protomyces AAGGAATTGACGGGAGGGCCACCAC-GGAGTAA-EGTTTATGTCCGATTAATCTGCTCCGAAAAGCC
Protomyces AAGGAATTGACGGGAGGGCCACCAC-GGAGTAAAGCTTTATGTCCGATTAATCTGCTCCGAAAAGCC
Protomyces AAGGAATTGACGGGAGGGCCACCAC-GGAGTAAAGCTTTATGTCCGATTAATCTGCTCCGAAAAGCC
N. irregularis AAGGAATTGACGGGAGGGCCACCAC-GGAGT-----
    
```

(f) Long insertions and introns

```

Phaffomyces TTTGATAGTTTTTTGTTACACGGGCAACCTCGGTAAATCTGATGCTACTACGTCGGTAAAGCCTTCGGGGT
Phaffomyces TTTGATAGTTTTTTGTTACACGGGCAACCTCGGTAAATCTGATGCTACTACGTCGGTAAAGCCTTCGGGGT
Phaffomyces TTTGATAGTTTTTTGTTACACGGGCAACCTCGGTAAATCTGATGCTACTACGTCGGTAAAGCCTTCGGGGT
Phaffomyces TTTGAT-----CAACCTCGGTAAATCTGATGCTACTACG-----
Phaffomyces TTTGATAGTTTTTTGTTACACGGGCAACCTCGGTAAATCTGATGCTACTACGTCGGTAAAGCCTTCGGGGT
Phaffomyces TTTGATAGTTTTTTGTTACACGGGCAACCTCGGTAAATCTGATGCTACTACGTCGGTAAAGCCTTCGGGGT
    
```

(g) Deletions



### **Tools within PHYMYCO-DB**

We designed PHYMYCO-DB with specific tools to facilitate online use. Firstly, users can easily select sequences by browsing our interface through the hierarchical taxonomic lineages arborescence (GenBank taxonomy), and then download them in a FASTA format file. The number of sequences stored in the database for each taxonomic level is given in brackets. Secondly, users can download an alignment file using a filter to find an alignment with the gene and the taxonomic rank requested. Special attention must be paid to the fact that some sequence characteristics in PHYMYCO-DB format are inherited from the extraction of GenBank sequences. For example, in some cases (e.g. Agaromycotina, a subphylum of Ascomycota), information on sequences taxonomy was associated to a 'no rank' tag in GenBank. To avoid the problem that these sequences are mistakenly placed in another taxonomic group, they were qualified as 'undefined' at the subphylum rank in PHYMYCO-DB. For the next lower taxonomic rank, no known tag problem exists. Environmental sequences have, by definition, no clear taxonomic ranking. Therefore, they were also qualified as 'undefined', but only until the lowest taxonomic rank. These are important features to take into account when using the PHYMYCO-DB.

Thirdly, users can launch a ClustalW 2.0 alignment on our back-end computer clusters by uploading their own personal sequences in a FASTA or ALN format file. A future PHYMYCO-DB release will offer the possibility to select the multiple alignment tool (i.e. ClustalW, MUSCLE & MAFFT likely). Currently, users can choose to append an outgroup or sequences from a particular PHYMYCO-DB taxonomic group. We anticipate that this tool will be very efficient when combined with phylogenetic analyses for investigating the sequence diversity of fungal amplicons from an environmental sample even to identify new fungal lineages.

### **PHYMYCO-DB as a tool for phylogenetic identifications and inferences**

Based on a well-developed theoretical corpus, phylogenies can be computed using several different approaches [e.g. 40]. From a mathematical point of view, the maximum likelihood phylogenetic reconstruction provides the best possible tree for a given explicit sequence evolution model. The model that best fits the aligned sequence data can be selected, after using the popular Modeltest [41]. Achieving a good alignment is therefore of tremendous importance for good interpretation. Alignments should be refined using 'influence function' that allow the removal of outlier columns from the matrix (i.e. nucleotide position where the phylogenetic signal differs from the general phylogenetic information recorded in the dataset) [42]. It must be emphasized that the sequence-based identification using SSU rRNA gene could be at the species level or at higher taxonomic levels depending on the fungal affiliation.

Following the above strategy, we provide an analysis of chytrid diversity as a proof of concept. Sequencing of the SSU rRNA gene achieved by targeting chytrids from deep marine hydrothermal samples (ciPCR). First, the alignment of SSU rRNA gene sequences of the Chytridiomycota from PHYMYCO-DB was used to design specific primers manually. Two sets of designed primers were covering the V3 and V4 variable regions and were suitable for pyrosequencing: C130 (5'TACCTTACTACTTGGATAACCG3') with SR8R (5'TCAAAGTAAAAGTCCTGGATC3') modified from Vilgalys lab webpage (<http://www.biology.duke.edu/fungi/mycolab/primers.htm>), and MH2 (5'TTCGATGGTAGGATAGAGG3') [43] with SR8R. Another set of primers, expected to be universal for fungi and to produce longer amplicons, were also tested: MH2 with NS7R (5'ATCACAGACCTGTTATTGCC3') modified from [44]. Primers specificity was checked with a sample from a hydrothermal site from which several sequences of chytrids were retrieved [45].

**Figure 3** : SSU rRNA phylogenetic positions of deep-sea Chytridiomycota (colored terminals) along with the closest known related SSU rRNA fungal sequences. Topology was built using MrBayes v.3.1.2 (Scale bar: 0.1 estimated substitutions per site, 3000000 generations sampled every 100 generations) from a ClustalW 2.1 alignment. The model GTR+I+G has been designated by jModelTest 0.1. Node support values are given in the following order: Maximum Parsimony/Maximum Likelihood (both calculated with PAUP 4.0 $\beta$ 10 version, 500 bootstraps)/MrBayes. *Corallochytrium limacisporum* (L42528), a putative *choanoflagellate*, was used as outgroup. *Maunachytrium keaense* (it is not part of PHYMYCO-DB) was also used to help build the tree. All sequences are listed with their GenBank accession numbers. The topologies were congruent (apart from dashed lines indicated in the figure). Bootstrap values of >50% are indicated at the nodes, bootstrap values of >70% are in bold. The sequences belonging to the *Lobulomycetaceae* family are indicated with their BLASTn percentage of maximum identity compared to the three deep-sea Chytridiomycota OTUs.

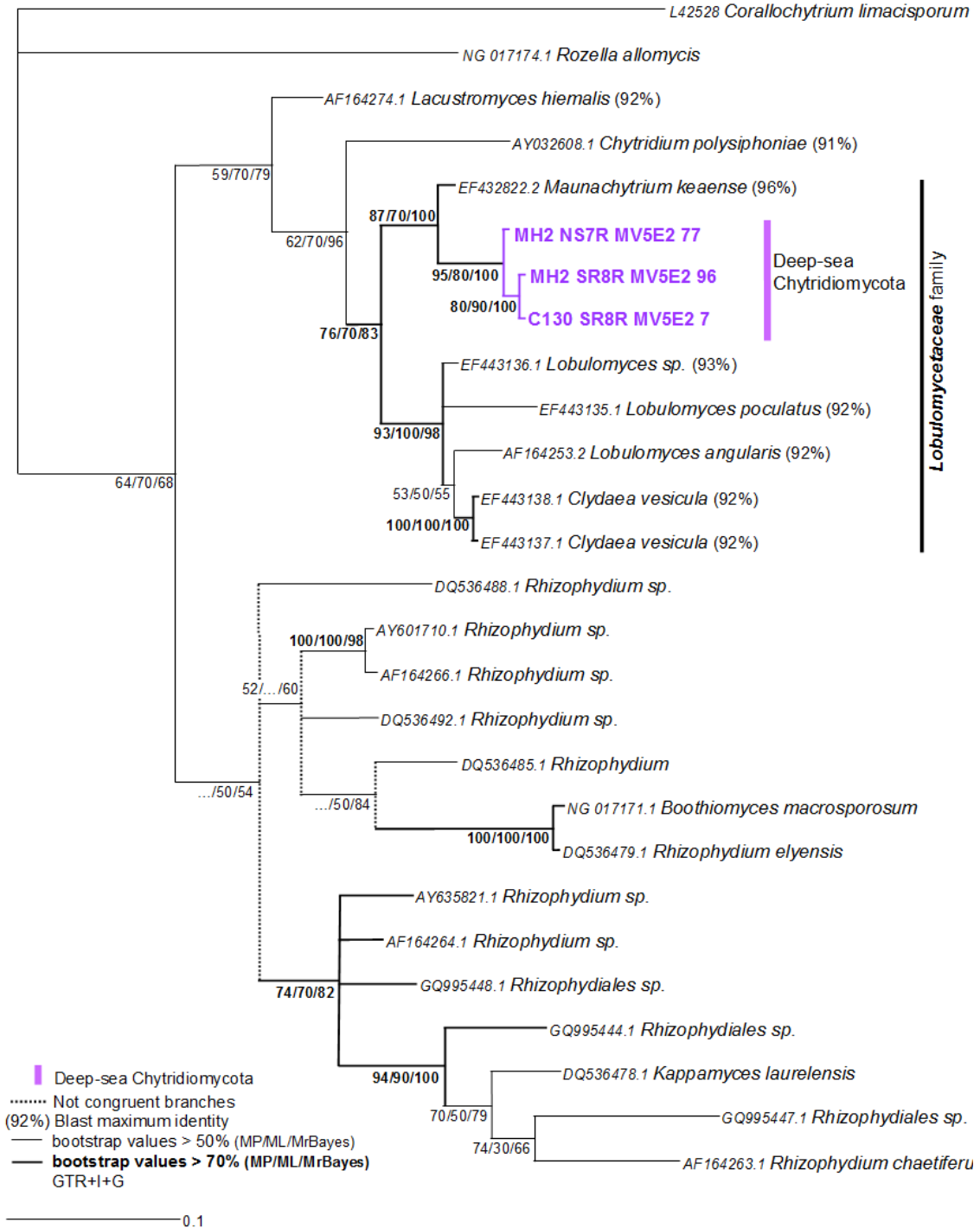


Figure 3

The resulting sequences (accession number JN986721 to JN986723) were analyzed using the corresponding 'reference' alignment in PHYMYCO-DB and the sequences having the highest similarity score in BLASTn. The phylogeny computed highlights the presence of a new group within the Chytridiomycota phylum (Fig. 3). The three OTUs found, each containing redundant sequences with high identity level (>98%), form a monophyletic group, whose closest relative is a sequence from the genus *Maunachytrium*. These OTUs constitute a new clade in the Lobulomycetaceae family [46]. BLASTn searches of these environmental sequence return the *Maunachytrium* sequence as the best hit, with a maximal identity of 96%. The widely used BLAST-based annotation of the environmental sequences, would end with an assignation to *Maunachytrium keaense* or *Maunachytrium* sp. In contrast to the phylogenetic analyses, we conclude that these sequences form a new clade within the Lobulomycetaceae family, outside the *Maunachytrium*, *Lobulomyces* (maximal identity 93%) and *Clydaea* (maximal identity 92%) genera. This exercise thus highlights the discrepancy between phylogenetically based annotation and BLASTn annotation.

## Discussion

The release of PHYMYCO-DB is expected to provide comprehensive access to fungal sequences for two phylogenetic markers (SSU rRNA and EF1- $\alpha$  genes) obtained from cultivated isolates, as well as environmental samples. As a result of deep sequence cleaning, the aligned sequences available in PHYMYCO-DB are of high quality (Fig. 1). To our knowledge, this curation strategy provides a novel approach to the problem of database pollution. As such, we anticipate that it will compliment other existing databases such as the "Assembling the Fungal Tree Of Life" project (AFTOL) [47], UNITE [33, 48] and MaarjAM [34] which 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.

Curation and annotation of ITS is made possible through the web-based-workbench of PlutoF [49]. Initially, the UNITE system contained ITS and nLSU/28S rRNA gene sequences from Basidiomycota and Ascomycota. ITS 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 [50], as is the case for Ascomycota and Basidiomycota. ITS are often used to resolve phylogenetic relationships at the species level or at the infraspecific level [51]. However, as ITS displays high sequence variability, even within a given organism as in Glomeromycota [i.e. 13], obtaining reliable alignments with this marker can be difficult [52]. This is because 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 [i.e. 1, 53]. The MaarjAM database has focused on SSU rRNA gene of arbuscular mycorrhizal fungi (Glomeromycota), with associated metadata. The existence of these databases and the potential emergence of others should be encouraged. It enables the community to have access to reliable sequences.

For fungal sequence annotations and phylogenetic interpretations of fungal environmental sequences, one of the main advantages of PHYMYCO-DB is to facilitate the primer design and subsequent phylogenetic analyses of amplicons as shown in the example above (Fig. 3). The use of PHYMYCO-DB to perform expert analyses appears to be complementary to BLASTn, the latter allowing a quick look of the query sequence proximity compared to the available sequences. From our example, we can speculate that different apparent polyphyletic groups might result in wrong annotation. We anticipate that the use of PHYMYCO-DB will help to limit incorrect SSU rRNA and EF1- $\alpha$  genes fungal annotation propagation in sequence databases.

### Availability and future directions

The PHYMYCO-DB is available via a web-based interface at <http://phymycodb.genouest.org/> on the GenOuest bioinformatics platform web site. The web interface is divided into 2 parts. The first part, entitled “*DB admin*”, is restricted to the administrators for use in 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. For the PHYMYCO-DB release, we have focused on the SSU rRNA and EF1- $\alpha$  genes, i.e. concentrating on the targets most commonly used to analyze fungal diversity and evolution. We offer the opportunity to download specific sequences, sets of sequences chosen by user, alignment files prepared from subsets of the PHYMYCO-DB and also to perform sequence alignments using ClustalW 2.0. The next PHYMYCO-DB releases will include (i) the provision of alignment files in which outlier nucleotides identified from influence functions [42] will be highlighted, so that users can then delete these sites (ii) taxonomic modifications within Chytridiomycota and Zygomycota after Hibbett et al. (2007) [54] and after Jones et al. (2011) [5]. PHYMYCO-DB will continue to expand with new genes. We are currently investigating  $\beta$ -tubulin (*tub1*, *tub2*), actin (*act1*), and RNA polymerase II subunits (*rpb1* and *rpb2*) as potential interesting targets. PHYMYCO-DB will also be improved by incorporating all the finished fungal genomes available, and increasing the diversity of tools to perform multiple alignments.

### Acknowledgements

We thank the GenOuest Bioinformatics core facility.

### References

- [1]Vandenkoornhuysen P, Baldauf SL, Leyval C, Straczek J, Young JPW (2002a). Extensive fungal diversity in plant roots. *Science* 295: 2051.

- [2] Purvis A, Hector A (2000) Getting the measure of biodiversity. *Nature* 405: 212-219.
- [3] Taylor JW, Jacobson DJ, Kroken S, Kasuga T, Geiser DM (2000) Phylogenetic species recognition and species concept in fungi. *Fung Genet Biol* 31: 21-32.
- [4] Vandenkoornhuyse P, Husband R, Daniell TJ, Watson IJ, Duck JM et al. (2002b). Arbuscular mycorrhizal community composition associated with two plant species in a grassland ecosystem. *Mol Ecol* 11: 1555-1564.
- [5] Jones, MDM, Forn, I, Gadelha, C, Egan, MJ, Bass, D, et al. (2011) Discovery of novel intermediate forms redefines the fungal tree of life. *Nature* 474, 200–203
- [6] Powell JR, Monaghan MT, Öpik M, Rillig MC (2011) Evolutionary criteria outperform operational approaches in producing ecologically relevant fungal species inventories. *Molecular Ecology* 20: 655-666.
- [7] Hawksworth DL, Rossman AY (1997) Where are all the undescribed fungi? *Phytopath* 87: 888-891.
- [8] Blackwell M (2011) The Fungi: 1, 2, 3 ... 5.1 million species? *Am J Bot* 98: 426-438
- [9] Woese CR (2000) Interpreting the universal phylogenetic tree. *Proc Natl Acad Sci USA* 97: 8392-8396.
- [10] Pace NR (2009) Mapping the Tree of Life: Progress and Prospects. *Microbiology and Mol Biol Rev* 73: 565-576.
- [11] Choi IG, Kim SH (2007) Global extent of horizontal gene transfer. *Proc Natl Acad Sci USA* 104: 4489-4494
- [12] Boon E, Zimmerman E; Lang BF, Hijri, M (2010) Intra-isolate genome variation in arbuscular mycorrhizal fungi persists in the transcriptome. *J Evol Biol* 23:1519-1527
- [13] Sanders IR, Alt M, Groppe K, Boller T, Wiemken A (1995) Identification of ribosomal DNA polymorphisms among and within spores of the Glomales - Applications to study on the genetic diversity of Arbuscular Mycorrhizal fungal communities. *New Phytol* 130: 419-427
- [14] Lim YW, Sturrock R, Leal I, Pellow K, Yamaguchi T, Breuil C (2008) Distinguishing



homokaryons and heterokaryons in *Phellinus sulphurascens* using pairing tests and ITS polymorphisms. *Antonie van Leeuwenhoek* 93 : 99-110.

- [15]Benson DA, Karsch-Mizrachi I, Lipman DJ, Ostell J, Sayers EW (2011) GenBank. *Nucl Acids Res* 39: D32-D37
- [16]Kulikova T, Akhtar R, Aldebert P, Althorpe N, Andersson M, et al. (2007) The EMBL nucleotide sequence database in 2006. *Nucl Acids Res* 35: D16-D20.
- [17]Kaminuma E, Kosuge T, Kodama Y, Aono H, Mashima J et al. (2011) DDBJ progress report. *Nucl Acids Res* 39: D22-D27.
- [18]Avice JC (2004) *Molecular Markers, Natural History, and Evolution*. Sunderland, Massachusetts: Sinauer Associates.
- [19]Baldauf SL (1999) A search for the origins of Animals and Fungi: comparing and combining molecular data. *Am Nat* 154: S178-S188
- [20]Baldauf SL, Roger AJ, Wenk-Siefert I, Doolittle WF (2000) A kingdom-level phylogeny of eukaryotes based on combined protein data. *Science* 290: 972-977
- [21]Helgason T, Watson IJ, Young JPW (2003) Phylogeny of the Glomerales and Diversisporales (Fungi: Glomeromycota) from actin and elongation factor 1-alpha sequences. *FEMS Microb Let* 229: 127-132.
- [22]Baldauf SL, Palmer JD (1993) Animals and fungi are each others closest relatives: congruent evidence from multiple proteins. *Proc Natl Acad Sci USA* 90: 11558-11562
- [23]Moon CD, Miles CO, Jarlfors U, Schardl CL (2002) The evolutionary origins of three new *Neotyphodium* endophyte species from grasses indigenous to the Southern Hemisphere. *Mycologia* 94: 694–711.
- [24]Tanabe Y, Saikawa M, Watanabe MM, Sugiyama J (2004). Molecular phylogeny of Zygomycota based on EF-1 $\alpha$  and RBP1 sequences: limitations and utility of alternative markers to rDNA. *Mol Phyl Evol* 30: 438-449.
- [25]Bidartondo MI (2008) Preserving accuracy in GenBank. *Science* 319: 1616

- [26]Bridge PD, Roberts PJ, Spooner BM, Panchal G (2003) On the unreliability of published DNA sequences. *New Phytol* 160: 43-48
- [27]Vilgalys R (2003) Taxonomic misidentification in public DNA databases. *New Phytol* 160: 4-5.
- [28]Bridge PD, Spooner BM, Roberts PJ (2004) Reliability and use of published sequence data. *New Phytol* 161: 15-17
- [29]Hawksworth DL (2004) 'Misidentification' in fungal DNA sequence databanks. *New Phytol* 161: 13-15.
- [30]Hawksworth DL (2009) Separate name for fungus's sexual stage may cause confusion. *Nature* 458: 29.
- [31]Holst-Jensen A, Vrålstad T, Schumacher T (2004) On reliability. *New Phytol* 161: 11-13.
- [32]Nilsson RH, Ryberg M, Kristiansson E, Abarenkov K, Larsson KH et al. (2006) Taxonomic reliability of DNA sequences in public sequence databases: a fungal perspective. *PloS One* 1: e59.
- [33]Kõljalg U, Larsson KH, Abarenkov K, Nilsson RH, Alexander IJ, et al. (2005) UNITE: a database providing web-based methods for the molecular identification of ectomycorrhizal fungi. *New Phytol* 166: 1063-1068.
- [34]Öpik M, Vanatoa A, Vanatoa E, Moora M, Davison J et al. (2010) The online database MaarjAM reveals global and ecosystemic distribution patterns in arbuscular mycorrhizal fungi (Glomeromycota). *New Phytol* 188: 223-241.
- [35]Cole JR, Wang Q, Cardenas E, Fish J, Chai B, et al. (2009) The Ribosomal Database Project: Improved alignments and new tools for rRNA analysis. *Nucl Acids Res* 37: D141-D145
- [36]Pruesse E, Quast C, Knittel K, Fuchs BM, Ludwig W, (2007) SILVA: a comprehensive online resource for quality checked and aligned ribosomal RNA sequence data compatible with ARB. *Nucl Acids Res* 35: 7188-7196.

- [37]DeSantis TZ, Hugenholtz P, Larsen N, Rojas M, Brodie EL, Keller K, Huber T, Dalevi D, Hu P, Andersen GL (2006) Greengenes, a Chimera-Checked 16S rRNA Gene Database and Workbench Compatible with ARB. *Appl Env Microb* 72: 5069-5072
- [38]Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA et al. (2007) Clustal W and Clustal X version 2.0. *Bioinformatics* 23: 2947-2948.
- [39]Bhattacharya D, Friedl T, Helms G (2002) Vertical evolution and intragenic spread of lichen-fungal group I intron. *J Mol Evol* 55: 74-84
- [40]Felsenstein J (2004) *Inferring phylogenies*. Sinauer Associates, Sunderland, Massachusetts.
- [41]Posada D, Crandall KA (1998) MODELTEST: testing the model of DNA substitution. *Bioinformatics* 14: 817-818.
- [42]Bar-Hen A, Mariadassou M, Poursat MA, Vandenkoornhuyse P (2008) Influence function for robust phylogenetic reconstructions. *Mol Biol Evol* 25: 869-873
- [43]Vandenkoornhuyse P, Leyval C (1998) SSU rDNA sequencing and PCR-fingerprinting reveal genetic variation within *Glomus mossae*. *Mycologia* 90: 791-797.
- [44]White TJ, Bruns T, Lee S, Taylor J (1990) Amplification and direct sequencing of fungal ribosomal RNA genes for phylogenetics. In: Innis MA, Gelfand DH, Sninsky JJ, White TJ editors. *PCR protocols, a guide to methods and applications*. San Diego, California Academic Press pp. 315-322.
- [45]Le Calvez T, Burgaud G, Mahé S, Barbier G, Vandenkoornhuyse P (2009) Fungal diversity in deep-sea hydrothermal ecosystems. *Appl Env Microb* 75: 6415-6421.
- [46]Simmons DR, James TY, Meyer AF, Longcore JE (2009) Lobulomycetales, a new order in the Chytridiomycota. *Mycological Res* 113: 450-460.
- [47]Lutzoni F, Kauff F, Cox CJ, McLaughlin D, Celio G, et al. (2004) Assembling the fungal tree of life: progress, classification, and evolution of subcellular traits. *American J Bot* 91: 1446-1480.
- [48]Abarenkov K, Nilsson RH, Larsson KH, Alexander IJ, Eberhardt U, et al. (2010a). The

- UNITE database for molecular identification of fungi - recent updates and future perspectives. *New Phytol* 186: 281-285
- [49]Abarenkov K, Tedersoo L, Nilsson RH, Vellak K, Saar I, et al. (2010b) PlutoF-a Web Based Workbench for Ecological and Taxonomic Research, with an Online Implementation for Fungal ITS Sequences. *Evol Bioinformatics* 6: 189-196
- [50]Anderson IC, Parkin PI (2007). Detection of active soil fungi by RT-PCR amplification of precursor rRNA molecules. *J Microb Meth* 68: 248–253
- [51]Xu P, Han Y, Wu J, Lv H, Qiu L et al. (2007) Phylogenetic Analysis of the Sequences of rDNA Internal Transcribed Spacer (ITS) of *Phytophthora sojae*. *J Genet Genom* 34: 180-188.
- [52]D’Auria G, Pushker R, Rodriguez-Valera F (2006) IWoCS analyzing ribosomal intergenic transcribed spacers configuration and taxonomic relationships. *Bioinformatics* 22: 527–531
- [53]Bass D, Howe A, Brown N, Barton H, Demidova M et al. 2007. Yeast forms dominate fungal diversity in the deep oceans. *Proc Royal Soc B* 274: 3069-3077
- [54]Hibbett DS, Binder M, Bischoff JF, Blackwell M, Cannon PF et al. (2007) A higher-level phylogenetic classification of the Fungi. *Mycol Res* 111: 509-547.



# CHAPITRE 2

Fungal diversity

in deep-sea hydrothermal ecosystems



## Chapitre 2

---

### Fungal diversity in deep-sea hydrothermal ecosystems

Cette étude est antérieure à mon travail de thèse. Nous avons choisi de l'intégrer au manuscrit de thèse car j'ai contribué à différentes étapes de sa réalisation lors d'un contrat d'ingénieur d'étude. Ainsi, j'ai participé à la génération des séquences environnementales, aux analyses phylogénétiques et à la rédaction du manuscrit. De plus, les résultats qui ont été obtenus sont à l'origine de ce travail de thèse.

Le premier inventaire des champignons dans les écosystèmes hydrothermaux marins profonds a été obtenu par cette étude. La diversité fongique a été étudiée en analysant les séquences du gène de l'ARNr 18S amplifiées par PCR culture-indépendante à partir d'ADN extrait d'échantillons hydrothermaux et des souches obtenues par la mise en culture. Nos travaux ont révélé une diversité insoupçonnée d'espèces appartenant à trois des cinq phyla fongiques. Parmi celle-ci, des séquences se positionnant parmi les Chytridiomycota ont été amplifiées et montre une divergence marquée avec les espèces décrites. Enfin, une discussion sur la diversification des champignons en milieu marin est proposée.





## Fungal Diversity in Deep-Sea Hydrothermal Ecosystems<sup>∇†</sup>

Thomas Le Calvez,<sup>1</sup> Gaëtan Burgaud,<sup>2</sup> Stéphane Mahé,<sup>1</sup>  
Georges Barbier,<sup>1</sup> and Philippe Vandenkoornhuise<sup>2\*</sup>

UMR 6553 CNRS ECOBIO, FR/90 IFR 2116 CAREN, Université de Rennes 1, Campus de Beaulieu, 35042 Rennes Cedex, France,<sup>1</sup> and Laboratoire de Biodiversité et Ecologie Microbienne (EA 3882), Université de Bretagne Occidentale, ESMISAB, Technopôle Brest Iroise, 29280 Plouzané, France<sup>2</sup>

Received 19 March 2009/Accepted 18 July 2009

Deep-sea hydrothermal ecosystems are considered oases of life in oceans. Since the discovery of these ecosystems in the late 1970s, 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. These organisms are considered key organisms in terrestrial ecosystems because of their ecological functions and especially their ability to degrade organic matter. The lack of knowledge about them in the sea reflects the widely held belief that fungi are terrestrial organisms. The first inventory of such organisms in deep-sea hydrothermal environments was obtained in this study. Fungal diversity was investigated by analyzing the small-subunit rRNA gene sequences amplified by culture-independent PCR using DNA extracts from hydrothermal samples and from a culture collection that was established. Our work revealed an unsuspected diversity of species in three of the five fungal phyla. We found a new branch of *Chytridiomycota* forming an ancient evolutionary lineage. Many of the species identified are unknown, even at higher taxonomic levels in the *Chytridiomycota*, *Ascomycota*, and *Basidiomycota*. This work opens the way to new studies of the diversity, ecology, and physiology of fungi in oceans and might stimulate new prospecting for biomolecules. From an evolutionary point of view, the diversification of fungi in the oceans can no longer be ignored.

Since the discovery of hydrothermal vent ecosystems 30 years ago, unexpected species diversity has been revealed that has shed light on the functional coupling between the geosphere and the biosphere. When submersibles dive to the seafloor, they bring numerous organisms back to the surface, and this has resulted in the description of nearly two new species per month (10). Deep-sea hydrothermal ecosystems are considered hotspots of microbial diversity on the seafloor. Indeed, they are ecosystems that produce biomass using the wide range of chemical compounds released by the polymetallic sulfite chimneys or “black smokers” that represent the huge quantity of chemical energy that is available (26). The vent fluid, having been heated close to a magma chamber, can have a temperature of 400°C when it is emitted. It is also characterized by a lack of dissolved oxygen, strong acidity (pH 2 to 3), a high concentration of electron donors (i.e., reduced compounds such as methane and hydrogen sulfide), and the presence of heavy metals (36). Continual mixing with the cold ocean water (2 to 4°C) that is rich in electron acceptors creates a dynamic chemical disequilibrium that is a source of energy for microorganisms that control the rates of redox reactions (16).

Each ridge displays varied geochemistry, and the vent fluids differ, even at scales as small as the fractures, pipes, and porosities in the black smokers, creating diverse microhabitats for biota (10, 16, 26). Microbes have colonized these different

microsites; an impressive diversity of *Bacteria* and *Archaea* species has been described, and even more diversity has been inferred. Some organisms are chemolithoautotrophic primary producers and are 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, reportedly are able to harvest energy from geothermal radiation (3). 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.

A comparison of hydrothermal vent cycling of carbon with the carbon cycling in terrestrial ecosystems suggests that key organisms in hydrothermal vents, the fungi, have not been studied yet. Contemporary organisms living in land ecosystems are well adapted. However, the first colonizers of land faced a harsh physical environment (29). Thus, it has been suggested that early colonization of land was possible only when there was a macroevolutionary jump (i.e., acquisition of multiple competences during a single evolutionary step) (24, 29). It has been suggested that establishment of eukaryotes on land may have been possible only as a consequence of the association of a fungus and a phototroph (24, 29). The earliest known fossil fungus (460 million years) was thought to form arbuscular mycorrhizae (27). The idea that the colonization of land was made possible by this symbiotic association is supported the findings of Simon et al. (31). Based on molecular clock estimates, some authors have suggested that fungi appeared a long time before Devonian land colonization. The major radiations in the fungi may even have appeared during the Precambrian (1,460 to 960 million years ago) (13), 300 million or more years before colonization of land. These analyses are currently under

\* Corresponding author. Mailing address: UMR 6553 CNRS ECOBIO, FR/90 IFR 2116 CAREN, Université de Rennes 1, Campus de Beaulieu, Bat. 14a, 35042 Rennes Cedex, France. Phone: 33-2 23 23 50 07. Fax: 33-2 23 23 68 28. E-mail: philippe.vandenkoornhuise@univ-rennes1.fr.

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

∇ Published ahead of print on 24 July 2009.

discussion because of the absence of fossil evidence to support the conclusions and because of possible bias in the molecular clock estimates (34). However, assuming that the protein estimates are correct, it could be hypothesized that fungi emerged and diversified in water before they colonized land. An ancestral characteristic, a flagellated gamete without a wall, is found in old lineages of fungi. This is generally considered an advantage for dispersal and reproduction in water. The loss of flagella in fungi appears to have occurred in a single evolutionary step (20) and has been used as an argument to support the dogma that higher fungi diversified on land before a secondary colonization of water ecosystems, assuming that some terrestrial fungi secondarily invaded salt marshes and mangroves and adapted to life in salt and brackish water. It is a common belief among ecologists, microbiologists, and mycologists that fungi are found (almost) exclusively on land. A recent study showed that the loss of flagella may have occurred at least four times during fungal evolution (15), which suggests an evolutionary paradigm that is more complicated than previously thought. Despite the view that fungi diversified in terrestrial ecosystems, chytrids are considered primarily aquatic fungi. Furthermore, 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 (17). We advocate an alternative hypothesis for the loss of flagella. It can be proposed that the loss of motile gametes in fungi was compensated for by the resistance and long-range dispersal of spores. We suggest that this evolutionary innovation in eukaryotes should have led to colonization and long-term persistence in many new environments, including land, even if the primitive transition of fungi to terrestrial life could have occurred without the previous loss of flagellated spores. From the molecular clock predictions (13) and partially in agreement with our computations (P. Vandenkoornhuyse, unpublished data), we suggest that this evolutionary event might have occurred approximately 1,200 to 1,000 million years ago and, therefore, in oceans. If this is true, we would expect to find a large variety of fungi, including fungi producing nonflagellated spores, in oceanic samples even in isolated habitats.

Consequently, the main aim of the present work was to test for the presence of fungi in deep-sea hydrothermal ecosystems by establishing a culture collection and using a culture-independent approach involving direct amplification of the fungal small-subunit (SSU) rRNA gene from environmental samples and then analysis and evaluation of the diversity of the fungi in these environments. The deep-sea hydrothermal ecosystem was an interesting target to test our hypotheses for various reasons. First, it has been suggested that the hydrothermal ecosystem is the "cradle of life" (14, 18, 39), although more recently it has been shown that life emerged first in a mesophilic environment and diversified in thermophilic conditions (5). Second, hydrothermal conditions were very common during the Precambrian (28). Third, use of general eukaryotic primers to perform culture-independent PCRs (ciPCRs) has led to the detection of fungi close to hydrothermal vents previously (21, 22). Fungal sequences have also been detected in deep-sea anoxic sediments (2, 8, 9), and barophilic fungal strains have been isolated from this environment (7).

## MATERIALS AND METHODS

**Studied sites and sampling.** Samples were collected during the following two oceanographic cruises: (i) HERO (30 September 1991 to 4 November 1991) on the East Pacific Rise at the Elsa site (12°48'N, 103°57'W; depth, 2,630 m) and (ii) MARVEL (29 August 1997 to 13 September 1997) on the Mid-Atlantic Ridge at the Menez Gwen site (37°51'N, 31°31'W; depth, 860 m) and Lucky Strike site (37°17'N, 32°16'W; depth, 1,700 m). Deep-sea sampling was performed using DSV *Nautile*. The support research vessels were R/V *Nadir* for the HERO cruise and R/V *Atalante* for the MARVEL cruise ([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 and then closed and placed in the insulated crate of the submersible to maintain a low temperature until subsequent examination on board.

**Sample preparation and fungal cultures.** On board in the lab, sterility was obtained with a Bunsen burner and a vertical laminar flow hood. Solid samples were removed from the container with sterile strips, placed on a sterile petri dish, rinsed well with sterile seawater, and then crushed with a sterile pestle and mortar. After this, the samples were aliquoted and preserved by deep-freeze storage (−80°C) until they were used.

During the HERO cruise, Sabouraud chloramphenicol solid medium (AES Laboratory) was used for aerobic enrichment cultures incubated at 30°C with atmospheric pressure. During the MARVEL cruise, five solid culture media were used. Ac medium contained (per liter) 5 g potato starch (Sigma), 0.5 g yeast extract, 1 g peptone, 30 g sea salts, and 6.05 g piperazine-*N,N'*-bis(2-ethanesulfonic acid) (PIPES) buffer (Sigma). The starch was replaced by 5 g cellobiose (Sigma) in Cc medium, by 5 g glucose (Sigma) in Gc medium, by 5 g arabic gum and 5 g olive oil in Lc medium, and by 5 g xylan oat spelt in Xc medium. Pc medium contained (per liter) 9 g brain heart infusion (Difco), 23 g NaCl, and 6.05 g PIPES buffer. The pH was adjusted to 7.5 with 4 N NaOH. The five media were supplemented with (per liter) 15 g agar and 500 mg chloramphenicol. 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) 1 g glucose, 1 g yeast extract, 1 g peptone, 1 g starch, and 30 g sea salts. Pure cultures were obtained and kept in a culture collection by continuous culture and then freeze-dried at −80°C. More details about sampling and cultures are available in a recently published paper (6).

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

**Cloning and sequencing.** The fungal SSU rRNA genes were amplified by PCR using primers MH2 (5'TTCGATGGTAGGATAGAGG3') and MH4 (5'GTCTCACTAAGCCATTC3') (37) for environmental sequences and primers NS1 (5'GTAGTCATATGCTTGCTCTC3') and ITS5R (5'CCTTGTACGACTTTACTTCC3') primers for cultivated strains (40). The AU2/AU4 primer set (38) did not amplify any of the samples analyzed. DNA was amplified by touchdown PCR with the initial step at 94°C for 1.30 min followed by 37 cycles of 94°C for 30 s, 48°C for 1.25 min (with a 0.1°C decrease at each cycle), and 72°C for 1.5 min. The PCR amplification program ended with a final elongation step at 72°C for 10 min. For the samples which did not yield any product we used a primer set modified from the primer set described by Smit et al. (32), EF3 (5'TCCTCTA AATGACCAGTTTG3') and EF4 (5'GGAAGGGNTGTATTATTAGAT3'), and the same PCR conditions except for the annealing temperature (53°C instead of 48°C). For these ciPCRs, a dilution of a soil DNA extract and DNA-free water were used as positive and negative controls, respectively. The amplification fragments were purified with a High Pure PCR product purification kit (Roche) and were cloned in the pGEM-T vector (Promega) and DH5 $\alpha$  competent cells (Gibco BRL Life Technologies). For each clone library, 16 to 120 randomly selected positive clones were sequenced (ABI-PRISM Dye Terminator cycle sequencing Ready Reaction kit; Perkin Elmer) on both strands using primers T7 and SP6 for environmental sequences and primers NS1, NS3 (5'GCAAGTCTGGTGCCAGCAGCC3'), and ITS5R for the culture sequences (6). The contig of the two sequence strands was calculated using Sequencher 4.6 (GeneCodes), and alignment uncertainties were checked manually. Chimeric artifacts were deleted from the data set using CHIMERA\_CHECK 2.7 (<http://rdp.cme.msu.edu/html/analyses.html>; Ribosomal Database Project).

**rRNA fungal database.** A fungal SSU rRNA gene database was created so that the detected sequences could be analyzed (<http://phymycodb.genouest.org/>). Al-

gorithms were used to recover SSU rRNA gene sequences for each fungal phylum from the GenBank/EMBL/DBJ databases. Filters were used to eliminate the sequences that were too short (<1,000 bp), too long (>2,500 bp), or of poor quality or that included at least 10 consecutive undetermined nucleotides (T. Le Calvez, L. Guillot, A. Dufresne, and P. Vandenkoornhuys, submitted for publication). A multiple-sequence alignment was constructed for each phylum using Clustal X 1.81 with 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 1,733 sequences from *Basidiomycota*, 215 sequences from *Chytridiomycota*, 4,117 sequences from *Ascomycota*, 292 sequences from *Zygomycota*, and 621 sequences from *Glomeromycota*, which represented all the different branches of fungal phylogeny. Each phylum was then subjected to a multiple-sequence alignment procedure, followed by neighbor-joining analyses. Phylogenetic trees were visualized by using Treeview 1.6.6. The phylogenetic neighbors closest to the environmental sequences analyzed were selected, and then phylogenetic analyses were performed (see below).

**Diversity and phylogenetic analyses.** Rarefaction curves were computed for each sample. The number of species was determined for 100 random combinations of 1 to  $n$  sequences by using 100 bootstrap pseudoreplicates (<http://viceroy.ceb.uconn.edu/EstimateS>) implemented in EstimateS. A multiple-sequence alignment procedure was performed using CLUSTALX 1.81 (35), and the alignment was refined by eye. This analysis included our environmental sequences, sequences from the fungal cultures isolated, and the most relevant representative SSU rRNA gene sequences of the *Chytridiomycota*, *Basidiomycota*, and *Ascomycota* (93 sequences, including 18 representative environmental sequences and 26 sequences from fungal cultures). The alignment was performed with 1,145 nucleotides as described above. Other multiple-sequence alignment methods tested gave similar results (not shown). Phylogenetic analyses were performed as follows. CLUSTALX 1.81 was used to obtain neighbor-joining (NJ-K2P) phylogenies with distance correction, gap omission, and 1,000 bootstrap pseudoreplicates. PAUP 4.0 $\beta$ 10 was used for maximum parsimony analysis using a heuristic tree search with 500 random-addition replicates and with tree bisection and reconnection as the branching algorithm, with each of the 500 bootstrap iterations utilizing 10 random-addition replicates and tree bisection and reconnection branch swapping. In addition, the data were analyzed using the ML-GTR + I + G procedure with 100 iterations. Modeltest 3.7 software (25) was used to select the model.

After these analyses were performed, phylotypes were specified using a cutoff of 98% (pairwise distance computed using PAUP 4.0 $\beta$ 10).

**Quantitative PCR (qPCR) assays.** PCRs were performed with 10- $\mu$ l (final volume) mixtures using iQ SYBR green Supermix (Bio-Rad), which contained SYBR green PCR buffer, 2.7  $\mu$ M dATP, 2.7  $\mu$ M dTTP, 2.7  $\mu$ M dGTP, 2.7  $\mu$ M dCTP, and 0.42 U of iTaq DNA polymerase (Bio-Rad). Then 0.35  $\mu$ M primer MH2 (5'TTCGATGGTAGGATAG3') and 0.35  $\mu$ M primer FungqPCR1 (5'TGTCGGGATTGGGTAATTT3') were added to the mixture.

All reactions were performed in optical tubes (Bio-Rad); 8.5  $\mu$ l of the master mixture was added first, followed by 1.5  $\mu$ l of template. The tubes were sealed with microseal film (Bio-Rad). All reactions were performed with a Chromo 4 thermocycler (MJ Research), using an initial denaturation at 94°C for 3 min to activate the enzyme, followed by 35 or 40 cycles of denaturation at 94°C for 30 s and annealing-extension at 48°C for 45 s and then by plate reading. The dissociation curve for temperatures 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 two preliminary runs were compared in a final run. Seven different plasmid concentrations ( $10^9$  to  $10^3$  copies in 1.5  $\mu$ l) were used to construct a standard curve for absolute quantification. The numbers of copies in the standards were calculated using the following formula: molecules/ $\mu$ l =  $a/(\text{plasmid length} \times 660) \times (6.022 \times 10^{23})$ , where  $a$  is the plasmid concentration (in  $\mu$ g/ $\mu$ l), 660 is the average molecular weight of one base pair, and  $6.022 \times 10^{23}$  is the molar constant.

**Nucleotide sequence accession numbers.** The sequences reported in this paper have been deposited in the GenBank database under accession numbers EF638466 to EF638515 for sample MV2E1, EF638516 to EF638564 for sample H18E12, EF638565 to EF638614 for sample MV2E2, EF638615 to EF638636 for sample MV2E3, EF638637 to EF638654 for sample MV5E1, EF638655 to EF638686 for sample MV5E2, and EF638687 to EF638706 for cultivable strains. The matrix is available upon request from the corresponding author.

## RESULTS

**Diversity of the culture collection.** Cultivation and isolation of microbes (*Bacteria* and *Archaea*) from deep-sea hydrothermal samples are difficult, and the procedure includes many constraining manipulations. Nevertheless, we were able to establish a fungal collection with different samples from several sites without using complex culture procedures, such as pressurized fermentors (see Materials and Methods). The fungi were isolated using a classic medium supplemented with sea salt at a pressure of 1 atm and 25°C. Physiological analyses demonstrated that the strains were able to grow in deep-sea salinity (3% [wt/vol] NaCl) and at low temperatures (6). 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 (15). A subset of the cultivated *Ascomycota* belonged to the classes *Dothideomycetes* (Fig. 1, phylotypes 14 to 16) and *Sordariomycetes* (Fig. 1, phylotypes 17, 18, and 20), and at least one new undescribed phylotype belonging to each of these groups was discovered. Black yeasts were also detected, including two phylotypes of *Exophiala* (order *Chaetothyriales*) (Fig. 1, phylotypes 10 to 13), a group that includes pathogens and saprophytes. In a recent study, black yeasts were identified in mussel tissues present in hydrothermal samples from the Fidji Basin, where they appeared to be mussel parasites, possibly regulating the trophic chain (37). Finally, one phylotype was obtained from cultures that clustered with the mitosporic *Ascomycota* (Fig. 1, phylotype 19), which have been described as "common" pathogenic fungi in tidal seawater (4, 30). As expected based on previous information, the spores produced by the cultivated isolates were not flagellated, even though some isolates appeared to be specific to the oceanic habitat (for details about the culture collection, physiology, and morphology, see reference 6).

**Fungal diversity determined by a culture-independent approach.** We investigated the fungal SSU rRNA genes from environmental samples (Table 1) directly amplified by ciPCRs using fungus-specific primers. A customized database was constructed to analyze in depth the SSU rRNA gene sequences detected. This database contained all available fungal SSU rRNA sequences that passed filters for length and quality (144 *Chytridiomycota* sequences, 266 *Zygomycota* sequences, 418 *Glomeromycota* sequences, 2,327 *Basidiomycota* sequences, and 4,270 *Ascomycota* sequences), and there was at least one representative of each known fungal genus in the GenBank database. All of 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 constructed using these computations along with BLASTn analyses (1), which included the closest relatives of the environmental SSU rRNA sequences analyzed (for each phylum, neighbor-joining phylogenetic trees obtained using the entire data sets are shown in Fig. S1, S2, S3, and S4 in the supplemental material). A total of 20 distinct phylotypes were detected using the ciPCR results and the fungal culture collection (Fig. 1). The ciPCR results revealed previously unsuspected diversity of fungal phylotypes in the vent ecosystems; there were nine distinct phylotypes, five of which were new at the genus level or a higher taxonomic level in the *Chytridiomycota* (Fig. 1, phylotypes 1 and 2; see Fig.



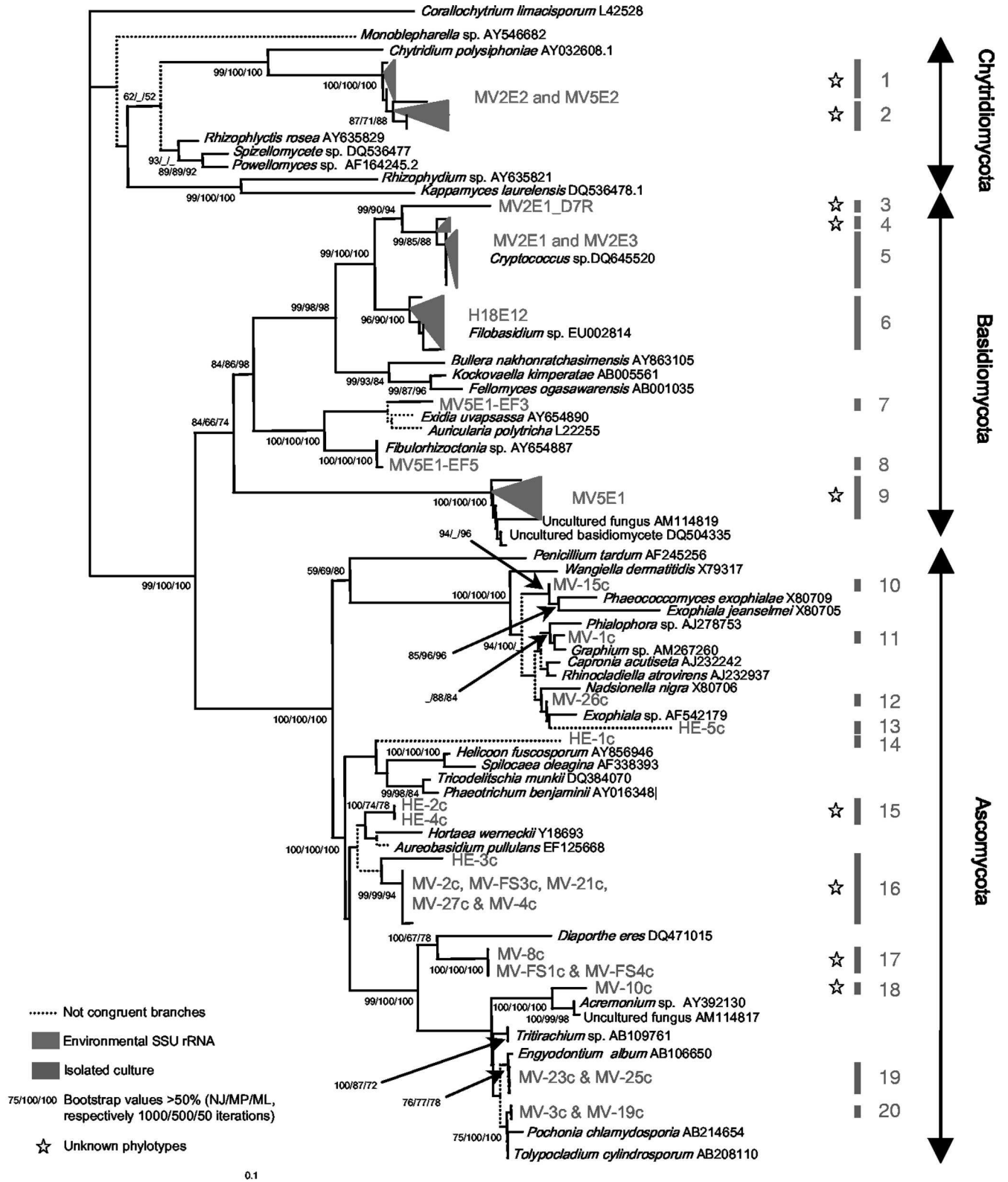


FIG. 1. Phylogenetic positions of deep-sea hydrothermal fungi. This consensus tree includes environmental SSU rRNA sequences isolated from environmental ecosystems (phylotypes 1 to 9) and isolated cultures from the same samples (phylotypes 10 to 20), along with the closest known related SSU rRNA fungal sequences. The tree was constructed using the neighbor-joining algorithm. Bootstrap values of >50% are indicated at the nodes (estimated using 1,000, 500, and 50 iterations for the neighbor-joining [NJ], maximum parsimony [MP], and maximum likelihood [ML] analyses, respectively). Stars indicate the phylotypes considered new phylotypes. The scale bar indicates 0.1 change per position computed using

TABLE 1. Samples analyzed in this study

Sample	Source	Location	Depth (m)	No. of phylotypes	No. of SSU rRNA sequences	No. of copies of rRNA per $\mu\text{g}$ of genomic DNA <sup>a</sup>
MV2E1	Exterior of <i>B. azoricus</i> shell	Menez Gwen, Mid-Atlantic Ridge	860	8	59	$1.91 \times 10^5$
MV2E2	<i>B. azoricus</i> (body)	Menez Gwen, Mid-Atlantic Ridge	860	5	47	$1.483 \times 10^6$
MV2E3	Surrounding friable rocks	Menez Gwen, Mid-Atlantic Ridge	860	6	25	$3.01 \times 10^6$
MV5E1	Sulfide of surrounding rocks	Lucky Strike, Mid-Atlantic Ridge	1,700	3	18	NA <sup>b</sup>
MV5E2	Exterior of <i>B. azoricus</i>	Lucky Strike, Mid-Atlantic Ridge	1,700	2	32	$1.35 \times 10^7$
H18E12	<i>A. pompejana</i> shell (epiderm)	Elsa, East Pacific Rise	2,630	2	49	NA

<sup>a</sup> Determined by qPCR.

<sup>b</sup> NA, not applicable.

S4 in the supplemental material) and *Basidiomycota* (Fig. 1, phylotypes 3, 4, and 9).

On the one hand, the *Basidiomycota* phylotypes detected belonged to the *Agaricomycotina* (Fig. 1, phylotypes 3 to 8), mostly to the genus *Cryptococcus*, a phylotype previously detected in another hydrothermal area (33), and the *Filobasidium* (East Pacific Rise) anamorphs, which have economic, agricultural, and medical importance. Phylotype 9 (Fig. 1), a close relative of uncultured fungi, was detected previously in deep-sea and hydrothermal sediments (2) and fluids (22). Moreover, homobasidiomycete yeasts belonging to the order *Auriculariales* (Fig. 1, phylotype 7) were also found. On the other hand, the *Chytridiomycota* phylotypes retrieved belonged to species that have not been described previously (Fig. 1, phylotypes 1 and 2; see Fig. S4 in the supplemental material) and, in particular, to an apparently ancient evolutionary lineage in the order *Chytridiales*. These sequences shared only 95.7% of similarity with their closest relative, *Chytridium polysiphoniae* (Fig. 1). The phylogenetic position of *C. polysiphoniae* is uncertain because it is a pathogenic fungus (19) and such fungi tend to show a long-branch attraction bias (Fig. 1). 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 Fig S4 in the supplemental material). It is very probable that the phylogenetic position of these newly discovered fungi was determined correctly because the branches in this group were short, a result that would not be expected if the terminal organisms were pathogens. This result is also supported by a Bayesian phylogenetic reconstruction (result not shown).

Different fungal communities were discovered when the fungal diversities of the Mid-Atlantic Ridge and the East Pacific Rise were compared; these two ridges are known to have distinct endemic animal communities (40) (*Bathymodiolus azoricus* and *Alvinella pompejana* samples, respectively). However, *Agaricomycotina* (Fig. 1, phylotypes 3 to 8) and *Dothideomycetes* (Fig. 1, phylotypes 14 to 16) sequences were found in

both ecosystems. *Agaromycotina* fungi seem to be widely distributed because they were also found in deep-sea sediments (11, 23) and methane seeps (33).

**Estimation of fungal SSU rRNA gene copy number in environmental samples.** qPCR analyses were performed to indirectly assess the fungal SSU rRNA gene copy numbers in environmental samples obtained from sites on the Mid-Atlantic Ridge and East Pacific Rise to test the possibility of contamination of the environmental samples by exogenous fungi. In these samples the numbers of fungal SSU rRNA gene copies would be expected to be low. Therefore, a set of specific fungal primers was designed on the basis of the diversity analyses, as described in Materials and Methods. Strikingly, the number of SSU rRNA gene copies ranged from  $1.91 \times 10^5$  to  $1.35 \times 10^7$  per  $\mu\text{g}$  of genomic DNA. Thus, this analysis clearly allowed us to reject the hypothesis that of all the fungi had an exogenous origin (i.e., that there was contamination of the samples by the water column or during sampling and conditioning). As the highest level of fungi was not observed in the bodies of animals (*B. azoricus* or *A. pompejana*) but on the exterior of the shell (*B. azoricus*) and the surrounding rock (Table 1), we hypothesized that most of these fungi were not biotrophic.

## DISCUSSION

This study reports the first inventory of fungal diversity in deep-sea hydrothermal environments. Unsuspected diversity, including new species in three fungal phyla, was found. One of the main results of the study described here is the evidence of an old *Chytridiomycota* lineage, unknown either at the genus level or at a higher taxonomic level (Fig. 1; see Fig. S4 in the supplemental material). As protein clock analyses (13) suggested previously that fungi emerged in oceans approximately 1 billion years ago during the Proterozoic era of the Precambrian, deep branches, such as this unknown phylotype, were expected based on our working hypotheses. It is thus possible

the NJ-K2P model. The dotted lines indicate branches not recovered in all three analyses. Designations that begin with MV followed by a hyphen indicate isolates from MARVEL cruise samples (Mid-Atlantic Ridge), and designations that begin with HE followed by a hyphen indicate isolates from "HERO" cruise samples (East Pacific Rise) (Table 1). The cultures obtained from sample MV2E1 are MV-1c to MV-4c, MV-21c, MV-23c, MV-FS1c, and MV-FS3c; the cultures obtained from sample MV2E2 are MV-8c, MV-10c, MV-25c, and MV-FS4c; the cultures obtained from sample MV2E3 are MV-15c, MV-19c, MV-26c, and MV-27c; the culture obtained from sample H18E9 is HE-5c; the cultures obtained from sample H18E11 are HE-1c to HE-3c; and the culture obtained from sample H18E12 is HE-4c. In the *Ascomycota*, the terminal HE-1c and HE-3c cultures were not defined as new phylotypes since their phylogenetic affinities were unclear.

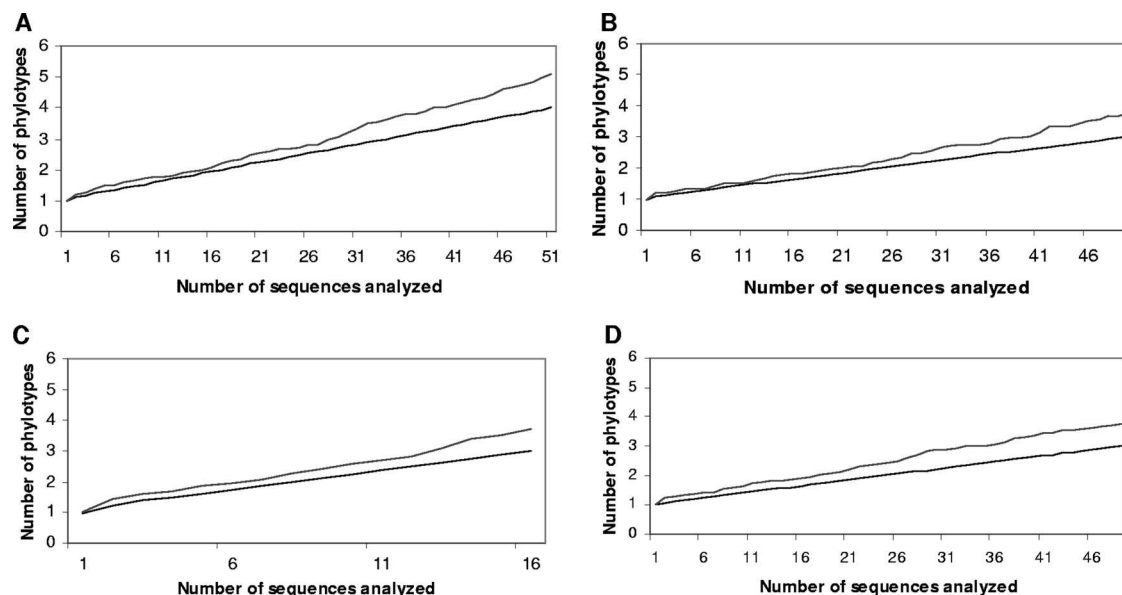


FIG. 2. Estimates of fungal community diversity as a function of sampling effort for environmental samples. Rarefaction curves were computed using 100 bootstrap replicates (upper line), and the expected richness function (lower line) is the number of phylotypes estimated (random sampling without replacement) from our sequence data set generated using 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 when sequence analyses of the ciPCR products obtained using primers MH2 and MH4 were performed.

that the emergence and initial diversification of fungi occurred in a marine environment. This study supports this hypothesis, but the data are not conclusive. The loss of flagella and therefore motility in fungal gametes, which likely occurred more than once (15), does not invalidate our conclusion since this loss could be compensated for 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 they colonized 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 three of the five fungal phyla, which supports our assumption even if the possibility of secondary colonizations from land to marine ecosystems cannot be excluded. 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 deeply branching *Chytridiomycota* (Fig. 1; see Fig. S4 in the supplemental material) are being addressed in an ongoing metagenomic analysis of the fungi living in deep-sea hydrothermal ecosystems.

**Fungal diversity in marine hydrothermal ecosystems.** Despite the effectiveness of the primer set used here for amplification of *Ascomycota* (37), we did not find any *Ascomycota* when ciPCR was used. Rarefaction curves resulting from resampling and pseudoresampling (bootstrap) procedures indicated that the number of sequences analyzed might have been insufficient to represent the entire fungal diversity in all samples. One or two additional phylotypes might be discovered (Fig. 2) by doubling the sequencing effort. We suggest that the abundance of *Ascomycota* in the environmental samples 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 physicochemical conditions (i.e., gradients of temperature, pH, etc.). Low redundancy of phylotypes was observed in 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 (12) are substantial underestimates. Given the high fungal SSU rRNA gene copy number obtained for a given sample (up to  $1.91 \times 10^5$  copies/ $\mu\text{g}$ ), it can also be suggested that the fungi might be involved in important ecological functions. Moreover, three of the phylotypes found in this work were detected by ciPCR in previous studies (2, 22). This also argues for rejection of the hypothesis that these sequences originated from contamination.

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

#### ACKNOWLEDGMENTS

We acknowledge Eugene Koonin, J. Peter W. Young, and Myriam Bormans for valuable comments on the manuscript and for discussions and suggestions. We thank Laetitia Guillot ("Ouest Génomole" Bioin-

formatics) for constructing the fungal database. We are also grateful to the "Ouest Génopole" sequencing service.

This work was supported by GIS Génomique Marine and by a grant from the Total Corporate Foundation for Biodiversity and the Sea.

#### REFERENCES

- Altschul, S. F., W. Gish, W. Miller, E. W. Myers, and D. J. Lipman. 1990. Basic local alignment search tool. *J. Mol. Biol.* **215**:403–410.
- Bass, D., A. Howe, N. Brown, H. Barton, M. Demidova, H. Michelle, L. Li, H. Sanders, S. C. C. Watkinson, S. Willcock, and T. A. A. Richards. 2007. Yeast forms dominate fungal diversity in the deep oceans. *Proc. R. Soc. B* **274**:3069–3077.
- Beatty, J. T., J. Overmann, M. T. Lince, A. K. Manske, A. S. Lang, R. E. Blankenship, C. L. Van Dover, T. A. Martinson, and E. G. Plumley. 2005. An obligately photosynthetic bacterial anaerobe from a deep-sea hydrothermal vent. *Proc. Natl. Acad. Sci. USA* **102**:9306–9310.
- Berbee, M. L., B. Kloekener, and J. W. Taylor. 1992. Yeast biology. *Science* **257**:1610–1611.
- Boussau, B., S. Blanquart, A. Necsula, N. Lartillot, and M. Gouy. 2008. Parallel adaptations to high temperatures in the Archaean eon. *Nature* **456**:942–945.
- Burgaud, G., T. Le Calvez, D. Arzur, P. Vandenkoornhuys, and G. Barbier. 2009. Diversity of culturable marine filamentous fungi from deep-sea hydrothermal vents. *Environ. Microbiol.* **11**:1588–1600.
- Damare, S., C. Raghukumar, and S. Raghukumar. 2006. Fungi in deep-sea sediments of the Central Indian Basin. *Deep-Sea Res. Part I* **53**:14–27.
- Dawson, R. C., and N. R. Pace. 2002. Novel kingdom-level eukaryotic diversity in anoxic environments. *Proc. Natl. Acad. Sci. USA* **99**:8324–8329.
- Edgcomb, V. P., D. T. Kysela, A. Teske, A. de Vera Gomez, and M. L. Sogin. 2002. Benthic eukaryotic diversity in the Guaymas Basin hydrothermal vent environment. *Proc. Natl. Acad. Sci. USA* **99**:7658–7662.
- Fisher, C. R., K. Takai, and N. Le Bris. 2007. Hydrothermal vent ecosystems. *Oceanography* **20**:14–23.
- Gadanho, M., and J. P. Sampaio. 2005. Occurrence and diversity of yeasts in the Mid-Atlantic Ridge hydrothermal fields near the Azores archipelago. *Microb. Ecol.* **50**:408–417.
- Hawksworth, D. L. 2001. The magnitude of fungal diversity: the 1.5 million species estimate revisited. *Mycol. Res.* **105**:1422–1432.
- Heckman, D. S., D. M. Geiser, B. R. Eidell, R. L. Stauffer, N. L. Kardos, and S. B. Hedges. 2001. Molecular evidence for the early colonization of land by fungi and plants. *Science* **293**:1129–1133.
- Huber, C., and G. Wächtershäuser. 1998. Peptides by activation of amino acids with CO on (Ni,Fe)S surfaces: implications for the origin of life. *Science* **281**:670–672.
- James, T. Y., F. Kauff, C. L. Schoch, P. B. Matheny, V. Hofstetter, C. J. Cox, G. Celio, C. Gueidan, E. Fraker, J. Miadlikowska, H. Thorsten Lumbsch, A. Rauhut, V. Reeb, A. E. Arnold, A. Amtoft, J. E. Stajich, K. Hosaka, G. H. Sung, D. Johnson, B. O'Rourke, M. Crockett, M. Binder, J. M. Curtis, J. C. Slot, Z. Wang, A. W. Wilson, A. Schübler, J. E. Longcore, K. O'Donnell, S. Mozley-Standridge, D. Porter, P. M. Letcher, M. J. Powell, J. W. Taylor, M. M. White, G. W. Griffith, D. R. Davies, R. A. Humber, J. B. Morton, J. Sugiyama, A. Y. Rossman, J. D. Rogers, D. H. Pfister, D. Hewitt, K. Hansen, S. Hambleton, R. A. Shoemaker, J. Kohlmeyer, B. Volkman-Kohlmeyer, R. A. Spotts, M. Serdani, P. W. Crous, K. W. Hughes, K. Matsuura, E. Langer, G. Langer, W. A. Untereiner, R. Lücking, B. Büdel, D. M. Geiser, A. Aptroot, P. Diederich, I. Schmitt, M. Schultz, R. Yahr, D. S. Hibbett, F. Lutzoni, D. J. McLaughlin, J. W. Spatafora, and R. Vilgalys. 2006. Reconstructing the early evolution of Fungi using a six-gene phylogeny. *Nature* **443**:818–822.
- Jørgensen, B. B., and A. Boetius. 2007. Feast and famine—microbial life in the deep-sea bed. *Nat. Rev. Microbiol.* **5**:770–781.
- Kohlmeyer, J., and E. Kohlmeyer. 1979. Marine mycology: the higher fungi. Academic Press, New York, NY.
- Koonin, E. V., and W. Martin. 2005. On the origin of genomes and cells within inorganic compartments. *Trends Genet.* **21**:647–654.
- Küpfer, F. C., I. Maier, D. T. Müller, S. Loiseaux-De Goer, and L. Guillou. 2006. Phylogenetic affinities of two eukaryotic pathogens of marine macroalgae, *Eurychasma dicksonii* (Wright) Magnus and *Chytridium polysiphoniae* Cohn. *Cryptogam. Algal.* **27**:165–184.
- Liu, Y. J., M. C. Hodson, and B. D. Hall. 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 Evol. Biol.* **6**:74–86.
- López-García, P., H. Philippe, F. Gail, and D. Moreira. 2003. Autochthonous eukaryotic diversity in hydrothermal sediment and experimental microcolonizers at the Mid-Atlantic Ridge. *Proc. Natl. Acad. Sci. USA* **100**:697–702.
- López-García, P., A. Vereshchaka, and D. Moreira. 2007. Eukaryotic diversity associated with carbonates and fluid-seawater interface in Lost City hydrothermal field. *Environ. Microbiol.* **9**:546–554.
- Nagahama, T., M. Hamamoto, T. Nakase, H. Takami, and K. Horikoshi. 2001. Distribution and identification of red yeasts in deep-sea environments around the northwest Pacific Ocean. *Antonie van Leeuwenhoek* **80**:101–110.
- Pirozynski, K. A., and D. W. Malloch. 1975. The origin of land plants: a matter of mycotrophism. *BioSystems* **6**:153–164.
- Posada, D., and K. Crandall. 1998. Modeltest: testing the model of DNA substitution. *Bioinformatics* **14**:817–818.
- Ramirez-Llodra, E., T. M. Shank, and C. R. German. 2007. Biodiversity and biogeography of hydrothermal vent species: thirty years of discovery and investigations. *Oceanography* **20**:30–41.
- Redecker, D., R. Kodner, and L. E. Graham. 2000. Glomalean fungi from the Ordovician. *Science* **289**:1920–1921.
- Robert, F., and M. Chaussidon. 2006. A paleotemperature curve for the Precambrian oceans based on silicon isotopes in cherts. *Nature* **443**:969–972.
- Selosse, D. A., and F. Le Tacon. 1998. The land flora: a phototroph-fungus partnership? *Trends Ecol. Evol.* **13**:15–20.
- Shearer, C. A., E. Descals, B. Kohlmeyer, J. Kohlmeyer, L. Marvanová, D. Padgett, D. Porter, H. Raja, J. Schmit, H. Thorton, and H. Voglymayr. 2007. Fungal biodiversity in aquatic habitats. *Biodivers. Conserv.* **16**:49–67.
- Simon, L., J. Bousquet, R. C. Lévesque, and M. Lalonde. 1993. Origin and diversification of endomycorrhizal fungi and coincidence with vascular land plants. *Nature* **363**:67–69.
- Smit, E., P. Leeftang, B. Glandorf, J. D. van Elsas, and K. Wernars. 1999. Analysis of fungal diversity in the wheat rhizosphere by sequencing of cloned PCR-amplified genes encoding 18S rRNA and temperature gradient gel electrophoresis. *Appl. Environ. Microbiol.* **65**:2614–2621.
- Takishita, K., M. Tsuchiya, J. D. Reimer, and T. Maruyama. 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.
- Taylor, J. W., and M. L. Berbee. 2006. Dating divergences in the fungal tree of life: review and new analyses. *Mycologia* **98**:838–849.
- Thompson, J. D., T. J. Gibson, F. Plewniak, F. Jeanmougin, and D. G. Higgins. 1997. The Clustal X Windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tool. *Nucleic Acids Res.* **24**:4876–4882.
- Tivey, M. K. 2007. Generation of seafloor hydrothermal vent fluids and associated mineral deposits. *Oceanography* **20**:50–65.
- Vandenkoornhuys, P., and C. Leyval. 1998. SSU rDNA sequencing and PCR-fingerprinting reveal genetic variation within *Glomus mossae*. *Mycologia* **90**:791–797.
- Vandenkoornhuys, P., S. L. Baldauf, J. Stracek, C. Leyval, and J. P. W. Young. 2002. Extensive fungal diversity in plant roots. *Science* **295**:2051.
- Wächtershäuser, G. 1990. Evolution of the first metabolic cycle. *Proc. Natl. Acad. Sci. USA* **87**:200–204.
- White, T. J., T. Bruns, S. Lee, and J. W. Taylor. 1990. Amplification and direct sequencing of fungal ribosomal RNA genes for phylogenetics, p. 315–322. *In* M. Innis, D. Gelfand, J. Sninsky, and T. White (ed.), *PCR protocols: a guide to methods and applications*. Academic Press, New York, NY.





# CHAPITRE 3

Ancient lineages of Chytridiomycota

& Opisthokonta revealed

in deep oceanic hydrothermal ecosystems



## Chapitre 3

---

### Ancient lineages of Chytridiomycota & Opisthokonta revealed in deep oceanic hydrothermal ecosystems

Les résultats obtenus lors de l'étude de la diversité du règne fongique au niveau de sources hydrothermales par Le Calvez *et al.* (2009) ont soulevé des interrogations vis-à-vis de la présence de lignées basales fongiques et mené à la formulation d'hypothèses sur l'histoire évolutive des champignons dans le milieu marin. Cela a lancé l'étude de la diversité des Chytridiomycota en milieu hydrothermal marin profond.

Ainsi des amorces spécifiques aux Chytridiomycota ont été dessinées pour décrire cette diversité par une approche de génomique environnementale et une analyse phylogénétique du marqueur moléculaire 18S rRNA. Six phylotypes de Chytridiomycota ont été obtenus dont quatre représentent des espèces non-décrites, ce qui reflète la faible connaissance de ce phylum. De manière inattendue, des phylotypes se positionnant en dehors du clade des Opisthokonta, voire au niveau des racines des Unikonta, ont également été détectés. Ces résultats montrant une diversité originale au niveau des sources hydrothermales profondes offrent de nombreuses perspectives pour la description de nouvelles espèces et la résolution des lignées profondes des Opisthokonta et des Unikonta.

## Ancient lineages of Chytridiomycota & Opisthokonta revealed in deep oceanic hydrothermal ecosystems

Stéphane Mahé and Philippe Vandenkoornhuys

### Summary

One of the way to apprehend biodiversity is to model the evolutionary relationships among all living organisms by constructing phylogenetic trees. Considering the Opisthokonta clade, the roots of the fungal kingdom may be resolved by increasing the molecular description of early diverging organisms to infer evolutionary relationships (Carr & Baldauf, 2011). Recently, an unsuspected diversity of fungi has been revealed in deep oceanic hydrothermal ecosystems leading to new evolutionary hypotheses about fungal emergence and diversification (Le Calvez *et al.*, 2009). As a result of that study we hypothesized that the Chytridiomycota phylum, one of the earliest fungal evolutionary lineages, are more diverse than previously believed. Primer sets were designed to specifically amplify the 18S rRNA genes by ciPCR (*i.e.* culture-independent PCR) and a variety of environmental samples from hydrothermal ecosystems were analysed. Positive amplifications were obtained from a subset of these samples. Although the sequence diversity in a given sample was low, the rarest sequences remained likely to be revealed. Phylogenetic analyses highlighted the presence of nine phylotypes (97% inside identity). All of these contained redundant sequences from at least two different amplicons. Six of them belonged to the Chytridiomycota, the largest proportion of which remains unknown and seems dependent on Oceanic environments (*e.g.* sediments). The three other phylotypes exhibited even deeper evolutionary branches. One formed an unknown lineage within Apusozoa and the other two formed deeper branches in the Unikonta and possibly Bikonta yet known. Evidence for the existence of an ancestral lineage of the Unikonta suggests an evolutionary paradigm different from the one currently accepted, to explain diversification within this evolutionary branch.

## 1. INTRODUCTION

The fungal kingdom forms a major eukaryotic lineage mainly known for its ecological function of decomposer, involved in the nutrient and carbon cycling vital to ecosystems (Wardle & Van der Putten 2002). Our current vision of the fungal kingdom has mainly arisen from studies of cultured isolates sampled in terrestrial ecosystems (Richards *et al.* 2012). However, the description of novel species of fungi is generally hindered by their minute size, poor morphological features in certain organisms, or because they are cryptic (*i.e.* distinct species with the same morphology) or uncultivable (Redecker 2002; Blackwell 2011). Moreover, basal lineages are less represented in culture collections and have therefore been less frequently studied than the upper fungi (Powell 1993; James *et al.* 2000; White *et al.* 2006). This lack of knowledge of fungal diversity has direct consequences on the testing of evolutionary hypotheses (Mueller & Schmit 2007). Hibbett *et al.* (2007) published a classification based on recent molecular phylogenetic analyses, which proposes taxa rearrangements at different taxonomic levels. In particular, the basal lineages previously clustered in the Chytridiomycota phylum have undergone significant revision. Indeed, the results suggest a Chytridiomycota *sensu stricto* phylum, separated from the newly validated phyla Neocallimastigomycota and Blastocladiomycota (James *et al.* 2006a, 2006b). The present fungal tree of life and taxonomical classification is based on 100,000 isolated and cultured species (Kirk *et al.* 2008), which offers a narrow view of global fungal diversity, traditionally estimated at between 712,000 (Schmit & Mueller 2007) and 1,500,000 (Hawksworth 1991, 2001). One molecular data analysis has indicated 3.5-5.1 million species, showing an even greater gap in current knowledge (O'Brien *et al.* 2005). Owing to the increasing number of molecular studies, several scientists request the integration of environmental molecular OTUs (Operational Taxon Units) into fungal phylogenetic classification (*e.g.*, Hibbett *et al.* 2011; Kruger *et al.* 2012). This constitutes a step to convert taxonomy into systematics, the aim being to include and clarify the existing evolutionary relationships among organisms (Carr & Baldauf 2011).

Environmental molecular studies have already shown the existence of novel fungal clades in terrestrial and marine ecosystems (*e.g.*, Vandenkoornhuysen *et al.* 2002; Bass *et al.* 2007; Schadt *et al.* 2003; Porter *et al.* 2008; Le Calvez *et al.* 2009; Nagano *et al.* 2010). These newly detected clades are initiating a new era in the understanding of fungal evolution and diversification. Extension of the classification system to environmental sequences should facilitate communication concerning fungal diversity studies and what follows from them (Hibbett *et al.*, 2011). This systematics approach made it possible to situate the environmental sequences of 'Soil clone group 1' (Porter *et al.* 2008) in a novel class Archaeorhizomycetes in the Ascomycota phylum (Rosling *et al.* 2011). Recently, and of even greater interest, the Cryptomycota phyl. nov. (Jones *et al.* 2011b) has been described. It encompasses the genus *Rozella* (Held 1981) and the ubiquitous environmental 'LKM11' clade, previously named as such from the first sequence retrieved (Van Hanneken 1999). The cells containing environmental sequences from this group have been visualised using fluorescent detection via DNA probes but have not yet been cultured (Jones *et al.* 2011a). These Cryptomycota form the deepest described fungal lineage, which likely appeared before Chytridiomycota, thus modifying our understanding of early fungal radiation and perception of the fungal tree of life (Jones *et al.* 2011a).

Apart from the question of early fungal evolution, there is of course the question of fungal diversity of ancient lineages. Chytridiomycota have been detected in the marine environment by using molecular tools (Le Calvez *et al.*, 2009). Indeed, only three species have been formally described in marine and brackish ecosystems *Rhizophidium littoreum*, *Thalassochytrium gracilariopsis* and *Chytridium polysiphonae* (Gleason *et al.* 2011). Richards *et al.* (2012) recorded 18S rRNA gene sequences of fungi from different studies of marine ecosystems and examined their phylogenetic affinities with sequences from cultured fungi. Among 36 novel marine lineages highlighted, (*i.e.*, identity to cultured fungi < 97%), 2/3 belonged to basal lineages, mainly Chytridiomycota, and constituted the most divergent branches. The detection of Chytridiomycota started with molecular eukaryotic inventories in different marine ecosystems such as intertidal sediments (Dawson & Pace 2002), open sea surface water (Massana *et al.* 2004), oxygen-depleted seawater from an intertidal pool (Stoeck & Epstein 2003), anoxic seawater in a fjord (Zuendorf *et*

*al.* 2006) and in the open sea (Stoeck *et al.* 2006; Edgcomb *et al.* 2011). However, in general, the selected primers only describe a subset of the diversity and sampling saturation is not attained (Edgcomb *et al.*, 2002 ; Stoeck *et al.*, 2006 ; Richards & Bass, 2005). Recently, fungi-specific studies have been carried out in deep-sea ecosystems. Several sequences signatures of Chytridiomycota were reported such as anaerobic bacterial mats (1575m depth, Bass *et al.* 2007), in association with *Bathymodiolus azoricus* in hydrothermal vents (860m and 1,700m depth, Le Calvez *et al.* 2009) and methane cold seep sediments (2,400m depth, Thaler *et al.* 2012). The 18S rRNA gene sequences amplified (Bass *et al.* 2007 ; Le Calvez *et al.* 2009) did not show any matches with described fungi and can therefore be identified as new marine fungi (Richards *et al.* 2012). Concerns have been voiced about the primers used and the DNA targets chosen to describe fungal diversity (*i.e.*, bias in amplification due to primer specificity for a subset of template DNA; lack of a quality phylogenetic signal when using *ITS* (*i.e.*, Internal Transcribed Spacer) as targeted DNA), suggesting that fungal diversity, including Chytridiomycota, in marine environments is likely much greater than currently demonstrated (Nagano *et al.* 2010 ; Richards *et al.* 2012).

Le Clavez *et al.* (2009) suggested a marine emergence and diversification of fungi in oceans. We therefore hypothesized that i) Chytridiomycota are more diverse than previously believed and ii) analysing this important ancient fungal group would lead to a better understanding of fungal and Opisthokonts (*i.e.* animal and fungi) origin and diversification. With this aim, we selected and designed a primer set specific to Chytridiomycota 18S rRNA using PHMYCO-DB, a database dedicated to fungal sequences for phylogenetic analyses (Mahé *et al.*, submitted). We analysed samples collected at four deep-sea hydrothermal vents on the East Pacific Rise and the Mid-Atlantic Ridge. We chose this set of samples because (i) Chytridiomycota had been found in this environment (Le Calvez *et al.*, 2009) (ii) these hydrothermal ecosystems are oases of life on the ocean floor, (iii) hydrothermal conditions were mundane during the Precambrian (Robert & Chaussidon, 2006) and also because this environment is considered as a possible cradle of life (Wächtershäuser, 1990; Huber & Wächtershäuser, 1998). The diversity detected in the present study revealed a new phylotypes in the Chytridiomycota phylum and deeply-positioned lineages forming an ancestor group in the Opisthokonta or forming a sister clade to the Opisthokonta.



## 2. MATERIAL AND METHODS

### (a) Sampling and DNA extraction

Samples were collected from hydrothermal vents on the East Pacific Rise in the Guaymas basin (27°01' N, 111°24'W) and on the Mid-Atlantic Ridge at Lucky Strike (37°17'N, 32°16'W), Rainbow (36°06'N, 33°11'W) and the Trans-Atlantic Geotraverse (T.A.G) (26°08'N, 44°49'W). The methods used to collect the samples are described in Le Calvez *et al.* (2009). The nature and location of the samples are presented in Table 1. Prior to chemical cell lysis, samples frozen in liquid nitrogen were ground with a bead beater (MM 301; Retsch) using tungsten beads (3mm diameter) with two shaking steps for 2mn at 30Hz. DNA was extracted using the DNeasy Plant mini kit (Qiagen) according to the manufacturer's recommendations.

Table 1: Samples processed.

Samples	Location	Depth (m)	Nature
GY12T4(5)	Guaymas	2000	Sediment core (20-25cm)
MV5E2	Lucky Strike	1700	Mussel external shell side
EX6E5	Rainbow	2300	<i>Chorocaris chacei</i>
EX8E6	Rainbow	2300	<i>Rimicaris exoculata</i>
EX9E1	Rainbow	2300	Chimney fragment (sulfide)
EX9E5	Rainbow	2300	<i>Bathymodiolus azoricus</i>
EX9E8	Rainbow	2300	<i>Bathymodiolus azoricus</i>
EX9E9	Rainbow	2300	<i>Bathymodiolus azoricus</i>
EXA5E3	T.A.G	3645	<i>Rimicaris exoculata</i>

### (b) Design of primers and environmental clone libraries

Primers were selected to specifically amplify Chytridiomycota 18S rRNA gene sequences with a size compatible for 454 mass sequencing, *i.e.* ranging from 450 bp to 600 bp amplicons. Chytridiomycota 18S rRNA gene sequences were downloaded from PHYMYCO-DB to test several

primer sets. The choice of amplified fragment was also oriented to encompass the V4 area (Wuyts *et al.* 2000). We chose MH2 (TTCGATGGTAGGATAGAGG) (Le Calvez *et al.* 2009) and SR8RChytrid (TCAAAGTAAAAGTCCTGGATC) modified from SR8R published on the Vilgalys lab webpage (<http://www.biology.duke.edu/fungi/mycolab/primers.htm>) to target Chytridiomycota. Coverage of Chytridiomycota molecular diversity was maximised by using two degenerate versions of SR8RChytrid, that we called SR82RChytrid (TCAAAGTAAAAGTCCTGGTTC') and SR83RChytrid (TCAAAGTAACAGTCCTAGATC). DNA was amplified by touchdown PCR. The overall PCR yield was low and the amplification intensity was not constant across samples. FailSafe™ PCR PreMixes (Epicentre® Biotechnologies) gave similar results. Tests with bacteria-specific primers showed that the PCR reactions were not impeded by inhibitors and the DNA was not degraded, *i.e.* the amplification signals were very strong. We assumed from these PCR reactions that the DNA matrix targets were very scarce, and were diluted by the other DNA matrices. So the small amount of amplicons was not sequenced by 454. Instead we used a classical cloning and sequencing strategy. Amplicons were purified with magnetic beads (Agencourt AMPure). They were then cloned in the pGEM-T vector (Promega) and DH5α competent cells (Gibco BRL Life Technologies), according to the manufacturer's recommendations.

### **(c) Phylogenetic analyses**

A multiple-sequence alignment was performed with CLUSTAL X 2.1 (Larkin *et al.* 2007) and then refined by eye. Environmental sequences were appended to a Chytridiomycota-aligned matrix downloaded from PHYMYCO-DB (Mahé *et al.*, submitted). The environmental sequences which grouped together were then checked for their percentage identity. When the nucleotide character difference was lower than 3%, sequences were considered belonging to a same single OTU for a given clone library. OTUs were then locally aligned against GenBank (Benson *et al.* 2011) using BLASTn to select the best BLAST hit sequences. The final matrix consisted of the unique OTUs for each clone library along with their best BLAST hit sequences or their closest neighbours in the Chytridiomycota matrix (PHYMYCO-DB), and sequences chosen according to their positioning in

the Opisthokonta clade. Finally, phylogenetic analyses were carried out as follows. A neighbor-joining phylogeny using Kimura 2 parameters as model distance was calculated using CLUSTAL X 2.1 with 1,000 bootstrap iterations (Larkin *et al.* 2007). To perform the maximum likelihood (ML) analysis, we selected the best possible model with jModelTest (Posada & Crandall 1998). After this step, the ML phylogeny was computed using the GTR+I+G model with 1,000 bootstrap iterations using PHYML (Guindon *et al.* 2003). A Bayesian phylogenetic analysis was performed using MrBayes 3.2.1 (Ronquist & Huelsenbeck 2003) and the same GTR+I+G model, with 2,000,000 generation samples and a sample frequency of 100 generations.

### 3. RESULTS AND DISCUSSION

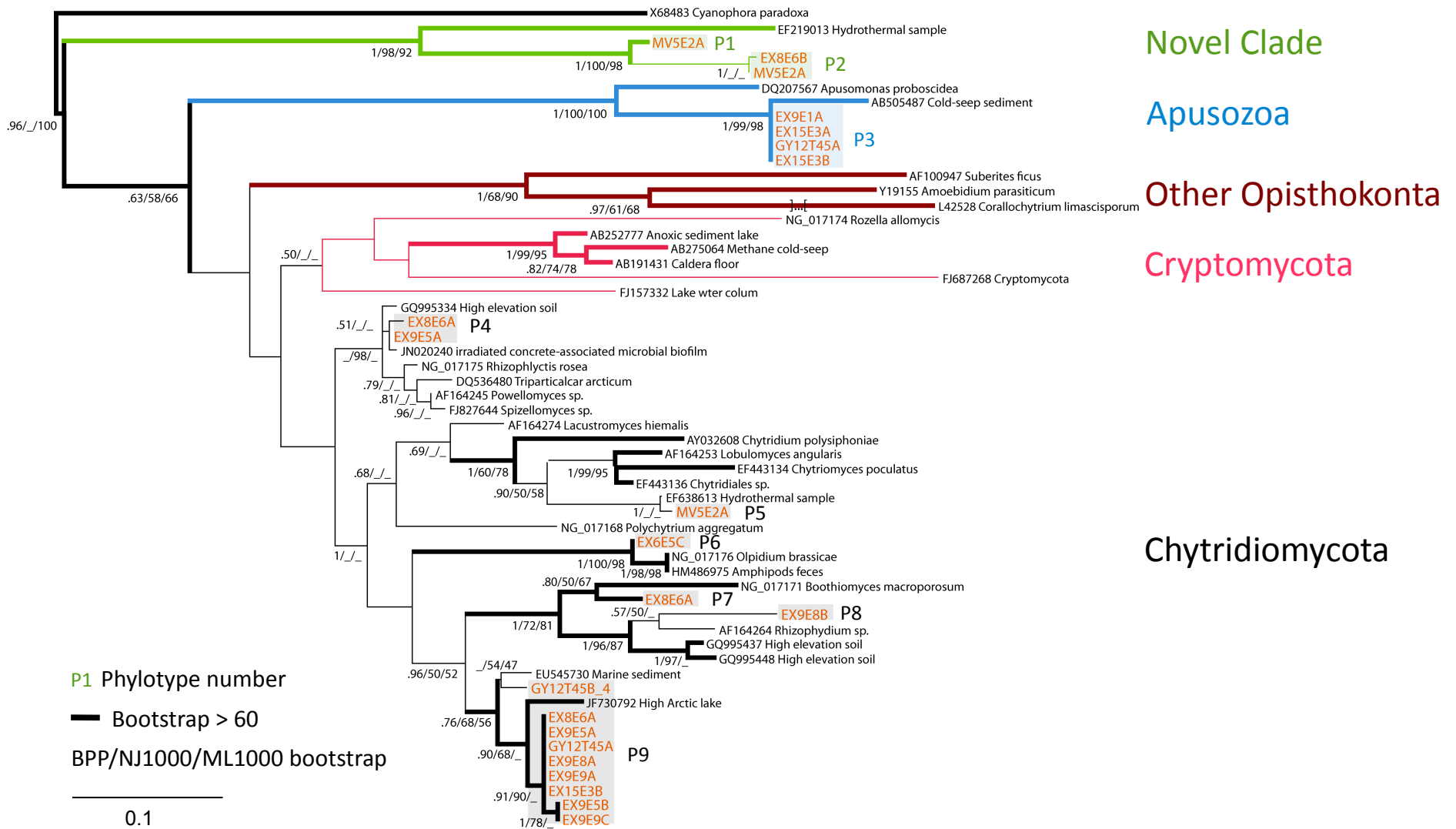
#### (a) *Data mining*

A range of environmental samples from hydrothermal vents was tested. Nine samples were selected, each with two replicates of the 18rRNA gene amplification. Clone libraries were constructed and 226 clones were sequenced. Within these 226 clones, 136 belonged to fungi with 130 Chytridiomycota, four Ascomycota and two Basidiomycota, 52 corresponded to sequences of Apusozoa and 38 to sequences of an undefined group of Eukaryota. As the diversity covered a larger proportion of the Eukaryota domain, than expected, we had to choose an outgroup among the Bikonta. *Cyanophora paradoxa* (Glaucophyta) is a good candidate as it exhibits close relationships with the Unikonta (Cavalier-Smith 2004). It fulfilled the incontrovertible condition of aligning with our matrix, which might appear trivial but the first part of the V4 area is difficult to align as the initial helixes are not present in all Eukaryota (Wuyts *et al.* 2000) and compatibility must be tested. The three methods of reconstruction produced congruent topologies. The phylogenetic analyses did not recover strong bootstrap support or Bayesian posterior probabilities for the internal nodes. Nevertheless, the deep branches of the fungal kingdom were coherent with other studies of 18S rRNA gene sequences (Cavalier-Smith & Chao 1995; Moreira *et al.* 2007, Bass *et al.* 2007; Richards *et al.* 2012). Higher bootstrap supports for these deep branches can only be obtained when several molecular markers are combined (*e.g.* James *et al.* 2006a).

**(b) Fungal diversity sampled**

The selected primer set successfully amplified the deep fungal branches, higher fungal sequences being rare among the clones (*i.e.* fewer than 3%). To focus on the roots of the fungal kingdom, OTUs belonging to Ascomycota and Basidiomycota were not included in the phylogenetic analyses. These were only locally aligned using BLASTn. One OTU was found in samples EX9E5 and EX9E8 (*Cf.* Table 1) and showed high percentage identity (> 99%) with *Malassezia* yeast (Basidiomycota) which has been characterized as a 'hydrothermal and/or anaerobic fungal group' (López-García *et al.* 2007). Indeed, sequences of this group are often retrieved from such ecosystems, *e.g.* hydrothermal fluid-seawater mix in Lost City (López-García *et al.* 2007), anoxic sediments in the Arabian sea (Jebaraj *et al.* 2010), hydrothermal chimney fragment (sulphide) in Lucky Strike (Le Calvez *et al.* 2009). However, *Malassezia* yeasts are also considered to be continental species. A second OTU was found in samples EX6E5 and EX9E9 and could be assigned to the Pezizomycotina class (Ascomycota).

The phylogenetic analyses revealed nine phylotypes in the hydrothermal vents studied but the actual number may be larger. Indeed, the diversity of OTUs found in the different samples was low, ranging from 1 to 4 per sample but the rarest 18S rRNA gene sequences would not be caught by the strategy used here and the diversity observed in amplicon replicates was not strictly identical. For example, for sample EX8E6, 25 clones clustered in three phylotypes in the first replicate and 8 clones clustered in one phylotype, different from those in the first replicate, in the second one. This suggests that some samples may harbour a larger number of OTUs. We did not detect any particular assemblages related to the nature of the sample.



**Figure 1. Phylogeny demonstrating the phylogenetic positions of environmental sequences from samples collected at hydrothermal vents on the East Pacific Rise and Mid-Atlantic Ridge.** The topology shown here, was obtained by Maximum Likelihood (ML) analysis (GTR+I+G) and it's congruent with Neighbor-Joining (NJ) and Bayesian analyses. Bootstrap values are written for ML and NJ analyses calculated over 1,000 iterations and Bayesian posterior probabilities (BPP). Thickened branches present at least two of these statistic values: BPP>0.60, NJ>60% or ML>60%. OTUs in orange were retrieved in this study. P1 to P9 designate environmental phylotypes recovered in this study. Coloured branches referred to clade as follows: black, Chytridiomycota; pink, Cryptomycota; brown, other Opisthokonta; blue, Apusozoa and green, novel clade.

Most of the phylotypes identified belonged to the Chytridiomycota phylum (figure 1). The obtained phylogenetic relationships demonstrate the paraphyly of this clade which is congruent with another phylogeny based on all the rRNA genes (James *et al.* 2006b). Six phylotypes were recovered from three different orders and *Olpidium*. Phylotype 4 is closely related to *Rhizophlyctis rosea* which belongs to the order Rhizophlyctidiales, a group reported from soils and known as cellulose degraders (Letcher *et al.* 2008). Identical environmental sequences have been found in high elevation soils (Freeman *et al.* 2009) and in an irradiated concrete-associated microbial biofilm (Ragon *et al.* 2011). Phylotype 6 is also close to a described genus : *Olpidium*. One of its species, *Olpidium rostriferum*, has been described as a parasite of marine green algae and diatoms, but other species in this genus are known as parasites of plant roots and invertebrates in soils and freshwater (Gleason *et al.* 2011). The molecular phylogenetic classification of *Olpidium* in the Chytridiomycota phylum is debated (James *et al.* 2006a; Hibbett *et al.* 2007) because as a parasite, it has a high rate of substitution producing long branches which can lead to erroneous positioning (Philippe & Laurent 1998). The four other chytrid phylotypes correspond to environmental phylotypes that diverge from described species. Phylotype 5 belongs to the order Lobulomycetes whose species are reported as parasites and saprophytes in soils, marine and freshwater environments (Simmons *et al.* 2009). It is closely related to an environmental sequence associated with the body of *Bathymodiolus azoricus* in the hydrothermal site of Menez Gwen (MAR) (Le Calvez *et al.* 2009) whereas here, it comes from a sample recovered by scratching the external surface outside of *B. azoricus* shell. This order also contains one of the three marine species described as Chytridiomycota, *i.e.*, *Chytridium polysiphonae* (Küpper *et al.* 2006), a parasite of brown algae. The two other phylotypes form environmental clades which are unrelated to described species, but statistically supported as belonging to the Rhizophydiales order. Species from this order are reported as parasites of algae and as pollen decomposers in soils (Letcher *et al.* 2006). One of them is *Rhizophyidium littoreum*, a marine facultative parasite of algae and/or saprophyte (Gleason *et al.* 2011), but none of our sequences was closely related to it. Indeed, phylotype 7 appears to be a novel environmental clade, sharing only 95% identity with the closest

sequence. Phylotype 6 is related to sequences from lakes (AY649695, LeFranc *et al.* 2005; DQ244005, Lefèvre *et al.* 2007). In lakes, Chytridiomycota can have a great impact on phytoplanktonic populations via parasitism (Ibelings *et al.* 2004; Kagami *et al.* 2007) and also significantly participate in organic matter cycling in a complementary manner to the microbial loop (Lefèvre *et al.* 2008). Finally, phylotype 9 is recovered from six of the nine samples. Thus it seems to be ubiquitous in hydrothermal vents as its sequences were found at one site on the East Pacific Rise and at several sites on the Mid-Atlantic Ridge. In five samples, these sequences were associated with animals, *i.e.* deep-sea shrimps (*Rimicaris exoculata*) or deep-sea mussels (*Bathymodiolus rimicaris*) and in one sample, they were found in sediments. The only close relatives are environmental sequences from marine sediments from the sea of Japan (Park *et al.* 2008) and a high Arctic lake (JF730792). The closest cultured and described species belong to the Rhizophlyctidiales and Spizellomycetales orders with only 93% identity. All these OTUs retrieved from hydrothermal vents likely present characteristics and ecological statuses associated with the Chytridiomycota taxonomical classification. However, to our knowledge Chytridiomycota are not reported as parasites of crustaceans or molluscs. So further studies are needed to determine the exact ecological status of the Chytridiomycota found associated with animals in hydrothermal deep-sea vents.

### **(c) Other Eukaryota diversity**

The positioning of Cryptomycota and the other Opisthokonta clades was not supported by robust bootstrap values or Bayesian posterior probabilities (figure 1). Additional phylogenetic markers might help to improve bootstrap support. At least, the branching order is consistent with other studies (Cavalier-Smith & Chao 1995; Moreira *et al.* 2007; Richards *et al.* 2012). We retrieved new sequences branching at the base of the whole Opisthokonta group, with *Apusomonas proboscidea* as closest known relative. The bootstrap support here is not high but remain acceptable. The apusozoan phylotype seems to be ubiquitous as its sequences were found on the Mid-Atlantic Ridge at the Rainbow and T.A.G. sites and on the East Pacific Rise, associated with animals,

sediments or mineral substrates. It also contains an environmental sequence recovered from microbial mats of deep-sea cold-seep sediments (Takishita *et al.* 2010). Apusozoa are gliding heterotrophic unicellular organisms with two flagella that live freely and feed by phagocytosis on bacteria (Adl *et al.* 2005; Cavalier-Smith & Chao 2010). They are observed in soils, and in marine and freshwater environments (Cavalier-Smith & Chao 2010). Apusozoa are essential to understand the Eukaryota diversification and the divergence between Unikonta and Bikonta (Moreira *et al.* 2007). The positioning of this phylum is ambiguous and depends on the molecular markers used. Several studies using different molecular markers place it with no statistical support between Unikonta and Bikonta (*e.g.* Stechmann & Cavalier-Smith 2003; Moreira *et al.* 2007). One analysis based on 18S rRNA (Cavalier-Smith & Chao 1995) and an analysis based on six molecular markers (Kim *et al.* 2006), recovered strong support for the hypothesis of a clade constituted by Apusozoa and Opisthokonta with Amoebozoa as sister clade.

An early diverging lineage composed of two phylotypes (phylotypes 1 & 2) branched before the Opisthokonta and Apusozoa clades. The phylogenetic position was validated both by very high bootstrap supports and Bayesian posterior probabilities at the nodes. These phylotypes were identified from two different samples. Phylotype 2 was represented by eight clones in sample EX8E6 and 16 clones in sample MV5E2 and phylotype 1 by 14 clones in sample MV5E2. Thus, they should not be chimeric sequences. It was noted that they did not produce long branches. Thus an erroneous position due to long branch attraction bias was unlikely. These sequences are found in animal samples from two sites on the Mid-Atlantic Ridge. The existence of such a lineage is intriguing as no closely related sequences are represented in the GenBank database. The sequence with the best BLAST hit score (87% identity), which is also the closest phylogenetic neighbour (*Cf.* figure 1), corresponds to a sequence identified from a ridge flank abyssal hill at a hydrothermal site on the East Pacific Rise (Takishita *et al.* 2010). This particular sequence BLAST with diverse Chytridiomycota and some *incertae sedis* early branching fungi from the former Zygomycota phylum (Hibbett *et al.* 2007) but the identity scores were low (90-91%). The presence of this ancient lineage therefore suggests that diversification within the Unikonta is the result of a more



complex evolutionary history. Hydrothermalism was supposed to be mundane during the precambrian and earlier geological periods (Robert & Chaussidon, 2005). Thus the fact that these sequences come from a hydrothermal environment is compatible with a long evolutionary history.

More analyses are required to better understand phylotypes 1, 2 and 3 (figure 1). A genome analysis should provide important clues and facilitate comprehension of these unknown organisms from an evolutionary and functional perspective and also permit more accurate positioning within the Eukaryota domain.

#### **(d) Primer set relevance and data analysis**

This study reports the first molecular inventory dedicated to fungal basal lineages in marine ecosystems. Previous molecular studies of marine fungi focused on the entire fungal kingdom and Chytridiomycota sequences represented only a small subset of the described fungal diversity (Bass *et al.* 2007; Le Calvez *et al.* 2009). These studies indicated that the environmental OTUs retrieved from hydrothermal samples belonged mainly to Ascomycota and Basidiomycota. Here, the use of a Chytridiomycota-specific primer set provided successful access to the diversity of the basal fungal lineages and also, unexpectedly, amplified the diversity of organisms branching before the animal-fungal divergence. Moreover, amplification of Ascomycota and Basidiomycota sequences was limited. The diversity described by this set of primers is original as it does not focus on a taxonomic clade but investigates deep unresolved branches.

Due to the secondary structure of the V4 area (Wuyts *et al.* 2000), this primer can be used to study closely related species unless these organisms produce long branches. Indeed, no sequences from the newly described phylum Cryptomycota were retrieved here. We aligned MH2 with the Cryptomycota environmental sequences listed in Richards *et al.* 2011, and less than 1/3 presented at least one mismatch, the other aligning perfectly. On the other hand, most of these sequences diverge from the three versions modified from SR8RChytrid, the closest version being SR82RChytrid. For example, three of the Cryptomycota sequences used in our phylogeny present

one or two mismatches and the one producing a long branch diverges by three substitutions and one deletion. Thus, we cannot conclude whether Cryptomycota are present at these hydrothermal sites or not. For the phylogenetic analyses, some clades cannot be used as they do not align with Chytridiomycota sequences on the V4 region. For example, *Nuclearia simplex* (Nucleariidae) seems to have an additional helix in this V4 region, *Dictyostelium discoideum* and *Entamoeba nuttalli* (Amoebozoa) are highly divergent and Microsporidia do not possess this V4 region (Wuyts *et al.* 2000). Despite these facts, and as shown herein, this primer set is able to amplify sequences that branch deeply .

#### **(e) Marine fungi diversification**

The Chytridiomycota diversity retrieved here reflects different possible evolutionary histories concerning the relationships between terrestrial and marine species. The diversity observed in hydrothermal vents is in accordance with the great capacity of the Chytridiomycota to adapt to habitat conditions considered anthropologically as stressful or extreme (Gleason *et al.* 2010). Only two OTUs (*i.e.*, phylotypes 4 and 6, figure 1) are very closely related to described terrestrial species with short branches which would favour the hypothesis that marine-terrestrial transitions in both directions occurred (Bass *et al.* 2007; Richards *et al.* 2012). Two other OTUs (*i.e.*, phylotypes 7 and 9, figure 1), closely linked to environmental sequences found in freshwater environments, also support this hypothesis. They are also symptomatic of the lack of description of basal fungi as there is no close described sequence. And finally, two OTUs (*i.e.*, phylotypes 5 and 8, figure 1) present distant relationships with described species and terrestrial environmental sequences. We make the hypothesis that in addition to these marine-terrestrial transitions, the dispersion of species have led to long time isolation which results in speciation events. For the moment, not enough data are available to determine whether the emergence of fungi occurred in marine environment (Le Calvez *et al.* 2009) or in freshwater (Richards *et al.* 2012). The only accepted hypothesis is that they emerged in aquatic environments because the early diverging fungi display a zoosporic mobile stage and lack aerial spore dispersal (James *et al.* 2006a).

The parasitic behaviour of Chytridiomycota is likely overestimated due to an anthropocentric perception. It is clear from the literature that Chytridiomycota play an important role in nutrient cycling and they might also perform this function in deep-sea hydrothermal vents. These are known to harbour a species-rich microbial community (Fischer *et al.* 2007) performing chemoautotrophy, *i.e.*, microbes use reduced compounds from the vent fluids to fix inorganic carbon from the seawater (Ramirez-Llodra *et al.* 2007). Thus, organic matters available should be in sufficient quantity to sustain populations of saprophytes. Knowledge of nutrient cycling in these ecosystems is scanty but we can speculate that fungi could be important in this process.

#### **4. CONCLUSION**

This primer set allowed an initial description of the evolutionary roots of the fungal kingdom in hydrothermal vents and marine environments. An interesting diversity of Chytridiomycota is reported, attesting for the great capacity of adaptation of the species in this group and reinforcing the hypothesis of several marine-terrestrial transitions during their evolutionary history, along with evidence for events of marine diversification. Other Eukaryotic sequences were retrieved, revealing the ubiquitous presence of Apusozoa. Finally, a novel lineage was recovered which will require further study to determine its taxonomical classification and biological and physiological characteristics.

#### **ACKNOWLEDGEMENTS**

We thank Alix Mas and Elodie Maluenda for their help for the construction of the clone libraries.

#### **REFERENCES**

Adl SM, Simpson AGB, Farmer MA, Andersen RA, Anderson OR, et al. (2005) The new higher level classification of eukaryotes with emphasis on the taxonomy of protists. *The Journal of Eukaryotic Microbiology* 52: 399–451.

- Bass D, Howe A, Brown N, Barton H, Demidova M, et al. (2007) Yeast forms dominate fungal diversity in the deep oceans. *Proceedings of the Royal Society B: Biological sciences* 274: 3069–3077.
- Benson DA, Karsch-Mizrachi I, Lipman DJ, Ostell J, Sayers EW (2011) GenBank. *Nucleic Acids Research* 39: D32-D37.
- Blackwell M (2011) The Fungi: 1, 2, 3 ... 5.1 million species? *American Journal of Botany* 98: 426–438.
- Carr M & Baldauf SL (2011) The protistan origins of animals and fungi in *The Mycota: Evolution of Fungi and fungal-like organisms XIV*. Pöggeler S & Wöstemeyer J, eds. Springer Heidelberg Dordrecht London New York.
- Cavalier-smith T & Chao E (1995) The opalozoan Apusomonas is related to the common ancestor of animals, fungi, and choanoflagellates. *Proceedings of the Royal Society B: Biological sciences* 261: 1–6.
- Cavalier-Smith T & Chao EE (2010) Phylogeny and evolution of Apusomonadida (Protozoa: Apusozoa): new genera and species. *Protist* 161: 549–576.
- Dawson SC & Pace NR (2002) Novel kingdom-level eukaryotic diversity in anoxic environments. *Proceedings of the National Academy of Sciences U.S.A.* 99: 8324–8329.
- Edgcomb VP, Kysela DT, Teske A, de Vera Gomez A & Sogin ML (2002) Benthic eukaryotic diversity in the Guaymas Basin hydrothermal vent environment. *Proceedings of the National Academy of Sciences of the United States of America* 99: 7658–7662.
- Edgcomb V, Orsi W, Bunge J, Jeon S, Christen R, et al. (2011a) Protistan microbial observatory in the Cariaco Basin, Caribbean. I. Pyrosequencing vs Sanger insights into species richness. *The ISME journal* 5: 1344–1356.
- Fisher CR, Takai K, Le Bris N (2007) Hydrothermal vent ecosystems. *Oceanography* 20: 14-23.
- Freeman KR, Martin AP, Karki D, Lynch RC, Mitter MS, et al. (2009) Evidence that chytrids dominate fungal communities in high-elevation soils. *Proceedings of the National Academy of Sciences of the United States of America*. 106: 18315-18320.
- Gleason FH, Schmidt SK & Marano AV (2010) Can zoosporic true fungi grow or survive in extreme

or stressful environments? *Extremophiles* 14:417–425.

Gleason FH, Küpper FC, Amon JP, Picard K, Gachon CMM et al. (2011) Zoosporic true fungi in marine ecosystems : a review. *Marine and Freshwater research* 62: 383–393.

Guindon S & Gascuel O (2003) A Simple, Fast, and Accurate Algorithm to Estimate Large Phylogenies by Maximum Likelihood. *Systematic Biology* 52:696–704.

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.

van Hannen EJ, Mooij W, Agterveld MPV, Gons HJ & Laanbroek HJ (1999) Detritus-Dependent Development of the Microbial Community in an Experimental System : Qualitative Analysis by Denaturing Gradient Gel Electrophoresis. *Applied and environmental microbiology* 65: 2478–2484

Hawksworth DL (1991) The fungal dimension of biodiversity: Magnitude, significance, and conservation. *Mycological Research* 95: 641-655.

Hawksworth DL (2001) The magnitude of fungal diversity: the 1,5 million species estimate revisited. *Mycological Research* 105: 1422–1432.

Held AA (1981) *Rozella rozellopsis*: Naked endo-parasitic fungi which dress-up as their hosts. *The Botanical Review* 47: 451–515.

Hibbett DS, Binder M, Bischoff JF, Blackwell M, Cannon PF et al. (2007) A higher-level phylogenetic classification of the Fungi. *Mycological research* 111: 509–547.

Hibbett DS., Ohman A, Glotzer D, Nuhn M, Kirk P & Nilsson RH (2011) Progress in molecular and morphological taxon discovery in Fungi and options for formal classification of environmental sequences. *Fungal Biology Reviews* 25: 38–47.

Ibelings BW, De Bruin A, Kagami M, Rijkeboer M, Brehm M et al. (2004) Host parasite interactions between freshwater phytoplankton and chytrid fungi (Chytridiomycota). *Journal of Phycology* 40: 437–453.

James TY, Porter D, Leander CA, Vilgalys R & Joyce E (2000) Molecular phylogenetics of the Chytridiomycota supports the utility of ultrastructural data in chytrid systematics. *Canadian Journal*

*al of Botany* 78: 336–350.

- James TY, Kauff F, Schoch CL, Matheny PB, Hofstetter V, et al. **(2006a)** Reconstructing the early evolution of Fungi using a six-gene phylogeny. *Nature* 443: 818–822.
- James TY, Letcher PM, Longcore JE, Mozley-Standridge SE, Porter D et al. **(2006b)** A molecular phylogeny of the flagellated fungi (Chytridiomycota) and description of a new phylum (Blastocladiomycota). *Mycologia* 98: 860–871.
- Jebaraj CS, Raghukumar C, Behnke A & Stoeck T **(2010)** Fungal diversity in oxygen-depleted regions of the Arabian Sea revealed by targeted environmental sequencing combined with cultivation. *FEMS Microbiology Ecology* 71: 399–412.
- Jones MDM, Forn I, Gadelha C, Egan MJ, Bass D et al. **(2011a)** Discovery of novel intermediate forms redefines the fungal tree of life. *Nature* 474: 200–203.
- Jones MDM, Richards TA, Hawksworth DL & Bass D **(2011b)** Validation and justification of the phylum name *Cryptomycota* phyl. nov. *IMA Fungus* 2: 173–175.
- Kagami M, Bruin A, Ibelings BW & Donk E **(2007)** Parasitic chytrids: their effects on phytoplankton communities and food-web dynamics. *Hydrobiologia* 578: 113–129.
- Kim E, Simpson AGB & Graham LE **(2006)** Evolutionary relationships of Apusomonads inferred from taxon-rich analyses of 6 nuclear encoded genes. *Molecular Biology and Evolution* 23: 2455–2466.
- Kirk PM, Cannon PF, Minter DW & Stalpers JA **(2008)** *Dictionary of the Mycota*. 10<sup>th</sup> Edition. Cromwell Press, Trowbridge.
- Krüger M, Krüger C, Walker C, Stockinger H & Schüssler A **(2012)** Phylogenetic reference data for systematics and phylotaxonomy of arbuscular mycorrhizal fungi from phylum to species level. *New phytologist* 193: 970–984.
- Küpper FC, Maier I, Muller DG, 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.

Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, et al. (2007) Clustal W and Clustal X version 2.0. *Bioinformatics* 23: 2947–2948.

Le Calvez T, Burgaud G, Mahé S, Barbier G & Vandenkoornhuyse P (2009) Fungal diversity in deep-sea hydrothermal ecosystems. *Applied and Environmental Microbiology* 75: 6415–6421.

Lefèvre E, Bardot C, Noël C, Carrias J-F, Viscogliosi E et al. (2007) Unveiling fungal zooflagellates as members of freshwater picoeukaryotes: evidence from a molecular diversity study in a deep meromictic lake. *Environmental Microbiology* 9: 61–71.

Lefèvre E, Roussel B, Amblard C & Sime-Ngando T (2008) The molecular diversity of freshwater picoeukaryotes reveals high occurrence of putative parasitoids in the plankton. *PLoS one* 3: e2324.

Lefranc M, Thénot A, Lepère C & Debroas D (2005) Genetic Diversity of Small Eukaryotes in Lakes Differing by Their Trophic Status. *Applied Environmental Microbiology* 71: 5935.

Letcher PM, Powell MJ, Churchill PF & Chambers JG (2006) Ultrastructural and molecular phylogenetic delineation of a new order, the Rhizophydiales (Chytridiomycota). *Mycological Research* 110: 898-915.

Letcher PM, Powell MJ, Barr DJS, Churchill PF, Wakefield WS et al. (2008) Rhizophlyctidales—a new order in Chytridiomycota. *Mycological Research* 112: 1031-1048.

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.

Massana R, Balagué V, Guillou L & Pedrós-Alió C (2004) Picoeukaryotic diversity in an oligotrophic coastal site studied by molecular and culturing approaches. *FEMS Microbiology Ecology* 50: 231–243.

Moreira D, von der Heyden S, Bass D, López-García P, Chao E et al. (2007) Global eukaryote phylogeny: Combined small- and large-subunit ribosomal DNA trees support monophyly of Rhiza-

ria, Retaria and Excavata. *Molecular Phylogenetics and Evolution* 44: 255–266.

Moreira D, von der Heyden S, Bass D, López-García P, Chao E et al. (2007) Global eukaryote phylogeny: Combined small- and large-subunit ribosomal DNA trees support monophyly of Rhizaria, Retaria and Excavata. *Molecular Phylogenetics and Evolution* 44: 255–266.

Mueller GM & Schmit JP (2007) Fungal biodiversity: what do we know? What can we predict? *Biodiversity and Conservation* 16: 1–5.

Nagano Y, Nagahama T, Hatada Y, Nunoura T, Takami H et al. (2010) Fungal diversity in deep-sea sediments – the presence of novel fungal groups. *Fungal Ecology* 3: 316–325.

O'Brien HE, Parrent JL, Jackson JA, Moncalvo J-M & Vylgalis R (2005) Fungal community analysis by large-scale sequencing of environmental samples. *Applied and Environmental Microbiology* 71: 5544-5550.

Park S-J, Park B-J, Pham VH, Yoon D-N, Kim S-K et al. (2008) Microeukaryotic Diversity in Marine Environments, an Analysis of Surface Layer Sediments from the East Sea. *The Journal of Microbiology* 46: 244-249.

Philippe H & Laurent J (1998) How good are deep phylogenetic trees? *Current Opinion in Genetics and Development* 8: 616–623.

Porter TM, Schadt CW, Rizvi L, Martin AP, Schmidt SK et al. (2008) Widespread occurrence and phylogenetic placement of a soil clone group adds a prominent new branch to the fungal tree of life. *Molecular Phylogenetics and Evolution* 46: 635–644.

Posada D & Crandall KA (1998) MODELTEST: testing the model of DNA substitution. *Bioinformatic Applications Note* 14: 817-818.

Powell MJ (1993) Looking at mycology with a Janus face – A glimpse at Chtridiomycetes active in the environment. *Mycologia* 85: 1-20.

Ragon M, Restoux G, Moreira D, Møller AP, López-García P (2011) Sunlight-Exposed Biofilm Microbial Communities Are Naturally Resistant to Chernobyl Ionizing-Radiation Levels. *PLoS*



ONE 6: e21764.

- Ramirez-Llodra E, Shank TM & German CR (2007) Biodiversity and Biogeography of Hydrothermal Vent Species. *Oceanography* 20:30-41.
- Redecker D (2002) New views on fungal evolution based on DNA markers and the fossil record. *Research in Microbiology* 153: 125–130.
- Richards TA & Bass D (2005) Molecular screening of free-living microbial eukaryotes: diversity and distribution using a meta-analysis. *Current Opinion in Microbiology* 8: 240–252.
- Richards TA., Jones MDM., Leonard G & Bass D (2012) Marine fungi: their ecology and molecular diversity. *Annual Review of Marine Science* 4: 495–522.
- Robert F & Chaussidon M (2006) A palaeotemperature curve for the Precambrian oceans based on silicon isotopes in cherts. *Nature* 443: 969-972.
- Ronquist, F & Huelsenbeck, JP (2003) MrBayes3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* 19: 1572–1574.
- Rosling A, Cox F, Cruz-Martinez K, Ihrmark K, Grelet G-A et al. (2011) Archaeorhizomycetes: Unearthing an ancient class of ubiquitous soil fungi. *Science* 333: 876–879.
- Schadt CW, Martin AP, Lipson DA & Schmidt SK (2003) Seasonal dynamics of previously unknown fungal lineages in tundra soils. *Science* 301: 1359–1361.
- Schmit JP & Mueller GM (2007) An estimate of the lower limit of global fungal diversity. *Biodiversity and Conservation* 16: 99–111.
- Simmons DR, James TY, Meyer AF & Longcore JE (2009) Lobulomycetales, a new order in the Chytridiomycota. *Mycological Research* 113: 450-460.
- Stechmann A & Cavalier-Smith T (2003) Phylogenetic analysis of eukaryotes using heat-shock protein Hsp90. *Journal of Molecular Evolution* 57: 408–419.
- Stoeck T & Epstein S (2003) Novel Eukaryotic Lineages Inferred from Small-Subunit rRNA Analyses of Oxygen-Depleted Marine Environments. *Applied and Environmental Microbiology* 69: 2657-2663.

- Stoeck T, Hayward B, Taylor GT, Varela R & Epstein SS (2006) A multiple PCR-primer approach to access the microeukaryotic diversity in environmental samples. *Protist* 157: 31–43.
- Takishita K, Kakizoe N, Yoshida T & Maruyama T (2010) Molecular Evidence that Phylogenetically Diverged Ciliates Are Active in Microbial Mats of Deep-Sea Cold-Seep Sediment. *Journal of Eukaryotic Microbiology* 57: 76-86.
- Thaler AD, Van Dover CL & Vilgalys R (2012) Ascomycete phylotypes recovered from a Gulf of Mexico methane seep are identical to an uncultured deep-sea fungal clade from the Pacific. *Fungal Ecology* 5: 270–273.
- Vandenkoornhuyse P, Baldauf SL, Leyval C, Straczek J & Young JPW (2002) Extensive fungal diversity in plant roots. *Science* 295: 2051.
- Wächtershäuser G (1990) Evolution of the first metabolic cycle. *Proceedings of the National Academy of Sciences, USA* 87: 200-204.
- Wardle DA & Van der Putten WH (2002) Biodiversity, ecosystem functioning and above-ground-below-ground in *Biodiversity and ecosystem functioning: synthesis and perspectives*. Loreau M, Naeem S & Inchausti P, eds. Oxford University Press.
- White MM, James TY, O'Donnell K, Cafaro MJ, Tanabe Y & Sugiyama J (2006) Phylogeny of the Zygomycota based on nuclear ribosomal sequence data. *Mycologia* 98: 872–884.
- Wuyts J, De Rijk P, van de Peer Y, Pison G, Rousseeuw P et al. (2000) Comparative analysis of more than 3000 sequences reveals the existence of two pseudoknots in area V4 of eukaryotic small subunit ribosomal RNA. *Nucleic Acids Research* 28: 4698–4708.
- Zuendorf A, Bunge J, Behnke A, Barger KJA & Stoeck T (2006) Diversity estimates of microeukaryotes below the chemocline of the anoxic Mariager Fjord, Denmark. *FEMS Microbiology Ecology* 58: 476–491.



# DISCUSSION GENERALE ET PERSPECTIVES



## Discussion générale et perspectives

---

Les Chytridiomycota n'ont pas beaucoup été étudiés, et de ce fait, les connaissances à leur sujet, tous domaines confondus, ne sont pas très étendues. A l'exception de quelques cas dont l'étude est induite par leur dimension anthropocentrique. Si l'on fait une recherche rapide avec des mots clés correspondant aux cinq phyla, en utilisant des plate-formes de bases de données (*e.g.* Web of Knowledge, Pub Med...), on constate qu'il y a quasiment un facteur dix, entre le nombre de publications sur les Chytridiomycota et celles sur les Basidiomycota et Ascomycota. Cependant, pendant ces dix dernières années, une attention croissante a été portée sur ce phylum de champignons, et les études qui en découlent ont permis d'enrichir les connaissances sur leur diversité et leurs fonctions écosystémiques entre autres. Par exemple, il a été montré que dans certains écosystèmes, ils dominaient la communauté fongique (*e.g.* Freeman *et al.*, 2009) ou encore qu'ils pouvaient intervenir à différents niveaux dans les réseaux trophiques lentiennes en tenant des rôles divers. L'objectif de cette thèse a été de décrire leur diversité via la génomique environnementale dans des écosystèmes hydrothermaux marins profonds, leur présence ayant été détectée dans ces milieux. L'analyse des données obtenues a nécessité l'utilisation d'une base de données publique de séquences moléculaires, qui dans son état actuel ne répondait pas à toutes nos attentes. Ainsi, il a été créé une base de données visant spécifiquement les marqueurs moléculaires fongiques couramment utilisés pour la construction de phylogénies et validés selon des critères de qualité. Cette partie discutera des principaux résultats générés lors de cette thèse et proposera des perspectives de recherche.

### PHYMYCO-DB : un outils pour les analyses phylogénétiques fongiques

La base de données PHYMYCO-DB a été mise en place pour centraliser des séquences fongiques directement utilisables pour des analyses phylogénétiques. Cela sous-entend non seulement de sélectionner des gènes porteurs d'informations évolutives mais également que ces

séquences soient contrôlées vis-à-vis de leur qualité. En effet, celles-ci sont extraites d'une base de données publique dans laquelle elles sont déposées sans examen de leur contenu nucléotidique ni des renseignements descriptifs comme par exemple, l'organisme source. Ainsi il y a un risque qu'une séquence erronée soit enregistrée, ce qui l'invalide pour une éventuelle utilisation dans une autre étude. Cependant, certaines erreurs ne sont pas détectables lorsqu'un alignement local avec BLASTn est réalisé, ce qui peut engendrer leur propagation, et elles ne sont perceptibles que lorsqu'un alignement multiple est effectué. Dans notre analyse, près de 20% des séquences récupérées après le filtrage automatique, ont été écartées, ce qui est assez révélateur de l'utilité de bases de données comme PHYMYCO-DB. Ce résultat illustre également une pollution beaucoup plus importante des banques de séquences qu'imaginée.

Plusieurs autres bases de données de ce type existent et sont complémentaires. Les motivations sur lesquelles reposent leur création sont assez similaires mais les informations qu'elles contiennent peuvent être destinées à des usages différents. Toutefois, l'objectif commun est d'aider à la caractérisation de la diversité fongique que ce soit d'un point de vue taxonomique ou phylogénétique, et c'est en lien direct avec les importantes modifications subies par la taxonomie du règne des champignons récemment. En effet, leur complémentarité va probablement permettre d'accélérer la description de cette diversité mais également de la valider avec beaucoup plus de robustesse. Par exemple, si un nouveau champignon Asco- ou Basidiomycota est décrit, le séquençage des *ITS* va permettre de lui attribuer un 'code-barre' moléculaire qui sera confronté à une base de données comme UNITE pour confirmer son état nouveau et le séquençage d'un marqueur d'évolution va permettre de le positionner au sein de l'arbre de la vie des champignons *e.g.* avec PHYMYCO-DB. Ce qui est en accord avec la volonté des mycologues d'intégrer les séquences nucléotidiques à la description des espèces. De plus, le choix de séquences semble se porter sur les gènes de l'ARNr pour la variété des informations qu'ils apportent.

Si certaines erreurs sont inhérentes aux champignons et aux problèmes d'identification des espèces, les problèmes de séquençage ou de mauvaises annotations peuvent survenir pour tous types d'organismes. Ce sont d'ailleurs des bases de données sur le gène qui code l'ARNr 16S des

bactéries et des archaeobactéries qui ont été construites les premières. La suite logicielle ARB a commencé à être développée au milieu des années 90 car déjà à cette époque le nombre de séquences disponibles était conséquent et elles nécessitaient un traitement spécifique avant leur exploration (Pruesse *et al.*, 2007). Elle a depuis été étendue aux Eukaryota. A la différence de PHYMYCO-DB, le contrôle de la qualité est réalisé par filtration automatique uniquement. ARB a été testé pour la réalisation de la phylogénie sur les lignées basales (*Cf.* Chapitre 3) et les résultats obtenus étaient assez proches de ceux présentés ici. La différence majeure réside dans le temps nécessaire à l'obtention des données. PHYMYCO-DB s'avère être plus rapide et plus simple d'utilisation, ce qui est probablement dû au fait qu'elle soit restreinte à un plus petit groupe taxonomique.

Cette étude a donc montré que PHYMYCO-DB est un outil adapté à l'étude des relations phylogénétiques chez les champignons. Elle a notamment été très utile pour le dessin des amorces, pour que celles-ci couvrent la plus grande partie possible de la diversité des séquences de Chytridiomycota décrits ainsi que les séquences environnementales.

### La génomique environnementale appliquée aux Chytridiomycota

Ce travail constitue la première étude focalisée sur les chytrides dans un écosystème donné en utilisant les outils de la génomique environnementale. En effet, les études de diversité précédentes qui ont détecté la présence de Chytridiomycota, ciblent l'ensemble du domaine des Eukaryota ou du règne des champignons. Dans la plupart des cas, elles montrent que les Chytridiomycota présentent une diversité assez faible et sont peu abondants. Toutefois, il est à noter que cette méthode semble la plus adaptée pour les décrire puisque les autres méthodes y font rarement référence, *e.g.* les études de diversité par une approche culturale ou par des observations au microscope. Cela peut venir de plusieurs facteurs tels que ceux décrits dans la revue bibliographique, *i.e.* leur taille microscopique, les difficultés à les isoler... Un autre biais peut être induit selon le type d'écosystème où se trouve la communauté étudiée. En effet, les Chytridiomycota nécessitant la présence d'eau pour leur maintien, du moins sur une certaine période, il est probable que les écosystèmes les plus couramment étudiés ne présentent pas des



conditions optimales pour leur développement.

Suite aux résultats obtenus lors de cette thèse, un biais additionnel lié à la méthode moléculaire a été déterminé. Les amorces utilisées lors d'études plus larges d'un point de vue taxonomique ne sont pas divergentes des séquences du gène de l'ARNr 18S des Chytridiomycota vu que leurs séquences sont amplifiées. L'une des hypothèses est qu'ils font partie de la biosphère rare qui est difficile à étudier via la constitution de banques de clones. De plus le séquençage de ces clones est rarement mené jusqu'à saturation. La difficulté à obtenir des amplicons et surtout la non répétabilité de l'amplification par PCR constituent des arguments en la faveur de cette hypothèse. En effet, il est probable que le ratio organismes autres que Chytridiomycota *versus* Chytridiomycota soit fortement déséquilibré en faveur des autres organismes, diminuant ainsi la probabilité de rencontre entre les amorces et la matrice d'ADN cible. Le pyroséquençage est une solution proposée à l'étude de cette biosphère rare puisqu'il permet d'étendre l'échantillonnage des séquences jusqu'à saturation selon les organismes ciblés. Toutefois l'une des conditions incontournables est l'obtention d'une quantité suffisante d'amplicons à séquencer. De plus, et *a posteriori*, il semblerait que dans le cas de notre étude, l'utilisation du pyroséquençage aurait été superflu car la diversité par échantillon est assez faible. Par contre, l'emploi des amorces dessinées ici, combinées avec du pyroséquençage s'avérerait pertinent dans le cas où les Chytrides dominent la communauté fongique et représentent une grande proportion de la communauté eucaryote comme dans l'étude de Freeman *et al.* (2009). De manière plus prudente, l'emploi de ces amorces combinées à du clonage semble efficace pour décrire la diversité des Chytridiomycota dans tous types d'écosystèmes.

#### *Nouvelles technologies/stratégies potentiellement applicables*

Les analyses metatranscriptomiques permettent une analyse directe de l'ensemble des ARN extraits, sans passer par une étape de PCR dont les amorces peuvent induire des biais, comme vu précédemment. Le pyroséquençage d'ARN donne accès aux ARNr et aux autres familles d'ARN y compris les messagers qui renseignent sur les fonctions exercées. Par l'analyse des ARNr séquencés on peut décrire très finement la diversité globale des (micro)organismes qui composent un écosystème (Quaiser *et al.*, en préparation). De manière plus prospective, avec un équipement

comme le LabChip XT (Caliper), il devrait devenir possible, avec l'amélioration des systèmes microfluidiques développés, de séparer préférentiellement les ARNr 18S d'un mélange d'ARN totaux par sélection de taille. Par cette stratégie, *i.e.* sans passer par une étape d'amplification, se profile donc la possibilité d'un séquençage de masse d'ARN. Nous pourrions alors envisager la description de la diversité avec un nouveau regard. Néanmoins un verrou subsiste : la nécessité de subdiviser le jeu de séquences par zones alignables pour ensuite analyser ces zones une par une et en déduire la diversité, et cela même, si par pyroséquençage, des fragments de 500-700 nucléotides peuvent être produits. L'un des défis ici sera donc d'être capable d'analyser ce type de données.

### Importance du choix du couple d'amorces

La description de la diversité par une approche moléculaire a montré que le choix du couple d'amorces est déterminant. Comparer des diversités obtenues par des couples d'amorces ciblant des niveaux taxonomiques différents, correspond à observer la diversité à la manière d'un changement d'échelle taxonomique. A un niveau élevé, les amorces doivent être porteuses d'un caractère universel. Cela explique que les études eucaryotes, qui pour la plupart ciblent des protistes, ont pu mettre en évidence la présence de champignons. A un niveau intermédiaire, comme le règne des champignons, les amorces sont dessinées dans des zones assez conservées pour pouvoir récupérer des séquences qui présentent un degré de divergence entre elles, encore assez fort. Par contre, dans cette étude, l'objectif premier étant de décrire un phylum, il a été important de se concentrer sur une zone conservée au sein de ce phylum mais suffisamment divergente des clades proches. Mais la particularité de ce phylum est qu'il se trouve en position basale du règne des champignons et donc proche de la divergence champignons-animaux. Cela implique de porter une attention particulière sur le degré de divergence des autres phyla fongiques, laissant une ouverture possible sur les nœuds antérieurs aux champignons et au règne des animaux. Les résultats montrent qu'effectivement les amorces dessinées ciblent une zone captant des séquences d'organismes se positionnant en amont des Chytridiomycota. Cette zone antérieure à la divergence champignons-animaux est une zone qui selon la littérature n'est pas très

bien résolue. Il est donc montré que des amorces ciblant un clade proche d'une zone de divergence ont une capacité intrinsèque de décrire des clades en amont de cette zone. Le choix de l'amorce détermine donc la diversité qui est décrite. Cela implique que lorsque la diversité est décrite dans un écosystème donné, celle-ci est dépendante du couple d'amorces et ne reflète pas une diversité exhaustive de la communauté d'organismes présente. L'utilisation de plusieurs couples d'amorces sur un même métagénome environnemental montre bien ce phénomène.

### Les clades en amont de la divergence champignons-animaux

Le couple MH2/SR8RChytrid a permis de mettre en évidence la présence de séquences d'organismes qui se branchent de manière basale dans les Opisthokonta et/ou de manière basale dans les Unikonta. Concernant les Apusozoa, les études s'attachant à résoudre leur positionnement au sein des Eukaryota ont souvent abouti à des résultats non soutenus statistiquement, à l'exception d'une étude avec le gène de l'ARNr 18S (Cavalier-Smith & Chao, 1995) et d'une analyse multigénique (Kim *et al.*, 2006). Il y est montré que les Apusozoa se positionnent en clade frère des Opisthokonta après l'embranchement des Amoebozoa. Nos résultats montrent que le nœud qui porte les Opisthokonta et les Apusozoa est moyennement soutenu. Par contre la présence de séquences d'Apusozoa dans notre jeu de données nous permet de dire que le groupe inconnu mis en évidence dans les échantillons d'origine hydrothermal se positionne de manière antérieure aux Opisthokonta. Il est également probable que ce groupe se branche à la base des Unikonta. Toutefois nous ne pouvons ni confirmer ni infirmer cette hypothèse puisque le fragment du gène de l'ARNr amplifié n'a pas permis d'intégrer des séquences d'Amoebozoa à l'alignement, celles-ci étant trop divergentes. De ce fait, ce groupe est très intéressant de part les informations qu'il peut apporter pour résoudre les racines des Opisthokonta, avec par exemple le positionnement des Apusozoa, mais également l'articulation entre les Unikonta et les Bikonta.

Des études complémentaires pourraient apporter plus d'informations sur ce nouveau groupe. Tout d'abord le séquençage du gène de l'ARNr en son entier pourrait permettre de le

positionner de manière plus robuste dans l'arbre de la vie. L'utilisation de marqueurs fluorescents pourraient permettre de visualiser les organismes auxquels appartiennent ces séquences comme cela a été fait pour la mise en évidence du phylum des Cryptomycota (Jones *et al.*, 2011). Enfin, une approche moléculaire consistant à séquencer son génome permettrait de faire une analyse plus poussée de son positionnement phylogénétique, ainsi qu'une analyse de prédiction de fonctions.

### Les Chytridiomycota en milieu hydrothermal marin profond

La diversité décrite dans cette étude montre que les Chytridiomycota présents dans les écosystèmes hydrothermaux étudiés, sont répartis d'un point de vue taxonomique sur une grande partie du phylum. En effet, les séquences obtenues appartiennent à trois des six ordres, sans compter *Olpidium* et le phylotype le plus divergent retrouvé sur plusieurs sites, qui pourrait constituer un nouvel ordre de Chytridiomycota. Ce profil est comparable à ceux décrits dans des lacs (Lefèvre *et al.*, 2007 ; 2008), en nombre de phylotypes, mais dans notre étude, la composition taxonomique est plus diverse. Cela pourrait suggérer qu'un événement de diversification dans ces lignées basales se serait produit en milieu marin. En effet, des séquences appartenant à des organismes décrits sont effectivement retrouvées, mais des séquences environnementales ou distantes des espèces décrites sont récupérées en plus grande proportion. Le fait d'avoir une minorité d'espèces décrites est assez révélateur de l'absence de connaissance sur ce phylum. Cela n'est pas seulement lié au fait que ce soit la première étude des Chytridiomycota en milieu hydrothermal puisque la même tendance est observée en lac (Lefèvre *et al.*, 2007 ; 2008) et dans les sols à haute altitude (Freeman *et al.*, 2009). Pour cette dernière étude, 27 nouveaux phylotypes ont été mis en évidence et seulement trois étaient proches de taxons connus. Cependant la majorité de ces nouvelles séquences sont regroupées dans un clade très divergent qui comprend les Spizellomycetales et les Rhizophlyctidiales. La diversité semble être moins étendue taxonomiquement, ce qui conforte l'hypothèse de diversification marine, sans pour autant la confirmer. L'obtention d'arguments confirmant ou non cette diversification marine viendra de l'extension de la diversité décrite, en prospectant différents écosystèmes.

A cette diversité est liée la question sur leurs fonctions écosystémiques, et de la même manière on retrouve un biais lié aux études. En effet, les Chytridiomycota ont pour le moment été décrits comme étant des saprophytes et des parasites, ces derniers étant largement majoritaires. En partant du principe qu'ils sont peu étudiés et difficiles à étudier de manière générale, cela peut s'expliquer par le fait qu'un parasite est plus facilement détectable puisqu'il induit des symptômes généralement visibles et menant à la mort de l'hôte (Ibelings, 2004). Il est probable que les Chytridiomycota présentent d'autres interactions biologiques dans les écosystèmes comme le commensalisme ou encore la symbiose mutualiste, puisque la longueur des branches de la majorité des séquences obtenues est assez courte. Cette diversité en milieu hydrothermal soulève la question concernant les fonctions exercées et il serait intéressant de les regarder de plus près. Par exemple, par l'analyse des génomes il est possible de prédire les fonctions encodées par les gènes et cela peut être étendu aux fonctions écologiques potentielles de ces organismes.

#### Diversification fongique : milieu marin versus milieu terrestre

La diversité fongique en milieu marin profond est étudiée avec des techniques moléculaires depuis quatre ou cinq ans. L'analyse de Richards *et al.* (2012) montre qu'un grand nombre de champignons répartis dans les Ascomycota, Basidiomycota et Chytridiomycota, sont présents dans le milieu marin. De manière générale, les lignées basales présentent une divergence beaucoup plus marquée avec les espèces décrites et les séquences environnementales obtenues, que les lignées des Dikarya. Deux hypothèses peuvent être formulées. La première repose sur le caractère ubiquiste des champignons supérieurs et le fait qu'ils aient été beaucoup plus étudiés, d'où une plus grande probabilité de les retrouver dans deux écosystèmes différents. La seconde serait de considérer un événement de diversification, surtout concernant les lignées basales, avant la colonisation terrestre. Tout d'abord, à l'encontre de la première hypothèse, les lignées basales sont peu connues d'où une plus grande probabilité de récupérer des séquences plus divergentes. Ensuite, concernant les lignées basales, la diversité décrite en milieu marin est répartie sur l'ensemble du phylum alors que pour les lignées des Dikarya, la diversité marine semble restreinte à certains groupes taxonomiques. Richards *et al.* (2012) et Bass *et al.* (2007) proposent des

transitions marine-continentale dans les deux sens avec des événements de diversification dans les deux milieux. Cette hypothèse est validée au regard de l'imbrication des séquences d'origine marine avec les séquences d'origine continentale, ainsi que les tailles de branches courtes généralement constatées. Par contre, l'hypothèse d'une diversification continentale en raison d'un nombre plus important de séquences marines appartenant aux Dikarya et d'une plus grande proportion de lignées basales en eau douce par rapport au milieu marin (Richards *et al.*, 2012) est peu parcimonieuse et n'est pas confirmée par les travaux développés dans ce travail de thèse. De plus il est clair que les couples d'amorces généralement utilisés produisent une amplification biaisée vers les Dikarya et donc ne reflète pas la diversité réelle. Ce fait est une conséquence de la composition biaisée des banques de séquences publiques internationales. Le cas des Cryptomycota est intéressant et reste à éclaircir puisque nous n'avons pas pu décrire leur diversité avec les amorces MH2/SR8RChytrid. Il serait intéressant également d'utiliser ce couple d'amorces sur différents échantillons prélevés dans différents systèmes marins, à commencer par ceux déjà explorés, pour pouvoir comparer les proportions Dikarya/lignées basales. Pour le moment, les hypothèses de diversification sont recevables aussi bien pour le milieu continental que marin. Ainsi l'hypothèse de plus d'un événement de transition marine-continentale dans les deux sens est plausible. Par contre, aucun élément n'est disponible pour le moment pour déterminer une émergence plutôt continentale ou plutôt marine.

Les branches anciennes découvertes dans ce travail représente sans doute une avancée importante des connaissances. Ces branches témoignent d'une évolution plus complexe qu'imaginée quant à l'émergence et la diversification des Opisthokonta. Ce travail soulève de nouvelles questions et révèle la nécessité de développer des efforts de recherche dans la prospection de la diversité de ces groupes basaux.

Tout au long de cette thèse, le phylum des Chytridiomycota a été abordé à travers le prisme de l'évolution et il a été montré qu'ils occupaient une place stratégique pour la compréhension de la complexité de l'histoire évolutive du règne des champignons. Depuis la mise en évidence des Cryptomycota, ils ont une position encore plus déterminante puisqu'ils représentent une transition entre ces organismes et les autres champignons. Cette étude pionnière ouvre la voie à des études

qui pourraient se développer sur la complexité évolutive mais également sur l'importance écologique de cette diversité cryptique. Les difficultés techniques et méthodologiques résolues au cours de ce travail offrent aussi des solutions pour contourner un certain nombre d'écueils.

REFERENCES

BIBLIOGRAPHIQUES





## REFERENCES BIBLIOGRAPHIQUES

---

- Abarenkov K, Nilsson RH, Larsson KH, Alexander IJ, Eberhardt U, et al. (2010a) The UNITE database for molecular identification of fungi - recent updates and future perspectives. *New Phytologist* 186: 281-285.
- Abarenkov K, Tedersoo L, Nilsson RH, Vellak K, Saar I, et al. (2010b) PlutoF-a Web Based Workbench for Ecological and Taxonomic Research, with an Online Implementation for Fungal ITS Sequences. *Evolutionary Bioinformatics* 6: 189-196.
- Adl SM, Simpson AGB, Farmer MA, Andersen RA, Anderson OR, et al. (2005) The new higher level classification of eukaryotes with emphasis on the taxonomy of protists. *The Journal of Eukaryotic Microbiology* 52: 399–451.
- Alker AP, Smith GW & Kim K (2001) Characterization of *Aspergillus sydowii* (Thom & Church), a fungal pathogen of Caribbean sea fan corals. *Hydrobiologia* 460: 105-111.
- Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic Local Alignment Search Tool. *Journal of Molecular Biology* 215: 403-410.
- Amaral Zettler LA, Nerad TA, O’Kelly CJ & Sogin ML (2001) The nucleariid amoebae: More protists at the animal-fungal boundary. *Journal of Eukaryotic Microbiology* 48: 293-297.
- Anderson IC, Parkin PI (2007) Detection of active soil fungi by RT-PCR amplification of precursor rRNA molecules. *Journal of Microbial Methods* 68: 248–253.
- Atkins Michael S, McArthur AG & Teske AP (2000) Ancyromonadida: A new phylogenetic lineage among the protozoa closely related to the common ancestor of metazoans, fungi , and choanoflagellates ( Opisthokonta ). *Journal of Molecular Evolution* 278–285.
- Baldauf SL & Palmert JD (1993) Animals and fungi are each other’s closest relatives: Congruent evidence from multiple proteins. *Evolution* 90: 11558–11562.
- Baldauf SL & Doolittle WF (1997) Origin and evolution of the slime molds (Mycetozoa). *Proceedings of the National Academy of Sciences of the United States of America* 94: 12007-

12012.

- Baldauf SL (1999) A search for the origins of Animals and Fungi: comparing and combining molecular data. *American Naturalist* 154: S178-S188.
- Baldauf SL, Roger AJ, Wenk-Siefert I, Doolittle WF (2000) A kingdom-level phylogeny of eukaryotes based on combined protein data. *Science* 290: 972-977.
- Baldauf SL (2003) The deep roots of Eukaryotes. *Science* 300 : 1703-1706.
- Baldauf SL, Bhattacharya D, Cockrill J, Hugenholtz P, Pawlowski J et al. (2004) The Tree of Life, an overview in *Assembling the Tree of Life*. Cracraft J & Donoghue MJ, eds. Oxford University Press.
- Barghoorn ES & Linden DH (1944) Marine fungi: Their taxonomy and biology. *Farlowia* 1: 395-467.
- Bar-Hen A, Mariadassou M, Poursat MA, Vandenkoornhuysen P (2008) Influence function for robust phylogenetic reconstructions. *Molecular Biology and Evolution* 25: 869-873.
- Barr DJS (1980) An outline for the reclassification of the Chytridiales, and for a new order, the Spizellomycetales. *Canadian Journal of Botany* 58: 2380-2394.
- Bass D, Howe A, Brown N, Barton H, Demidova M, et al. (2007) Yeast forms dominate fungal diversity in the deep oceans. *Proceedings of the Royal Society B: Biological sciences* 274: 3069–3077.
- Bass D & Richards TA (2011) Three reasons to re-evaluate fungal diversity “on Earth and in the ocean.” *Fungal Biology Reviews* 25: 159–164.
- Beatty JT, Overmann J, Lince MT, Manske AK, Langet AS et al. (2005) An obligately photosynthetic bacterial anaerobe from a deep-sea hydrothermal vent. *Proceedings of the National Academy of Sciences of the United States of America* 102: 9306-9310.
- Benson DA, Karsch-Mizrachi I, Lipman DJ, Ostell J, Sayers EW (2011) GenBank. *Nucleic Acids Research* 39: D32-D37.
- Benton MJ & Ayala FJ (2003) Dating the tree of life. *Science* 300: 1698–1700.
- Berbee ML, Kloeckener B, Taylor JW (1992) Yeast biology. *Science* 257: 1610-1611.
- Bhattacharya D, Friedl T, Helms G (2002) Vertical evolution and intragenic spread of lichen-fungal group I intron. *Journal of Molecular Evolution* 55: 74-84.

- Bidartondo MI, Bruns TD, Blackwell M, Edwards I & Taylor AFS (2008) Preserving accuracy in GenBank. *Science* 319: 1616-1616.
- Biddle JF, House CH & Brenchley JE (2005) Microbial stratification in deeply buried marine sediment reflects changes in sulfate/methane profiles. *Geobiology* 3: 287-295.
- Blackwell M (2011) The Fungi: 1, 2, 3 ... 5.1 million species? *American Journal of Botany* 98: 426–438.
- Blair JE (2009) The Fungi in *The Timetree of Life*. Hedges SB & Kumar S, eds. Oxford University Press.
- Boon E, Zimmerman E; Lang BF, Hijri, M (2010) Intra-isolate genome variation in arbuscular mycorrhizal fungi persists in the transcriptome. *Journal of Evolutionary Biology* 23:1519-1527.
- Boussau B, Blanquart S, Necsulea A, Lartillot N & Gouy M (2008) Parallel adaptations to high temperatures in the Archaean eon. *Nature* 456: 942-945.
- Bowman SM & Free SJ (2006) The structure and synthesis of the fungal cell wall. *BioEssays* 28: 799–808.
- Bridge PD, Roberts PJ, Spooner BM & Panchal G (2003) On the unreliability of published DNA sequences. *New Phytologist* 160: 43-48
- Bridge, PD & Worland MR (2004) First report of an entomophthoralean fungus on an arthropod host in Antarctica. *Polar Biology* 27: 190-192
- Bridge PD & Newsham KK (2009) Soil fungal community composition at Mars Oasis, a southern maritime Antarctic site, assessed by PCR amplification and cloning. *Fungal Ecology* 2: 66-74.
- Brown MW, Spiegel FW & Silberman JD (2009) Phylogeny of the “forgotten” cellular slime mold, *Fonticula alba*, reveals a key evolutionary branch within Opisthokonta. *Molecular Biology and Evolution* 26: 2699–2709.
- Bugni TS & Ireland CM (2004) Marine-derived fungi: a chemically and biologically diverse group of microorganisms. *Natural Product Reports* 21: 143–163.
- 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*

11: 1588–1600.

- Burgaud G, Arzur D, Durand L, Cambon-Bonavita M-A & Barbier G (2010) Marine culturable yeasts in deep-sea hydrothermal vents: species richness and association with fauna. *FEMS Microbiology Ecology* 73: 121–133.
- Burki F, Shalchian-Tabrizi K, Minge M, Skjaeveland A, Nikolaev SI, Jakobsen KS & Pawlowski J (2007) Phylogenomics reshuffles the eukaryotic supergroups. *PloS one* 2: e790.
- Carr M & Baldauf SL (2011) The protistan origins of animals and fungi in The Mycota: Evolution of Fungi and fungal-like organisms XIV. Pöggeler S & Wöstemeyer J, eds. Springer Heidelberg Dordrecht London New York.
- Cavalier-smith T & Chao E (1995) The opalozoan Apusomonas is related to the common ancestor of animals, fungi, and choanoflagellates. *Proceedings of the Royal Society B: Biological sciences* 261: 1–6.
- Cavalier-Smith T, Chao EE, Stechmann A, Oates B, & Nikolaev S (2008) Planomonadida ord. nov. (Apusozoa): ultrastructural affinity with micronuclearia podoventralis and deep divergences within planomonas gen. nov. *Protist* 159: 535-562.
- Cavalier-Smith T (2009) Megaphylogeny, Cell Body Plans, Adaptive Zones: Causes and Timing of Eukaryote Basal Radiations. *Journal of Eukaryotic Microbiology* 56: 26–33.
- Cavalier-Smith T & Chao EE (2010) Phylogeny and evolution of Apusomonadida (Protozoa: Apusozoa): new genera and species. *Protist* 161: 549–576.
- Choi IG, Kim SH (2007) Global extent of horizontal gene transfer. *Proceedings of the National Academy of Sciences of the United States of America* 104: 4489-4494.
- Cole JR, Wang Q, Cardenas E, Fish J, Chai B, et al. (2009) The Ribosomal Database Project: Improved alignments and new tools for rRNA analysis. *Nucleic Acids Research* 37: D141-D145.
- Crous PW, Gams W, Stalpers JA, Robert V & Stegehuis G (2004) MycoBank: an online initiative to launch mycology into the 21st century. *Studies in Mycology* 50: 19–22.
- Damare S, Raghukumar C & Raghukumar S (2006) Fungi in deep-sea sediments of the Central Indian Basin. *Deep Sea Research Part I* 53: 14–27.

- D'Auria G, Pushker R, Rodriguez-Valera F (2006) IWoCS analyzing ribosomal intergenic transcribed spacers configuration and taxonomic relationships. *Bioinformatics* 22: 527-531.
- Dawson SC & Pace NR (2002) Novel kingdom-level eukaryotic diversity in anoxic environments. *Proceedings of the National Academy of Sciences U.S.A.* 99: 8324–8329.
- DeSantis TZ, Hugenholtz P, Larsen N, Rojas M, Brodie EL et al. (2006) Greengenes, a Chimera-Checked 16S rRNA Gene Database and Workbench Compatible with ARB. *Applied Environmental Microbiology* 72: 5069-5072.
- Díez B, Pedros-Alio C & Massana R (2001) Study of genetic diversity of eukaryotic picoplankton in different oceanic regions by small-subunit rRNA gene cloning and sequencing. *Applied and Environmental Microbiology* 67: 2932-2941.
- Drummond AJ, Ho SYW, Phillips MJ & Rambaut A (2006) Relaxed phylogenetics and dating with confidence. *PLoS Biology* 4: e88.
- Edgcomb VP, Kysela DT, Teske A, de Vera Gomez A & Sogin ML (2002) Benthic eukaryotic diversity in the Guaymas Basin hydrothermal vent environment. *Proceedings of the National Academy of Sciences of the United States of America* 99: 7658–7662.
- Edgcomb V, Orsi W, Bunge J, Jeon S, Christen R, et al. (2011a) Protistan microbial observatory in the Cariaco Basin, Caribbean. I. Pyrosequencing vs Sanger insights into species richness. *The ISME journal* 5: 1344–1356.
- Edgcomb VP, Beaudoin D, Gast R, Biddle JF & Teske A (2011b) Marine subsurface Eukaryotes: the fungal majority. *Environmental Microbiology* 13: 172–183.
- Felsenstein J (2004) *Inferring phylogenies*. Sinauer Associates, Sunderland, Massachusetts.
- Fernandes P, Domitrovic T, Kao CM & Kurtenbach E (2004) Genomic expression pattern in *Saccharomyces cerevisiae* cells in response to high hydrostatic pressure. *FEBS Letters* 556: 153–160.
- Fisher CR, Takai K, Le Bris N (2007) Hydrothermal vent ecosystems. *Oceanography* 20: 14-23.
- Fiore-Donno AM, Nikolaev SI, Nelson M, Pawlowski J, Cavalier-Smith T et al. (2010) Deep phylogeny and evolution of slime moulds (Mycetozoa). *Protist* 161: 55–70.
- Freeman KR, Martin AP, Karki D, Lynch RC, Mitter MS, et al. (2009) Evidence that chytrids dominate

- fungal communities in high-elevation soils. *Proceedings of the National Academy of Sciences of the United States of America*. 106: 18315-18320.
- Gadanho M & Sampaio JP (2005) Occurrence and diversity of yeasts in the mid-atlantic ridge hydrothermal fields near the Azores Archipelago. *Microbial Ecology* 50: 408–417.
- Gao Z, Li B, Zheng C & Wang G (2008) Molecular detection of fungal communities in the Hawaiian marine sponges *Suberites zeteki* and *Mycale armata*. *Applied and Environmental Microbiology* 74: 6091–6101.
- Gao Z, Johnson ZI & Wang G (2010) Molecular characterization of the spatial diversity and novel lineages of mycoplankton in Hawaiian coastal waters. *The ISME journal* 4: 111–120.
- Giraud T, Refrégier G, Le Gac M, de Vienne DM & Hood ME (2008) Speciation in fungi. *Fungal Genetics and Biology* 45: 791–802.
- Gleason F, Kagami M, Lefevre E & Sime-Ngando T (2008) The ecology of chytrids in aquatic ecosystems: roles in food web dynamics. *Fungal Biology Reviews* 22: 17–25.
- Gleason FH & Lilje O (2009) Structure and function of fungal zoospores: ecological implications. *Fungal Ecology* 2: 53–59.
- Gleason FH, Schmidt SK & Marano AV (2010) Can zoosporic true fungi grow or survive in extreme or stressful environments? *Extremophiles* 14:417–425.
- Gleason FH, Küpper FC, Amon JP, Picard K, Gachon CMM et al. (2011) Zoosporic true fungi in marine ecosystems : a review. *Marine and Freshwater research* 62: 383–393.
- Golubic S, Radtke G & Le Campion-Alsumard T (2005) Endolithic fungi in marine ecosystems. *Trends in Microbiology* 13: 229–235.
- Guindon S & Gascuel O (2003) A Simple, Fast, and Accurate Algorithm to Estimate Large Phylogenies by Maximum Likelihood. *Systematic Biology* 52:696–704.
- Gunderson JH, Elwood H, Ingold A, Kindle K & Sogin ML (1987) Phylogenetic relationships between Chlorophytes, Chrysophytes, and Oomycetes. *Proceedings of the National Academy of Sciences of the United States of America* 84: 5823–5827.
- Gutiérrez MH, Pantoja S, Tejos E & Quinones RA (2011) The role of fungi in processing marine

organic matter in the upwelling ecosystem off Chile. *Marine Biology* 158: 205-219.

Hampel V, Hug L, Leigh JW, Dacks JB, Lang BF, Simpson AGB & Roger AJ (2009) Phylogenomic analyses support the monophyly of Excavata and resolve relationships among eukaryotic "supergroups". *Proceedings of the National Academy of Sciences of the United States of America* 106: 3859–3864.

van Hannen EJ, Mooij W, Agterveld MPV, Gons HJ & Laanbroek HJ (1999) Detritus-Dependent Development of the Microbial Community in an Experimental System : Qualitative Analysis by Denaturing Gradient Gel Electrophoresis. *Applied and environmental microbiology* 65: 2478–2484

Hawksworth DL (1991) The fungal dimension of biodiversity: Magnitude, significance, and conservation. *Mycological Research* 95: 641-655.

Hawksworth DL (2001) The magnitude of fungal diversity: the 1,5 million species estimate revisited. *Mycological Research* 105: 1422–1432.

Hawksworth DL (2004) 'Misidentifications' in fungal DNA sequence databanks. *New Phytologist* 161: 13-15.

Hawksworth DL (2005) Universal fungus register offers pattern for zoology. *Nature* 438: 2005.

Hawksworth DL (2009) Separate name for fungus's sexual stage may cause confusion. *Nature* 458: 29.

Heckman DS, Geiser DM, Eidell BR, Stauffer RL, Kardos NL et al. (2001) Molecular evidence for the early colonization of land by fungi and plants. *Science* 293: 1129–1133.

Hedges SB, Blair JE, Venturi ML & Shoe JL (2004) A molecular timescale of eukaryote evolution and the rise of complex multicellular life. *BMC Evolutionary Biology* 4: 2.

Held AA (1981) *Rozella rozellopsis*: Naked endo-parasitic fungi which dress-up as their hosts. *The Botanical Review* 47: 451–515.

Helgason T, Watson IJ, Young JPW (2003) Phylogeny of the Glomerales and Diversisporales (Fungi: Glomeromycota) from actin and elongation factor 1-alpha sequences. *Federation of European Microbiological Societies Microbiological Letters* 229: 127-132.



- Hibbett DS, Binder M, Bischoff JF, Blackwell M, Cannon PF et al. (2007) A higher-level phylogenetic classification of the Fungi. *Mycological research* 111: 509–547.
- Hibbett DS, Ohman A & Kirk PM (2009) Fungal ecology catches fire. *New Phytologist* 184: 279–282.
- Hibbett DS., Ohman A, Glotzer D, Nuhn M, Kirk P & Nilsson RH (2011) Progress in molecular and morphological taxon discovery in Fungi and options for formal classification of environmental sequences. *Fungal Biology Reviews* 25: 38–47.
- Holst-Jensen A, Vrålstad T & Schumacher Y (2004) On reliability. *New Phytologist* 161:11-13
- 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.
- Hyde KD, Jones EBG, Leana E, Pointing SB, Poonyth AD et al. (1998) Role of fungi in marine ecosystems. *Biodiversity and Conservation* 7: 1147–1161.
- Hyde KD, Bussaban B, Paulus B, Crous PW, Lee S et al. (2007) Diversity of saprobic microfungi. *Biodiversity and Conservation* 16: 7–35.
- Ibelings BW, De Bruin A, Kagami M, Rijkeboer M, Brehm M et al. (2004) Host parasite interactions between freshwater phytoplankton and chytrid fungi (Chytridiomycota). *Journal of Phycology* 40: 437–453.
- Imhoff JF, Labes A & Wiese J (2011) Bio-mining the microbial treasures of the ocean: New natural products. *Biotechnology Advances* 29: 468–482.
- Itoh T, Martin W & Nei M (2002) Acceleration of genomic evolution caused by enhanced mutation rate in endocellular symbionts. *Proceedings of the National Academy of Sciences of the United States* 99: 12944–12948.
- James TY, Porter D, Leander CA, Vilgalys R & Joyce E (2000) Molecular phylogenetics of the Chytridiomycota supports the utility of ultrastructural data in chytrid systematics. *Canadian Journal of Botany* 78: 336–350.
- James TY, Kauff F, Schoch CL, Matheny PB, Hofstetter V, et al. (2006a) Reconstructing the early evolution of Fungi using a six-gene phylogeny. *Nature* 443: 818–822.
- James TY, Letcher PM, Longcore JE, Mozley-Standridge SE, Porter D et al. (2006b) A molecular

phylogeny of the flagellated fungi (Chytridiomycota) and description of a new phylum (Blastocladiomycota). *Mycologia* 98: 860–871.

James TY & Berbee ML (2012) No jacket required--new fungal lineage defies dress code: recently described zoosporic fungi lack a cell wall during trophic phase. *BioEssays* 34: 94–102.

Jebaraj CS & Raghukumar C (2009) Anaerobic denitrification in fungi from the coastal marine sediments off Goa, India. *Mycological Research* 113: 100–109.

Jebaraj CS, Raghukumar C, Behnke A & Stoeck T (2010) Fungal diversity in oxygen-depleted regions of the Arabian Sea revealed by targeted environmental sequencing combined with cultivation. *FEMS Microbiology Ecology* 71: 399–412.

Jobard M, Rasconi S & Sime-Ngando T (2010) Diversity and functions of microscopic fungi: a missing component in pelagic food webs. *Aquatic Sciences* 72: 255–268.

Jones EBG (2011) Fifty years of marine mycology. *Fungal Diversity* 50: 73–112.

Jones MDM, Forn I, Gadelha C, Egan MJ, Bass D et al. (2011a) Discovery of novel intermediate forms redefines the fungal tree of life. *Nature* 474: 200–203.

Jones MDM, Richards TA, Hawksworth DL & Bass D (2011b) Validation and justification of the phylum name *Cryptomycota* phyl. nov. *IMA Fungus* 2: 173–175.

Jørgensen BB, Boetius A (2007) Feast and famine – microbial life in the deep-sea bed. *Nature Reviews Microbiology* 5: 770–781.

Kagami M, Bruin A, Ibelings BW & Donk E (2007) Parasitic chytrids: their effects on phytoplankton communities and food-web dynamics. *Hydrobiologia* 578: 113–129.

Kagami M, von Elert E, Ibelings BW, de Bruin A & Van Donk (2009) The parasitic chytrid, *Zygorhizidium*, facilitates the growth of the cladoceran zooplankton, *Daphnia*, in cultures of the inedible alga, *Asterionella*. *Proceedings of the Royal Society B: Biological sciences* 274: 1561–1566.

Kaminuma E, Kosuge T, Kodama Y, Aono H, Mashima J et al. (2011) DDBJ progress report. *Nucleic Acids Research* 39: D22–D27.

Kappe R, Okeke CN, Fauser C, Maiwald M & Sonntag HG (1998) Molecular probes for the detection of pathogenic fungi in the presence of human tissue. *Journal of Medical Microbiology* 47:

811–820.

- Kerry BR (2000) Rhizosphere interactions and three exploitation of microbial agents for the biological control of plant-parasitic nematodes. *Annual Review of Phytopathology* 38:423–441.
- Kilpatrick AM, Briggs CJ & Daszak P (2010) The ecology and impact of chytridiomycosis: an emerging disease of amphibians. *Trends in Ecology and Evolution* 25: 109–118.
- Kim E, Simpson AGB & Graham LE (2006) Evolutionary relationships of Apusomonads inferred from taxon-rich analyses of 6 nuclear encoded genes. *Molecular Biology and Evolution* 23: 2455–2466.
- Kirk PM (2000) World catalogue of 340 K fungal names on-line. *Mycological Research* 104:516–517.
- Kirk PM, Cannon PF, Minter DW & Stalpers JA (2008) *Dictionary of the Mycota*. 10<sup>th</sup> Edition. Cromwell Press, Trowbridge.
- Kohlmeyer J & Kohlmeyer E (1979) *Marine mycology: the higher fungi*, Academic Press London, UK.
- Krings M, Taylor TN & Dotzler N (2011) The fossil record of the Peronosporomycetes (Oomycota). *Mycologia* 103: 455–457.
- Krüger M, Krüger C, Walker C, Stockinger H & Schüssler A (2012) Phylogenetic reference data for systematics and phylotaxonomy of arbuscular mycorrhizal fungi from phylum to species level. *New phytologist* 193: 970–984.
- Küpper FC, Maier I, Muller DG, 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.
- Lai X, Cao L, Tan H, Fang S, Huang Y & Zhou S (2007) Fungal communities from methane hydrate-bearing deep-sea marine sediments in South China Sea. *The ISME journal* 1: 756–762.
- Lang BF, O’Kelly C, Nerad T, Gray MW & Burger G (2002) The closest unicellular relatives of animals. *Current biology* 12: 1773–1778.
- Langmuir CH & Forsyth DW (2007) Mantle Melting Beneath Mid-Ocean Ridge. *Oceanography* 20: 78–89.

- Lara E, Moreira D & López-García P (2010) The environmental clade LKM11 and *Rozella* form the deepest branching clade of fungi. *Protist* 161: 116–121.
- Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, et al. (2007) Clustal W and Clustal X version 2.0. *Bioinformatics* 23: 2947–2948.
- Le Calvez T, Burgaud G, Mahé S, Barbier G & Vandenkoornhuyse P (2009) Fungal diversity in deep-sea hydrothermal ecosystems. *Applied and Environmental Microbiology* 75: 6415–6421.
- Lefèvre E, Bardot C, Noël C, Carrias J-F, Viscogliosi E et al. (2007) Unveiling fungal zooflagellates as members of freshwater picoeukaryotes: evidence from a molecular diversity study in a deep meromictic lake. *Environmental Microbiology* 9: 61–71.
- Lefèvre E, Roussel B, Amblard C & Sime-Ngando T (2008) The molecular diversity of freshwater picoeukaryotes reveals high occurrence of putative parasitoids in the plankton. *PloS one* 3: e2324.
- Lefranc M, Thénot A, Lepère C & Debroas D (2005) Genetic Diversity of Small Eukaryotes in Lakes Differing by Their Trophic Status. *Applied Environmental Microbiology* 71: 5935.
- Lenardon MD, Munro CA & Gow NAR (2010) Chitin synthesis and fungal pathogenesis. *Current Opinion in Microbiology* 13: 416–423.
- Lepère Cecile, Masquelier S, Mangot J-F, Debroas D & Domaizon I (2010) Vertical structure of small eukaryotes in three lakes that differ by their trophic status: a quantitative approach. *The ISME journal* 4: 1509–1519.
- Lesaulnier C, Papamichall D, McCorkle S, Ollivier B, Skiena S et al. (2008) Elevated atmospheric CO<sub>2</sub> affects soil microbial diversity associated with trembling aspen. *Environmental Microbiology* 10: 926–941.
- Letcher PM & Powell MJ (2002) A taxonomic summary of Chytriomycetes (Chytridiomycota). *Mycotaxon* 84: 447-487
- Letcher PM, Powell MJ, Chambers JG, Longcore JE, Churchill PF et al. (2005) Ultrastructural and molecular delineation of the Chytridiaceae (Chytridiales). *Canadian Journal of Botany* 83: 1561-1573.
- Letcher PM, Powell MJ, Churchill PF & Chambers JG (2006) Ultrastructural and molecular

phylogenetic delineation of a new order, the Rhizophydiales (Chytridiomycota). *Mycological Research* 110: 898-915.

Letcher PM, Powell MJ, Barr DJS, Churchill PF, Wakefield WS et al. (2008) Rhizophlyctidales—a new order in Chytridiomycota. *Mycological Research* 112: 1031-1048.

Liggenstoffer AS, Youssef NH, Couger MB & Elshashed MS (2010) Phylogenetic diversity and community structure of anaerobic gut fungi (phylum Neocallimastigomycota) in ruminant and non-ruminant herbivores. *The ISME Journal* 4: 1225-1235.

Liu YJ, Hodson MC & Hall BD (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.

Liu YJ, Steenkamp ET, Brinkmann H, Forget L, Philippe H & Lang BF (2009) Phylogenomic analyses predict sistergroup relationship of nucleariids and fungi and paraphyly of zygomycetes with significant support. *BMC Evolutionary Biology* 9: 272.

López-García P, Rodríguez-Valera F, Pedrós-Alió C & Moreira D (2001) Unexpected diversity of small eukaryotes in deep-sea Antarctic plankton. *Nature* 409: 603–607.

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 of the United States of America* 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.

Lutzoni F, Kauff F, Cox CJ, McLaughlin D, Celio G et al. (2004) Assembling the fungal tree of life: progress, classification, and evolution of subcellular traits. *American Journal of Botany* 91: 1446-1480

Massana R, Balagué V, Guillou L & Pedrós-Alió C (2004) Picoeukaryotic diversity in an oligotrophic coastal site studied by molecular and culturing approaches. *FEMS Microbiology Ecology* 50: 231–243.

- Massana R & Pedrós-Alió C (2008) Unveiling new microbial eukaryotes in the surface ocean. *Current Opinion in Microbiology* 11: 213–218.
- McLaughlin DJ, Hibbett DS, Lutzoni F, Spatafora JW & Vilgalys R (2009) The search for the fungal tree of life. *Trends in Microbiology* 17: 488–497.
- McNeill J, Barrie FR, Burdet HM, Demoulin V, Hawksworth *et al.* (eds.) (2006) International Code of Botanical Nomenclature (Vienna Code): adopted by the Seventeenth International Botanical Congress Vienna, Austria, July 2005. *Regnum Vegetabile* 146, A.R.G. Gantner Verlag KG, Ruggell, Liechtenstein.: [i]-xviii, 1-568.
- Moreira D and López-García P (2003) Are hydrothermal vents oases for parasitic protists? *TRENDS in Parasitology* 19:556-558.
- Moreira D, von der Heyden S, Bass D, López-García P, Chao E *et al.* (2007) Global eukaryote phylogeny: Combined small- and large-subunit ribosomal DNA trees support monophyly of Rhizaria, Retaria and Excavata. *Molecular Phylogenetics and Evolution* 44: 255–266.
- Mozley-Standridge SE, Letcher PM, Longcore JE, Porter D & Simmons DR (2009) Cladochytriales—a new order in the Chytridiomycota. *Mycological Research* 113:498–507.
- Mueller GM & Schmit JP (2007) Fungal biodiversity: what do we know? What can we predict? *Biodiversity and Conservation* 16: 1–5.
- Müller DG, Küpper FC and Küpper H (1999) Infection experiments reveal broad host ranges of *Eurychasma dicksonii* (Oomycota) and *Chytridium polysiphoniae* (Chytridiomycota), two eukaryotic parasites in marine brown algae (Phaeophyceae) *Phycological Research* 47:217-223.
- Nagahama T, Hamamoto M, Nakase T, Takami H & Horikoshi K (2001) Distribution and identification of red yeasts in deep-sea environments around the northwest Pacific Ocean. *Antonie van Leeuwenhoek* 80: 101–110.
- Nagahama T, Takahashi E, Nagano Y, Abdel-Wahab MA & Miyazaki M (2011) Molecular evidence that deep-branching fungi are major fungal components in deep-sea methane cold-seep sediments. *Environmental Microbiology* 13: 2359–2370.
- Nagano Y, Nagahama T, Hatada Y, Nunoura T, Takami H *et al.* (2010) Fungal diversity in deep-sea

sediments – the presence of novel fungal groups. *Fungal Ecology* 3: 316–325.

Nilsson RH, Ryberg M, Kristiansson E, Abarenkov K, Larsson KH et al. (2006) Taxonomic reliability of DNA sequences in public sequence databases: a fungal perspective. *PLoS One* 1: e59.

Norvell LL, Hawksworth DL., Petersen RH & Redhead SA (2010) IMC9 Edinburgh Nomenclature Sessions. *IMA Fungus* 1: 143–147.

Nyvall P, Pedersen M & Longcore JE (1999) *Thalassochytrium gracilariopsisidis* (Chytridiomycota), gen. et sp. nov., endosymbiotic in *Gracilariopsis* sp. (Rhodophyceae). *Journal of Phycology* 35: 176-185.

O'Brien HE, Parrent JL, Jackson JA, Moncalvo J-M & Vylgalis R (2005) Fungal community analysis by large-scale sequencing of environmental samples. *Applied and Environmental Microbiology* 71: 5544-5550.

Öpik M, Vanatoa A, Vanatoa E, Moora M, Davison J et al. (2010) The online database MaarjAM reveals global and ecosystemic distribution patterns in arbuscular mycorrhizal fungi (Glomeromycota). *New Phytologist* 188: 223-241.

Padovan ACB, Sanson GFO, Brunstein A & Briones MRS (2005) Fungi evolution revisited: application of the penalized likelihood method to a Bayesian fungal phylogeny provides a new perspective on phylogenetic relationships and divergence dates of Ascomycota groups. *Journal of Molecular Evolution* 60: 726–735.

Park S-J, Park B-J, Pham VH, Yoon D-N, Kim S-K et al. (2008) Microeukaryotic Diversity in Marine Environments, an Analysis of Surface Layer Sediments from the East Sea. *The Journal of Microbiology* 46: 244-249.

Parniske M (2008) Arbuscular mycorrhiza: the mother of plant root endosymbioses. *Nature reviews. Microbiology* 6: 763–775.

Patterson DJ, Cooper J, Kirk PM, Pyle RL & Remsen DP (2010) Names are key to the big new biology. *Trends in Ecology & Evolution* 25: 686–691.

Peay KG, Kennedy PG & Bruns TD (2008) Fungal community ecology: A hybrid beast with a molecular master. *BioScience* 58: 799–810.

- Philippe H & Laurent J (1998) How good are deep phylogenetic trees? *Current Opinion in Genetics and Development* 8: 616–623.
- Porter TM, Schadt CW, Rizvi L, Martin AP, Schmidt SK et al. (2008) Widespread occurrence and phylogenetic placement of a soil clone group adds a prominent new branch to the fungal tree of life. *Molecular Phylogenetics and Evolution* 46: 635–644.
- Posada D & Crandall KA (1998) MODELTEST: testing the model of DNA substitution. *Bioinformatic Applications Note* 14: 817-818.
- Powell MJ (1993) Looking at mycology with a Janus face – A glimpse at Chtridiomycetes active in the environment. *Mycologia* 85: 1-20.
- Pruesse E, Quast C, Knittel K, Fuchs BM, Ludwig W et al. (2007) SILVA: a comprehensive online resource for quality checked and aligned ribosomal RNA sequence data compatible with ARB. *Nucleic Acids Research* 35: 7188–7196.
- Raghukumar C (2000) Fungi from marine habitats: an application in bioremediation. *Mycological research* 104: 1222-1226
- Raghukumar C, Raghukumar S, Sheelu G, Gupta SM, Nagender Nath B et al. (2004) Buried in time: culturable fungi in a deep-sea sediment core from the Chagos Trench, Indian Ocean. *Deep Sea Research Part I: Oceanographic Research Papers* 51: 1759–1768.
- Ragon M, Restoux G, Moreira D, Møller AP, López-García P (2011) Sunlight-Exposed Biofilm Microbial Communities Are Naturally Resistant to Chernobyl Ionizing-Radiation Levels. *PLoS ONE* 6: e21764.
- Ramirez-Llodra E, Shank TM & German CR (2007) Biodiversity and Biogeography of Hydrothermal Vent Species. *Oceanography* 20:30-41.
- Redecker D, Kodner R & Graham LE (2000) Glomalean fungi from the Ordovician. *Science* 289: 1920-1921
- Redecker D (2002) New views on fungal evolution based on DNA markers and the fossil record. *Research in Microbiology* 153: 125–130.
- Remy W, Taylor TN, Hass H & Kerp H (1994) Four hundred-million-year-old vesicular arbuscular mycorrhizae. *Proceedings of the National Academy of Sciences of the United States of*



*America* 91: 11841–11843.

Richards TA & Bass D (2005) Molecular screening of free-living microbial eukaryotes: diversity and distribution using a meta-analysis. *Current Opinion in Microbiology* 8: 240–252.

Richards TA., Jones MDM., Leonard G & Bass D (2012) Marine fungi: their ecology and molecular diversity. *Annual Review of Marine Science* 4: 495–522.

Ronquist, F & Huelsenbeck, JP (2003) MrBayes3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* 19: 1572–1574.

Rosling A, Cox F, Cruz-Martinez K, Ihrmark K, Grelet G-A et al. (2011) Archaeorhizomycetes: Unearthing an ancient class of ubiquitous soil fungi. *Science* 333: 876–879.

Ruiz-Herrera J, González-Prieto JM & Ruiz-Medrano R (2002) Evolution and phylogenetic relationships of chitin synthases from yeasts and fungi. *FEMS Yeast Research* 1: 247–256.

Ruiz-Trillo I, Inagaki Y, Davis LA, Sperstad S, Landfald B et al. (2004) *Capsaspora owczarzaki* is an independent opisthokont lineage. *Current Biology* 14: R946-R947.

Saitou N. & Nei, N (1987) The Neighbor Joining method: a new method for reconstructing phylogenetic trees. *Molecular Biology and Evolution* 4: 293-298.

Sanderson MJ (2003) Molecular data from 27 proteins do not support a Precambrian origin of land plants. *American Journal of Botany* 90: 954-956.

Schadt CW, Martin AP, Lipson DA & Schmidt SK (2003) Seasonal dynamics of previously unknown fungal lineages in tundra soils. *Science* 301: 1359–1361.

Schmit JP & Mueller GM (2007) An estimate of the lower limit of global fungal diversity. *Biodiversity and Conservation* 16: 99–111.

Schüßler A, Schwarzott D & Walker C (2001) A new fungal phylum, the Glomeromycota: phylogeny and evolution. *Mycological Research* 105: 1413-1421.

Selosse MA & Le Tacon F (1998) The land flora: a phototroph-fungus partnership? *Trends in Ecology & Evolution* 13: 15–20.

Shearer CA, Descals E, Kohlmeyer B, Kohlmeyer J, Marvanova L et al. (2007) Fungal biodiversity in aquatic habitats. *Biodiversity and Conservation* 16: 49-67.

- Sime-Ngando T, Lefèvre E & Gleason FH (2011) Hidden diversity among aquatic heterotrophic flagellates: ecological potentials of zoosporic fungi. *Hydrobiologia* 659: 5-22.
- Simmons DR, James TY, Meyer AF & Longcore JE (2009) Lobulomycetales, a new order in the Chytridiomycota. *Mycological Research* 113: 450-460.
- Simon L, Bousquet J, Levesque RC & Lalonde M (1993) Origin and diversification of endomycorrhizal fungi and coincidence with vascular land plants. *Nature* 363: 67-69.
- Simonato F, Campanaro S, Lauro FM, Vezzi A, D'Angelo M et al. (2006) Piezophilic adaptation: a genomic point of view. *Journal of Biotechnology* 126: 11–25.
- Singh P, Raghukumar C, Verma P & Shouche Y (2010) Phylogenetic diversity of culturable fungi from the deep-sea sediments of the Central Indian Basin and their growth characteristics. *Fungal Diversity* 40: 89–102.
- Smith SE, Facelli E, Pope S & Smith FA (2010) Plant performance in stressful environments: interpreting new and established knowledge of the rôles of arbuscular mycorrhizas. *Plant and Soil* 326: 3-20.
- Stajich JE, Dietrich FS & Roy SW (2007) Comparative genomic analysis of fungal genomes reveals intron-rich ancestors. *Genome Biology* 8: R223.
- Stajich JE, Berbee ML, Blackwell M, Hibbett DS, James TY et al. (2009) The fungi. *Current Biology* 19: R840–R845.
- Stechmann A & Cavalier-Smith T (2003) Phylogenetic analysis of eukaryotes using heat-shock protein Hsp90. *Journal of Molecular Evolution* 57: 408–419.
- Steenkamp ET, Wright J & Baldauf SL (2006) The protistan origins of animals and fungi. *Molecular Biology and Evolution* 23: 93–106.
- Stoeck T & Epstein S (2003) Novel Eukaryotic Lineages Inferred from Small-Subunit rRNA Analyses of Oxygen-Depleted Marine Environments. *Applied and Environmental Microbiology* 69: 2657-2663.
- Stoeck T, Taylor GT & Epstein SS (2003) Novel Eukaryotes from the Permanently Anoxic Cariaco Basin (Caribbean Sea). *Applied and Environmental Microbiology* 69: 5656-5663.

- Stoeck T, Hayward B, Taylor GT, Varela R & Epstein SS (2006) A multiple PCR-primer approach to access the microeukaryotic diversity in environmental samples. *Protist* 157: 31–43.
- Stoeck T, Bass D, Nebel M, Christen R, Jones MDM et al. (2010) Multiple marker parallel tag environmental DNA sequencing reveals a highly complex eukaryotic community in marine anoxic water. *Molecular Ecology* 19 Suppl 1: 21–31.
- Takishita K, Tsuchiya M, Reimer JD & 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.
- Takishita K, Yubuki N, Kakizoe N, Inagaki Y & Maruyama T (2007) Diversity of microbial eukaryotes in sediment at a deep-sea methane cold seep: surveys of ribosomal DNA libraries from raw sediment samples and two enrichment cultures. *Extremophiles* 11: 563–576.
- Takishita K, Kakizoe N, Yoshida T & Maruyama T (2010) Molecular Evidence that Phylogenetically Diverged Ciliates Are Active in Microbial Mats of Deep-Sea Cold-Seep Sediment. *Journal Eukaryotic Microbiology* 57: 76-86.
- Tanabe Y, Saikawa M, Watanabe MM & Sugiyama J (2004) Molecular phylogeny of Zygomycota based on EF-1 $\alpha$  and RPB1 sequences: limitations and utility of alternative markers to rDNA. *Molecular Phylogenetics and Evolution* 30: 438–449.
- Tanabe Y, Watanabe MM & Sugiyama J (2005) Evolutionary relationships among basal fungi (Chytridiomycota and Zygomycota): Insights from molecular phylogenetics. *Journal of Genetic Applied Microbiology* 51: 267-276.
- Taylor TN, Remy W & Hass H (1992) Fungi from the lower Devonian Rhynie chert: Chytridiomycetes. *American Journal of Botany* 79: 1233-1241.
- Taylor TN, Hass H & Kerp H (1997) A cyanolichen from the lower Devonian Rhynie chert. *American Journal of Botany* 84: 992-1004.
- Taylor TN, Hass H & Kerp H (1999) The oldest fossil Ascomycetes. *Nature* 399: 648.
- Taylor JW & Berbee ML (2006) Dating divergences in the Fungal Tree of Life: review and new analyses. *Mycologia* 98: 838–849.
- Thaler AD, Van Dover CL & Vilgalys R (2012) Ascomycete phylotypes recovered from a Gulf of

- Mexico methane seep are identical to an uncultured deep-sea fungal clade from the Pacific. *Fungal Ecology* 5: 270–273.
- Van Dover CL, Ward ME, Scott JL, Underdown J, Anderson B et al. (2007) A fungal epizootic in mussels at a deep-sea hydrothermal vent. *Marine Ecology* 28: 54–62.
- Vandenkoornhuyse P, Baldauf SL, Leyval C, Straczek J & Young JPW (2002) Extensive fungal diversity in plant roots. *Science* 295: 2051.
- Vilgalys R (2003) Taxonomic misidentification in public DNA databases. *New Phytologist* 160: 4-5.
- Voigt K & Kirk PM (2011) Recent developments in the taxonomic affiliation and phylogenetic positioning of fungi: impact in applied microbiology and environmental biotechnology. *Applied Microbiology and Biotechnology* 90: 41–57.
- Wainright PO, Hinkle G, Sogin ML & Stickel SK (1993) Monophyletic origins of the Metazoa: an evolutionary link with fungi. *Science* 260: 340–342.
- Wardle DA & Van der Putten WH (2002) Biodiversity, ecosystem functioning and above-ground-below-ground in *Biodiversity and ecosystem functioning: synthesis and perspectives*. Loreau M, Naeem S & Inchausti P, eds. Oxford University Press.
- White MM, James TY, O'Donnell K, Cafaro MJ, Tanabe Y & Sugiyama J (2006) Phylogeny of the Zygomycota based on nuclear ribosomal sequence data. *Mycologia* 98: 872–884.
- Wuyts J, De Rijk P, van de Peer Y, Pison G, Rousseeuw P et al. (2000) Comparative analysis of more than 3000 sequences reveals the existence of two pseudoknots in area V4 of eukaryotic small subunit ribosomal RNA. *Nucleic Acids Research* 28: 4698–4708.
- Zuendorf A, Bunge J, Behnke A, Barger KJA & Stoeck T (2006) Diversity estimates of microeukaryotes below the chemocline of the anoxic Mariager Fjord, Denmark. *FEMS Microbiology Ecology* 58: 476–491.
- Zuker M (2003) Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Research* 31: 3406–3415.



# ANNEXES



# Active root-inhabiting microbes identified by rapid incorporation of plant-derived carbon into RNA

Philippe Vandenkoornhuys<sup>\*</sup>, Stéphane Mahé<sup>\*</sup>, Philip Ineson<sup>†</sup>, Phil Staddon<sup>†</sup>, Nick Ostle<sup>‡</sup>, Jean-Bernard Cliquet<sup>§</sup>, André-Jean Francez<sup>\*</sup>, Alastair H. Fitter<sup>†</sup>, and J. Peter W. Young<sup>†¶</sup>

<sup>\*</sup>Centre National de la Recherche Scientifique, Unité Mixte de Recherche (UMR) 6553 EcoBio, IFR90/FR2116, Centre Armoricaire de Recherche sur l'Environnement, Université de Rennes I, Campus de Beaulieu, 35042 Rennes Cedex, France; <sup>†</sup>Department of Biology, University of York, P.O. Box 373, York YO10 5YW, United Kingdom; <sup>‡</sup>Centre for Ecology and Hydrology, Lancaster Environment Centre, Bailrigg, Lancaster LA1 4AP, United Kingdom; and <sup>§</sup>UMR 950 Laboratoire d'Écophysiologie Végétale, Agronomie et nutrition, Université de Caen, Esplanade de la Paix, 14032 Caen Cedex, France

Edited by James M. Tiedje, Michigan State University, East Lansing, MI, and approved September 5, 2007 (received for review June 22, 2007)

Plant roots harbor a large diversity of microorganisms that have an essential role in ecosystem functioning. To better understand the level of intimacy of root-inhabiting microbes such as arbuscular mycorrhizal fungi and bacteria, we provided <sup>13</sup>C<sub>2</sub> to plants at atmospheric concentration during a 5-h pulse. We expected microbes dependent on a carbon flux from their host plant to become rapidly labeled. We showed that a wide variety of microbes occurred in roots, mostly previously unknown. Strikingly, the greatest part of this unsuspected diversity corresponded to active primary consumers. We found 17 bacterial phylotypes co-occurring within roots of a single plant, including five potentially new phylotypes. Fourteen phylotypes were heavily labeled with the <sup>13</sup>C. Eight were phylogenetically close to Burkholderiales, which encompass known symbionts; the others were potentially new bacterial root symbionts. By analyzing unlabeled and <sup>13</sup>C-enriched RNAs, we demonstrated differential activity in C consumption among these root-inhabiting microbes. Arbuscular mycorrhizal fungal RNAs were heavily labeled, confirming the high carbon flux from the plant to the fungal compartment, but some of the fungi present appeared to be much more active than others. The results presented here reveal the possibility of uncharacterized root symbioses.

ribosomal RNA | stable isotope probing | symbiosis | arbuscular mycorrhiza | endophytes

Plants are the dominant primary producers in most terrestrial ecosystems. In the soil, they are escorted by a myriad of microorganisms living freely or in intimate interaction with their roots (1, 2). These microorganisms can be pathogenic, parasitic, saprotrophic, or mutualistic. Among the root symbionts, arbuscular mycorrhizal (AM) fungi are well known and have been observed colonizing the roots of most plant species in many ecosystems (3). Recent studies show that high diversity is the norm even where plant diversity is low (4–6). These AM fungi are biotrophs, unable to grow in the absence of a living plant, and often display a broad host range although there is growing evidence for differences in host preference (5–8). They have been demonstrated to improve plant mineral nutrition (3) and stress resistance (3). AM fungi are important for the global carbon cycle because up to 20% of photoassimilates can be translocated to them (9). We also know that the diversity of AM fungi can determine plant community structure and ecosystem productivity (10). The plant–bacteria symbioses are variably documented. The best studied symbiosis is the rhizobium–legume interaction, but a single plant root can harbor a large variety of fungi (1) and bacteria (2), as well as several different archaea (2). So far we have no information about the functions of most of these root-living microbes. The strategy chosen herein, stable isotope probing (SIP)–RNA analysis, enabled us to highlight an unsuspected diversity of microbes living in roots. We identified microbes that are active and direct utilizers of

photosynthetic carbon from the plant by demonstrating a differential carbon flow to them.

## Results and Discussion

Several methods have been developed recently to analyze the functional diversity of microorganisms *in situ* without preliminary cultivation or isolation. Here we use SIP–RNA (11) based on the fractionation of heavily labeled [<sup>13</sup>C]RNAs from a mixture after providing CO<sub>2</sub> enriched in the stable isotope <sup>13</sup>C. The enriched RNAs can be assigned to microbial taxon by analyzing the small subunit ribosomal RNA (SSU rRNA). RNA becomes labeled more rapidly and heavily than DNA because ribosomes are more abundant and turn over faster than DNA, and transcription is not semiconservative. Hence, SIP–RNA allows analysis of shorter term responses than SIP–DNA (i.e., primary consumers are targeted before the label can reach secondary consumers).

We used a pulse of <sup>13</sup>CO<sub>2</sub> to label turfs lifted from an upland grassland in the United Kingdom (experiment A) and a regenerating peatland in France (experiment B). The vegetation included grasses (predominantly *Agrostis capillaris*, *Festuca rubra*, *Poa pratensis*) and white clover (*Trifolium repens*) for the upland grassland and *Agrostis stolonifera*, *Eriophorum angustifolium*, and *Hydrocotyle vulgaris* for the peatland. The purpose of experiment A was to identify and compare the AM fungal communities associated with the co-occurring species, *A. capillaris* and *T. repens*, and to characterize their behavior with regard to the carbon flux from the plant. Experiment B surveyed a wider range of microorganisms (bacteria as well as AM fungi) associated with the roots of *A. stolonifera*. To focus on the primary consumers of current photosynthates, the time of exposure to <sup>13</sup>C assimilates must be reduced to a minimum. In a previous field experiment that provided a <sup>13</sup>CO<sub>2</sub> pulse at atmospheric concentration to plants, the maximum enrichment of microbial RNA was reached 3 h after the end of a 6-h pulse (12). Furthermore, microorganisms have been demonstrated to rapidly metabolize

Author contributions: P.V., P.I., A.-J.F., A.H.F., and J.P.W.Y. designed research; P.V., S.M., P.S., and N.O. performed research; N.O. and J.-B.C. contributed new reagents/analytic tools; P.V., S.M., P.I., A.H.F., and J.P.W.Y. analyzed data; and P.V., S.M., and J.P.W.Y. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

Abbreviations: AM, arbuscular mycorrhizal; CSTFA, cesium trifluoroacetate; ML, maximum likelihood; MP, maximum parsimony; NJ, neighbor joining; SIP, stable isotope probing; SSU, small subunit.

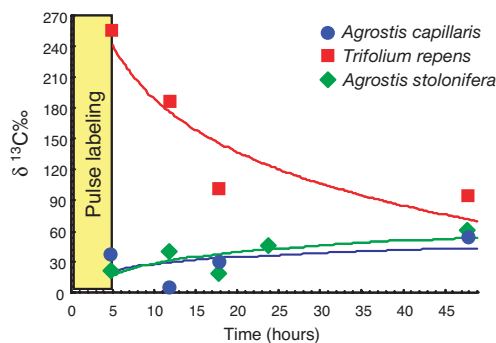
Data deposition: The sequences reported in this paper have been deposited in the GenBank database (accession nos. AF481533–AF481698, experiment A; AY273534–AY273619, experiment B; EF040887–EF041036, bacteria; EF041037–EF041100, AM fungi).

¶To whom correspondence should be addressed. E-mail: jpy1@york.ac.uk.

This article contains supporting information online at [www.pnas.org/cgi/content/full/0705902104/DC1](http://www.pnas.org/cgi/content/full/0705902104/DC1).

© 2007 by The National Academy of Sciences of the USA



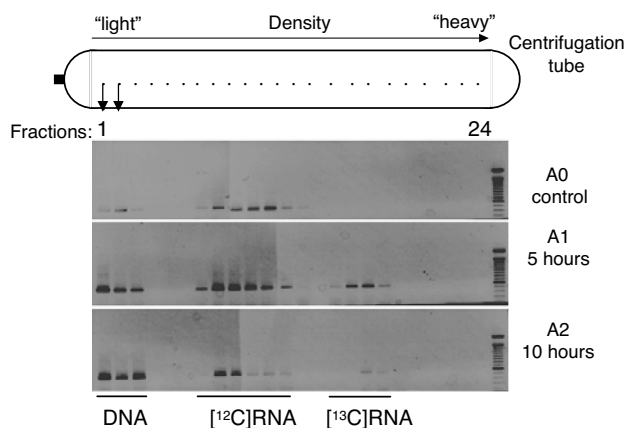


**Fig. 1.** Isotopic signatures ( $\delta^{13}\text{C}\text{‰}$ ) of roots after  $^{13}\text{CO}_2$  pulse labeling. Each dot represents a mean of four measures (i.e., not true replicates). Results of  $\delta^{13}\text{C}\text{‰}$  were obtained for *T. repens* (squares) and *A. capillaris* (circles) in experiment A and from *A. stolonifera* (diamonds) in experiment B. Fitted regression curves are shown. Natural values of  $\delta^{13}\text{C}\text{‰}$  (control before labeling) were  $-35$  for *T. repens* roots,  $-28$  for *A. capillaris*, and  $-30$  for *A. stolonifera*.

$^{13}\text{C}$ -labeled primary metabolites released directly in soil (13). Bearing these results in mind, we reduced the pulse to 5 h with harvesting directly after completion.

The overall isotopic signature ( $\delta^{13}\text{C}$ ) demonstrated that the roots of all three plant species were enriched in  $^{13}\text{C}$  immediately after the end of the pulse labeling (Fig. 1). Carbon translocation was especially rapid to the roots of *T. repens*, consistent with the high AM mycelial carbon respiration that was recently demonstrated in this species (14). It is likely that demand by rhizobia in root nodules also contributed to the early accumulation of  $^{13}\text{C}$  in clover roots. By contrast,  $^{13}\text{C}$  built up in *Agrostis* roots over several days, although our results demonstrated that microbes still exploited a  $^{13}\text{C}$ -rich pool of recent photosynthates.

To identify the AM fungal and bacterial communities colonizing the roots, we used primers specific for the amplification of the SSU rRNA of AM fungi (4) and specific for bacteria (modified from ref. 15). Isopycnic ultracentrifugation of the RNA extracts by using cesium trifluoroacetate (CsTFA) was followed by fractionation of the gradient and RNA precipitation. Direct PCR on each fraction showed that only low-density fractions 1, 2, and 3 gave a positive signal indicating the presence of DNA (data not shown). This DNA was used to assess the overall diversity of AM fungi in the roots for experiment A. From these direct PCR amplifications, it can be concluded that fractions 4–24 did not contain detectable DNA, so any ampli-



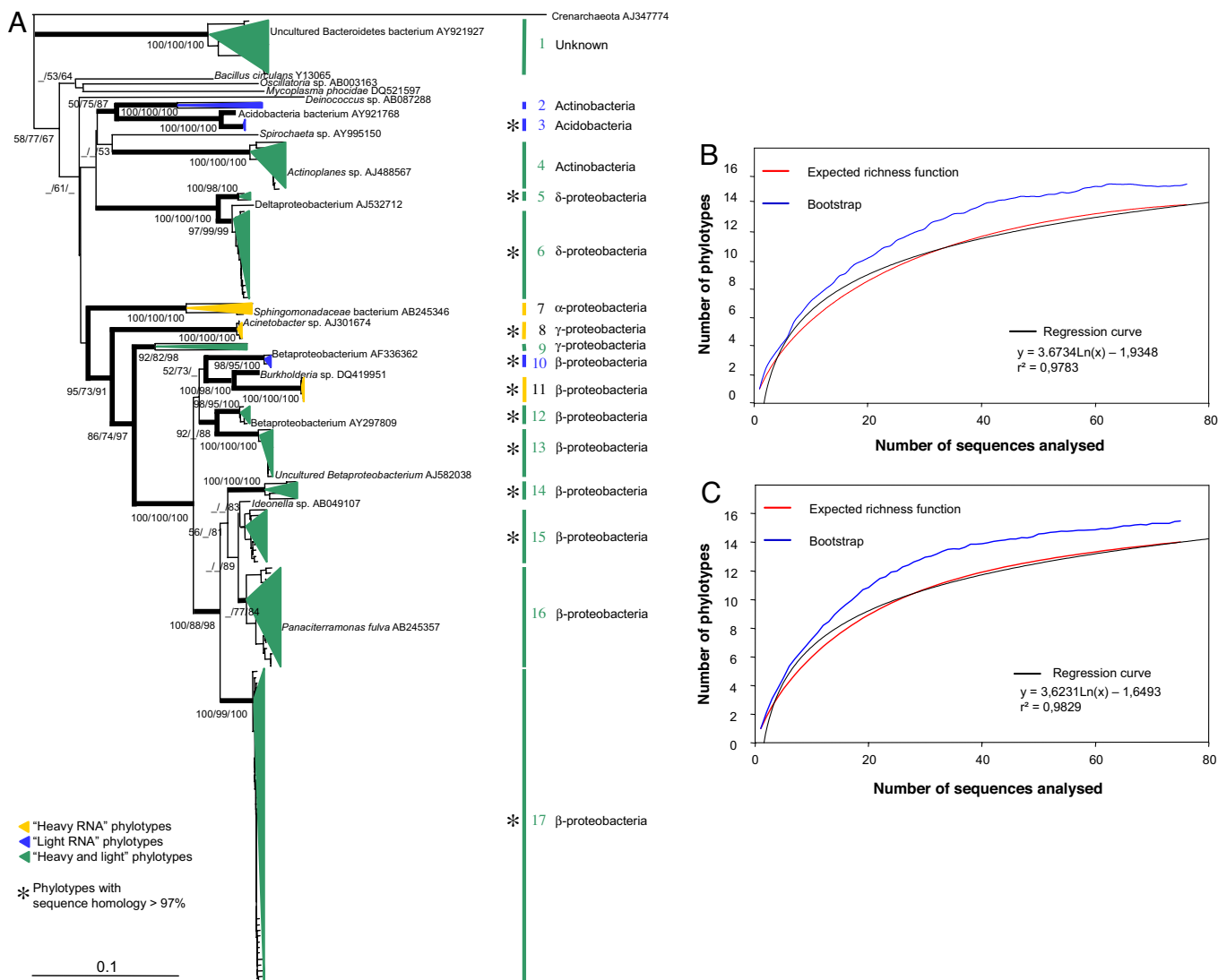
**Fig. 2.** RT-PCR products from each fraction collected from the CsTFA gradient. RNA extracts from *T. repens* were amplified by RT-PCR specific for AM fungi, using samples before (A0), immediately after (A1), and 5 h after (A2) the 5-h  $^{13}\text{CO}_2$  labeling period.

fication in the RT-PCR reflected RNA only. In the unlabeled control, RT-PCR products were seen in fractions 7–13 (Fig. 2). After the  $^{13}\text{CO}_2$  pulse, additional bands appeared in the higher density fractions (fractions 15–18 for A1 and 17–18 for A2) (Fig. 2), which we interpret as being amplified from the  $^{13}\text{C}$ -labeled (heavy) RNA, whereas unlabeled RNA was found in fractions 7–12 (Fig. 2). We argue that light RNA did not contaminate the heavy fractions because no positive clones were recovered from fractions 17 and 18 of the unlabeled control. The gap between the light and heavy RNAs (e.g., A1 fraction 14) (Fig. 2) indicates that the labeling was intense. This finding is supported by estimated buoyant densities of 1.78–1.80 g/ml for the unlabeled and 1.82–1.85 g/ml for the labeled RNA, comparable with values reported for pure  $^{12}\text{C}$ rRNA and pure  $^{13}\text{C}$ rRNA, respectively (16). In addition, there was no delay: heavy RNA was found immediately after the end of the 5-h pulse labeling. Similar results were obtained in experiment B. The main conclusion from these observations is that the AM fungi were preferentially using assimilates provided by plants (labeled molecules), rather than previously fixed carbon (unlabeled). Traces of heavy RNA were still detected 43 h after the end of the  $^{13}\text{CO}_2$  pulse labeling (experiment A) (data not shown). For the bacteria, we found similar results, with positive RT-PCRs and heavily labeled RNA in fraction 17 (heavy RNAs). Compared with the results for the AM fungi, the only difference is a positive signal that is uninterrupted from fractions 5–17 (experiment B) (data not shown). Thus, depending on the bacteria and their ecological status, different proportions of  $^{13}\text{C}$ -labeled photosynthates were incorporated. From an ecological point of view, bacterial symbionts (i.e., root-inhabiting bacteria, whether mutualistic or parasitic) can be expected to be heavily labeled in the same way as the AM fungi, whereas saprotrophic bacteria should receive little direct photosynthate.

Cloning and sequencing of the various PCR products identified 17 phylotypes (phylogenetically related sequences) of AM fungi colonizing the roots (Fig. 3). By computing rarefaction curves using a random resampling procedure and bootstrap (17) from the 25 to 30 sequenced clones from each clone library, we demonstrate that we have not underestimated the diversity of AM fungal phylotypes in any of the RNA samples [supporting information (SI) Fig. 5]. In the case of the DNA samples, however, a greater sampling effort probably would have uncovered additional phylotypes because the corresponding curves have not reached their asymptotes (SI Fig. 5). These results were confirmed by using a Bayesian estimator of the diversity for noninvasive sampling (18). All of the sequences were related to recognized *Glomus*, *Acaulospora*, and *Scutellospora* species, although none was close enough to be assigned to them. Only 4 phylotypes have been reported from other field populations in previous studies, whereas 13 phylotypes (Glo26–Glo29, Glo46, Glo57–Glo60, Acau13–Acau15, and Scut5) were potentially new (Fig. 3). In experiment A, seven phylotypes were found exclusively within the AM fungal community colonizing *A. capillaris* and three for *T. repens*, whereas only three were shared by the two plants (SI Table 1). Despite the co-occurrence of the plant species in the same turf, the AM fungal community composition differed among host plants (SI Table 1), in agreement with previous studies (5–7).

For both plant species in experiment A, the heavy RNA corresponded to a subset of the diversity found in the light RNA or DNA (SI Table 1). In each *T. repens* sample, half of the AM fungal phylotypes were recovered from the heavy RNA and hence had received and metabolized  $^{13}\text{C}$ -labeled assimilates. The active AM fungi were different in the two root samples. *Glomus* Glo8, active in A1, is present in A2, but was not detected in the heavy RNA fraction, whereas *Glomus* Glo3 was apparently active in A2, but inactive in A1. It is likely that these *Glomus*





**Fig. 4.** Phylogenetic affinities of bacterial SSU rRNA sequences and statistical analyses of sampling effort. (A) Tree derived by ML analysis of all 151 bacterial sequences amplified from *A. stolonifera* roots, with representatives of major bacterial groups and the highest BLAST hit of each phylotype. The outgroup is an uncultured crenarchaeote (AJ347774). Branches in bold were congruent among MP, ML, and NJ phylogenetic reconstructions (at least two bootstrap values of >70%). Bootstrap values of >50% are indicated at the nodes (NJ/MP/ML estimated from 1,000/500/100 iterations, respectively). (Scale bar: 0.1 substitutions per site.) Phylotypes are presented as colored triangles proportional to the number of clones found and their phylogenetic depth: yellow, heavy RNA fraction; blue, light RNA fraction; green, sequences found in both RNA fractions. Asterisks indicate phylotypes in which all sequences have >97% identity. (B) Estimates of bacterial community diversity as a function of sampling effort for *A. stolonifera* root samples. Rarefaction curves are computed from 100 replicates of bootstrap (blue), and the expected richness function (red) is the number of phylotypes estimated (random sampling without replacement) from the sequences analyzed for heavy SSU RNA (i.e., <sup>13</sup>C-labeled). (C) Estimates of bacterial community diversity for light SSU RNA.

of host-plant preference and competition among AM fungal colonizers (5, 6, 20).

The *A. stolonifera* root-colonizing bacteria were analyzed from RNA in fraction 7 (unlabeled) and fraction 18 (heavily labeled) of experiment B. The statistical analyses of the bacterial phylotypes allow us to conclude that the sampling was large enough to describe the diversity for both light and heavy fractions (Fig. 4 B and C). The regression curves show that one or two additional phylotypes should be expected if the number of sequences were doubled. The G + C composition was similar for all RNA phylotypes from both light and heavy fractions and thus should not introduce any bias. The phylogenetic reconstructions segregate bacteria into 17 phylotypes (with  $\leq 3\%$  internal sequence divergence, except for 6 phylotypes that were slightly more heterogeneous). Five strongly supported phylotypes cannot be identified because they are too distant from any published

sequences other than from environmental samples (Fig. 4). Phylotype 1 forms a particularly deep branch within the bacteria, whereas phylotypes 10–13 are within the class Betaproteobacteria. We can argue that these bacteria are not parasitic because (i) thorough examination of the plant roots indicated that they were healthy, and (ii) no long branches were observed in the phylogeny (typical of parasites because of higher mutation rate) (21). A number of sequences were close to the Burkholderiales, a group that includes both pathogens and symbionts. Several *Burkholderia* spp. have developed symbioses with plants and are found in roots, leaves, and stems (22). Furthermore, they colonize AM fungi as endosymbionts (23), and several are described as symbionts of *Rhizopus*, a fungus belonging to the Zygomycota, conferring on the fungus a certain level of pathogenicity toward plants (24). Another group of putative Burkholderiales sequences were close to those of nitrogen-fixing *Ideonella* sp. isolated from rice stems and roots (25).



The metagenomic approach using stable isotopes (26) has allowed us to describe the community diversity of AM fungi and bacteria and to relate this diversity to a functional trait, the consumption of newly formed photosynthates. An extension of our approach would probably reveal the presence of additional microorganisms, including archaea and a wider range of fungi (1). Because we have focused on highly labeled RNAs, we conclude that the microorganisms involved in the process are closely linked with their host plant and are likely to be plant-dependent. However, our results bring more questions than answers. We have much to learn about the ecology of microorganisms in roots, especially of the potentially new bacterial groups described herein.

## Materials and Methods

**Field Sites and Plant Growth.** In experiment A, six turfs (45 × 35 cm) were collected from the Natural Environment Research Council Soil Biodiversity field experiment at Sourhope near Kelso (Scotland), a seminatural grassland ecosystem (March 2002). Turfs were placed in plastic containers (45 × 35 cm) and acclimated for 3 months at 18°C lit 12 h per day before labeling.

In experiment B, three turfs of the same size were collected from a peatland experiment of the European research program RECIPE, located close to Baupte in Normandy (France) and acclimated similarly (March 2005).

**<sup>13</sup>C Labeling and Total RNA Extraction.** In experiment A, a turf in its container was put into an acrylic air-flow chamber and labeled with <sup>13</sup>CO<sub>2</sub> (99 atom %) at atmospheric concentration. Air flow to the turf (≈5 liters/min) and CO<sub>2</sub> delivery were controlled by measuring the concentration of CO<sub>2</sub> in the vent gas, using an infrared gas analyzer. Core samples (4.1-cm diameter) were taken from the turf: A0, before the pulse labeling; A1, immediately after the 5-h labeling; and A2, A3, A4, and A5, 10, 16, 24, and 48 h after the start of the labeling, respectively. From each core, roots of *T. repens* and *A. capillaris* were immediately washed in tap water, three times in 0.1% Triton X-100, and five times in sterilized distilled water and then frozen. All roots appeared to be living and healthy (checked under binocular microscope). Total RNA was extracted from ≈30 mg (fresh weight) by using the RNeasy Plant Mini Kit (Qiagen, Valencia, CA). Experiment B was conducted similarly, using roots of *A. stolonifera*.

**Isotopic Signature.** A sample of the remaining roots was dried. The <sup>13</sup>C enrichment in root samples reported in δ<sup>13</sup>C‰ (δ<sup>13</sup>C: <sup>12</sup>C ratio) was calibrated relative to internal gas standards and solid reference against Pee Dee Belemnite standard. All isotopic signatures (δ<sup>13</sup>C) were determined by continuous-flow/combustion/isotope ratio mass spectrometry. Although the intensity of labeling might vary with root age, samples were large and representative enough to even out any such effect. Exponential regression curves were estimated by using Statistica 7.1.

**Separation of Labeled RNA from Unlabeled RNA.** Before the experiment, the time, speed, and temperature of the ultracentrifugation and quantities of CsTFA and RNA were optimized to separate [<sup>13</sup>C]RNA from [<sup>12</sup>C]RNA. The isopycnic ultracentrifugation was performed using Quick Seal centrifugation tubes (Beckman, Fullerton, CA) filled with 13.5 ml of an aqueous solution of 1.9 g/ml CsTFA (Amersham Biosciences, Piscataway, NJ), to which 50 ng of RNA was added (optimal quantity). Tubes were spun at 48,000 rpm in a 90Ti rotor (Beckman) at 4°C for 48 h. We noticed that RNA degradation occurs (i.e., RNA is broken) when tubes are spun at speeds >55,000 rpm. Twenty-four fractions of the ultracentrifugation gradient were taken, starting from the top of the tube, by puncturing the wall every 0.1 in (2.54 mm), using a needle, syringe, and guide. RNA was precipitated with two volumes of isopropanol. After centrifugation, the pellet was washed in 75% ethanol, dried, and redissolved

in 20 μl of ultrapure DNase- and RNase-free water (Sigma–Aldrich, St. Louis, MO). Densities of fractions from two gradients run under the same conditions as the experiments, but without nucleic acid, were estimated by weighing measured volumes.

**RT-PCR.** RT-PCR was carried out on 5 μl of each of the 24 fractions from each sample (A0–5, B0–1) by using the Titan One Tube RT-PCR kit (Roche Molecular Systems, Alameda, CA), with primers NS31 and AM1 for the specific amplification of AM fungal SSU rDNA (4) and (for experiment B) the primers Eub\_519f and Eub\_1390r (specific for bacteria (15) in the modified forms 5'-GTTTCAGCMGCCGCGGT-3' and 5'-GTTTGACGGGGT-GTGT-3', respectively. For AM fungal sequences, reverse transcription at 51°C for 30 min was followed by PCR amplification with 2 min denaturation at 94°C, 35 cycles of 30 sec at 94°C, 1 min at 58°C or 52°C (AM fungi or bacteria, respectively), 50 sec at 68°C, and final extension for 10 min at 68°C.

**Cloning and Sequencing.** The RT-PCR yield was low for the later samples (A3–5), so analysis was limited to A1, A2, and B1. The amplified fragments were cloned in pGEM-T vector (Promega, Madison, WI) and DH5α-competent cells (Gibco/BRL, Carlsbad, CA). From each clone library, 25 to 30 randomly selected positive clones were sequenced (ABI-PRISM Dye Terminator Cycle Sequencing Ready Reaction Kit; PerkinElmer, Beltsville, MD) for both strands by using primers T7 or SP6. The two sequences were aligned by using the programs Autoassembler (PerkinElmer) and Sequencher (GeneCodes, Ann Arbor, MI) for experiments A and B, respectively. No chimeric artifacts were found among the clones sequenced of AM fungi, whereas 53 artifacts were detected and deleted from the data set among 204 bacteria sequences by using CHIMERA-CHECK 2.7 (Ribosomal Database Project II; <http://rdp.cme.msu.edu>).

**Diversity and Phylogenetic Analyses.** Rarefaction curves were computed for each data set. The number of species was quantified for 100 random combinations of 1 to *N* sequences and also by performing 100 bootstrap pseudoreplicates implemented in Estimates (17).

Multiple alignments, one for the AM fungal phylogenetic analyses containing all of the sequences plus a set of eight AM fungal sequences from GenBank/EMBL/DBJ and one for the 151 bacterial sequences, were performed by using CLUSTALX 1.81 (27) and refined by eye. For the two data sets, the phylogenetic analyses were done as follows: CLUSTALX 1.81 was used for neighbor joining (NJ) phylogenies, with distance correction using K2P and complete omission of gaps. For AM fungi and bacteria, bootstrapping was repeated 200 and 1,000 times, respectively. PAUP 4.0β10 was used for maximum parsimony (MP) by using a heuristic tree search with 500 replicates (for AM fungi, 300 for bacteria) of random addition, tree bisection and reconnection as branching algorithm, and 15 random-addition swaps per replicate. For the bacteria, an identical likelihood score was found for the constructions by using tree bisection and reconnection or subtree pruning-grafting as swapping algorithms. Then, the bacterial set was studied by using maximum-likelihood (ML) procedures under the HKY model with 200 bootstrap iterations. For the AM fungal data set, the ML phylogeny was constructed under the GTR + I + G model (100 bootstrap iterations). Modeltest 3.7 software (28) was used to select the model.

We thank Ouest Génopole and Allen Mould for the sequencing and S. Diquelou and S. Lemauiel for labeling of experiment B. This work was supported by the Soil Biodiversity Thematic Program of the Natural Environment Research Council and the Ecosphère Continentale–Fonctionnement et Dynamique de la Biosphère Continentale Program of the Centre National de la Recherche Scientifique.

1. Vandenkoornhuysen P, Baldauf SL, Leyval C, Straczek J, Young JPW (2002) *Science* 295:2051.
2. Chelius MK, Triplett EW (2001) *Microb Ecol* 41:252–263.
3. Smith SE, Read DJ (1997) *Mycorrhizal Symbiosis* (Academic, San Diego), 2nd Ed.
4. Helgason T, Daniell TJ, Husband R, Fitter AH, Young JPW (1998) *Nature* 394:431.
5. Vandenkoornhuysen P, Husband R, Daniell TJ, Watson IJ, Duck JM, Fitter AH, Young JPW (2002) *Mol Ecol* 11:1555–1564.
6. Vandenkoornhuysen P, Ridgway KP, Watson IJ, Fitter AH, Young JPW (2003) *Mol Ecol* 12:3085–3095.
7. Husband R, Herre EA, Turner SL, Gallery R, Young JPW (2002) *Mol Ecol* 11:2669–2678.
8. Gollote A, van Tuinen D, Atkinson D (2004) *Mycorrhiza* 14:111–117.
9. Bago B, Pfeffer PE, Shachar-Hill Y (2000) *Plant Physiol* 124:949–957.
10. van der Heijden MGA, Klironomos JN, Ursic M, Moutoglou M, Streitwolf-Engel R, Boller T, Wiemken A, Sanders I (1998) *Nature* 396:69–72.
11. Manefield M, Whiteley AS, Griffiths RI, Bailey MJ (2002) *Appl Environ Microbiol* 68:5367–5373.
12. Rangel-Castro I, Killham K, Ostle N, Nicol GW, Anderson IC, Scrimgeour CM, Ineson P, Meharg A, Prosser J (2005) *Environ Microbiol* 7:828–838.
13. Padmanabhan P, Padmanabhan S, Derito C, Gray A, Gannon D, Snape JR, Tsai CS, Park W, Jeon C, Madsen EL (2003) *Appl Environ Microbiol* 69:1614–1622.
14. Johnson D, Leake JR, Ostle N, Ineson P, Read DJ (2002) *New Phytol* 153:327–334.
15. Orphan VJ, Taylor LT, Hafenbradl D, Delong EF (2000) *Appl Environ Microbiol* 66:700–711.
16. Whiteley AS, Thomson B, Lueders T, Manefield M (2007) *Nat Protoc* 2:838–844.
17. Colwell RK, Mao CX, Chang J (2004) *Ecology* 85:2717–2727.
18. Petit E, Valière N (2006) *Conserv Biol* 20:1062–1073.
19. Kiers ET, van der Heijden MGA (2006) *Ecology* 87:1627–1636.
20. Fitter AH (2006) *New Phytol* 172:3–6.
21. Gribaldo S, Philippe H (2002) *Theor Popul Biol* 61:391–408.
22. Coenye T, Vandamme P (2003) *Environ Microbiol* 5:719–729.
23. Bianciotto V, Bandi C, Minerdi D, Sironi M, Tichy HV, Bonfante P (1996) *Appl Environ Microbiol* 62:3005–3010.
24. Partida-Martinez LP, Hertweck C (2005) *Nature* 437:823–824.
25. Elbeltagy A, Nishioka K, Sato T, Suzuki H, Ye B, Hamada T, Isawa T, Mitsui H, Minamisawa K (2001) *Appl Environ Microbiol* 67:5285–5293.
26. Dumont MG, Murrell JC (2005) *Nat Rev Microbiol* 3:499–504.
27. Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG (1997) *Nucleic Acids Res* 25:4876–4882.
28. Posada D, Crandall KA (1998) *Bioinformatics* 14:817–818.

## REVIEW AND SYNTHESIS

# Integration of molecular functions at the ecosystemic level: breakthroughs and future goals of environmental genomics and post-genomics

Philippe Vandenkoornhuys\*,  
Alexis Dufresne, Achim Quaiser,  
Gwenola Gouesbet, Françoise  
Binet, André-Jean Francez,  
Stéphane Mahé, Myriam  
Bormans, Yvan Lagadeuc and  
Ivan Couée

UMR 6553 ECOBIO, Centre  
National de la Recherche  
Scientifique, Université de  
Rennes 1, Campus de Beaulieu,  
bâtiment 14A, F-35042 Rennes  
Cedex, France

\*Correspondence: E-mail:  
philippe.vandenkoornhuys@  
univ-rennes1.fr

P. Vandenkoornhuys, A.  
Dufresne and I. Couée have  
contributed equally to this  
work.

### Abstract

Environmental genomics and genome-wide expression approaches deal with large-scale sequence-based information obtained from environmental samples, at organismal, population or community levels. To date, environmental genomics, transcriptomics and proteomics are arguably the most powerful approaches to discover completely novel ecological functions and to link organismal capabilities, organism–environment interactions, functional diversity, ecosystem processes, evolution and Earth history. Thus, environmental genomics is not merely a toolbox of new technologies but also a source of novel ecological concepts and hypotheses. By removing previous dichotomies between ecophysiology, population ecology, community ecology and ecosystem functioning, environmental genomics enables the integration of sequence-based information into higher ecological and evolutionary levels. However, environmental genomics, along with transcriptomics and proteomics, must involve pluridisciplinary research, such as new developments in bioinformatics, in order to integrate high-throughput molecular biology techniques into ecology. In this review, the validity of environmental genomics and post-genomics for studying ecosystem functioning is discussed in terms of major advances and expectations, as well as in terms of potential hurdles and limitations. Novel avenues for improving the use of these approaches to test theory-driven ecological hypotheses are also explored.

### Keywords

Biodiversity, ecosystem functioning, environmental bioinformatics, environmental genomics, functional ecology, metagenomics, molecular ecology, systems biology.

Ecology Letters (2010) 13: 776–791

## INTRODUCTION

All individuals and populations of individuals forming species live and forage within space and time limits. Understanding the interactions and functions of these organisms within their environment is the purpose of ecology, for which a large range of research strategies has been developed. However, exhaustive analysis of all the functional compartments in a given ecosystem presents a major challenge. Microorganisms (i.e. viruses, bacteria, Archaea and micro-eukaryotes), which are essential entities

of biogeochemical cycles on the planetary scale (e.g. Falkowski *et al.* 2008), and represent approximately half of the total carbon contained in living organisms (Shively *et al.* 2001), are still considered as a black box in many ecological studies. Although we know more and more about the importance of microorganisms in nature, the current absence of crucial pieces of information is due not only to the tremendous diversity of genes, metabolisms and species of microorganisms but also to our incapacity to culture over 90% of them (Amann *et al.* 1995; Pace 1997). One of the major challenges facing ecology is therefore to obtain a holistic perception of ecosystems including a comprehensive understanding of microbial communities. Environmental genomics is one of the most promising approaches that can meet this challenge.

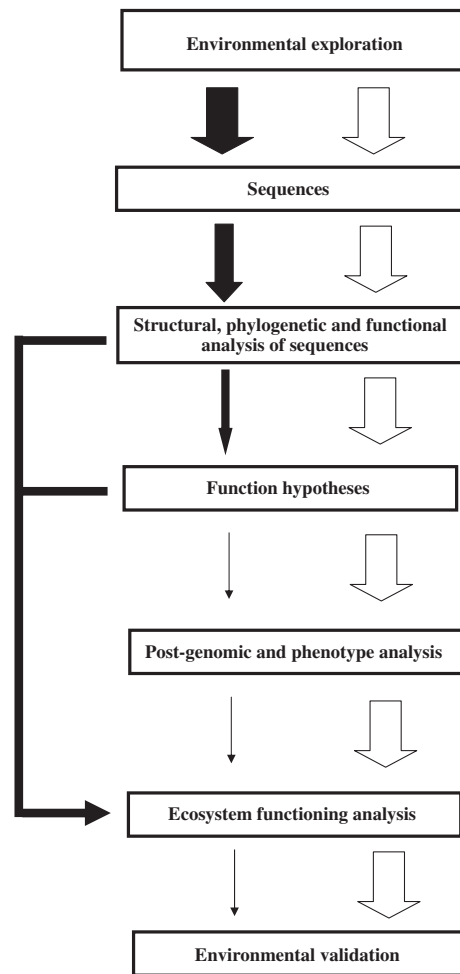
Re-use of this article is permitted in accordance with the Terms and Conditions set out at <http://www3.interscience.wiley.com/authorresources/onlineopen.html>

In the wider sense, environmental genomics in association with post-genomics (i.e. *transcriptomics* and *proteomics*; see the glossary for italicized terms) consists in studying large-scale sequence-based information obtained from a variety of environmental samples, at organism, population or community levels, in order to gain novel insights into evolutionary ecology, organism–environment interactions and processes of ecosystem functioning. As such information contains both synchronic (related to current functioning at a given point in time) and diachronic (related to historical and evolutionary dynamics) aspects, the deciphering of genomes, transcriptomes and proteomes is the most powerful and most large-scale approach to date that may link ecology, evolution and Earth history.

Environmental genomics and post-genomics are not restricted to bacteria and archaea community genomics, and can encompass studies of various other biological systems. For example: (1) mixed prokaryotic-eukaryotic microorganism communities, (2) small-size eukaryotes, especially pico- and nano-eukaryotes, (3) intricate multi-species networks of higher eukaryotic organisms, such as root mats or mixed-species insect swarms, (4) higher eukaryotic organism tissues containing their naturally associated parasitic or mutualistic symbionts and (5) non-model species that cannot be grown or raised under laboratory conditions. In other fields of research such as toxicology and ecotoxicology, environmental genomics generally refers to gene–environment or genome–environment interactions, thus including the study of model species, such as yeast or *Arabidopsis thaliana*, under strong environmental constraints (Teixeira *et al.* 2007) or from an evolutionary perspective (Delneri *et al.* 2008), or even studies of the human genome (Ballatori *et al.* 2003).

This review is focussed on environmental genomics and post-genomics in an ecological context, where analyses of large-scale sequence information can reveal how functions and signals are propagated and integrated at the different ecological levels – individual, population, community, ecosystem – and across various temporal and spatial scales. The aim of environmental genomics, transcriptomics and proteomics in an ecological context is to understand the ecosystem ‘dark matter’ (Marcy *et al.* 2007) after translation into nucleic acid and protein sequences (Fig. 1; Box S1), by taking advantage of the fact that these sequences convey functional information, interact with ecosystem parameters through environmental signalling and acclimation processes, and have been shaped by evolutionary pressures, thus offering a glimpse of past environments.

Given the great expectations associated with this recent field of research, we also discuss the validity of environmental genomics and post-genomics for studying ecosystem functioning, in terms of major advances and limitations, and then explore new avenues for improving these approaches to test theory-driven ecological hypotheses.



**Figure 1** Real-life and ideal fluxes of analysis and information in environmental genomics. Current throughputs of analysis and information-processing are given as black arrows, whereas the ideal throughputs to be achieved are shown as white arrows. Arrow thickness reflects the efficiency of the analyses.

## ENVIRONMENTAL GENOMICS AND THE UNIFICATION OF DIFFERENT FIELDS OF ECOLOGY AND BIOLOGY

Clear connections exist between the hierarchic levels of ecological organization from individual to population to community to ecosystem. However, ecosystem ecology, which requires a mechanistic approach, is mainly based on physiological ecology (e.g. measurements of C, N or P fluxes). Ecosystem ecology is thus disconnected from the other ecological levels, and from the rest of ecology, although ignoring the question ‘who’s doing what?’ could be justified by the scale of the analysis. Along with this fact, and as pointed out by Fitter (2005), this dichotomy in ecology [...] *has been framed in terms of functional redundancy* [...], thus placing the ecological function as a cornerstone, while



individuals are only considered as vectors of this ecological function. Hence, the consequences at the ecosystem level of changes at the population level are poorly known (Fitter 2005). Environmental genomics allows the diversity of organisms to be linked to the functions they display by providing the theoretical possibility of accessing at least partially every single species of a given ecosystem. As underlined by Ungerer *et al.* (2008), genomic approaches [...] offer new insights into higher-level biological phenomena that previously occupied the realm of ecological investigation only [...]. By removing previous dichotomies between ecophysiology, population ecology, community ecology, and phylogenetics on the one hand and ecosystem functioning on the contrary, environmental genomics along with genome-wide expression approaches greatly contributes to the merging of scientific fields and is a source of novel ecological concepts and hypotheses (see major breakthrough & new frontiers sections). However, linking diversity with the entire set of functions carried out by organisms in their natural habitat remains a major challenge.

#### INTEGRATION OF DIVERSITY AND FUNCTIONS FROM MOLECULAR DATA

For over two decades, culture-independent molecular analyses have been used to analyse microbial community and population diversity, and also to study particular functions, such as denitrification or nitrogen fixation. In current environmental genomics studies, the metabolic and physiological potentialities of uncultured (micro)organisms are revealed by analyses of metagenomes (see Box S1 for details), i.e. the collection of genomes recovered from the same environmental sample, or from single-cell environmental genomes (see 'major breakthroughs' section). Despite analytical and technological limits (Table 1), advances in bioinformatics have improved the assembly of large fragments of genomes, the identification of RNA and protein-coding genes within these fragments and the determination of their biochemical and biological potential functions in complex mixtures of sequences from co-occurring organisms. The general aim of these analyses is to decipher taxonomic composition, metabolism, physiology and interactions in natural consortia of organisms in order to unravel evolutionary and ecological processes together with biotic interactions, as well as their changes over time and space. In other words, environmental genomics tackles the questions 'who's doing what, how, when and where?' Furthermore, the correlations between the genetic and functional diversity of communities and environmental conditions can be used to integrate this sequence information into ecosystem processes (Box S1). However, it must be stressed that these approaches, although fruitful, 'only' provide hypotheses which must

then be tested by other means (Figs. 1, 2). Analyses of genome sequences do not in fact reveal which functions are really expressed or identify the active organisms in a given process. The relevance of functional predictions and the validity of functional models based on genomics data can be improved by coupling environmental genomics with (*meta*)transcriptomics and (*meta*)proteomics approaches. It has also been shown that environmental genomics approaches can be coupled with direct probing or labelling of ecological processes. In an elegant work, Mou *et al.* (2008) used an experimental metagenomic approach to investigate the assimilation and mineralization of dissolved organic carbon by adding thymidine analogue bromodeoxyuridine as substrate in order to detect and extract the DNA of the individuals involved in the ecological process under study. The authors were able to elucidate the factors controlling heterotrophic communities (i.e. trophic interactions and physical conditions) and the rules controlling the assemblages of microorganisms within the studied ecosystem. This work presented convincing results arguing in favour of the ecological theory which predicts that heterogeneous environments are conducive to the establishment of generalist species with broad ecological niches (Kassen 2002). Other experimental metagenomic analyses using stable-isotope probing (Dumont & Murrell 2005) have greatly advanced our understanding of the actors in methane cycling (Cébron *et al.* 2007). Use of RNA stable-isotope probing has also led to new findings and hypotheses related to plant-microbe interactions and has highlighted that plants interact within their roots with many more microorganisms than previously believed (Vandenkoornhuysen *et al.* 2007). The selected studies above demonstrate that these approaches are not a mere technological tour de force. They provide novel insights into community structures and generate numerous functional hypotheses. The following section describes other striking examples of the application of environmental genomics to develop our understanding of ecosystem functioning.

#### MAJOR BREAKTHROUGHS OF ENVIRONMENTAL GENOMICS

One of the most innovating aspects of environmental genomics is the capacity to predict new functions and to infer relationships between functions, whether novel or not, and particular species or specific communities. A classic example is the discovery of a new class of light-driven proton pumps in uncultured marine proteobacteria (Béjà *et al.* 2000). These proteins, named proteorhodopsins, might sustain a photoheterotrophic lifestyle in many planktonic bacteria and archaea species (de la Torre *et al.* 2003; Frigaard *et al.* 2006) inhabiting various sunlit aquatic environments (Béjà *et al.* 2001; Sabehi *et al.* 2003; Venter *et al.* 2004;

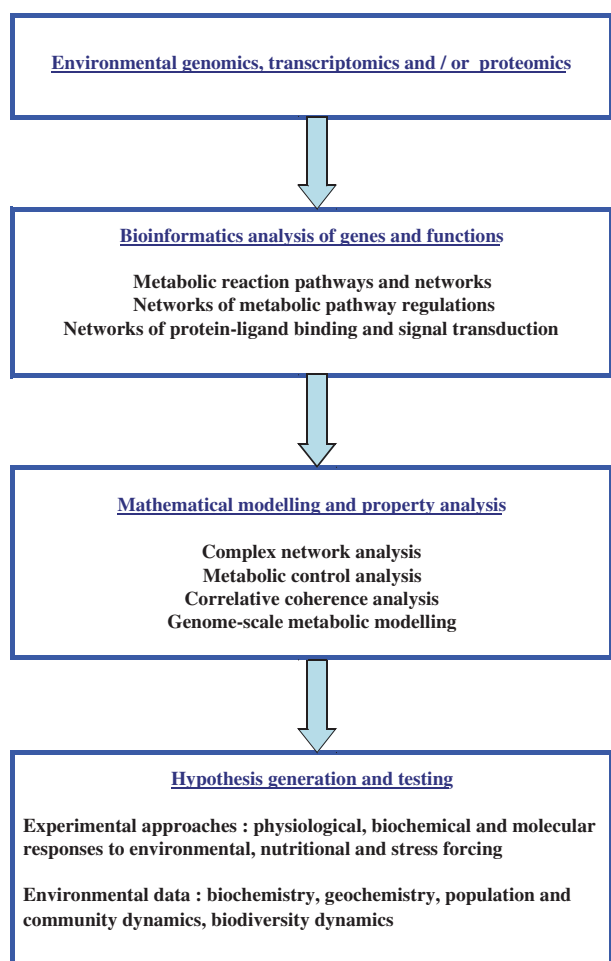


**Table 1** Advantages and limitations in environmental genomics and post-genomics

Stage of analysis	Advantages	Limitations
Sampling	No culture- or growth-related bias	Spatio-temporal heterogeneity
	Direct environmental sampling; large multi-species sampling; large multi-tissue sampling	Cost of representative or exhaustive sampling
Sequencing	Analysis of complex experimental designs involving populations and communities	Careful ecological assessment of environmental sampling and of experimental designs
	Possible long-term storage of DNA, RNA, or protein samples	Availability of reliable protocols for the extractions of nucleic acids and proteins
	High-throughput technologies for DNA, RNA and proteins	Possibilities of sequencing bias; poor sequencing of less-represented genomes
Information processing and functional analysis of organisms, communities and ecosystems	Decreasing cost of sequencing and mass spectrometry	Cost of sequencing for large sample collections, in relation to the exhaustiveness of sampling
	Long-term public databases	Exponential increase of the amount of sequence data; cost and maintenance of database infrastructure
	Biodiversity and phylogenetic analysis	Taxonomic bias in databases
	Functional profiling of naturally occurring organisms and communities	Assembly of short genomic fragments giving a partial view of organismal functional capacities
	Link function and diversity and answer the question 'who is doing what?'	Functional bias in database; computational demand for bioinformatics analyses; poor quality of annotations and amplification of annotation errors
	Discovery of novel ecologically relevant functions	Functional inferences from genomics data in the absence of transcriptomic and/or proteomic data; biased conclusions on the basis of apparent absence of function
	Identifying links between diversity, functional changes and environmental variables	Experimental bottleneck of functional characterization of new genes
	Evolvability of genomics data analysis through improvement of annotations	Computational cost of re-annotating sequences
Re-analysis of genomics data in the light of novel environmental data	Comprehensive environment variable surveys; environment variable databases; environment-dedicated bioinformatics tools; exponential increase of environmental data; increased complexity of the comparison between environmental data and genomics data	
Comparison of present-day ecosystem functioning with earth history and paleo-ecosystem functioning		
Combination of synchronic and diachronic analysis		
Identifying links between diversity, functional changes and environmental variables	Confusing the reality of ecosystem functioning with the reconstructed image from environmental genomics	

Atamna-Ismael *et al.* 2008). However, the physiological and ecological roles of every type of proteorhodopsin need to be fully described (Fuhrman *et al.* 2008).

The strength of environmental genomics was also shown when mesophilic Crenarchaeota could be linked to ammonium oxidation. Few specific bacterial groups were known



**Figure 2** Mathematical modelling in environmental genomics analysis. Reconstructed networks from environmental genomics data (Box S2) can be analysed by various methods of mathematical modelling (Getz 2003; Feist *et al.* 2008; Westerhoff & Palsson 2008; Fuhrman 2009), that can assess and quantify their dynamic properties and generate hypotheses on community and ecosystem functioning. Hypothesis testing can then be carried out by experimental and environmental verification approaches, with the subsequent possibility of iterations between the different steps of the process. The main steps in this flowchart are derived from the description of the systems biology paradigm by Palsson (2006).

to use ammonium as an energy source. Parallel application of environmental genomics approaches to marine plankton and soil samples led to identification of genes encoding for an ammonium monooxygenase on genomic fragments affiliated to Archaea (Venter *et al.* 2004; Treusch *et al.* 2005). In an impressive follow-up study, Leininger *et al.* (2006) not only showed that one subgroup of mesophilic Crenarchaea actively catalyses ammonium nitrification but also established that archaeal *amoA* genes were much more abundant than the corresponding bacterial genes in different

soil samples, thus suggesting that they are major players in ammonia oxidation in diverse soil ecosystems. This discovery produced a downright jump-start for an enormous number of studies of Crenarchaeota in other terrestrial and marine environments, most of the results indicating the prevalence of Archaea over Bacteria in this first step of nitrification. The hypothesis that Archaea play an important role in the overall N-cycle was therefore considerably strengthened. These are two impressive examples of how the detection of key protein-coding genes on a genomic fragment can challenge long-lasting ecological paradigms.

In the above studies, the authors sequenced long fragments of DNA bearing taxonomically or functionally informative genes. In contrast, community-centered approaches, followed for instance by Tyson *et al.* (2004) and Venter *et al.* (2004), have demonstrated the possibility of inferring the structure and the potential activity of microbial assemblages using *shotgun sequencing*.

The biofilm analysed by Tyson and co-workers flourishes at the surface of highly acidic, metal-rich drainage waters in an iron mine. Because of the very reduced biodiversity in this extreme environment, the authors were able to reconstruct two near-complete genomes and they deduced the potential biological functions of the organisms in the biofilm in relation to water chemistry. In particular, they were able to hypothesize that bacteria of the *Leptospirillum* group III, which were relatively sparse in the biofilm, were probably the only group of N<sub>2</sub>-fixing organisms and therefore the single possible point of entry of nitrogen in the biofilm.

Environmental genomics tools have also been applied to ecosystems harbouring more diverse microbial communities. In one of the largest environmental genomics study ever undertaken, Rusch *et al.* (2007) produced a total of 7.7 million reads from samples of surface waters collected during the Global Ocean Sampling expedition off the eastern American coast, in the Gulf of Mexico, the Panama canal and in the eastern part of the equatorial Pacific Ocean. Despite a strong sequencing effort, 53% of the reads remained unassembled, which could be ascribed to the high levels of diversity within the samples. However, despite this high level of genetic polymorphism, this impressive dataset was dominated by very few genera of bacteria such as *Pelagibacter*, *Prochlorococcus* and *Synechococcus*, which were found at many sites along the transects. Two other abundant genera, *Burkholderia* and *Shewanella*, only appeared in the Sargasso sea (Venter *et al.* 2004). These five genera were also found to be among the most abundant in the dataset when 16S rRNA sequence clusters were used to characterise the diversity. A large fraction of the diversity fell within ribotypes, with the presence of distinct populations in different environments. Likewise, computations of the similarities between community genomes were used to

assess genetic distances between sampled environments. Samples from unique habitats such as a hypersaline pond and a freshwater lake were the most distant in terms of genomic composition whilst similar habitats such as the Sargasso sea or tropical open ocean waters contained more similar microbial metagenomes.

Environmental constraints exert a strong selection pressure on living (micro)organisms. These factors drive the selection of guilds that are best adapted for habitat colonisation. Thus, application of environmental genomics on a 'global' scale (e.g. through sampling along a gradient of environmental fluctuation or through comparison of different ecosystems) offers an unprecedented way of linking environmental parameters with the specific and functional diversity of microbial assemblages (see also Tringe *et al.* 2005; Dinsdale *et al.* 2008).

Metagenomic studies have offered a broad view of the organization of genetic diversity in various microbial communities as well as insights into the metabolism of their dominant members. However, the paucity of fully assembled genomes from metagenome sequencing has hampered our ability to link diversity and functions. The need to target specific groups of organisms in an environmental sample has led to the development of numerous methods and protocols for isolating populations ranging from a few thousand cells to only one cell and for obtaining enough DNA template for sequencing (Rodrigue *et al.* 2009; Woyke *et al.* 2009). Recently, Zehr *et al.* (2008), by deciphering the genome sequence of a new group of unicellular nitrogen-fixing marine cyanobacteria dubbed UCYN-A, have provided an excellent example of how the combination of isolation techniques and environmental genomics helps to link ecosystem functioning with the genetic makeup and metabolic features of organisms. UCYN-A cyanobacteria were first detected through the amplification of transcripts of the *nifH* gene (dinitrogenase reductase subunit of nitrogenase; Zehr *et al.* 2001) in environmental samples. Unlike other unicellular diazotrophic cyanobacteria, UCYN-A cyanobacteria express the *nifH* gene during daytime when oxygen production by photosystem II (PSII) inhibits nitrogen fixation (Church *et al.* 2005). Despite repeated efforts, no member of this group could be maintained in culture. The authors used flow cytometry to isolate about 5000 cells from a natural population of the UCYN-A group and subjected the genomic DNA to isothermal whole genome amplification and pyrosequencing. As expected for a diazotroph, the UCYN-A metagenome encodes a complete nitrogen fixation pathway. Surprisingly, although numerous sequences of Photosystem I genes were detected, no genes coding for the PSII proteins were found. The authors provided strong evidence that cyanobacteria of the UCYN-A group do not possess a complete photosynthetic

apparatus and also seems to lack all the genes necessary for CO<sub>2</sub> fixation. Thus, the UCYN-A group appears to be the sole known cyanobacterial lineage unable to produce oxygen. This would explain how UCYN-A cyanobacteria concomitantly perform N<sub>2</sub> fixation and photosynthesis. Several studies had suggested that members of the UCYN-A group were abundant in oceans and might contribute markedly to biological nitrogen fixation (Montoya *et al.* 2004). The inability of some marine diazotrophs to fix CO<sub>2</sub> will certainly require a refinement of established models of N and C cycling in oceans as it deviates from the stoichiometrical relationships previously assumed for biological N fixation and photosynthetic C incorporation (Mahaffey *et al.* 2005).

Environmental genomics has become a standard approach in the study of aquatic habitats, owing to their relative simplicity. In comparison, soils and sediments appear to be more spatially heterogeneous and phylogenetically diverse. Estimates of soil diversity are often in the range of hundreds to thousands of microbial species per gram of soil (Torsvik *et al.* 2002). Soil and sediments are often considered to constitute one of the largest reservoirs of microbial diversity on Earth. Notwithstanding the difficulties of obtaining representative samples or limitations associated with DNA extraction and purification (Table 1), sequencing of metagenomes from soil communities also requires much greater effort to obtain significant sequence coverage. Consequently, terrestrial habitats have mainly been targeted by metagenomic studies in the prospect of finding new molecules of biomedical or agricultural interest (Daniel 2005). International programs such as TerraGenome have been started with the aim of sequencing the metagenomes of reference soils (see <http://www.terragenome.org/>).

The use of high-throughput sequencing technologies has also led to tremendous progress in understanding the intricate associations between symbiotic microorganisms and their eukaryotic hosts. Woyke *et al.* (2006) described the functioning of a complex symbiosis between the marine oligochaete *Olavius algarvensis* and a microbial consortium consisting of two sulphur-oxidizing gamma-proteobacteria and two sulphate-reducing delta-proteobacteria. The worm is characterized by the complete absence of a digestive apparatus and a reduced excretory system. Thus, nutrition of the host, as well as the degradation of toxic by-products of its metabolism, is entirely dependent on the activity of the bacterial consortium. Analysis of the metagenomic data provided valuable insights into the metabolism of the different bacterial partners and into the network of interactions established between the worm and its symbionts. The host is supplied with C, N, S and P compounds by the symbiotic bacteria, and host organic osmolytes and waste products are used as C and N sources for symbiont metabolism. Analysis of the

protein-coding genes of the symbionts has confirmed the existence of syntrophic cycling of sulphur elements between the sulphur-oxidizing and the sulphate-reducing symbionts.

Finally, organism-centered studies of isolable multicellular eukaryotes (Martin *et al.* 2008; Vera *et al.* 2008; Rasmussen & Noor 2009) have shown the usefulness of environmental genomics for analysing such organisms in their ecological and evolutionary context. Altogether, these examples of function-, organism-, community- or environment-centered approaches shed light on how environmental genomics and post-genomics allow the integration of molecular data with ecological metrics and open new windows on the complex interplays between genomes, phenotypes, populations and environment. All these results, which have already induced advances in ecology, are based on a battery of bioinformatics tools (see Box S2 for details) to analyse sequence data. However, there are still limitations, which are discussed below, along with recommendations to avoid mis-analyses and mis-interpretations.

## CURRENT LIMITATIONS OF ENVIRONMENTAL GENOMICS FUNCTIONAL INTEGRATION

### Sampling and sequencing

Technological and conceptual limitations of environmental genomics (Table 1) are not trivial, and require thorough consideration to further improve analyses. Confrontation with various environmental samples (such as seawater, freshwater, soils, sediments, bacterial mats, plant and animal tissues) has resulted in the considerable improvement of extraction protocols and methods, and of sample preparations, which must be environment-compatible, contamination-free, non-degradative, non-combinatorial, and complete. Considerable progress has also been made in the quality of massive sequencing in terms of throughput, cost, read length, and read quality. Current sequencing methods can generally yield deep and representative environmental sequences of high quality. Moreover, these methods are constantly improving and bioinformatics analysis of sequences is constantly reducing sequencing noise and bias (Quince *et al.* 2009). However, the quality and representativity of sequencing may remain hampered by the complexity of some environmental samples, in terms of organism diversity and abundance as well as size and composition (e.g. percentage of repeats) of the individual genomes.

### Gene identification and functional characterisation

The first task of finding genes in environmental genomics or metagenomics data is sometimes compounded by the great

diversity of genomes that is revealed and by the myriad novel genes they contain (Table 1). Whereas gene identification has become less and less problematic for bacteria and archaea genomes, the difficulties must not be underestimated in the case of higher eukaryotic genomes (Levasseur *et al.* 2008) due to the modular nature of eukaryotic genes and to the short sequences produced by second-generation sequencing platforms which complicate the prediction of *open-reading frames*.

Another major challenge in environmental genomics is the subsequent step of correctly identifying functions on the basis of sequence data. Classically, the identification of gene functions is heavily dependent on comparisons, using standard tools such as BLAST (Basic Local Alignment Search Tool, Altschul *et al.* 1997), with sequences from other organisms or metagenomes present in genome databases such as GenBank. The inference of gene function is then derived from functional annotations of these similar sequences. Bioinformatics analyses are thus becoming a major bottleneck in environmental genomic studies (Fig. 1), as the production of sequences outpaces the computational capacities available in most laboratories. Moreover, as highlighted by Pálsson (2006), “*it should be emphasised that every gene annotation based on in silico methods is hypothesised and such annotation is subject to revision, until the gene has been cloned, expressed, and the function of the gene product directly evaluated*”. Thus, most *bona fide* annotations are derived from genes of model organisms, where biochemical analysis and reverse genetics can readily be carried out. Furthermore, the sequenced organisms available in databases represent a small and strongly biased subset of the biodiversity revealed by cultivation-independent methods. However, it is worth noting that several recent initiatives such as the Moore Foundation Marine Microbial Genome Sequencing Project, the Genomic Encyclopedia of Bacteria and Archaea Project, or the Fungal Genome Initiative will contribute to improve the list of sequenced organisms and to obtain a better coverage of the known biodiversity.

This duality between the great phylogenetic diversity of environmental genes (Yooseph *et al.* 2007) and the limited number of well-characterised genes in the databases is likely to result in high proportions of genes with ‘unknown’ or ‘hypothetical’ functions in environmental genomes. This may also cause a strong bias towards identification of the best-known, and maybe most straight-forward, functions, such as those related to central metabolism. Finally, numerous causes of incorrect annotations in model species have been identified (Galperin & Koonin 1998). This is why some authors have voiced concern that comparison of environmental genomes with imprecise or erroneous annotations in databases may lead to exponentially amplified errors and inappropriate functional predictions (López-García & Moreira 2008).



### The concept of function and the difficulty of function assignment

Most studies of gene-function relationships have focussed on the cell and organismal levels. Even at these levels, the difficulty of precisely defining the multi-faceted concept of function has been emphasised (Danchin *et al.* 2004) and gene functions may be more complex than those hypothesised from database annotations. A well-annotated gene, with a well-defined function, may yield various products through alternative splicing and post-translational modifications, and/or multi-functional products. For instance, a gene may code for multiple enzymatic activities, with multiple subcellular localizations (Silva-Filho 2003), or with combined enzymatic and regulatory functions (Takeda *et al.* 2009). Complete understanding and annotation of gene product functions are therefore extremely difficult to achieve (Danchin *et al.* 2004).

Moreover, many annotations that are based purely on sequence homology are likely to be incorrect, since biochemical characterization of gene products previously identified by similarity searches has often yielded surprises, especially in terms of ligand/substrate specificities or of subcellular targeting. Conversely, an apparent absence of gene families on the basis of homology searches does not necessarily mean an absence of function since independent emergence of catalytic processes can occur in independent protein phylogenetic backgrounds, thereby creating sets of analogous enzymes (Galperin *et al.* 1998). Finally, whereas *homologous identification* can be extremely precise on the basis of short sequences, as in the identification of short expressed tags vs. genome data from the same organism, *heterologous identification* of unknown genes vs. gene databases from more or less related organisms can be hazardous. Thus, as an exaggerated example, BLASTX analysis (search of protein databases for all the translated possibilities of a DNA sequence) of the complete gene sequence of *Nicotiana tabacum* ornithine decarboxylase (polyamine biosynthesis pathway) versus the *Arabidopsis thaliana* protein database yields a significant identification with diamino-pimelate decarboxylase (lysine biosynthesis pathway). This instance of heterologous mis-identification between related species may be ascribed to the fact that *Arabidopsis thaliana* lacks an archetypal ornithine decarboxylase (Hummel *et al.* 2004).

It is clear that all the above-described situations are likely not only to occur but also to be compounded at the ecosystem level where multiple environmental variables drive the expression of gene functions and direct the role played by organisms in ecosystem processes. Furthermore, our ability to determine the links between biodiversity and ecosystem functioning might be hampered by the importance of horizontal transfers of protein-coding genes – for

instance through viruses or plasmids – between phylogenetically distant Bacteria and Archaea (Koonin & Wolf 2008).

### Genome–environment interactions and the plasticity of gene expression

Although identification of a given function at the gene level may indicate selection of this gene in the organisms present in the ecosystem, it does not give information on the patterns of gene expression. In other words, there are always important differences between who is there in the ecosystem and who is at work in the ecosystem. As far as possible, genomics data must be complemented with transcriptomics or proteomics data, which correspond to measurements of steady-state levels of transcripts or proteins (Box S2; Fig. 2).

Although its adaptive value has been subjected to criticism (Feder & Walser 2005), mRNA expression is an important step in gene-to-functional protein expression (Stranger *et al.* 2007), and an important response to the perception of environmental clues (Hummel *et al.* 2004). Improvement of RNA isolation and application of massive sequencing to the analysis of cDNA from environmental samples (Frias-Lopez *et al.* 2008) or non-model species (Vera *et al.* 2008) have circumvented the limitations of *DNA array technologies*. In spite of some successful applications (Parro *et al.* 2007), DNA array technologies cannot be readily applied to most environmental samples, since they imply *a priori* knowledge of the species and communities under investigation. It must be kept in mind however that environmental transcriptomics suffers from some drawbacks, such as the variable half-lives of mRNA, and the fact that, in bacteria and archaea, mRNAs represent a small proportion of the total RNA and cannot be enriched by poly-dT affinity, since they lack the polyA tail found in eukaryotic mRNA. Moreover, functional characterization of cDNAs is confronted with the same limitations of annotation as those described above for gene function analysis (Table 1). Finally, transcriptomics generally gives a comprehensive view of expression levels across the individuals of the sampled population (Stranger *et al.* 2007). More detailed analysis of environmental transcriptomics data should eventually take into account the impact of individual genetic variations on gene expression (Stranger *et al.* 2007).

Analysis at the protein level may provide the most representative snapshots of organism or community functionalities. Proteomics and metaproteomics approaches have indeed been carried out with success on environmental samples (Ram *et al.* 2005). Nonetheless, reliable extraction of proteins from natural environments can be more challenging than for nucleic acids, especially in terms of the quality and quantity of the sampled proteomes. High

throughput analysis of metaproteomes can be carried out by mass spectrometry, which however requires comparison with databases containing gene sequences originating from the same organisms or from very closely related organisms, as mass spectrometry data are very sensitive to changes in protein sequences. Thus, metaproteomics studies must be coupled to metagenome sequencing to detect significant numbers of protein matches (Ram *et al.* 2005).

### From environmental genomics to environmental phenotypes

As most metabolic and functional schemes of ecosystem functioning are dependent on heterologous comparisons with databases containing significant numbers of *in silico* annotated genes, such schemes should be clearly labelled as hypothetical (Fig. 2). This hypothetical nature does not undermine the core value of such analyses, but should be taken as an incentive to validate hypotheses and integrate these hypothetical schemes into further ecosystem-level studies. In other words, caution must be taken not to indulge in direct integration of sequence analysis, which may short-circuit important validation steps (Fig. 1). Moreover, due to regulatory, biochemical and supramolecular interactions, the number and scope of organism and ecosystem functions derivable from a single genome or from community genomes does not scale with the mere catalogue of genes contained in those genomes.

The identification of new environmental genes should be followed by further functional, biochemical, and physiological characterization. This can first be carried out on candidate genes, selected on the basis of their outstanding interest or representativity in relation to ecosystem knowledge. This was the case for proteorhodopsin genes. They were identified in analyses of environmental DNA, and their products were biochemically characterised after *over-expression* (Béjà *et al.* 2000). Furthermore, environmental genomics data can be complemented with laboratory organism-centered approaches, not only in the case of isolable multicellular eukaryotic organisms, but also in the case of microbial communities. Thus, enrichment cultures and the cultivation of selected microbial strains may be useful for further genomic and physiological characterisation (Giovannoni *et al.* 2005) or to test important physiological and ecosystemic hypotheses (López-García & Moreira 2008). In this context, important progress has been made to develop culture protocols and media to cultivate recalcitrant microorganisms of ecological interest (Ben-Dov *et al.* 2009).

More generally, environmental genomics results must be critically confronted with ecological ecosystem knowledge (Mou *et al.* 2008; Zehr *et al.* 2008) and/or tested through modelling procedures (Röling *et al.* 2007). Procedures for

environmental validation, corresponding to a kind of ecosystem phenotype characterization, should be better defined, in the same way that model species genomics should be complemented with organism phenotype characterisation (Fig. 2). However, it may be extremely difficult to carry out high-throughput post-genomics functional characterisation, such as protein over-expression and biochemical analysis, mutant-based gene/function analysis or natural variation-based gene/function analysis, in the context of environmental genomics (Wullschleger *et al.* 2007). However, it has to be stressed that bioinformatics approaches and tools can yield broad and useful information, especially functional information, even with a genome coverage as low as 0.1X (Rasmussen & Noor 2009), when long enough sequence tags are obtained from random pyrosequencing. This is true even for communities of organisms that do not correspond to any available genomic sequence in the databases. Moreover, novel ideas and methods are constantly improving the relevance of environmental genomic analyses to address ecological questions.

### IMPROVEMENT OF GENOMICS APPROACHES FROM AN ECOLOGICAL POINT OF VIEW

#### The importance of ecological and evolutionary criteria for functional identification

The difficulties of homology-based functional identification have been recognized for some time, but various improvements using protein domain detection and gene context approaches (Singh *et al.* 2009) have been made. Phylogenetic analyses have been particularly valuable in going beyond basic homology comparisons and accounting for the evolutionary history of genes (Levasseur *et al.* 2008). Thus, combinations of phylogenetic tree construction, integration of experimental data and differentiation of *orthologs* and *paralogs*, have been proposed to address annotation errors. As a result, a number of software platforms and databases have been developed recently (see Box S2). These enable phylogenetic analysis and utilisation of gene clusters, such as COGs (clusters of orthologous groups; Tatusov *et al.* 2003), to infer gene function by superimposing experimental information on the phylogenetic trees (Levasseur *et al.* 2008). The use of phylogenetic data for functional reconstruction from environmental genomics is particularly interesting in the light of relationships between community phylogenetic structure and ecosystem processes (Prinzinger *et al.* 2008). However, the quality of this kind of phylogeny-based analysis is strongly dependent on the scope of the initial phylogenomics database and on relationships between the environmental species under study and the set of species present in the databases.

### The importance of bioinformatics and statistical controls

Given the unfinished status of gene and protein databases, it may be important to develop experimental bioinformatics controls, especially when the species in the environmental genomics data do not have phylogenetically related counterparts in the databases. Thus, controls can be carried out with artificially-reconstructed genomes (Yang & Bennetzen 2009) or communities (Quince *et al.* 2009). In robustness controls, a known genome of a control species could also be re-analysed by comparison with gene and protein databases from which this given species, its genus, or its family would be artificially removed. This approach could be used to estimate the accuracy of functional assignments when an unknown genome is compared with phylogenetically unrelated genomes, and thus to select the most robust functional assignments. Environmental genomics approaches often imply the parallel comparative analysis of various samples corresponding to gradients of ecological factors, such as light, salinity, or anthropic pressure (Raes *et al.* 2007; Dinsdale *et al.* 2008; ). The complexity of environmental genomics data therefore requires the specific development and/or adaptation of statistical analysis tools as described in Rodriguez-Brito *et al.* (2006).

### Expected improvements of functional annotations and genome assembly

As described above, a great number of functional annotations are hypothetical and subject to revision. Conversely, continuous revision can be expected to improve environmental genomics data analysis. However, systematic and standardized processes for database revision are still lacking, and need to be developed for all the different genomics approaches, whether model-species-based or environmental, in order to avoid possible erroneous revisions. Moreover, novel methods, such as those taking into account not only the nature of direct gene products but also regulatory interactions, protein-protein interactions, and protein-metabolite interactions (Palsson 2006), are likely to improve annotations. Developing comparisons of metagenomics data with metatranscriptomics and metaproteomics data can also be expected to improve *in silico* identification of genes and annotations. Finally, full and accurate annotation of model species genomes, corresponding to different major phyla, remains to be carried out and may further improve environmental genomics data analyses. However, the diversity and variability encountered in environmental genomics data may eventually surpass the range of model species genomics data and even modify the very concept of species and of model species (Medini *et al.* 2008). Moreover, model species databases will be progressively complemented with databases for single-species genomes of ecological

interest, especially if single-cell genomics (Marcy *et al.* 2007; Rodrigue *et al.* 2009; Woyke *et al.* 2009) can be developed in an ecological context. These environmental genomics data on single species, obtained through direct sampling of individuals, cultivation or single-cell approaches, will be extremely useful not only for annotation but also to assemble metagenomics data.

### Further analysis of the complete wealth of environmental genomics data

In the same way that they can be re-analysed in the light of improved annotations, stored environmental genomics data can be re-analysed to extract meaningful new information. For instance, the comparative analysis of *promoter sequences*, which are involved in gene expression regulation, has been extremely limited in the case of environmental genomics data. Promoter sequences involve consensus sequences and *regulatory cis-acting elements* that can be highly conserved across species or highly variable, depending on evolutionary constraints and selection pressures (Zhu & Snyder 2002). Furthermore, databases of promoters are being developed (Zhu & Snyder 2002). Therefore, it could be possible to classify gene sets from environmental genomics data according to the cis-acting regulatory elements that are present in their promoters, thereby generating classes of co-activated or co-inhibited genes. Insofar as cross-species consensus sequences are available for use, such classification could point to co-regulated genes at the community level. Moreover, such information on co-regulation at the ecological level could lead to experimental verification using *ChIP-on-chip approaches* on the proteins that regulate these networks of co-regulated genes (Buck & Lieb 2004). Similarly, it will be possible in the future to carry out deeper analyses of environmental genomics data for other regulatory levels, such as the generation of multiple transcripts from a single gene (Méreau *et al.* 2009) or the systematic analysis of regulatory RNAs (Shi *et al.* 2009). Finally, in parallel to environmental genomics, the miniaturization and automation of sensors and probes have also resulted in the development of powerful analytical tools that make it possible to carry out high-frequency temporal, as well as proximal, monitoring of natural habitats. Such tools are essential to monitor environment variables at scales of time and space relevant to community activities and molecular functions. Analytical microsensors are able to monitor fine variations or gradients of various physico-chemical parameters (Krawczyk-Barsch *et al.* 2008). Likewise, isotopic (nanoSIMS) and microscopic techniques (FISH, TEM) can measure the activities of (micro)organisms in their habitats (Dekas *et al.* 2009). Progress has also been achieved in the setting-up of controlled experiments, in which the complexity of communities and the geochem-

ical environments can be manipulated. The use of environmental genomics approaches that combine accurate monitoring and experimentally controlled environments may contribute to build appropriate models of ecosystem functioning (Fig. 2).

### Present and future importance of mathematical modelling for environmental validation

Environmental genomics data are complex in scale and scope. Even the pivotal task of inferring community-level functions from individual functions of genes requires the parallel analysis and integration of hundreds or thousands of genes and individual functions, and an understanding of their functional and regulatory interactions. For the reasons given above, genomics-based data must be compared and integrated with higher-level environmental data, such as experimental data or fluxes of biogeochemical cycles. The richness and complexity of these data raise the problem of transforming functions into equations. However, it is important to be able to describe reconstructed functional networks mathematically, in order to analyse their properties in greater detail (Palsson 2006). Mathematical properties can be used to generate functional hypotheses (Fig. 2) through complex networks analysis (e.g. Fuhrman 2009), metabolic control analysis (Westerhoff & Palsson 2004), correlative coherence analysis (Getz 2003), or genome-scale metabolic modelling (Feist *et al.* 2008). These hypotheses can then be tested experimentally or tested for their fit to environmental data, such as geochemical fluxes, biodiversity fluctuations, or biomass production. Finally, models of reconstructed networks can be improved by iterative interactions between modelling, experimental results and ecosystemic data (Fig. 2).

### NEW FRONTIERS OF ENVIRONMENTAL GENOMICS

The present state-of-the-art shows that environmental genomics has already generated new concepts and tackled questions that were impossible to address before. Improvement of multidisciplinary integration of bioinformatics, genetics, statistics, physiology, ecology, and evolutionary sciences, is likely to raise further questions and to offer the possibility to reinvestigate existing paradigms.

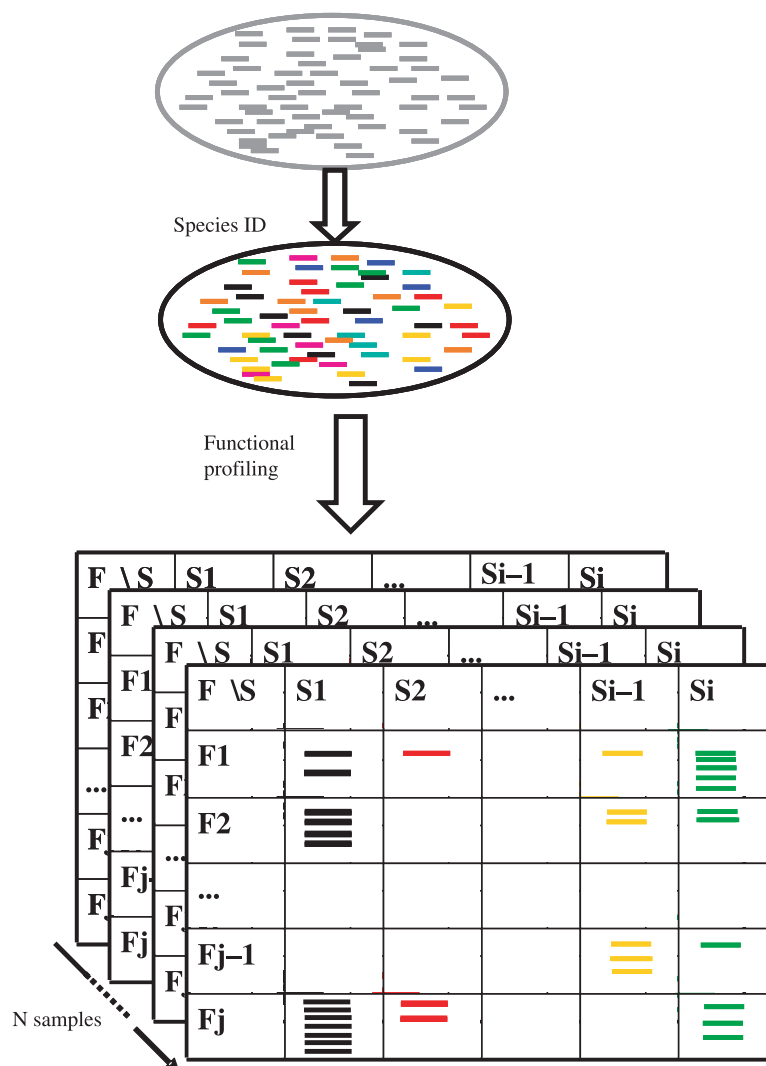
Environmental genomics is leading to a better understanding of diversity at different ecological scales ranging from population to ecosystem by demonstrating that the environmental gene pool is several orders of magnitude greater than previously believed (Yooseph *et al.* 2007). It is clear, from these findings, that the availability of one complete genome sequence for each described taxon would be insufficient to explain the complexity of species (Medini *et al.* 2008). Despite the fact that species are considered as

fundamental units of biology and are thus as important as the cell or individual, the definition of a species and the adoption of a unified species concept is still under debate, although interesting essays on this topic have been published (Mishler & Brandon 1987; de Queiroz 2007). Ribosomal RNA gene analyses have been long considered as sufficient tools to describe diversity because (1) these genes are shared by all living organisms, (2) they contain robust phylogenetic information and (3) they are useful, easy-to-apply tools for application of the phylogenetic species concept (Mishler & Brandon 1987). Environmental sequencing has recently provided a global 'one-does-all' method providing a deep insight into the molecular list of all the sampled (micro)organisms, and describing the genes and functions displayed in more or less complex communities. From this, it becomes possible to consider a genome as a trait and to delimit species as '*separately evolving metapopulation lineages (or, more properly, segments thereof)*' (de Queiroz 2007) by analysing this trait rather than core genes, such as ribosomal RNA genes. It also has to be stressed that the adoption of an explicit species concept directly affects the actual assessment of diversity and thus the fit of (1) models of community dynamics and (2) theories of species assembly. The use of the genome as a trait to describe a species could involve, among other criteria, *gene synteny* and the level of similarity. However, at present, this can be envisaged only for small-genome organisms, such as bacteria, archaea and some eukaryotes.

Besides these considerations, novel fields of research that cannot be studied by other means than environmental genomics are now open to investigation. Pioneer papers, at the intersection of ecology and evolutionary biology, have paved the way for the genomics of co-evolution including mutualism, symbioses and parasitism. For instance, Martin *et al.* (2008) analysed mycorrhizal symbiosis and provided important insights into the behaviour and capacities of the fungal symbiont. In a similar line of research, the behavioural evolution and capacities of insect heritable bacteria have been explored (e.g. Moran *et al.* 2008). Such studies have demonstrated the existence of obligate and facultative mutualists displaying functions ranging from nutrition, protection against biotic or abiotic stresses, to symbiont-manipulating reproduction regimes. The local biotic environment of these bacteria may promote speciation as a result of reproductive and ecological isolation (Moran *et al.* 2008). These studies thus (1) address new questions of co-evolution and macroevolution, and (2) further our understanding of the responses of the partnership to biotic or abiotic environmental stresses.

To date, functional and mechanistic objectives have not taken into account variation at the population level although this information is generally accessible in a number of environmental genomics projects. Usually, deep sequence coverage can detect single nucleotide polymorphisms





**Figure 3** Spatio-temporal three-dimensional organisation of sequence-derived datasets. The set of environmental genomic, *cDNA*, or protein sequences (grey bars) is ascribed to a set of  $i$  Species ( $S$ ), thus resulting in species-labelled sequences (colour bars). The aim of functional analysis and profiling is to ascribe species-labelled sequences to a set of  $j$  functional categories ( $F$ ), thus resulting in a ‘potential function  $\times$  species’ understanding of the ecosystem. The third dimension of the matrix corresponds to spatio-temporally replicated samples, such as samples subjected to various environmental constraints, or samples at different points in time. This kind of dataset can be analysed not only to understand the mechanisms induced by a forcing variable, but also to select and parameterize the components that have to be included in a model.

(SNPs) and structural variations, such as copy number variants (CNVs) (Stranger *et al.* 2007), which can affect individual fitness. However, the field of population genomics (i.e. population studies analysing genome-wide genetic markers) is mainly developing apart from environmental genomics, despite the fact that the theoretical corpus of population genetics is well adapted to deal with environmental genomics data. Reciprocally, predictions and hypotheses can be derived from genomic neutrality tests of population differentiation due to environmental changes (i.e. population differentiation shown through association(s) between an environmental constraint and specific genetic markers). In this case, the genetic marker can be supposed to be a genetic trait of adaptation (Schmidt *et al.* 2008), which can thus be regarded and tested as a possible factor involved in individual fitness. This kind of idea may be considered as one of the purposes of comparative genomics or metagenomics projects.

One major result of environmental genomics projects is the possibility of reconstructing and modelling potential metabolic and regulatory networks. However, these data cannot be readily used to formalise models of ecosystem functioning, as no data can be directly assigned to parameter variables: spatio-temporal variations must be taken into account if ecosystem functioning is to be comprehensively modelled from three-dimensional data matrices, as shown in Fig. 3. Experimental metagenomics, metatranscriptomics and metaproteomics projects testing the consequences of different environmental constraints on physico-chemical measurements can define the most important variables to include in a formal model of ecosystem functioning. Statistical modelling of a given ecosystem requires the kind of data presented in Fig. 3 and metadata, such as biogeochemical analyses, must be included to help the interpretations. It is also possible to model environmental genomics data from a stoichiometric approach or from a kinetic

approach (e.g. Röling *et al.* 2007). Incorporation of spatio-temporal variations into the model would, in itself, lead to a change of scale. Even if environmental genomics is generally focussed at a small scale, it can be speculated that the data contain fractal properties of self-similarity (i.e. sub-units at multiple levels reflecting the structure of the whole object) and fractional dimensionality. These fractal properties could be tested to allow further rescaling at higher levels. As far as we know, such approaches have not yet been used. Such a model could in return be a source of testable hypotheses of ecosystem functioning, and could be used to predict changes in a given ecosystem.

## ACKNOWLEDGEMENTS

This work was supported at least partially by funding from the Centre National de la Recherche Scientifique (CNRS, EC2CO funding programme), the Centre armoricain de recherches en environnement (CAREN, Rennes), the Total Corporate Foundation for Biodiversity and the Sea, and the Brittany regional council. For fruitful and ground-breaking discussions, the authors wish to thank the participants at the CNRS summer school 'High-throughput methods of molecular biology in the environmental sciences' (Roscoff, France 2005) and the CNRS Jacques Monod conference 'Environmental Genomics' (Roscoff, France 2007), with special thanks to Eric Allen, Oded Béjà, Catherine Boyen, Frédérique Barloy-Hubler, Françoise Bringel, Ana Caicedo, Daniela Delneri, Patrick Forterre, Francis Martin, Frédéric Partensky, and Peter Young. The three anonymous referees are also acknowledged for their insight and valuable comments.

## REFERENCES

- Altschul, S.F., Madden, T.L., Schäffer, A.A., Zhang, J., Zhang, Z., Miller, W. *et al.* (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.*, 25, 3389–3402.
- Amann, R.L., Ludwig, W. & Schleifer, K.H. (1995). Phylogenetic identification and *in situ* detection of individual microbial cells without cultivation. *Microbiol. Rev.*, 59, 143–169.
- Atamna-Ismaeel, N., Sabehi, G., Sharon, I., Witzel, K.P., Labrenz, M., Jürgens, K. *et al.* (2008). Widespread distribution of proteorhodopsins in freshwater and brackish ecosystems. *ISME J.*, 2, 656–662.
- Ballatori, N., Boyer, J.L. & Rockett, J.C. (2003). Exploiting genome data to understand the function, regulation, and evolutionary origins of toxicologically relevant genes. *Environ. Health Perspect.*, 111, 871–875.
- Béjà, O., Aravind, L., Koonin, E.V., Suzuki, M.T., Hadd, A., Nguyen, L.P. *et al.* (2000). Bacterial rhodopsin: evidence for a new type of phototrophy in the sea. *Science*, 289, 1902–1906.
- Béjà, O., Spudich, E.N., Spudich, J.L., Leclerc, M. & DeLong, E.F. (2001). Proteorhodopsin phototrophy in the ocean. *Nature*, 411, 786–789.
- Ben-Dov, E., Kramarsky-Winter, E. & Kushmaro, A. (2009). An *in situ* method for cultivating microorganisms using a double encapsulation technique. *FEMS Microbiol. Ecol.*, 68, 363–371.
- Buck, M.J. & Lieb, J.D. (2004). ChIP-chip: considerations for the design, analysis, and application of genome-wide chromatin immunoprecipitation experiments. *Genomics*, 83, 349–360.
- Cébron, A., Bodrossy, L., Chen, Y., Singer, A.C., Thompson, I.P., Prosser, J.I. *et al.* (2007). Identity of active methanotrophs in landfill cover soil as revealed by DNA-stable isotope probing. *FEMS Microbiol. Ecol.*, 62, 12–23.
- Church, M.J., Short, C.M., Jenkins, B.D., Karl, D.M. & Zehr, J.P. (2005). Temporal patterns of nitrogenase gene (*nifH*) expression in the oligotrophic North Pacific Ocean. *Appl. Environ. Microbiol.*, 71, 5362–5370.
- Danchin, E., Vitiello, V., Vienne, A., Richard, O., Gouret, P., McDermott, M.F. *et al.* (2004). The major histocompatibility complex origin. *Immunol. Rev.*, 198, 216–232.
- Daniel, R. (2005). The metagenomics of soil. *Nat. Rev. Microbiol.*, 3, 470–478.
- Dekas, A.E., Poretsky, R.S. & Orphan, V.J. (2009). Deep-sea archaea fix and share nitrogen in methane-consuming microbial consortia. *Science*, 326, 422–426.
- Delneri, D., Hoyle, D.C., Gkargkas, K., Cross, E.J.M., Rash, B., Zeef, L. *et al.* (2008). Identification and characterization of high-flux-control genes of yeast through competition analyses in continuous cultures. *Nat. Genet.*, 40, 113–117.
- Dinsdale, E.A., Edwards, R.A., Hall, D., Angly, F., Breitbart, M., Brulc, J.M. *et al.* (2008). Functional metagenomic profiling of nine biomes. *Nature*, 452, 629–632.
- Dumont, M.G. & Murrell, J.C. (2005). Stable isotope probing – linking microbial identity to function. *Nat. Rev. Microbiol.*, 3, 499–504.
- Falkowski, P.G., Fenchel, T. & DeLong, E.F. (2008). The microbial engines that drive Earth's biogeochemical cycles. *Science*, 320, 1034–1039.
- Feder, M.E. & Walser, J.C. (2005). The biological limitations of transcriptomics in elucidating stress and stress responses. *J. Evol. Biol.*, 18, 901–910.
- Feist, A.M., Herrgard, M.J., Thiele, I., Reed, J.L. & Palsson, B.O. (2008). Reconstruction of biochemical networks in microorganisms. *Nat. Rev. Micro.*, 7, 129–143.
- Fitter, A.H. (2005). Darkness visible: reflections on underground ecology. *J. Ecol.*, 93, 231–243.
- Frias-Lopez, J., Shi, Y., Tyson, G.W., Coleman, M.L., Schuster, S.C., Chisholm, S.W. *et al.* (2008). Microbial community gene expression in ocean surface waters. *Proc. Natl Acad. Sci. USA*, 105, 3805–3810.
- Frigaard, N.U., Martinez, A., Mincer, T.J. & DeLong, E.F. (2006). Proteorhodopsin lateral gene transfer between marine planktonic Bacteria and Archaea. *Nature*, 439, 847–850.
- Fuhrman, J.A. (2009). Microbial community structure and its functional implications. *Nature*, 459, 193–199.
- Fuhrman, J.A., Schwalbach, M.S. & Stingl, U. (2008). Proteorhodopsins: an array of physiological roles? *Nat. Rev. Microbiol.*, 6, 488–494.
- Galperin, M.Y. & Koonin, E.V. (1998). Sources of systematic error in functional annotation of genomes: domain rearrangement, non-orthologous gene displacement and operon disruption. *In Silico Biol.*, 1, 55–67.

- Galperin, M.Y., Walker, D.R. & Koonin, E.V. (1998). Analogous enzymes: independent inventions in enzyme evolution. *Genome Res.*, 8, 779–790.
- Getz, W.M. (2003). Correlative coherence analysis: variation from intrinsic and extrinsic sources in competing populations. *Theor. Popul. Biol.*, 64, 89–99.
- Giovanoni, S.J., Tripp, H.J., Givan, S., Podar, M., Vergin, K.L., Baptista, D. *et al.* (2005). Genome streamlining in a cosmopolitan oceanic bacterium. *Science*, 309, 1242–1245.
- Hummel, I., Bourdais, G., Gouesbet, G., Couée, I., Malmberg, R.L. & El Amrani, A. (2004). Differential gene expression of Arginine Decarboxylase ADC1 and ADC2 in *Arabidopsis thaliana*: characterization of transcriptional regulation during seed germination and seedling development. *New Phytol.*, 163, 519–531.
- Kassen, R. (2002). The experimental evolution of specialists, generalists, and the maintenance of diversity. *J. Evol. Biol.*, 15, 173–190.
- Koonin, E.V. & Wolf, Y.I. (2008). Genomics of bacteria and archaea: the emerging dynamic view of the prokaryotic world. *Nucleic Acids Res.*, 36, 6688–6719.
- Krawczyk-Barsch, E., Grossmann, K., Arnold, T., Hofmann, S. & Wobus, A. (2008). Influence of uranium (VI) on the metabolic activity of stable multispecies biofilms studied by oxygen micro-sensors and fluorescence microscopy. *Geochim. Cosmochim. Acta*, 72, 5251–5265.
- Leininger, S., Ulrich, T., Schloter, M., Schwark, L., Qi, J., Nicol, G.W. *et al.* (2006). Archaea predominate among ammonia-oxidizing prokaryotes in soils. *Nature*, 442, 806–809.
- Levasseur, A., Pontarotti, P., Poch, O. & Thompson, J.D. (2008). Strategies for reliable exploitation of evolutionary concepts in high throughput biology. *Evol. Bioinform. Online*, 4, 121–137.
- López-García, P. & Moreira, D. (2008). Tracking microbial biodiversity through molecular and genomic ecology. *Res. Microbiol.*, 159, 67–73.
- Mahaffey, C., Michaels, A.F. & Capone, D.G. (2005). The conundrum of marine N<sub>2</sub> fixation. *Am. J. Sci.*, 305, 546–595.
- Marcy, Y., Ouverney, C., Bik, E.M., Lösekann, T., Ivanova, N., Martin, H.G. *et al.* (2007). Dissecting biological “dark matter” with single-cell genetic analysis of rare and uncultivated TM7 microbes from the human mouth. *Proc. Natl Acad. Sci. USA*, 104, 11889–11894.
- Martin, F., Aerts, A., Ahrén, D., Brun, A., Danchin, E.G.J., Duchaussoy, F. *et al.* (2008). The genome of *Laccaria bicolor* provides insights into mycorrhizal symbiosis. *Nature*, 452, 88–92.
- Medini, D., Serruto, D., Parkhill, J., Relman, D.A., Donati, C., Moxon, R. *et al.* (2008). Microbiology in the post-genomic era. *Nat. Rev. Microbiol.*, 6, 419–430.
- Méreau, A., Anquetil, V., Cibois, M., Noiret, M., Primot, A., Vallée, A. *et al.* (2009). Analysis of splicing patterns by pyrosequencing. *Nucleic Acids Res.*, 37, e126.
- Mitra, S., Klar, B. & Huson, D.H. (2009). Visual and statistical comparison of metagenomes. *Bioinformatics*, 25, 1849–1855.
- Montoya, J.P., Holl, C.M., Zehr, J.P., Hansen, A., Villareal, T.A. & Capone, D.G. (2004). High rates of N<sub>2</sub> fixation by unicellular diazotrophs in the oligotrophic Pacific Ocean. *Nature*, 430, 1027–1032.
- Moran, N.A., McCutcheon, J.P. & Nakabachi, A. (2008). Genomics and evolution of heritable bacterial symbionts. *Annu. Rev. Genet.*, 42, 165–190.
- Mou, X., Sun, S., Edwards, R.A., Hodson, R.E. & Moran, M.A. (2008). Bacterial carbon processing by generalist species in the coastal ocean. *Nature*, 451, 708–711.
- Pace, N.R. (1997). A molecular view of microbial diversity and the biosphere. *Science*, 276, 734–740.
- Palsson, B.O. (2006). *Systems Biology, Properties of Reconstructed Networks*, 1st edn. Cambridge University Press, Cambridge, New York.
- Parro, V., Moreno-Paz, M. & González-Toril, E. (2007). Analysis of environmental transcriptomes by DNA microarrays. *Environ. Microbiol.*, 9, 453–464.
- Prinzinger, A., Reiffers, R., Braakhekke, W.G., Hennekens, S.M., Tackenberg, O., Ozinga, W.A. *et al.* (2008). Less lineages - more trait variation: phylogenetically clustered plant communities are functionally more diverse. *Ecol. Lett.*, 11, 809–819.
- de Queiroz, K. (2007). Species concepts and species delimitation. *Syst. Biol.*, 56, 879–886.
- Quince, C., Lanzén, A., Curtis, T.P., Davenport, R.J., Hall, N., Head, I.M. *et al.* (2009). Accurate determination of microbial diversity from 454 pyrosequencing data. *Nat. Methods*, 6, 639–641.
- Raes, J., Foerster, K.U. & Bork, P. (2007). Get the most out of your metagenome: computational analysis of environmental sequence data. *Curr. Opin. Microbiol.*, 10, 490–498.
- Ram, R.J., Verberkmoes, N.C., Thelen, M.P., Tyson, G.W., Baker, B.J., Blake, R.C. *et al.* (2005). Community proteomics of a natural microbial biofilm. *Science*, 308, 1915–1920.
- Rasmussen, D.A. & Noor, M.A.F. (2009). What can you do with 0.1x genome coverage?: A case study based on a genome survey of the scuttle fly *Megaselia scalaris* (Phoridae). *BMC Genomics*, 10, 382.
- Rodrigue, S., Malmstrom, R.R., Berlin, A.M., Birren, B.W., Henn, M.R. & Chisholm, S.W. (2009). Whole genome amplification and *de novo* assembly of single bacterial cells. *PLoS ONE*, 4, e6864.
- Rodriguez-Brito, B., Rohwer, F. & Edwards, R. (2006). An application of statistics to comparative metagenomics. *BMC Bioinformatics*, 7, 162.
- Röling, W.F., van Brekel, B.M., Bruggeman, F.J. & Westerhoff, H.G. (2007). Ecological control analysis: being(s) in control of mass flux and metabolite concentrations in anaerobic degradation processes. *Environ. Microbiol.*, 9, 500–511.
- Rusch, D.B., Halpern, A.L., Sutton, G., Heidelberg, K.B., Williamson, S., Yooshep, S. *et al.* (2007). The Sorcerer II Global Ocean Sampling expedition: northwest Atlantic through eastern tropical Pacific. *PLoS Biol.*, 5, 77.
- Sabehi, G., Massana, R., Bielawski, J.P., Rosenberg, M., DeLong, E.F. & Béjà, O. (2003). Novel Proteorhodopsin variants from the Mediterranean and Red Seas. *Environ. Microbiol.*, 5, 842–849.
- Schmidt, P.S., Serrão, E.A., Pearson, G.A., Riginos, C., Rawson, P.D., Hilbish, T.J. *et al.* (2008). Ecological genetics in the North Atlantic: environmental gradients and adaptation at specific loci. *Ecology*, 89, 91.
- Shi, Y., Tyson, G.W. & DeLong, E.F. (2009). Metatranscriptomics reveals unique microbial small RNAs in the ocean’s water column. *Nature*, 459, 266–269.
- Shively, J.M., English, R.S., Baker, S.H. & Cannon, G.C. (2001). Carbon cycling: the prokaryotic contribution. *Curr. Opin. Microbiol.*, 4, 301–306.
- Silva-Filho, M.C. (2003). One ticket for multiple destinations: dual targeting of proteins to distinct subcellular locations. *Curr. Opin. Plant Biol.*, 6, 589–595.

- Singh, A.H., Doerks, T., Letunic, I., Raes, J. & Bork, P. (2009). Discovering functional novelty in metagenomes: examples from light-mediated processes. *J. Bacteriol.*, 191, 32–41.
- Stranger, B.E., Forrest, M.S., Dunning, M., Ingle, C.E., Beazley, C., Thorne, N. *et al.* (2007). Relative impact of nucleotide and copy number variation on gene expression phenotypes. *Science*, 315, 848–853.
- Takeda, K., Komuro, Y., Hayakawa, T., Oguchi, H., Ishida, Y., Murakami, S. *et al.* (2009). Mitochondrial phosphoglycerate mutase 5 uses alternate catalytic activity as a protein serine/threonine phosphatase to activate ASK1. *Proc. Natl Acad. Sci. USA*, 106, 12301–12305.
- Tatusov, R.L., Fedorova, N.D., Jackson, J.D., Jacobs, A.R., Kiryutin, B., Koonin, E.V. *et al.* (2003). The COG database: an updated version includes eukaryotes. *BMC Bioinformatics*, 4, 41.
- Teixeira, M.C., Duque, P. & Sá-Correia, I. (2007). Environmental genomics: mechanistic insights into toxicity of and resistance to the herbicide 2,4-D. *Trends Biotechnol.*, 25, 363–370.
- de la Torre, J.R., Christianson, L.M., Bèjà, O., Suzuki, M.T., Karl, D.M., Heidelberg, J. *et al.* (2003). Proteorhodopsin genes are distributed among divergent marine bacterial taxa. *Proc. Natl Acad. Sci. USA*, 100, 12830–12835.
- Torsvik, V., Øvreås, L. & Thingstad, T.F. (2002). Prokaryotic diversity—magnitude, dynamics, and controlling factors. *Science*, 296, 1064–1066.
- Treusch, A.H., Leininger, S., Kletzin, A., Schuster, S.C., Klenk, H. & Schleper, C. (2005). Novel genes for nitrite reductase and Amo-related proteins indicate a role of uncultivated mesophilic crenarchaeota in nitrogen cycling. *Environ. Microbiol.*, 7, 1985–1995.
- Tringe, S.G., von Mering, C., Kobayashi, A., Salamov, A.A., Chen, K., Chang, H.W. *et al.* (2005). Comparative metagenomics of microbial communities. *Science*, 308, 554–557.
- Tyson, G.W., Chapman, J., Hugenholtz, P., Allen, E.E., Ram, R.J., Richardson, P.M. *et al.* (2004). Community structure and metabolism through reconstruction of microbial genomes from the environment. *Nature*, 428, 37–43.
- Ungerer, M.C., Johnson, L.C. & Herman, M.A. (2008). Ecological genomics: understanding gene and genome function in the natural environment. *Heredity*, 100, 178–183.
- Vandenkoornhuysen, P., Mahé, S., Ineson, P., Staddon, P., Ostle, N., Cliquet, J.B. *et al.* (2007). Active root-inhabiting microbes identified by rapid incorporation of plant-derived carbon into RNA. *Proc. Natl Acad. Sci. USA*, 104, 16970–16975.
- Venter, J.C., Remington, K., Heidelberg, J.F., Halpern, A.L., Rusch, D., Eisen, J.A. *et al.* (2004). Environmental genome shotgun sequencing of the Sargasso Sea. *Science*, 304, 66–74.
- Vera, J.C., Wheat, C.W., Fescemyer, H.W., Frilander, M.J., Crawford, D.L., Hanski, I. *et al.* (2008). Rapid transcriptome characterization for a nonmodel organism using 454 pyrosequencing. *Mol. Ecol.*, 17, 1636–1647.
- Westerhoff, H.V. & Palsson, B.O. (2004). The evolution of molecular biology into systems biology. *Nature Biotech.*, 22, 1249–1252.
- Woyke, T., Teeling, H., Ivanova, N.N., Huntemann, M., Richter, M., Gloeckner, F.O. *et al.* (2006). Symbiosis insights through metagenomic analysis of a microbial consortium. *Nature*, 443, 950–955.
- Woyke, T., Xie, G., Copeland, A., González, J.M., Han, C., Kiss, H. *et al.* (2009). Assembling the marine metagenome, one cell at a time. *PLoS ONE*, 4, 5299.
- Wullschlegel, S.D., Leakey, A.D.B. & St Clair, S.B. (2007). Functional genomics and ecology—a tale of two scales. *New Phytol.*, 176, 735–739.
- Yang, L. & Bennetzen, J.L. (2009). Structure-based discovery and description of plant and animal helitrons. *Proc. Natl Acad. Sci. USA*, 106, 12832–12837.
- Yooseph, S., Sutton, G., Rusch, D.B., Halpern, A.L., Williamson, S.J., Remington, K. *et al.* (2007). The Sorcerer II Global Ocean Sampling expedition: expanding the universe of protein families. *PLoS Biol.*, 5, e16.
- Zehr, J.P., Waterbury, J.B., Turner, P.J., Montoya, J.P., Omoregie, E., Steward, G.F. *et al.* (2001). Unicellular cyanobacteria fix N<sub>2</sub> in the subtropical North Pacific Ocean. *Nature*, 412, 635–638.
- Zehr, J.P., Bench, S.R., Carter, B.J., Hewson, I., Niazi, F., Shi, T. *et al.* (2008). Globally distributed uncultivated oceanic N<sub>2</sub>-fixing cyanobacteria lack oxygenic photosystem II. *Science*, 322, 1110–1112.
- Zhu, H. & Snyder, M. (2002). “Omic” approaches for unraveling signaling networks. *Curr. Opin. Cell Biol.*, 14, 173–179.

## GLOSSARY

*cDNA*: complementary DNA, reverse-transcribed or copied from an RNA template.

*ChIP-on-chip* approaches: a method combining Chromatine Immuno Precipitation (ChIP) with microarray (chip) technology to study interactions between proteins and DNA. Mainly used to determine the locations of binding sites, and to understand gene expression and regulation.

*Environmental genomics*: analysis of large-scale sequence-based information (such as DNA) obtained from a variety of environmental samples, at cell, organism, population, and community levels.

*Environmental post-genomics*: gene functional characterisation approaches and genome-wide expression analyses in an environmental context. Includes *transcriptomics* (analysis of the complete set of transcripts), mainly by mass sequencing of transcript-derived cDNA, through the development of second-generation sequencing machines, and *proteomics* (analysis of the complete set of proteins), mainly by coupled liquid or gas-chromatography/ tandem mass spectrometer (LC and GC-MS-MS) for *de novo* identification. Along with the metagenome, the *metatranscriptome* and the *metaproteome* can be analysed, when considering a community of organisms.

*Fosmid*: particular vector designed for the insertion of large-size DNA fragments.

*Gene synteny*: The order of genes on a chromosome region and its conservation.

*Homologous and heterologous identification*: similarity-based analysis of unknown genes by comparison with same-species (homologous) or cross-species (heterologous) sequences.



*Insert*: a DNA fragment of interest that has been cloned within a vector, such as a plasmid (non-chromosomal bacterial DNA) or a fosmid.

*Large insert libraries*: collections of cloned DNA inserts of long size (>20 000 bp).

*Metagenomics*: analysis of a mixed set of genomes from a community of organisms.

*Microarray, DNA array, DNA chip*: membrane or glass-slide surface where known DNA sequences are fixed, each at specific XY coordinates, to act as anchors for their complementary sequences. Mainly used for simultaneous expression surveys of a great number of genes.

*Microfluidics*: devices designed to manipulate and analyse microliter volumes of fluids in order to assay the composition within small samples.

*Open reading frame*: nucleotide sequence located between a start codon and a stop codon, thus potentially translatable into a polypeptide.

*Orthologs*: homologous genes found in different genomes and resulting from the duplication of an ancestral sequence during a speciation event.

*Overexpression*: molecular biology method resulting in the enhanced expression of a gene of interest in order to characterise its product or its impact on phenotype

*Paralogs*: homologous sequences resulting from an internal duplication event and belonging to the same species genome.

*Promoter*: sequence region upstream of a gene, to which RNA polymerase binds for gene transcription, and involved in gene expression regulation. Promoters can work in concert with other regulatory elements.

*Regulatory cis-acting element*: short DNA sequence in a promoter interacting with transcription factors to regulate expression of the downstream gene.

*Shotgun sequencing*: random sequencing of a high number of short anonymous fragments of genomes

## SUPPORTING INFORMATION

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

**Box S1** Strategies in environmental genomics.

**Box S2** Selection of bioinformatics tools for analysing environmental genomics data.

As a service to our authors and readers, this journal provides supporting information supplied by the authors. Such materials are peer-reviewed and may be re-organized for online delivery, but are not copy-edited or typeset. Technical support issues arising from supporting information (other than missing files) should be addressed to the authors.

Editor, Nancy Johnson

Manuscript received 19 October 2009

First decision made 16 November 2009

Second decision made 31 January 2010

Manuscript accepted 24 February 2010