



HAL
open science

Phylogénomique des structures multiprotéiques eucaryotes impliquées dans le cycle cellulaire et contribution à la phylogénie des eucaryotes.

Laura Eme

► To cite this version:

Laura Eme. Phylogénomique des structures multiprotéiques eucaryotes impliquées dans le cycle cellulaire et contribution à la phylogénie des eucaryotes.. Evolution [q-bio.PE]. Université de la Méditerranée, 2011. Français. NNT: . tel-03368378

HAL Id: tel-03368378

<https://hal.science/tel-03368378>

Submitted on 6 Oct 2021

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.

Université de la Méditerranée
Faculté des Sciences de Luminy

THÈSE DE DOCTORAT

Bioinformatique, Biochimie structurale et Génomique

Présentée et soutenue publiquement par

Laura EME

le 1^e juin 2011

Phylogénomique des structures
multiprotéiques eucaryotes impliquées
dans le cycle cellulaire et contribution à la
phylogénie des eucaryotes

Jury

Céline BROCHIER-ARMANET

Emmanuel DOUZERY

Manolo GOUY

André GILLES

Matthieu PIEL

Pascal RIHET

Directrice de thèse

Rapporteur

Rapporteur

Examineur

Examineur

Examineur

REMERCIEMENTS

En premier lieu, je tiens à remercier vivement Emmanuel DOUZERY et Manolo GOUY d'avoir accepté d'évaluer ce travail en tant que rapporteurs, ainsi qu'André GILLES, Matthieu PIEL et Pascal RIHET pour leur participation à mon jury de thèse.

Je remercie Frédéric BARRAS de m'avoir accueillie au sein de son laboratoire durant ces quelques années, ainsi que l'ensemble des membres du LCB pour leur convivialité.

Merci également à Emmanuel TALLA pour m'avoir accueillie au sein de l'équipe Génome, Evolution et Bioinformatique.

Je remercie chaleureusement Céline BROCHIER-ARMANET pour son investissement sans faille durant toutes ces années et pour la qualité exceptionnelle de son enseignement. Merci de m'avoir proposé ce projet stimulant aux multiples facettes, j'ai aimé en relever le défi. Ça a été un réel plaisir de travailler à tes côtés.

Je remercie très sincèrement David MOREIRA qui a suivi mon travail depuis le début. Merci pour ta disponibilité et tes conseils toujours précieux. Merci également pour ton aide dans la recherche (fructueuse!) d'un stage postdoctoral. Je suis ravie d'avoir eu l'opportunité de travailler avec toi.

Un merci tout particulier à Simonetta GRIBALDO pour son enthousiasme communicatif et les échanges précieux que nous avons pu avoir, à Purificación LOPEZ-GARCIA pour nos - trop rares mais néanmoins vivifiantes - discussions, ainsi qu'à Elie DESMOND pour nos conversations passionnées entre eucaryotistes!

Merci également à Amel LATIFI et Cheng-Cai ZHANG pour cette enrichissante collaboration.

Je remercie sincèrement Pascal HINGAMP pour sa participation à mon comité de suivi de thèse, ses nombreuses lettres de recommandation, et pour m'avoir fait partager sa stimulante vision de l'enseignement.

Merci à Rémi PLANEL pour le développement de Sesame, la gestion du cluster et pour m'avoir aidée dans mes galères unix-iennes!

Un clin d'œil à l'ensemble des membres passés et présents de l'équipe GEB, auprès de qui j'ai passé d'agréables années.

Une attention particulière à Aurélie TRILLES, que j'ai réellement apprécié encadrer pendant son stage de M1. Je te souhaite une bonne continuation.

Une pensée personnelle pour mes collègues et amies, Céline PETITJEAN et Rym AGREBI, pour leurs attentions petites et grandes, et pour tous les bons moments passés ensemble.

Un grand merci à Séverine GAGNOT et Jonathan LOMBARD pour leurs précieuses relectures.

Enfin, une profonde pensée pour mes proches, famille et amis, qui m'ont constamment soutenu tout au long de ce parcours.

INTRODUCTION

1	L'ARBRE UNIVERSEL DU VIVANT	11
2	HISTOIRE EVOLUTIVE DES EUCARYOTES.....	15
3	ORIGINE EVOLUTIVE DES EUCARYOTES.....	16
3.1	RELATIONS ENTRE EUCARYA ET LES AUTRES DOMAINES	16
3.1.1	<i>Lien entre Eucarya et Archaea.....</i>	16
3.1.2	<i>Lien entre Eucarya et Bacteria.....</i>	16
3.2	HYPOTHESES SUR LES LIENS UNISSANT EUCARYA ET ARCHAEA.....	16
3.2.1	<i>L'hypothèse 3D</i>	19
3.2.2	<i>L'hypothèse 2D</i>	19
3.2.2.1	Les hypothèses de fusion	20
3.2.2.2	Les hypothèses de symbioses entre Alpha-proteobacteria et Archaea	20
3.2.2.3	Les symbioses de trois espèces procaryotes	21
3.3	ARGUMENTS EN FAVEUR DE L'UNE OU L'AUTRE DES HYPOTHESES	23
4	LA PHYLOGENIE DES EUCARYA.....	27
4.1	LES PREMIERES PHYLOGENIES MOLECULAIRES.....	27
4.2	L'UTILISATION DE NOUVEAUX MARQUEURS.....	27
4.3	LES CAUSES D'IRRÉSOLUTION ET LES ARTEFACTS DE RECONSTRUCTION	28
4.3.1	<i>Artefacts méthodologiques</i>	28
4.3.2	<i>Radiation évolutive.....</i>	29
4.3.3	<i>Saturation moléculaire</i>	30
4.3.4	<i>Les transferts latéraux entre eucaryotes</i>	30
4.4	COMMENT PALLIER AU MANQUE DE RESOLUTION ?.....	31
4.4.1	<i>Amélioration des modèles et méthodes</i>	31
4.4.2	<i>Améliorer l'échantillonnage taxonomique.....</i>	31
4.4.3	<i>La détection d'évènements génétiques rares</i>	32
4.4.4	<i>De la phylogénie à la phylogénomique.....</i>	32
4.5	LA PHYLOGENIE DES EUCARYA ACTUELLEMENT ADMISE ET LES GRANDES QUESTIONS	34
4.5.1	<i>Opisthokonta</i>	35
4.5.2	<i>Amoebozoa.....</i>	36
4.5.3	<i>Plantae.....</i>	37
4.5.4	<i>Rhizaria.....</i>	38
4.5.5	<i>Excavata</i>	40
4.5.6	<i>Chromalveolata</i>	41
4.5.6.1	Haptophyta	41
4.5.6.2	Cryptophyta	41
4.5.6.3	Alveolata	42
4.5.6.4	Heterokonta	43
4.5.6.5	Les endosymbioses secondaires et tertiaires.....	43
4.5.7	<i>Les relations entre grands groupes, la racine de l'arbre des Eucarya.....</i>	47
5	LECA, LE DERNIER ANCETRE COMMUN DES EUCARYA	49
5.1	DEUX APPROCHES POUR INFERER LES CARACTERISTIQUES DE LECA.....	49
5.1.1	<i>Reconstruction de protéomes ancestraux complets.....</i>	49
5.1.1.1	Le protéome de LECA : aspect quantitatifs	49
5.1.1.2	Le protéome ancestral mitochondrial	50
5.1.1.3	Avantages et inconvénients de ces approches.....	50
5.1.2	<i>Etude phylogénomique de l'évolution des systèmes biologiques</i>	51
5.1.3	<i>Les SME.....</i>	52
5.1.4	<i>LECA, une cellule très complexe.....</i>	54
5.1.4.1	Caractéristiques génomiques.....	54
5.1.4.2	Processus informationnels	54
5.1.4.3	Le métabolisme	54

5.1.4.4	Compartiments cellulaires	54
5.1.4.5	Processus cellulaires	55
5.2	OBJET DE CETTE THESE	56

AXE 1 : EVOLUTION DES SYSTEMES EUCARYOTES

1	EVOLUTION DE LA CYTOKINESE.....	65
1.1	LE MIDBODY	65
1.2	LA DIVERSITE DES MECANISMES DE CYTOKINESE CHEZ LES EUCARYA	67
1.3	OBJECTIF	71
1.4	UN MIDBODY CHEZ L'ANCETRE DES METAZOAIREES	73
1.5	UN MIDBODY CHEZ L'ANCETRE DES UNIKONTES ?	73
1.6	UN MIDBODY CHEZ LECA ?.....	73
1.6.1	<i>Les orthologues chez les Bikontes.....</i>	73
1.6.2	<i>La majorité des composants existaient chez LECA.....</i>	75
1.6.3	<i>Fonction des composants chez LECA.....</i>	76
1.6.3.1	Distribution fonctionnelle des composants ancestraux	76
1.6.3.2	Conservation de la composition en domaines fonctionnels des orthologues.....	76
1.6.3.3	Fonction cellulaire des orthologues	77
1.7	ET CHEZ LES PROCARYOTES ?	78
1.8	DISCUSSION ET PERSPECTIVES	81
2	EVOLUTION DU CYCLOSOME	85
2.1	INTRODUCTION	85
2.1.1	<i>Les ubiquitine-ligases.....</i>	85
2.1.2	<i>L'Anaphase Promoting Complex.....</i>	86
2.2	STRUCTURE DE L'APC/C.....	87
2.2.1	<i>Le cœur</i>	87
2.2.2	<i>Les adaptateurs, responsables de la reconnaissance du substrat</i>	89
2.3	ROLES DE L'APC/C.....	90
2.3.1	<i>Dégradation des cyclines</i>	90
2.3.2	<i>Séparation de chromatides-sœurs.....</i>	91
2.3.3	<i>La vie sans APC/C.....</i>	92
2.3.4	<i>Qu'en est-il de l'APC/C chez les autres eucaryotes ?.....</i>	93
2.4	ANALYSE PHYLOGENOMIQUE DE L'APC/C ET DE SES PRINCIPALES CIBLES	93
2.4.1	<i>La plupart des composants et adaptateurs existaient chez LECA.....</i>	98
2.4.1.1	La majorité des composants a été héritée de LECA et conservée au cours de l'évolution des eucaryotes	99
2.4.1.2	Des composants hérités de LECA ont subi de multiples pertes	99
2.4.2	<i>Plusieurs composants sont issus de duplications ou d'innovations récentes.....</i>	105
2.4.3	<i>Certaines lignées ont subi des pertes importantes de composants</i>	105
2.4.3.1	Vers la définition d'un APC/C minimal ?	105
2.4.3.2	Pertes massives.....	106
2.4.4	<i>La majorité des cibles existaient chez LECA.....</i>	107
2.4.4.1	L'APC/C est le résultat d'innovations eucaryotes.....	108
2.4.5	<i>Conclusion.....</i>	109
3	EVOLUTION DU SYSTEME SRX/PRX	113
3.1	INTRODUCTION	113
3.1.1	<i>Les peroxirédoxines.....</i>	113
3.1.1.1	Les peroxyrédoxines à deux cystéines typiques.....	113
3.1.1.2	Les peroxyrédoxines à deux cystéines atypiques.....	115
3.1.1.3	Les peroxyrédoxines à une cystéine.....	115
3.1.2	<i>Suroxydation des 2-Cys Prx typiques</i>	115
3.1.3	<i>Rôle du cycle de suroxydation.....</i>	117
3.1.3.1	Rôle dans la transduction du signal.....	117
3.1.3.2	Rôle de chaperonnes	117

3.2	ANALYSES PHYLOGENETIQUES DE SRX ET DE 2-CYS PRX	118
3.2.1	<i>Matériel et méthodes</i>	119
3.2.1.1	Recherche d'homologues	119
3.2.1.2	Construction des jeux de données	119
3.2.1.3	Reconstructions phylogénétiques	119
3.2.1.4	Recherche des séquences d'adressage	120
3.2.2	<i>Analyses phylogénétiques</i>	121
3.2.2.1	Répartition taxonomique	121
3.2.2.1.1	Prx	121
3.2.2.1.2	Srx	124
3.2.2.2	Phylogénies	126
3.2.2.2.1	Prx	126
3.2.2.2.2	Srx	130
3.2.2.3	Localisation subcellulaire	132
3.2.3	<i>Scénario évolutif</i>	134
3.2.3.1	Scenarion proposé par Basu <i>et al.</i> 2005	134
3.2.3.2	Notre scénario sur l'émergence de Srx et des 2-Cys Prx sensibles.....	137
3.2.3.2.1	Hypothèse 1 : Origine cyanobactérienne	137
3.2.3.2.2	Hypothèse 2 : Origine eucaryote	137
3.2.3.2.3	Origine des deux copies chez les Plantae.....	138
3.3	CONCLUSION ET PERSPECTIVES	140

AXE 2 : PHYLOGENIE DES EUCARYA

1	INTRODUCTION	143
2	METHODES	145
2.1	JEUX DE DONNEES ANALYSES.....	145
2.2	ANALYSES PHYLOGENETIQUES	150
2.2.1	<i>Méthodes de reconstruction</i>	150
2.2.2	<i>Support statistique</i>	150
2.2.3	<i>Recodage</i>	150
2.2.4	<i>Modèles d'évolution</i>	150
2.2.5	<i>Saturation et Vitesse d'évolution des marqueurs</i>	151
3	RESULTATS	151
3.1	LES CONCATENATIONS "MIDBODY"	151
3.1.1	<i>Effet du recodage</i>	151
3.1.2	<i>La supermatrice M75-67 (concaténation de 75 protéines)</i>	154
3.1.3	<i>La supermatrice M35-67 (concaténation de 35 protéines)</i>	154
3.2	CONCATENATION "RODRIGUEZ-EZPELETA" MISE A JOUR	157
3.3	CONCATENATIONS DE 230 PROTEINES (M75-56 + RE)	163
3.4	LA SUPERMATRICE APC.....	166
4	POINTS DE DEBATS	169
4.1	LA MONOPHYLIE DES PLANTES	170
4.2	PLACEMENT DES MICROSPORIDIUMS AU SEIN DES FUNGI	172
4.3	LA MONOPHYLIE DES EXCAVATA	173
4.4	LE PLACEMENT D'EMILIANIA.....	173
4.5	RELATIONS AU SEIN DES HOLOZOA.....	174
4.6	ECDYSOZOA OU COELOMATA ?	175
4.7	MONOPHYLIE DES AMOEBOZOA.....	175
4.8	MONOPHYLIE DES UNIKONTA	175
4.9	L'HYPOTHESE CHROMALVEOLES.....	175
5	CONCLUSION	176

AXE 3 : RELATIONS AVEC LES ARCHAEA

1	INTRODUCTION	179
2	DESCRIPTION DU SITE D'ECHANTILLONNAGE ET DIVERSITE BIOLOGIQUE.....	181
2.1	DIVERSITE MICROBIENNE	181
2.2	ISOLATION DE GRANDS CONTIGS DE THAUMARCHAEOTA	183
3	MATERIEL ET METHODES.....	186
3.1	CONSTRUCTION DES JEUX DE DONNEES	186
3.1.1	<i>Données nucléotidiques</i>	186
3.1.2	<i>Données protéiques</i>	186
3.2	RECONSTRUCTIONS PHYLOGENETIQUES.....	187
4	RESULTATS	188
4.1	RELATIONS PHYLOGENETIQUES AU SEIN DES SEQUENCES DE THAUMARCHAEOTA	188
4.2	POSITION PHYLOGENETIQUE DES THAUMARCHAEOTA AU SEIN DES ARCHAEA.....	190
4.3	RELATIONS ENTRE EUCARYA ET ARCHAEA.....	193
4.3.1	<i>Hypothèse 3D</i>	193
4.3.2	<i>Hypothèse 2D</i>	193
5	CONCLUSION.....	196
 CONCLUSION		 261
BIBLIOGRAPHIE.....		265
ANNEXE 1 : Article 1		285
ANNEXE 2 : Article 2		xxx
ANNEXE 3 : Article 3		xxx

INTRODUCTION

1 L'arbre universel du vivant

Dans les années 1960 et 1970, le séquençage et la comparaison des premières séquences porteuses d'information biologique (protéines et ADN) ont conduit **Zuckerlandl et Pauling** à proposer que ces séquences représentent une source d'information riche pouvant être utilisée pour reconstruire leur histoire évolutive et celle des organismes qui les portent (Zuckerlandl and Pauling, 1965). L'ère de la phylogénie moléculaire est née.

Le pouvoir révolutionnaire de ces méthodes moléculaires devient évident treize ans plus tard, quand **Carl Woese** et **George Fox** développent une technique de comparaison des séquences d'ARN de la petite sous-unité du ribosome (ARNr SSU). Cet ARN (ARN 16S chez les procaryotes et 18S chez les eucaryotes) fut choisi en raison de son abondance dans les cellules, de sa distribution universelle, et parce qu'il évolue lentement, ce qui implique la possibilité de l'utiliser pour l'étude des relations entre organismes très éloignés. Le principe de la technique de Woese et Fox est simple : après purification de l'ARNr SSU d'un organisme particulier, la préparation est soumise à une enzyme (la ribonucléase T1), qui digère la molécule d'ARNr SSU en petits fragments (ou "oligonucléotides"). Les oligonucléotides sont ensuite séparés les uns des autres par électrophorèse en deux dimensions afin de donner une empreinte (Figure 1).

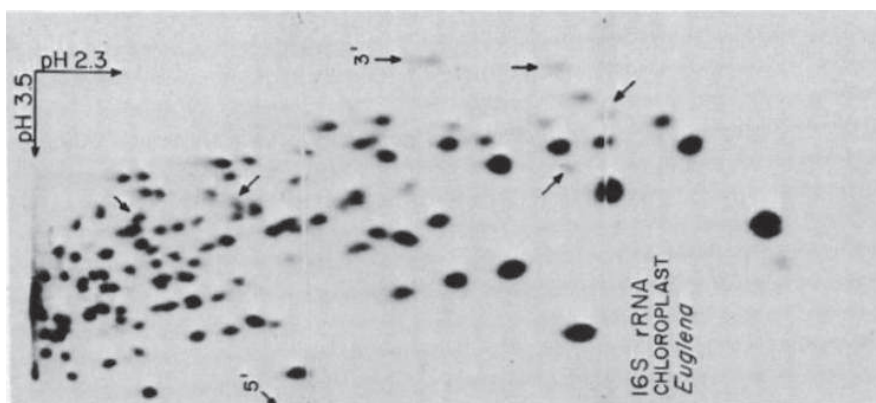


Figure 1. Empreinte électrophorétique en deux dimensions d'une digestion par l'ARNr SSU de chloroplaste (Zablen et al., 1975).

Chaque espèce peut donc être caractérisée par le profil de migration des oligonucléotides. En 1977, Woese et Fox publient un article dans lequel ils comparent les profils de digestion des séquences nucléotidiques de l'ARNr de la petite sous-unité de 13 procaryotes et eucaryotes. La comparaison des profils de migration pour chaque paire d'espèces étudiées permet l'obtention d'une constante d'association qui traduit la similitude entre oligonucléotides issus de ces deux organismes.

La phylogénie basée sur ces profils a révolutionné la vision du monde vivant que l'on avait jusqu'alors. En effet, elle a permis de montrer non seulement que certains de ces organismes (ex : les méthanogènes) ne font pas partie du groupe défini par les autres bactéries, mais également qu'ils sont aussi éloignés de ces dernières que des eucaryotes (Woese and Fox, 1977). Ainsi, ces observations ont permis de découvrir que le monde vivant cellulaire ne se divise pas seulement – comme on le croyait jusqu'alors – en deux domaines, sur la base de la présence ou non d'un noyau (Chatton, 1937; Stanier, 1961), mais plutôt en trois grandes lignées évolutives (appelées aussi "domaines") : les Eucarya, les Bacteria et les Archaea (Woese et al., 1990). Par ailleurs, ces résultats, associés à ceux de la biologie moléculaire, ont permis de confirmer l'unité du monde vivant : tous les êtres vivants actuels sont formés des

mêmes macromolécules, utilisent la même machinerie de synthèse protéique et le même code génétique, ce qui suggère qu'ils descendent tous d'un même ancêtre commun, LUCA (Last Universal Common Ancestor) (Figure 2).

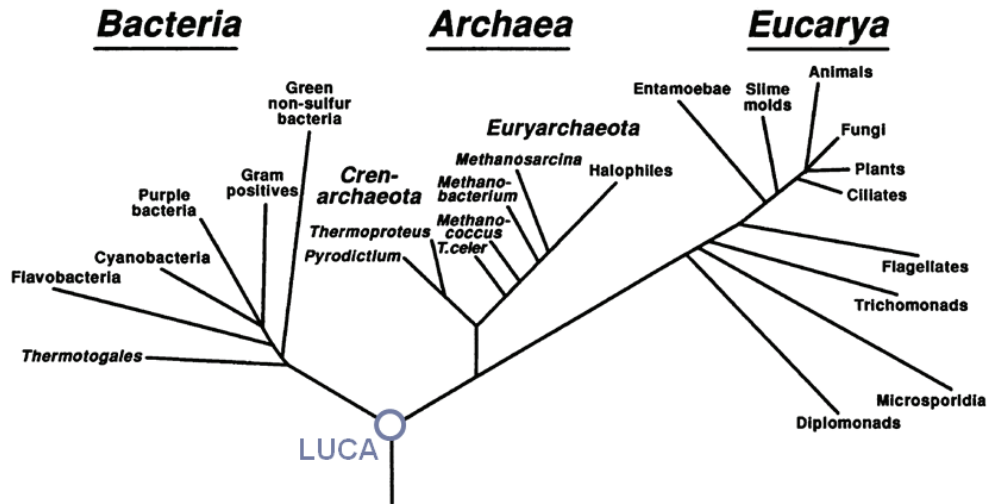


Figure 2. Phylogénie universelle du vivant. Adapté de Woese 1994.

Il est important de garder à l'esprit que cela ne préjuge en rien d'une origine unique ou multiple de la vie sur Terre – sujet toujours très discuté. Néanmoins, si d'autres formes de vie ont vu le jour sur notre planète, cela implique qu'elles n'ont pas laissé de descendants vivant encore aujourd'hui (Figure 3).

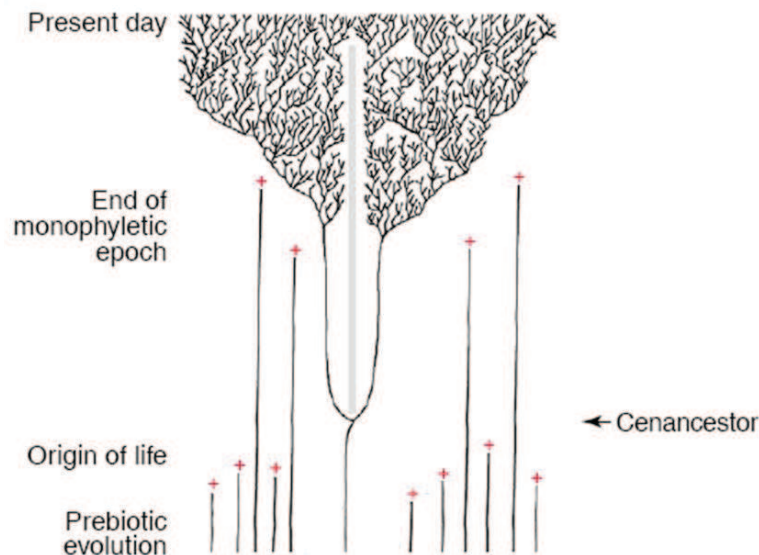


Figure 3. Scénario sur l'origine et l'évolution précoce de la vie sur Terre (adapté de (Zhaxybayeva and Gogarten, 2004)). Au cours de l'évolution prébiotique, de nombreuses lignées auraient émergé indépendamment, et certaines auraient disparu (croix rouges). Ainsi, malgré l'émergence potentiellement multiple de la vie, tous les organismes actuels descendent d'un ancêtre commun unique.

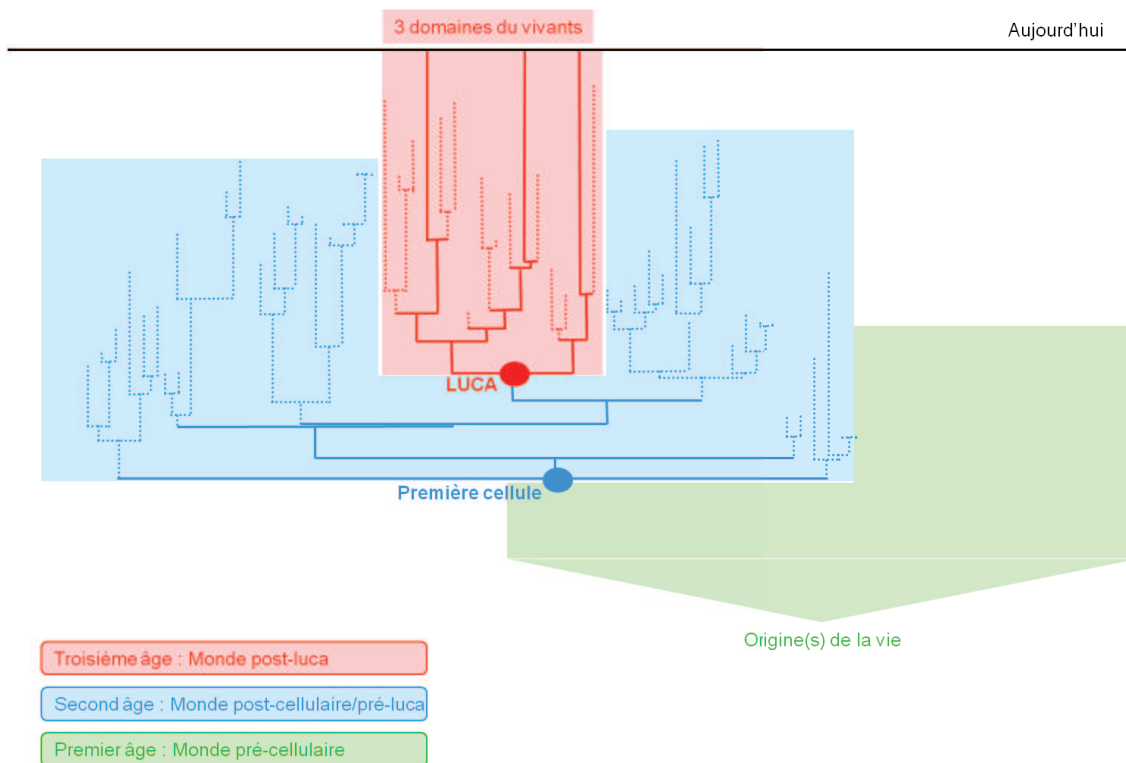


Figure 4. Évolution de la vie sur terre depuis son (ou ses) origine(s). Trois grandes étapes possiblement chevauchantes peuvent être distinguées : une phase incluant les premières formes de vie non cellulaires ("monde pré-cellulaire"), une qui regroupe des descendants de la première cellule ("monde post-cellulaire/pré-LUCA") et qui s'achève avec LUCA, et, enfin, la troisième période, qui commence avec LUCA et se poursuit jusqu'à nos jours ("monde post-LUCA"). Aucune échelle temporelle n'est indiquée car on ne connaît pas la durée de chaque phase (Forterre et al., 2005)

La notion de "dernier" ancêtre commun (*i.e.* le plus récent par rapport à aujourd'hui) est particulièrement importante, car elle permet d'éviter la confusion extrêmement fréquente qui est faite entre LUCA et la première cellule (Figure 4) : c'est après LUCA que les trois grands domaines du vivant reconnus actuellement, Archaea, Bacteria et Eucarya, ont divergé. La question de savoir à quoi ressemblait LUCA fait l'objet de nombreux débats dans la communauté scientifique. Carl Woese (Woese and Fox, 1977) a souligné très tôt l'importance cruciale de cette question. En effet, LUCA est un organisme clé dans l'histoire de la vie sur Terre, car il constitue un jalon intermédiaire entre les origines de la vie et les organismes actuels (Figure 4). Connaître LUCA permettrait de déterminer les mécanismes moléculaires qui s'étaient déjà mis en place et les étapes évolutives qui ont eu lieu entre la première cellule et l'apparition de cet ancêtre universel ("monde post-cellulaire/pré-LUCA") et d'entreprendre des études pour essayer d'approfondir notre compréhension de l'émergence des trois grands domaines du vivant qui lui ont succédé ("monde post-LUCA").

L'avancée des connaissances sur la nature de LUCA représente une longue quête qui pourrait s'appuyer, d'une part sur la détermination des relations de parenté entre les trois domaines du vivant, afin de pouvoir inférer la présence ou non de certains caractères chez LUCA, et d'autre part, sur la caractérisation du dernier ancêtre commun exclusif à chaque domaine du vivant que sont LECA (*Last Eukaryotic Common Ancestor*, pour "dernier ancêtre commun des eucaryotes"), LACA (*Last Archaeal Common Ancestor*, pour "dernier ancêtre commun des archées") et LBCA (*Last Bacterial Common Ancestor*, pour "dernier ancêtre commun des bactéries"). En effet, la caractérisation de ces ancêtres permettrait de poser un point de

repère intermédiaire entre LUCA et les organismes actuels, et de déterminer le moment où se sont mises en places les caractéristiques des organismes modernes.

Mon travail de doctorat s'inscrit dans ces problématiques. Je me suis plus particulièrement intéressée à la caractérisation du dernier ancêtre des eucaryotes (*i.e.* LECA), et l'évolution de ce domaine en général.

2 Histoire évolutive des eucaryotes

L'étude de l'origine et l'évolution des eucaryotes peut s'organiser autour de trois grands axes :

- Déterminer leurs relations avec les Archaea et les Bacteria. Il s'agit à la fois de s'intéresser à leurs relations de parenté, mais aussi aux relations d'échange de matériel génétique entre organismes n'appartenant pas au même domaine. L'objectif est donc de distinguer, au niveau du matériel génétique, l'héritage vertical de l'acquisition par transfert horizontal ;
- Déterminer les relations entre les différentes lignées d'eucaryotes, c'est-à-dire ici aussi, leurs relations de parenté, et les échanges de matériel génétique qui ont pu avoir lieu entre lignées distantes. L'objectif est donc de comprendre comment s'est faite leur diversification et quelles en ont été les conséquences biologiques ;
- Caractériser LECA du point de vue biologique, et donc de distinguer les composants/systèmes biologiques qui se sont mis en place avant ou après l'existence de cet ancêtre. Si leur apparition est antérieure à LECA, mais que ces caractères sont spécifiques des eucaryotes, alors cela signifie qu'ils sont apparus dans le "stem" des eucaryotes, c'est-à-dire dans la branchée située entre le premier eucaryote et le dernier ancêtre commun des eucaryotes (*i.e.* avant la diversification qui a donné lieu aux grands groupes eucaryotes actuels) (Figure 5, trait bleu à la base des eucaryotes). S'ils sont partagés avec les procaryotes, il est intéressant de suivre leur évolution au sein des eucaryotes. Si leur apparition est légèrement postérieure à LECA, alors ces composants ont pu jouer un rôle clé dans l'émergence précoce des grands groupes eucaryotes. Enfin si leur apparition est plus récente, ils ont plutôt eu un rôle dans la diversification au sein des grands groupes. Ces aspects sont d'une grande importance, aussi bien du point de vue biologique et qu'évolutif.

Le travail réalisé autour de ces trois axes aura un impact majeur sur notre compréhension de l'émergence et l'évolution des caractères eucaryotes-spécifiques, ainsi que sur notre compréhension des aspects fondamentaux de la biologie des cellules eucaryotes. Au cours de ma thèse, je me suis attachée à travailler sur ces chacun de ces trois aspects.

3 Origine évolutive des eucaryotes

3.1 Relations entre Eucarya et les autres domaines

3.1.1 Lien entre Eucarya et Archaea

Depuis la découverte des Archaea, il est couramment admis qu'Archaea et Eucarya partagent un ancêtre commun exclusif. En effet, il existe de nombreux caractères et mécanismes moléculaires communs aux archées et aux eucaryotes (Brown and Doolittle, 1997; Gribaldo et al., 2010) :

- De très nombreux composants des systèmes dits "informationnels", c'est-à-dire impliqués dans la transmission et l'expression de l'information génétique (réparation, réplication, transcription et traduction), présentent une plus grande similarité de séquence entre Archaea et Eucarya qu'avec leurs homologues bactériens, lorsqu'ils existent (protéines ribosomales, facteurs de traduction, ARN polymérase, ARN nucléolaires, exosome, histones, etc.) (Bell and Jackson, 2001; Cubonova et al., 2005; Edgell and Doolittle, 1997; Gaspin et al., 2000; Gribaldo and Brochier-Armanet, 2006; Hartung and Hopfner, 2009; Lecompte et al., 2002; Leipe et al., 1999; Londei, 2005; Omer et al., 2000; Reeve et al., 1997; Werner, 2008).
- Certains systèmes "opérationnels", c'est-à-dire impliqués dans les fonctions liées à l'organisation et au maintien de la cellule, et au métabolisme) semblent également homologues entre Archaea et Eucarya (ATPases vacuolaires, voies de sécrétion, etc.) (Gogarten et al., 1989) (Gribaldo and Brochier-Armanet, 2006; Gribaldo and Cammarano, 1998; Lindas et al., 2008; Samson et al., 2008).

3.1.2 Lien entre Eucarya et Bacteria

Les Eucarya partagent également un lien évolutif particulier avec les bactéries, mais de nature différente, puisqu'il s'agit cette fois d'acquisition de matériel génétique *via* l'endosymbiose mitochondriale et l'endosymbiose chloroplastique. Ces événements ont certainement été des événements déterminants dans l'évolution de la cellule eucaryote. Le concept d'endosymbiose a été proposé il y a plus d'un siècle (Mereschkowsky, 1905), mais était généralement considéré comme «fantaisie ludique» ou «trop invraisemblable pour être mentionné dans la bonne société» (Wilson, 1925) jusqu'à la découverte de l'ADN mitochondrial et chloroplastique dans les années 1960. Les phylogénies basées sur cet ADN ont démontré depuis que les mitochondries sont bien apparues à partir de l'endosymbiose unique d'une α -protéobactérie (Gray et al., 1999; Gray and Doolittle, 1982), et que les chloroplastes sont issus de l'endosymbiose d'une cyanobactérie (Nelissen et al., 1995).

3.2 Hypothèses sur les liens unissant Eucarya et Archaea

Deux grandes théories s'affrontent quant à la nature du lien évolutif entre Eucarya et les deux autres domaines :

La première (appelée ici "hypothèse des trois domaines" (3D)) propose que les Eucarya et les Archaea sont deux lignées sœurs distinctes qui descendent d'un ancêtre commun exclusif (Figure 5). Ce point de vue est classiquement dépeint dans l'arbre universel du vivant qui représente trois domaines – les Bacteria, les Archaea et les Eucarya - et place la racine dans la branche conduisant aux Bacteria (comme suggéré par l'analyse des gènes anciennement dupliqués (Gogarten et al., 1989; Gribaldo and Cammarano, 1998; Iwabe et al., 1989)). Une des implications directes du scénario 3D est que les caractéristiques communes

entre Archaea et Eukarya représentent des traits ancestraux qui ont été hérités de leur ancêtre commun exclusif.

La seconde hypothèse (appelée ici « hypothèse des deux domaines » (2D)) propose que les Eucarya résultent de l'association d'une bactérie et d'une archée.

Selon ce scénario, l'arbre universel de la vie est composé de seulement deux domaines principaux - les Archaea et les Bacteria - les Eucarya représentant un domaine secondaire qui a émergé après les deux premiers (Figure 6).

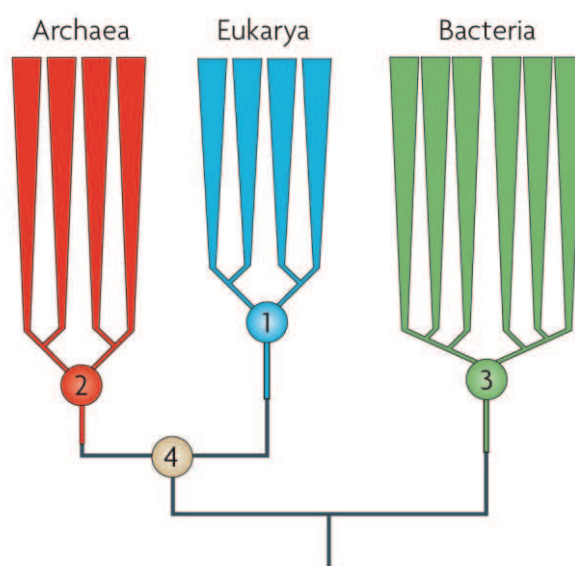


Figure 5. Scénario "3D" (hypothèse des trois domaines) (Gribaldo et al., 2010). Les eucaryotes, les archées et les bactéries forment trois domaines principaux, chacun ayant un ancêtre qui lui est propre (1, 2 et 3). Les Eucarya et les Archaea sont deux lignées-sœurs distinctes, ce qui implique qu'ils partagent un ancêtre exclusif (4), dont la nature reste indéterminée. La barre verticale bleue représente l'émergence des caractères spécifiques eucaryotes. Ce scénario implique que l'ancêtre 4 était vraisemblablement de type "procaryote" et qu'il possédait tous les caractères homologues partagés entre Archaea et Bacteria. Toutefois, cela ne signifie pas que cet ancêtre était de nature "archée". Par exemple, il ne possédait vraisemblablement pas les lipides typiques qui définissent les Archaea actuelles et qui sont probablement apparus le long de la branche menant à ce domaine (barre verticale rouge). En outre, les caractéristiques des eucaryotes pourraient avoir été présentes dans cet ancêtre et avoir ensuite été perdues dans la branche menant aux Archaea, comme certains éléments le suggèrent (Csuros and Miklos, 2009; de Crecy-Lagard et al., 2010; Lecompte et al., 2002).

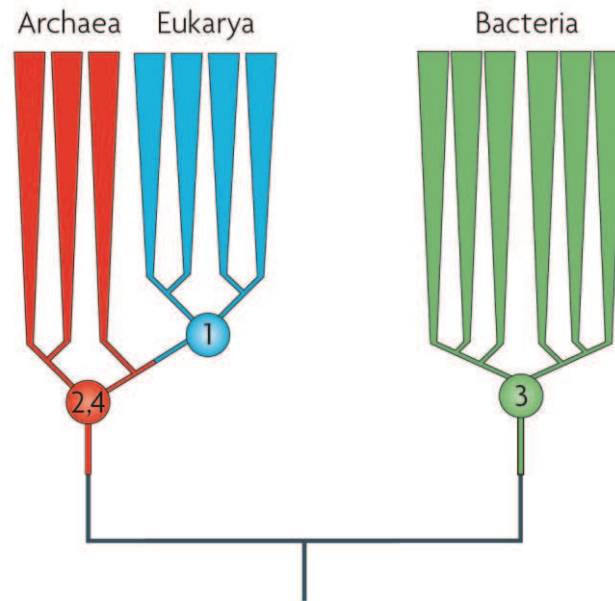


Figure 6. Scénario "2D" (hypothèse des deux domaines) (Gribaldo et al., 2010). Dans ce scénario, Archaea et Bacteria sont les deux domaines principaux, tandis que les Eucarya forment un domaine secondaire qui résulte de l'association d'une archée et d'une bactérie (pour simplifier, seule la relation entre Archaea et Eucarya est figurée ici). Dans ce scénario, le dernier ancêtre commun des eucaryotes (1) descend directement d'une archée ayant toutes les caractéristiques des Archaea modernes. Par conséquent, le dernier ancêtre commun des Archaea est aussi le dernier ancêtre commun des Archaea et des Eucarya (2,4). Tout comme dans le scénario 3D, il existe une longue branche à la base des Eucarya (barre bleue), le long de laquelle les caractères eucaryotes font leur apparition en dérivant de caractères archéens.

Dans les deux cas, il existe un "stem" à la base des Eucarya où se sont mis en place les caractères spécifiques de ce domaine. La différence majeure entre les deux théories se situe dans le fait que dans le cadre de l'hypothèse 2D, ces caractères se sont formés à partir de la transformation de caractères typiquement archéens, alors que dans le cas de l'hypothèse 3D, les caractères archéens d'un côté, et eucaryotes de l'autre, se sont mis en place indépendamment.

3.2.1 L'hypothèse 3D

Cette hypothèse considère que les topologies de type 3D obtenues lors des reconstructions phylogénétiques basées sur l'analyse des gènes universels ne sont pas artefactuelles, et reflètent une séparation réelle entre les domaines (Poole and Neumann, 2011). Elle implique également que les caractères présents chez les eucaryotes et chez les archées ont été hérités de leur ancêtre commun (Figure 5, point 4) (ou hérités de LUCA et perdus chez les bactéries), et que les caractères spécifiques de chacun des deux domaines sont apparus indépendamment dans chacun des branches menant à ces domaines, et n'impliquent donc pas la transformation de caractères spécifiques de l'un (*i.e.* caractères archéens) en caractères spécifiques de l'autre (*i.e.* caractères eucaryotes).

Des auteurs comme De Duve (De Duve, 1969, 2007) ont proposé que la phagocytose, qui exige un cytosquelette dynamique, un système endomembranaire et la perte de la paroi cellulaire rigide des procaryotes, ait évolué avant l'endosymbiose, permettant plus tard, l'internalisation des endosymbiontes procaryotes. Les modèles postulant une origine tardive des mitochondries, c'est-à-dire l'internalisation de la proto-mitochondrie par un proto-eucaryote, ont été fortement soutenus par l'observation d'eucaryotes anaérobies amitochondriaux collectivement appelées "Archezoa", émergeant à la base des arbres des eucaryotes basés sur l'ARNr SSU (Cavalier-Smith, 1987b; Golding and Gupta, 1995; Sogin, 1991). A la suite de ces publications, l'hypothèse selon laquelle la proto-mitochondrie était une α -protéobactérie qui a été internalisée par un eucaryote primitif anaérobie stricte est devenue largement acceptée (le « modèle Archezoa »).

Durant les années 1990, de nouvelles données et de nouvelles méthodes ont permis de montrer que les organismes appartenant aux Archezoa descendent en fait tous d'ancêtres ayant possédé des mitochondries et sont donc devenus secondairement amitochondriaux (de Duve, 2007; Embley, 2006). L'hypothèse Archezoa a ainsi été rejetée dans son ensemble et l'absence d'eucaryotes "primitifs" n'ayant jamais eu de mitochondrie fut alors considérée comme invalidant l'hypothèse selon laquelle les Eucarya formeraient un groupe frère de celui des Archaea (Embley, 2006). En effet, il n'existe aucun intermédiaire connu permettant de retracer l'apparition des caractères eucaryotes à partir de l'ancêtre qu'ils partageraient avec les Archaea.

3.2.2 L'hypothèse 2D

De nombreux scénarios ont été proposés dans le cadre des hypothèses 2D. Ils diffèrent selon la nature des partenaires archéens et bactériens qui se seraient associés pour donner naissance aux Eucarya.

Sur la base de la phylogénie de l'ARNr SSU, les Archaea ont été historiquement divisées en deux embranchements principaux : les Crenarchaeota (comprenant les lignées hyperthermophiles et thermophiles), ainsi que des lignées mésophiles présentes dans les milieux marins) et les Euryarchaeota (comprenant des lignées trouvées dans des environnements variés et présentant une grande diversité de métabolismes, tels que la méthanogénèse) (Schleper et al., 2005; Woese, 1987).

Depuis, d'autres phylums ont été proposés : les Nanoarchaeota (un seul représentant, *Nanoarchaeum equitans*, minuscule hyperthermophile anaérobie vivant en symbiose obligatoire à la surface d'une crenarchée) (Huber et al., 2002), les Korarchaeota (lignée d'organismes anaérobies non-

cultivés hyperthermophiles, vivant dans des environnements marins ou dans des sources chaudes) (Barns et al., 1996) et les Thaumarchaeota (d'abord considérées comme des crenarchées mésophiles, organismes aérobies vivant dans des environnements très variés, et ayant probablement un rôle important dans le cycle de l'azote)(Brochier-Armanet et al., 2008).

Les différentes hypothèses 2D peuvent être classées en trois catégories : (1) les hypothèses suggérant une véritable fusion de deux (ou plusieurs) procaryotes, (2) les hypothèses suggérant une association symbiotique de deux procaryotes, (3) les hypothèses suggérant une association symbiotique de trois espèces procaryotes.

3.2.2.1 Les hypothèses de fusion

Durant les années 1990, de nouvelles données montrent qu'en plus des gènes informationnels eucaryotes qui sont phylogénétiquement proches de ceux des Archaea, de nombreux gènes opérationnels eucaryotes, dont beaucoup ne sont pas impliqués dans la fonction mitochondriale, se sont avérés d'origine bactérienne (Kurland et al., 2003).

Sur cette base, **Zillig**, puis **Rivera et Lake** et **Gupta** ont formulé un modèle chimérique en vertu duquel une véritable fusion de cellules bactériennes gram-négatives et archées a été à l'origine des Eucarya (Gupta and Golding, 1996; Lake et al., 1984; Lake and Rivera, 1994; Rivera and Lake, 1992, 2004; Zillig et al., 1989). D'après l'analyse comparative des ribosomes des Bacteria, Eucarya, Euryarchaeota et Crenarchaeota (Lake et al., 1984) et l'analyse d'insertions dans le facteur d'élongation EF-1 alpha partagées entre Eucarya et Crenarchaeota (Rivera and Lake, 1992), **Rivera et Lake** ont suggéré que l'archée-hôte (qui sera aussi à l'origine du noyau) appartenait au groupe des Crenarchaeota. En effet, ces analyses semblaient indiquer que les Archaea sont paraphylétiques et que les Eucarya sont groupe-frère des Crenarchaeota. **Lake** a alors proposé d'élever ces dernières au rang de domaine et de les renommer "éocytes" (Lake et al., 1984).

Les critiques faites à ce modèle portent sur plusieurs points :

- Comment expliquer la fusion de deux cellules procaryotes d'un point de vue mécanique ?
- Qu'advient-il de l'organisme résultant de la fusion ? D'un point de vue biologique, ce modèle de fusion impliquerait une hybridation inter-domaine difficile à concevoir.

3.2.2.2 Les hypothèses de symbioses entre Alpha-proteobacteria et Archaea

En 1998, s'appuyant sur les associations entre archées et bactéries observées dans la nature, **Martin et Muller** proposent non pas une fusion mais une symbiose. Ainsi, la cellule eucaryote serait issue d'une symbiose d' α -protéobactéries productrices d'hydrogène et d'euryarchées méthanogènes consommatrices d'hydrogène ("hypothèse hydrogène", Figure 7 (Martin and Muller, 1998)). A l'origine, le rôle initial de la mitochondrie aurait donc été la production d'hydrogène et non pas la respiration aérobie.

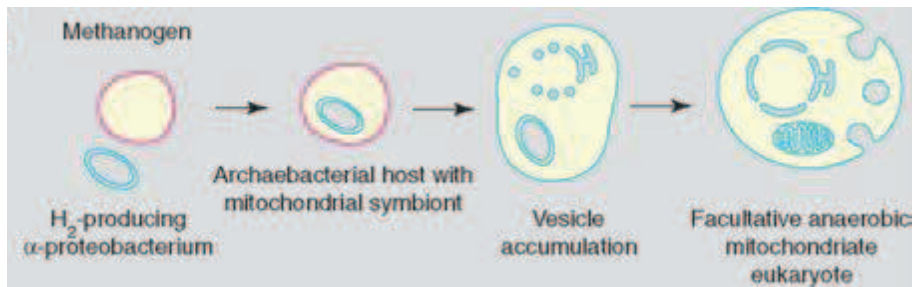


Figure 7. Hypothèse hydrogène (Martin and Muller, 1998). Une α -protéobactérie productrice d'hydrogène serait entrée en symbiose avec une euryarchée méthanogène consommatrice d'hydrogène.

En 2003, **Searcy** suggère que les eucaryotes sont le résultat de la symbiose d' α -protéobactéries oxydant le soufre et d'archées sulfato-réductrices de type *Thermoplasma* ("Sulfur syntrophy hypothesis" (Searcy, 2003)).

3.2.2.3 Les symbioses de trois espèces procaryotes

Il y a une quarantaine d'années, **Lynn Margulis** a réactualisé la théorie endosymbiotique, et a suggéré que les eucaryotes aient été formés à la suite d'associations endosymbiotiques successives de procaryotes (Margulis, 1970; Sagan, 1967). L'hôte aurait été une archée semblable aux membres du genre *Thermoplasma* (une lignée d'Euryarchaeota thermophiles sans paroi, qui respirent le soufre et se développent dans des environnements acides) (Figure 8). Cette archée serait entrée successivement en symbiose avec des bactéries de la famille des Spirochaeta qui seraient à l'origine des flagelles eucaryotes et de l'appareil mitotique. L'organisme ainsi formé aurait alors acquis la proto-mitochondrie (Figure 8). Au cours des années, Margulis a modifié plusieurs fois son modèle notamment quant à l'ordre des symbioses (Margulis, 1996; Margulis and Bermudes, 1985; Margulis et al., 2000). Il est surprenant de voir que ce modèle a encore été récemment actualisé (Margulis et al., 2006) bien que, jusqu'à présent, aucun gène codant pour la tubuline (*i.e.* un des composants du flagelle eucaryote) n'ait été découvert chez les spirochètes (Satir et al., 2007). De plus, il n'existe aucune indication que les gènes des *Thermoplasma* et/ou des *Spirochaetes* soient particulièrement abondants dans les génomes eucaryotes.

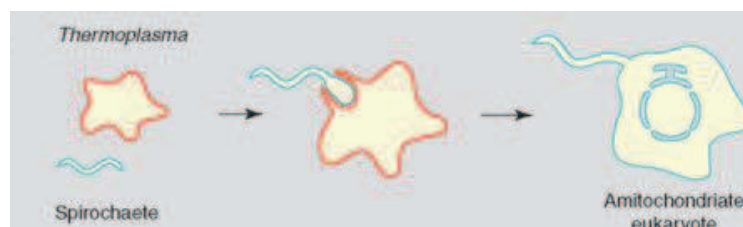


Figure 8. Hypothèse de symbiose de Lynn Margulis. Une thermoplasmatale serait rentrée en symbiose avec un spirochète qui serait à l'origine du flagelle eucaryote et de l'appareil mitotique.

Moreira et López-García ont proposé un scénario évolutif complexe et une association métabolique tripartite : une δ -protéobactérie productrice d'hydrogène aurait subi l'endosymbiose d'une archée méthanogène consommatrice d'hydrogène, puis cet organisme

aurait ensuite été l'hôte de l'endosymbiose mitochondriale. C'est l'hypothèse "syntrophique" (Figure 9) (Lopez-Garcia and Moreira, 2006; Moreira and Lopez-Garcia, 1998).

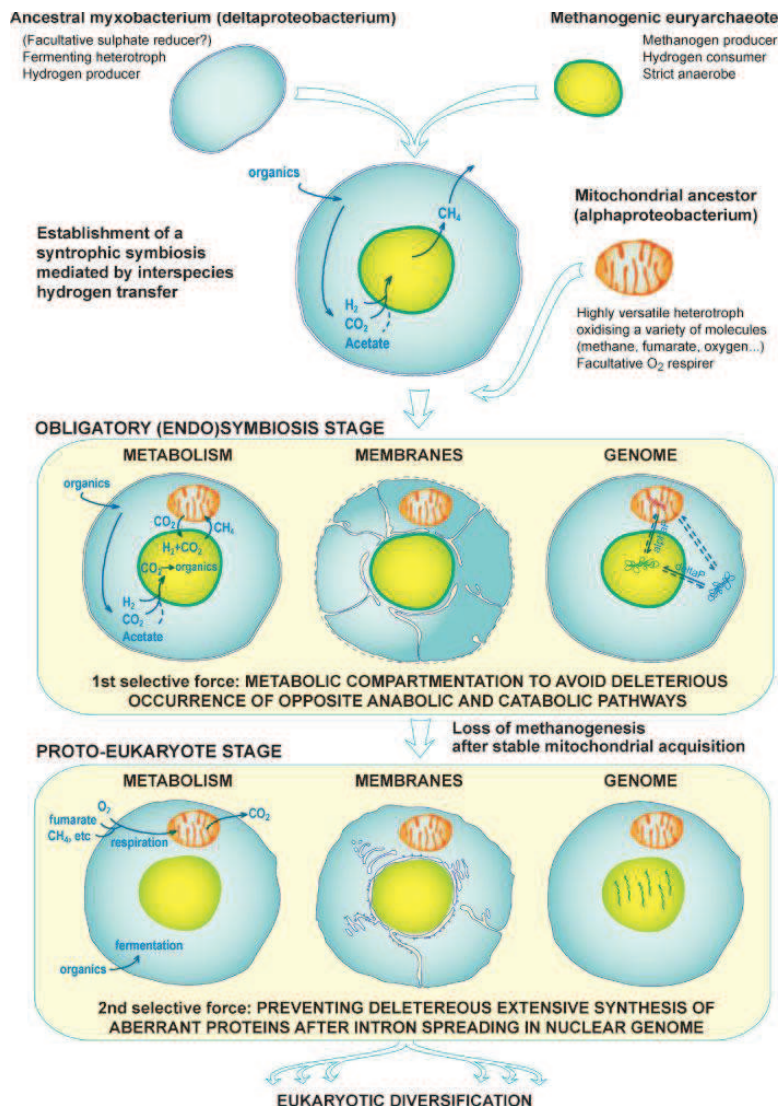


Figure 9. Hypothèse syntrophique (Lopez-Garcia and Moreira, 2006). Schéma de l'évolution des voies métaboliques, de la membrane et du génome conduisant à l'émergence des eucaryotes à partir d'un consortium symbiotique de myxobactéries (Deltaprotéobactéries) et d'archées méthanogènes. La fermentation de la myxobactérie libère de l'hydrogène qui est utilisé par l'archée méthanogène pour son métabolisme énergétique. Les mitochondries provenant d'ancêtres alpha-protéobactériens recyclent le méthane.

Le modèle de **Horiike et al.** (Horiike et al., 2001; Horiike et al., 2004) suggère, quant à lui, une symbiose d' α -protéobactéries, de γ -protéobactéries et d'archées, toutefois sans définir les propriétés métaboliques des différents partenaires.

Aussi diverses soient elles, l'ensemble de ces hypothèses 2D ont pour point commun de considérer que la topologie des arbres phylogénétiques basés sur des marqueurs universels qui soutiennent l'hypothèse 3D est le résultat d'artefacts de reconstruction.

D'autres hypothèses sur l'origine des eucaryotes ont aussi été avancées (impliquant par exemple les virus), et sont résumées notamment dans les revues de **Martin** (Martin et al., 2001) et **Searcy** (Searcy, 2003).

3.3 Arguments en faveur de l'une ou l'autre des hypothèses

Un des arguments le plus souvent avancé en faveur du scénario (2D) est qu'il explique l'origine des eucaryotes en se référant aux espèces archées et bactériennes modernes connues, sans avoir besoin de postuler l'existence de lignées proto-eucaryotes aujourd'hui disparues (en particulier, des lignées pré-mitochondriales) qui aurait précédé la diversification des eucaryotes actuels, alors que ce postulat est requis par le scénario 3D (Dagan and Martin, 2007).

Cependant, même dans le cadre des hypothèses 2D, il faut expliquer l'apparition de caractères eucaryotes à partir d'organismes porteurs de caractéristiques archées. Il existe alors deux scénarios :

- La mitochondrie est le premier caractère eucaryote-spécifique à être apparu. Dans ce cas se pose le problème du mécanisme d'endosymbiose. Comment imaginer cette endosymbiose sans le développement auparavant de la capacité de phagocytose, qui est *a priori* inconnue chez les Archaea ?
- La mitochondrie n'est pas le premier caractère eucaryote-spécifique à être apparu, et dans ce cas, la question de l'existence de lignées proto-eucaryotes amitochondriales aujourd'hui éteintes se pose toujours (Field and Dacks, 2009; Poole and Penny, 2007).

La question du moment de l'apparition de la mitochondrie mise à part, la complexité des caractères eucaryotes-spécifiques qui sont présumés avoir été présents dans le dernier ancêtre commun des eucaryotes (Koonin, 2010) représente un argument fort en faveur de l'existence de lignées proto-eucaryotes qui ont ensuite disparu. En effet, dans le cas contraire, toutes les caractéristiques eucaryotes-spécifique auraient dû apparaître dans la même lignée sans qu'aucune diversification n'ait lieu, ce qui semble assez improbable (Gribaldo et al., 2010; Poole and Neumann, 2011).

Par conséquent, tout comme le scénario 3D, les scénarios 2D impliquent l'existence d'une grande branche (d'un "stem"), le long de laquelle les spécificités eucaryotes ont dû évoluer (Poole and Neumann, 2011).

Philippe et collaborateurs ont donc tenté d'apporter une explication à l'absence de lignées de "proto-eucaryotes". Ils proposent une hypothèse selon laquelle les lignées eucaryotes primitives n'auraient pas laissé de descendants au cours de l'évolution, laissant la place aux lignées actuelles issues de la radiation observée à la base des Eucarya actuels ("hypothèse du Big-Bang" Figure 10) (Philippe et al., 2000a).

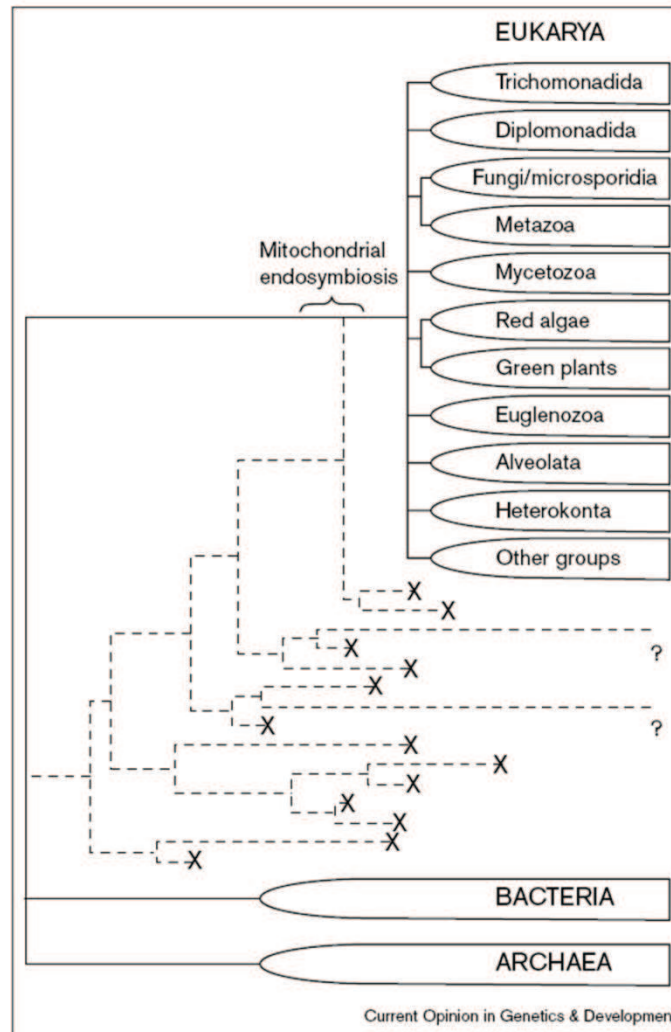


Figure 10. L'hypothèse du "big bang" sur l'évolution des eucaryotes (Philippe et al., 2000a). Les grands phylums eucaryotes actuels ont émergé d'une radiation ayant eu lieu suite à l'endosymbiose mitochondriale. Une très longue branche sépare les eucaryotes des procaryotes. Celle-ci ne signifie pas qu'aucune lignée eucaryote n'est apparue durant cette période, mais qu'ils ont très probablement disparu (pointillés).

Un second argument en faveur des scénarios 2D est que les génomes eucaryotes, en plus des gènes spécifiques à ce domaine, possèdent une fraction de gènes qui sont plus proches de leurs homologues bactériens ou partagés exclusivement avec les bactéries et une fraction des gènes qui sont plus proches de leurs homologues archéens ou uniquement partagés avec les Archaea, ce qui témoignerait d'une origine chimérique des eucaryotes.

Toutefois, les scénarios 3D ne sont pas incompatibles avec une contribution génétique considérable des bactéries : les gènes qui sont plus proches de leurs homologues bactériens ou exclusivement partagés avec les bactéries pourraient être d'origine mitochondriale, ou bien acquis par transfert horizontal à partir de donneurs bactériens, ou bien encore hérités du dernier ancêtre commun universel (LUCA), puis perdus ou remplacés chez les Archaea. De la même façon, les gènes eucaryotes qui sont plus proches de leurs homologues archéens ou partagés uniquement avec les Archaea pourraient dériver d'un ancêtre commun exclusif Archaea-Eucarya, ou bien être acquis par THG à partir de donneurs archéens, ou bien encore hérités de LUCA puis perdus ou remplacés chez les bactéries. Enfin, les gènes eucaryotes-spécifiques sont soit apparus dans la lignée eucaryote ou bien ont été perdus ou remplacés chez les bactéries et les archées.

Un argument avancé en faveur de l'hypothèse 2D est qu'il existe un exemple d'endosymbiose d'une gamma-protéobactérie dans une bêta-protéobactérie sans phagocytose. Néanmoins, aucune association symbiotique n'a été observée à ce jour entre une archée et une bactérie - condition indispensable pour la réalisation du scénario 2D (von Dohlen et al., 2001).

Enfin, un argument en défaveur des scénarios 2D est qu'ils impliquent la perte ou le remplacement le long du stem eucaryote des caractères typiques des Archaea comme par exemple les lipides uniques qui constituent leur membrane, ainsi que des adaptations spécifiques telles que la méthanogénèse ou l'hyperthermophilie, en fonction de la lignée archée impliquée dans l'association symbiotique. Cependant, des hypothèses tenant compte de ces pertes ont été émises (Lopez-Garcia and Moreira, 2006; Martin, 2005; Martin and Muller, 1998; Moreira and Lopez-Garcia, 1998).

De plus, si l'on favorise les hypothèses 2D n'impliquant pas l'existence de proto-eucaryotes, alors cela implique que les gènes eucaryotes censés être d'origine archéenne ont perdu le signal phylogénétique remontant à leur origine. Cela paraît difficile à expliquer alors que cela n'a pas eu lieu pour les gènes eucaryotes d'origine mitochondriale.

En pratique, ces deux hypothèses peuvent être testées par la phylogénie moléculaire, mais dans la réalité les relations évolutives anciennes sont très difficiles à retrouver. Le signal phylogénétique qui pourrait fournir des informations sur les événements de spéciation anciens donnant naissance aux bactéries, archées et eucaryotes est très faible, car il a été largement effacé par les substitutions successives qui ont lieu au cours du temps. De plus, la reconstruction des relations anciennes est particulièrement sensible aux artefacts de reconstruction d'arbres qui peuvent influencer sur la topologie qui en résulte (cf. p.28). En conséquence, il a été suggéré par certains auteurs que la topologie 3D de l'arbre du vivant obtenu à partir de l'analyse phylogénétique de gènes universels est en fait le résultat d'un artefact de reconstruction qui fait que les Archaea semblent former un groupe monophylétique (Figure 5), alors qu'en réalité les Eucarya devraient se brancher à l'intérieur de ce domaine (Figure 6) (Cox et al., 2008; Foster et al., 2009; Lake, 1988; Rivera and Lake, 1992, 2004; Tourasse and Gouy, 1999).

De nombreuses études récentes basées sur l'analyse phylogénétique simultanée d'un grand nombre de gènes ont abordé la question de la relation entre Archaea et Eucarya. Étonnamment, malgré l'utilisation de données très similaires, ces études donnent des

résultats contradictoires : trois études récentes appuient l'idée que les Archaea et les Eucarya sont deux lignées monophylétiques indépendantes (scénario 3D), alors que quatre études favorisent le scénario 2D, mais sont en désaccord entre elles sur la lignée d'archées qui a donné naissance aux eucaryotes (Gribaldo et al., 2010).

4 La phylogénie des Eucarya

4.1 Les premières phylogénies moléculaires

Au début des années 90, les premières phylogénies moléculaires des eucaryotes (basées sur l'ARNr SSU) divisaient ceux-ci en cinq groupes principaux (correspondant aux règnes des animaux, des champignons, des plantes, des alvéolés et des straménopiles), auxquels s'ajoutaient une succession de lignées de protistes amitochondriaux émergeant à la base de l'arbre (Figure 11). Cet arbre a conduit à l'idée qu'il existe des eucaryotes "supérieurs" qui forment la "couronne eucaryote" en haut de l'arbre, tandis que les eucaryotes "inférieurs" émergeant à la base de l'arbre correspondent à des lignées "primitives" (diplomonadés, parabasalidés et microsporidies) dépourvus de certains caractères tels que les mitochondries, l'appareil de Golgi ou les peroxisomes, et confirmant ainsi l'hypothèse Archezoa discutée précédemment (Figure 11) (Sogin, 1991).

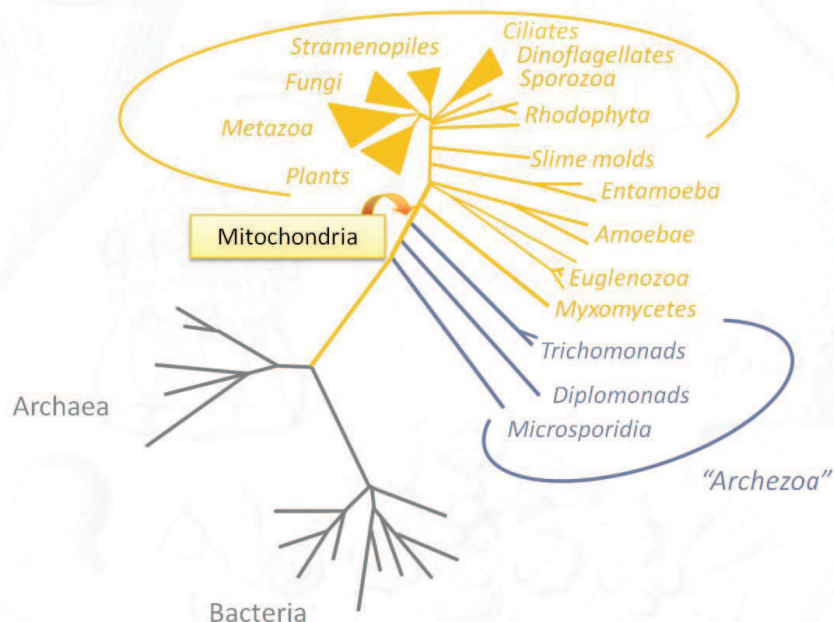


Figure 11. Phylogénie des trois domaines inférée à partir de l'ARNr SSU. Adapté de Sogin 1991. Les protistes amitochondriaux divergent précocement, tandis que les multicellulaires sont positionnés en "haut" de l'arbre.

4.2 L'utilisation de nouveaux marqueurs

Au cours des années 1990, l'utilisation de quelques marqueurs phylogénétiques alternatifs (principalement EF1 α , HSP70 et HSP90, α -tubuline, β -tubuline, actine), seuls ou concaténés, a conduit à une autre image de l'évolution des eucaryotes (Philippe et al., 2000b; Stechmann and Cavalier-Smith, 2003a), et la majorité des relations de parenté entre grands groupes eucaryotes a été remise en question (Budin and Philippe, 1998; Philippe, 2000; Philippe and Germot, 2000; Philippe et al., 2000a; Philippe et al., 2000b; Roger et al., 1999). Au cours des années suivantes, les données de biologie moléculaire montrant que tous les eucaryotes amitochondriaux possèdent des traces d'un ancêtre mitochondrial ont conduit à ce que leur position soit largement débattue. Pour autant, les analyses phylogénétiques basées sur les marqueurs individuels n'ont pas permis de replacer ces espèces de façon fiable. La difficulté à

retrouver les relations de parenté entre les groupes peut avoir plusieurs causes détaillées ci-dessous.

4.3 Les causes d'irrésolution et les artefacts de reconstruction

4.3.1 Artefacts méthodologiques

Une source importante de biais en phylogénie moléculaire est liée à l'hétérogénéité des vitesses d'évolution entre les sites. En particulier, les modèles d'évolution sont le plus souvent homogènes (tous les sites d'une protéine évoluent suivant les mêmes modalités) et stationnaires (ces modalités, de plus, sont constantes au cours du temps), alors que l'hétérogénéité du processus d'évolution entre les différentes positions de l'alignement et entre les lignées semble être un aspect essentiel de l'évolution des séquences protéiques et nucléiques. Cette hétérogénéité peut être, entre autres, à l'origine d'artefacts d'attraction des longues branches (ALB).

En effet, si certaines séquences évoluent significativement plus vite que d'autres, alors elles sont plus susceptibles de partager des états de caractères identiques dérivés par hasard (fausses synapomorphies) que les séquences qui évoluent lentement. En parallèle, ces dernières ont tendance à conserver les caractères ancestraux, et donc à partager entre elles des symplesiomorphies. En l'absence de modèles d'évolution adéquats, ce phénomène n'est pas pris en compte et les séquences évoluant rapidement seront regroupées dans une même région de l'arbre indépendamment de leurs véritables relations de parenté. Un facteur aggravant l'ALB est lié à l'utilisation d'un groupe extérieur distant. De plus, le placement de ces séquences peut être statistiquement fortement soutenu. C'est par exemple le cas des microsporidies placées à la base de l'arbre avec plusieurs autres taxons évoluant vite et tous étant attirés par la grande branche du groupe extérieur (Figure 11).

Une observation intéressante a été faite lorsque des marqueurs alternatifs à l'ARNr SSU ont commencé à être utilisés, à savoir que les vitesses d'évolution relatives des différentes lignées varient selon les marqueurs considérés (Figure 12) (Philippe et al., 2000b). Ainsi, selon les marqueurs, ce ne sont pas les mêmes taxa qui sont affectés par l'ALB, et certains marqueurs sont plus performants que d'autres selon le groupe d'organismes que l'on cherche à positionner.

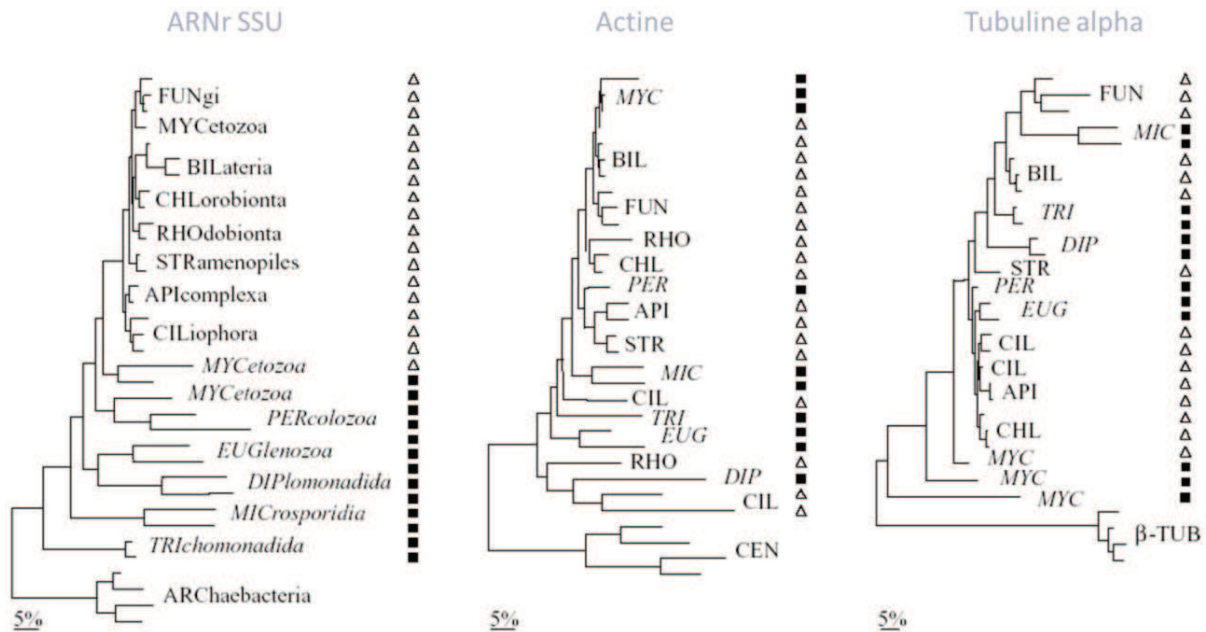


Figure 12. Illustration de l'incongruence entre des phylogénies basées sur des marqueurs différents liée à un artefact d'attraction des longues branches. Adapté de (Philippe et al. 2000). Afin de faire ressortir les incongruences entre les marqueurs, les espèces qui émergent à la base dans l'arbre ARNr SSU sont en italiques et marquées d'un carré noir. Selon les marqueurs, les espèces évoluent à des vitesses différentes.

4.3.2 Radiation évolutive

Une cause d'irrésolution est liée à l'intervalle de temps dans lequel ont eu lieu les événements de spéciations ayant donné naissance aux différents phylums (Figure 13).

En effet, si cet intervalle de temps a été court, alors il est possible que le nombre de substitutions qui s'est produit entre chaque événement de spéciation soit trop faible pour pouvoir inférer leurs relations de parenté.

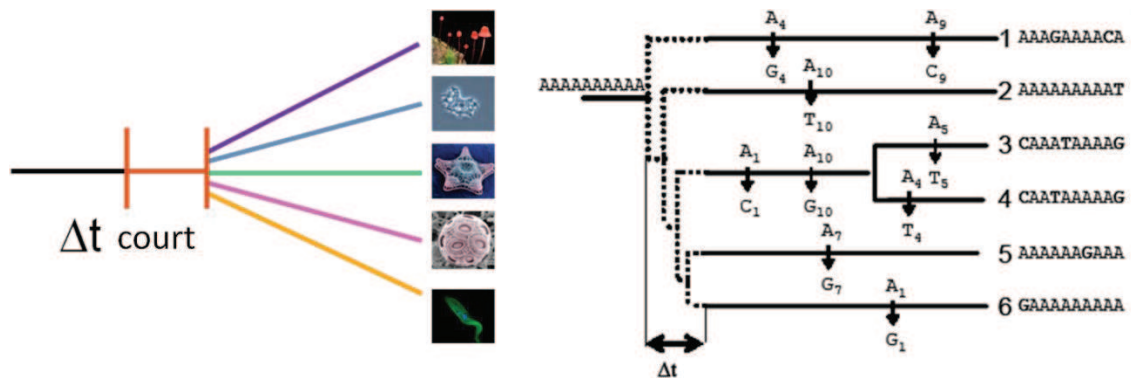
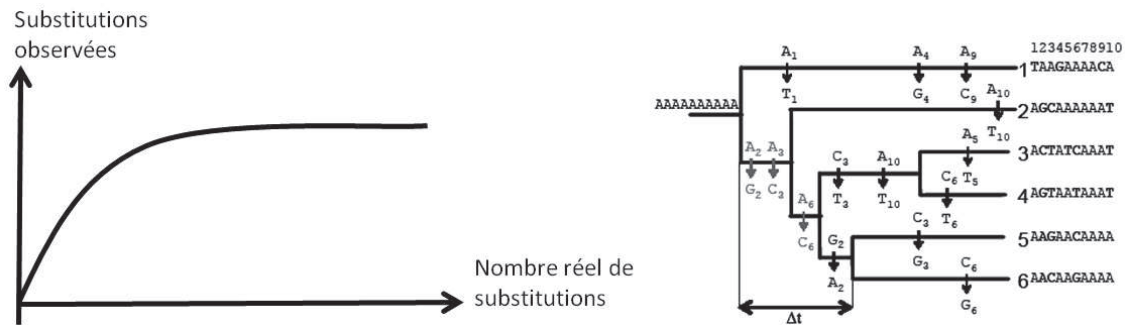


Figure 13. Phénomène de radiation évolutive. Adapté depuis (Gribaldo and Brochier, 2009).

On parle alors de radiation (Figure 13). C'est l'hypothèse popularisée par Philippe et collaborateurs pour expliquer l'émergence des grands groupes eucaryotes (Figure 10) (Philippe et al., 2000a).

4.3.3 Saturation moléculaire

Une seconde explication est le phénomène de saturation moléculaire (Figure 14). Celui-ci est lié aux substitutions multiples qui se produisent lors de l'évolution des séquences. En effet, les marqueurs moléculaires ayant une taille finie, au cours de l'évolution une même position peut être affectée par plusieurs événements de substitution successifs. La principale conséquence de ce phénomène est que le signal phylogénétique le plus ancien est progressivement remplacé par un signal phylogénétique plus récent.



4.3.4 Les transferts latéraux entre eucaryotes

Bien que considérés comme moins fréquents que chez les procaryotes, des transferts horizontaux de gènes peuvent avoir lieu entre organismes eucaryotes (Keeling and Palmer, 2008).

Entre autres, une partie de ces transferts de gènes a eu lieu à la suite des endosymbioses chloroplastiques secondaire et tertiaire (cf. p.41). En effet, au cours ces évènements, les organismes eucaryotes internalisés par d'autres eucaryotes ont cédé beaucoup de leurs gènes (chloroplastiques et nucléaires) au génome nucléaire de leur hôte. Ce phénomène est appelé "transfert de gènes endosymbiotiques" (Andersson, 2005; Archibald, 2005; Esser et al., 2004; Martin et al., 2002). Il est largement répandu et complique beaucoup la reconstruction de l'arbre des eucaryotes.

4.4 Comment pallier au manque de résolution ?

Plusieurs solutions ont été proposées pour surmonter les difficultés évoquées précédemment et progresser dans notre compréhension de l'évolution ancienne.

4.4.1 Amélioration des modèles et méthodes

Le développement d'approches permettant de prendre en compte les différences de vitesse d'évolution entre les sites a permis d'améliorer de manière significative les reconstructions phylogénétiques. Par exemple, **Adachi et Hasegawa** ont développé une méthode pour prendre en compte l'existence de sites invariants (Adachi and Hasegawa, 1995), et **Yang** a proposé un moyen de modéliser l'hétérogénéité des vitesses d'évolution par l'utilisation d'une loi gamma (Yang, 1993)

Ces corrections apportées aux modèles ont des performances qui s'avèrent largement supérieures à celles des modèles "à taux uniformes", et sont maintenant utilisés de façon routinière. A cette disparité des vitesses d'évolution des sites se superpose une hétérogénéité des patterns de substitution, c'est-à-dire que les matrices de substitutions couramment utilisées en phylogénie se révèlent inadaptées pour certains sites de l'alignement. En effet, en général, chaque position subit une contrainte spécifique sélective, et, par conséquent, seul un sous-ensemble des 20 acides aminés est susceptible d'être fixé à cette position durant l'évolution. Dans ce contexte, **Lartillot et Philippe** ont proposé en 2004 un modèle probabiliste (le modèle CAT) permettant de prendre en compte l'hétérogénéité substitutionnelle des sites de l'alignement en les affectant à des catégories possédant leurs propres modalités d'évolution (Figure 15). Le nombre de catégories est estimé à partir des données (Lartillot and Philippe, 2004) et fait partie des paramètres du modèle.

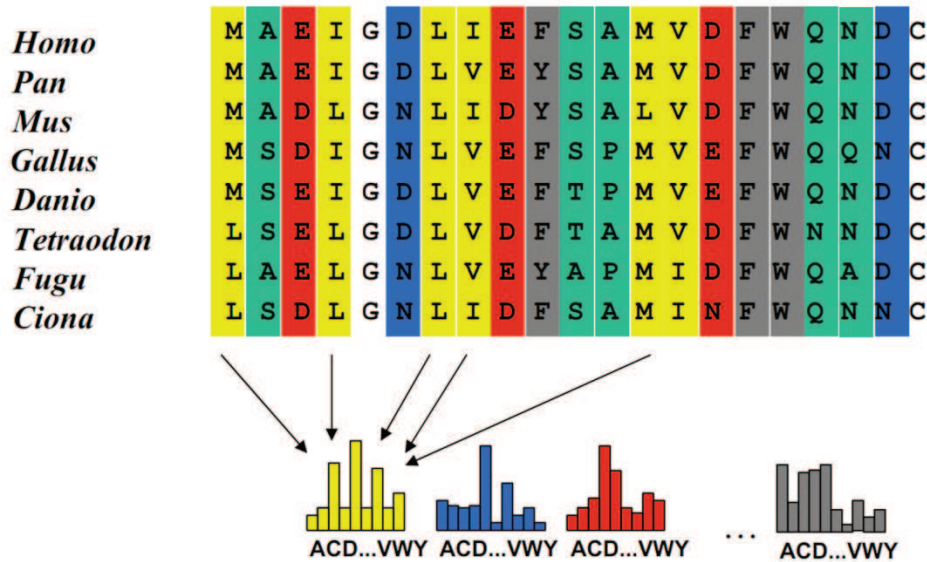


Figure 15. Principe du modèle CAT. La matrice de substitution est différente d'un site à l'autre. Le modèle est dit hétérogène et chaque site de l'alignement appartient à une catégorie (ou classe) regroupant d'autres sites partageant le même profil de fréquence stationnaire et la même matrice de substitution.

4.4.2 Améliorer l'échantillonnage taxonomique

Une autre solution est d'augmenter l'échantillonnage taxonomique au sein des différents groupes afin d'accéder à leur diversité génétique. De plus, le séquençage de multiples

espèces permettrait d'effectuer un choix rationnel des espèces incluses dans les analyses, en particulier en choisissant des espèces dont le taux d'évolution est plutôt lent, afin de mieux détecter les événements de substitutions multiples et de limiter l'artefact d'attraction des longues branches (Hedtke et al., 2006; Philippe and Douzery, 1994; Zwickl and Hillis, 2002).

4.4.3 La détection d'évènements génétiques rares

La recherche d'évènements génétiques rares (ex : les insertions/délétions partagées, les fusions de gènes, ou les transferts horizontaux de gènes) s'étant produits à la base d'un groupe taxonomique constituent des signatures de ce groupe et une piste intéressante (Stechmann and Cavalier-Smith, 2002) (exemple, Figure 16). Cependant, cette approche doit être utilisée avec prudence, car il a été montré que des convergences inattendues peuvent se produire (Baptiste and Philippe, 2002).



Figure 16. Insertion dans le site actif de la polyubiquitine chez les Rhizaria. Cette insertion est uniquement retrouvée chez deux groupes de Rhizaria : les Cercozoa et les Foraminifera. et pourrait constituer une signature de ce groupe (Archibald et al., 2003).

4.4.4 De la phylogénie à la phylogénomique

Les analyses phylogénétiques basées sur un seul gène sont souvent affectées par l'erreur stochastique due au nombre limité de positions informatives (vis-à-vis de la problématique d'intérêt) contenu dans les alignements.

Sur la base de ce constat, une des méthodes les plus intuitives pour tenter d'améliorer la résolution consiste à unir l'information phylogénétique provenant de plusieurs gènes (parfois plusieurs centaines). L'objectif est de combiner le signal faible porté par chacun d'eux qui reflète la véritable histoire des organismes pour en faire émerger un signal fort tout en atténuant les biais pouvant être présents dans chacun des gènes (sous la condition que les biais ne soient pas systématiquement dans le même sens).

Cette analyse simultanée de plusieurs marqueurs phylogénétiques peut se faire principalement par deux approches (Delsuc et al., 2005) (Figure 17) : (i) la méthode des

superarbres, consistant en la reconstruction d'une phylogénie pour chaque gène puis à la combinaison des arbres qui résultent en une super-phylogénie, (ii) et celle des supermatrices, qui consiste à concaténer les différents alignements en une matrice unique de caractères, sur laquelle on applique une méthode standard de reconstruction d'arbre.

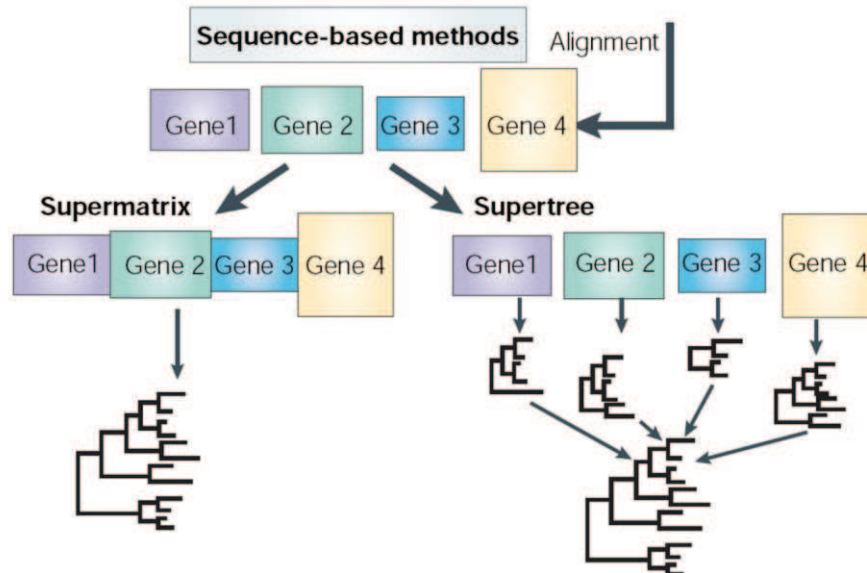


Figure 17. Deux méthodes alternatives d'analyse phylogénomique des séquences. (Delsuc et al., 2005)

L'approche par concaténation de gènes (*i.e.* supermatrice) a commencé à être utilisée à la fin des années 1990. En effet, cette approche est dépendante des données génomiques disponibles dans les banques et on parle d'ailleurs souvent de phylogénomique. C'est au début des années 2000 qu'ont été réalisées les premières concaténations massives de gènes pour l'étude de la phylogénie des eucaryotes (Baptiste et al., 2002) et depuis lors, cette méthode s'est progressivement imposée pour résoudre les relations de parenté anciennes entre organismes. Cette approche pose néanmoins des problèmes d'ordre méthodologique, spécialement au niveau des modèles d'évolution utilisés. En effet, lors de ce type d'analyse, on combine des objets qui n'évoluent pas nécessairement selon les mêmes modalités.

4.5 La phylogénie des Eucarya actuellement admise et les grandes questions

Durant la dernière décennie, les analyses moléculaires ont fourni des hypothèses sur les relations entre environ 75 petits groupes d'eucaryotes "robustes", c'est-à-dire dont la monophylie ne fait *a priori* pas de doute (Patterson, 1999). Ces groupes sont définis par des caractéristiques ultrastructurales (*i.e.* modes d'organisation cellulaire et subcellulaire révélés par microscopie électronique, tels que la forme des crêtes mitochondriales, la longueur et l'orientation des corps basaux et leur association avec d'autres organites, le nombre de membranes délimitant le chloroplaste s'il y en a un, etc.) (Patterson, 1999). En parallèle, ces groupes sont fortement soutenus par des analyses moléculaires (Parfrey et al., 2006; Yoon et al., 2008).

C'est le cas par exemple des Nucleariidae, qui ont pour particularité ultrastructurale d'avoir des mitochondries à crêtes discoïdes, des dictyosomes, des pseudopodes sans microtubules soutenus par des éléments microfibrillaires, mais aucun extrusome (Patterson, 1999). Pendant la mitose, l'enveloppe nucléaire reste intacte et les microtubules pénètrent dans le noyau.

Grâce aux analyses phylogénomiques, la plupart de ces 75 lignées est désormais regroupée en un petit nombre de clades de rang supérieur (Adl et al., 2005; Keeling et al., 2005; Simpson and Roger, 2004) (Figure 18).

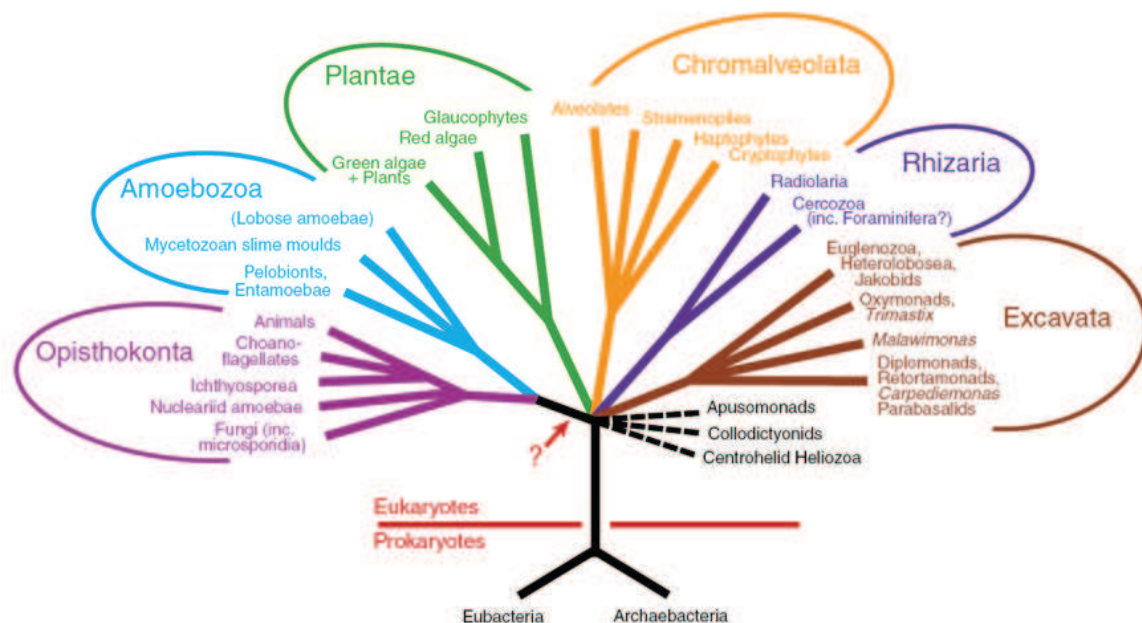


Figure 18. Arbre schématique illustrant l'organisation de la plupart des eucaryotes en six grands groupes (Simpson and Roger, 2004). Les relations entre la plupart des grands groupes et la position de la «racine» de l'arbre sont indiquées comme incertaines (à l'exception du regroupement des Opisthokonta et Amoebozoa). La flèche et le point d'interrogation indiquent un placement précis possible de la racine, sur la base de données de fusion de gènes (cf. p.47).

4.5.1 Opisthokonta

Le supergroupe des Opisthokonta contient les **animaux** et les **champignons**, ainsi que plusieurs groupes de protistes, comprenant les **choanoflagellés** (unicellulaires libres), les **Ichthyosporea** (une large gamme de parasites), et un groupe d'amibes libres, les **Nucleariids**. Les choanoflagellés, les gamètes mâles d'animaux et la plupart des zoospores de chytrides (les seuls champignons flagellés) se propulsent tous avec un unique flagelle à leur extrémité postérieure. Les animaux sont pratiquement les seuls prédateurs multicellulaires sur Terre. Les champignons sont à dominante osmotrophe, et jouent des rôles cruciaux en tant que décomposeurs ou en tant que symbiotes ou parasites des plantes. Parmi les champignons, les microsporidies infectent un large éventail d'animaux, y compris les insectes, et représentent probablement un des groupes de parasites les plus divers (Simpson and Roger, 2004).



Figure 19. Quelques exemples d'Opisthokonta. De gauche à droite : le choanoflagellé *Salpingoeca* (© William Bourland), les métazoaires *Saimiri sciureus* (© Luc Viatour) et *Cyanea capillata* (© Patrick Keeling), le champignon *Morchella* (© Patrick Keeling), et la microsporidie *Fibrillanosema crangonycis* (© Leon White). Source : <http://tolweb.org/>.

La disposition du flagelle est quasiment unique et il a été proposé qu'elle ait été héritée de l'ancêtre commun des Opisthokonta (Cavalier-Smith, 1987a). Au début des années 1990, plusieurs phylogénies moléculaires ont montré le lien spécifique entre animaux, choanoflagellés et champignons (Baldauf and Palmer, 1993; Wainright et al., 1993). Les autres groupes ont été progressivement intégrés dans les Opisthokonta depuis, en grande partie grâce aux analyses de l'ARNr (Ragan et al., 1996; Ruiz-Trillo et al., 2006; Steenkamp et al., 2006). La monophylie de ce groupe est également soutenue par une insertion dans le facteur d'élongation EF 1-alpha (Figure 20) (Baldauf and Palmer, 1993; Steenkamp et al., 2006). Les études moléculaires récentes indiquent que les choanoflagellés et Ichthyosporea seraient plus étroitement liés aux animaux qu'aux champignons, mais de façon générale, les relations précises au sein des Opisthokonta restent encore à préciser.

		153	168	198	236		
OPISTHOKONTA	ANIMALIA	Deuterostomia	<i>Homo</i>	NKMD STEPPYS QKRYE ...	GDNMLEPSANMPWFKG WKVTRK----	DGNASG TTLEALDCILPP	
		Ecdysozoa	<i>Ciona</i>	NKMD NTEPPYS EQRFE ...	GDNMLETSNNMPWFKG WAIERK----	EGNASG KTLYNALDAIIPP	
		Lophotrochozoa	<i>Caenorhabditis</i>	NKMD STEPPYS EARFT ...	GDNMLEVSSNNMPWFKG WAVERK----	EGNASG KTLLEALDSIIPP	
			<i>Drosophila</i>	NKMD SSEPPYS EARYE ...	GDNMLEPSTNNMPWFKG WEVGRK----	EGNADG KTLVLDALDAIIPP	
			<i>Dugesia</i>	NKMD STEPPYS EPRFD ...	GDNMIDESNNMPWFKG WEITRKNKAKEIKTTG	KTLLEALDSLEPP	
			<i>Chaetopleura</i>	NKMD STTPPFS QPRFE ...	GDNMLEVSSNTAWFKG WNIERK----	EGNASG KTLFEALDSILPP	
		Radiata	<i>Eugymnanthea</i>	NKMD NTEPPYS EARFK ...	GDNMLEPSTNNMPWFKG WEIERK----	AGKASG KTLLEALDAVVP	
		Parazoa	<i>Geodia</i>	NKMD STEPPYS QARYD ...	GDNMLEESPNNMPWFKG WNVVERK----	EGNASG KTLFNFPLDSILPP	
		FUNGI	Basidiomycota	<i>Puccinia</i>	NKMD TT--KWS EQRFE ...	GDNMLEESTNNMPWFKG WTKETK----	AGVSKG KTLLEALDAIEIPP
			Ascomycota	<i>Neurospora</i>	NKMD TT--QWS QTRFE ...	GDNMLEPSTNNMPWFKG WEKATK----	AGKATG KTLLEALDAIEIPP
	Zygomycetes		<i>Mucor</i>	NKMD TT--KWS QDRYN ...	GDNMLEESTNNMPWFKG WTKETK----	AGSKTG KTLLEALDSIIPP	
	Trichomycetes		<i>Smittium</i>	NKMD SN--KYS EERFT ...	GDNMIEASTNNMPWFKG WTKETK----	SGVSKG VTLLEALDAIEIPP	
	Chytridiomycota		<i>Chytridium</i>	NKMD TT--KWS EDRYN ...	GDNMLEEASTNNMPWFKG WTKETK----	AGSSTG KTLLEALDAIEIPP	
	Microsporidia		<i>Glugea</i>	NKMD TIDEKNR ISRFD ...	GINIVEKDGKFEWFKG WKPVSG----	AG-DSI FTLEALNSQIPP	
	ALLIES		Choanoflagellata	<i>Monosiga</i>	NKMD STEPPYS ESRFN ...	GDNMIEASEKLPWYKG WEITRK----	EGNATG KTLLEALDAIIPP
			Amoebidium	<i>Amoebidium</i>	NKMD SI--KYS QDRFN ...	GDNMVEPTNNMPWYKG WEVERK----	EGNATG KTLLEALDAIIPP
		Ichthyosporea	<i>Ichthyophonus</i>	NKMD SV--KYS EDRFK ...	GDNMVAPTNNMPWYKG WTCERK----	EGNTSG FTLEALDNQIAPP	
		Corallochytrium	<i>Corallochytrium</i>	NKMD SI--KYS KDRFD ...	GDNMIEASTNNMPWYKG WEKD----	GSVGG KTLLEALDAVSP	
	Nucleariida	<i>Nuclearia</i>	NKMD TC--KYS EERFN ...	GDNMLEPTNNMPWYKG WEIDRK----	NGKVMG KTLVGLALDAIEIPP		
		<i>Ministeria</i>	NKMD SI--KYD EARFT ...	GDNMLDASTNNMPWYKG WEVDRK----	NGKASG KTLLEALDAVLEIPP		
	APUSOZOA	Ancyromonadidae	<i>Ancyromonas</i>	NKMD DKSNNYS KARFD ...	GDNMTEPSANMPWYSG	PTLLEALDACEVP	
	AMOEBOZOA	Apusomonadidae	<i>Apusomonas</i>	NKMD DKTVKYS KDRYE ...	GDNMMEPSPQMGWFKG	GTLLEALDAITPP	
		Mycetozoa	<i>Dictyostelium</i>	NKMD EKSTNYS QARYD ...	GDNMLERSDKMEWYKG	PTLLEALDAIVEP	
			<i>Lobosa</i>	<i>Acanthamoeba</i>	NKMD NV--NWA ENRYN ...	GDNMVDRTDKMPWYKG	PTLLEALDGIKPP
<i>Hartmannella</i>		NKMD SEVKYS QERYD ...	GDNMLEKSTNLPWYKG	PTLVEALDALKPP			
Conosa		<i>Entamoeba</i>	NKMD AI--QYK QERYE ...	GDNMIEPSTNNMPWYKG	PTLLEALDSVTFP		
Phalanstereia		<i>Phalansterium</i>	NKMD DKTVNWG EPRYQ ...	GDNMLERSANLPWYKG	PTLLEALDNLVP		
FLANFAE		<i>Arabidopsis</i>	NKMD ATTPKYS KARYD ...	GDNMIEERSTNLDWYKG	PTLLEALDQINEP		
		<i>Cyanophoraceae</i>	<i>Cyanophora</i>	NKMD EKSVNYS QPRFE ...	GDNMLEPSSNLPWYKG	PTLVEALDQVEEP	
HETEROKONTA		Bangiophyceae	<i>Porphyra</i>	NKMD DKVNNYS KERYE ...	GDNMLEKSTNNMPWYTG	PTLFEVLDAMKPP	
		Blastocystis	<i>Blastocystis</i>	NKMD DKSNNYS EARYK ...	GDNMIEHSAANMPWYKG	PTLLEALDNVHPP	
ALVEOLATA	Oomycetes	<i>Phytophthora</i>	NKMD DSSVMYK QARYE ...	GDNMIDRSTNNMPWYKG	PTLLEALDNLMAP		
	Apicomplexa	<i>Plasmodium</i>	NKMD TV--KYS EDRYE ...	GDNMLEKSDKTPWYKG	PTLLEALDTMQPP		
EUGLENOZOA	Cilliophora	<i>Paramecium</i>	NKMD EKTVNYA QGRYD ...	GDNMLEKSAFNGWYKG	PTLLEALDAVTFP		
	Euglenida	<i>Euglena</i>	NKFD DKTVKYS QARYE ...	GDNMIEEASNNMPWYKG	LTLEALDNLEPP		
HETEROLOBOSEA	Kinetoplastida	<i>Leishmania</i>	NKMD DKTVKYS QARYE ...	GDNMIEEASNNMPWYKG	PTLLEALDNLEPP		
	Schizopyrenida	<i>Naegleria</i>	NKFD DTSVYK EDRYN ...	GDNMIEKSTNNMPWYKG	PCLLEALDNLEPP		
PARABASALIDEA	Acrasida	<i>Acrasida</i>	NKMD DKSVMYK EDRYK ...	GDNMLEKSTNNMPWYKG	PTLLEALDALEPP		
	Trichomonadida	<i>Trichomonas</i>	NKMD DKTVNYN KARFD ...	GDNMTEKSPNNMPWYKG	PTLLEALDSLQPP		
DIPLOMONADIDA	Hypermastigida	<i>Trichonympha</i>	NKMD DNTVNYA ESRYK ...	GDNMTEKSDKMPWYKG	LTLEALDTLEPP		
	Giardiinae	<i>Giardia</i>	NKMD DGVVYK KERYD ...	GDNIMEKSDKMPWYKG	PCLLEALDGLKPP		
OXYMONADIDA	Pyrsonymphiidae	<i>Dinenympha</i>	NKMD DKSVNWA ESRYN ...	GDNMLDRSTNNMPWYKG	PTLLEALDLLEVP		
ARCHAEA	Crenarchaeota	<i>Sulfolobus</i>	NKMD LATPYD EKRFK ...	GDNVTHKSTNNMPWYKG	PTLLEALDLLEVP		
	Euryarchaeota	<i>Thermoplasma</i>	NKMD ATSPYS EKRYN ...	GDNVTKPSPNNMPWYKG	PSSLQQLDAFKVP		

Figure 20. Insertion de ~12 acides aminés dans la protéine EF-1 α retrouvée dans toutes les lignées d'Opisthokonta et absente dans toutes les autres lignées (Steenkamp et al., 2006).

4.5.2 Amoebozoa

Les Amoebozoa rassemblent des dizaines de lignées d'amibes aérobies ou anaérobies, la majorité se déplaçant par vagues cytoplasmiques internes à l'aide de pseudopodes et étant dépourvue de flagelle. La plupart des Amoebozoa sont des hétérotrophes vivant en liberté, qui se nourrissent d'autres cellules grâce à leurs pseudopodes (certaines grandes amibes peuvent même manger de petits animaux). Elles sont souvent rencontrées dans les sols et les milieux aquatiques. Leur taille varie beaucoup : nombreuses sont celles qui mesurent seulement 10 à 20 μm , mais certaines peuvent atteindre 800 μm (*Amoeba proteus*).

Ce groupe comprend les **myxomycètes** (auxquels appartiennent les amibes sociales telles que *Dictyostelium discoideum*, dont les individus unicellulaires sont capables de s'agréger pour former un organisme "multicellulaire", dans lequel certaines cellules se différencient pour créer la tige, tandis que d'autres forment le corps fructifère. Ce dernier se transformera en spores qui seront disséminées), les **Lobosea** (qui incluent des amibes à thèques et des amibes nues telles que *Acanthamoeba castellanii*) et certains organismes amitochondriaux (les **Pelobiontes**, tels que *Mastigamoeba balamuthi* et les **Entamoebae**, qui sont des organismes pathogènes, tels que *Entamoeba histolytica*).

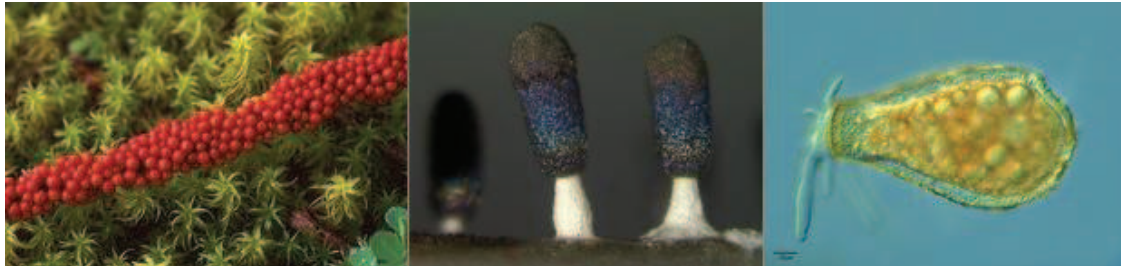


Figure 21. Quelques exemples d'Amoebozoa. De gauche à droite : le myxomycète *Leocarpus fragilis* (© 2008 Mavi Rodriguez Garcia), le myxomycète *Diachea leucopodia* (© 2006 Randy Darrah), et un lobosoa *Nebela tubulosa* (© 2009 Antonio Guillén). Source : <http://tolweb.org/>.

Le soutien pour la monophylie des **Amoebozoa** est assez récent puisque les premières études phylogénétiques de séquences d'ARNr suggéraient que les divers Amoebozoa étaient en fait des groupes non apparentés (Pawlowski et al., 1996; Pawlowski et al., 1999; Stiller and Hall, 1999). L'amélioration de l'échantillonnage taxonomique et des méthodes tendent à montrer que ces organismes sont liés. Une concaténation de 132 protéines a permis d'établir la monophylie du groupe incluant *D. discoideum*, *E. histolytica* et *M. balamuthi* (Baptiste et al., 2002).

4.5.3 Plantae

Le groupe des Plantae (ou Archaeplastida) inclut les plantes terrestres et les algues vertes (telles que *Chlamydomonas reinhardtii*), les algues rouges (rhodophytes), et les glaucophytes. Ses membres se distinguent par la présence d'un plaste dit "primaire", résultant de "l'endosymbiose primaire". Celle-ci correspond à l'internalisation d'une cyanobactérie par l'ancêtre de ces organismes, puis à sa réduction génomique.

L'incorporation du plaste primaire a eu un effet très important sur la biologie et le potentiel génétique des organismes hôtes. Presque toutes les Plantae sont phototrophes; quelques-uns sont des parasites non-photosynthétiques (ex : les *Orobanche* qui sont des plantes herbacées parasitant d'autres plantes), mais même ces organismes conservent des plastes sous forme réduite.

Les phylogénies de plusieurs gènes des plastes et l'organisation des génomes plastidiques suggèrent que les plastes de ces groupes forment une seule lignée émergeant au sein des cyanobactéries (Bachvaroff et al., 2005; Helmchen et al., 1995; Martin et al., 1998; Reyes-Prieto and Bhattacharya, 2007; Rodriguez-Ezpeleta et al., 2005; Rodriguez-Ezpeleta et al., 2007a), et indiquent ainsi que l'endosymbiose primaire n'a eu lieu qu'une fois, chez l'ancêtre de ces trois groupes.

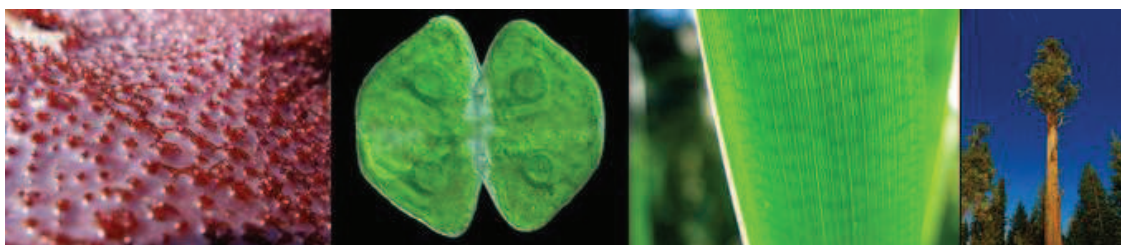


Figure 22. Exemples de Plantae. De gauche à droite : l'algue rouge *Chondracanthus* (© Patrick Keeling), l'algue verte *Cosmarium* (© Patrick Keeling) et les plantes terrestres *Typha latifolia* (© Patrick Keeling) et *Calocedrus decurrens* (© 2008 Fort Photo). Source : <http://tolweb.org/>.

La monophylie de ce supergroupe a également été établie par l'analyse phylogénétique du facteur d'élongation EF-2 (Moreira et al., 2000) et la concaténation de gènes nucléaires (Burki et al., 2008; Hackett et al., 2007; Reyes-Prieto and Bhattacharya, 2007; Rodriguez-Ezpeleta et al., 2005). Si la monophylie de ce groupe est rarement remise en question, elle peut néanmoins s'avérer difficile à retrouver selon les données analysées (Hampl et al., 2009; Minge et al., 2009; Nozaki et al., 2007; Nozaki et al., 2003; Parfrey et al., 2010), ou bien être faiblement soutenue (Burki et al., 2007). En parallèle, les relations de parenté entre les trois lignées de plantes restent très débattues (Deschamps and Moreira, 2009; Reyes-Prieto and Bhattacharya, 2007).

4.5.4 Rhizaria

Le groupe des Rhizaria a été proposé en 2002 par Cavalier-Smith, et réunit une grande diversité d'organismes unicellulaires libres (dont la majorité se nourrit en utilisant des pseudopodes), ainsi que certains parasites de plantes et d'animaux (Cavalier-Smith, 2002). Les relations exactes entre tous ces organismes sont encore incertaines. Les rhizaires libres les plus connus sont les foraminifères, les radiolaires et les cercozoaires (Moreira et al., 2007).

Les **foraminifères** sont des amibes planctoniques ou benthiques marines, dont beaucoup possèdent des structures externes minéralisées qui fossilisent et en font d'importants indicateurs géologiques (Figure 23).

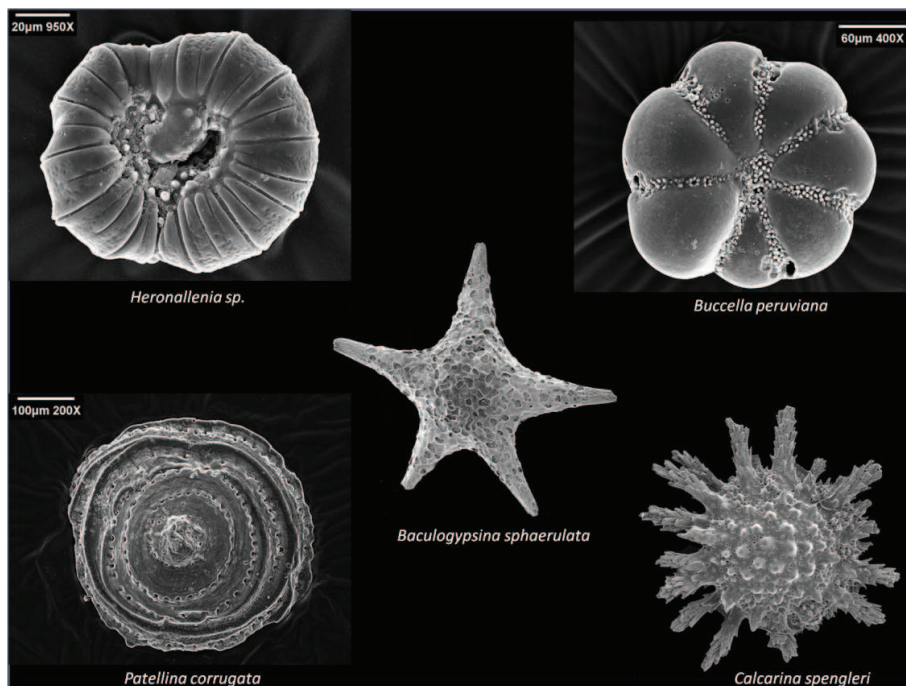


Figure 23. Divers foraminifères (images obtenues par microscopie électronique à balayage). Images : <http://www.flickr.com>.

Les **radiolaires** sont de grandes cellules planctoniques marines, présentant généralement un surprenant squelette interne de dioxyde de silicium à symétrie radiale.

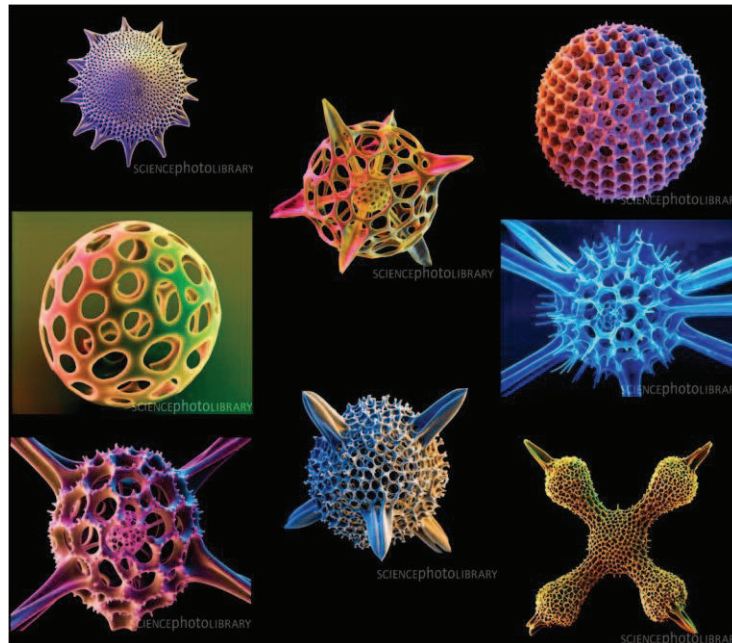


Figure 24. Squelettes de radiolaires vus par microscopie électronique à balayages (images colorisées). Images : <http://www.sciencephoto.com/>.

Le groupe des **cercozoaires** comprend plusieurs centaines d'espèces d'amibes et de flagellés dont la position phylogénétique a longtemps été incertaine. Ces organismes ont adopté des stratégies de vie très différentes, incluant la phagotrophie, le parasitisme et la photoautotrophie. Parmi les membres de ce groupe, les chlorarachniophytes (ex : *Bigelowiella natans*) ont attiré l'attention car ils hébergent des plastes particuliers issus d'une endosymbiose secondaire d'algue verte dont le noyau a fortement dégénéré pour donner un nucléomorphe. Des vestiges du réticulum endoplasmique de l'algue sont également retrouvés. Bien que capable d'effectuer la photosynthèse, ces organismes consomment aussi des proies qu'ils capturent via leurs filopodes (ce sont des mixotrophes).

Récemment, le cercozoa *Paulinella chromatophora*, a reçu une attention considérable car il peut être en mesure de fournir des indications biologiques sur la transition d'un endosymbionte cyanobactérien vers un organite photosynthétique (Marin et al., 2005; Nowack et al., 2008; Yoon et al., 2006). La cellule de *P. chromatophora* contient deux chromatophores qui ressemblent à des cyanobactéries cylindriques (Kies, 1974; Lauterborn, 1895) et il a été débattu pour savoir si ces chromatophores sont des cyanobactéries endosymbiotiques ou des organites photosynthétiques (Keeling, 2004; McFadden, 2001). Le génome du chromatophore d'une souche de *P. chromatophora* a été récemment séquencé, révélant que sa taille (~1 Mb) a été réduite par rapport à celle des génomes de *Prochlorococcus* et *Synechococcus* (~1,6 à 3,6 Mb) dont il est proche. Il lui manque également plusieurs gènes importants pour les cyanobactéries, dont un petit nombre de gènes photosynthétiques (Nowack et al., 2008). De plus, il a été montré que l'un des gènes photosynthétiques (*PsaE*) a été transféré depuis le chromatophore vers le génome nucléaire de *Paulinella* ce qui indique qu'une autre acquisition endosymbiotique primaire d'un organite photosynthétique est en cours (Nakayama and Ishida, 2009).

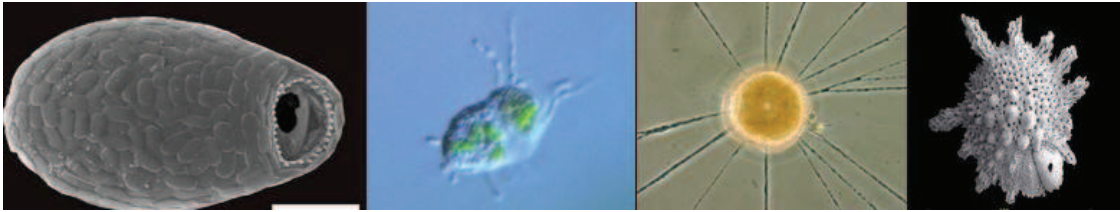


Figure 25. Exemples de Rhizaria. De gauche à droite : l'euglyphide (Cercozoa) *Corythion dubium* (© Edward Mitchell), le chlorarachniophyte (Cercozoa) *Gymnochlora* (© Patrick Keeling), les foraminifères *Allogromia* (© Jan Pawlowski and José Fahrni) et *Calcarina spengleri* (© 2008 Michael). Source : <http://tolweb.org/>.

La plupart des autres Rhizaria libres sont des protistes flagellés ou amiboïdes hétérotrophes qui consomment d'autres microbes, et ont probablement un rôle très important dans les habitats benthiques et les sols.

La monophylie de ce supergroupe est soutenue depuis peu de temps par les phylogénies des ARN ribosomiaux et de l'actine (Moreira et al., 2007; Nikolaev et al., 2004) et par la découverte d'une insertion dans le site actif de la polyubiquitine (Figure 16) (Archibald et al., 2003). A ce jour, aucun génome complet n'est disponible pour ce groupe, et seules les séquences EST sont disponibles et ont permis leur inclusion dans quelques récentes analyses phylogénomiques (Burki et al., 2008).

4.5.5 Excavata

Les Excavata représentent un groupe diversifié de protistes flagellés hétérotrophes ou parasites, les plus connues étant *Trypanosoma*, *Giardia* et *Trichomonas*. De nombreux excavés sont anaérobies ou microaérophiles et ne contiennent pas de mitochondries mais des mitosomes ou des hydrogénosomes (mitochondries dégénérées qui ne sont plus capables de produire de l'énergie à partir de l'oxygène). Une lignée d'excavés, les euglénides, comprend des espèces photosynthétiques possédant des plastes secondaires provenant de l'endosymbiose d'une algue verte (cf. p.43) (Breglia et al., 2007; Leander et al., 2007).

Le regroupement des lignées formant les Excavata repose sur une synthèse des données morphologiques et moléculaires. Beaucoup de ces organismes partagent une structure d'alimentation similaire (Simpson and Patterson, 2001). D'autres ne possèdent pas cette structure, mais, sur la base des phylogénies moléculaires, sont manifestement proches des lignées qui en possèdent (Simpson, 2003; Simpson et al., 2006; Simpson et al., 2002). Ces différents éléments conduisent à supposer l'existence d'un ancêtre commun exclusif, et quelques analyses phylogénomiques récentes fournissent un support pour la monophylie de l'ensemble du groupe. Ce support reste toutefois assez faible et encore très dépendant du choix des taxa et des marqueurs utilisés (Burki et al., 2008; Hampl et al., 2009; Rodriguez-Ezpeleta et al., 2007a; Simpson, 2003).

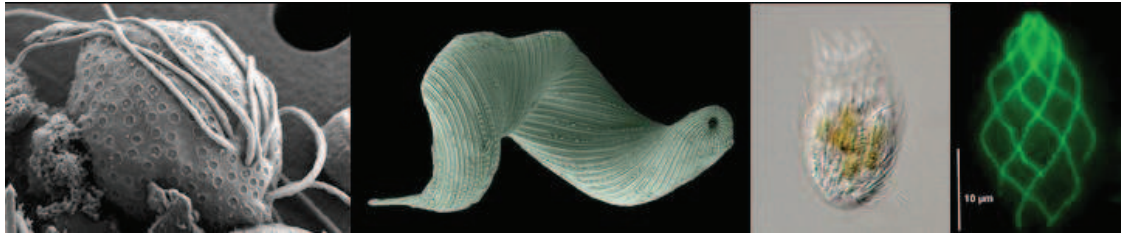


Figure 26. Exemples d'Excavata. De gauche à droite : l'oxymonade *Saccinobaculus minor* (© Kevin Carpenter and Patrick Keeling), l'euglénide photosynthétique *Lepocinclis spirogyra* (© 2003 Brian S. Leander), l'hétérolobose *Stephanopogon minuta* (© Naoji Yubuki et Brian S. Leander) et le parabasalide *Holomastigotes elongatum* (© Guy Brugerolle). Source : <http://tolweb.org/>.

4.5.6 Chromalveolata

Les chromalvéolés regroupent quatre grands groupes de protistes : les Haptophyta, les Cryptophyta, les Heterokonta et les Alveolata (Cavalier-Smith, 1999).

4.5.6.1 Haptophyta

Les **Haptophyta** sont des unicellulaires de très petite taille (nanoplancton), très répandus dans tous les océans, c'est un des groupes majeurs du phytoplancton. Ils jouent donc un rôle très important dans le cycle du carbone. Il existe quelques espèces terrestres ou d'eau douce. On estime à environ 500 le nombre d'espèces actuelles et il existe de nombreux groupes fossiles (les coccolithes, par exemple, sont importants en paléontologie car les cellules sont recouvertes d'écailles calcifiées). Le représentant actuel le plus connu est *Emiliana huxleyi*.

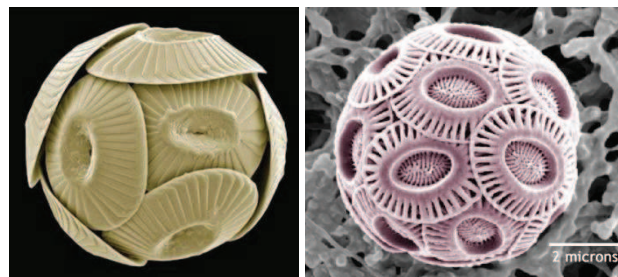


Figure 27. Coccolithus pelagicus et Emiliana huxleyi. Source : <http://tolweb.org> et Dr. Markus Geisen

Les haptophytes ont un ou deux plastes qui contiennent toujours des vestiges du réticulum de l'algue rouge à l'origine de ce plaste. Malgré leur capacité à réaliser la photosynthèse, certaines espèces ont néanmoins conservé la faculté de phagocyter des proies.

4.5.6.2 Cryptophyta

Ces algues vivent dans les eaux douces et marines. Elles représentent les proies de nombreux ciliés (Gustafson et al., 2000). Elles ont été peu étudiées et les données génomiques disponibles font défaut : le seul génome nucléaire en cours de séquençage est celui de *Guillardia theta*.

4.5.6.3 Alveolata

Les **Alveolata** forment un groupe monophylétique d'eucaryotes unicellulaires qui ont adopté des modes de vie très divers, tels que la prédation, le parasitisme intracellulaire et ou la photoautotrophie. Les alvéolés ont pour caractéristiques communes d'avoir des mitochondries à crêtes tubulaires et des vésicules sous-membranaires positionnées juste sous leur membrane formant des sortes d'alvéoles (d'où leur nom) (Adl et al., 2005).

La plupart des alvéolés font partie de l'un des trois sous-groupes principaux : les **ciliés**, les **dinoflagellés**, et **apicomplexes**.

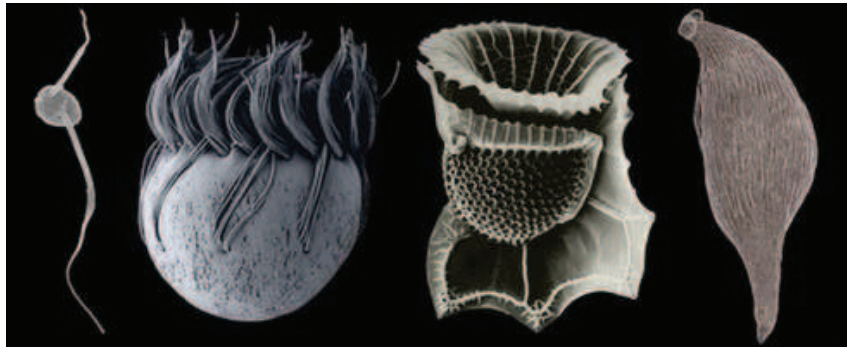


Figure 28. Exemples d'Alveolata. *Colpodella*, *Halteria* (ciliés), *Ornithocercus* (dinoflagellé), *Monocystis* (apicomplexe). Source : <http://tolweb.org>

Les **Apicomplexa** sont des parasites obligatoires. Le nom du groupe provient d'un appareil (le "complexe apical") situé à l'extrémité antérieure des parasites qui facilite la fixation sur la cellule hôte et son invasion.

Certains **Dinoflagellata** sont aussi des parasites, mais la plupart sont soit des prédateurs vivant en liberté soit des photoautotrophes dans les habitats aquatiques (Taylor, 1987). Beaucoup de dinoflagellés photosynthétiques sont en fait des mixotrophes car ils sont aussi des consommateurs de bactéries et d'autres microeukaryotes (Stoecker, 1999). Les dinoflagellés ont la caractéristique d'avoir des chromosomes condensés en permanence et de ne pas posséder de nucléosomes (Fensome et al., 1999).

Les **Ciliata** sont essentiellement des prédateurs, et tout comme les dinoflagellés et apicomplexes, ce groupe est extrêmement diversifié. Ils ont envahi de nombreuses niches écologiques (formes planctoniques ou benthiques de mer ou d'eau douce, vivant dans les sols humides etc.) et se nourrissent de bactéries, d'algues ou d'autres ciliés qu'ils ingèrent par leur cytostome (= "bouche cellulaire") (Lecointre and Le Guyader). Les principales caractéristiques de ces cellules sont (i) la présence d'une ciliature complexe et importante (d'où leur nom) qui leur permet des mouvements rapides ou la capture de proies; et (ii) leur polarisation avec la présence d'un avant et d'un arrière et une division qui est transversale (Frankel, 2000). Ce type de division implique des mécanismes morphogénétiques importants car la partie avant doit régénérer un arrière et réciproquement. Ils possèdent également deux types de noyaux : les micronoyaux et les macronoyaux. Les micronoyaux sont diploïdes et servent uniquement à la transmission de l'information au cours de la reproduction. Ils sont transcriptionnellement inactifs, ont des chromosomes usuels des eucaryotes et se divisent par des mitoses classiques. Les macronoyaux dérivent des micronoyaux. Ils portent des chromosomes fragmentés et sont transcriptionnellement actifs. Les macronoyaux se divisent par un processus de partition des molécules d'ADN qui n'est pas une mitose. Les micronoyaux forment donc une lignée de noyaux germinaux spécifiquement dévolus à la reproduction

alors que les macronoyaux sont une lignée somatique qui disparaît à chaque reproduction sexuée (Eisen et al., 2006).

La monophylie de ce groupe a été retrouvée dès l'analyse de l'ARNr SSU (Gajadhar et al., 1991), et n'a jamais été remise en question depuis.

4.5.6.4 Heterokonta

Les Heterokonta (aussi appelés straménopiles) regroupent de nombreuses lignées d'organismes (environ 100 000 espèces) qui présentent une grande variété de plans d'organisations unicellulaires et pluricellulaires. Ils ont envahi pratiquement tous les biotopes (milieux marins, d'eaux douces mais aussi terrestres).



Figure 29. Exemples d'hétérokontes. De gauche à droite : *Cymbella tumida*, *Macrocystis integrifolia*, *Saprolegnia* (fly fungus). Source : <http://tolweb.org>

Les hétérokontes sont caractérisés par la présence à un moment de leur cycle d'une cellule possédant deux flagelles, dont l'un porte des mastigonèmes (appendices protéiques mis en place à partir de l'appareil de Golgi) (Adl et al., 2005).

Ce groupe contient des organismes qui ont longtemps été classés avec les champignons (comme les oomycètes) ou les algues (comme les algues brunes ou jaunes). Les analyses moléculaires et ultra-structurales ont permis de rapprocher de ce groupe de nombreuses espèces de protistes dont on connaît mal la biologie (Cavalier-Smith and Chao, 2006; Patterson, 1999; Patterson and Sogin, 1992). La monophylie de ce groupe a été largement retrouvée depuis dans les analyses phylogénétiques basées sur les concaténations de gènes.

Il est aujourd'hui largement admis que les Heterokonta forment un groupe monophylétique avec les Alveolata. Celui-ci est soutenu par de nombreux arbres de gènes (y compris celui de l'ARNr SSU) et par les analyses individuelles ou concaténées de gènes nucléaires codant pour des protéines (Baldauf et al., 2000; Harper et al., 2005; Philippe et al., 2004; Van de Peer and De Wachter, 1997).

4.5.6.5 Les endosymbioses secondaires et tertiaires

Comme décrit précédemment, l'endosymbiose d'une cyanobactérie a permis l'acquisition de la photosynthèse et est peut être en cours chez *Paulinella chromatophora*. Ces "plastés primaires" existent chez les organismes du groupe des Plantae.

Plus récemment au cours de l'évolution, des eucaryotes non photosynthétiques ont acquis des plastés d'une autre façon, c'est-à-dire en internalisant un eucaryote photosynthétique au lieu d'une bactérie. Quand le symbiote photosynthétique est un membre des Plantae, l'événement est appelé "endosymbiose secondaire", tandis qu'il est appelé "endosymbiose tertiaire" ("quaternaire", ...) s'il désigne l'internalisation d'un symbiote qui est déjà le produit d'une précédente endosymbiose eucaryote-eucaryote (EEE). L'existence des EEE est soulignée par la présence de trois ou quatre membranes (voire plus) entourant certains

plastes, qui impliquent le ciblage des protéines plastidiques codées par le noyau *via* de multiples séquences d'adressage (Cavalier-Smith, 2003), et par le fait que, dans certaines lignées, l'endosymbionte conserve un noyau vestigial ("nucléomorphe") (Ludwig and Gibbs, 1987).

Les Cryptophyta, Alveolata, Heterokonta, et Haptophyta (ci-après collectivement dénommés CASH, d'après (Baurain et al., 2010)) sont quatre groupes qui comprennent à la fois des lignées photosynthétiques et non photosynthétiques.

Les plastes des CASH proviennent probablement d'un seul événement initial d'endosymbiose secondaire d'une algue rouge (Gould et al., 2008; Palmer, 2003; Reyes-Prieto et al., 2007). En effet, toutes les espèces photosynthétiques des CASH contiennent de la chlorophylle *c*, qui est absente chez toutes les autres algues (y compris les rouges). De plus, dans les phylogénies basées sur les gènes plastidiques (Yoon et al., 2002b), ainsi que sur certaines protéines codées dans le noyau et impliquées dans la fonction du plaste (Fast et al., 2001; Patron et al., 2004; Petersen et al., 2006), la monophylie des CASH est retrouvée.

En revanche, il existe une controverse sur la façon dont ont été hérités les plastes des CASH (et des gènes codés dans le noyau et participant au fonctionnement des plastes). Ont-ils été hérités de façon strictement verticale à partir de l'ancêtre de ces groupes, ceci impliquant une seule endosymbiose secondaire ("hypothèse chromalvéolés" (Cavalier-Smith, 1999, 2003; Keeling, 2009) (Figure 30) ? Ou bien se propagent-ils horizontalement, comme suggéré par les hypothèses d'EEEs "en série" ou "multiples" (Figure 31) (Archibald, 2009; Cavalier-Smith et al., 1994; Sanchez-Puerta et al., 2007) ?

En proposant une endosymbiose unique à l'origine des lignées CASH, l'hypothèse chromalvéolés est la plus parcimonieuse, mais repose sur deux hypothèses qui sont très discutées : les événements EEE seraient extrêmement rares, et, deuxièmement, les nombreuses espèces CASH qui ne possèdent pas de plaste auraient toutes perdu leurs plastes secondairement et indépendamment.

Dans l'hypothèse concurrente (EEE "en série"), les plastes des lignées CASH proviendraient également de l'endosymbiose secondaire unique d'une algue rouge dans un hôte eucaryote, mais le plaste se serait ensuite propagé dans des lignées CASH distantes via des EEE successives, aboutissant à des "chimères" évolutives de type "poupées russes" (Archibald, 2009; Baurain et al., 2010; Bodyl, 2005; Cavalier-Smith et al., 1994; Sanchez-Puerta et al., 2007).

En conséquence, la monophylie observée lors de l'analyse des protéines adressées au plaste mais codées dans le noyau ne serait pas due à l'héritage vertical. Au lieu de cela, ces gènes seraient relocalisés dans le noyau du nouvel hôte à partir du nucléomorphe, par coévolution avec les plastes transférés en série (Petersen et al., 2006). Ces dernières années, la plausibilité des hypothèses "en série" a augmenté avec des exemples confirmés d'EEE d'ordre supérieur et indépendantes (c'est-à-dire résultant de l'endosymbiose d'un cryptophyte, d'un haptophyte ou encore d'un hétérokote selon les lignées) dans les dinoflagellés (Hackett et al., 2004; Inagaki et al., 2000; Tengs et al., 2000).

À l'heure actuelle, aucune des deux hypothèses n'a été clairement validée par les études phylogénétiques. La monophylie des lignées CASH a été testée intensivement, pourtant sans succès. En effet, les cryptophytes et haptophytes n'ont jamais été replacés de façon robuste par les analyses phylogénomiques basées sur les gènes nucléaires (Burki et al., 2009; Burki et al., 2007; Burki et al., 2008; Hackett et al., 2007; Patron et al., 2007).

En outre, les études récentes basées sur l'analyse phylogénomique des gènes nucléaires indiquent que Alveolata et Heterokonta seraient plus proches des Rhizaria que des Cryptophyta et Haptophyta (Burki et al., 2009; Burki et al., 2007; Burki et al., 2008; Hackett et

al., 2007; Rodriguez-Ezpeleta et al., 2007a), conduisant à l'ajout de ce super groupe dans les chromalvéolés alors que la plupart des Rhizaria sont non photosynthétiques et que les quelques espèces photosynthétiques portent des plastides d'origine algue verte (Gould et al., 2008; Palmer, 2003; Reyes-Prieto et al., 2007).

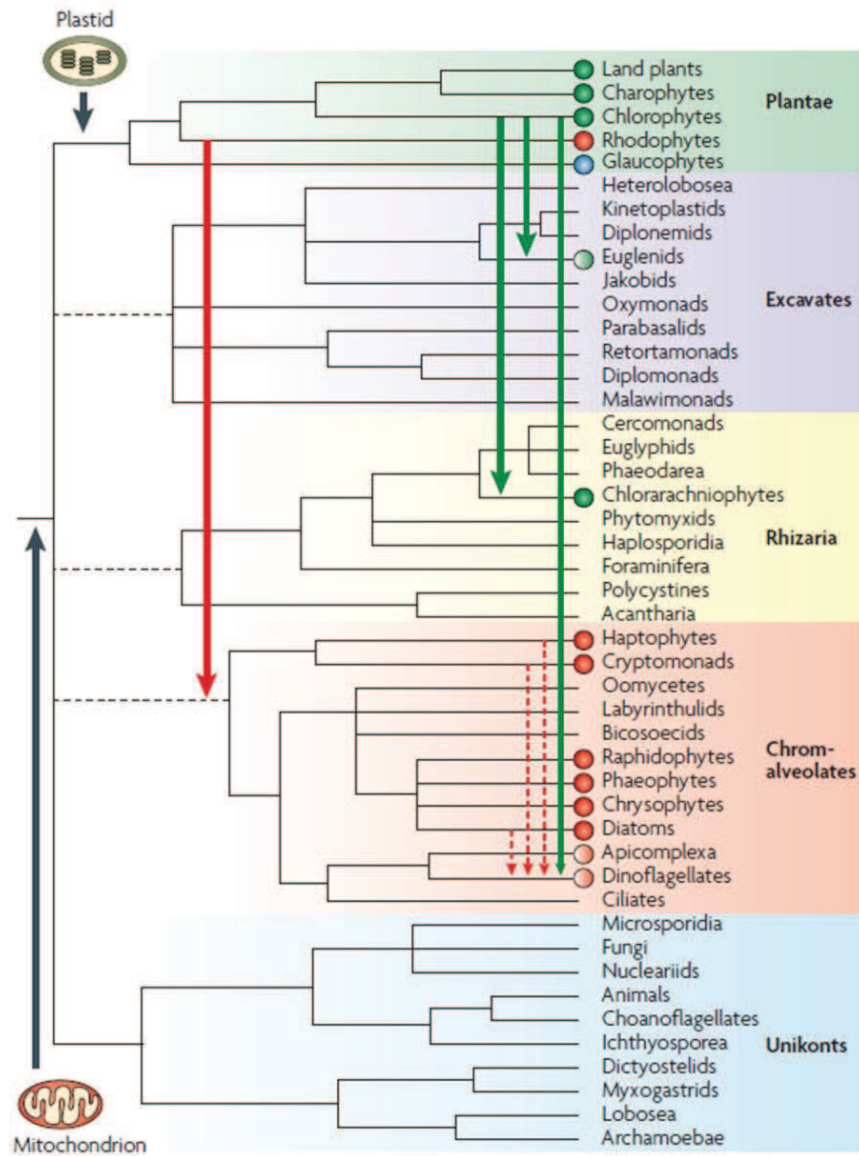


Figure 30. Hypothèse chromalvéolés (Keeling and Palmer, 2008). Les traits en pointillés indiquent les clades incertains. Les événements qui ont conduit à l'endosymbiose mitochondriale et aux endosymbioses chloroplastiques sont indiqués par les flèches verticales. Les flèches noires indiquent les endosymbioses primaires à l'origine des mitochondries et des plastides. Les flèches rouges et vertes indiquent les transferts de plastides rouges ou verts, par endosymbioses secondaires (traits pleins) et tertiaires (traits pointillés). Les lignées dans lesquelles un plaste existe actuellement sont indiquées par des cercles de couleur (rouge ou vert, respectivement pour les algues rouges ou les algues vertes et tous les plastides qui en descendent, et bleu pour les glaucophytes). Les cercles correspondant aux lignées contenant un mélange important d'espèces photosynthétiques et non-photosynthétiques sont partiellement remplis. Le plaste des dinoflagellés est indiqué en rouge pour plus de simplicité, parce qu'une seule lignée contient un plaste secondaire vert.

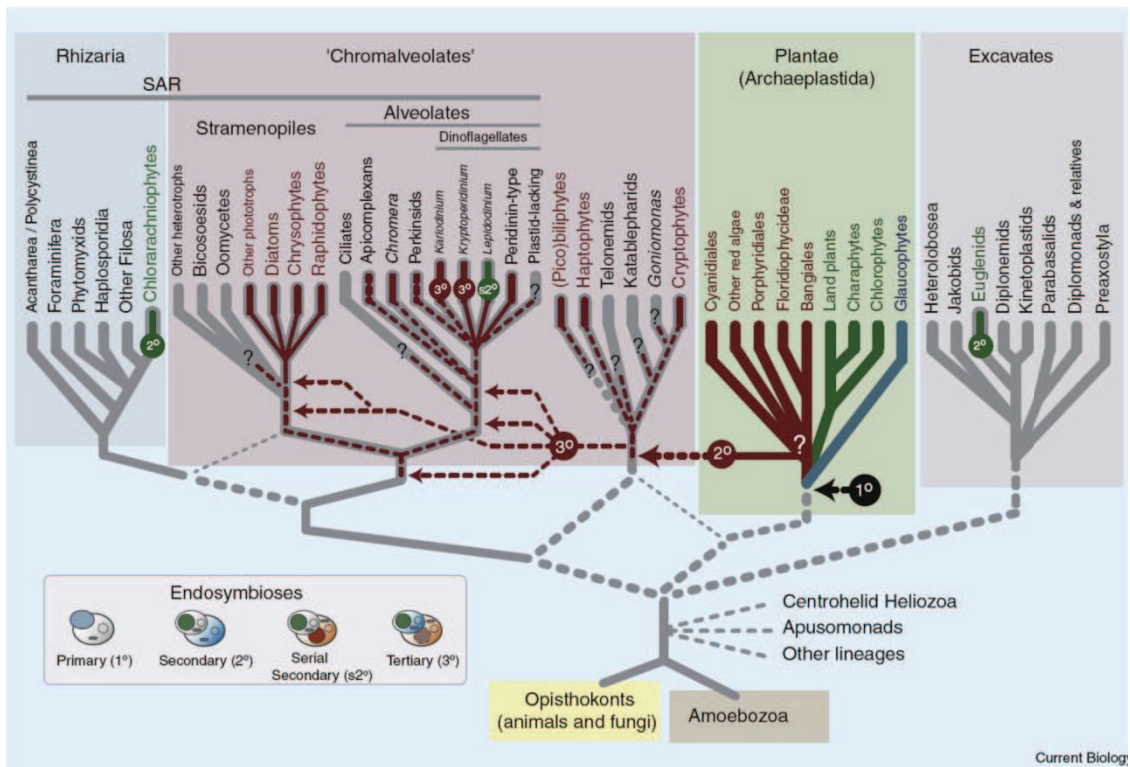


Figure 31. Hypothèse de transfert horizontal du plaste rouge par endosymbioses eucaryote-eucaryote en série (Archibald, 2009). Les évènements d'endosymbiose primaire, secondaire et tertiaires sont représentés sur une topologie consensus des analyses phylogénomiques récentes de gènes nucléaires. Les pointillés représentent les relations et évènements incertains.

4.5.7 Les relations entre grands groupes, la racine de l'arbre des Eucarya

Si la monophylie de tout ou partie de ces grands groupes est généralement admise, les liens qui les unissent restent toujours très débattus. Entre autres, les tentatives de résoudre l'ordre des spéciations ayant donné naissance aux grandes lignées eucaryotes se sont révélées non concluantes, notamment à cause des artefacts de reconstruction liés à l'inclusion de groupes extérieurs procaryotes très distants.

Une méthode alternative est d'utiliser des caractères génétiques rares tels que des fusions de gènes, des insertions/délétions et la présence ou l'absence de gènes particuliers. Ainsi, Philippe *et al.* en 2000 ont constaté que la fusion des gènes codant pour la dihydrofolate réductase (DHFR) et la thymidylate synthase (TS) était absente chez les Opisthokonta, les Amoebozoa et les procaryotes, et présente chez tous les autres groupes eucaryotes (Philippe *et al.*, 2000b). Cavalier-Smith *et al.* propose que cette fusion constitue un caractère dérivé, et donc que la racine de l'arbre ne peut pas se placer à l'intérieur du groupe présentant cette fusion. En parallèle, on retrouve chez les Opisthokonta et les Amoebozoa, la fusion de trois gènes codant des enzymes intervenant dans la synthèse de nucléotides pyrimidines : la carbamoyl phosphate synthase, la dihydroorotase et l'aspartate carbamoyltransferase (Cavalier-Smith, 2002). Cette fusion étant absente chez les autres eucaryotes et les procaryotes, Stechmann et Cavalier-Smith ont considéré cet événement de double fusion, lui aussi, comme un caractère dérivé. La racine se placerait donc entre ces deux méga-groupes eucaryotes (c'est-à-dire Opisthokonta et Amoebozoa d'un côté; tous les autres grands groupes de l'autre) (Stechmann and Cavalier-Smith, 2003b).

En 2002, Cavalier-Smith a proposé les noms "Unikonta" et "Bikonta" sur la base d'un scénario évolutif de l'appareil flagellaire. En effet, dans la plupart des cellules eucaryotes l'appareil flagellaire est le centre de l'organisation du cytosquelette. Le cœur du système est constitué d'un un ou plusieurs corps basaux, qui peuvent donner naissance ou non à des flagelles (Cavalier-Smith, 2002). Cavalier-Smith émet l'hypothèse que LECA avait un seul corps basal et un unique flagelle. Il suggère également que les Unikonta (= "flagelle unique") ont conservé cette organisation ancestrale, alors que les Bikonta (= "biflagellé") descendent d'un ancêtre qui a acquis un second flagelle, l'un étant antérieur, l'autre postérieur. Les bikontes sont également caractérisés par un processus particulier lors de la division cellulaire : le corps basal nouvellement formé devient systématiquement le corps antérieur, tandis que l'ancien (*i.e.* celui hérité de la cellule mère) donne le corps basal postérieur. Ainsi, ces caractères dérivés combinés à la fusion des gènes DHFR-TS excluraient la possibilité que la racine se situe à l'intérieur des bikontes.

Cependant, un des problèmes de cette hypothèse vient du fait que de nombreux unikontes ont en réalité plusieurs corps basaux (Roger and Simpson, 2009). Les Opisthokonta flagellés ont typiquement deux corps basaux - un flagellé, l'autre non. De plus, il a été récemment découvert une amibe biflagellée apparentée aux Amoebozoa. Ainsi, si l'on reporte le nombre de corps basaux sur une phylogénie des eucaryotes (Figure 32), alors l'idée que les "unikontes" étaient ancestralement uniflagellés ne semble plus tenir (Roger and Simpson, 2009). Si la position phylogénétique de cette amibe (*Breviata*) est confirmée, alors l'hypothèse la plus parcimonieuse est que l'ancêtre des Unikontes était lui-même biflagellé, tout comme l'ancêtre de tous les eucaryotes et qu'un des flagelle a été perdu au cours de l'évolution de certains unikontes.

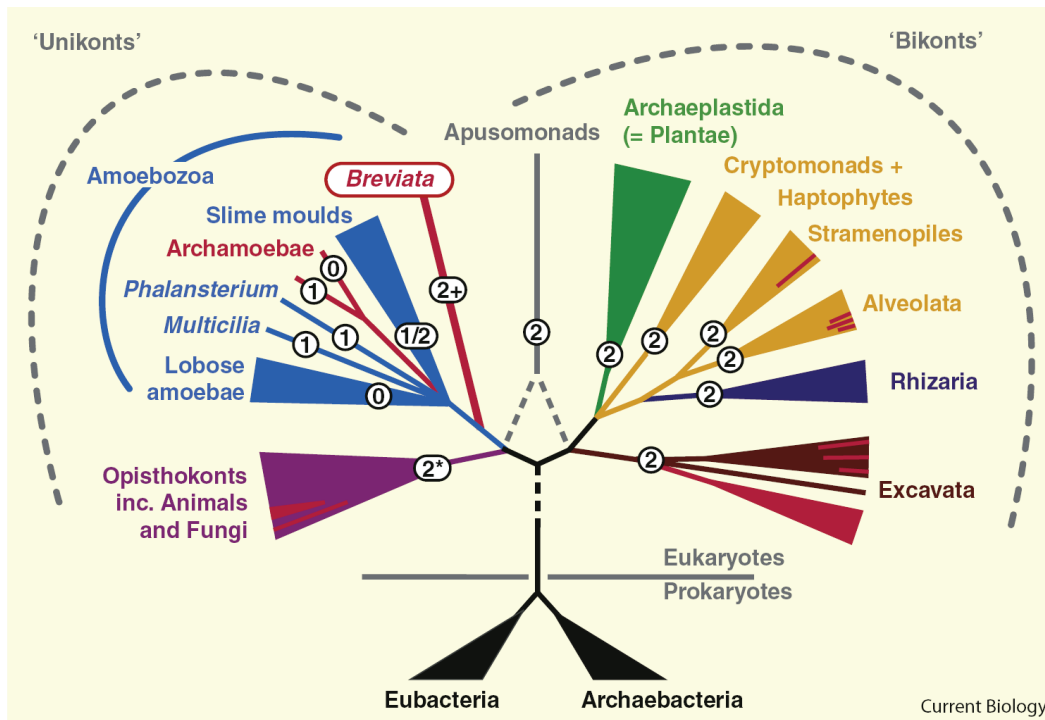


Figure 32. Nombre de corps basaux dans les différentes lignées eucaryotes. L'arbre est enraciné selon l'hypothèse Unikonta-Bikonta. Les nombres indiquent le nombre de corps basaux inférés chez l'ancêtre des différentes lignées. Le signe (+) indique que *Breviata* possède potentiellement plus de deux corps basaux. L'astérisque (*) indique qu'un des deux corps basaux est non flagellé. Les lignes pointillées indiquent une incertitude concernant le placement de cette branche (Roger and Simpson, 2009).

La question de la position de la racine est donc toujours ouverte, et aucun consensus n'est pour l'instant atteint. C'est pourquoi, tout en gardant à l'esprit ce point crucial, et dans l'attente de nouvelles avancées sur le sujet, j'ai adopté, pour mes analyses, la position la plus communément admise et avons raciné notre arbre de référence entre unikonts et bikonts. J'ai néanmoins considéré d'autres enracinements potentiels, et les implications qu'ils peuvent avoir sur mes résultats.

5 LECA, le dernier ancêtre commun des Eucarya

Quel que soit l'endroit où se place la racine, LECA représente le dernier ancêtre commun de l'ensemble des eucaryotes vivants actuellement, et dresser son "portrait-robot" peut permettre de répondre à des questions relatives à l'origine et l'évolution des eucaryotes. Il est également indispensable de bien connaître cet ancêtre ainsi que l'état des différentes fonctions biologiques qui existaient chez lui pour mieux comprendre comment ces fonctions ont évolué par la suite dans les différentes lignées eucaryotes. Cet aspect constitue le principal axe de recherche que j'ai développé pendant ma thèse.

5.1 Deux approches pour inférer les caractéristiques de LECA

L'avènement de la génomique a fourni les données permettant d'entreprendre des études à grande échelle par des l'analyse de protéomes complets. Il est alors possible de reconstruire l'état ancestral du protéome le long d'un arbre des espèces, aussi bien d'un point de vue quantitatif que qualitatif.

5.1.1 Reconstruction de protéomes ancestraux complets

5.1.1.1 Le protéome de LECA : aspect quantitatifs

L'étude quantitative de l'évolution du protéome d'un groupe consiste à comparer le répertoire protéique codé par plusieurs génomes. Le cœur de protéines présentes dans l'ensemble des organismes comparés peut être inféré comme étant le nombre minimal de protéines que possédait leur dernier ancêtre commun.

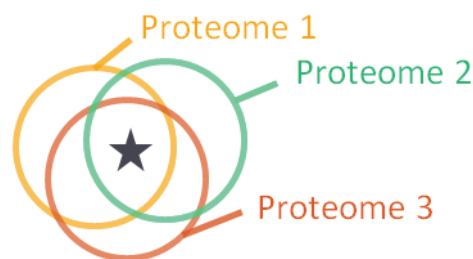


Figure 33. Inférence du nombre minimum de gènes présents chez l'ancêtre, par comparaison des protéomes. L'étoile représente le cœur de protéines commun aux trois organismes.

En 2005, **Makarova et collaborateurs** entreprennent une reconstruction du répertoire de gènes présents chez LECA, et estiment que cet ancêtre devait déjà posséder plus de 4100 gènes, contre seulement 2150 chez le premier eucaryote, mettant ainsi en avant un processus évolutif guidé par un taux de duplications particulièrement élevé chez les Eucarya comparés aux Archaea et aux Bacteria (Makarova et al., 2005).

Leur approche consiste à inférer la présence d'un gène chez LECA si celui-ci était retrouvé dans au moins deux des quatre groupes suivants : plantes (représentées par *Arabidopsis thaliana* et *Oryza sativa*), animaux (représentés par *Caenorhabditis elegans*, *Drosophila melanogaster* et *Homo sapiens*), champignons-microsporidies (*Saccharomyces cerevisiae*, *Schizosaccharomyces pombe*, *Magnaporthe grisea*, et *Encephalitozoon cuniculi*) et apicomplexa-diplomonadida (représentés par *Plasmodium falciparum* et *Giardia lamblia*). Le dernier "groupe" est représenté par des lignées non apparentées (un apicomplexe et un excavé), mais les auteurs le justifient par le fait que ces deux organismes étant des parasites,

ceux-ci sont enclins à des pertes massives de gènes. Si elles sont différentes entre les deux lignées, les considérer comme un seul groupe permettrait de compenser ces pertes. De plus, ils infèrent comme étant présent chez LECA des gènes qui sont partagés uniquement par les champignons et les animaux (c'est-à-dire des innovations *a priori* récentes, ou en tout cas postérieures à LECA et spécifiques des Opisthokonta), en théorie pour compenser les pertes potentielles ayant eu lieu dans les autres lignées. Ce type d'inférence est intéressant, mais néanmoins, ces deux derniers aspects de leur approche peuvent paraître un peu curieux. En effet, dans le meilleur des cas, le nombre de gènes inférés comme étant présents chez LECA est proche de la réalité. En revanche, la nature des gènes (et donc des fonctions biologiques) qui sont prédits chez l'ancêtre est, du moins pour une part, très certainement incorrecte (puisqu'ils infèrent comme étant présents chez LECA des gènes issus d'innovations opisthokontes).

5.1.1.2 Le protéome ancestral mitochondrial

Une analyse du même type a été menée pour tenter de répondre en partie à la question de l'origine des Eucarya. Afin de trancher entre les différentes hypothèses concernant les propriétés métaboliques de l'endosymbionte ancêtre de la mitochondrie (cf. p.19), **Gabaldón et Huynen** ont reconstruit le protéome mitochondrial ancestral (Gabaldon and Huynen, 2003). Les auteurs ont employé une approche phylogénomique massive en reconstruisant de façon automatique des milliers d'arbres phylogénétiques de protéines codées par six génomes d'alpha-protéobactéries, pour ensuite sélectionner ceux dont la topologie indiquait la présence d'un membre de cette famille de protéines dans l'ancêtre mitochondrial, c'est-à-dire qui montraient une relation étroite entre les protéines d'alpha-protéobactéries et leurs orthologues eucaryotes et qui n'étaient pas le résultat de THG récents. En outre, le report des fonctions de ces familles de protéines sur les cartes métaboliques a permis l'inférence du métabolisme proto-mitochondrial. L'image qui émerge est celle d'un endosymbionte aérobie facultatif catabolisant des composés fournis par l'hôte. Certaines voies métaboliques proto-mitochondriales non directement liées à l'adénosine triphosphate (ATP) ont été conservées chez les eucaryotes modernes (ex : le métabolisme du fructose/mannose, la synthèse des lipides, des nucléotides, et des vitamines) (Gabaldon and Huynen, 2003). Ceci suggère des avantages multiples pour la cellule hôte et pourrait laisser penser qu'ont existé des eucaryotes amitochondriaux qui n'auraient pas laissé de descendants dû à l'avantage évolutif conféré à d'autres eucaryotes par l'acquisition de la mitochondrie.

5.1.1.3 Avantages et inconvénients de ces approches

Ce type d'approche à grande échelle présente l'avantage d'analyser un très grand nombre de données et l'étude de toutes les fonctions biologiques à la fois. L'inconvénient est directement lié à la gestion d'un si grand nombre de données. En effet, il est indispensable dans ce cas d'employer des méthodes automatiques pouvant introduire des biais à tous les niveaux de l'analyse (récupération des homologues, identification des orthologues, alignement, sélection des bornes ...). Ainsi, les phylogénies reconstruites de cette façon ne permettent pas toujours de prendre en compte l'ensemble des phénomènes de paralopies cachées, de transferts horizontaux de gènes, de divergences importantes, etc. Ces difficultés croissent d'autant plus que l'on étudie des événements évolutifs anciens, notamment chez les eucaryotes, au sein desquels ont eu lieu de nombreuses duplications et pertes de gènes et des recombinaisons de domaines protéiques. L'emploi de ces méthodes peut conduire à de fortes altérations dans la qualité des reconstructions, et donc des inférences qui en sont déduites.

5.1.2 Etude phylogénomique de l'évolution des systèmes biologiques

Une autre approche consiste essayer de caractériser les systèmes biologiques présents chez LECA, c'est-à-dire à réaliser l'analyse évolutive intégrée des systèmes biologiques d'intérêt. Contrairement aux approches présentées précédemment, il s'agit de se focaliser sur l'histoire évolutive d'un complexe protéique, d'une voie métabolique ou d'une structure cellulaire, et non pas sur l'ensemble du protéome d'un organisme. Cela permet, d'une part, de prédire le nombre de composants du système qui étaient déjà présents chez LECA, mais également de réaliser des inférences fines sur la fonction de ces composants dans cet ancêtre, puisque la seule inférence de la présence d'un composant ne suffit pas à prédire qu'il avait le même rôle dans la cellule. Le choix des systèmes étudiés se fait en fonction de la problématique évolutive ciblée. Pour ma part, il s'agit de systèmes ayant pu jouer un rôle dans l'origine et la diversification des lignées eucaryotes actuelles.

Cette approche peut être considérée comme intégrative : c'est celle que j'ai favorisée. Plus concrètement, elle consiste à reconstruire de façon précise la phylogénie de chaque composant du système, ce qui permet la détection des transferts horizontaux, des paralogies cachées, des artefacts de reconstructions, etc., qui sont fréquents à cette échelle évolutive. Il s'agit ensuite d'intégrer, sur ces phylogénies, un ensemble de données fonctionnelles telles que la composition en domaines fonctionnels, le contexte génomique, les données d'expression et d'interactions, etc.

Cette méthode permet de répondre à des questions importantes sur le système étudié, telles que (Eisen and Fraser, 2003; Eisen and Hanawalt, 1999) :

- Quand (*i.e.* dans quelle lignée évolutive) est-il apparu ? Sous quelle forme (*i.e.* quels composants étaient déjà présents) ?
- Quelles étaient les fonctions des composants ancestraux ?
- Quels innovations/pertes/remplacements/duplications/TGH sont survenus dans chaque lignée ?
- Est-ce le système considéré existe dans tel organisme (actuel ou ancestral) d'intérêt ? Si oui, sous quelle forme (prédictions fonctionnelles) ?
- Y-a-t-il eu un changement de fonction biochimique/biologique de certains composants dans certaines lignées ?
- Y-a-t-il une coévolution des différents composants ?
- Ce système a-t-il un lien évolutif avec un autre système ?

Les résultats de ces analyses peuvent permettre de mieux comprendre l'origine et l'évolution des systèmes biologiques eucaryotes et par extension d'éclairer la mise en place de la cellule eucaryote avant LECA, et les étapes précoces de la diversification ayant engendré les grandes lignées eucaryotes.

Il est aujourd'hui possible de favoriser cette approche car, depuis quelques années, les données nécessaires à son entreprise deviennent de plus en plus nombreuses. Le rythme de séquençage des génomes s'est considérablement accru donnant ainsi accès à l'information génétique de nombreux organismes. Le nombre de séquences génomiques eucaryotes reste encore réduit en comparaison des données disponibles sur les Procaryotes (155 génomes complets eucaryotes et environ 1800 projets en cours ou projets EST, contre près de 1500 génomes complets disponibles pour les procaryotes, et 8000 projets en cours début 2011) (Genomes OnLine Database, www.genomesonline.com). Cependant, une bonne couverture taxonomique des principaux groupes à l'intérieur de chacun des trois domaines commence à être atteinte, en particulier au sein des Eucarya (Figure 34), offrant la possibilité d'entreprendre des analyses de génomique comparative à grande échelle.

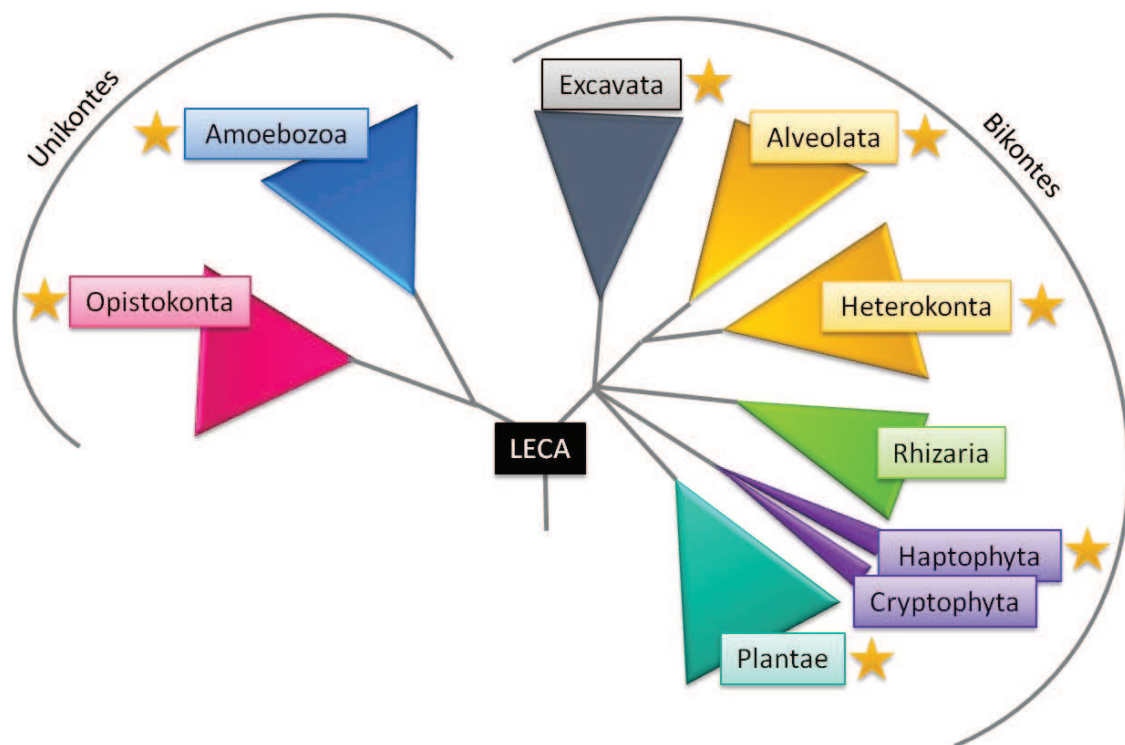


Figure 34. Couverture taxonomique des génomes complets eucaryotes. Les groupes pour lesquels au moins un génome complet est disponible sont marqués par une étoile. En plus des groupes qui sont représentés ici, il existe des lignées dont la position dans l'arbre est mal connue et pour lesquels aucun génome n'est disponible.

En parallèle, les dernières années ont vu le développement de nouvelles techniques de protéomique à haut débit (spectrométrie de masse en tandem MS/MS, chromatographie en phase liquide suivie de MS/MS, ...). Celles-ci ont permis d'entreprendre l'identification des protéines constituant certains systèmes, structures ou processus biologiques (tels que la réplication, la réparation, la traduction, la transcription, la division cellulaire, les voies métaboliques, structures cellulaires, ...). En particulier, il est maintenant possible de déterminer l'ensemble des composants de gros complexes protéiques.

Enfin, bien que ces études soient généralement limitées à quelques organismes modèles, nous connaissons le fonctionnement cellulaire de quelques-unes de ces systèmes biologiques.

Cet accroissement des données génomiques disponibles couplé à une meilleure connaissance des systèmes biologiques permet d'entreprendre l'analyse comparée des systèmes biologiques à grande échelle (*i.e.* à l'échelle du vivant), et d'appliquer les méthodes de phylogénomique intégrée décrites ci-dessus.

5.1.3 Les SME

Parmi les systèmes, structures ou processus biologiques pouvant nous renseigner sur l'origine et l'évolution des Eucarya en particulier, les Structures Multiprotéiques eucaryotes (SME) occupent une place privilégiée. Ces SME correspondent généralement à de gros complexes protéiques composés de plusieurs dizaines, voire plusieurs centaines de protéines différentes. Elles ont un rôle dans de nombreux processus fondamentaux de la cellule eucaryote, aussi bien au niveau de processus informationnels, c'est-à-dire impliqués dans l'expression et transmission de l'information génétique (ex : maturation de l'ARNm, transcription, traduction,

réplication), que de processus opérationnels, c'est-à-dire du maintien et de l'organisation de la cellule (ex : cytosquelette, voies métaboliques, ...).

De par leurs rôles cruciaux pour la cellule eucaryote, ces complexes sont considérés comme étant probablement très anciens, autrement dit, apparus très tôt au cours de l'évolution des eucaryotes. Néanmoins, il faut garder à l'esprit que dans la majorité des cas, ces structures n'ont été observées et/ou caractérisées que chez un ou quelques organismes modèles. Rien ne permet donc d'affirmer d'emblée que ces structures sont présentes chez tous les Eucarya. Le cas échéant, elles peuvent néanmoins présenter des variations de forme.

Chez les procaryotes, qui possèdent une structure cellulaire plus simple, il peut exister des processus similaires à ceux réalisés par certains SME chez les eucaryotes (par exemple, l'assemblage des ribosomes ou la division cellulaire). Pour autant, les systèmes qui les réalisent sont (généralement) plus simples et non homologues aux systèmes eucaryotes (*i.e.* aux SME). C'est pourquoi on considère que les SME constituent des marqueurs essentiels dans la distinction entre cellules eucaryotes et procaryotes. Comprendre l'origine et l'évolution des SME est donc une étape fondamentale pour appréhender l'origine et l'évolution de la cellule eucaryote elle-même.

Jusqu'à ces dernières années, et malgré leur importance, assez peu d'études systématiques concernant l'origine et l'évolution de ces structures ont été menées, la principale limite étant le peu de données disponibles, tant au niveau protéomique des structures (composition souvent non élucidée), qu'au niveau génomique (quantité de données génomiques insuffisante pour réaliser des études comparatives à grande échelle). Cette situation commence à changer grâce l'accès aux données de génomique et de protéomique (cf. ci-dessus). Ceci ouvre pour la première fois la possibilité de réaliser des études comparatives des SME permettant de proposer des hypothèses sur leur origine et leur évolution au sein des Eucarya, et par extension, de nous renseigner sur l'évolution de ce groupe.

SME	Taille du protéome	Référence / source
Appareil méiotique	~250	(Kumar et al., 2002)
Appareil mitotique	305	(Sauer et al., 2005)
Anaphase Promoting Complexe / Cyclosome	18	(Peters, 2006)
Centromères	~40	(Obuse et al., 2004)
Complexes ARN messenger	76	(Angenstein et al., 2005)
Complexes modification de la chromatine	18	(Lee et al., 2004)
Complexes télomériques	~25	http://www-ermm.cbcu.cam.ac.uk/02004556a.pdf
Corps intermédiaire de la division cellulaire ('midbody')	172	(Skop et al., 2004)
Flagelle / corps basal	>400	(Li et al., 2004; Pazour, 2004)

Membrane nucléaire	148	(Dreger et al., 2001)
Nucléole	~350	(Andersen et al., 2002; Scherl et al., 2002)
Pore nucléaire	68	(Rout et al., 2000)
RISC / Editosome	>15	(Sontheimer, 2005; Tang, 2005)
Spliceosome	~300	(Jurica and Moore, 2003; Rappsilber et al., 2002; Zhou et al., 2002)
Structure chromosomique ('scaffold')	79	(Gassmann et al., 2005)

Tableau 1. Protéomes de quelques structures multiprotéiques eucaryotes.

5.1.4 LECA, une cellule très complexe

Bien qu'encore rares, les études de certaines SME (dont plusieurs sont des analyses phylogénomiques) ont permis d'esquisser un portrait de LECA apparemment très complexe (Figure 35). Ci-dessous sont listées quelques caractéristiques que cet organisme possédait vraisemblablement.

5.1.4.1 Caractéristiques génomiques

- Les introns (Dacks and Doolittle, 2001)
- Les chromosomes linéaires (Aravind et al., 2006)
- Les télomères et télomérases (Aravind et al., 2006)

5.1.4.2 Processus informationnels

- Une machinerie ARN complexe (splicing majeur et mineur, splicing alternatif, machinerie RNAi, RNase MRP) (Collins and Chen, 2009)
- Le système de terminaison de la transcription (Cavalier-Smith, 1981)
- Le système ubiquitine (Hughes and Rusten, 2007)

5.1.4.3 Le métabolisme

- La voie de synthèse du shikimate (Richards et al., 2006)
- La voie de synthèse des stérols (Desmond and Gribaldo, 2009)
- La voie de synthèse du mévalonate (Lombard and Moreira, 2011)

5.1.4.4 Compartiments cellulaires

- Le noyau, la membrane nucléaire et les pores nucléaires (Bapteste et al., 2005; Mans et al., 2004)
- Le nucléole (Cavalier-Smith, 2002)
- L'appareil Golgi (Dacks et al., 2003)
- Le flagelle (Jekely and Arendt, 2006; Mitchell, 2007)
- Le peroxyosome (Gabaldon et al., 2006)
- La mitochondrie (Embley, 2006)
- Le système endomembranaire (Dacks et al., 2009)

5.1.4.5 Processus cellulaires

- La phagotrophie (Cavalier-Smith, 2002)
- L'autophagie (Hughes and Rusten, 2007)
- Le système de trafic membranaire (Dacks and Field, 2007; Koumandou et al., 2007; Leung et al., 2008)
- La formation de pseudopodes (Richards and Cavalier-Smith, 2005)
- La méiose (Ramesh et al., 2005)

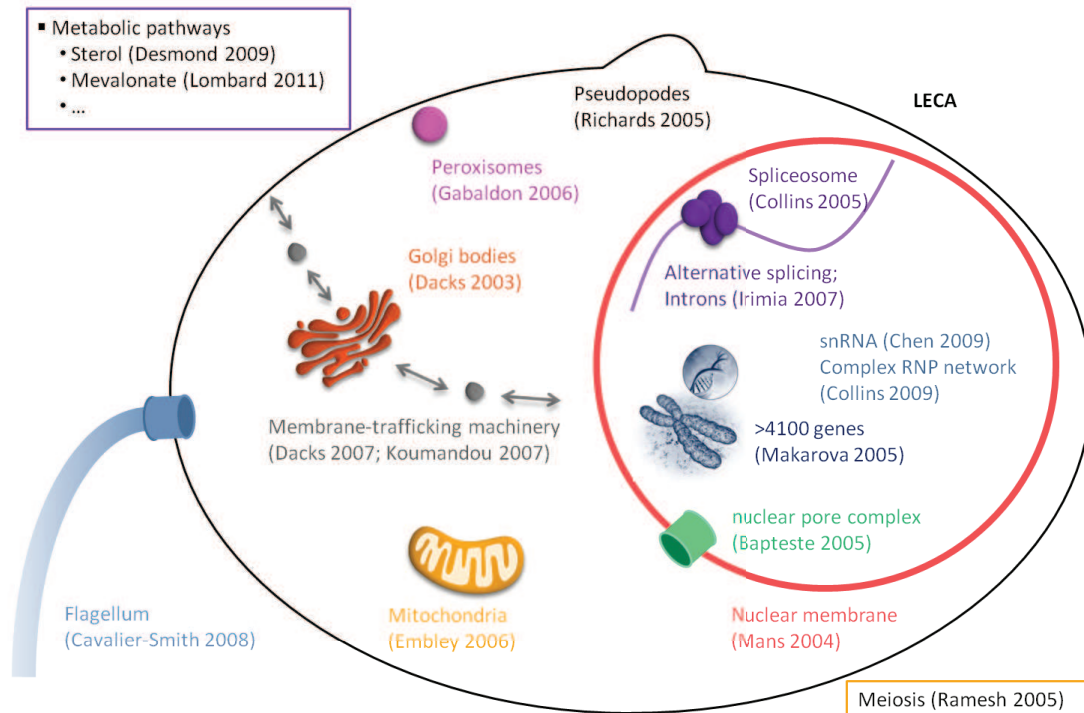


Figure 35. Représentation schématique de LECA et de quelques une de ses caractéristiques.

L'idée d'un LECA avec des structures et processus intracellulaires complexes est en accord avec l'estimation du nombre important de gènes présents chez cet organisme (plus de 4100, d'après l'analyse de Makarova *et al.* (Makarova et al., 2005)).

5.2 Objet de cette thèse

C'est dans ce contexte que s'est inscrit mon projet de thèse. Il avait pour objectif l'étude par des méthodes de phylogénomique de certaines SME ayant contribué à l'émergence et l'évolution des Eucarya. Mon travail s'est articulé autour de deux axes principaux.

Une partie orientée "phylogénomique" à proprement parler, consistant en l'étude de plusieurs SME eucaryotes par une approche de phylogénomique intégrée. Je me suis tout particulièrement intéressée à des structures/systèmes impliqués dans certaines étapes du cycle cellulaire, puisque jusque-là, aucune étude à grande échelle évolutive n'avait été menée sur ces structures. En effet, bien que très étudié, le cycle cellulaire reste encore mal caractérisé dans de nombreuses lignées eucaryotes, et son déroulement chez LECA était totalement inconnu. Ces analyses avaient donc pour objectif d'améliorer nos connaissances sur l'origine et l'évolution de certaines parties du cycle cellulaire eucaryote depuis LECA jusqu'aux lignées actuelles.

- Dans un premier temps, je me suis intéressée à l'évolution de diverses structures eucaryotes impliquées dans la cytokinèse (*i.e.* dernière étape de la division cellulaire, qui correspond à la séparation des cellules-filles). Les résultats sont présentés dans le chapitre p.65. Ils ont donné lieu à la publication dans Plos One, jointe en Annexe 1.
- J'ai ensuite réalisée l'étude phylogénomique de l'Anaphase Promoting Complex (ou cyclosome), une ubiquitine ligase E3 responsable de l'adressage de certaines protéines-clés du cycle cellulaire vers le protéasome 26S afin qu'elles y soient dégradées. Cette analyse est présentée au chapitre p.85. Elle fait l'objet d'une publication en cours de soumission (jointe en Annexe 2).

Un des intérêts de l'étude de chaque SME est qu'en tant que complexe multiprotéique et entité fonctionnelle, il est attendu que certains composants aient co-évolué. De plus, dans le cas des SME anciennes et conservées, leurs composants peuvent porter des traces de l'histoire évolutive des organismes qui les portent. Il est alors intéressant d'analyser le signal phylogénétique porté par ces composants et d'étudier dans quelle mesure ils peuvent nous renseigner sur l'histoire évolutive des eucaryotes. Dans une deuxième partie, j'ai donc réalisé un travail axé sur la phylogénie des eucaryotes, en utilisant comme marqueurs, les composants issus des SME analysées précédemment. Ces résultats sont reportés au chapitre p.143.

Au cours de ma thèse, j'ai également eu l'opportunité de participer à l'étude d'un système enzymatique de type eucaryote impliqué dans la réponse au stress oxydant chez les cyanobactéries, dont les résultats sont présentés au chapitre p.113. Ce travail a fait l'objet d'une publication en cours de révision dans *New Phytologist* (Annexe 3). *Ce travail a été réalisé en collaboration avec Amel Latifi et Cheng-Cai Zhang de l'équipe "Différenciation cellulaire microbienne et signalisation" (Laboratoire de Chimie Bactérienne, CNRS, Marseille).*

Enfin, j'ai eu l'opportunité de participer à un travail sur la phylogénie des Archaea, basé sur de nouvelles données de métagénomique d'archées hyperthermophiles, présenté au chapitre p.179. Ce travail fait l'objet d'une publication en cours de rédaction qui sera soumise dans *Environmental Microbiology*. *Ce travail a été réalisé en collaboration avec Laila Reigstad, du Centre de Géobiologie de l'université de Bergen (Norvège), ainsi que Christa Schleper et Anja Spang du Département de Génétique et Ecologie de l'Université de Vienne (Autriche).*

AXE 1 :

**EVOLUTION DES SYSTEMES
BIOLOGIQUES EUCARYOTES**

J'ai réalisé l'analyse phylogénomique de deux systèmes multiprotéiques eucaryotes ayant un rôle dans la division cellulaire. La première analyse concerne le midbody, structure impliquée dans la cytotinèse, dont l'analyse est présentée au début de ce chapitre. La seconde analyse que j'ai réalisée concerne l'Anaphase Promoting Complex, une ubiquitine-ligase entraînant la dégradation de protéines-clés du cycle cellulaire, dont les résultats sont présentés en deuxième partie de ce chapitre. Enfin, bien qu'il ne s'agisse pas à proprement parler de l'étude d'une structure multiprotéique eucaryote, j'ai eu l'opportunité de participer à l'analyse d'un système enzymatique impliqué dans la réponse au stress oxydant dont l'histoire évolutive atypique suggère des transferts horizontaux entre le dernier ancêtre commun des cyanobactéries et celui des Eucarya. Ces résultats seront présentés en fin de partie.

Pour l'étude de ces systèmes, j'ai utilisé une approche de phylogénomique intégrée. Le principe de cette approche peut paraître simple à première vue mais l'expérience a montré qu'il est en réalité difficile à mettre en œuvre, et demande de nombreuses analyses préliminaires que je détaille ci-après.

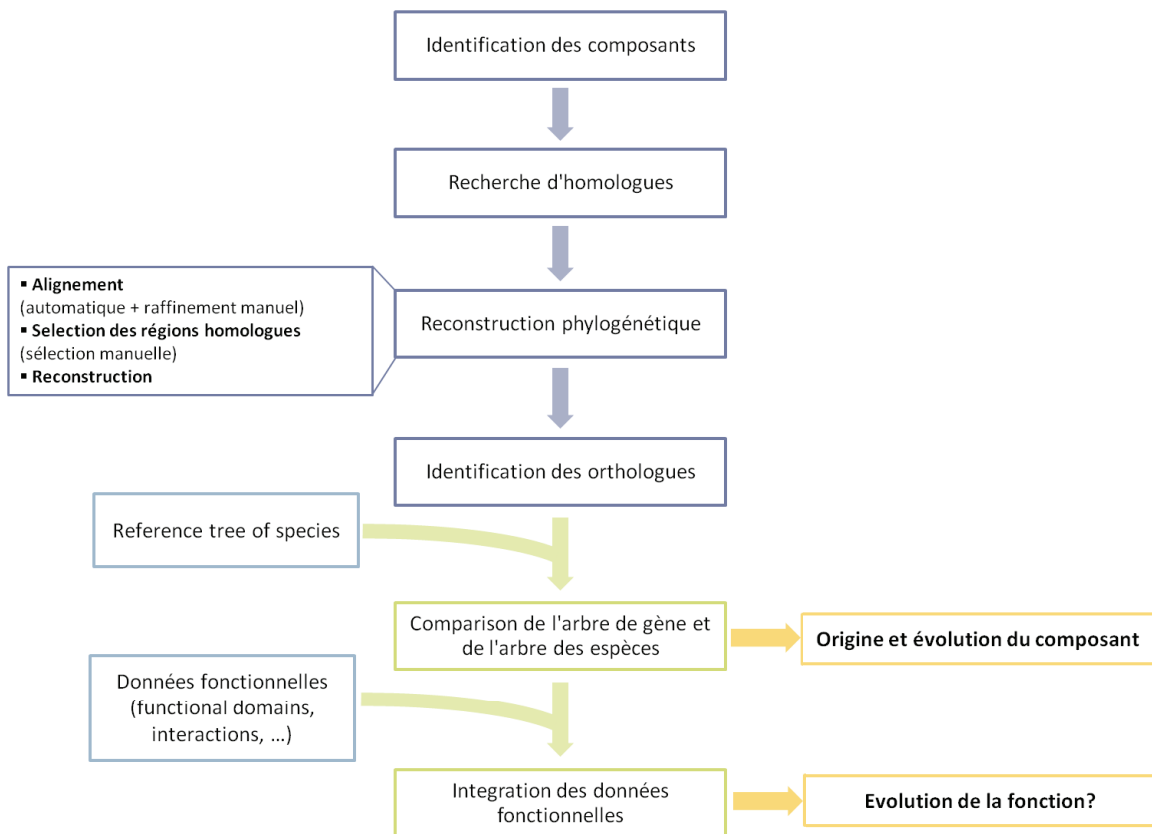


Figure 36. Principe de l'étude phylogénomique des systèmes biologiques.

RECHERCHE DES SEQUENCES HOMOLOGUES

Lors de l'analyse de ces deux SME, j'ai recherché, les séquences homologues de chaque composant par Blast dans les banques de données publiques (NCBI, JGI, TIGR ...). Cette étape doit être réalisée en plusieurs fois afin d'être certain de récupérer l'ensemble des homologues. En effet, il arrive fréquemment que les homologues soient divergents (c'est-à-dire présentant un faible degré de similarité). Il est donc nécessaire d'utiliser des séquences requêtes (ou "graines") adaptées aux lignées ciblées (par exemple, partir d'une séquence d'alvéolé afin de détecter des séquences d'hétérokotes qui ne l'auraient pas été en partant d'une séquence animale). Il peut également être nécessaire de réaliser des Psi-Blast (ou blast itératifs) qui permettent de trouver des homologues lointains par construction d'un profil de manière itérative.

Parmi les résultats de Blast, il faut ensuite identifier les séquences qui correspondent réellement à des homologues. Cette étape délicate est indispensable et ne peut pas être réalisée de façon automatique. Elle repose sur la prise en compte de nombreux critères : (i) le score et la e-value associés au hit; (ii) le pourcentage d'identité (résidus identiques) et similarité (résidus identiques + substitutions conservatives d'après la matrice de substitutions) entre les séquences; (iii) la longueur de l'alignement comparé à la longueur des séquences; (iv) la longueur de la séquence cible par rapport à la séquence requête; (v) la présence éventuelle de motifs fonctionnels, etc.

CONSTRUCTIONS DES JEUX DE DONNEES

Dans un premier temps, les séquences ont été alignées de façon automatique avec les logiciels ClustalW (Thompson et al., 2002), Muscle (Edgar, 2004) ou encore MAFFT (Katoh et al., 2002). Lorsque les jeux de données étaient suffisamment petits (≤ 50 séquences), nous avons également utilisé T-coffee (Notredame et al., 2000), qui donne généralement de bons résultats quand les alignements contiennent peu de séquences. Ces quatre algorithmes ont été utilisés avec les paramètres par défaut. En fonction de la qualité de l'alignement obtenue pour chaque jeu de données, nous avons utilisé l'un ou l'autre des alignements pour la suite de l'analyse.

Le logiciel ED fait partie du package MUST (un package de logiciels dédiés à la phylogénie) (Philippe, 1993), et permet de visualiser et d'éditer un ensemble de séquences alignées et de consulter les informations associées à chaque séquence si celle-ci provient d'une fiche Genpept. L'accès aux annotations est important pour la prise de décision lors du nettoyage des jeux de données. ED permet également de générer à tout moment un arbre par Neighbor-Joining. Cet arbre est visualisable par le logiciel TREEPLOT, qui fait aussi partie du package. Cela permet des analyses phylogénétiques préliminaires rapides et l'identification des relations de parenté entre séquences, des orthologues et paralogues, des séquences très divergentes, etc., qui peuvent aider à l'expertise de l'alignement.

La vérification de la qualité de l'alignement consiste à :

- "Nettoyer" les jeux de données en éliminant les séquences manifestement non homologues, redondantes, ou correspondant à des transcrits alternatifs ;
- Mettre en évidence l'absence de séquences dans certaines lignées. Ces absences sont systématiquement vérifiées par de nouvelles recherches dans les banques de données afin de compléter les jeux de données ;
- Enfin, il est également indispensable de vérifier manuellement l'alignement et de le corriger si cela s'avère nécessaire.

L'expertise et l'amélioration des jeux de données est l'étape la plus longue et la plus fastidieuse de ce type d'analyse, mais elle est absolument indispensable, et ne peut pas être complètement automatisée lorsqu'on veut favoriser l'aspect qualitatif des analyses par

rapport à l'aspect quantitatif. En particulier, les deux premiers points sont cruciaux, puisque malgré le soin apporté à la constitution initiale des jeux de données, je n'ai jamais obtenu des jeux de données "irréprochables" en une seule étape.

RECONSTRUCTIONS PHYLOGENETIQUES

La qualité des inférences phylogénétiques est intimement liée à celle des alignements multiples sur lesquelles elles sont fondées. Pour une analyse phylogénétique, chaque colonne (*i.e.* site) de l'alignement doit contenir des acides aminés homologues. Or, il existe presque toujours des régions de l'alignement multiple où l'homologie entre les résidus n'est pas certaine et qui doivent être omises des analyses phylogénétiques. Grâce au logiciel NET (également disponible dans le package MUST), j'ai sélectionné dans les alignements les zones à prendre en compte pour la suite des analyses. En tant que pré-filtre, il est possible d'utiliser une fonction du logiciel qui permet d'éliminer automatiquement les régions contenant plus de 5% de gaps (zones d'insertions/délétions), où l'alignement est souvent ambigu. L'observation des zones restantes permet d'identifier de façon plus rigoureuse les zones dont l'homologie est non ambiguë. Une fois ces zones choisies, NET permet de générer les fichiers utilisables par différents logiciels de reconstruction phylogénétique.

J'ai utilisé plusieurs méthodes de reconstruction d'arbres : des algorithmes basés sur le maximum de vraisemblance (PhyML (Guindon and Gascuel, 2003) et Treefinder (Jobb et al., 2004)), ainsi que des méthodes bayésiennes (MrBayes (Ronquist and Huelsenbeck, 2003) et Phylobayes (Lartillot et al., 2009)).

Les modèles d'évolution ont été sélectionnés différemment selon les analyses afin d'être le plus adapté possible aux données étudiées. J'ai notamment utilisé l'outil ProposeModel de Treefinder, qui permet de proposer automatiquement le modèle qui maximise la vraisemblance de l'arbre.

Une estimation de la robustesse des branches des arbres reconstruits en maximum de vraisemblance a été systématiquement effectuée par bootstrap (100 réplicats du jeu de données initial). Pour les arbres reconstruits en bayésien, le support est directement donné par les valeurs de probabilités postérieures.

La reconstruction des arbres phylogénétiques permet l'identification rigoureuse des orthologues et paralogues. En effet, pour pointer l'origine d'un gène, il est nécessaire de faire l'étude des orthologues. L'arbre permet également d'inférer l'existence, ou non, d'un homologue à chaque nœud de l'arbre des espèces.

INTEGRATION DES DONNEES FONCTIONNELLES

La présence d'un orthologue n'est pas suffisante pour prédire que la fonction existe (ou existait, s'il s'agit d'un organisme ancestral). Il est nécessaire de croiser cette donnée avec d'autres informations. Il peut s'agir, par exemple, de déterminer la composition en domaines fonctionnels de chacun des orthologues. Si celle-ci est strictement identique au sein de tous les orthologues, alors il est probable que leur fonction biochimique soit également identique, et qu'elle ait été héritée du composant ancestral. Au contraire, il est possible d'observer un changement de composition dans une partie de l'arbre, suggérant un changement de fonction dans les espèces concernées.

1

Evolution de la cytokinèse

1 Evolution de la cytokinèse

Ce travail a été réalisé en collaboration avec David Moreira du Laboratoire d'Ecologie, Systématique et Evolution, Université Paris-Sud, Orsay. Il a fait l'objet d'une publication disponible en Annexe 1.

1.1 Le midbody

Le midbody, ou "Mittelkoerper", a été décrit pour la première fois en 1891 par Walter Flemming (Flemming, 1891) comme un résidu de la mitose, auquel les cytologistes donnent souvent le nom de "corps de Flemming". Depuis, ce complexe n'a été caractérisé que chez quelques Eucarya, et plus précisément chez quelques organismes modèles animaux (*D. melanogaster* (Silverman-Gavrila and Wilde, 2006), *H. sapiens* (Zhu et al., 2005), *M. musculus* (Errico et al., 2004), etc.). Cette structure transitoire formant un simili-organite est impliquée dans le processus de cytokinèse, c'est-à-dire l'étape finale de la division cellulaire conduisant à la séparation des deux cellules filles (Figure 37).

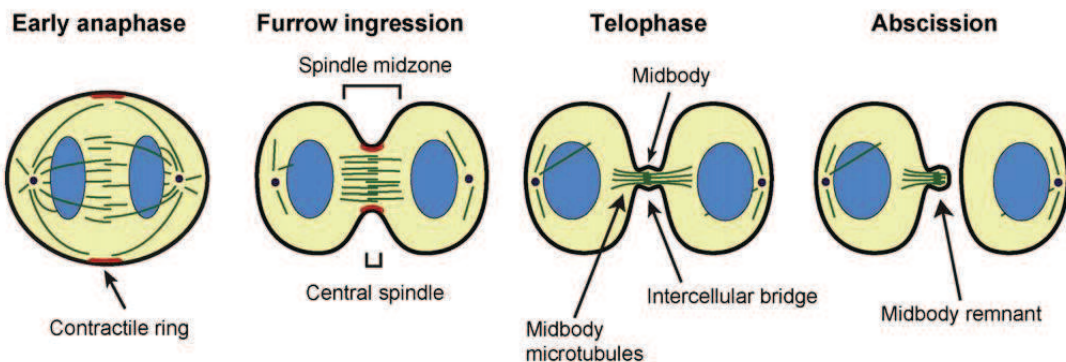


Figure 37. Les différents stades de la cytokinèse chez les animaux. Vert : microtubules, bleu : chromatine, noir : membrane plasmique; rouge : anneau contractile, violet : centrosomes. Depuis (Guizetti and Gerlich, 2010).

Le midbody est constitué du chevauchement des microtubules antiparallèles non kinétochoriens émanant des deux pôles, et d'un assemblage de nombreuses protéines formant une matrice compacte et dense. Il s'agit aussi bien de protéines de structure (s'associant aux microtubules ou à l'actine (Otegui et al., 2005)), que de protéines fonctionnelles (en particulier des kinésines, moteurs moléculaires indispensables à la constitution, au positionnement et au fonctionnement du midbody (Glotzer, 2001)). On y trouve également des vésicules dérivées du Golgi, potentiellement impliquées dans le transport de protéines et de membrane, ainsi que des protéines impliquées dans le trafic membranaire et les activités sécrétoires (Skop et al., 2004).

Les constituants du midbody se mettent en place dans une zone centrale et sont rapprochés physiquement par le resserrement de l'anneau contractile qui forme le sillon de division. Les cellules filles peuvent rester connectées pendant des heures par le pont intercellulaire que forme le midbody, ce qui permet le passage de molécules entre ces deux cellules (Eggert et al., 2006; Glotzer, 2001; Otegui et al., 2005) (Figure 38). Ce pont intercellulaire est apparemment essentiel aux étapes finales de la cytokinèse, car des protéines nécessaires à la scission des deux cellules-filles s'y trouvent localisées. Malgré l'importance du midbody, les mécanismes moléculaires conduisant à sa formation, à sa stabilisation et finalement à sa disparition, restent encore mal connus.

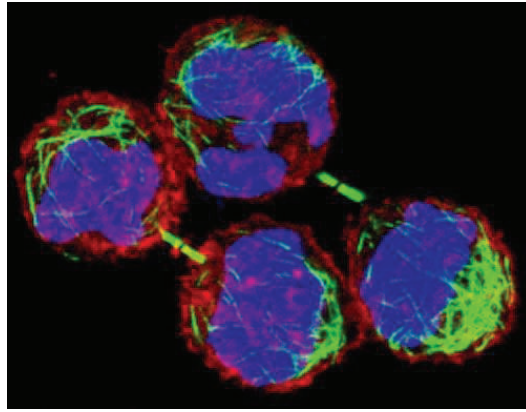


Figure 38. Cellules de Hamster à un stade tardif de division. La membrane plasmique (marquage rouge anti-actine) est presque totalement formée autour de chaque cellule, tandis que le midbody forme un pont intercellulaire entre les deux cellules (marquage vert anti-tubuline). L'ADN est coloré en bleu (DAPI). Image issue de (Skop et al., 2004)

Une étude, récemment publiée, a identifié le protéome du midbody des cellules de Hamster (Skop et al., 2004). Sur les 577 protéines identifiées par spectrométrie de masse, 417 ont été écartées car elles provenaient d'une possible contamination par des protéines mitochondriales, nucléaires ou ribosomiques. Les 160 protéines restantes ont alors été classées en cinq catégories fonctionnelles : "Protéines associées au trafic membranaire et activités sécrétoires" (53 protéines), "Protéines associées à l'actine" (46), "Protéines associées aux microtubules" (18), "Kinases" (18) et "Autres" (25) (Figure 39).

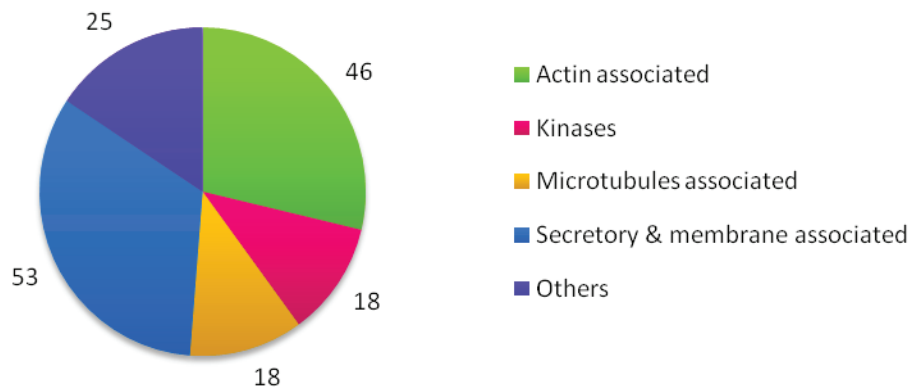


Figure 39. Classification fonctionnelle des 160 composants du midbody de Hamster. D'après (Skop et al., 2004)

Afin d'identifier le rôle de chacun des composants identifiés, Skop et collaborateurs ont inactivé fonctionnellement (par ARN interférant) les gènes homologues chez *C. elegans* et analysé les phénotypes observés au cours des premières divisions embryonnaires. Un résultat notable est que l'inactivation de près de 60% des gènes a entraîné des défauts de cytokinèse chez *C. elegans* (défauts mitotiques précoces ou tardifs, et/ou dans la lignée germinale). Cette combinaison d'approches de protéomique et génomique comparative a considérablement enrichi le répertoire connu des gènes indispensables à la cytokinèse : sur les 93 gènes dont la

mutation entraîne un phénotype anormal, seulement 16 étaient déjà connus chez *C. elegans*. Par ailleurs, 36% des 160 protéines étaient déjà connues pour leur rôle dans la cytokinèse dans d'autres systèmes modèles (*D. melanogaster*, *S. pombe*, *Dictyostelium discoideum*, mammifères, plantes, etc.), ce qui suggère une possible conservation fonctionnelle de ces gènes au sein de différents groupes eucaryotes. Enfin, bien que la fonction précise de chaque composant pendant la cytokinèse ne soit pas encore définie, ce type d'approche a permis de mettre en lumière des liens inattendus : par exemple, entre des protéines communes à la cytokinèse et aux fonctions neuronales (ex : KIF4). Il est également à noter que 14% (22/160) des protéines ainsi identifiées sont impliquées dans des maladies humaines (entre autres : surdit , scl roses, m lanomes ou leuc mies) (Skop et al., 2004).

1.2 La diversit  des m canismes de cytokin se chez les Eucarya

Jusqu'  pr sent, le midbody en tant que tel n'a  t  d crit que chez les m tazoaires.

Malgr  des similitudes qui peuvent exister au niveau m canistique, les eucaryotes pr sentent une grande vari t  de structures impliqu es dans les derni res  tapes de la division cellulaire (Figure 40).

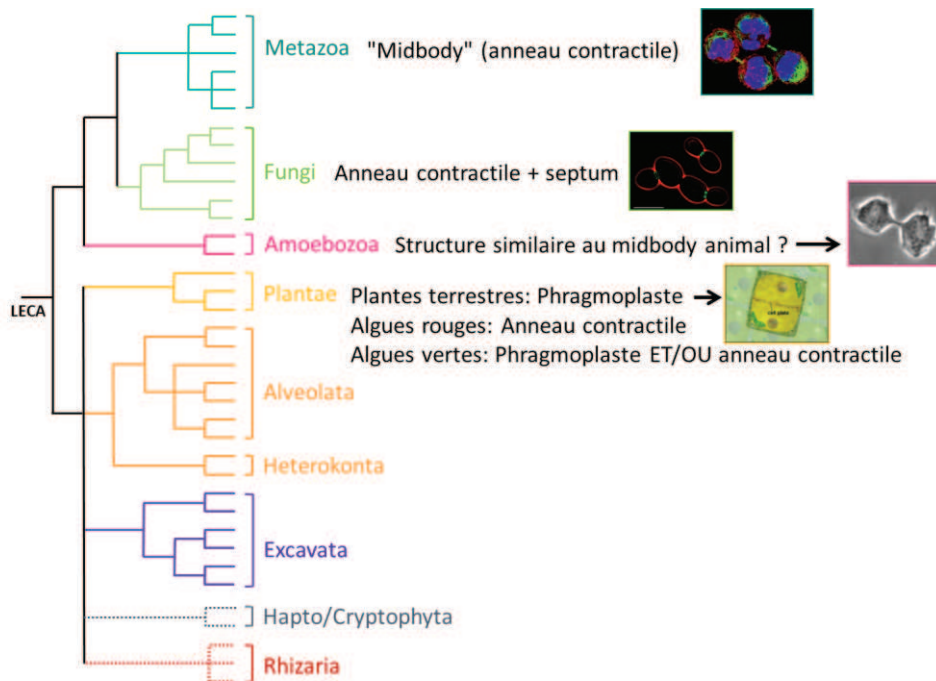


Figure 40. Diversit  des structures impliqu es dans la cytokin se chez les eucaryotes.

Par exemple, l'amibe *D. discoideum* cr e des ponts cytoplasmiques qui pr sentent certaines similitudes morphologiques avec le midbody animal (Neujahr et al., 1998) (Figure 41). Cependant, les donn es fonctionnelles sur les prot ines impliqu es dans ce m canisme sont quasiment inexistantes.

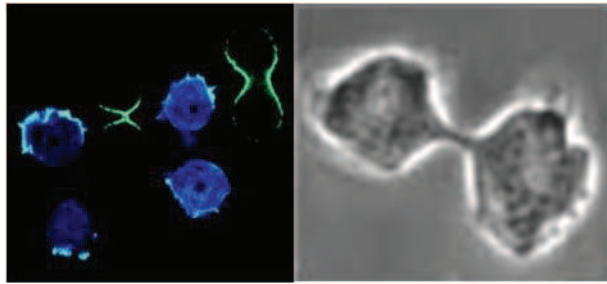


Figure 41. Cytocinèse chez *Dictyostelium discoideum*. A gauche : Le sillon de division est visible sur la deuxième et la quatrième cellule en partant de la gauche. (Arkowitz and Iglesias, 2008). A droite : le pont cytoplasmique est bien discernable. (Neujahr et al., 1998)

Les champignons présentent, tout comme les animaux, un sillon de division produit par le resserrement de l'anneau contractile d'actine, mais produit, en plus, une cloison riche en chitine au niveau du site de division (Walther and Wendland, 2003) (Figure 42). Dans les cellules de levure bourgeonnante (ex : *Saccharomyces cerevisiae*), l'anneau est positionné à l'interface entre la cellule mère et le bourgeon, alors que dans les cellules de levure à fission (ex : *Schizosaccharomyces pombe*), comme chez les animaux, l'anneau contractile est placé en position centrale, car ces deux types de cellules se divisent par fission médiane. Les deux types de levure synthétisent ensuite un septum de division au niveau de l'invagination créé par l'anneau de constriction. Enfin, le septum est dégradé, entraînant la séparation physique des cellules.

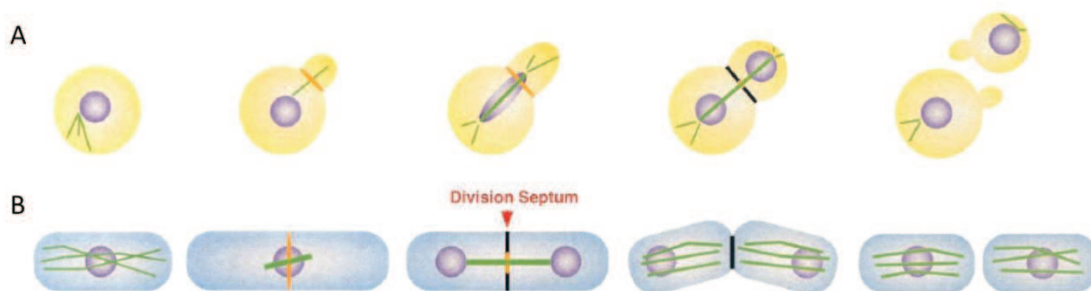


Figure 42. Les cellules de levures, tout comme les cellules animales, se divisent par l'utilisation d'un anneau contractile d'actomyosine. D'après (Guertin et al., 2002). (A) Levure bourgeonnante. (B) Levure à fission. Violet : chromatine; Vert : microtubules; Orange : anneau d'actomyosine.

La structure la plus distincte qui a été analysée en détail est le phragmoplaste présent chez les plantes terrestres. Il sert d'échafaudage pour l'assemblage de la plaque cellulaire et la formation d'une paroi séparant les deux cellules filles.

Le phragmoplaste est un assemblage complexe de microtubules, de microfilaments d'actine, et d'éléments du réticulum endoplasmique, qui s'assemblent, lors de l'anaphase et la télophase, en deux ensembles opposés perpendiculaires au plan de la future plaque cellulaire (Figure 43). Le phragmoplaste se forme à partir du fuseau mitotique entre les deux noyaux fils tandis que les enveloppes nucléaires se réassemblent autour d'eux. La plaque cellulaire prend initialement la forme d'un disque et est encadrée par

le phragmoplaste qui croit vers la périphérie. Les microtubules et les filaments d'actine du phragmoplaste servent à guider des vésicules dérivées du Golgi, et contiennent le matériel nécessaire à la croissance de la plaque cellulaire. Alors que du matériel est ajouté à la plaque cellulaire en croissance, les microtubules du phragmoplaste disparaissent au centre de la cellule et se régénèrent sur les bords de la plaque. Les deux structures se développent vers l'extérieur jusqu'à ce qu'ils atteignent la paroi de la cellule en division (Figure 43).

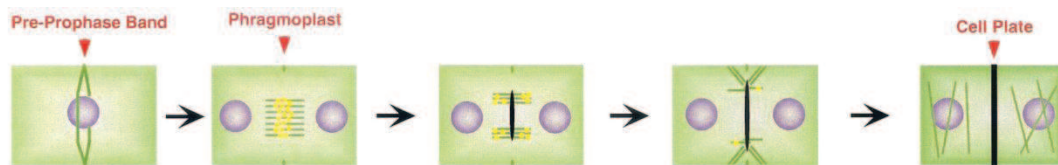


Figure 43. Les plantes terrestres, après séparation des noyaux, utilisent les microtubules pour délivrer des vésicules dérivées du Golgi à la région équatoriale. Les vésicules fusionnent pour former le phragmoplaste, qui, grâce à la fusion successive d'autres vésicules, se développe vers l'extérieur (vers le cortex cellulaire), pour former en fin de compte une barrière physique entre les cellules filles, appelée "plaque cellulaire". Mauve : chromatine; jaune : vésicules; vert : microtubules.

Ainsi, la division cellulaire ne fait pas intervenir d'anneau contractile d'actinomyosine, et, contrairement au midbody, le phragmoplaste croit de façon centrifuge. Néanmoins, le phragmoplaste et le midbody partagent quand même des caractéristiques communes, puisqu'ils contiennent tous les deux un certain nombre de protéines associées aux microtubules et des moteurs moléculaires de type kinésine qui stabilisent la structure (Jurgens, 2005).

Le phragmoplaste n'a été observé que chez les plantes terrestres (*i.e.* Embryophyte) et certaines espèces au sein de trois ordres d'algues (Figure 44) : les Coleochaetales (Marchant and Pickett-Heaps, 1973), les Charales (Pickett-Heaps, 1967) et les Zygnematales (Grolig, 1992). Il n'a, *a priori*, pas été documenté chez les Chlorokybales et les Klebsormidiales (Lopez-Bautista et al., 2003). Généralement, la division cellulaire médiée par le phragmoplaste n'est pas retrouvée dans les lignées appartenant aux Chlorophyta, à l'exception notable de deux espèces d'Ulvoophyceae (*Cephaleuros parasiticus* (Chapman and Henk, 1986) et *Trentepohlia odorata* (Chapman et al., 2001)).

Par ailleurs, diverses situations ont été décrites dans la lignée verte puisque, chez la plupart des autres algues vertes, la cytokinèse implique un sillon de division (Lopez-Bautista et al., 2003; Verma, 2001). Des situations intermédiaires ont également été observées, comme chez *Spirogyra* (une algue verte du sous-groupe des Zygnematales, Figure 44), dans lequel la cytokinèse est initiée par un sillon de division à base d'actine et complété par un assemblage de microtubules de type phragmoplaste. Toutefois, contrairement à ce qui se passe chez les plantes terrestres, cette structure "phragmoplaste-like" progresse vers le centre du plan de division cellulaire et non de façon centrifuge (Sawitzky and Grolig, 1995).

Fait intéressant, chez les algues rouges, qui sont des proches parents de la lignée verte, la présence d'un anneau contractile d'actine et d'un sillon de division a été observée (Broadwater, 1992), suggérant que le phragmoplaste représente en réalité une structure très dérivée.

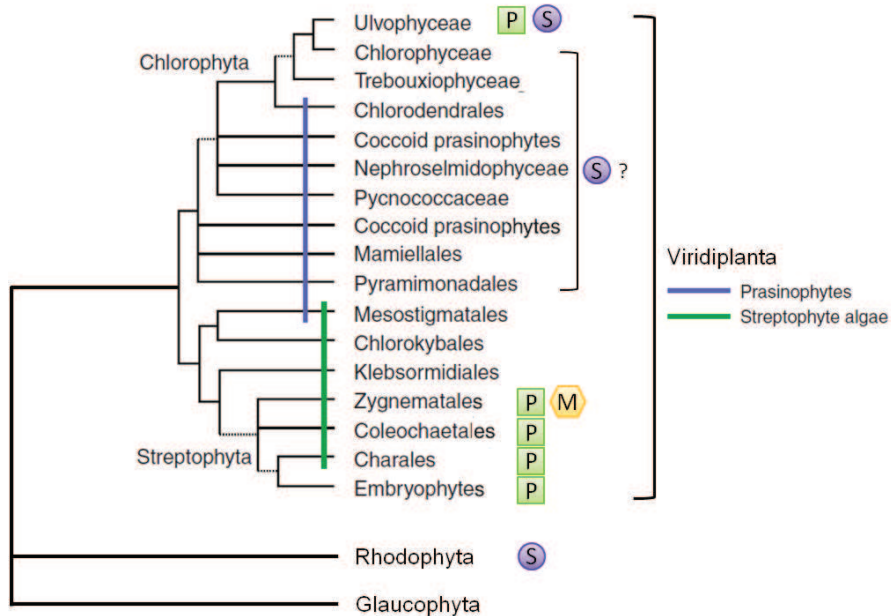


Figure 44. Phylogénie de référence des Plantae et structures impliquées dans la cytokinèse. Adapté d'après (Becker and Marin, 2009). Les carrés verts indiquent que certaines espèces de ce groupe ont été décrites comme se divisant par le moyen d'un phragmoplaste. Les cercles violets indiquent que certaines espèces de ce groupe ont été décrites comme se divisant par le moyen d'un sillon de division. L'hexagone jaune indique qu'une espèce de ce groupe a été décrite comme se divisant par une structure mixte. Le point d'interrogation indique que quelques espèces de ces groupes ont été *a priori* décrites comme se divisant par le biais d'un sillon, mais les données sont loin d'être exhaustives.

Enfin, très peu d'éléments sont connus sur le déroulement de la cytokinèse chez les autres eucaryotes, et en particulier chez les autres protistes.

1.3 Objectif

En résumé, il existe donc une diversité importante de structures impliquées dans la cytokinèse chez les eucaryotes, et dans le même temps une certaine similarité de composition entre des structures en apparence très différentes que sont le midbody animal et le phragmoplaste des plantes terrestres. La question s'est alors logiquement posée de savoir si ces structures sont homologues (descendant d'une structure ancestrale commune) ou analogues (ont des origines évolutives distinctes) (Jurgens, 2005; Otegui et al., 2005).

Pour répondre à cette question, j'ai entrepris l'analyse phylogénétique des composants de ces différentes structures. En effet, si elles ont des origines évolutives distinctes, il est attendu que peu d'orthologues soient partagés entre ces structures. Au contraire, si elles descendent d'une structure ancestrale commune, alors il est attendu de détecter de nombreux orthologues dans les différentes lignées eucaryotes.

La principale difficulté pour répondre à cette question vient de la partialité des données disponibles. En effet, la plupart des études réalisées sur ces structures, quand elles existent, concernent uniquement quelques composants étudiés par inactivation (ex : Centrin1 et FLA1 chez *Trypanosoma brucei*, ou GAPA chez *D. discoideum*) (LaCount et al., 2002; Mondal et al., 2010; Selvapandiyan et al., 2007).

Le midbody est la première (et la seule à ce jour) structure impliquée dans la cytokinèse dont la composition a été caractérisée (Skop et al., 2004). Les données expérimentales étant très rares dans les autres lignées, les approches *in silico* peuvent apporter certains éléments de réponse. Prenant alors comme point de départ les données de protéomique du midbody, j'ai réalisé l'analyse phylogénomique (*sensu* Eisen, (Eisen, 1998)) de ces composants.

Dans un premier temps, j'ai donc recherché les homologues de chacun des 160 composants, et cela dans 32 génomes représentatifs des grandes lignées eucaryotes dans les bases de données publiques (Figure 45). J'ai ensuite réalisé l'analyse phylogénétique des 160 jeux de données correspondants, et, à partir des arbres obtenus, j'ai discriminé les orthologues des paralogues. A cette étape, deux protéines ont été exclues (CDK4 et Ankyrin B/Ank2) car elles contenaient trop peu de signal phylogénétique pour permettre des inférences fiables. De plus, il s'est avéré que trois (dynamine 2, Dynamine-like I et Dynamine-like DLP1) et deux (calnexine et calréticuline) composants du midbody de hamster sont issus de duplications vertébrés-spécifiques. Au total, il restait donc 155 jeux de données pour la suite des analyses.

Sur la base des phylogénies, j'ai inféré le nombre d'orthologues présents dans chacun des ancêtres (*i.e.* à chaque nœud de l'arbre), en utilisant une approche de parcimonie. C'est-à-dire que si un composant est présent dans deux lignées, alors leur dernier ancêtre commun possédait déjà ce composant. Je me suis ainsi basée sur l'hypothèse forte, mais néanmoins souvent admise, que les transferts horizontaux sont rares entre eucaryotes.

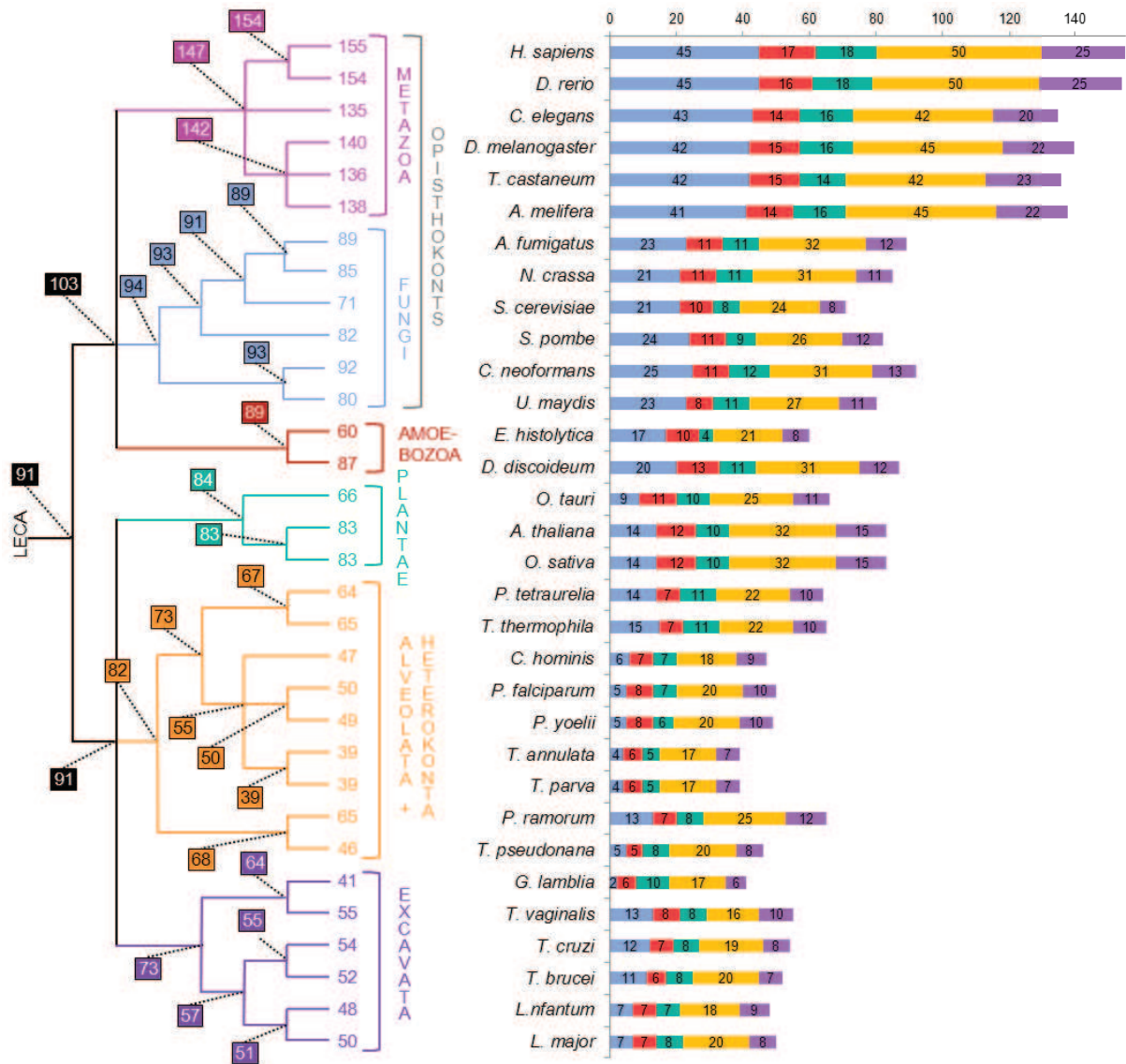


Figure 45. Répartition des orthologues dans les lignées eucaryotes. Le nombre d'orthologues des composants du midbody mammifère est indiqué pour chacune des 32 espèces eucaryotes étudiées. Les couleurs représentent les catégories fonctionnelles définies par Skop *et al* . Les relations phylogénétiques entre les lignées analysées sont indiquées sur la gauche. Les nombres aux nœuds indiquent le nombre d'orthologues inférés comme étant présents dans les ancêtres correspondants.

1.4 Un midbody chez l'ancêtre des métazoaires

Au sein des métazoaires, j'ai pu identifier de nombreux orthologues dans des espèces éloignées des mammifères, telles que *Drosophila melanogaster* (140 orthologues) et *Caenorhabditis elegans* (135 orthologues). Sur la base des phylogénies, j'ai pu inférer que la quasi-totalité des orthologues (147/155) était déjà présente chez l'ancêtre des métazoaires (Figure 45). En parallèle, les expériences d'inactivation systématique de ces orthologues chez *C. elegans*, ont montré que la majorité d'entre elles provoque des défauts de division cellulaire, suggérant une conservation fonctionnelle des composants entre les lignées du phylum animal. Il est donc vraisemblable que l'ancêtre de ce groupe possédait un midbody très similaire à celui existant aujourd'hui, et qu'il a été très conservé au cours de l'évolution chez les animaux.

1.5 Un midbody chez l'ancêtre des Unikontes ?

En ce qui concerne le reste des Unikontes, chez les Fungi et l'amoebozoa *Dictyostelium*, j'ai retrouvé un nombre d'orthologues à peu près similaire, c'est-à-dire proche de 90 (Figure 45), alors que chez l'autre amoebozoa *Entamoeba*, un nombre beaucoup plus restreint d'orthologues a été identifié (60). Deux explications sont possibles : soit il existe des pertes et/ou des remplacements non homologues, spécifiques à cet organisme, soit certains homologues existants chez *Entamoeba* n'ont pas été détectés. En effet, l'analyse des alignements a montré que les séquences d'*Entamoeba* étaient souvent très divergentes, ce qui pourrait expliquer qu'un biais de détection s'ajouterait à d'éventuelles différences biologiques.

Au sein des Fungi, il faut noter des pertes récentes assez nombreuses chez *Saccharomyces cerevisiae* qui ne possède que 73 orthologues contre 82 à 89 chez les autres ascomycètes. Cette constatation est importante car de nombreuses généralisations sont faites en biologie cellulaire à partir des observations réalisées chez cet organisme considéré comme modèle.

Il est intéressant de remarquer l'existence de 12 composants Unikontes-spécifiques (Figure 45). Ce résultat constitue une signature intéressante en faveur de l'existence de ce groupe qui est encore discuté d'un point de vue évolutif (Baptiste et al., 2002). A cela s'ajoutent les très nombreuses phylogénies basées sur les analyses individuelles des composants qui semblent soutenir la monophylie de ce groupe.

1.6 Un midbody chez LECA ?

1.6.1 Les orthologues chez les Bikontes

Chez les autres eucaryotes, il est intéressant de remarquer que le nombre d'orthologues détectés varie beaucoup même entre espèces proches (ex : 39 orthologues détectés chez l'apicomplexe *Theileria* et 50 chez *Plasmodium falciparum*, un autre apicomplexe), suggérant des pertes assez récentes. De même, ce nombre n'est pas directement corrélé à la distance évolutive entre les animaux et les autres lignées. En effet, les lignées de plantes terrestres possèdent en moyenne 83 orthologues, tout comme les Fungi (Figure 45), qui sont pourtant beaucoup plus proches des animaux.

Il n'existe pas non plus de corrélation entre le nombre d'homologues détectés dans les différentes lignées et la taille du génome des organismes. De façon générale, les animaux ont un génome de grande taille. Au contraire, *Giardia* et *Trichomonas*, ont de très petits génomes d'environ 10Mb. Or, pour ces deux organismes, n'a été détecté qu'un très petit nombre de composants, ce qui pourrait appuyer cette hypothèse. Cependant, chez *Entamoeba*, qui possède un plus gros génome (~20Mb), on observe également un plus petit nombre

d'homologues que dans les génomes de champignons ayant une taille similaire. On ne peut donc pas expliquer la totalité des différences de taille entre protéomes uniquement par des différences de taille de génomes.

Il est intéressant de noter des pertes extrêmement massives de composants appartenant à la catégorie des "actin-related" dans plusieurs lignées non apparentées, comme les Apicomplexa ou l'excavé *Giardia lamblia* (avec seulement 4 à 6 composants et 2 composants détectés respectivement, parmi les 45 de cette catégorie) (Figure 45). Cela peut s'expliquer par un rôle moins important de l'actine lors de la cytokinèse chez ces eucaryotes. En effet, chez les Apicomplexa, il a été montré que ce sont les microtubules et non les filaments d'actine qui sont responsables du bourgeonnement et de la division cellulaire (Striepen et al., 2007). En parallèle, chez *Giardia*, j'ai pu observer une accélération remarquable du taux d'évolution de sa séquence d'actine (Figure 46). Ces modifications concernant l'actine ont pu conduire à une convergence évolutive dans ces lignées par la perte (ou le remplacement) de la plupart des protéines associées à l'actine.

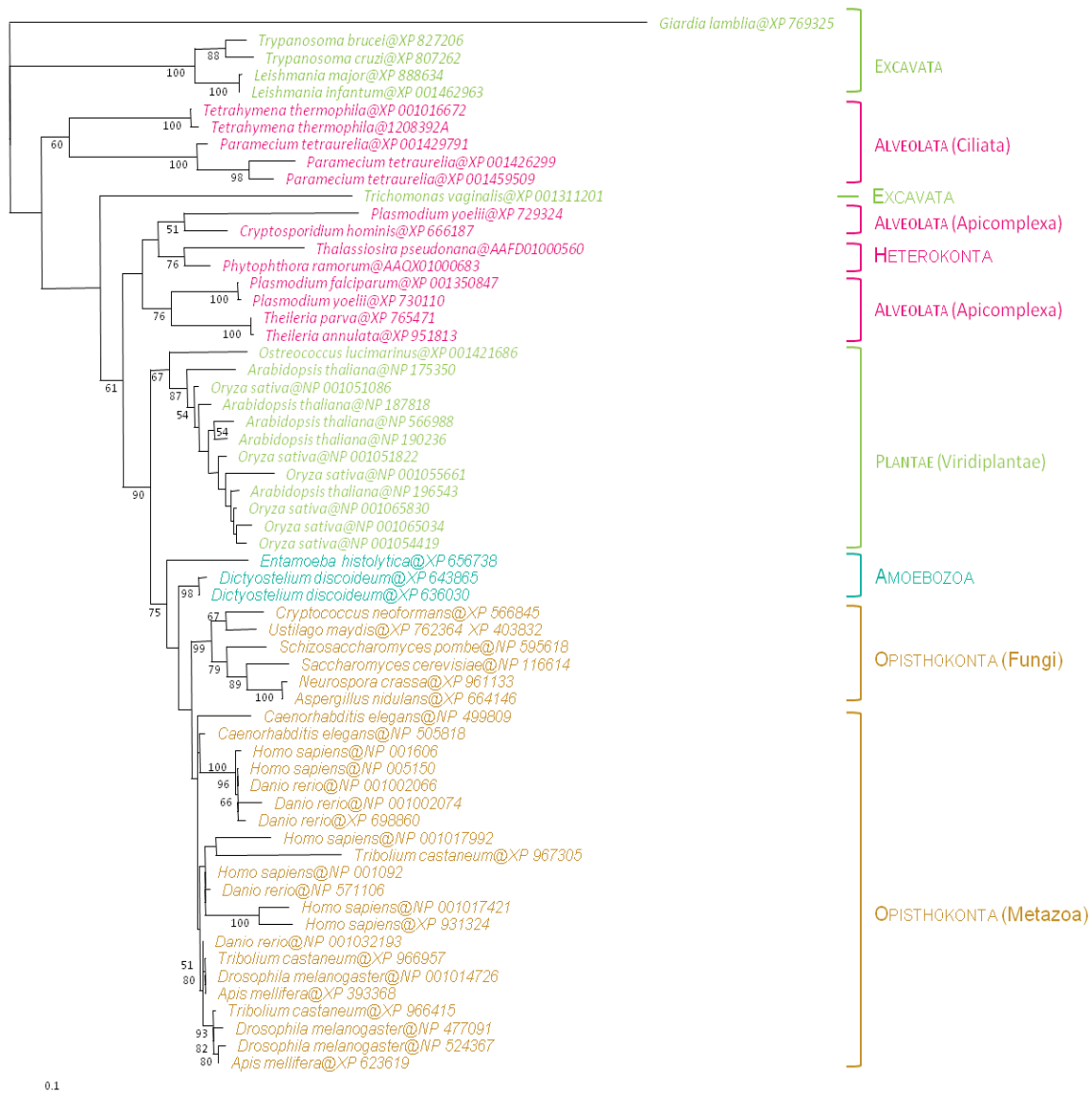


Figure 46. Phylogénie de l'actine (71 séquences, 360 positions analysées). Arbre en Maximum de Vraisemblance (PhyML) des homologues de l'actine présents dans les 32 lignées eucaryotes étudiées.

Les nombres aux nœuds représentent les valeurs de bootstrap (seules les valeurs >50% sont reportées). La barre d'échelle représente le nombre moyen de substitutions par site.

1.6.2 La majorité des composants existaient chez LECA

Sur la base des phylogénies, j'ai inféré que 91 composants du midbody étaient déjà présents chez LECA, c'est-à-dire qu'environ 60% des composants auraient une origine évolutive antérieure à la diversification des eucaryotes actuels. Cela permet de mettre en lumière une origine mixte du midbody animal, à savoir une majorité de composants "anciens", déjà présents chez LECA, et d'autres composants, plus récents, apparus dans la branche menant aux animaux (certains pouvant représenter des pertes ayant eu lieu chez les Bikontes).

A cette étape, il est important de garder à l'esprit que nous avons positionné la racine de l'arbre des eucaryotes entre les Unikontes et les Bikontes (cf. p.47). S'il s'avère qu'elle se place en fait ailleurs, alors cela modifierait le nombre de composants inférés comme étant présents chez LECA. Par exemple, si la racine se positionne en réalité entre les Excavata et le reste des eucaryotes, alors le nombre de composants inférés chez l'ancêtre des eucaryotes passerait de 91 à 73.

Par ailleurs, pour réaliser les inférences du nombre de composants ancestraux, j'ai fait l'hypothèse que les THG entre eucaryotes sont rares. Sur la base de cette hypothèse, la répartition taxonomique de quelques protéines (*i.e.* une dizaine) ne peut alors s'expliquer que par un grand nombre de pertes convergentes dans les différentes lignées eucaryotes (par exemple, Figure 47).

Il est possible que ces quelques phylogénies reflètent, en réalité, des cas de THG entre eucaryotes. Si tel est le cas, alors notre approche a pu ainsi aboutir à une légère surestimation du nombre de composants qui existaient chez LECA d'une dizaine de composants. Augmenter l'échantillonnage taxonomique serait nécessaire pour trancher entre ces deux hypothèses (*i.e.* pertes convergentes ou THG).

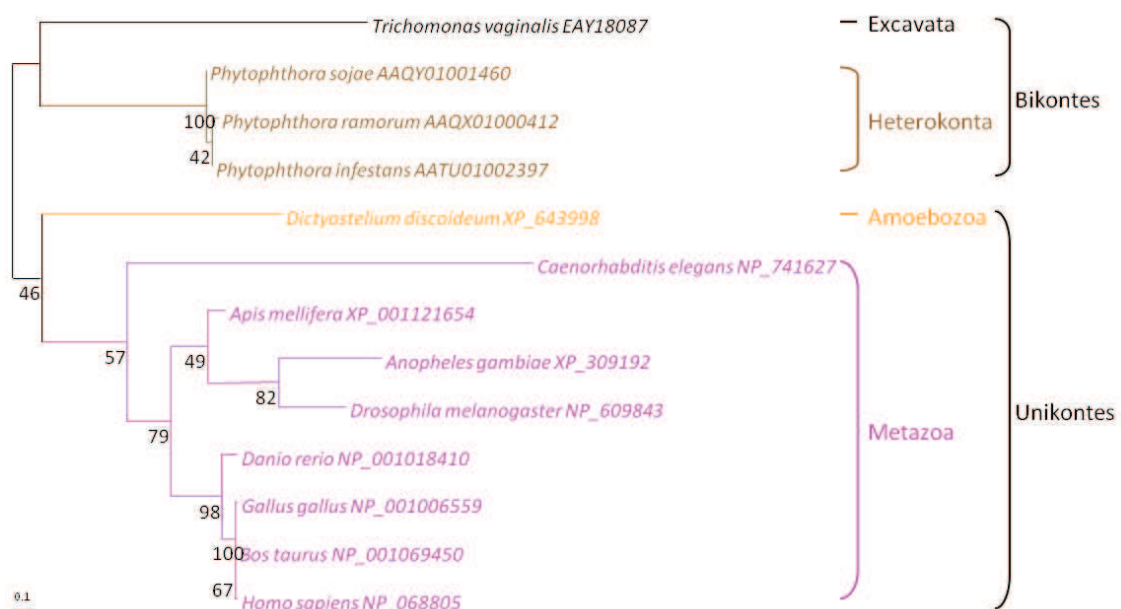


Figure 47. Arbre phylogénétique de la "MAP kinase scaffold protein 1", obtenu par Maximum de Vraisemblance (13 séquences, 111 positions analysées). La barre d'échelle représente le nombre moyen de changements par position. Les valeurs aux nœuds représentent les valeurs de bootstrap obtenues par PhyML.

1.6.3 Fonction des composants chez LECA

Concernant les orthologues présents chez LECA, deux hypothèses peuvent être émises : (i) ces composants existaient chez cet ancêtre mais avaient une fonction cellulaire différente, et ont été recrutés plus tard dans la lignée menant aux animaux pour former une structure de type midbody; (ii) ces composants avaient déjà une fonction cellulaire similaire, c'est-à-dire qu'ils étaient impliqués dans la fin de la division cellulaire. Pour trancher entre ces deux hypothèses, il a été nécessaire d'explorer les fonctions des orthologues dans les lignées non animales. Cependant, comme mentionné précédemment, les données expérimentales sont assez rares et j'ai donc utilisé des approches indirectes.

1.6.3.1 Distribution fonctionnelle des composants ancestraux

Les 91 composants présents chez LECA sont bien répartis dans les cinq catégories fonctionnelles définies par Skop *et al.* (Figure 48). Dans le cas contraire, si une ou plusieurs de ces catégories n'étaient pas du tout représentées chez cet ancêtre, il aurait été difficile d'imaginer l'existence d'une structure de fonction similaire à celle du midbody.

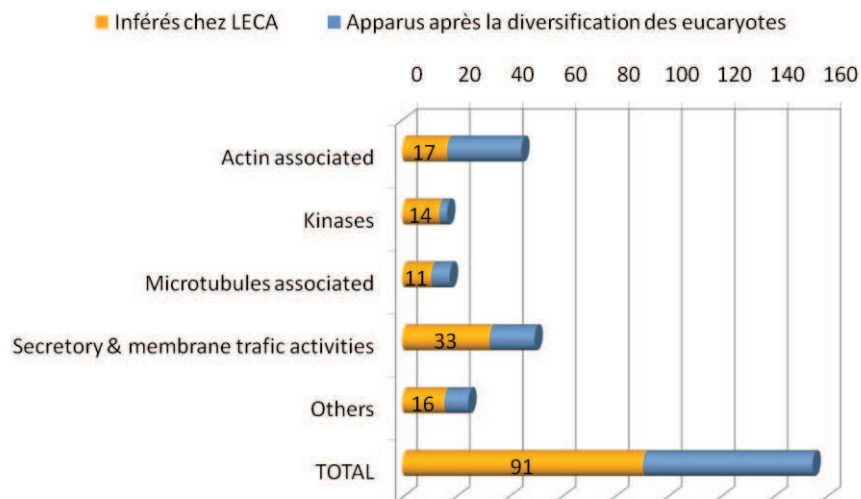


Figure 48. Origine des composants. Les composants inférés chez LECA sont représentés par les barres jaunes, tandis que ceux ayant une origine plus récente (c'est-à-dire dans la lignée menant de LECA aux mammifères) sont représentés par les barres bleues. Les nombres sont détaillés pour chacune des classes fonctionnelles définies par Skop *et al.*

1.6.3.2 Conservation de la composition en domaines fonctionnels des orthologues

Dans un premier temps, j'ai comparé la composition en domaines fonctionnels de chacun des orthologues à celle du composant de mammifère. Si les mêmes domaines sont détectés, et dans le même ordre, j'ai considéré que la composition était identique. Les résultats ont montré, qu'en moyenne, dans chacune des lignées eucaryotes étudiées, 80% des orthologues possèdent une composition en domaines fonctionnels identiques (Figure 49). Cela suggère une forte conservation de la fonction biochimique des orthologues au sein des diverses lignées eucaryotes. Néanmoins, cela n'indique pas nécessairement que ces composants sont impliqués dans des processus cellulaires identiques.

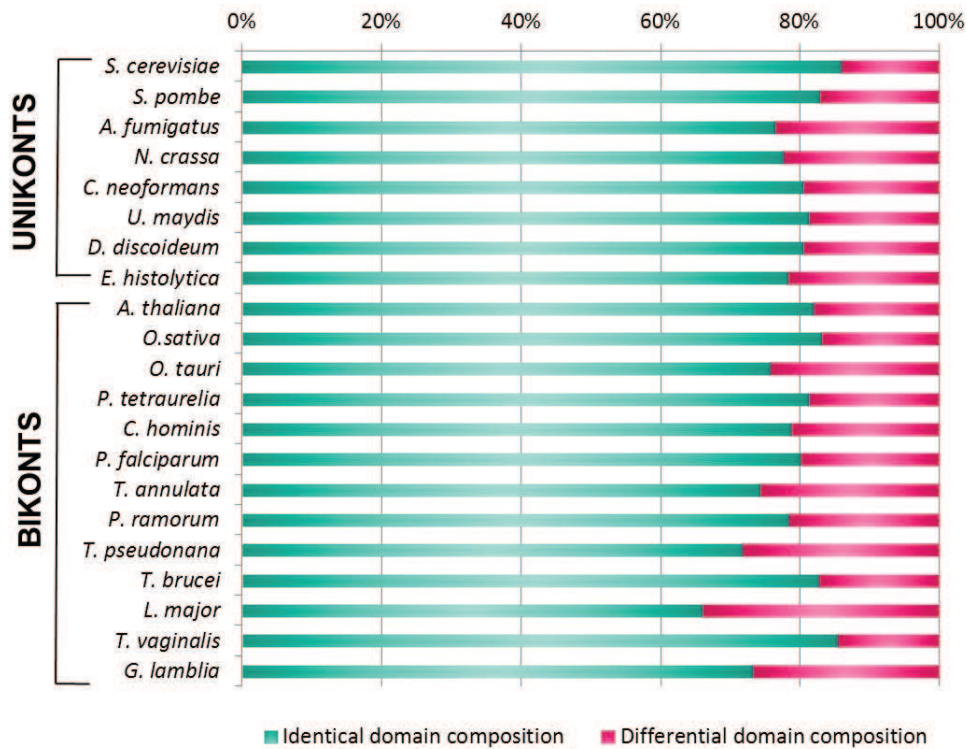


Figure 49. Conservation de la composition en domaines fonctionnels des orthologues du midbody dans les diverses lignées eucaryotes. Les orthologues de ces 21 espèces non-animales ayant une composition en domaines fonctionnels identique à celle de leurs homologues mammifères sont représentés par les barres vertes, tandis que ceux ayant une composition différente sont représentés par les barres magenta.

1.6.3.3 Fonction cellulaire des orthologues

Pour tester cette hypothèse, j'ai utilisé les données disponibles auprès de la base de données sur *Saccharomyces cerevisiae* (SGD, <http://www.yeastgenome.org/>). Cela m'a permis de voir que 31 des 71 orthologues de levure (44%) ont été caractérisés comme impliqués dans la cytotinèse ou la division cellulaire. Il est cependant nécessaire de noter que pas moins de 249 protéines sont annotées dans la SGD comme impliquées dans la division cellulaire. En parallèle, 55 de ces orthologues de levure (77%) ont été annotés comme interagissant avec au moins un des 71 autres. Ici, il faut rester prudent lors de l'utilisation des données d'interaction car il a été reporté qu'elles peuvent être surestimées selon la méthode utilisée (c'est le cas, par exemple, des méthodes de double-hybride).

Ces observations suggèrent néanmoins que, comme chez les métazoaires, la plupart des orthologues de levure interagissent ensemble et sont impliqués dans la division cellulaire, une caractéristique probablement héritée du dernier ancêtre commun des champignons et des animaux.

Parmi les orthologues mis en évidence chez les animaux, les champignons, et chez les plantes, huit ont été décrits dans la littérature comme participant aux structures impliquées dans la cytotinèse dans ces trois lignées (*i.e.* midbody, septum et phragmoplaste) (Otegui et al., 2005). Ce nombre est assez peu conséquent mais il faut garder à l'esprit qu'il ne s'agit pas de données exhaustives. Ces données sont compatibles avec l'hypothèse selon laquelle ces composants participent à des structures ayant une origine évolutive commune. En effet, si on fait l'hypothèse inverse, alors cela suppose que ces orthologues ont été recrutés de façon indépendante dans ces lignées pour participer à des processus cellulaires similaires.

Un lien évolutif supplémentaire entre les midbody et le phragmoplaste vient du fait que la croissance de ce dernier se fait principalement par la fusion des vésicules dérivées du Golgi (Verma, 2001). Or, Skop *et al.* ont montré que 38 composants du midbody mammifère sont des protéines associées au Golgi (Skop *et al.*, 2004).

1.7 Et chez les procaryotes ?

J'ai ensuite cherché à déterminer si les composants déjà présents chez LECA ont en réalité une origine encore plus ancienne en recherchant des homologues des différents composants chez les procaryotes. Des homologues procaryotes ont été détectés pour seulement neuf protéines. En outre, mes analyses phylogénétiques ont montré que pour quatre d'entre elles, il y a vraisemblablement eu un transfert horizontal entre eucaryotes et procaryotes (ex : Figure 50). Il serait intéressant de connaître le rôle de ces protéines chez ces derniers.

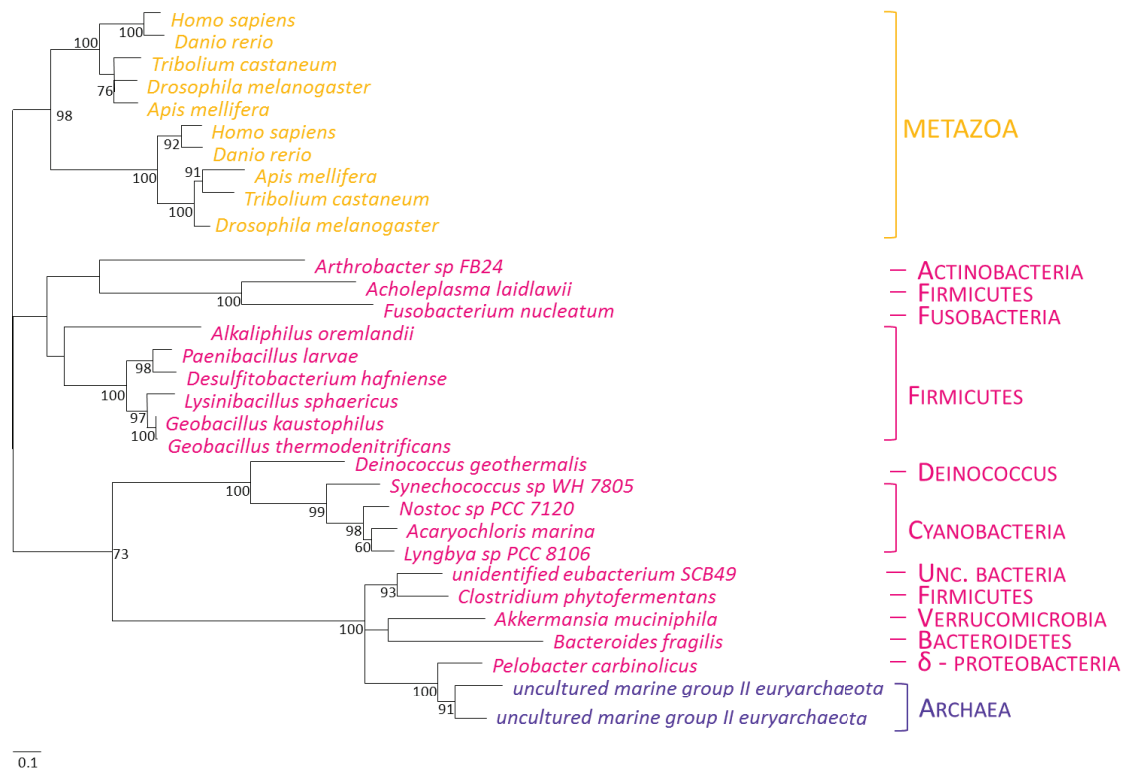


Figure 50. Arbre phylogénétique de la flotillin 1 obtenu en Maximum de Vraisemblance (33 séquences, 186 positions analysées). L'arbre a été reconstruit sur une sélection de séquences représentatives de la diversité des homologues. Cette protéine est l'un des quatre composants du midbody présent chez certains procaryotes mais clairement issu de THG entre les trois domaines du vivant. La barre d'échelle représente le nombre moyen de changements par position. Les nombres aux nœuds représentent les valeurs de bootstrap. Seules les valeurs supérieures à 50% ont été indiquées.

Ce petit nombre de composants partagés entre eucaryotes et procaryotes indique que presque toutes les protéines présentes chez LECA sont apparues après l'émergence de la lignée eucaryote. Pour autant, cela ne signifie pas que tous les composants résultent entièrement d'innovations eucaryotes (Figure 51). En effet, l'analyse de la composition en domaines protéiques des protéines du midbody a montré que seulement la moitié d'entre elles est composée de domaines qui sont spécifiques aux eucaryotes, (c'est-à-dire des domaines qui sont apparus après la divergence des eucaryotes). Ceci implique que la moitié seulement des protéines du midbody est effectivement issue d'innovations eucaryotes, tandis

que l'autre moitié est le résultat d'un assemblage de domaines plus anciens, parfois en association avec des domaines récents (*i.e.* eucaryotes-spécifiques).

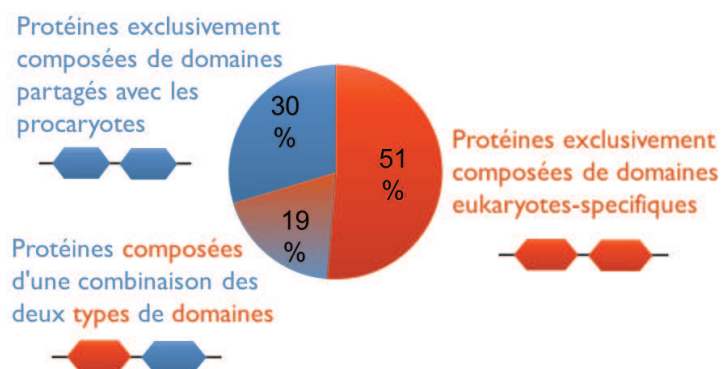


Figure 51. Composition en domaines fonctionnels des protéines du midbody. La section rouge représente les protéines composées uniquement de domaines fonctionnels eucaryotes-spécifiques (c'est-à-dire qu'ils n'ont pas d'homologues chez les procaryotes, ce qui signifie qu'ils sont probablement apparus après la divergence des eucaryotes). La section bleue représente les protéines exclusivement composées de domaines partagés par les procaryotes et les eucaryotes (c'est-à-dire suggérant que ces domaines peuvent avoir une origine plus ancienne, avant la divergence des eucaryotes). La section en dégradé de bleu et rouge représente les protéines composées par une combinaison des deux types de domaines.

Sur les 181 domaines présents dans les composants du midbody de Hamster, 60 ont été détectés dans des séquences procaryotes. Ces domaines sont globalement détectés dans de nombreux grands groupes procaryotes, et ne sont pas restreints à quelques groupes. Un tiers de ces domaines (21/60) sont retrouvés uniquement dans des séquences bactériennes, et les deux autres tiers (39/60) dans des séquences bactériennes et des séquences d'archées (Figure 52). Il est surprenant de noter qu'aucun domaine n'est exclusivement partagé par les archées et les eucaryotes, étant donné le lien évolutif spécifique unissant ces deux domaines (cf. p.16). Cependant, la plupart des hypothèses sur l'origine des eucaryotes postulent que le génome d'archée serait principalement à l'origine des gènes "informationnels", tandis que les gènes "opérationnels" seraient plutôt d'origine bactérienne. Il serait intéressant de comparer ces résultats avec ceux d'une étude de structure multiprotéiques eucaryotes (SME) liée à des processus informationnels, afin de voir si la tendance inverse est observée.

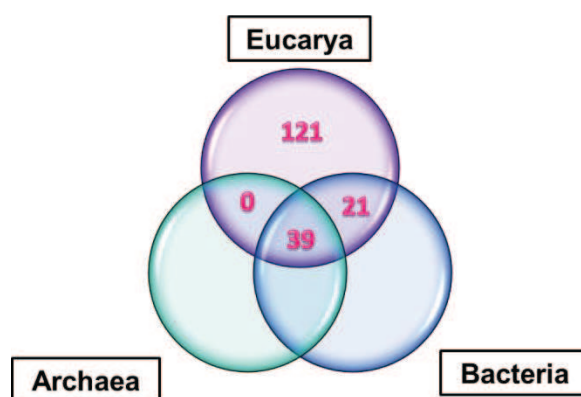


Figure 52. Schéma représentant le partage des domaines détectés au sein des composants du midbody de Hamster, entre les eucaryotes, les bactéries et les archées. Il est intéressant de remarquer qu'aucun domaine n'est spécifiquement partagé entre Eucarya et Archaea.

Un résultat intéressant est qu'il n'y a pas plus de protéines partageant des domaines avec les α -protéobactéries qu'avec d'autres groupes procaryotes, ce qui pourrait suggérer que la contribution de l'endosymbiose mitochondriale, sur la mise en place du midbody dans la

cellule eucaryote, est relativement faible. Les domaines partagés avec les bactéries sont donc probablement hérités de LUCA.

Dans la plupart des cas, je n'ai pas pu détecter par BLAST les séquences portant ces domaines, et donc étudier leur phylogénie. Il n'est donc pas possible de trancher entre l'hypothèse selon laquelle ces 60 domaines auraient été déjà présents chez LUCA, et celle d'une acquisition par THG.

1.8 Discussion et perspectives

En résumé, nos résultats confirment que :

- près de 60% des composants du midbody de mammifère étaient présents chez LECA ;
- Ceux-ci sont bien répartis dans les cinq catégories fonctionnelles ;
- En moyenne, 80% des orthologues identifiées dans les diverses lignées unikontes et bikontes affichent une forte conservation de la composition en domaines fonctionnels, indiquant des fonctions moléculaires similaires ;
- La plupart de ces orthologues interagissent dans la levure et sont vraisemblablement impliqué dans la cytocinèse ;
- Un certain nombre de ces orthologues font partie du phragmoplaste de façon certaine.

L'ensemble de ces résultats suggèrent que ces structures dérivent d'une structure bien plus ancienne datant de LECA. Cette analyse fournit donc des arguments supplémentaires en faveur de l'hypothèse d'un LECA "moderne", puisque qu'il possédait probablement un complexe multiprotéique de grande taille (comptant possiblement ~90 composants), issu d'une innovation eucaryote, impliqué dans les dernières étapes de la cytocinèse, et à partir duquel ont dérivé les différentes structures ayant ce rôle dans les eucaryotes actuels.

Nous avons également mis en évidence de nombreuses pertes récentes et différentes selon les lignées. Il est probable qu'en parallèle, au cours de l'évolution des eucaryotes, des recrutements de protéines aient eu lieu de façon spécifique dans ces différentes lignées. Ces gains potentiels et ces pertes pourraient expliquer la diversité observée chez les différents organismes au niveau des structures impliquées dans la cytocinèse.

Par ailleurs, nous avons observé de très nombreux cas de duplications ayant eu lieu dans différentes lignées selon le composant étudié (duplications vertébrés-spécifiques, métazoaires-spécifiques, Opisthokontes-spécifiques, quelques duplications Unikontes-spécifiques et des duplications Bikontes-spécifiques). Toutes ces duplications de types différents reflètent une plasticité importante existant entre les lignées. On sait que, chez les mammifères au moins, ces paralogues pourraient ne pas être impliqués dans le midbody (ils ne font pas partie du protéome du midbody de Hamster caractérisé par Skop *et al.*). Il serait utile d'avoir plus d'informations concernant leur rôle, pour mettre en évidence des éventuels changements de fonction. Ont-ils un rôle dans la cytocinèse d'autres tissus (chez les multicellulaires) ? A d'autres moments du développement ? Sont-ils impliqués dans des processus totalement différents de la cytocinèse ? De même, l'obtention de données de transcriptomique pourrait permettre d'affiner le rôle des différents paralogues.

Il paraît également nécessaire de disposer de données exhaustives de protéomique pour diverses lignées, et en particulier des bikontes, afin d'avoir plusieurs points de départ pour les analyses. Cela permettrait de réaliser une estimation plus précise du nombre de composants présents chez LECA, et également de mettre en évidence les gains ayant eu lieu dans ces différentes lignées, ainsi que les pertes dans la branche menant aux animaux.

Enfin, il est indispensable de confirmer les prédictions réalisées lors de nos analyses *in silico* par des données expérimentales. Par exemple, l'inactivation des orthologues que nous avons identifiés au cours de cette étude

permettrait de confirmer leur implication dans les dernières étapes de la division cellulaire, notamment chez des organismes distants. Un aspect important de ce type d'analyse est précisément l'identification de composants dans des lignées pour lesquelles les données expérimentales sont inexistantes ou rares, offrant ainsi des candidats prioritaires pour les tests expérimentaux.

2

Evolution du cyclosome

2 Evolution du cyclosome

2.1 Introduction

2.1.1 Les ubiquitine-ligases

L'Anaphase Promoting Complex, aussi appelé cyclosome (APC/C), a été caractérisé il y a une quinzaine d'années comme faisant partie de la famille des ubiquitine-ligases (aussi appelées E3) (Zachariae and Nasmyth, 1996; Zachariae et al., 1996). Les E3 sont impliquées dans le transfert d'une ou de plusieurs molécules d'ubiquitine sur un substrat protéique au cours d'une réaction en deux étapes impliquant au moins trois partenaires (Figure 53). En premier lieu, une enzyme d'activation (E1) active et transfère l'ubiquitine à une enzyme de conjugaison (E2). Ensuite, l'E3 réalise le transfert de l'ubiquitine fixée sur l'E2 vers une lysine de la protéine cible.

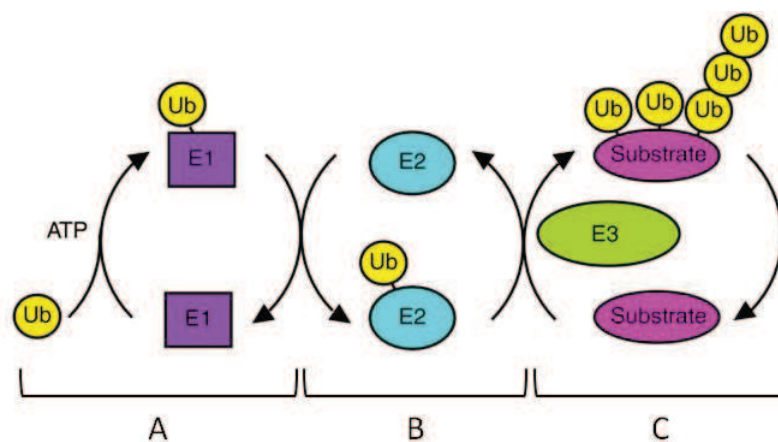


Figure 53. L'ubiquitination se produit par une cascade de trois enzymes (modifié depuis (Matyskiela et al., 2009)).

(A) L'ubiquitine (en jaune) est d'abord activée par une E1, ou "ubiquitine-activating protein" (carré violet), qui couple l'hydrolyse de l'ATP à la formation d'une liaison thioester entre la cystéine du site actif de l'E1 et l'extrémité carboxy-terminale de l'ubiquitine. **(B)** E1 transfère ensuite l'ubiquitine activée à la cystéine du site actif d'une E2, ou enzyme de conjugaison de l'ubiquitine (en bleu). **(C)** Enfin, l'E3, ou ubiquitine ligase (en vert), facilite le transfert de l'ubiquitine de E2 vers une lysine de la protéine cible (en magenta). Dans le cas de l'APC/C et de nombreux autres E3, cette dernière étape peut être répétée plusieurs fois avec le même substrat, ce qui entraîne l'ubiquitination de multiples lysines. En outre, des lysines spécifiques de l'ubiquitine elle-même peuvent être ubiquitinylées, aboutissant à l'assemblage de chaînes de polyubiquitine.

La plupart des E3 sont en mesure de polyubiquitinyler leur substrat par ajout de plusieurs monomères d'ubiquitine. Les protéines polyubiquitinylées sont alors adressées au protéasome 26S pour y être dégradées, alors que les protéines mono-ubiquitinylées peuvent voir leur fonction ou leur localisation subcellulaire modifiées par des protéines possédant des domaines de liaison à l'ubiquitine. Il existe plusieurs familles d'E3 définies sur la base des motifs qu'elles portent. Les deux types de E3 les plus répandues sont les E3 à domaine HECT qui reçoivent l'ubiquitine depuis E2 avant de la fixer sur le substrat (Figure 54a), et les E3 à domaine RING finger qui interviennent dans le transfert de l'ubiquitine depuis E2 directement sur le substrat (Figure 54b) (Pickart, 2001).

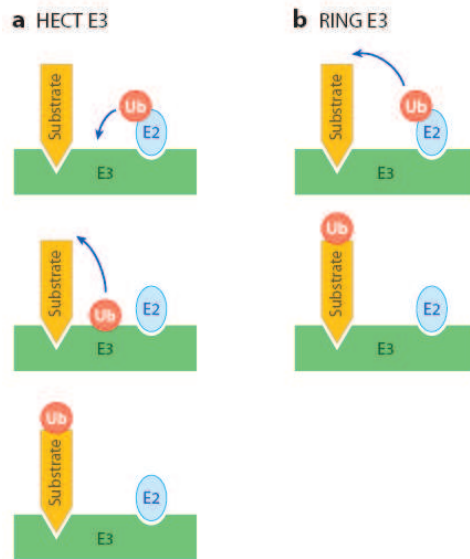


Figure 54. Les E3 à domaine HECT et celles à domaine RING fonctionnent différemment (Deshaies and Joazeiro, 2009). (a) Les E3 à domaine HECT acceptent l'ubiquitine depuis l'enzyme E2 puis transfèrent cette ubiquitine au substrat. (b) Au contraire, les E3 à domaine RING finger effectuent directement le transfert de l'ubiquitine depuis E2 vers le substrat.

2.1.2 L'Anaphase Promoting Complex

Le nombre de sous-unités composant les ubiquitine-ligases à domaine RING est très variable, allant de un à une douzaine (Thornton and Toczyski, 2006). Certaines d'entre elles intègrent également une sous-unité contenant un domaine Cullin. Parmi elles, l'APC/C est la plus grosse (1,5 MDa) et la plus complexe des E3 connues à ce jour. Elle est composée d'une ou plusieurs copies d'au moins une douzaine de sous-unités ainsi que de divers adaptateurs/co-activateurs (Tableau 2), dont la taille varie de 20 à 200 kDa (Manchado et al., 2010; Peters, 2006). La raison d'un si grand nombre de sous-unités reste obscure, notamment car d'autres ubiquitine-ligases à domaines Cullin et RING finger contenant beaucoup moins de sous-unités sont connues : par exemple, la Skp1-Cullin-F-box (SCF) est composée de seulement quatre protéines (Petroski and Deshaies, 2005). La masse moléculaire de l'APC/C le place dans la même catégorie de taille que le ribosome et le protéasome 26S (Peters, 2006). Il a été montré que l'APC/C de levure peut former des dimères dont l'activité *in vitro* est supérieure à celle des monomères (Passmore et al., 2003; Zachariae et al., 1996), ce qui soulève la possibilité intéressante que l'APC/C pourrait fonctionner comme un gigantesque dimère de 3,4 MDa *in vivo* (Passmore et al., 2005). La structure de ce dimère nouvellement acquise en microscopie électronique montre, cependant, que les deux complexes sont placés dos à dos, leurs centres catalytiques étant à l'opposé l'un de l'autre. Il est donc peu probable que les deux APC/C attachés coordonnent leurs efforts sur un même substrat. Néanmoins, dans la forme dimérique, la structure de l'APC/C diffère légèrement de celle du monomère, ce qui suggère que la dimérisation modifie la conformation de l'APC/C et est donc susceptible d'affecter sa processivité (da Fonseca et al., 2011).

Le fonctionnement de l'APC/C a été largement étudié chez l'animal et la levure, où il a été démontré qu'il a un rôle essentiel dans la progression du cycle cellulaire par le contrôle précis de la dégradation de protéines clés du cycle, par exemple les cyclines mitotiques, les inhibiteurs de l'anaphase, les protéines du fuseau mitotique, kinases, inhibiteurs de la réplication, etc. (Manchado et al., 2010).

Il a également été récemment suggéré que l'APC/C peut être actif dans des cellules non-mitotiques et impliqué dans les structures et les fonctions neuronales (Harper et al., 2002; Machado et al., 2010). Il contrôlerait la croissance axonale, l'organisation synaptique, la morphogenèse dendritique, le métabolisme et la survie neuronale en ciblant plusieurs substrats spécifiques.

Sous-unité	APC1	APC2	APC3	APC4	APC5	APC6	APC7	APC8
Domaine fonctionnel	PC_rep	Cullin	TPR	-	TPR	TPR	TPR	TPR
Sous-unité	APC9	APC10	APC11	APC12	APC13	APC14	APC15	APC16
Domaine fonctionnel	-	Apc10	RING	-	-	-	Apc15p	-
Sous-unité	Cdc20	Cdh1	Ama1	Rap	Cortex	Mrf1	Ccs52A	Ccs52B
Domaine fonctionnel	WD40	WD40	WD40	WD40	WD40	WD40	WD40	WD40

Tableau 2. Sous-unités et adaptateurs de l'APC/C décrits dans la littérature, et les domaines fonctionnels qui les composent.

2.2 Structure de l'APC/C

De nombreuses études basées sur des observations en microscopie électronique, des expériences *in vitro* et des expériences de génétique ont abouti à l'idée que l'APC/C est composé d'un cœur et de plusieurs adaptateurs/co-activateurs (Figure 55) (Ohi et al., 2007; Peters, 2006; Thornton and Toczyski, 2006).

2.2.1 Le cœur

Le cœur de l'APC/C, d'environ 1 MDa, correspond à l'association complexe d'une douzaine de sous-unités (chez les animaux et les champignons), réparties en trois régions fonctionnelles (Figure 55) :

Le **bras catalytique** abrite le site de liaison à l'E2. Il est constitué par Apc1 contenant le domaine RING qui permet justement la liaison à l'E2, et par Apc2 et son domaine Cullin qui se fixe à Apc1. Enfin Apc10 comporte une structure en tonneau β impliquée dans la liaison au substrat et qui ralentit la dissociation de celui-ci, augmentant ainsi l'efficacité du complexe (Schreiber et al., 2011).

Le second bras, ou **bras TPR**, permet de positionner à la fois l'E2 et le substrat afin de promouvoir le transfert de l'ubiquitine. Il est composé, chez les mammifères, de quatre sous-unités (Apc3, Apc6, Apc8 et Apc7, cette dernière étant absente chez la levure) qui contiennent des répétitions tetratricopeptide (TPR), et de plusieurs protéines accessoires (Apc9, Apc12, Apc13 et Apc15). Il a été proposé que les répétitions TPR interagissent avec les motifs riches en isoleucine-arginine des régions C-terminale des adaptateurs (Thornton and Toczyski, 2006). Le bras de TPR pourrait également contenir la protéine Apc16, une sous-unité qui a été très récemment mise en évidence (Hutchins et al., 2010). Enfin, l'emplacement de Apc14, une sous-unité essentielle chez la levure, reste indéterminée (Yoon et al., 2002a).

Une étude très récente alliant microscopie optique, spectrométrie de masse et cristallographie montre que les sous-unités TPR (quatre chez les animaux, trois chez la levure) sont présentes en deux exemplaires dans le complexe (Figure 56) (Schreiber et al., 2011). Il a été proposé que les sous-unités accessoires stabilisent l'association des sous-unités TPR avec le reste de l'APC/C (Ohi et al., 2007; Schwickart et al., 2004).

2.2.2 Les adaptateurs, responsables de la reconnaissance du substrat

L'activité et la spécificité de l'APC/C sont régulées par diverses protéines co-activatrices, qui s'associent à l'APC/C de manière transitoire à certaines périodes précises du cycle cellulaire. Les mieux étudiées sont les adaptateurs mitotiques Cdc20 et Cdh1, mais d'autres adaptateurs, par exemple spécifiques de la méiose, ont été identifiés chez les levures (Ama1 et Mfr1), et chez la drosophile (Rap et Cortex). Toutes ces protéines sont homologues et caractérisées par la présence d'éléments dans leur séquence, connus sous le nom de "C-box" et "IR-tail", qui interviennent dans leur liaison à APC/C (Passmore et al., 2003; Schwab et al., 2001; Vodermaier et al., 2003). En outre, tous les adaptateurs de l'APC/C contiennent un domaine WD40 C-terminal qui peut se replier en une structure en hélice, et qui permet l'interaction avec les D-box, KEN-box et O-box porté par les substrats (Manchado et al., 2010)

Il est probable que des variations de séquence dans les domaines WD40 des différents adaptateurs soient impliquées dans la reconnaissance de substrats différents. Ces différences dans la spécificité offrent un moyen de réguler temporellement la destruction de cibles différentes : Cdc20 cible un petit nombre de substrats clés pendant la métaphase, alors que Cdh1 possède une spécificité plus large, et cible les mêmes protéines que Cdc20 ainsi que de nombreuses autres durant la fin de la mitose et la phase G1 (Schwab et al., 1997; Visintin et al., 1997). La liaison de Cdc20 à l'APC/C est favorisée, entre autres, par la phosphorylation de multiples sous-unités de l'APC/C (Kraft et al., 2003), alors que le remplacement de Cdc20 par Cdh1, dépend de la déphosphorylation de ce dernier (Kramer et al., 2000).

2.3 Rôles de l'APC/C

2.3.1 Dégradation des cyclines

L'APC/C est indispensable pour la destruction des cyclines pendant la mitose. Au début de cette phase, l'association de Cdc20 avec l'APC/C est rendue possible par la phosphorylation de certaines des sous-unités du complexe APC/C par plusieurs kinases mitotiques telles que la kinase cycline-dépendante 1 (Cdk1) et la Polo-like kinase 1 (Plk-1). Dans le même temps, la forte activité de Cdk1 maintient la phosphorylation de Cdh1 qui ne peut pas interagir avec l'APC/C.

L'APC/C^{Cdc20} initie le processus de destruction des cyclines pendant la métaphase (c'est-à-dire au moment du "spindle assembly checkpoint", quand tous les chromosomes sont correctement fixés aux deux pôles du fuseau mitotique) (Figure 57, "SAC") et est indispensable à l'inactivation de Cdk1 et à la sortie de la mitose (Clute and Pines, 1999). L'APC/C^{Cdc20} polyubiquitine la cycline B afin que celle-ci soit dégradée par le protéasome 26S. Cdk1, qui n'est ainsi plus associé à la cycline B, subit un changement conformationnel qui l'empêche à la fois d'hydrolyser l'ATP et à ses substrats d'accéder au site actif. Ceci entraîne l'inactivation complète de Cdk1 (Jeffrey et al., 1995), ce qui permet à des protéines phosphatases de déphosphoryler les substrats de Cdk1. Cette condition préalable est essentielle pour le désassemblage du fuseau mitotique, la décondensation des chromosomes, la reformation d'une enveloppe nucléaire et la formation d'un sillon de cytokinèse, puis enfin la sortie de la mitose (Kraft et al., 2003).

Dans le même temps, Cdh1 est déphosphorylé et peut se fixer sur l'APC/C (Figure 57). L'APC/C^{Cdh1} maintient ensuite la dégradation des cyclines pendant la phase G1 et entraîne également la dégradation de Cdc20 et de Plk-1.

L'inactivation de Cdk1 a aussi des conséquences sur la suite du cycle cellulaire. Une période de faible activité de Cdk1 pendant la télophase et pendant la phase G1 est essentielle à la formation de complexes de pré-réplication sur les origines de réplication, sur lesquelles les ADN polymérases initieront la synthèse d'ADN en phase S (Peters, 2006). L'activité des Cdk1 inhibe l'assemblage de ces complexes, et la réplication de l'ADN est donc strictement dépendante des étapes précédentes, c'est-à-dire de l'inactivation de Cdk1 initiée par l'APC/C^{Cdc20} puis poursuivie par l'APC/C^{Cdh1} (Diffley, 2004). Ce rôle indirect de l'APC/C dans la progression de la phase S fait également partie du mécanisme qui limite la réplication de l'ADN à une fois par cycle cellulaire. La phase S ne dépend pas seulement de la période de faible activité de Cdk1, mais aussi d'une phase ultérieure d'activité élevée au cours de laquelle cette fois la synthèse d'ADN est initiée. Ainsi chaque origine de réplication ne peut être activée qu'une fois par cycle cellulaire (Diffley, 2004).

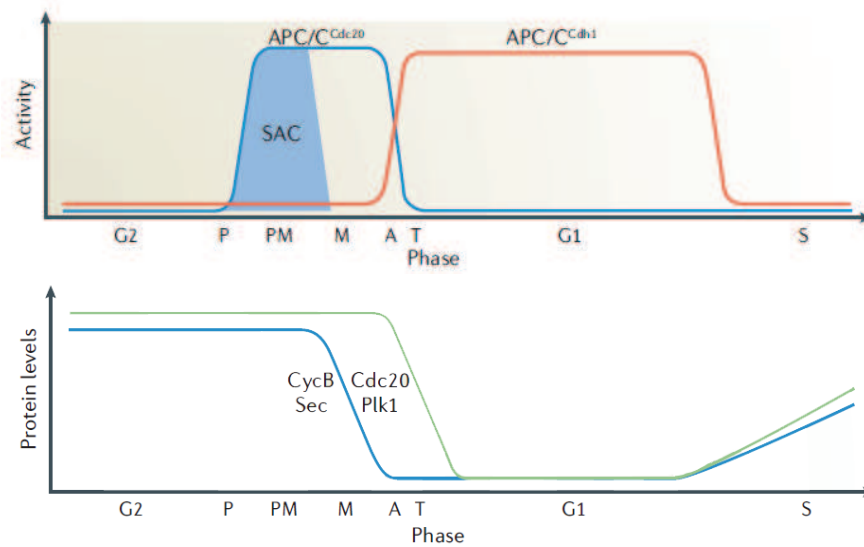


Figure 57. Activation de l'APC/C par Cdc20 et Cdh1 cours du cycle cellulaire. L'APC/C^{Cdc20} est assemblé en prophase (P). La protéolyse de la cycline B (CycB) et des sécurines (Sec) (cf. paragraphe suivant) dépend de l'APC/C^{Cdc20} et est retardée jusqu'à la métaphase (M) par le "spindle assembly checkpoint" (SAC). Au cours de l'anaphase (A) et de la télophase (T), l'APC/C^{Cdh1} est activé, et contribue à la dégradation de la cycline B et des sécurines. Il initie également la destruction d'autres protéines, telles que Polo-like kinase-1 (Plk1) et Cdc20, ce qui conduit à l'inactivation de l'APC/C^{Cdc20}.

2.3.2 Séparation de chromatides-sœurs

La fonction la plus connue et peut-être la plus importante de l'APC/C, et celle qui lui a donné son nom ("anaphase promoting complex"), est son rôle dans l'initiation de l'anaphase (Figure 59). Pour permettre la ségrégation des chromosomes, il est nécessaire de se défaire des éléments qui retiennent les chromatides-sœurs ensemble, c'est-à-dire des composants du complexe cohésine.

Le complexe cohésine permet de maintenir les chromatides-sœurs attachés, de la réplication de l'ADN jusqu'à l'étape de ségrégation des chromosomes. Son cœur est constitué par quatre protéines (Scc1, Scc3, Smc1 et Smc3) qui assurent la cohésion des chromatides (Figure 58) (Onn et al., 2008). Scc2 et Scc4 forment un complexe (le "loading complex") nécessaire à l'association du complexe cœur avec la chromatine, même si rien n'est connu sur la façon dont Scc2/4 exerce cette fonction. La protéine Eco1 est ensuite nécessaire à l'établissement de la cohésion, mais pas à son maintien, pour lequel les protéines Pds5 et Wpl1 sont indispensables.

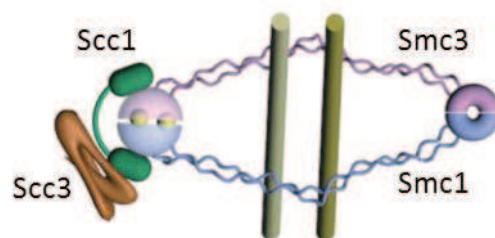


Figure 58. Architecture du complexe cohésine (Onn et al., 2008). Les chromatides sont maintenues ensemble par le complexe formé par Scc1, Scc3, Smc1 et Smc3.

En métaphase, l'APC/C^{Cdc20} initie le processus de destruction des éléments du complexe cohésine en ubiquitinylant la "sécurine", une petite protéine qui sert d'inhibiteur de la protéase qu'est la "séparase". Une fois activée, la séparase clive la sous-unité Scc1 du complexe cohésine (Nasmyth and Haering, 2009), ce qui aboutit à la séparation des chromatides-sœurs. Bien que la destruction de la sécurine soit essentielle pour l'activation des séparases, il a été montré que la levure, les cellules humaines en culture et même les souris peuvent vivre sans sécurine (Jallepalli et al., 2001; Pflieger et al., 2005; Wang et al., 2001; Yamamoto et al., 1996). Il doit donc exister des mécanismes sécurine-indépendants qui contrôlent l'activité des séparases, au travers, par exemple, de leur phosphorylation et leur association ultérieure à Cdk1-cycline B, qui sont suffisants pour l'inhibition des séparases, comme chez *Xenopus laevis* (Gorr et al., 2005; Stemmann et al., 2001). L'APC/C^{Cdc20} pourrait donc contribuer à l'activation des séparases par l'ubiquitinylation à la fois de la sécurine et de la cycline B.

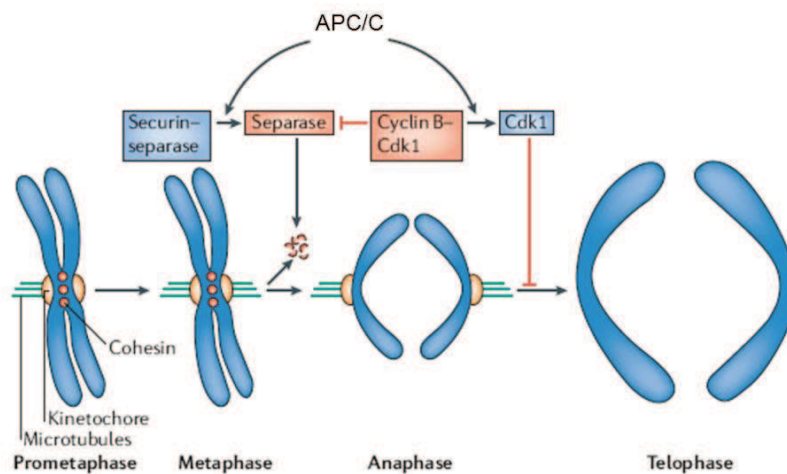


Figure 59. Régulation de l'anaphase et de la sortie de la mitose par l'APC/C (adapté depuis (Peters, 2006)). En métaphase, lorsque tous les kinétochores sont attachés aux microtubules, l'APC/C ubiquitine la cycline B et la sécurine activant ainsi la protéase "séparase" et inactivant la Cdk1. La séparase clive alors les complexes cohésine (cercles rouges) qui maintiennent les chromatides-sœurs ensemble, ce qui initie la séparation de ces mêmes chromatides-sœurs. En parallèle, l'inactivation de la Cdk1 conduit à la déphosphorylation de ses substrats par des protéines phosphatases, permettant ainsi la sortie de la mitose.

2.3.3 La vie sans APC/C

L'inactivation génétique de l'APC/C s'est montrée létale chez toutes les espèces animales chez lesquelles l'expérience a été faite (souris, drosophile et nématode) ainsi que chez la levure (Wirth et al., 2004). De nombreuses protéines autres que les sécurines et les cyclines mitotiques sont également dégradées *via* l'action de l'APC/C. Toutefois, des études indiquent que les seules fonctions essentielles de l'APC/C chez la levure sont la dégradation des cyclines B et des sécurines. Si la nécessité de détruire ces deux protéines est contournée par des manipulations génétiques, alors les cellules de levure peuvent proliférer en l'absence d'APC/C fonctionnel (Shirayama et al., 1999; Thornton and Toczyski, 2003). Ce résultat est remarquable car il montre que l'inactivation des autres substrats de l'APC/C n'est pas essentielle chez la levure (ou bien que la dégradation *via* l'APC/C est seulement un des nombreux mécanismes qui peuvent inactiver ces protéines).

2.3.4 Qu'en est-il de l'APC/C chez les autres eucaryotes ?

Comme évoqué précédemment, la composition, la structure et le rôle de l'APC/C ont surtout été étudiés chez quelques espèces modèles animales et levures. Néanmoins, chez *Arabidopsis thaliana*, la recherche d'homologues par comparaison de séquence a révélé 11 ORFs codant potentiellement pour des sous-unités de l'APC/C (*i.e.* Apc1 à 8, et Apc10 à 12) (Capron et al., 2003a; Eloy et al., 2006; Lima Mde et al., 2010), dont certaines ont été étudiées expérimentalement (Capron et al., 2003b; Serralbo et al., 2006). De façon intéressante, les profils d'expression divergent entre les différents gènes codant pour les sous-unités, ce qui soulève la possibilité qu'il puisse exister plusieurs versions de l'APC/C en fonction des sous-unités le composant. Elles auraient des fonctions distinctes selon le stade de développement et le tissu et/ou le compartiment cellulaire (Lima Mde et al., 2010). En outre, le génome d'*A. thaliana* contient cinq homologues de type Cdc20 ainsi que trois protéines de type Cdh1 qui sont capables de se lier aux cyclines mitotiques tout au long du cycle cellulaire (Fulop et al., 2005). Il a été montré que ces divers homologues de Cdc20 et Cdh1 sont exprimés différemment durant le cycle cellulaire et en fonction du type cellulaire ou tissulaire (Lima Mde et al., 2010).

En revanche, très peu d'informations sont disponibles chez les représentants des autres grandes lignées eucaryotes. La seule exception concerne l'excavé *Trypanosoma brucei*, dont il a été montré que le génome code pour sept homologues des sous-unités de l'APC/C de levure (Apc1 à 3, Apc6, Apc8, Apc10 et Apc1) (Kumar and Wang, 2005). Toutefois, seulement deux d'entre eux (Apc1 et Apc3) sont démontrés expérimentalement comme indispensables pour assurer la transition métaphase-anaphase (Kumar and Wang, 2005).

2.4 Analyse phylogénomique de l'APC/C et de ses principales cibles

La conservation apparente des éléments constitutifs de l'APC/C entre eucaryotes lointains soulève la question de l'origine et l'évolution de cette ubiquitine-ligase atypique par sa taille et le nombre de sous-unités qui la composent. Pour tenter de répondre à cette question, j'ai effectué une analyse phylogénomique de l'ensemble des sous-unités et adaptateurs (16 et 6 protéines, respectivement) décrits dans la littérature dans toutes les lignées eucaryotes pour lesquels des génomes complets (ou presque complets) sont disponibles. J'ai donc recherché les orthologues dans 65 taxons représentant tous les grands groupes eucaryotes à l'exception des Rhizaria et des Cryptophyta (pour lesquels aucun génome n'est disponible et les EST sont trop peu nombreuses). J'ai également inclus dans l'étude plusieurs des cibles majeures (directes ou indirectes) de l'APC/C, à savoir : la séparase, la sécurine, les cyclines A et B, les Cdk-1 et -2, et les 9 composants du complexe cohésine. En parallèle, j'ai effectué une recherche de domaines fonctionnels pour chacun des homologues détectés (Tableau 3, Tableau 4 et Tableau 5).

L'approche méthodologique est similaire à celle employée pour l'analyse du midbody et est détaillé dans l'article disponible en Annexe 2.

		Core subunits							
		Catalytic arm			Structural complex				
		APC10	APC11	APC2	APC1	APC4	APC5		
Opisthokonta	Choanoflagellata	<i>Monosiga brevicollis</i>	APC10	APC11	APC2	APC1	APC4	APC5	
		<i>Salpingoeca rosetta</i>	APC10		APC2/Cullin	PC_rep	-		
	Metazoa	<i>Trichoplax adhaerens</i>	APC10		Cullin	PC_rep			
		<i>Homo sapiens</i>	APC10	zf-C3HC4	APC2/Cullin	PC_rep	-	TPR	
		<i>Danio rerio</i>	APC10	zf-C3HC4	APC2/Cullin	PC_rep	-	TPR	
		<i>Branchiostoma floridae</i>	APC10	zf-C3HC4	APC2/Cullin	PC_rep	-	TPR	
		<i>Drosophila melanogaster</i>	APC10	zf-C3HC4	APC2/Cullin	PC_rep	-	TPR	
		<i>Lottia gigantea</i>	APC10	zf-C3HC4	APC2/Cullin	PC_rep	-	TPR	
		<i>Apis mellifera</i>	APC10	zf-C3HC4	APC2/Cullin	PC_rep	-	TPR	
		<i>Caenorhabditis elegans</i>	APC10	zf-C3HC4	APC2/Cullin	PC_rep	-	-	
		<i>Brugia malayi</i>	APC10	zf-C3HC4	APC2/Cullin	PC_rep	-	-	
		<i>Nematostella vectensis</i>	APC10	zf-C3HC4	APC2/Cullin	PC_rep	-	TPR	
		<i>Helobdella robusta</i>	APC10	zf-C3HC4	APC2/Cullin	PC_rep	-	TPR	
		<i>Daphnia pulex</i>	APC10	zf-C3HC4	APC2/Cullin	PC_rep	-	-	
		<i>Capsaspora owczarzakii</i>	APC10	zf-C3HC4	APC2/Cullin	PC_rep	-	-	
		Fungi	<i>Cryptococcus neoformans</i>	APC10	zf-C3HC4	APC2/Cullin	PC_rep	-	TPR
	<i>Ustilago maydis</i>		APC10	zf-C3HC4	-	PC_rep	-	-	
	<i>Aspergillus fumigatus</i>		APC10	zf-C3HC4	-	PC_rep	-	-	
	<i>Schizosaccharomyces pombe</i>		APC10	zf-C3HC4	APC2/Cullin	PC_rep	-	-	
	<i>Saccharomyces cerevisiae</i>		APC10	zf-C3HC4	APC2/Cullin	PC_rep	-	-	
	<i>Neurospora crassa</i>		APC10	zf-C3HC4	APC2/Cullin	PC_rep	-	-	
	<i>Encephalitozoon cuniculi</i>		APC10	zf-C3HC4	APC2/Cullin	PC_rep	-	-	
	<i>Enterocytozoon bieneusi</i>		APC10			PC_rep			
	<i>Nosema ceranae</i>		APC10	zf-C3HC4		PC_rep			
	<i>Encephalitozoon intestinalis</i>		APC10			PC_rep			
	Chytrids		<i>Batrachochytrium dendrobatidis</i>	APC10			PC_rep		
			<i>Spizellomyces punctatus</i>	APC10	zf-C3HC4	Cullin	PC_rep	-	-
	Apusozoa	<i>Thecamonas trahens</i>		zf-C3HC4	APC2/Cullin	PC_rep	-	-	
	Amoebozoa	<i>Dictyostelium discoideum</i>	APC10	zf-C3HC4	APC2/Cullin	PC_rep	-	TPR	
		<i>Entamoeba histolytica</i>	APC10	zf-C3HC4	APC2/Cullin	PC_rep	-	-	
	Excavata	Metamonada	<i>Giardia intestinalis</i>	APC10	zf-C3HC4		PC_rep		
			<i>Trichomonas vaginalis</i>		zf-C3HC4				
		Euglenozoa	<i>Leishmania major</i>	APC10	zf-C3HC4	Cullin	PC_rep		
			<i>Leishmania infantum</i>	APC10	zf-C3HC4	Cullin	PC_rep		
		Heterolobosoa	<i>Trypanosoma cruzi</i>	APC10	zf-C3HC4	Cullin	PC_rep		
			<i>Trypanosoma brucei</i>	APC10	zf-C3HC4	Cullin	PC_rep		
	Alveolata	Ciliata	<i>Naegleria gruberi</i>	APC10	zf-C3HC4	APC2/Cullin	PC_rep		
			<i>Tetrahymena thermophila</i>	APC10		APC2/Cullin	PC_rep		
			<i>Paramecium tetraurelia</i>	APC10	zf-C3HC4	APC2/Cullin	PC_rep	-	
		Apicomplexa	<i>Oxytricha trifallax</i>	APC10	zf-C3HC4	Cullin	PC_rep	-	
			<i>Plasmodium yoelii</i>	APC10	zf-C3HC4	APC2/Cullin	PC_rep		
			<i>Plasmodium falciparum</i>	APC10	zf-C3HC4				
			<i>Cryptosporidium hominis</i>	APC10	zf-C3HC4				
			<i>Babesia bovis</i>		zf-C3HC4		PC_rep		
			<i>Theileria annulata</i>						
			<i>Toxoplasma gondii</i>						
<i>Blastocystis hominis</i>	APC10	zf-C3HC4		PC_rep					
Heterokonta	<i>Ectocarpus siliculosus</i>	APC10							
	Oomycota	<i>Phytophthora ramorum</i>	APC10	zf-C3HC4	APC2/Cullin	PC_rep	-		
		<i>Phytophthora infestans</i>	APC10	zf-C3HC4	Cullin		-		
	Bacillariophyta	<i>Thalassiosira pseudonana</i>	APC10		APC2/Cullin	PC_rep	-		
		<i>Phaeodactylum tricorutum</i>	APC10	zf-C3HC4	APC2/Cullin	PC_rep	TPR		
	<i>Aureococcus anophagefferens</i>	APC10	zf-C3HC4	APC2/Cullin					
Plantae	Vridiplantae	<i>Oryza sativa</i>	APC10		APC2/Cullin		-		
		<i>Arabidopsis thaliana</i>	APC10	zf-C3HC4	APC2/Cullin	PC_rep	-		
		<i>Chlamydomonas reinhardtii</i>	APC10	zf-C3HC4	APC2/Cullin	PC_rep	-		
		<i>Ostreococcus tauri</i>	APC10	zf-C3HC4	Cullin	PC_rep	-		
		<i>Ostreococcus lucimarinus</i>	APC10	zf-C3HC4	APC2/Cullin	PC_rep	-		
		<i>Physcomitrella patens</i>	APC10	zf-C3HC4	APC2/Cullin	PC_rep	-		
		<i>Chlorella vulgaris</i>	APC10	zf-C3HC4	APC2/Cullin	PC_rep	-		
		<i>Micromonas pusilla</i>	APC10	zf-C3HC4	APC2/Cullin	PC_rep	-		
		<i>Selaginella moellendorffii</i>	APC10	zf-C3HC4	APC2/Cullin	PC_rep	-		
		<i>Selaginella moellendorffii</i>	APC10	zf-C3HC4	APC2/Cullin	PC_rep	-		
	Rhodophyta	<i>Cyanidioschyzon merolae</i>	APC10	zf-C3HC4	APC2/Cullin	PC_rep	-		
		<i>Galdieria sulphuraria</i>	APC10	zf-C3HC4	APC2/Cullin	PC_rep	-		
Haptophyta	<i>Emiliania huxleyi</i>	APC10	zf-C3HC4	APC2/Cullin	PC_rep	-			

Tableau 3a. Composition en domaines fonctionnels des homologues des composants de L'APC/C (1/2). Les domaines détectés grâce à HMMPFAM sont indiqués pour chaque homologue. Un tiret (-) indique qu'aucun domaine n'a été détecté pour cet homologue.

		Core subunits										
		TPR arm and associated proteins										?
		APC8	APC6	APC3	APC7	APC12	APC13	APC16	APC9	APC15	APC14	
Opisthokonta	Choanoflagellata	<i>Monosiga brevicollis</i>	APC8	APC6	APC3	APC7	APC12	APC13	APC16	APC9	APC15	APC14
		<i>Salpingoeca rosetta</i>	APC8/TPR	TPR	TPR	TPR						
	Metazoa	<i>Trichoplax adhaerens</i>	APC8/TPR	TPR	TPR							
		<i>Homo sapiens</i>	APC8/TPR	TPR	TPR	TPR		-				
		<i>Danio rerio</i>	APC8/TPR	TPR	TPR	TPR	-	-	-			
		<i>Branchiostoma floridae</i>	APC8/TPR	TPR	TPR	TPR	-	-	-			
		<i>Drosophila melanogaster</i>	APC8/TPR	TPR	TPR	TPR	-	-				
		<i>Lottia gigantea</i>	APC8/TPR	TPR	TPR	TPR	-	-				
		<i>Apis mellifera</i>	APC8/TPR	TPR	TPR	TPR	-	-				
		<i>Caenorhabditis elegans</i>	APC8/TPR	TPR	TPR	TPR	-	-				
		<i>Brugia malayi</i>	APC8/TPR	TPR	TPR	TPR						
		<i>Nematostella vectensis</i>	APC8/TPR	TPR	TPR	TPR	-	-				
		<i>Helobdella robusta</i>	APC8/TPR	TPR	TPR	TPR						
		<i>Daphnia pulex</i>	APC8/TPR	TPR	TPR	TPR						
		Capsaspora	<i>Capsaspora owczarzaki</i>	APC8/TPR	TPR	TPR	TPR					
	Fungi	<i>Cryptococcus neoformans</i>	APC8/TPR	TPR	TPR							
		<i>Ustilago maydis</i>	APC8/TPR	TPR	TPR							
		<i>Aspergillus fumigatus</i>	APC8/TPR	TPR	TPR							
		<i>Schizosaccharomyces pombe</i>	APC8/TPR	TPR	TPR		-	-				
		<i>Saccharomyces cerevisiae</i>	APC8/TPR	TPR	TPR		-	-			Apc15p	-
		<i>Neurospora crassa</i>	APC8/TPR	TPR	TPR		-	-			Apc15p	
		<i>Encephalitozoon cuniculi</i>	APC8/TPR	TPR	TPR		-	-				
		<i>Enterocytozoon bieneusi</i>	TPR	TPR	TPR							
		<i>Nosema ceranae</i>		TPR	TPR							
		<i>Encephalitozoon intestinalis</i>	TPR	TPR	TPR							
	Chytrids	<i>Batrachochytrium dendrobatidis</i>	TPR	TPR	TPR							
		<i>Spizellomyces punctatus</i>	APC8/TPR	TPR	TPR	TPR						
	Apusozoa	<i>Thecamonas trahens</i>	APC8/TPR	TPR	TPR	TPR						
	Amoebozoa	<i>Dictyostelium discoideum</i>	APC8/TPR	TPR	TPR							
		<i>Entamoeba histolytica</i>	APC8/TPR	TPR	TPR	TPR	-	-				
	Excavata	Metamonada	<i>Giardia intestinalis</i>									
			<i>Trichomonas vaginalis</i>									
		Euglenozoa	<i>Leishmania major</i>	APC8/TPR	TPR	TPR	TPR					
<i>Leishmania infantum</i>			TPR	TPR	TPR							
<i>Trypanosoma cruzi</i>			TPR	TPR	TPR							
Heterolobosoa	<i>Trypanosoma brucei</i>	TPR	TPR	TPR								
	<i>Naegleria gruberi</i>	TPR	TPR	TPR								
Alveolata	Ciliata	<i>Tetrahymena thermophila</i>	TPR	TPR	TPR							
		<i>Paramecium tetraurelia</i>	APC8/TPR	TPR	TPR							
		<i>Oxytricha trifallax</i>	APC8/TPR	TPR	TPR							
	Apicomplexa	<i>Plasmodium yoelii</i>	TPR	TPR	TPR							
		<i>Plasmodium falciparum</i>		TPR	TPR							
		<i>Cryptosporidium hominis</i>			TPR							
		<i>Babesia bovis</i>	APC8/TPR	TPR	TPR							
	<i>Theileria annulata</i>											
	<i>Toxoplasma gondii</i>											
	<i>Blastocystis hominis</i>											
Heterokonta	Oomycota	<i>Ectocarpus siliculosus</i>	TPR	TPR								
		<i>Phytophthora ramorum</i>	TPR	TPR	TPR	TPR						
		<i>Phytophthora infestans</i>	APC8/TPR	TPR	TPR	TPR	-					
	Bacillariophyta	<i>Thalassiosira pseudonana</i>	APC8/TPR	TPR	TPR	TPR	-	-				
		<i>Phaeodactylum tricorutum</i>	TPR	TPR	TPR							
	<i>Aureococcus anophagefferens</i>	APC8/TPR	TPR	TPR	TPR	-	-					
Plantae	Vridiplantae	<i>Oryza sativa</i>	APC8/TPR	TPR	TPR	TPR						
		<i>Arabidopsis thaliana</i>	APC8/TPR	TPR	TPR	TPR	-	-				
		<i>Chlamydomonas reinhardtii</i>	APC8/TPR	TPR	TPR	TPR	-	-				
		<i>Ostreococcus tauri</i>	APC8/TPR	TPR	TPR							
		<i>Ostreococcus lucimarinus</i>	APC8/TPR	TPR	TPR	TPR						
		<i>Physcomitrella patens</i>	TPR	TPR	TPR							
		<i>Chlorella vulgaris</i>	APC8/TPR	TPR	TPR	TPR	-	-				
		<i>Micromonas pusilla</i>	APC8/TPR	TPR	TPR							
		<i>Selaginella moellendorffii</i>	APC8/TPR	TPR	TPR	TPR						
	Rhodophyta	<i>Cyanidioschyzon merolae</i>	APC8/TPR	TPR	TPR	TPR						
<i>Galdieria sulphuraria</i>		APC8/TPR	TPR	TPR								
Haptophyta	<i>Emiliania huxleyi</i>	TPR	TPR	TPR								
		APC8/TPR	TPR	TPR								

Tableau 2b. Composition en domaines fonctionnels des homologues des composants de L'APC/C (1/2). Les domaines détectés grâce à HMMPFAM sont indiqués pour chaque homologue. Un tiret (-) indique qu'aucun domaine n'a été détecté pour cet homologue.

			Co-activators								
			Mitosis co-activators			Meiotic co-activator					
			CDC20	CDH1	incertain	Mfr1	Ama1	Rap	Cortex		
Opisthokonta	Choanoflagellata	<i>Monosiga brevicollis</i>	WD40	WD40							
		<i>Salpingoeca rosetta</i>	WD40	WD40							
	Metazoa	<i>Trichoplax adhaerens</i>	WD40	WD40							
		<i>Homo sapiens</i>	WD40	WD40							
		<i>Danio rerio</i>	WD40	WD40							
		<i>Branchiostoma floridae</i>	WD40	WD40							
		<i>Drosophila melanogaster</i>	WD40	WD40							
		<i>Lottia gigantea</i>	WD40	WD40					WD40	WD40	
		<i>Apis mellifera</i>	WD40	WD40							
		<i>Caenorhabditis elegans</i>	WD40	WD40							
		<i>Brugia malayi</i>		WD40							
		<i>Nematostella vectensis</i>	WD40	WD40							
		<i>Helobdella robusta</i>	WD40	WD40							
		<i>Daphnia pulex</i>		WD40							
	Capsaspora	<i>Capsaspora owczarzaki</i>	WD40	WD40							
	Fungi	<i>Cryptococcus neoformans</i>	WD40	WD40							
		<i>Ustilago maydis</i>	WD40	WD40							
		<i>Aspergillus fumigatus</i>	WD40	WD40							
		<i>Schizosaccharomyces pombe</i>	WD40	WD40				WD40			
		<i>Saccharomyces cerevisiae</i>	WD40	WD40				WD40	WD40		
		<i>Neurospora crassa</i>	WD40	WD40					WD40		
		<i>Encephalitozoon cuniculi</i>	WD40	WD40					WD40		
		<i>Enterocytozoon bieneusi</i>	WD40		WD40						
		<i>Nosema ceranae</i>	WD40	WD40							
		<i>Encephalitozoon intestinalis</i>	WD40	WD40							
		Chytrids	<i>Batrachomyces dendrobatidis</i>	WD40		WD40					
			<i>Spizellomyces punctatus</i>	WD40	WD40						
	Apusozoa	<i>Thecamonas trahens</i>	WD40	WD40							
	Amoebozoa	<i>Dictyostelium discoideum</i>	WD40	WD40							
		<i>Entamoeba histolytica</i>	WD40	WD40							
	Excavata	Metamonada	<i>Giardia intestinalis</i>				WD40				
			<i>Trichomonas vaginalis</i>								
Euglenozoa		<i>Leishmania major</i>					WD40				
		<i>Leishmania infantum</i>	WD40								
		<i>Trypanosoma cruzi</i>	WD40								
Heterolobosoa		<i>Trypanosoma brucei</i>	WD40								
	<i>Naegleria gruberi</i>	WD40									
Alveolata	Ciliata	<i>Tetrahymena thermophila</i>	WD40	WD40							
		<i>Paramecium tetraurelia</i>	WD40	WD40							
		<i>Oxytricha trifallax</i>	WD40	WD40							
	Apicomplexa	<i>Plasmodium yoelii</i>		WD40							
		<i>Plasmodium falciparum</i>			WD40						
		<i>Cryptosporidium hominis</i>			WD40						
		<i>Babesia bovis</i>		WD40							
		<i>Theileria annulata</i>									
	<i>Toxoplasma gondii</i>										
	<i>Blastocystis hominis</i>										
Heterokonta		<i>Ectocarpus siliculosus</i>	WD40								
	Oomycota	<i>Phytophthora ramorum</i>	WD40	WD40							
		<i>Phytophthora infestans</i>	WD40	WD40							
	Bacillariophyta	<i>Thalassiosira pseudonana</i>	WD40	WD40							
		<i>Phaeodactylum tricornutum</i>	WD40	WD40							
		<i>Aureococcus anophagefferens</i>	WD40	WD40							
Plantae	Vridiplantae	<i>Oryza sativa</i>	WD40	WD40							
		<i>Arabidopsis thaliana</i>	WD40	WD40							
		<i>Chlamydomonas reinhardtii</i>	WD40	WD40							
		<i>Ostreococcus tauri</i>	WD40	WD40							
		<i>Ostreococcus lucimarinus</i>	WD40	WD40							
		<i>Physcomitrella patens</i>	WD40								
		<i>Chlorella vulgaris</i>	WD40	WD40							
		<i>Micromonas pusilla</i>	WD40	WD40							
	Rhodophyta	<i>Selaginella moellendorffii</i>	WD40	WD40							
		<i>Cyanidioschyzon merolae</i>	WD40	WD40							
	Haptophyta	<i>Galdieria sulphuraria</i>	WD40	WD40							
		<i>Emiliania huxleyi</i>	WD40	WD40							

Tableau 4. Composition en domaines fonctionnels des homologues des activateurs de L'APC/C. Les domaines détectés grâce à HMMPFAM sont indiqués pour chaque homologue. Un tiret (-) indique qu'aucun domaine n'a été détecté pour cet homologue.

		Core complex					
		Smc1	Smc3	Scs1	Scs3		
Opisthokonta	Choanoflagellata	<i>Monosiga brevicollis</i>	-	SMC_hinge/SMC_N	Rad21_Rec8/Rad21_Rec8_N	-	
		<i>Salpingoeca rosetta</i>	SMC_hinge/SMC_N	-	Rad21_Rec8/Rad21_Rec8_N	NA	
	Metazoa	<i>Trichoplax adhaerens</i>	-	-	Rad21_Rec8/Rad21_Rec8_N	HEAT/STAG	
		<i>Homo sapiens</i>	-	-	Rad21_Rec8/Rad21_Rec8_N	HEAT	
		<i>Danio rerio</i>	-	-	Rad21_Rec8/Rad21_Rec8_N	HEAT	
		<i>Branchiostoma floridae</i>	-	-	Rad21_Rec8_N	STAG	
		<i>Drosophila melanogaster</i>	-	-	Rad21_Rec8/Rad21_Rec8_N	HEAT	
		<i>Lottia gigantea</i>	-	SMC_N	Rad21_Rec8/Rad21_Rec8_N	HEAT	
		<i>Apis mellifera</i>	SMC_hinge/SMC_N	SMC_hinge/SMC_N		HEAT	
		<i>Caenorhabditis elegans</i>	SMC_hinge/SMC_N	SMC_hinge/SMC_N	Rad21_Rec8/Rad21_Rec8_N	STAG	
		<i>Brugia malayi</i>	SMC_hinge/SMC_N	SMC_hinge/SMC_N	Rad21_Rec8/Rad21_Rec8_N	HEAT	
		<i>Nematostella vectensis</i>	-	-	Rad21_Rec8/Rad21_Rec8_N	STAG	
		<i>Helobdella robusta</i>	-	-	Rad21_Rec8_N	HEAT	
		<i>Daphnia pulex</i>	-	SMC_hinge/SMC_N	Rad21_Rec8/Rad21_Rec8_N	HEAT	
	Capsaspora	<i>Capsaspora owczarzakii</i>	-	SMC_hinge/SMC_N	Rad21_Rec8/Rad21_Rec8_N	STAG	
	Fungi	<i>Cryptococcus neoformans</i>	SMC_hinge/SMC_N	SMC_hinge/SMC_N	Rad21_Rec8/Rad21_Rec8_N	STAG	
		<i>Ustilago maydis</i>	SMC_hinge/SMC_N	-	Rad21_Rec8/Rad21_Rec8_N	HEAT	
		<i>Aspergillus fumigatus</i>	-	-	Rad21_Rec8/Rad21_Rec8_N	STAG	
		<i>Schizosaccharomyces pombe</i>	SMC_hinge/SMC_N	-	Rad21_Rec8/Rad21_Rec8_N	STAG	
		<i>Saccharomyces cerevisiae</i>	-	-	Rad21_Rec8/Rad21_Rec8_N	STAG	
		<i>Neurospora crassa</i>	SMC_hinge/SMC_N	SMC_N	Rad21_Rec8/Rad21_Rec8_N	HEAT	
		<i>Encephalitozoon cuniculi</i>	SMC_hinge/SMC_N	SMC_hinge/SMC_N		-	
		<i>Enterocytozoon bieneusi</i>	-	-		-	
		<i>Nosema ceranae</i>	-	SMC_hinge/SMC_N		-	
		<i>Encephalitozoon intestinalis</i>	SMC_hinge/SMC_N	SMC_hinge/SMC_N		-	
		<i>Batrachochytrium dendrobatidis</i>	SMC_hinge/SMC_N	SMC_hinge/SMC_N	Rad21_Rec8_N	STAG	
		<i>Spizellomyces punctatus</i>	-	SMC_hinge/SMC_N	Rad21_Rec8/Rad21_Rec8_N	HEAT	
	Apusozoa	<i>Thecamonas trahens</i>	SMC_hinge/SMC_N	SMC_hinge/SMC_N	Rad21_Rec8/Rad21_Rec8_N	HEAT	
	Amoebozoa	<i>Dictyostelium discoideum</i>	-	SMC_hinge/SMC_N	Rad21_Rec8/Rad21_Rec8_N	STAG	
		<i>Entamoeba histolytica</i>	SMC_hinge/SMC_N	-	Rad21_Rec8/Rad21_Rec8_N		
	Excavata	Metamonada	<i>Giardia intestinalis</i>	SMC_hinge/SMC_N	SMC_hinge/SMC_N		
			<i>Trichomonas vaginalis</i>	-	-	Rad21_Rec8/Rad21_Rec8_N	HEAT
		Euglenozoa	<i>Leishmania major</i>	SMC_hinge/SMC_N	SMC_N	Rad21_Rec8/Rad21_Rec8_N	STAG
			<i>Leishmania infantum</i>	SMC_hinge/SMC_N	-	Rad21_Rec8/Rad21_Rec8_N	STAG
			<i>Trypanosoma cruzi</i>	SMC_hinge/SMC_N	SMC_N	Rad21_Rec8/Rad21_Rec8_N	STAG
		<i>Trypanosoma brucei</i>	SMC_hinge/SMC_N	SMC_hinge/SMC_N	Rad21_Rec8/Rad21_Rec8_N	STAG	
Heterolobosoa	<i>Naegleria gruberi</i>	-	-	Rad21_Rec8/Rad21_Rec8_N	HEAT		
Alveolata	Ciliata	<i>Tetrahymena thermophila</i>	-	-			
		<i>Paramecium tetraurelia</i>	SMC_hinge/SMC_N	-			
	<i>Oxytricha trifallax</i>	SMC_N	-	?			
	Apicomplexa	<i>Plasmodium yoelii</i>	-	SMC_hinge/SMC_N		STAG	
		<i>Plasmodium falciparum</i>	SMC_hinge/SMC_N	SMC_hinge/SMC_N		STAG	
		<i>Cryptosporidium hominis</i>	SMC_hinge/SMC_N	SMC_N	Rad21_Rec8/Rad21_Rec8_N		
		<i>Babesia bovis</i>	SMC_hinge/SMC_N	SMC_hinge/SMC_N		STAG	
<i>Theileria annulata</i>	-	-		VAR1			
<i>Toxoplasma gondii</i>	SMC_hinge/SMC_N	SMC_hinge/SMC_N					
Blastocystae	<i>Blastocystis hominis</i>	SMC_N	-				
Phaeophyceae	<i>Ectocarpus siliculosus</i>	SMC_hinge/SMC_N	-	Rad21_Rec8_N	HEAT		
	Oomycota	<i>Phytophthora ramorum</i>	-	SMC_hinge/SMC_N	Rad21_Rec8/Rad21_Rec8_N	HEAT	
<i>Phytophthora infestans</i>		SMC_hinge/SMC_N	SMC_hinge/SMC_N	Rad21_Rec8/Rad21_Rec8_N	HEAT		
Bacillariophyta	<i>Thalassiosira pseudonana</i>	SMC_hinge/SMC_N	-	Rad21_Rec8_N			
	<i>Phaeodactylum tricorutum</i>	SMC_hinge/SMC_N	-	Rad21_Rec8/Rad21_Rec8_N			
	<i>Aureococcus anophagefferens</i>	SMC_hinge/SMC_N	SMC_hinge/SMC_N	Rad21_Rec8_N			
Plantae	Vridiplantae	<i>Oryza sativa</i>	SMC_hinge/SMC_N	-	Rad21_Rec8/Rad21_Rec8_N	HEAT	
		<i>Arabidopsis thaliana</i>	-	-	Rad21_Rec8/Rad21_Rec8_N	STAG	
		<i>Chlamydomonas reinhardtii</i>	SMC_hinge/SMC_N	SMC_N	Rad21_Rec8_N	STAG	
		<i>Ostreococcus tauri</i>	-	SMC_hinge/SMC_N	Rad21_Rec8/Rad21_Rec8_N	HEAT	
		<i>Ostreococcus lucimarinus</i>	SMC_hinge/SMC_N	SMC_hinge/SMC_N	Rad21_Rec8/Rad21_Rec8_N	HEAT	
		<i>Physcomitrella patens</i>	SMC_hinge/SMC_N	-	Rad21_Rec8_N	HEAT	
		<i>Chlorella vulgaris</i>	-	SMC_N	Rad21_Rec8/Rad21_Rec8_N	HEAT	
		<i>Micromonas pusilla</i>	-	SMC_hinge/SMC_N	Rad21_Rec8/Rad21_Rec8_N	HEAT	
	Rhodophyta	<i>Selaginella moellendorffii</i>	-	SMC_hinge/SMC_N	Rad21_Rec8_N	HEAT	
		<i>Cyanidioschyzon merolae</i>	SMC_hinge/SMC_N	SMC_hinge/SMC_N	Rad21_Rec8/Rad21_Rec8_N	HEAT	
		<i>Galdieria sulphuraria</i>	SMC_hinge/SMC_N	-	?	HEAT	
		<i>Emiliania huxleyi</i>	-	SMC_N	Rad21_Rec8/Rad21_Rec8_N	HEAT	

Tableau 5a. Composition en domaines fonctionnels des homologues des composants du complexe cohésine. Les domaines détectés grâce à HMMPFAM sont indiqués pour chaque homologue. Un tiret (-) indique qu'aucun domaine n'a été détecté pour cet homologue. Un point d'interrogation indique qu'un homologue existe potentiellement, mais que la séquence récupérée par TblastN était trop partielle pour être incluse dans l'étude.

			Loading complex		Cohesion establishment	Maintenance			Dissolution		
			Scc2	Scc4	Eco1	Fds5	Wpl1	Rad61	Separase	Securin	
Opisthokonta	Choanoflagellata	<i>Monosiga brevicollis</i>	HEAT			HEAT			Peptidase_C50		
		<i>Salpingoeca rosetta</i>	HEAT		-	HEAT	WAPL		Peptidase_C50		
		<i>Trichoplax adhaerens</i>	HEAT	TPR	-	HEAT	WAPL		?		
	Metazoa	<i>Homo sapiens</i>	HEAT	TPR	-	HEAT	WAPL		Peptidase_C50	Securin	
		<i>Danio rerio</i>	HEAT	-	-	-	WAPL		Peptidase_C50	Securin	
		<i>Branchiostoma floridae</i>	HEAT	-	-	HEAT	WAPL		Peptidase_C50		
		<i>Drosophila melanogaster</i>	HEAT	TPR	-	-	WAPL		Peptidase_C50	-	
		<i>Lottia gigantea</i>	HEAT	-	-	HEAT	WAPL		Peptidase_C50		
		<i>Apis mellifera</i>	HEAT	TPR	-	HEAT	WAPL				
		<i>Caenorhabditis elegans</i>	HEAT	TPR	-	HEAT	WAPL		Peptidase_C50	-	
		<i>Brugia malayi</i>	HEAT	-	-	HEAT	WAPL		Peptidase_C50		
		<i>Nematostella vectensis</i>	HEAT	TPR	-	HEAT	WAPL		Peptidase_C50		
		<i>Helobdella robusta</i>	HEAT	TPR	-	HEAT	WAPL		Peptidase_C50		
		<i>Daphnia pulex</i>	HEAT	TPR	-	HEAT	WAPL		Peptidase_C50		
		Fungi	<i>Capsaspora owczarzaki</i>	HEAT	TPR	-	HEAT	WAPL		Peptidase_C50	
			<i>Cryptococcus neoformans</i>	HEAT		-	HEAT	WAPL		Peptidase_C50	
			<i>Ustilago maydis</i>	HEAT			HEAT	WAPL		Peptidase_C50	
			<i>Aspergillus fumigatus</i>	HEAT	TPR	-	HEAT	WAPL		Peptidase_C50	
	<i>Schizosaccharomyces pombe</i>		HEAT	-	-	HEAT	WAPL		Peptidase_C50	-	
	<i>Saccharomyces cerevisiae</i>		HEAT	-	-	HEAT		-	Peptidase_C50	-	
	<i>Neurospora crassa</i>		HEAT	-	-	HEAT	WAPL		Peptidase_C50		
	<i>Encephalitozoon cuniculi</i>								Peptidase_C50		
	<i>Enterocytozoon bieneusi</i>								Peptidase_C50		
	<i>Nosema ceranae</i>								Peptidase_C50		
	<i>Encephalitozoon intestinalis</i>								Peptidase_C50		
	<i>Batrachochytrium dendrobatidis</i>		HEAT		-	HEAT	WAPL		Peptidase_C50		
	<i>Spizellomyces punctatus</i>	HEAT	-	-	HEAT	WAPL		Peptidase_C50			
	Apusozoa	<i>Thecamonas trahens</i>	HEAT		-				Peptidase_C50		
	Amoebozoa	<i>Dictyostelium discoideum</i>	HEAT		-	HEAT			Peptidase_C50		
		<i>Entamoeba histolytica</i>							Peptidase_C50		
	Excavata	Metamonada	<i>Giardia intestinalis</i>						Peptidase_C50		
			<i>Trichomonas vaginalis</i>	HEAT					Peptidase_C50		
		Euglenozoa	<i>Leishmania major</i>			-	HEAT			Peptidase_C50	
			<i>Leishmania infantum</i>			-	HEAT			Peptidase_C50	
			<i>Trypanosoma cruzi</i>			-	HEAT			Peptidase_C50	
<i>Trypanosoma brucei</i>				-	HEAT			Peptidase_C50			
Heterolobosoa		<i>Naegleria gruberi</i>	HEAT		-	HEAT			Peptidase_C50		
Alveolata	Ciliata	<i>Tetrahymena thermophila</i>						Peptidase_C50			
		<i>Paramecium tetraurelia</i>						Peptidase_C50			
	Apicomplexa	<i>Oxytricha trifallax</i>							Peptidase_C50		
		<i>Plasmodium yoelii</i>							Peptidase_C50		
		<i>Plasmodium falciparum</i>							Peptidase_C50		
		<i>Cryptosporidium hominis</i>							Peptidase_C50		
		<i>Babesia bovis</i>									
<i>Theileria annulata</i>											
<i>Toxoplasma gondii</i>							Peptidase_C50				
Bacillariophyta	Blastocystae	<i>Blastocystis hominis</i>						Peptidase_C50			
	Phaeophyceae	<i>Ectocarpus siliculosus</i>	HEAT		-	HEAT			Peptidase_C50		
		<i>Phytophthora ramorum</i>	HEAT		-	-					
	Oomycota	<i>Phytophthora infestans</i>	HEAT		-	HEAT			Peptidase_C50		
		<i>Thalassiosira pseudonana</i>	HEAT		-	HEAT			Peptidase_C50		
		<i>Phaeodactylum tricorutum</i>	HEAT		-				Peptidase_C50		
	<i>Aureococcus anophagefferens</i>	-						Peptidase_C50			
Plantae	Vridiplantae	<i>Oryza sativa</i>	HEAT	-	-	HEAT	WAPL		Peptidase_C50		
		<i>Arabidopsis thaliana</i>	HEAT	TPR	-	HEAT	WAPL		Peptidase_C50		
		<i>Chlamydomonas reinhardtii</i>	-			HEAT			Peptidase_C50		
		<i>Ostreococcus tauri</i>	HEAT	TPR	-	HEAT	WAPL		Peptidase_C50		
		<i>Ostreococcus lucimarinus</i>	HEAT	TPR	-	HEAT	WAPL		Peptidase_C50		
		<i>Physcomitrella patens</i>	HEAT	TPR	-	HEAT	WAPL		Peptidase_C50		
		<i>Chlorella vulgaris</i>	HEAT		-	HEAT	-		Peptidase_C50		
		<i>Micromonas pusilla</i>	HEAT	-	-	HEAT					
		<i>Selaginella moellendorffii</i>	HEAT	TPR	-	HEAT	WAPL		Peptidase_C50		
	Rhodophyta	<i>Cyanidioschyzon merolae</i>			-				Peptidase_C50		
		<i>Galdieria sulphuraria</i>	?		-	HEAT			Peptidase_C50		
	Haptophyta	<i>Emiliania huxleyi</i>	HEAT		-	-			Peptidase_C50		

Tableau 4b. Composition en domaines fonctionnels des homologues des composants du complexe cohésine. Les domaines détectés grâce à HMMPFAM sont indiqués pour chaque homologue. Un tiret (-) indique qu'aucun domaine n'a été détecté pour cet homologue. Un point d'interrogation indique qu'un homologue existe potentiellement, mais que la séquence récupérée par TBLASTN était trop partielle pour être incluse dans l'étude.

2.4.1 La plupart des composants et adaptateurs existaient chez LECA

Sur la base de l'analyse des phylogénies individuelles des composants, et en utilisant la même approche que lors de l'analyse du midbody, j'ai pu inférer que 12 composants de l'APC/C (sur 16) et 16) et deux adaptateurs (sur 6) étaient déjà présents chez LECA (Figure 8 et

Figure 61). De plus, la recherche de domaines fonctionnels montre que la composition en domaines est très conservée au sein des homologues, ce qui suggère que ceux-ci possèdent la même fonction biochimique, et que cette fonction a vraisemblablement été héritée de LECA.

2.4.1.1 La majorité des composants a été héritée de LECA et conservée au cours de l'évolution des eucaryotes

Les trois composants du **bras catalytique** (Apc10, Apc1 et Apc2), la plus grosse sous-unité de la **plateforme** (Apc1) et trois des quatre sous-unités TPR du **bras TPR** (Apc3, Apc6 et Apc8) ont vraisemblablement héritées de LECA et été extrêmement conservées dans l'ensemble des lignées et (Tableau 6). L'analyse phylogénétique des **adaptateurs** Cdc20 et Cdh1 a montré que ces deux paralogues étaient, eux aussi, déjà présents chez LECA (Tableau 7).

2.4.1.2 Des composants hérités de LECA ont subi de multiples pertes

Apc4 et Apc5, les deux autres composants de la **plateforme**, sont souvent absents (ex : chez les microsporidies, les excavés et les alvéolés) mais sont néanmoins retrouvés dans plusieurs lignées d'unikontes et de bikontes (Tableau 6) ce qui suggère qu'ils étaient déjà présents chez leur dernier ancêtre commun (*i.e.* LECA), et qu'ils ont été secondairement perdus au cours de la diversification des eucaryotes.

De même, la quatrième sous-unité du **bras TPR** (Apc7) a été perdue dans de nombreuses lignées de champignons, d'excavés, d'alvéolés et certaines lignées de plantes (Tableau 6). Toutefois, sa présence dans de nombreuses lignées distantes telles que les animaux, les amoebozoa, les hétérokontes et les plantes, ainsi que l'absence de THG suggérés par la phylogénie du composant, suggèrent qu'Apc7 était déjà présent chez LECA et que son absence dans des lignées non apparentées reflète des pertes convergentes.

Apc12 et Apc13, des **protéines accessoires du bras TPR**, ont été elles aussi perdues massivement dans certaines lignées, mais ici aussi, leur présence dans des lignées éloignées suggère une présence ancestrale chez LECA (Tableau 6). Les phylogénies de ces deux composants sont très mal résolues (due au très faible nombre de positions utilisables pour la reconstruction phylogénétique : respectivement 22 et 25 positions), et il n'est pas possible d'exclure des THG. Néanmoins, étant donné la répartition taxonomique observée, l'hypothèse de THG impliquerait un très grand nombre d'évènements de transferts qui paraissent difficilement imaginables si on fait l'hypothèse que de tels évènements sont rares entre eucaryotes.

		Core subunits						
		APC10	Catalytic arm		Structural complex			
			APC11	APC2	APC1	APC4	APC5	
Opisthokonta	Choanoflagellata	<i>Monosiga brevicollis</i>	XP_001746604		XP_001749209	XP_001742225	XP_001744221	
		<i>Salpingoeca rosetta</i>	Acc Num NA		Acc Num NA	Acc Num NA		
	Metazoa	<i>Trichoplax adhaerens</i>	XP_002109549	XP_002115875	XP_002114958	XP_002109033	XP_002111892	XP_002109398
		<i>Homo sapiens</i>	NP_055700	NP_057560	NP_037498	NP_073153	NP_037499	NP_057321
		<i>Danio rerio</i>	XP_001923028	NP_001091950	CAX13097	CAX12586	NP_001008589	NP_055916
		<i>Branchiostoma floridae</i>	XP_002598258	XP_002606314	XP_002600222	XP_002610275	XP_002609002	XP_002610040
		<i>Drosophila melanogaster</i>	NP_611223	NP_001097122	NP_611862	NP_573025	NP_572522	AAM51081
		<i>Lottia gigantea</i>	Acc Num NA	Acc Num NA	Acc Num NA	Acc Num NA	Acc Num NA	Acc Num NA
		<i>Apis mellifera</i>	XP_396738	XP_001122138	XP_395411	XP_001122449	XP_393301	XP_393771
		<i>Caenorhabditis elegans</i>	NP_001021777 NP_505679	XP_777934 NP_497937	NP_498762	NP_496383	NP_499074	NP_499485
		<i>Brugia malayi</i>	XP_001901978	XP_001901276 XP_001898974	XP_001894849	XP_001893638	XP_001900733	XP_001893623
		<i>Nematostella vectensis</i>	XP_001630621	XP_001628272	XP_001625156	XP_001632264	XP_001624909	XP_001626465
	<i>Helobdella robusta</i>	Acc Num NA	Acc Num NA	Acc Num NA	Acc Num NA	Acc Num NA	Acc Num NA	
	<i>Daphnia pulex</i>	Acc Num NA	Acc Num NA	Acc Num NA	Acc Num NA	Acc Num NA	Acc Num NA	
	Capsaspora	<i>Capsaspora owczarzaki</i>	Acc Num NA	Acc Num NA	Acc Num NA	Acc Num NA	Acc Num NA	Acc Num NA
	Fungi	<i>Cryptococcus neoformans</i>	XP_571694	XP_777934	XP_568721	XP_570920	XP_572593	XP_776363
		<i>Ustilago maydis</i>	XP_756827	XM_753342	XP_759373	XP_758574	XP_761900	XP_756294
		<i>Aspergillus fumigatus</i>	XP_752943	XP_746794	XP_749651	EDP53140	XP_752937	EDP53116
		<i>Schizosaccharomyces pombe</i>	NP_595803	NP_593423	NP_001018806	NP_595158	XP_001713100	XP_001713082
		<i>Saccharomyces cerevisiae</i>	1GQP_A	NP_010276	EEU06984	EDN62645	NP_010403	EDV10820
		<i>Neurospora crassa</i>	CAC28560	XP_001728250	XP_964316	XP_959459	XP_956179	XP_964327
		<i>Encephalitozoon cuniculi</i>	NP_597441			NP_586231		
		<i>Enterocytozoon bienersi H348</i>	XP_002649405	XP_001827913(?)		XP_002649879		
		<i>Nosema ceranae BRL01</i>	XP_002995625			XP_002996689		
		<i>Encephalitozoon intestinalis ATCC 50506</i>	XP_003072909			XP_003073796		
		<i>Batrachochytrium dendrobatidis</i>	Acc Num NA	Acc Num NA	Acc Num NA	Acc Num NA	Acc Num NA	Acc Num NA
		<i>Spizellomyces punctatus</i>	Acc Num NA	Acc Num NA	Acc Num NA	Acc Num NA	Acc Num NA	Acc Num NA
	Apusozoa	<i>Thecamonas trahens</i>	Acc Num NA	Acc Num NA	Acc Num NA	Acc Num NA	Acc Num NA	Acc Num NA
	Amoebozoa	<i>Dictyostelium discoideum</i>	XP_646955	XP_637455	XP_643183	XP_638254	XP_638336	XP_641250
		<i>Entamoeba histolytica</i>	XP_654652	XP_651657		XP_648486		
	Excavata	Metamonada	<i>Giardia intestinalis</i>		EET00603			
			<i>Trichomonas vaginalis</i>	XP_001313595 XP_001579634	XP_001325785	XP_001306780	XP_001326479	
		Euglenozoa	<i>Leishmania major</i>	XP_888563	XP_843556	XP_843480	XP_001684803	
<i>Leishmania infantum</i>			XP_001462876	XP_001469259	XP_001469183	XP_001467045		
<i>Trypanosoma cruzi</i>			XP_811195	XP_813702	XP_821644	XP_811561		
		<i>Trypanosoma brucei</i>	XP_828039	XP_827311	XP_827412	XP_845449		
Heterolobosoa	<i>Naegleria gruberi</i>	EFC45514		EFC48750	EFC50559			
Alveolata	Ciliata	<i>Tetrahymena thermophila</i>	XP_001028080	XP_001014455	XP_001031380	XP_001031504	XP_001021620	
		<i>Paramecium tetraurelia</i>	XP_001453141 XP_001450294	XP_001433857	XP_001426605	Acc Num NA	XP_001458727	
	Apicomplexa	<i>Oxytricha trifallax</i>	Acc Num NA	Acc Num NA	Acc Num NA	Acc Num NA		
		<i>Plasmodium yoelii</i>	XP_725062	XP_724743				
		<i>Plasmodium falciparum</i>	XP_001350578	XP_966227				
		<i>Cryptosporidium hominis</i>		XP_668394		XP_668328		
		<i>Babesia bovis</i>						
		<i>Theileria annulata</i>						
		<i>Toxoplasma gondii</i>	XP_002367626	XP_002368820		XP_002371936		
		<i>Blastocystis hominis</i>	CBK23071					
Phaeophyceae	<i>Ecotocarpus siliculosus</i>	Acc Num NA	Acc Num NA	Acc Num NA	Acc Num NA	Acc Num NA	Acc Num NA	
	<i>Phytophthora ramorum</i>	Acc Num NA	Acc Num NA	Acc Num NA	Acc Num NA	Acc Num NA	Acc Num NA	
Oomycota	<i>Phytophthora infestans</i>	EEY65500		EEY69744	EEY66226	EEY60933	EEY59982	
Bacillariophyta	<i>Thalassiosira pseudonana</i>	XP_002290626	XP_002293598	XP_002291443	XP_002292917		XP_002294732	
	<i>Phaeodactylum tricornutum</i>	XP_002182838	XP_002186283	XP_002181491				
	<i>Aureococcus anophagefferens</i>	Acc Num NA		Acc Num NA			Acc Num NA	
Plantae	Vridiplantae	<i>Oryza sativa</i>	AAU10698	NP_001049871	BAG91185	EEE63377	EEC74136	EEE53648
		<i>Arabidopsis thaliana</i>	NP_565433	AAL13436	NP_178543	NP_196175	NP_193884	NP_172146
		<i>Chlamydomonas reinhardtii</i>	XP_001693736	XP_001692946	XP_001698597	XP_001693798	XP_001694053	
		<i>Ostreococcus tauri</i>	AAV68616	AAV68620	AAV68613	AAV68612	CAL52461	AAV68614
		<i>Ostreococcus lucimarinus</i>	XP_001420017	XP_001420379	XP_001420260	XP_001418248	XP_001416348	XP_001418321
		<i>Physcomitrella patens</i>	XP_001760400	XP_001761005	XP_001758242	XP_001770276	XP_001774147	XP_001775896
		<i>Chlorella vulgaris</i>	Acc Num NA	Acc Num NA	Acc Num NA	Acc Num NA	Acc Num NA	Acc Num NA
		<i>Micromonas pusilla</i>	EEH60930	EEH54465	EEH54093	EEH56332	EEH58620	EEH53809
		<i>Selaginella moellendorffii</i>	Acc Num NA	Acc Num NA	Acc Num NA	Acc Num NA	Acc Num NA	Acc Num NA
		Rhodophyta	<i>Cyanidioschyzon merolae</i>	Acc Num NA	Acc Num NA	Acc Num NA	Acc Num NA	Acc Num NA
	<i>Galdieria sulphuraria</i>	Acc Num NA	Acc Num NA	Acc Num NA	Acc Num NA	Acc Num NA	Acc Num NA	
Haptophyta	<i>Emiliania huxleyi</i>	Acc Num NA	Acc Num NA	Acc Num NA	Acc Num NA	Acc Num NA	Acc Num NA	

Tableau 6a. Distribution taxonomique des homologues des sous-unités de l'APC/C (1/2). Le numéro d'accès Genpept est indiqué. Acc Num NA signifie que le numéro d'accès n'existe pas pour cette séquence (c'est le cas lorsque les séquences ont été récupérées par TBLastN).

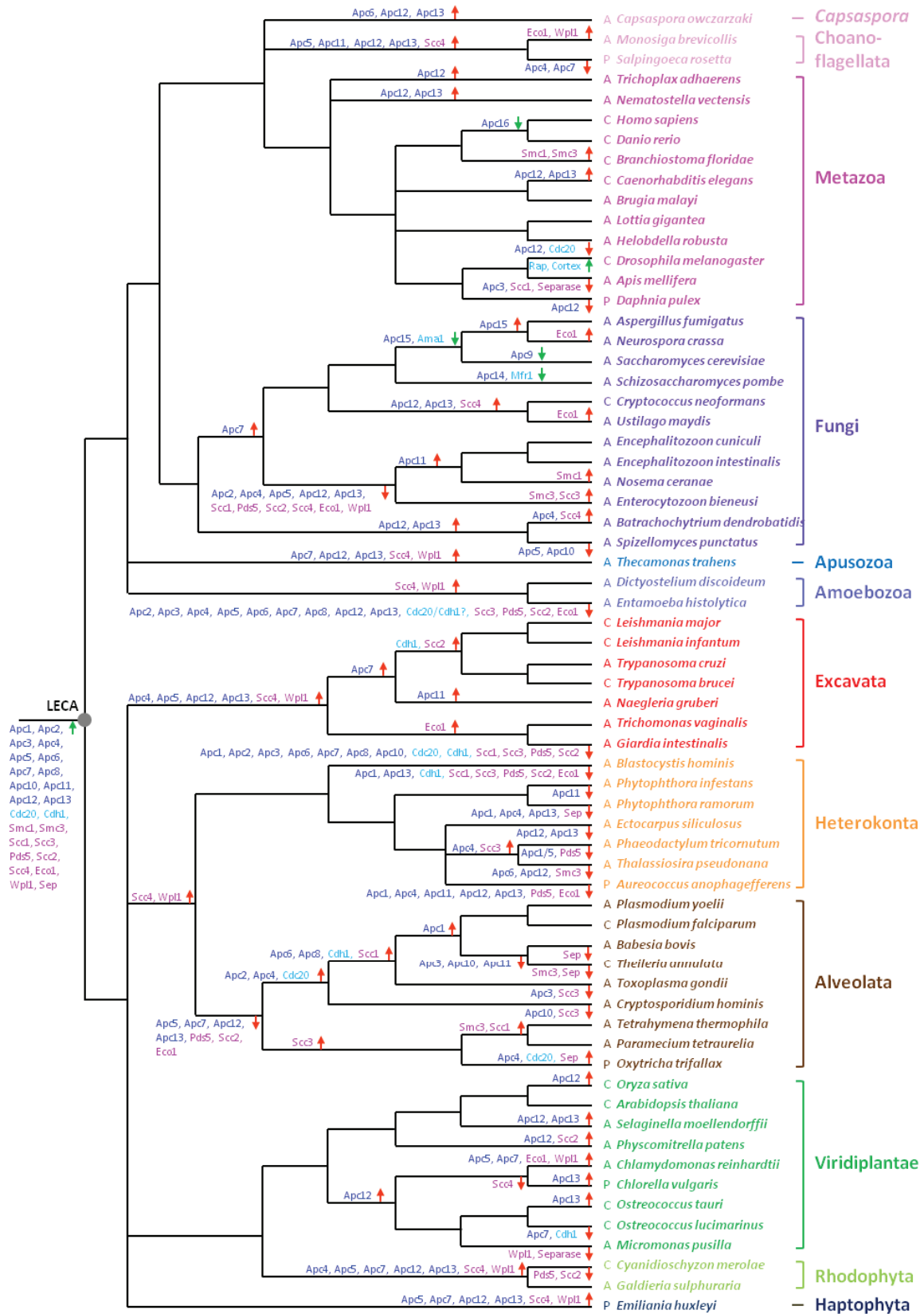
		Core subunits										
		TPR arm and associated proteins										?
		APC8	APC6	APC3	APC7	APC12	APC13	APC16	APC9	APC15	APC14	
Opisthokonta	Choanoflagellata	Monosiga brevicollis	XP_001747393	XP_001745786	XP_001747295	XP_001744593						
		Salpingoeca rosetta	Acc Num NA	Acc Num NA	Acc Num NA							
		Trichoplax adhaerens	XP_002109327	XP_002115543	XP_002110345	XP_002107763		XP_002115489				
		Homo sapiens	AAC70920	NP_001072113	EAV57688	O9UJX3	NP_644815	NP_056206	NP_775744			
		Danio rerio	NP_957227	ABC87083	NP_958857	NP_001107057	NP_001004005	XP_687958	NP_001017878			
		Branchiostoma floridae	XP_002590511	XP_002603477	XP_002598631	XP_002605850	XP_002590834	XP_002590107				
		Metazoa	Drosophila melanogaster	NP_610036	NP_477397	NP_648093 AAA57340	AAL39766	NP_609940	NP_001033954			
			Lottia gigantea	Acc Num NA	Acc Num NA	Acc Num NA	Acc Num NA	Acc Num NA	Acc Num NA			
			Apis mellifera	XP_396943	XP_623316	Acc Num NA	XP_396165	XP_001122028	XP_001120410			
			Caenorhabditis elegans	NP_497203	NP_495712	NP_001021714	NP_500116					
			Brugia malayi	XP_001900333	XP_001900464	XP_001898600	XP_001896953	XP_001896246	XP_001897845			
			Nematostella vectensis	XP_001635295	XP_001637375	XP_001635851	XP_001633228					
			Helobdella robusta	Acc Num NA	Acc Num NA	Acc Num NA	Acc Num NA		Acc Num NA			
			Daphnia pulex	Acc Num NA	Acc Num NA	Acc Num NA	Acc Num NA		Acc Num NA			
		Capsaspora	Capsaspora owczarzaki	Acc Num NA	Acc Num NA	Acc Num NA	Acc Num NA					
			Cryptococcus neoformans	XP_568792	XP_570972	XP_566728						
			Ustilago maydis	XP_756798	XP_760206	XP_758236						
			Aspergillus fumigatus	XP_748088	EDP51110	XP_755196		XP_747203	XP_753351			
			Schizosaccharomyces pombe	NP_593300	NP_593301	CAA30532		NP_593643	NP_595754		NP_595636	
			Saccharomyces cerevisiae	NP_012036	CAY81061	CAY77700		NP_116694	1		NP_013203	
			Neurospora crassa	XP_961540	CAC18281	XP_956598		XP_964014	XP_960666		NP_594611	
			Encephalitozoon cuniculi	NP_585899	NP_585771	NP_586422						
			Enterocytozoon bieneusi H348		XP_002650349	XP_002650295						
			Nosema ceranae BRL01	XP_002995905	XP_002995208	XP_002995969						
			Encephalitozoon intestinalis ATCC 50506	XP_003073104	XP_003072974	XP_003073964						
			Batrachochytrium dendrobatidis	Acc Num NA	Acc Num NA	Acc Num NA	Acc Num NA					
			Spizellomyces punctatus	Acc Num NA	Acc Num NA	Acc Num NA	Acc Num NA					
		Apusozoa	Thecamonas trahens	Acc Num NA	Acc Num NA	Acc Num NA	Acc Num NA					
		Amoebozoa	Dictyostelium discoideum	XP_644962	XP_001134535	XP_638853	XP_629638	XP_629236	XP_642090			
			Entamoeba histolytica									
			Giardia intestinalis									
		Metamonada	Trichomonas vaginalis	XP_001327110	XP_001583340	XP_001312985	XP_001316770					
					XP_001579241		XP_001581999					
	Excavata	Leishmania major	XP_001681866	XP_001684662	XP_001687518							
		Leishmania infantum	XP_001463991	XP_001466919	XP_001463004							
		Euglenozoa	Trypanosoma cruzi	XP_821942	XP_811257	XP_809564						
			Trypanosoma brucei	XP_001219085	XP_845336	XP_823254						
		Heterolobosia	Naegleria gruberi	EFC45000	EFC50413	EFC50492						
		Tetrahymena thermophila	XP_001011597	XP_001026497	XP_001032153							
	Alveolata	Ciliata	Paramecium tetraurelia	XP_001425074	XP_001456952	XP_001434341	XP_001430646	XP_001480999	XP_001481175			
						1#						
			Oxytricha trifallax	Acc Num NA	Acc Num NA	Acc Num NA						
		Apicomplexa	Plasmodium yoelii									
			Plasmodium falciparum			XP_001351576#						
			Cryptosporidium hominis	XP_668598	XP_667497#	XP_668239#						
			Babesia bovis									
			Theileria annulata									
			Toxoplasma gondii									
	Blastocystae	Blastocystis hominis	CBK23690	Acc Num NA								
	Phaeophyceae	Ectocarpus siliculosus	Acc Num NA	Acc Num NA	Acc Num NA	Acc Num NA						
		Oomycota	Phytophthora ramorum	Acc Num NA	Acc Num NA	Acc Num NA	Acc Num NA	Acc Num NA				
			Phytophthora infestans	EEY68460	EEY64266	EEY65802	EEY63812	EEY58244	EEY70446			
			Thalassiosira pseudonana	XP_002291008		XP_002287562	XP_002292854		XP_002290825			
	Bacillariophyta	Phaeodactylum tricornutum	XP_002182840	XP_002176654	XP_002176724	XP_002177784	XP_002177784	XP_002177671	XP_002177799			
		Aureococcus anophagefferens	Acc Num NA	Acc Num NA	Acc Num NA	Acc Num NA						
		Oryza sativa	NP_001047624	EEE58661	EAZ01716	AAU44035		NP_001060376				
		Arabidopsis thaliana	NP_190398	NP_565188	NP_849994	NP_188253	NP_850309	AAN10198	NP_565057			
		Chlamydomonas reinhardtii	XP_001698877	XP_001693404	XP_001701084				XP_001691845			
		Ostreococcus tauri	CAL54508	AAV68615	CAL53302	AAV68618						
		Ostreococcus lucimarinus	XP_001418934	XP_001415756	XP_001418358				XP_001416938			
		Phycomitrella patens	XP_001765887	XP_001784365	XP_001771285	XP_001758114	XP_001753812		XP_001767480			
		Chlorella vulgaris	Acc Num NA	Acc Num NA	Acc Num NA							
		Micromonas pusilla	EEH57771	EEH59359	EEH56226	EEH57219		EEH55073				
		Selaginella moellendorffii	Acc Num NA	Acc Num NA	Acc Num NA	Acc Num NA						
		Cyanidioschyzon merolae	Acc Num NA	Acc Num NA	Acc Num NA							
	Rhodophyta	Galdieria sulphuraria	Acc Num NA	Acc Num NA	Acc Num NA							
	Haptophyta	Emiliania huxleyi	Acc Num NA	Acc Num NA	Acc Num NA							

Tableau 5b. Distribution taxonomique des homologues des sous-unités de l'APC/C (2/2). Le numéro d'accès Genpept est indiqué. Acc Num NA signifie que le numéro d'accès n'existe pas pour cette séquence (c'est le cas lorsque les séquences ont été récupérées par TBLastN).

				Co-activators						
				Mitosis co-activators		Meiotic co-activator				
				CDC20	CDH1	uncertain	Mfr1	Ama1	Rap	Cortex
Choanoflagellata	Monosiga brevicollis	XP_001747841	Acc Num NA							
	Salpingoeca rosetta	Acc Num NA	Acc Num NA							
	Trichoplax adhaerens	XP_002114324	XP_002114322							
	Homo sapiens	NP_001246	NP_057347							
	Danio rerio	CAN88634	XP_001922908							
	Branchiostoma floridae	AAO85336	XP_002603872							
	Drosophila melanogaster	AAO23567	NP_611854						NP_524852	NP_523484
	Loftia gigantea	Acc Num NA	Acc Num NA							
	Apis mellifera	XP_624567	XP_623291							
	Caenorhabditis elegans	NP_495051	NP_496075							
	Brugia malayi	XP_001893850	XP_001899995							
	Nematostella vectensis	XP_001625300	XP_001637197							
			XP_001623504							
		Helobdella robusta	Acc Num NA	Acc Num NA						
		Daphnia pulex	Acc Num NA	Acc Num NA						
		Capsaspora	Acc Num NA	Acc Num NA						
		Capsaspora owczarzaki	Acc Num NA	Acc Num NA						
		Cryptococcus neoformans	XP_572594	XP_572032						
		Ustilago maydis	XP_757794	XP_760064						
		Aspergillus fumigatus	EDP56709	XP_754786					XP_751769	
	Schizosaccharomyces pombe	NP_593161	NP_594674			NP_59508		NP_588462	NP_592334	
	Saccharomyces cerevisiae	NP_011399	EDN61598					NP_011741		
	Neurospora crassa	XP_965019	CAE76124					XP_956209		
	Encephalitozoon cuniculi	NP_584809				NP_597600				
	Enterocytozoon bieneusi H348	XP_002649951	XP_002650590							
	Toxema ceranae ERL01	XP_002965258	XP_002996735							
	Encephalitozoon intestinalis ATCC 50506	XP_003072772	XP_003072646							
	Batrachochytrium dendrobatidis	Acc Num NA	Acc Num NA							
	Spizellomyces punctatus	NA	NA							
		NA	NA							
Apusozoa	Thecamonas trahens	Acc Num NA	XP_002114232							
Amoebozoa	Dictyostelium discoideum	XP_638150	XP_637334							
	Entamoeba histolytica					XP_657064				
	Giardia intestinalis									
	Metamonada									
	Trichomonas vaginalis							XP_001320233	XP_001310738	
								XP_001306976		
Excavata	Leishmania major	XP_001683689								
	Leishmania infantum	XP_001463634								
	Trypanosoma cruzi	XP_819329								
	Trypanosoma brucei	XP_847480								
	Heterolobosoa									
	Naegleria gruberi	EFC37108	EFC42456							
			XP_001031466							
			XP_001028072							
			XP_001012331							
	Tetrahymena thermophila	XP_00102387	XP_001017096							
			XP_001033107							
			XP_001030793							
			XP_001010332							
			XP_001024950							
			XP_001456512							
			XP_001434711							
			XP_001423825							
			XP_001462391							
	Paramecium tetraurelia	XP_001428021	XP_001448596							
			XP_001460408							
			XP_001452380							
			XP_001449473							
			XP_001431487							
			XP_001446793							
	Ocytricha trifallax		Acc Num NA							
	Plasmodium yoelii					XP_728399				
	Plasmodium falciparum					XP_001347545				
	Cryptosporidium hominis					XP_665884				
	Sabesia bovis									
	Theileria annulata									
	Toxoplasma gondii									
Blastocystae	Blastocystis hominis	Acc Num NA								
Phaeophyceae	Ectocarpus siliculosus	NA	Acc Num NA							
		NA	Acc Num NA							
Oomycota	Phytophthora ramorum	Acc Num NA	Acc Num NA							
	Phytophthora infestans	EEY33557	EEY64288							
	Thalassiosira pseudonana	XP_002293859	XP_002296922							
		XP_002297004	XP_002293863							
Bacillariophyta	Phaeodactylum tricornutum	XP_002180582	XP_002179858							
		XP_002176387								
	Aureococcus anophagefferens	Acc Num NA	Acc Num NA							
		EEE69236	Acc Num NA							
	Oryza sativa	CAE03470	NP_001048543							
		EEE57631	NP_001048804							
		NP_568505								
	Arabidopsis thaliana	AA691049	NP_196888							
		NP_198109	NP_194022							
		NP_195053	NP_192929							
		NP_198060								
	Chlamydomonas reinhardtii	XP_001695107	XP_001697615							
	Ostreococcus tauri	AAV68609	CAL56578							
	Ostreococcus lucimarinus	XP_001417358								
		XP_001757257	XP_001769243							
		XP_001775444	XP_001774759							
	Physcomitrella patens	XP_001768400	XP_001774759							
		XP_001753296	XP_001769827							
		XP_001765142	XP_001773274							
	Chlorella vulgaris	Acc Num NA	Acc Num NA							
	Micromonas pusilla	EEH56365	EEH57236							
		NA								
	Selaginella moellendorffii	NA	Acc Num NA							
		NA								
	Rhodophyta	Acc Num NA	Acc Num NA							
	Cyanidioschyzon merolae	Acc Num NA	Acc Num NA							
	Galdieria sulphuraria	Acc Num NA	Acc Num NA							
Haptophyta	Emiliania huxleyi	Acc Num NA	Acc Num NA							

Tableau 7. Distribution taxonomique des homologues des adaptateurs de l'APC/C. Le numéro d'accès Genpept est indiqué. Acc Num NA signifie que le numéro d'accès n'existe pas pour cette séquence (c'est le cas lorsque les séquences ont été récupérées par TBLastN).

Figure 60 (page suivante). Histoire évolutive des sous-unités de l'APC/C, des adaptateurs, et ses cibles principales. L'arbre de référence est raciné entre unikontes et bikontes. L'état d'avancement du séquençage de chaque génome est indiqué : "C" : génome complet, «A» : en cours d'assemblage, "P" : en cours de séquençage. Les pertes (flèches rouges) et gains (flèches vertes) de composants ont été inférés à partir des analyses phylogénétiques. Les sous-unités de l'APC/C sont représentées en bleu foncé, les adaptateurs en bleu clair et les cibles principales en rose. Pour chaque lignée, les gains et les pertes de protéines sont indiquées par des flèches et vert foncé rouge, respectivement.



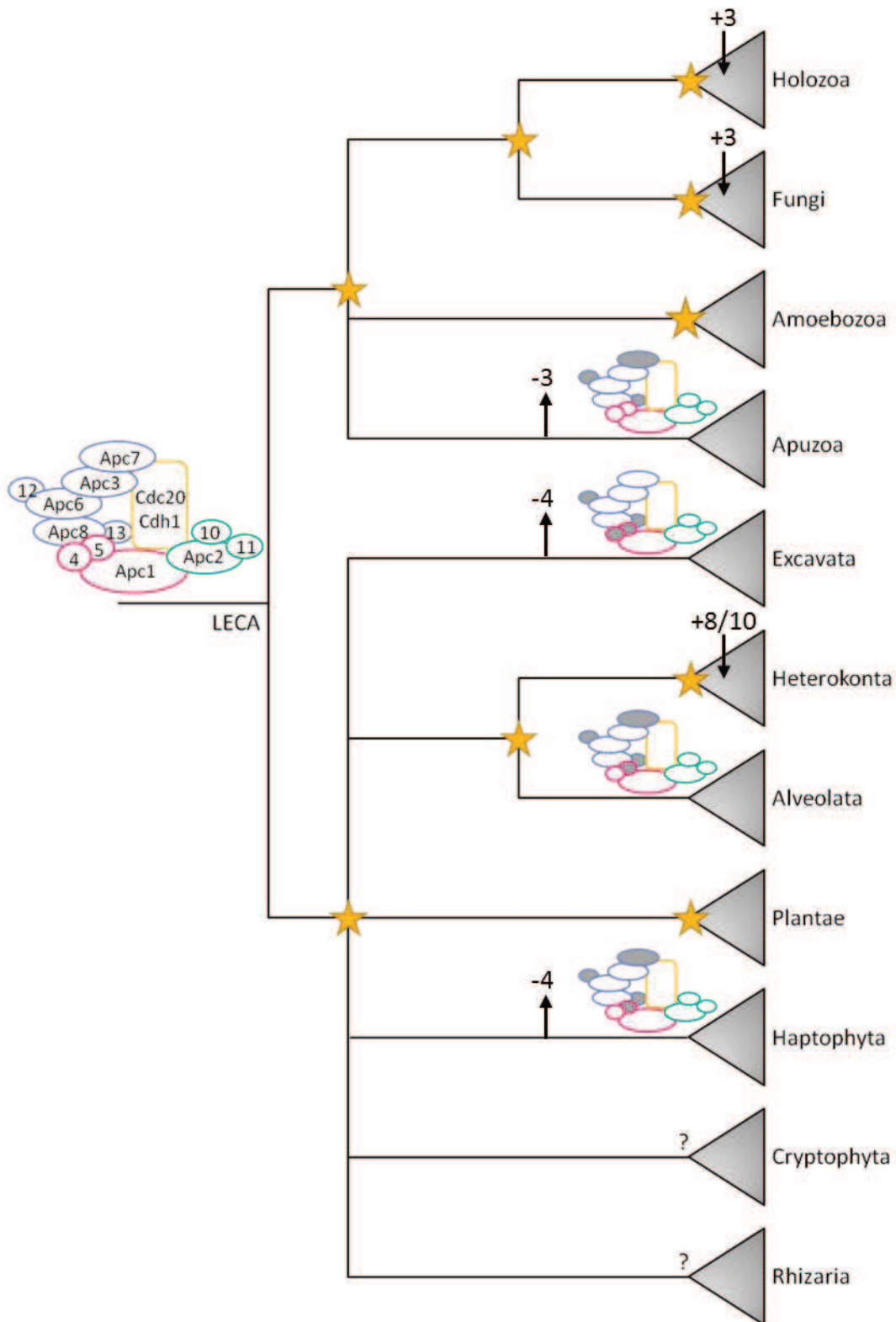


Figure 61. Inférence de la composition de l'APC/C et ses adaptateurs dans l'ancêtre de chaque grande lignée eucaryote. Les ancêtres marqués d'une étoile sont présumés avoir les mêmes sous-unités que LECA. Les composants gris sont ceux présumés avoir été perdus alors qu'ils existaient chez LECA. Il est intéressant de remarquer la convergence des pertes dans des groupes distants (ex : Apc5, Apc12, et Apc13). Sont également reportés les gains d'adaptateurs spécifiques de certaines lignées.

2.4.2 Plusieurs composants sont issus de duplications ou d'innovations récentes

Contrairement à l'étude des structures impliquées dans la cytokinèse, pour laquelle un seul protéome était disponible, l'APC/C a été étudié et décrit chez diverses lignées d'animaux et de champignons, ce qui a permis de mettre en évidence quelques composants spécifiques de ces lignées.

Par exemple, d'autres protéines accessoires du **bras TPR** (Apc9 et Apc15) ne sont présentes que dans quelques lignées d'ascomycètes ce qui suggère qu'elles sont issues d'innovations récentes postérieures à la diversification des champignons (Figure 60). Apc16 paraît, elle aussi, avoir une origine extrêmement récente, car elle n'est présente que chez les deux animaux représentant des Gnathostomata (Figure 60 et Tableau 6).

Par ailleurs, tous les **adaptateurs** autres que Cdc20 et Cdh1 décrits chez les animaux, les champignons et les plantes résultent de duplications de ces protéines qui se sont produites très récemment dans certaines lignées eucaryotes (Tableau 7). Par exemple, les orthologues de Rap et Cortex sont spécifiques des drosophiles, et sont issus de la duplication de Cdh1 et Cdc20 respectivement, tandis que Ama1 et Mfr1 sont limités aux Gnathostomata, et sont issus d'une duplication dans cette lignée de Cdc20 et Cdh1. Chez les plantes vertes, en accord avec la littérature, de multiples copies de Cdc20 et Cdh1 sont retrouvées, et sont issues elles-aussi de duplications spécifiques de ces lignées.

Il a été surprenant d'observer une expansion encore plus importante du répertoire des adaptateurs chez les deux ciliés *Tetrahymena thermophila* et *Paramecium tetraurelia*, pour lesquels j'ai identifié respectivement 8 et 10 exemplaires de Cdh1 qui n'ont pas encore été décrits dans la littérature. Cette expansion est d'autant plus surprenante que la troisième espèce de ciliés, *Oxytricha trifallax*, ne contient qu'une seule copie de Cdc20 et une de Cdh1. Cela soulève des questions très intéressantes sur le rôle de ces multiples adaptateurs chez ces unicellulaires qu'il serait nécessaire d'étudier expérimentalement.

2.4.3 Certaines lignées ont subi des pertes importantes de composants

Alors que les 12 sous-unités de l'APC/C inférées comme étant présentes chez LECA ont été conservées dans l'ensemble des grands groupes eucaryotes, quelques lignées présentent des pertes importantes de composants (Figure 60).

2.4.3.1 Vers la définition d'un APC/C minimal ?

Une étude récemment publiée n'avait détecté, par des méthodes automatiques, qu'un seul composant (Apc10, localisé dans le bras catalytique) chez *Encephalitozoon cuniculi* (Seidl and Schultz, Schultz, 2009). En revanche, mes analyses m'ont également permis de retrouver, chez les quatre quatre microsporidies étudiées, un composant du complexe structural (Apc1), ainsi que trois sous-sous-unités du bras TPR (Apc3, Apc6 et Apc8) et deux adaptateurs (Cdc20 et Cdh1) (Tableau 6 et Tableau 7). Le nombre de composants conservés est assez faible, néanmoins, la conservation d'au moins un élément de chaque partie fonctionnelle de l'APC/C suggère qu'une version minimaliste de l'APC/C pourrait exister chez ces champignons parasitiques (

Figure 61). Il serait intéressant de tester expérimentalement cette hypothèse.

En parallèle, un certain nombre de cibles ont été perdues, mais les protéines parmi les plus importantes fonctionnellement ont été conservées (Smc1, Smc3, Scc1 et la séparase), ce qui suggère que l'ensemble du système de contrôle de la ségrégation des chromosomes pourrait toujours être fonctionnel (cf. ci-après).

2.4.3.2 Pertes massives

Bien que 12 sous-unités soient codées par le génome de *Dictyostelium discoideum* (Amoebozoa), (Amoebozoa), l'absence de tous ces composants à l'exception de trois (Apc1, Apc10, Apc11) chez le deuxième représentant de ce groupe (le parasite anaérobie *Entamoeba histolytica*) suggère une perte massive de composants APC/C chez cette espèce. C'est également très probablement le cas pour *Giardia intestinalis*, qui ne possède qu'une sous-unité, alors que 8 sont inférées chez l'ancêtre des Excavata (

Figure 61). De la même façon, aucun adaptateur n'a été détecté chez cette espèce. Dans ces deux cas, et contrairement à la situation chez les microsporidies, l'absence de l'ensemble des protéines composant le bras TPR ne semble pas compatible avec un système APC/C pleinement opérationnel.

La situation est encore plus extrême chez les Apicomplexa, une des trois principales lignées d'alvéolés avec les ciliés et les dinoflagellés. Alors que les trois ciliés (*P. tetraurelia*, *T. thermophila* et *Oxytricha trifallax*) ont conservé un grand nombre de composants présents chez LECA, la plupart ou la totalité d'entre eux ont été perdus chez les apicomplexes (Tableau 6 et Tableau 7). Chez *Babesia*, *Theileria* et *Toxoplasma*, aucun adaptateur n'a été détecté non plus. Toutefois, il est important de noter que lorsque des orthologues existent dans ces lignées, ils présentent des séquences très divergentes de celles retrouvées dans les autres lignées eucaryotes. Il est donc possible que certains orthologues n'aient pas été détectés en raison de leur extrême degré de divergence. Dans tous les cas, les éventuelles pertes massives ou l'extrême divergence des composants de l'APC/C suggèrent que des changements importants se sont récemment produits dans ces groupes de parasites.

Pour certaines espèces d'apicomplexes, la réduction de l'APC/C pourrait être liée à leur mécanisme de division cellulaire particulier. C'est notamment le cas de *Theileria* qui, lors de la division de la cellule hôte, assure sa propagation dans les deux cellules filles en cooptant des composants de la machine de division de la cellule hôte (von Schubert et al., 2010). Le parasite se fait passer pour un chromosome de l'hôte, en utilisant les microtubules de la cellule hôte afin d'être tiré vers les extrémités de la cellule en division. Par la suite, *Theileria* recrute la protéine Plk1 (qui détermine la position du site de division) et la fixe à sa surface. La cellule hôte se divise alors exactement à l'endroit déterminé par Plk1, divisant ainsi le parasite qui se retrouve dans les deux cellules-filles (Figure 62). On peut alors se demander si, de la même façon, ces parasites pourraient ne pas recruter également certains composants de l'APC/C de la cellule qu'ils parasitent.

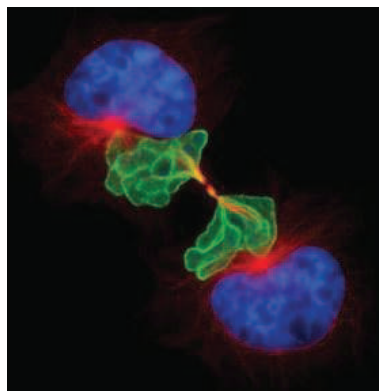


Figure 62. Durant la cytokinèse, *Theileria annulata* (en vert) est également distribuée dans chacune des cellules filles grâce à la cooptation de Plk1. Les noyaux sont en bleu et les microtubules en rouge.
Image : D. Dobbelaere, University of Bern, Switzerland.

2.4.4 La majorité des cibles existaient chez LECA

Les résultats précédents permettent d'inférer que la majorité des composants de l'APC/C était présente chez LECA et avaient la même composition en domaines fonctionnels que les composants actuels. Une hypothèse séduisante serait qu'une structure ayant la même fonction biochimique (*i.e.* fonction d'ubiquitine ligase) existait déjà chez cet ancêtre. Néanmoins, rien n'indique que cette structure avait la même fonction cellulaire (*i.e.* la régulation du cycle cellulaire). En revanche, si ce rôle cellulaire était bien le même, alors il est attendu d'inférer une grande partie de ses cibles comme étant aussi présentes chez cet ancêtre. Sous cette hypothèse, lorsque le système ancestral a été conservé dans une lignée (c'est-à-dire que de nombreux composants de l'APC/C sont présents), on s'attend à ce que les cibles soient également présentes dans cette lignée.

Pour tester cette hypothèse, j'ai réalisé les analyses phylogénétiques des cibles principales (directes ou indirectes) de l'APC/C. Les analyses phylogénétiques n'ont pas été concluantes dans le cas des cyclines A et B et de Cdk-1 et -2 parce que ces protéines appartiennent à de très grandes familles multigéniques avec une histoire évolutive complexe, avec de nombreuses duplications et pertes, ce qui empêche de conclure précisément sur leur origine.

Pour les autres cibles de l'APC/C, mes analyses ont permis d'inférer que la séparase ainsi que les 9 sous-unités du complexe cohésine étaient présentes dans LECA (c'est-à-dire les quatre composants du cœur, Smc1, Smc3, Scc1 et Scc3; les deux protéines du loading complex, Scc2 et Scc4; la protéine responsable de l'établissement de la cohésion, Eco1; et les deux protéines nécessaires au maintien de la cohésion, Pds5 et Wpl1) (Tableau 8). La seule exception est représentée par les sécurines qui ne sont présentes que chez certains champignons et animaux. De plus, elles ont abusivement été nommées de la même façon dans ces deux groupes car elles remplissent la même fonction (Zou et al., 1999), à savoir, la prévention de la séparation des deux chromatides-sœurs, mais les sécurines d'animaux ne sont en réalité pas homologues à celles des champignons. Cependant, il est intéressant de noter qu'elles se lient toutes deux aux séparases, qui, elles, sont bien homologues dans les deux lignées. Ceci suggère un remplacement non-homologue des sécurines dans un de ces deux groupes. Le fait de ne retrouver des sécurines que dans ces deux groupes alors que la séparase est conservée dans l'ensemble des grands groupes eucaryotes sous-entend soit l'existence d'analogues fonctionnels non homologues, soit que des mécanismes sécurine-indépendants suffisent à contrôler l'activité des séparases, comme suggéré par les expériences réalisées chez le Xénope (cf. p.91).

Chez les organismes ayant perdu beaucoup de composants de l'APC/C, les cibles ont elles aussi été perdues massivement : *Giardia* ne possède plus que les trois composants les plus importants du système cohésine/séparase (Smc1, Smc3 et la séparase). *Entamoeba* possède ces trois composants, et également Scc1, un autre composant du cœur du système. Les Apicomplexa, quant à eux, possèdent généralement aussi Scc3, le quatrième et dernier composant du cœur. Ainsi, malgré les pertes qui peuvent être importantes parmi les cibles, les principaux acteurs de la régulation de la ségrégation des chromosomes qui sont ciblés par l'APC/C sont conservés. Ceci suggère que leur dégradation nécessite d'être contrôlée, soit par l'APC/C (éventuellement via la cooptation de composants de l'hôte), soit peut-être, par une autre ubiquitine ligase.

De façon générale, dans chaque lignée, le nombre de cibles conservées est assez bien corrélé avec le nombre de composants de l'APC/C, ce qui renforce l'hypothèse selon laquelle l'ensemble du système formé par l'APC/C et ses cibles était déjà présent chez LECA.

2.4.4.1 L'APC/C est le résultat d'innovations eucaryotes

Deux composants du complexe cohésine, Smc1 et Smc3, sont largement retrouvés chez les archées et les bactéries, et sont impliqués dans la condensation et la ségrégation des chromosomes. A ces exceptions près, de façon similaire à ce qui a été observé pour le midbody, aucun homologue des autres sous-unités n'a été détecté chez les procaryotes. Ceci suggère que la majorité des composants de l'APC/C, de ses adaptateurs ainsi que de ses principales cibles est le résultat d'innovations eucaryotes qui ont eu lieu après la divergence de ce domaine, mais avant la diversification donnant les lignées eucaryotes actuelles.

		Core complex			Loading complex		Cohesion establishment		Maintenance		Dissolution		
		Smc1	Smc3	Sec1/Rad21	Sec3	Sec2	Sec4	Eco1	Pds5	Wpl1	Separase	Securin	
Chaozooflagellata	<i>Monosiga brevicollis</i>	Acc Num NA	XP_001749007	XP_001743156	XP_001744374	XP_001749937			XP_001745480		XP_001747241		
	<i>Salpingoeca rosetta</i>	Acc Num NA	Acc Num NA	Acc Num NA	Acc Num NA	Acc Num NA			Acc Num NA	Acc Num NA	Acc Num NA		
	<i>Trichoplax adhaerens</i>	Acc Num NA	XP_002116802	Acc Num NA	XP_002111813	XP_002114243	Acc Num NA		XP_002110319	XP_002112303	XP_002112136	?	
	<i>Homo sapiens</i>	NP_006297	NP_005436	NP_00130038	NP_005984 NP_005853	DAAO5331	NP_056144		NP_443143 NP_001017430	NP_001093870	NP_005890	NP_036423	NP_004210
	<i>Danio rerio</i>	AAT69048	NP_999854	NP_955889	NP_006994 XP_888812	XP_001920203	XP_001334031		XP_002863315 NP_001003872	AAH55602	XP_688905	XP_001337869	XP_695906
	<i>Branchiostoma floridae</i>			XP_002602455	XP_002603307	XP_002587265	XP_002609698		XP_002586734	XP_002811624	XP_002810333	XP_002607627	
	<i>Drosophila melanogaster</i>	NP_651211	NP_523374	AF186473_1	CAA74854	ABX00777	NP_650428		NP_648106 AAL28660	NP_810719	XP_002810333	NP_523935	NP_476678
	<i>Lottia gigantea</i>	Acc Num NA	Acc Num NA	Acc Num NA	Acc Num NA	Acc Num NA	XP_394668		Acc Num NA	Acc Num NA	Acc Num NA	Acc Num NA	
	<i>Apis mellifera</i>	XP_395059	XP_393700		XP_001123117	XP_001120878	XP_394668		XP_624624	XP_623860	XP_395593	XP_395593	
	<i>Caenorhabditis elegans</i>	NP_001040958	NP_499453	NP_509282	NP_506193	NP_493687	NP_492238		NP_498625	NP_497665	NP_500566	NP_491160	
Opisthokonta	<i>Brugia malayi</i>	XP_001884059	XP_001899737	XP_001895596	XP_001901686	EDP30362	XP_001892554		XP_001897363	XP_001896335	XP_001892742	XP_001894670	
	<i>Nematostella vectensis</i>	XP_001641659	XP_001626238	Acc Num NA	Acc Num NA	Acc Num NA	XP_001624621		XP_001639589	XP_001641695	XP_001626896	XP_001635666	
	<i>Helobdella robusta</i>	Acc Num NA	Acc Num NA	Acc Num NA	Acc Num NA	Acc Num NA	Acc Num NA		Acc Num NA	Acc Num NA	Acc Num NA	Acc Num NA	
	<i>Daphnia pulex</i>	Acc Num NA	Acc Num NA	Acc Num NA	Acc Num NA	Acc Num NA	Acc Num NA		Acc Num NA	Acc Num NA	Acc Num NA	Acc Num NA	
	<i>Capsaspora owczarzi</i>	Acc Num NA	Acc Num NA	Acc Num NA	Acc Num NA	Acc Num NA	Acc Num NA		Acc Num NA	Acc Num NA	Acc Num NA	Acc Num NA	
	<i>Cryptococcus neoformans</i>	XP_568851	XP_775936	XP_570356	XP_567136	XP_568838	XP_568838		XP_568915	XP_567466	XP_778346	XP_569001	
	<i>Ustilago maydis</i>	XP_750390	XP_760536	XP_758738	XP_758200	XP_761813				XP_759887	XP_760279	XP_759409	
	<i>Aspergillus nidulans</i>	EDP52912	XP_755741	EDP54236	EDP55114	XP_749676	XP_747813		XP_755006	EDP55189	EAL92045	NP_753688	
	<i>Schizosaccharomyces pombe</i>	NP_586049	NP_593260	NP_588151	NP_001281	NP_012238	NP_582917		NP_596778	NP_593355	NP_595195	NP_587903	NP_010388
	<i>Saccharomyces cerevisiae</i>	NP_116647	NP_012461	NP_010281	NP_010281	NP_010281	EDM63124		NP_116683	NP_013793	NP_010297	EDM61688	NP5859604
Fungi	<i>Neurospora crassa</i>	XP_961409	XP_957879	XP_964217	XP_960650	XP_961787	XP_961389		XP_956688	XP_955967	XP_955969	XP_955963	
	<i>Encephalitozoon cuniculi</i>	NP_584776	XP_955745		NP_584770							NP_585963	
	<i>Enterocytozoon bieneisi</i> H348	XP_002649733										XP_001827760	
	<i>Nosema ceranae</i> BRL01		XP_00295833		XP_00298603							XP_002986251	
	<i>Encephalitozoon intestinalis</i> ATCC 50506	XP_003072739	XP_003073682		XP_003072732							XP_003073153	
	<i>Batrachochytrium dendrobatidis</i>	Acc Num NA	Acc Num NA	Acc Num NA	Acc Num NA	Acc Num NA	Acc Num NA		Acc Num NA	Acc Num NA	Acc Num NA	Acc Num NA	
	<i>Spizellomyces punctatus</i>	Acc Num NA	Acc Num NA	Acc Num NA	Acc Num NA	Acc Num NA	Acc Num NA		Acc Num NA	Acc Num NA	Acc Num NA	Acc Num NA	
	<i>Thecamonas trahans</i>	Acc Num NA	Acc Num NA	Acc Num NA	Acc Num NA	Acc Num NA	Acc Num NA		Acc Num NA	Acc Num NA	Acc Num NA	Acc Num NA	
	<i>Dictyostelium discoideum</i>	XP_629977	XP_643274	XP_642955	XP_638863	XP_639229			XP_641425	XP_643624		XP_635002	
	<i>Amoeba histolytica</i>	XP_656581	XP_655216	XP_657283								XP_651118	
Amoebozoa	<i>Giardia intestinalis</i>	EET00911	EET02057									CAD67965	
	<i>Metamonada</i>												
	<i>Trichomonas vaginalis</i>	XP_001329543	XP_001319807	XP_001329298	ABC61972	XP_001581890			XP_001329031			XP_001582914	
	<i>Leishmania major</i>	XP_843456	XP_001681030	XP_001687586	XP_001686731				XP_001682791	XP_001681555		XP_001682914	
	<i>Leishmania infantum</i>	XP_001489160	XP_001463342	XP_001463072	XP_001489747				XP_001465213	XP_001463386		XP_001465208	
	<i>Trypanosoma cruzi</i>	XP_817950	XP_815055	XP_819913	XP_809775				XP_818808	XP_821239		XP_809313	
	<i>Trypanosoma brucei</i>	XP_827441	XP_845021	XP_846310	XP_822844				XP_001218988	XP_828716		CBH49022	
	<i>Naegleria gruberi</i>	XP_002682487	XP_002670278	XP_002678151	XP_002673603	XP_002683017			XP_002680835	XP_002671795		Acc Num NA	
	<i>Tetrahymena thermophila</i>	XP_001025504										XP_001013249	
	Excavata	<i>Paramecium tetraurelia</i>	XP_001426830										XP_001458342
<i>Oryzichia trifallax</i>		?	?	?									
<i>Plasmodium yoelii</i>		XP_726100	XP_725208		XP_730250							XP_726982	
<i>Plasmodium falciparum</i>		XP_001347988	XP_001351448		XP_001346712							XP_002808832	
<i>Cryptosporidium hominis</i>		XP_666991	XP_665748	XP_668359								XP_665902	
<i>Babesia bovis</i>		XP_001610390	XP_001611424		XP_001609890							Acc Num NA	
<i>Theileria annulata</i>		XP_954442			XP_952886								
<i>Toxoplasma gondii</i>		XP_002388310	AAZ57430									EEE32077	
<i>Blastocystis hominis</i>		CBK21095	CBK23679									CBK21518	
<i>Ectocarpus siliculosus</i>		CBN77803	CBN79764	CBJ33525	CBJ33097	CBJ31249			CBJ31471	CBJ25477		CBJ25688	
Phaeophyceae	<i>Phytophthora ramorum</i>	Acc Num NA	Acc Num NA	Acc Num NA	Acc Num NA	Acc Num NA		Acc Num NA	Acc Num NA				
	<i>Phytophthora infestans</i>	XP_002980013	XP_002967684	XP_002902993	XP_002904221	XP_002906797		XP_002902720	XP_002986972		XP_002906565		
	<i>Thalassiosira pseudonana</i>	XP_002296149		XP_002292911		XP_002287148		XP_002287134	XP_002291327		XP_002286533		
	<i>Phaeodactylum tricoratum</i>	XP_002177608	XP_002184568	XP_002178341		XP_002185061		XP_002177583			XP_002182954		
	<i>Aureococcus anophagefferens</i>	Acc Num NA	Acc Num NA	Acc Num NA	Acc Num NA	Acc Num NA					Acc Num NA		
	<i>Oryza sativa</i>	NP_001097404	NP_001045803	NP_001053148	NP_001054841	EEE68449	CAD40351		NP_001055381	EEE55732	NP_001064956	ECC74078	
	<i>Arabidopsis thaliana</i>	NP_191027	NP_001077968	NP_568586	NP_566119	NP_197056	BAF00935		NP_194868	NP_177883	NP_176236	NP_194028	
	<i>Chlamydomonas reinhardtii</i>	XP_001700607	XP_001698278	XP_001696730	XP_001698607	XP_001697971						XP_001698069	
	<i>Ostreococcus tauri</i>	XP_003078284	XP_003081202	XP_003079934	XP_003074534	XP_003075000	XP_003082234		XP_003081728	XP_003074597	XP_003081951	XP_003074856	
	<i>Ostreococcus lucimarinus</i>	XP_001416713	XP_001419553	XP_001418469	XP_001415613	XP_001416473	XP_001420595		XP_001420193	XP_001415641	XP_001420300	XP_001418421	
Plantae	<i>Physcomitrella patens</i>	XP_001787264	XP_001764554	XP_001765550	XP_001776000			XP_001773150	XP_001763476	XP_001778595	XP_001781920	XP_001759110	
	<i>Chlorella vulgaris</i>	Acc Num NA	Acc Num NA	Acc Num NA	Acc Num NA	Acc Num NA		Acc Num NA	Acc Num NA	Acc Num NA	Acc Num NA	Acc Num NA	
	<i>Micromonas pusilla</i>	XP_003055286	XP_003056807	XP_003062115	XP_003056489	XP_003056333	XP_003055504		XP_003063893	XP_003062863			
	<i>Selaginella moellendorffii</i>	XP_002990356	Acc Num NA	XP_002985553	XP_002986356	XP_002981324	Acc Num NA		XP_002980198	XP_002980499	XP_002977658	XP_002975469	
	<i>Cyanidioschyzon merolae</i>	Acc Num NA	Acc Num NA	CML311C	CMO331C				CMT193C			Acc Num NA	
	<i>Galdieria sulphuraria</i>	Acc Num NA	Acc Num NA	?	Acc Num NA	?			Acc Num NA	Acc Num NA		Acc Num NA	
	<i>Emiliania huxleyi</i>	Acc Num NA	?	Acc Num NA	Acc Num NA	Acc Num NA			?	Acc Num NA		Acc Num NA	

Tableau 8. Distribution taxonomique des orthologues des cibles principales de l'APC/C. Le numéro d'accès Genpept est indiqué. Acc Num NA signifie que le numéro d'accès n'existe pas pour cette séquence (c'est le cas lorsque les séquences ont été récupérées par TblastN). Les points d'interrogation indiquent une incertitude sur la présence ou non d'un orthologue

2.4.5 Conclusion

Cette analyse supporte l'hypothèse que la plupart des composants de l'APC/C, ses adaptateurs et ses principales cibles étaient déjà présents dans LECA. Il est donc tentant de conclure qu'un complexe similaire à celui qui régule le cycle cellulaire des eucaryotes actuels était déjà présent dans cet organisme ancestral.

Cette hypothèse est appuyée par le fait que quasiment tous les orthologues sont composés des mêmes domaines fonctionnels que ceux qui ont été expérimentalement prouvés comme faisant partie de l'APC/C ou de ses cibles, suggérant que tous les orthologues identifiés réalisent des fonctions moléculaires similaires.

En outre, les données expérimentales chez les plantes et l'excavé *T. brucei* ont montré que la plupart des composants identifiés dans notre analyse font partie de ou sont ciblés par l'APC/C. Ceci sous-entend donc que, non seulement la fonction moléculaire est conservée entre orthologues, mais que la fonction biologique l'est aussi. Cela suggère également que ces fonctions ont été héritées depuis LECA et par conséquent qu'un APC/C fonctionnel et simili-moderne existait chez cet ancêtre.

Ceci suppose aussi que les orthologues que j'ai identifiés dans des lignées pour lesquelles aucune donnée expérimentale n'est disponible sont probablement impliqués dans le contrôle du cycle cellulaire et, par conséquent, constituent des cibles intéressantes pour des travaux expérimentaux.

Bien que l'origine de la plupart des composants remonte à LECA et que la majorité d'entre eux a été conservée au cours de l'évolution des eucaryotes, des gains de composants ont pu être observés. Cependant, la plupart d'entre eux résultent de duplications de gènes codant pour des adaptateurs qui se sont produites de manière indépendante dans différentes lignées. Ceci est cohérent avec les récentes études sur les fonctions de l'APC/C chez certains eucaryotes (par exemple, son rôle dans la fonction neuronale ou, chez les plantes à fleurs, dans la progression de l'endocycle dans l'endosperme). Ceci suggère que la plupart des innovations depuis LECA a concerné l'acquisition de nouvelles fonctions de régulation, en augmentant le répertoire des adaptateurs, même s'il ne faut pas exclure la possibilité que des adaptateurs présents en un seul exemplaire dans certaines lignées soient capables d'interagir avec un plus large spectre de cibles que leurs homologues présents en multiples copies.

A l'opposé, des événements de pertes convergentes se sont produits secondairement dans différentes lignées, comme les Apicomplexa, *G. lamblia* ou *E. histolytica*. Les raisons expliquant ces pertes massives ne sont pas claires, et il n'est pas impossible que, pour un certain nombre de cas, une accélération extrême de leur taux d'évolution ait empêché leur détection. Cette hypothèse est étayée par les divergences importantes que présentent les quelques composants retrouvés chez les Apicomplexa. Cependant, chez des organismes qui évoluent également rapidement (tels que *E. cuniculi* et *T. vaginalis*), j'ai retrouvé les homologues sans difficulté. Ceci suggère donc, plutôt, que la plupart des éléments qui manquent chez les Apicomplexa, *G. lamblia* et *E. histolytica* reflètent des pertes réelles. Dans ce cas, on peut se demander si les éléments perdus ont été remplacés par des protéines non homologues qui jouent le même rôle ou si ces parasites recrutent les composants de leurs hôtes. Dans les deux cas, des études expérimentales dans ces lignées de parasites seront utiles pour élucider la nature de leur APC/C ou même de découvrir potentiellement des systèmes différents impliqués dans le contrôle du cycle cellulaire qui pourraient constituer de nouvelles cibles médicamenteuses.

Il est intéressant de faire le parallèle avec les résultats de l'analyse du midbody, qui montrent une tendance semblable d'évolution réductrice dans ces lignées. En effet, *E. histolytica*, *G.*

lamblia et les Apicomplexa ont subi des pertes massives de composants anciennement présents chez LECA et impliqués dans la cytotinèse. Les résultats combinés de ces deux analyses suggèrent des changements importants dans les différentes étapes du cycle cellulaire chez ces parasites.

3

Evolution du système SRX/PRX

3 Evolution du système SRX/PRX

Au cours de ma 3e année de thèse, j'ai eu l'opportunité de participer à l'étude d'un système enzymatique impliqué dans la réponse au stress oxydant dont l'histoire évolutive atypique suggère des transferts horizontaux entre le dernier ancêtre commun des cyanobactéries et celui des Eucarya.

Ce travail a été fait en collaboration avec Amel Latifi et Cheng-Cai Zhang de l'équipe "Différenciation cellulaire microbienne et signalisation" (Laboratoire de Chimie Bactérienne, CNRS, Marseille).

3.1 Introduction

3.1.1 Les peroxirédoxines

Les organismes aérobies ont développé des systèmes de défense efficaces et spécifiques qui permettent de détoxifier rapidement les agents oxydants dangereux comme le peroxyde d'hydrogène (H_2O_2) ou l'ion superoxyde ($O_2^{\cdot-}$). En particulier, ces organismes sont équipés d'une large gamme de molécules antioxydantes, comprenant les superoxyde-dismutases, les peroxydases et les catalases. Les peroxyrédoxines (Prx), des protéines antioxydantes catalysant la réduction d'une vaste gamme de peroxydes différents (H_2O_2 , hydroperoxydes d'alkyle et l'ion peroxydite) ont été identifiées chez les plantes à la fin des années 1990 (Baier and Dietz, 1996).

Les Prx forment une superfamille de peroxydases non hémiques à cystéine catalytique (ou C_p pour cystéine peroxydasique), et sont présentes dans les trois domaines du vivant (Brehelin et al., 2003). Cette famille est divisée en trois classes distinctes de par leur mécanisme de réduction des peroxydes et, en particulier, de par leur mode de recyclage (Latifi et al., 2009) : les Prx à deux cystéines typiques, les Prx à deux cystéines atypiques, et les Prx à une cystéine.

3.1.1.1 Les peroxyrédoxines à deux cystéines typiques

Les peroxyrédoxines à deux cystéines typiques (2-Cys Prx typiques) forment des homodimères de sous-unités assemblées tête-bêche (Figure 63).

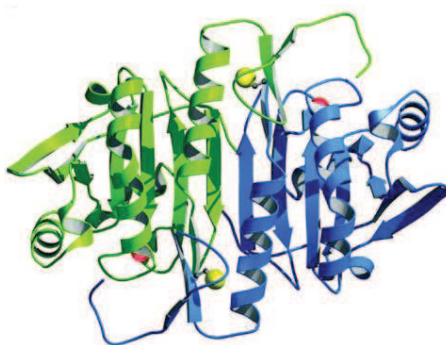


Figure 63. Structure tridimensionnelle d'un homodimère de 2-Cys Prx typiques de *Salmonella typhimurium* (gamma-protéobactérie) (Wood et al., 2003a). Les monomères sont représentés en bleu et vert. Le soufre (en jaune) de la chaîne latérale de la cystéine réactive d'une sous-unité réagit avec l'oxygène (en rouge) de la chaîne latérale de la cystéine de l'autre sous-unité.

Après réaction avec une molécule de H_2O_2 , la cystéine peroxydase (C_p-SH) d'une sous-unité est oxydée en un acide sulfénique (C_p-SOH) (Figure 64A). Cet acide sulfénique réagit alors avec une cystéine (dite cystéine de régénération, ou C_R) située sur l'extrémité C-terminale de la seconde sous-unité. Cela aboutit à la formation d'un pont disulfure entre ces deux sous-

unités et à la libération d'une molécule d'eau (Figure 64A). Les 2-Cys Prx ainsi oxydées perdent leur activité enzymatique et doivent ensuite être régénérées *via* la réduction du pont disulfure par un donneur d'électrons tel que la thiorédoxine ou la glutarédoxine (Figure 64A). Ce mécanisme a été étudié chez les bactéries (gamma-protéobactéries et firmicutes) et les eucaryotes (animaux et plantes vertes essentiellement) (revue de (Poole et al., 2000)).

L'assemblage des molécules de Prx se fait principalement sous forme de dimères ou de décimères (soit cinq dimères) (Figure 65), la forme décimérique étant la plus active (Parsonage et al., 2005; Wood et al., 2002; Wood et al., 2003b).

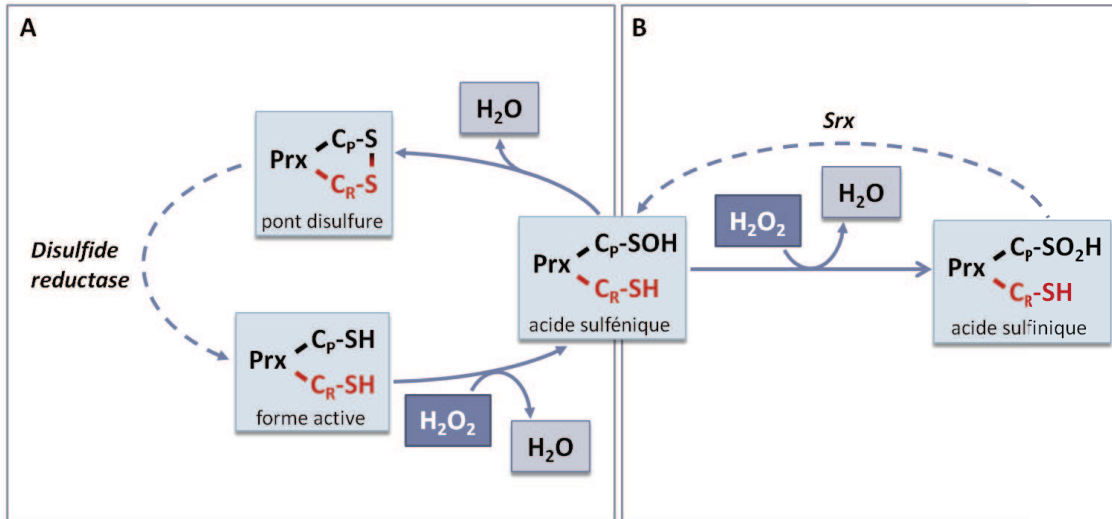


Figure 64. Cycle catalytique des 2-Cys Prx typiques. Adapté d'après (Wood et al., 2003a) et (Lowther and Haynes, 2010). (A) La forme active réagit avec une molécule de H₂O₂ et est oxydée en acide sulfénique. Les deux sous-unités du dimère forment ensuite un pont disulfure entre elles. La forme active est régénérée par réaction avec un donneur d'électron (disulfide réductase). (B) Dans certaines conditions, la cystéine peroxydase (C_P) peut être suroxydée en acide sulfonique conduisant à la perte de l'activité peroxydase. Cette réaction est réversible par l'action de la sulfirédoxine (Srx).



Figure 65. Décimère de 2-Cys Prx (Schroder et al., 2000). Chaque monomère est représenté d'une couleur différente.

3.1.1.2 Les peroxyrédoxines à deux cystéines atypiques

Les peroxyrédoxines à deux cystéines atypiques (2-Cys Prx atypiques) sont des enzymes monomériques au sein desquelles les deux cystéines qui participent à la réaction (C_P et C_R) se trouvent sur la même sous-unité (Seo et al., 2000).

3.1.1.3 Les peroxyrédoxines à une cystéine

Les peroxyrédoxines à une cystéine (1-Cys Prx) fonctionnent en homodimères mais ne possèdent pas de cystéine de régénération (C_R). Leur acide sulfénique est probablement régénéré par un partenaire externe contenant un thiol *via* un mécanisme qui n'a pas encore été élucidé (Kang et al., 1998).

3.1.2 Suroxydation des 2-Cys Prx typiques

Les 2-Cys Prx typiques peuvent être inactivées dans certaines conditions par oxydation de la cystéine peroxydase (C_P) en acide sulfénique ($-SO_2H$) conduisant à la perte de l'activité peroxydase (Figure 64B).

De telles inactivations ont été observées *in vitro* chez les mammifères (Cha et al., 2000; Rabilloud et al., 2001; Wagner et al., 2002), la levure (Koo et al., 2002; Park et al., 2000) et les plantes (Dietz et al., 2002; Konig et al., 2002), alors que ce n'est pas le cas des 2-Cys Prx typiques bactériennes étudiées (Wood et al., 2003a). Les 2-Cys Prx inactivables sont aussi dites "sensibles".

Une hypothèse expliquant pourquoi seules certaines Prx sont soumises à la suroxydation est venue de la comparaison de la structure de Prx "sensibles" (*Homo sapiens*, *Saccharomyces cerevisiae* et *Hordeum vulgare*) et de Prx "résistantes" (*Salmonella typhimurium*, *Streptococcus mutans* et *Escherichia coli*) (Wood et al., 2003a). La différence viendrait de deux caractéristiques uniques présentes dans les Prx sensibles : (i) une insertion de trois résidus associée à une séquence conservée Gly-Gly-Leu-Gly (motif GGLG) et (ii) une hélice supplémentaire formant une extension C-terminale associée à une séquence conservée Tyr-Phe (motif YF) (Figure 66).

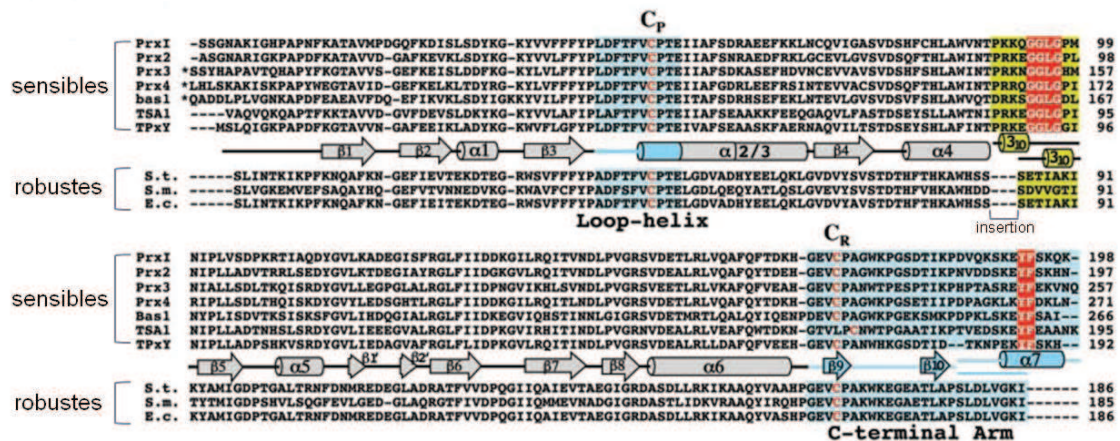


Figure 66. Alignement de séquences basées sur la structure de sept 2-Cys Prx typiques sensibles (groupe de séquences du haut) et robustes (groupe de séquences du bas) (Wood et al., 2003a). Les éléments de la structure secondaire sont représentés : hélices α (cylindres) et brins β (flèches). Les différences structurales apparaissent sous forme sur deux lignes superposées (jaune ou cyan). Les motifs GGLG et YF sont surlignés en rouge, et les cystéines peroxydase (C_P) et de régénération (C_R) sont représentés en rouge.

Les séquences sensibles sont des 2-Cys Prx d'*Homo sapiens* (Prx1, Prx2, Prx3 et Prx4), de levure (TSA1 et TPxY), et de plantes (Bas1). Les 2-Cys Prx robustes présentées sont des séquences de *S. typhimurium* (St), *Streptococcus mutans* (Sm) et *Escherichia coli* (E.c.).

Ces motifs structuraux supplémentaires entraveraient la capacité de la C_R de la seconde sous-unité à accéder à la C_P . Cette pause cinétique laisserait la forme C_P -SOH exposée à l'attaque d'une autre molécule de peroxyde résultant en une suroxydation en acide sulfénique (C_P -SO₂H) (Figure 67).

Ce modèle est compatible avec des expériences réalisées chez *Schizosaccharomyces pombe*, sur une version de la protéine où l'hélice C-terminale est tronquée (Koo et al., 2002). Dans cette expérience, l'attaque de la forme C_P -SOH (acide sulfénique) par la C_R se fait plus rapidement et empêche ainsi toute suroxydation.

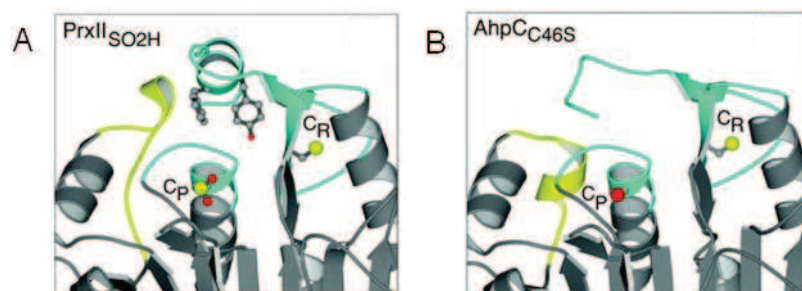


Figure 67. Différences structurales entre Prx sensible (A) et robuste (B). D'après (Wood et al., 2003a). Diagramme en ruban de la région du site actif d'une 2-Cys Prx sensible (PrxII humaine) (A), et d'une 2-Cys Prx robuste (AhpC de *Salmonella typhimurium*) (B). L'hélice représentée en jaune contient le motif GGLG dans la peroxyrédoxine sensible (A) mais pas dans la peroxyrédoxine résistante (B). De même, la région turquoise contient le motif YF en (A) mais pas en (B). Les chaînes latérales des C_P et C_R et celles du motif YF sont également représentées.

La comparaison des panneaux (A) et (B) montre les effets des motifs spécifiques de la Prx sensible sur l'enfouissement de l'hélice contenant la C_P . Elle qui devient beaucoup moins accessible à la C_R et donc plus sensible à l'attaque par une autre molécule de peroxyde.

Bien qu'on ait d'abord pensé que la réaction formant une 2-Cys Prx suroxydée était biologiquement irréversible, il a été montré, en 2003 (chez la levure), qu'elle est en fait réversible grâce à l'action d'une sulfinyl réductase : la sulfirédoxine (Srx), en présence d'ATP (Biteau et al., 2003).

Des études structurales récentes d'un complexe Prx-Srx (Figure 68) ont révélé que le motif GGLG des Prx sensibles se juxtapose au site de liaison à l'ATP de Srx et jouerait donc un rôle important dans la réaction de rétro-réduction (Jonsson et al., 2008).

3.1.3 Rôle du cycle de suroxydation

La découverte d'une possibilité de régénération de l'activité peroxydase de la 2-Cys Prx typique par la sulfirédoxine a soulevé l'hypothèse de l'existence d'un rôle physiologique de la forme suroxydée de cette peroxyrédoxine.

En parallèle, les valeurs de la constante catalytique (qui donne la mesure de l'efficacité de la catalyse par l'enzyme sur le substrat) de Srx chez l'homme, le rat et *Arabidopsis* sont très faibles (0.1 à 1.8/min), ce qui suggère que Srx est une enzyme "inefficace". Cette faible activité pourrait laisser penser que la réparation lente des molécules de Prx puisse avoir une importance physiologique.

3.1.3.1 Rôle dans la transduction du signal

Bien que le peroxyde d'hydrogène soit une source de stress oxydatif, il agit également comme second messager dans la transduction du signal par réaction avec les thiols des protéines impliquées dans la signalisation. Il constitue une molécule de signalisation performante car il est rapidement produit et sa concentration est facilement contrôlée par les enzymes antioxydantes (Hall et al., 2009).

Par exemple, des études récentes indiquent que, chez les animaux, in vivo, les 2-Cys Prx sont nécessaires à l'apoptose (Mu et al., 2002; Wonsey et al., 2002).

L'état d'oxydation des 2-Cys Prx jouerait alors le rôle de commutateur cellulaire : la transduction du signal se ferait par l'inactivation des 2-Cys Prx dans les conditions de signalisation conduisant à la perte de l'activité peroxydase. Ainsi, la sensibilité à l'inactivation par le peroxyde d'hydrogène serait une caractéristique des 2-Cys Prx typiques qui aurait été sélectionnée chez les organismes utilisant le peroxyde d'hydrogène comme signal.

3.1.3.2 Rôle de chaperonnes

En 2005, Moon *et al.* ont montré que, chez l'homme et chez la levure, les 2-Cys Prx typiques jouent également le rôle de protéines chaperonnes en protégeant des protéines (telles que l'insuline) d'une agrégation induite par un stress (Moon et al., 2005). Cette activité serait associée à la formation de complexes de 2-Cys Prx de haut poids moléculaire (pouvant être >2000 KDa), correspondant à l'empilement de plusieurs décamères superposés, alors que l'activité peroxydase, elle, est associée à la forme dimérique ou décamérique.

La suroxydation des 2-Cys Prx typiques sous forme d'acide sulfinique stabilise justement ces formes de haut poids moléculaire. Chez la levure, la forme suroxydée (observée sous forme de deux dodécamères) a une activité de chaperonne quatre fois supérieure à celle de la forme réduite qui a été observée majoritairement sous forme dimérique. Les formes suroxydées de ces Prx pourraient donc augmenter leur activité de protéine chaperonne grâce à la formation de complexes de haut poids moléculaires stables (Lim et al., 2008).

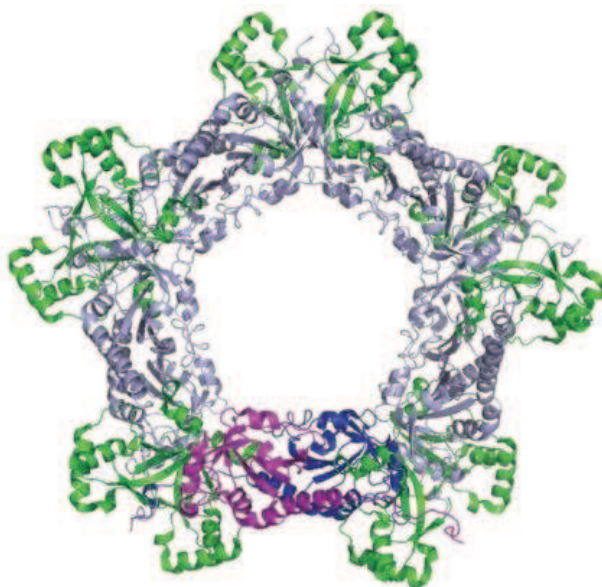


Figure 68. Le dimère de 2-Cys Prx (magenta et bleu foncé) s'associe avec d'autres dimères (bleu gris) dans le cadre du cycle catalytique normal pour former des oligomères d'ordre supérieur (ici un décimère) (Jonsson et al., 2008). L'état suroxydé est stabilisé sous cette forme. La modélisation des Srx (vert) sur un décimère de Prx montre que Srx peut s'associer à une telle structure sans changement important du décimère, de façon compatible avec le rôle du Srx dans la réduction des Prx suroxydés.

Pour résumer, dans des conditions extrêmes de stress oxydatif, les 2-Cys Prx sensibles peuvent être oxydées en acide sulfinique. L'hydroxydation peut conduire à la formation d'agrégats sphériques de masse moléculaire très élevée résultant en un commutateur entre activité enzymatique peroxydase et chaperonne moléculaire pouvant empêcher le dépliement et la précipitation de protéines (Barranco-Medina et al., 2009; Jang et al., 2006; Jang et al., 2004; Kumsta and Jakob, 2009). Cette fonction alternative est considérée comme une protection importante contre le stress oxydatif, et il a même été montré qu'elle pouvait bloquer l'entrée en apoptose (Moon et al., 2005).

Ainsi, la régénération des Prx médiée par Srx représente un processus physiologique important qui peut permettre aux cellules un retour à l'homéostasie en interrompant à la base la signalisation dépendant du peroxyde et l'activité chaperonne.

3.2 Analyses phylogénétiques de Srx et de 2-Cys Prx

Il a été longtemps supposé que les 2-Cys Prx sensibles n'existaient que chez les eucaryotes, jusqu'à ce qu'il ait été découvert des homologues de Srx (qui ont pour seul substrat connu les Prx suroxydés) présents chez certaines cyanobactéries (Basu and Koonin, 2005).

Amel Latifi (équipe "Différenciation cellulaire microbienne et signalisation", Laboratoire de Chimie Bactérienne, CNRS, Marseille) travaille sur la cyanobactérie *Anabaena sp. PCC 7120*, et a entrepris la caractérisation fonctionnelle de la protéine SrxA chez cette bactérie. Elle a pu confirmer expérimentalement que SrxA (la sulfirédoxine d'*Anabaena*) est capable de réduire spécifiquement la forme sulfinique de PrxA (une peroxyrédoxine à deux cystéines d'*Anabaena*), à la fois in vivo et in vitro. La réaction de réduction catalysée par la sulfirédoxine d'*Anabaena* est similaire à celle réalisée par la Srx eucaryote. De plus, un mutant dépourvu du gène codant pour SrxA a été montré comme étant plus sensible au stress que le type sauvage.

Dans le cadre de cette étude, je me suis intéressée à l'histoire évolutive du couple de protéines Srx/Prx afin de comprendre son origine et son évolution. L'article issu de cette collaboration est présenté en Annexe 3.

3.2.1 Matériel et méthodes

3.2.1.1 Recherche d'homologues

A partir des séquences de PrxA et SrxA d'*Anabaena sp. PCC 7120* (alr4641 et asl4146 respectivement), j'ai réalisé une recherche d'homologues par BLASTp contre une banque locale contenant l'ensemble des protéomes complets procaryotes disponibles en février 2010 (1002 protéomes de bactéries et 81 protéomes d'archées). Pour la recherche des homologues eucaryotes, j'ai utilisé de plusieurs séquences graines, et ai réalisé à la fois des requêtes par BLASTp au NCBI contre la banque *nr*, ainsi qu'une série de requêtes par tBLASTn contre des banques de génomes complets ou en cours de séquençage afin d'avoir la meilleure représentativité possible de la diversité eucaryote (banque du JGI, TBestDB, banque d'EST de *Porphyra yezoensis*, banque dédiée à *Galdieria sulphuraria* et celle dédiée au projet génome de *Cyanidioschyzon merolae*). Enfin, même pour les génomes *a priori* complets, j'ai également vérifié les absences par tBLASTn.

Les homologues ont été identifiés par examen visuel des différentes sorties de BLAST (aucun seuil de score ou d'e-value n'a été défini *a priori*).

3.2.1.2 Construction des jeux de données

Les homologues de SrxA et de PrxA ont été assemblés en deux jeux de données qui ont été alignés en utilisant ClustalW 2.0.12 (Larkin et al., 2007). Les alignements ont ensuite été édités et raffinés manuellement *via* le logiciel ED du package MUST (Philippe, 1993). Les régions des alignements pour lesquelles l'homologie entre les sites était douteuse ont été retirées manuellement avant l'analyse phylogénétique en utilisant le programme NET du même package MUST.

3.2.1.3 Reconstructions phylogénétiques

L'outil ProposeModel de Treefinder nous a orientés vers l'utilisation d'un modèle LG, avec estimation des fréquences et loi gamma (pour tenir compte de l'hétérogénéité des taux d'évolution suivant les sites).

Les arbres phylogénétiques ont donc été reconstruits pour chaque protéine en maximum de vraisemblance (ML) par le programme PhyML 3.0 (Guindon and Gascuel, 2003) en utilisant le modèle LG et une loi gamma (4 catégories de sites et estimation du paramètre alpha). La recherche de l'arbre ML a été réalisée par méthode NNI (Nearest Neighbor Interchange) et SPR (Subtree Pruning and Regrafting). La robustesse des branches des arbres inférés a été estimée par la procédure de bootstrap non paramétrique implémentée dans PHYML (100 réplicats). Une deuxième recherche par ML a été effectuée avec Treefinder (Jobb et al., 2004), en utilisant les mêmes paramètres que pour PhyML. L'approche bayésienne a également été utilisée *via* le programme MrBayes 3.1.2 (Ronquist and Huelsenbeck, 2003), avec un modèle mixte et une loi gamma (4 catégories). Les calculs ont été lancés sur 4 chaînes pendant 1000000 de générations, avec des arbres échantillonnés toutes les 100 générations. Les 2500 premiers arbres ont été éliminés ("burning").

3.2.1.4 Recherche des séquences d'adressage

La localisation cellulaire de chaque séquence eucaryote a été étudiée en utilisant les informations disponibles dans les entrées Genpept, combinées aux prédictions bioinformatiques de présence d'un peptide d'adressage au chloroplaste (CTP, Chloroplast Transit Peptide) en utilisant différents logiciels (Tableau 2 et Tableau 3) :

- ChloroP (Emanuelsson et al., 1999)
- TargetP (Emanuelsson et al., 2007)
- iPSORT (Bannai et al., 2002)
- WOLF PSORT (Horton et al., 2007)
- PCLR Chloroplast Localization Prediction version 0.9 (Schein et al., 2001)

3.2.2 Analyses phylogénétiques

3.2.2.1 Répartition taxonomique

3.2.2.1.1 Prx

2308 homologues de Prx ont été récupérés parmi lesquels 1329 présentent les deux cystéines caractéristiques des 2-Cys Prx typiques, respectivement à la position 56 et 178 dans PrxA (alr4641) (Figure 69).

La reconstruction phylogénétique en maximum de vraisemblance a montré que les séquences de 2-Cys Prx typiques forment un groupe monophylétique au sein de la superfamille des Prx, suggérant une origine évolutive unique (Figure 69).

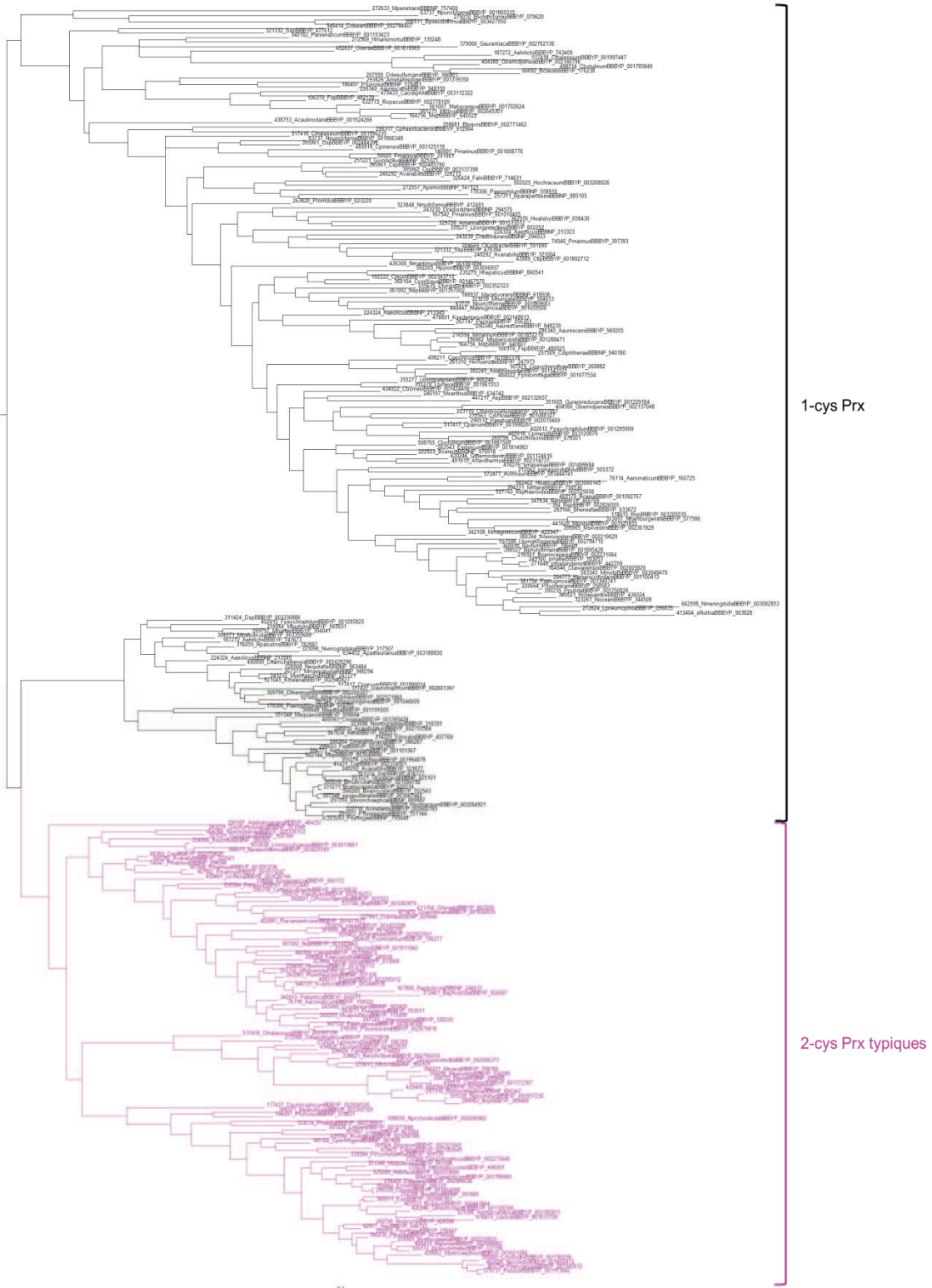


Figure 69. Arbre en maximum de vraisemblance de la superfamille des Prx. Cet arbre a été calculé sur une sélection de séquences homologues (238 séquences, 134 positions). Les séquences en rose correspondent aux 2-Cys Prx typiques, les autres sont des 1-Cys Prx. Arbre calculé avec PhyML, modèle LG, loi gamma (4 catégories).

Les 2-Cys Prx typiques sont retrouvées dans la quasi-totalité des grands groupes de bactéries (même si elles ne sont pas présentes chez toutes les espèces au sein de ces différents groupes) et chez quelques rares archées (Tableau 9).

Domaine	Phylum	Nombre de génomes codant pour une 2-Cys Prx typique	Nombre total de génomes	Pourcentage de génomes codant pour une 2-Cys Prx typique
Archaea	Crenarchaeota	0	25	0
	Euryarchaeota	2	53	4
	Korarchaeota	0	1	0
	Nanoarchaeota	0	1	0
	Thaumarchaeota	0	1	0
Bacteria	Acidobacteria	2	3	67
	Actinobacteria	41	93	44
	Aquificae	4	7	57
	Bacteroidetes	19	26	73
	Chlorobi	9	11	82
	Chloroflexi	0	13	0
	Cyanobacteria	39	39	100
	Deinococcus-Thermus	0	5	0
	Dictyoglomi	0	2	0
	Elusimicrobia	0	2	0
	Fibrobacteres	0	1	0
	Firmicutes	72	230	31
	Fusobacteria	2	4	50
	Gemmatimonadetes	0	1	0
	Nitrospirae	0	1	0
	PVC_Chlamydiae	15	16	94
	PVC_Planctomycetes	2	2	100
	PVC_Verrucomicrobia	1	3	33
	Proteobacteria	302	510	59
	Spirochaetes	7	18	39
Synergistetes	0	1	0	
Thermotogae	0	11	0	

Tableau 9. Répartition taxonomique des 2-Cys Prx typiques chez les procaryotes. Sont indiqués le nombre de génomes de chaque phylum codant au moins pour une 2-Cys Prx typique, le nombre total de génomes complets disponibles pour chaque phylum au moment de l'étude, et le pourcentage de génomes complets qui codent pour une 2-Cys Prx typique.

Chez les eucaryotes, bien que les 2-Cys Prx typiques ne soient pas systématiquement présentes chez toutes les espèces, des homologues sont retrouvés dans la quasi-totalité des représentants de l'ensemble des grands groupes (Opisthokonta, Amoebozoa, Excavata, Alveolata, Heterokonta, Rhizaria et Plantae) (Figure 70). Les seuls grands groupes chez qui je n'en ai pas détecté sont les Haptophyta et les Cryptophyta, mais les données disponibles pour ces groupes sont partielles.

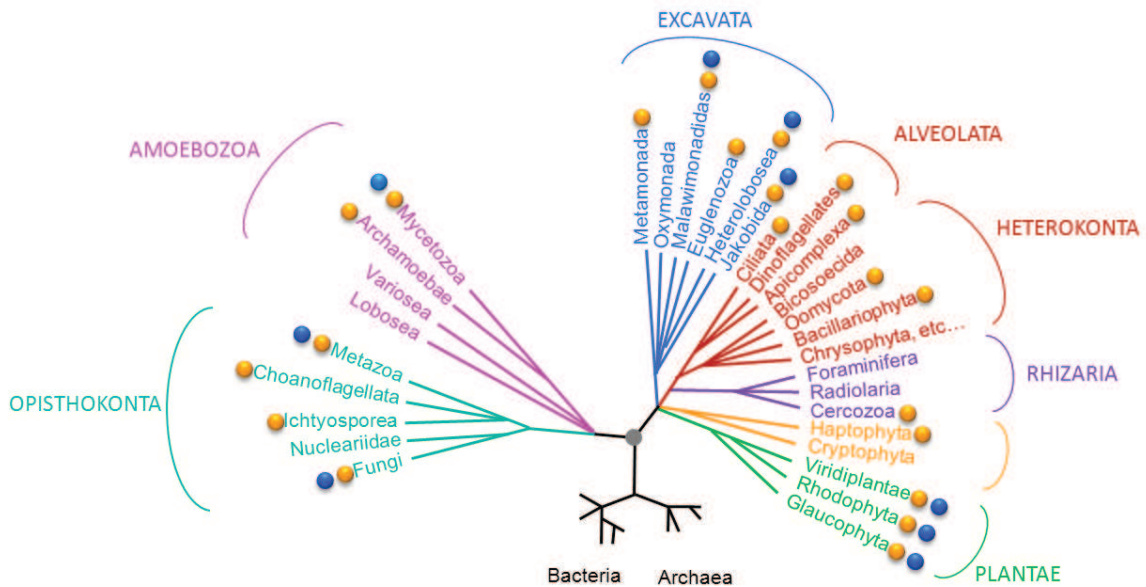


Figure 70. Distribution taxonomique des 2-Cys Prx typiques et des Srx chez les Eucarya. Les groupes marqués d'un point jaune contiennent au moins un représentant dont le génome code pour une 2-Cys Prx typique, tandis que ceux marqués d'un point bleu contiennent au moins un représentant dont le génome code pour une Srx.

3.2.2.1.2 Srx

Contrastant avec le cas des Prx, la répartition taxonomique des Srx est beaucoup plus restreinte et seuls 109 homologues ont été identifiés. Comme précédemment décrit dans la littérature (Basu and Koonin, 2005), nous avons identifié des homologues chez les Cyanobactéries et plusieurs groupes eucaryotes tels que les Metazoa, les Fungi, les Amoebozoa et les plantes terrestres.

En revanche, les recherches par tBLASTn sur les projets génomes en cours ont révélé des homologues inconnus jusque-là chez les algues vertes (*Chlamydomonas reinhardtii*, *Volvox carteri*, *Chlorella sp.*), les Rhodophyta (*Galdieria sulphuraria*), les Glaucophyta (*Cyanophora paradoxa*) et chez le groupe de protistes qui forment les Excavata (*Naegleria gruberi*, *Malawimonas californiana*, *Reclimonas americana*). Toutefois, aucun homologue n'a été détecté dans les autres lignées portant un plaste, telles que les Alveolata, les Heterokonta, les Haptophyta ou les Cryptophyta (Figure 70).

A cette étape, il est important de remarquer que tous les génomes codant pour une Srx codent aussi pour une 2-Cys Prx. En revanche, l'inverse n'est pas vrai, par exemple, toutes les plantes vertes n'ont pas systématiquement une copie de Srx (ex : *Ostreococcus*). De façon semblable, nous n'avons détecté Srx que dans une seule algue rouge (*Galdieria sulphuraria*) et une glaucophyte (*Cyanophora paradoxa*) (même si cela est probablement sous-estimé en raison de la rareté des données génomiques disponibles pour ces deux groupes), alors que

Prx est également détectée chez une seconde glaucophyte *Glaucocystis nostochinearum* et trois autres algues rouges.

Chez les cyanobactéries, les organismes ayant une copie de Srx sont répartis dans presque tous les sous-groupes, y compris les plus basaux, tels que les cyanobactéries Yellowstone (*Synechococcus* sp. JA-2-3B'a(2-13) et *Synechococcus* sp. JA-3-3Ab) (Figure 71). En revanche, contrairement aux 2-Cys Prx typiques qui sont présentes chez toutes les cyanobactéries sans exception, Srx est absente chez de nombreuses espèces au sein de chaque grand groupe.

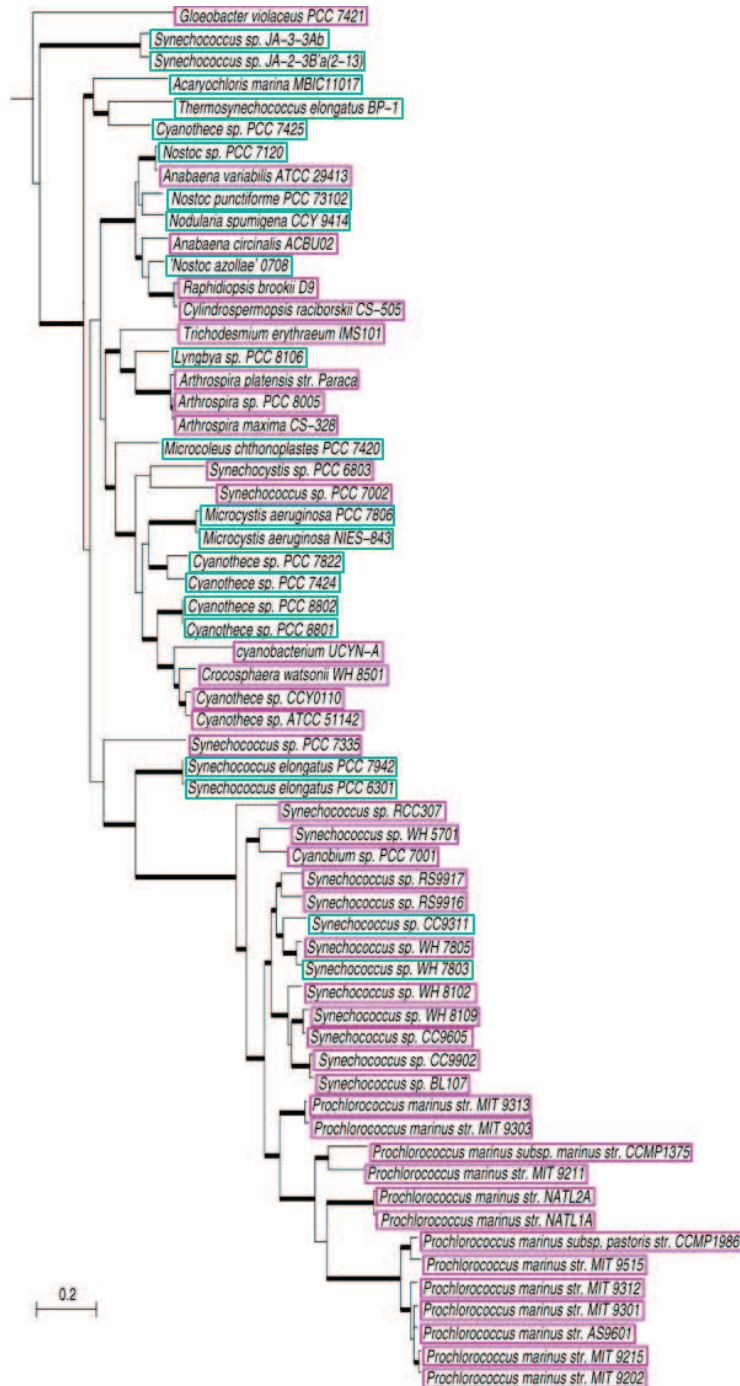


Figure 71. Distribution taxonomique des Srx chez les Cyanobacteria. La présence de Srx est reportée sur une phylogénie de référence des Cyanobacteria (Simonetta Gribaldo, communication personnelle). Les espèces dont le génome code pour Srx sont entourées en vert. Celles dont le génome ne code pas pour Srx sont entourées en rose. On peut voir que les espèces qui possèdent une Srx sont réparties dans presque tous les groupes de cyanobactéries, y compris des groupes très basaux.

Il est intéressant de noter que les recherches par BLAST ont également permis de récupérer des séquences procaryotes de la famille ParB. C'est une protéine présente dans un large échantillon taxonomique de bactéries et d'archées et qui joue un rôle dans la ségrégation de l'ADN pendant la division cellulaire. En dépit de la différence radicale entre les rôles fonctionnels des ParB et Srx, et le fait que les séquences des ParB sont beaucoup plus longues que celles des Srx (environ 300 acides aminés contre 100), l'alignement de l'une avec l'autre montre une identité de séquence élevée (41% entre les protéines Srx d'*Anabaena sp.* PCC 7120 et ParB d'*Ammonifex degensii*), ainsi que des motifs strictement conservés, mettant clairement en évidence l'homologie entre ParB et Srx. Il a d'ailleurs été postulé que les Srx dériveraient des ParB (Basu and Koonin, 2005).

3.2.2.2 Phylogénies

3.2.2.2.1 Prx

La phylogénie restreinte aux 2-Cys Prx typiques montre que les séquences se répartissent en deux groupes monophylétiques :

- Un groupe contenant la très grande majorité des séquences bactériennes (mais aucune séquence de cyanobactéries) et les quelques séquences d'archées (Figure 72, cadre rose);
- Un groupe contenant l'ensemble des séquences eucaryotes, au sein desquelles sont retrouvées les séquences de cyanobactéries et des séquences bactériennes non apparentées (principalement des protéobactéries, ainsi que quelques séquences éparses d'autres groupes bactériens, tels que les Verrucomicrobiae, Bacteroidetes, Planctomycètes et Spirochaetes) (Figure 72, cadre vert).

Il est intéressant de remarquer que les séquences qui constituent ce second groupe présentent les deux signatures correspondant aux motifs structuraux spécifiques des 2-Cys Prx sensibles à la suroxydation, à savoir GGLG et YF, situés respectivement aux positions 99-102 et 199-200 dans PrxA. Ceci signifie qu'il existe une origine évolutive unique des séquences sensibles (support SH-like pour la monophylie du groupe = 0.98) (Figure 72). En réalité, ces motifs ne sont pas aussi strictement conservés que décrits dans la littérature (Wood et al., 2003a), puisqu'on retrouve un motif GG[VIL]G (et non GGLG strict), et [YF] [FWL] (et non YF). Il est possible que le motif ait été défini de façon trop stricte dans la littérature sur la base d'un nombre trop restreint de séquences caractérisées expérimentalement. Une autre possibilité est que les séquences qui ne possèdent pas les motifs GGLG et YF "stricts" ont secondairement perdu leur capacité à être suroxydées.

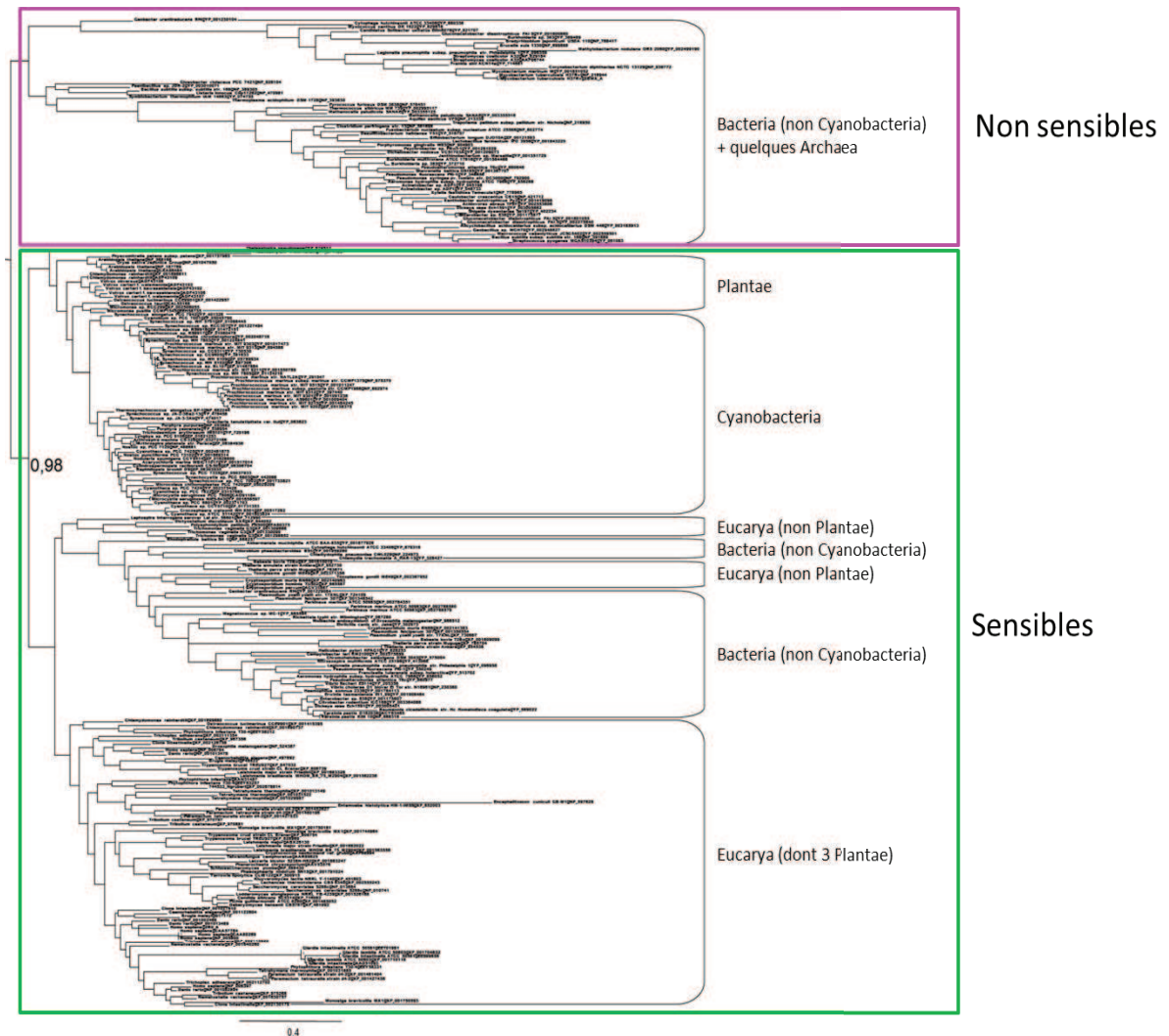


Figure 72. Arbre en maximum de vraisemblance de l'ensemble des 2-Cys Prx typiques. Cet arbre a été reconstruit à partir d'une sélection de séquences représentatives de la diversité des 1329 homologues (267 séquences, 179 positions). L'arbre a été raciné comme indiqué par la phylogénie des 2-Cys Prx et 1-cys Prx (Figure 69). Il a été reconstruit avec PhyML, modèle LG, loi gamma (4 catégories). Le cadre vert désigne les séquences de 2-Cys Prx qui portent le motif de suroxydation (c'est-à-dire l'ensemble des séquences eucaryotes et cyanobactériennes, ainsi que quelques séquences bactériennes autres), tandis que le cadre rose indique celles qui ne le portent pas (c'est-à-dire la quasi-totalité des séquences bactériennes et archées). La valeur de 0.98 est le soutien pour la monophylie des deux groupes (SH-like). La barre d'échelle représente 0.4 substitution par site.

Afin d'obtenir une image plus claire de l'histoire évolutive des 2-Cys Prx "sensibles", j'ai effectué une analyse phylogénétique fine de ces séquences (Figure 73). Les arbres obtenus (racinés comme indiqués par la Figure 72) montrent deux groupes monophylétiques bien soutenus :

- Le premier contient toutes les séquences de cyanobactéries et de presque toutes les séquences de plantes (plantes vertes, algues rouges, glaucophytes) (BV = 97% et PP = 1.0). L'ensemble des cyanobactéries complètement séquencées à ce jour possèdent une 2-Cys Prx sortant au sein de ce groupe.
- Le second groupe contient les autres séquences eucaryotes (dont quelques rares séquences de plantes) au sein desquels on retrouve un ensemble de séquences bactériennes non apparentées (~180 séquences de Proteobacteria, Verrucomicrobiae, Bacteroidetes, Planctomycètes et Spirochaetes) (Figure 73, boîte jaune), formant un groupe monophylétique avec les Prx d'Apicomplexa et Dinoflagellata (sous-groupes d'Alveolata). Ceci suggère fortement un THG de ceux-ci vers au moins une espèce bactérienne. Par la suite, plusieurs THG ont dû avoir lieu entre bactéries (ce que suggère le fait que les espèces de ce groupe proviennent de différentes lignées non apparentées).

L'AhpC de *Helicobacter pylori* fait partie de ce groupe de 2-Cys Prx bactériennes qui sont évolutivement plus proches des 2-Cys Prx eucaryotes que des autres 2-Cys Prx typiques bactériennes robustes (*i.e.* non sensibles). De façon intéressante, il a été constaté que l'AhpC de *H. pylori* agit également en tant que protéine chaperonne comme le fait la 2-Cys Prx de levure (Chuang et al., 2006). En condition de stress oxydant, certaines AhpC seront converties en chaperonnes de haut poids moléculaire afin de prévenir le mauvais repliement des protéines. Lorsque trop de protéines sont endommagées, toutes les AhpC sont converties en chaperonnes afin de restaurer les protéines dépliées. Cette similarité fonctionnelle avec les 2-Cys Prx eucaryotes semble en accord avec leur proximité phylogénétique.

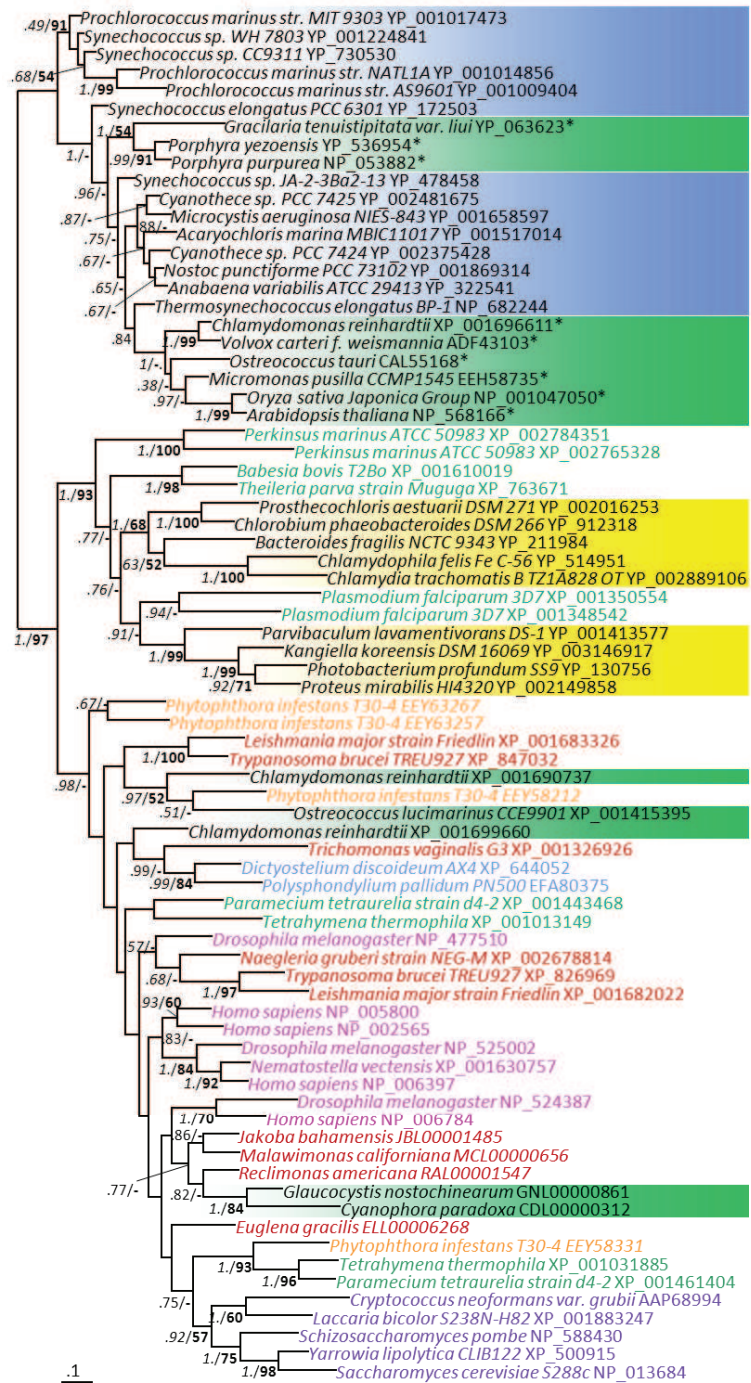


Figure 73. Arbre bayésien des 2-Cys Prx typiques sensibles (76 séquences, 167 positions). Cet arbre a été fait sur une sélection de séquences représentatives de la diversité des 2-Cys Prx sensibles. Il est raciné comme indiqué par la Figure 72. Les boîtes bleues, vertes et jaune indiquent respectivement les séquences de Cyanobacteria, de Plantae et les autres séquences bactériennes nichées au sein des eucaryotes. Les autres séquences eucaryotes sont représentées en rose (Metazoa), violet (Fungi), bleu (Amoebozoa), rouge (Excavata), orange (Heterokonta) et vert (Alveolata). Les séquences contenant un peptide d'adressage au chloroplaste sont indiquées par un astérisque (*). Les valeurs aux nœuds sont les PP et VB. Seules celles supérieures à 0.5 et 50%, respectivement, sont indiquées. La barre d'échelle représente 0.1 substitution par site.

3.2.2.2.2 Srx

En parallèle, j'ai reconstruit un arbre phylogénétique des membres de la famille Srx, enraciné avec des séquences ParB (Figure 74 et Figure 75). Fait intéressant, tout comme dans le cas des Prx, les séquences de plantes vertes sont plus proches des séquences de cyanobactéries que des autres séquences eucaryotes, formant avec elles un groupe monophylétique fortement soutenu (PP = 1 et VB = 87%). En revanche, au sein des deux sous-groupes principaux (plantes/cyanobactéries et autres eucaryotes), les relations de parentés sont relativement mal résolues.

Au sein des Viridiplantae, la seule exception concerne les séquences de *Volvox carteri* et *Chlamydomonas reinhardtii* qui n'émergent pas dans le groupe plantes/cyanobactéries mais au contraire, au sein du reste des eucaryotes.

Il est à noter qu'une séquence très partielle de Srx a été détectée chez *Galdieria sulphuraria*. En raison de sa petite taille elle n'a pas été incluse dans l'analyse, et il n'est donc pas possible de déterminer si elle est plus proche des séquences de plantes/cyanobactéries, ou des Srx "Eucarya-like".

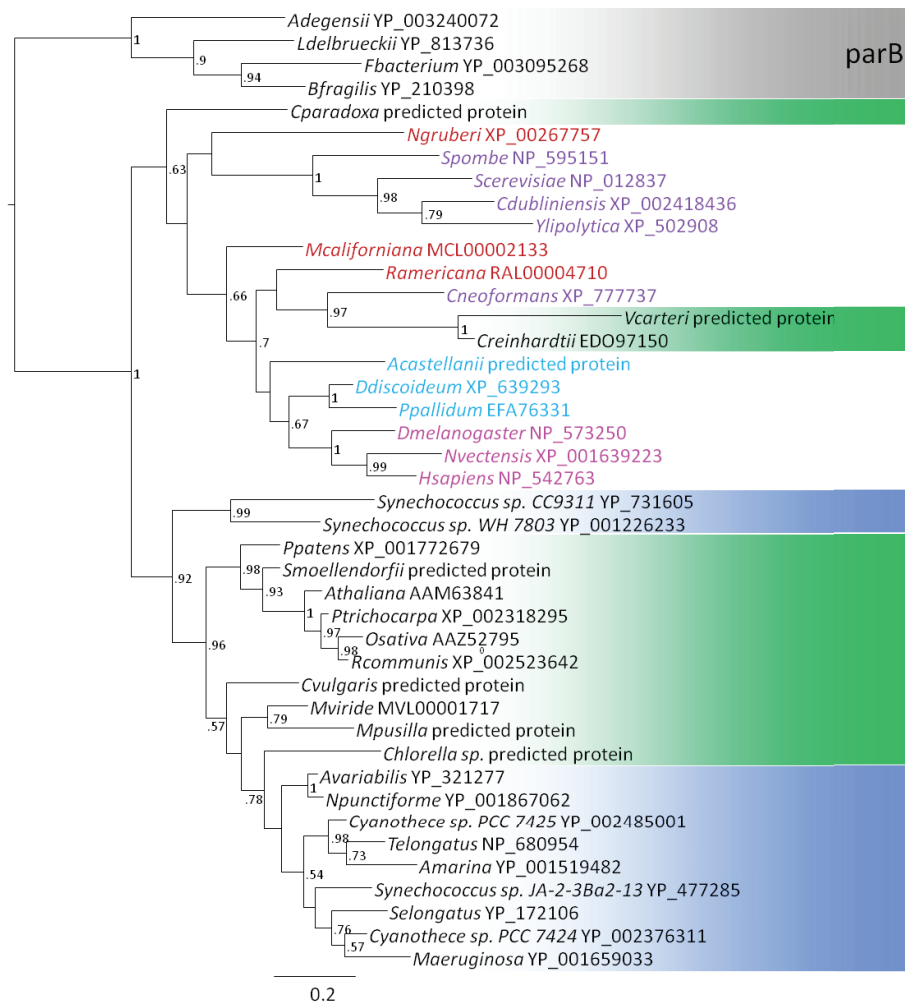


Figure 74. Arbre bayésien des homologues de SrxA raciné avec ParB (42 séquences, 77 positions). Cet arbre a été fait sur une sélection de séquences représentatives de la diversité des homologues de Srx. Les boîtes bleues et vertes indiquent respectivement les séquences de Cyanobacteria et de Plantae. Les séquences en rose, violet, bleu et rouge représentent respectivement les séquences de Metazoa, Fungi, Amoebozoa et Excavata. La boîte grise indique les séquences de ParB qui ont été utilisées comme groupe extérieur. Les valeurs aux nœuds sont les Posterior Probabilities (PP). Seules celles supérieures à 0.5 sont indiquées. La barre d'échelle représente 0.2 substitution par site.

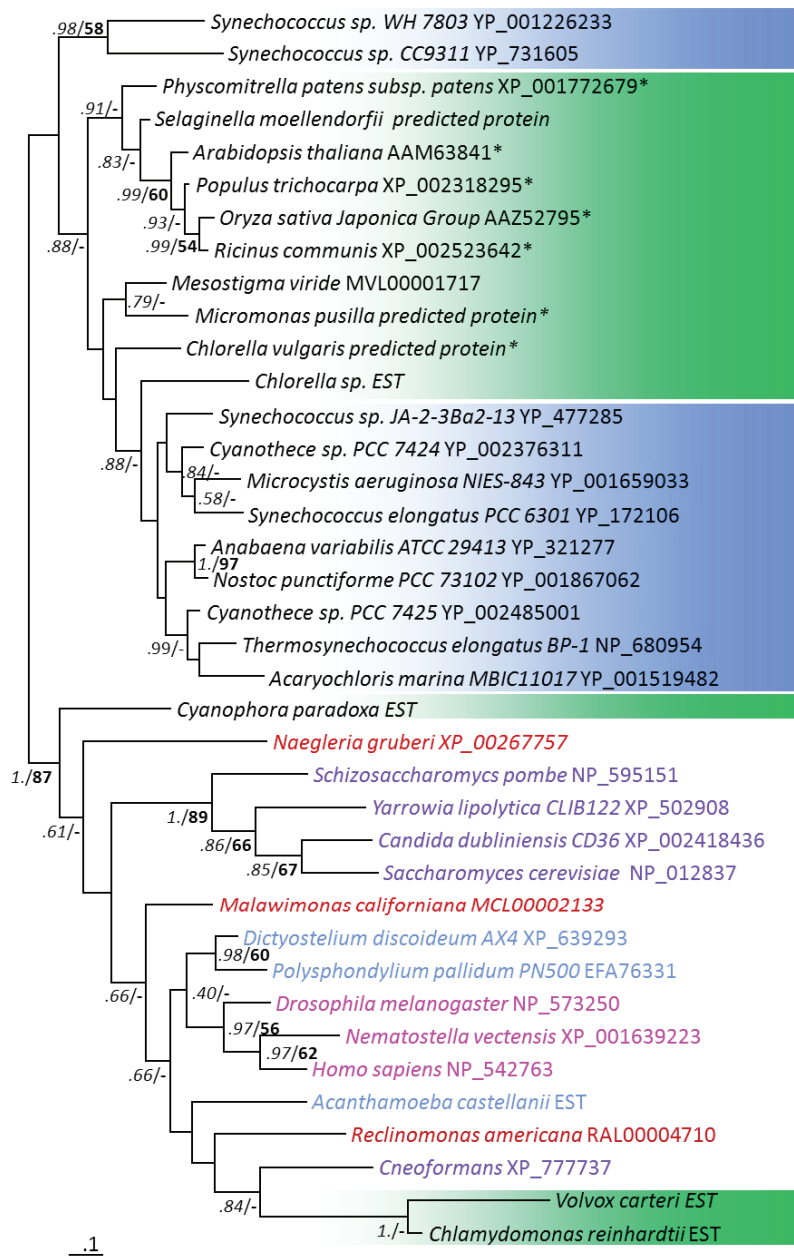


Figure 75. Arbre bayésien des homologues de Sr_x (38 séquences, 77 positions). L'arbre a été raciné comme indiqué par l'arbre précédent. Le code couleur est identique à celui de la **Figure 74**. Les séquences contenant un peptide d'adressage au chloroplaste sont indiquées par un astérisque (*). Les valeurs aux nœuds sont les PP et VB. Seules celles supérieures à 0.5 et 50%, respectivement, sont indiquées. La barre d'échelle représente 0.1 substitution par site.

3.2.2.3 Localisation subcellulaire

Il est intéressant de constater que, dans le cas de Srx et de Prx, la plupart des séquences de plantes qui sortent avec celles des cyanobactéries sont prédites comme possédant un peptide d'adressage au chloroplaste à leur extrémité N-terminale, contrairement aux autres séquences de plantes ou d'autres eucaryotes photosynthétiques (Tableau 10 et Tableau 11).

Organisme	Numéro d'accès	Adressage au plaste			
		ChloroP	TargetP	ipSort Chloro	PCLR
<i>Arabidopsis lyrata</i>	XP_002893835	x	x		x
<i>Arabidopsis thaliana</i>	AAM63841	x	x		x
<i>Chlorella sp.</i>	predicted protein			x	
<i>Cyanophora paradoxa</i>	predicted protein				
<i>Chlorella vulgaris</i>	predicted protein				
<i>Glycine max</i>	ACU16515	x	x	x	x
<i>Micromonas sp.</i>	predicted protein		x		
<i>Micromonas pusilla</i>	predicted protein			x	x
<i>Medicago truncatula</i>	ABC75369	x	x	x	x
<i>mesostigma viride</i>	predicted protein				
<i>Oryza sativa</i>	AAZ52795	x	x		x
<i>Physcomitrella patens</i>	XP_001772679				
<i>Physcomitrella patens</i>	predicted protein	x		x	
<i>Pisum sativum</i>	ADA57704	x	x		
<i>Populus trichocarpa</i>	XP_002318295	x			x
<i>Ricinus communis</i>	XP_002523642	x	x		x
<i>Sbicolor</i>	XP_002436578	x	x	x	x
<i>Selaginella moellendorffii</i>	predicted protein				
<i>Vitis vinifera</i>	XP_002276335	x	x	x	x
<i>Zea mays</i>	ACF81525				
<i>Chlamydomonas reinhardtii</i>	predicted protein				
<i>Volvox carteri</i>	predicted protein				

Tableau 10. Prédiction d'adressage au chloroplaste des Srx chez les eucaryotes photosynthétiques. En vert sont indiquées les séquences qui sont plus proches des séquences de cyanobactéries que des autres séquences eucaryotes.

Organisme	Numéro d'accès	Adressage au plaste (Fiches Genpept)	Adressage au plaste (logiciels de prédiction)			
			ChloroP	TargetP	ipSort Chloro	PCLR
<i>Arabidopsis thaliana</i>	NP_187769	x	x	x	x	x
<i>Arabidopsis thaliana</i>	NP_568166	x	x	x	x	x
<i>Arabidopsis thaliana</i>	AAG40040		x	x	x	x
<i>Arabidopsis thaliana</i>	AAM64537		x	x	x	x
<i>Arabidopsis thaliana</i>	CAA63909		x	x	x	x
<i>Chlamydomonas incerta</i>	ABA01158	x				
<i>Chlamydomonas reinhardtii</i>	XP_001696611	x	x	x		x
<i>Gracilaria tenuistipitata</i>	YP_063623	x			x	
<i>Micromonas pusilla</i>	EEH58735		x	x		x
<i>Micromonas sp</i>	XP_002506055		x			x
<i>Ostreococcus lucimarinus</i>	XP_001419466					
<i>Oryza sativa</i>	ABR25513					
<i>Oryza sativa</i>	EAY94067				x	
<i>Oryza sativa</i>	CAE01526				x	
<i>Oryza sativa</i>	EAZ30702		x		x	
<i>Oryza sativa</i>	NP_001047050		x	x	x	x
<i>Osterococcus tauri</i>	CAL55168		x	x		x
<i>Physcomitrella patens</i>	XP_001757965		x	x	x	
<i>Porphyra purpurea</i>	NP_053882	x				
<i>Phaeodactylum tricornutum</i>	YP_874429	x				
<i>Porphyra yezoensis</i>	YP_536954	x				
<i>Thalassiosira pseudonana</i>	YP_874512	x			x	
<i>Volvox carteri</i>	ADF43104	x				
<i>Volvox carteri</i>	ADF43103	x				
<i>Volvox carteri</i>	ADF43107	x				
<i>Volvox carteri</i>	ADI46867		x			x
<i>Volvox carteri</i>	ADF43102					
<i>Volvox carteri</i>	ADF43106					
<i>Volvox obversus</i>	ADF43108	x				
<i>Babesia bovis</i>	XP_001609099		x			x
<i>Babesia bovis</i>	XP_001610019					
<i>Bigelowiella natans</i>	BNL00002212					
<i>Cryptosporidium hominis</i>	XP_665587					
<i>Cyanophora paradoxa</i>	CDL00000312					
<i>Chlamydomonas reinhardtii</i>	XP_001690737					
<i>Chlamydomonas reinhardtii</i>	XP_001699660					
<i>Euglena gracilis</i>	ELL00006268					
<i>Glaucozystis nostochinearum</i>	GNL00000861					
<i>Ostreococcus lucimarinus</i>	XP_001415395					
<i>Plasmodium falciparum</i>	XP_001348542					
<i>Plasmodium falciparum</i>	XP_001350554					
<i>Phytophthora infestans</i>	EEY58212					
<i>Phytophthora infestans</i>	EEY58331					
<i>Phytophthora infestans</i>	EEY63257					
<i>Phytophthora infestans</i>	EEY63267					
<i>Phytophthora infestans</i>	AAN31487					
<i>Perkinsus marinus</i>	XP_002765328					
<i>Perkinsus marinus</i>	XP_002765329					
<i>Perkinsus marinus</i>	XP_002784351					
<i>Perkinsus marinus</i>	XP_002788581					
<i>Plasmodium yoelii</i>	XP_724109					
<i>Plasmodium yoelii</i>	XP_730667					
<i>Theileria annulata</i>	XP_952730					
<i>Theileria annulata</i>	XP_954538		x	x	x	x
<i>Toxoplasma gondii</i>	XP_002367952		x	x		x
<i>Toxoplasma gondii</i>	XP_002371356					
<i>Theileria parva</i>	XP_763671					
<i>Theileria parva</i>	XP_765704		x	x		x

Tableau 11. Prédiction d'adressage au chloroplaste des 2-Cys Prx chez les eucaryotes ayant un plaste ou un ancêtre ayant eu un plaste. En vert sont indiquées les séquences qui sont plus proches des séquences de cyanobactéries que des autres séquences eucaryotes.

3.2.3 Scénario évolutif

Les recherches dans les banques de séquences ont montré que la distribution taxonomique des protéines Srx et 2-Cys Prx sensibles est majoritairement chevauchante. Néanmoins, ceci n'est pas surprenant étant donné leur lien fonctionnel. Plus précisément, chaque génome codant pour Srx contient également un gène codant pour 2-Cys Prx sans aucune exception. L'inverse n'est cependant pas vrai, puisque certaines 2-Cys Prx sensibles sont retrouvées dans des génomes qui ne codent pas pour Srx, par exemple *Synechocystis PCC 6803*, *Helicobacter pylori*, les Heterokonta ou les Alveolata.

Ceci reflète le fait que d'autres enzymes non encore identifiées sont probablement capables de réduire la forme acide suroxydée de la cystéine catalytique de la 2-Cys Prx. Peu de temps après la découverte de Srx, Budanov *et al.* (Budanov et al., 2004) ont proposé que les membres d'une autre famille de protéines, appelées "sestrines" (Sesn), possèdent une activité réductase de la forme acide sulfonique de Prx, bien que Srx et Sesn ne montrent aucune similarité de séquence. Récemment, plusieurs études ont en fait montré que, contrairement à ce qui avait été suggéré, la sestrine n'est pas une réductase de la Prx suroxydée (Woo et al., 2009). Il reste donc encore à découvrir si d'autres enzymes que Srx sont capables de réaliser cette réaction.

Quoiqu'il en soit, La distribution taxonomique et les arbres phylogénétiques de 2-Cys Prx et de Srx suggèrent une histoire évolutive complexe, et posent la question de l'origine de ce système à deux partenaires.

3.2.3.1 Scénario proposé par Basu *et al.* 2005

Un premier scénario concernant l'histoire de Srx a été proposé par Basu et Koonin il y a quelques années (Basu and Koonin, 2005). Les auteurs suggèrent que Srx a évolué à partir de ParB par transfert horizontal de gènes (THG) des bactéries vers les eucaryotes. Étant donné que, dans cette étude, l'échantillonnage taxonomique laisse penser que Srx n'est présent que chez les eucaryotes du "crown-group" (animaux, plantes et champignons), les auteurs ont supposé que ce THG est survenu tard dans l'évolution des eucaryotes. Ils ont ensuite suggéré que la forte similarité entre les Srx de plantes et de cyanobactéries soit expliquée par un deuxième THG, cette fois à partir des plantes vers les cyanobactéries. Étant données les similitudes importantes des phylogénies de Srx et 2-Cys Prx, un scénario évolutif similaire à celui imaginé pour Srx doit être supposé pour Prx.

Plusieurs points peuvent nous conduire à rejeter cette hypothèse. Premièrement, il est maintenant admis que le soi-disant "crown-group" n'existe pas, et que, chez les Eucarya, les Plantae forment un groupe distant des Unikontes (c'est-à-dire Metazoa, Fungi, Amoebozoa et d'autres protistes comme *Monosiga brevicollis*). En outre, nous avons identifié des protéines Srx dans plusieurs lignées d'Excavata qui sont généralement considérées comme des lignées eucaryotes basales, ce qui suggère que Srx était probablement présent chez LECA.

Deuxièmement, cette hypothèse n'explique pas la présence, chez le glaucophyte *Cyanophora paradoxa* et chez les algues vertes que sont *C.reinhardtii* et *V.carteri*, d'une copie de Srx plus ressemblant aux Srx de type "eucaryote-like" qu'aux Srx de type Viridiplantae/Cyanobacteria. La présence de deux versions distinctes et divergentes de Srx (et de 2-Cys Prx) dans les plantes est difficile à expliquer dans le cadre de l'hypothèse d'un THG des plantes vers les cyanobactéries. En effet, cela implique une origine eucaryote unique des séquences de Srx (et de 2-Cys Prx) (Figure 76A), suivi d'un événement de duplication chez les plantes (Figure 76B) et de la divergence rapide d'une des deux copies à un point tel qu'elles ne formeraient plus

un groupe monophylétique avec les autres séquences eucaryotes dans les arbres phylogénétiques. La copie divergente aurait ensuite été transférée à une cyanobactérie (Figure 76C) et entre les cyanobactéries (Figure 76D). Enfin, ce scénario suppose également la perte de l'une ou l'autre copie de Srx (et de 2-Cys Prx) au cours de l'évolution dans les lignées de plantes (Figure 76E), afin d'expliquer pourquoi on retrouve généralement une seule copie dans ces organismes.

Troisièmement, un transfert des plantes vertes vers les Cyanobacteria impliquerait une forte contrainte temporelle. En effet, la plupart des sous-groupes de cyanobactéries présentent un Srx, y compris les lignées basales telles que *Synechococcus sp. JA-3-3AB*, suggérant alors l'une ou l'autre des hypothèses suivantes :

- Les plantes vertes étaient déjà diversifiées alors que les Cyanobacteria ne l'étaient pas, ce qui peut être exclu puisque leur diversification avait commencé avant l'endosymbiose primaire conduisant à l'émergence de Plantae;
- Les Cyanobacteria étaient également diversifiées et la présence de Srx dans presque tous les sous-groupes de cyanobactéries doit alors être expliquée par de nombreux THG secondaires entre ces sous-groupes.

Enfin, un THG des plantes vers les Cyanobacteria n'est manifestement pas compatible avec le fait que la majorité des Srx (et des 2-Cys Prx) de plantes vertes est prédit comme ayant un peptide de transit chloroplastique (Tableau 10 et Tableau 11). Au contraire, c'est la signature forte que les plantes vertes ont acquis les gènes codant pour ces protéines au travers de l'endosymbiose d'une cyanobactérie.

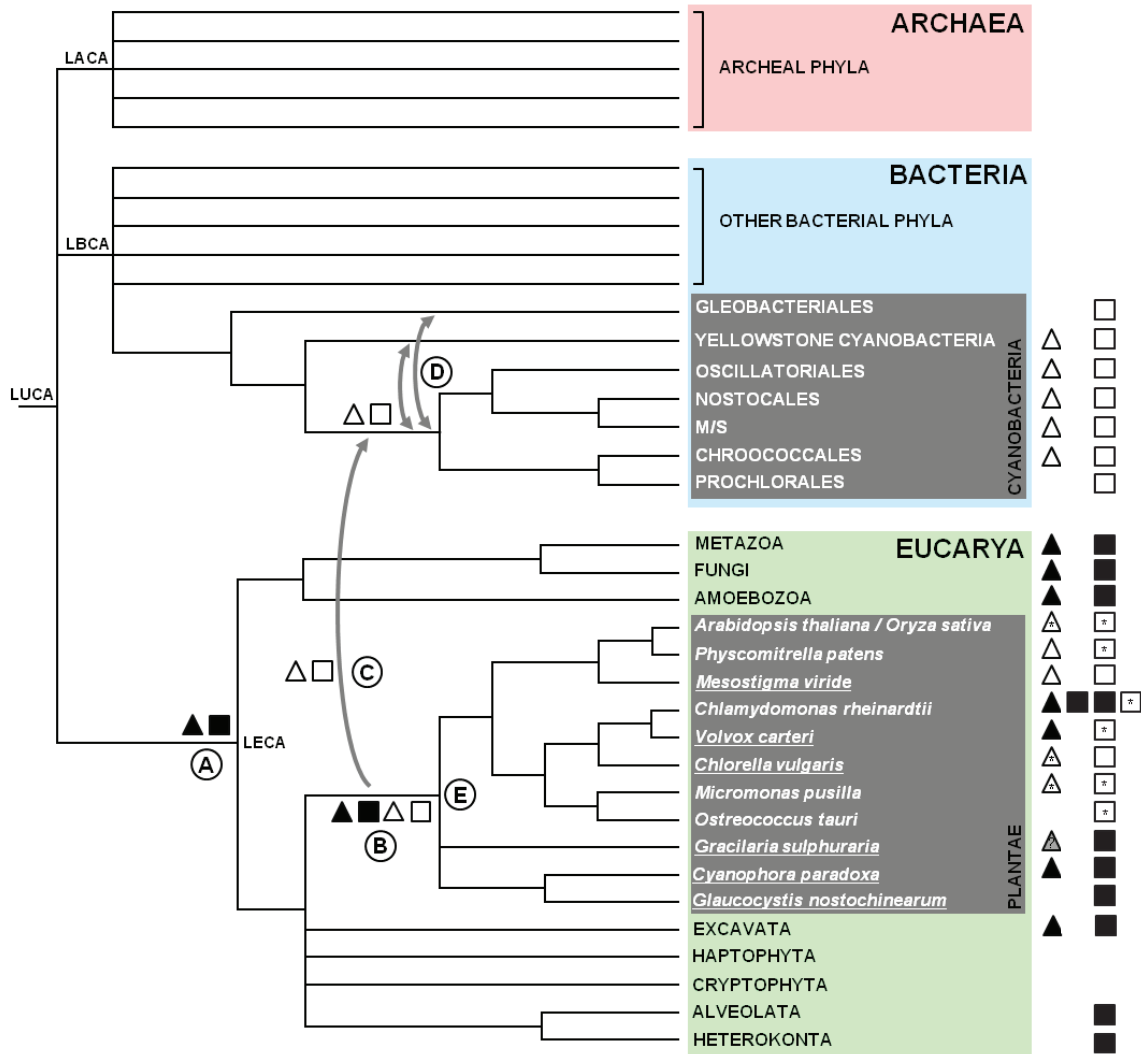


Figure 76. Scénario évolutif des séquences de Srx et de 2-Cys Prx sensibles proposé par Basu *et al.*

L'arbre présente les trois domaines de la vie cellulaire (Archaea, Eucarya et Bacteria). (A) Les séquences de Srx (triangles noirs) et de 2-Cys Prx sensibles (carrés noirs) sont apparues avant l'ancêtre des Eucarya. (B) Un événement de duplication a eu lieu chez l'ancêtre des Plantae, suivi d'une divergence rapide d'une des deux copies, qui a été transférée horizontalement à une cyanobactérie (C), puis entre cyanobactéries (D). Enfin, des copies de Srx et 2-Cys Prx ont été perdues alternativement au cours de l'évolution et de la diversification des lignées de plantes actuelles (E).

Les séquences de Srx et de Prx portant un CTP à leur extrémité N-terminale sont marquées d'un astérisque (*). Les espèces dont le génome complet n'est pas encore disponible sont soulignées. LECA, LACA, LBCA et LUCA désignent respectivement : le dernier ancêtre commun des Eucarya, des Archaea, des Bacteria et le dernier ancêtre commun universel.

3.2.3.2 Notre scénario sur l'émergence de Srx et des 2-Cys Prx sensibles

Nos analyses sont basées sur un échantillonnage taxonomique plus conséquent, et nous permettent de proposer un autre scénario évolutif.

Une version sensible de 2-Cys Prx est présente dans chaque cyanobactérie dont le génome était entièrement séquencé au moment de l'analyse (février 2010), et Srx est présente dans la majorité des grands groupes de cyanobactéries. Ceci sous-entend que l'émergence de la Srx et de la Prx sensible a eu lieu avant la diversification de cette lignée.

De la même façon, la large distribution des 2-Cys Prx sensibles chez les eucaryotes et la présence de Srx dans plusieurs groupes eucaryotes distants suggèrent que ces deux protéines étaient présentes chez LECA (Figure 70).

L'émergence de Srx et le passage d'une Prx robuste à une Prx sensible peuvent s'expliquer de deux façons.

3.2.3.2.1 Hypothèse 1 : Origine cyanobactérienne

Une duplication du gène *parB* a eu lieu chez les Cyanobacteria, l'une des copies a donné Srx et a ensuite été transférée aux Eucarya avant leur diversification (Figure 77A, triangle gris). Cela implique également que ce transfert ait précédé la diversification des Cyanobacteria puisque les Srx eucaryotes ne sortent pas au sein des Srx cyanobactériennes. En parallèle, le passage d'une Prx robuste à une Prx sensible a eu lieu chez l'ancêtre des Cyanobacteria (puisque les autres lignées procaryotes ne possèdent pas de Prx sensible), et le gène codant pour une Prx sensible a ensuite été transféré à l'ancêtre des eucaryotes (Figure 77A, carré gris).

3.2.3.2.2 Hypothèse 2 : Origine eucaryote

Selon cette hypothèse, ParB a été transféré aux eucaryotes ou hérité de LUCA, puis a ensuite divergé chez les Eucarya pour donner Srx, qui lui-même a alors été transféré au dernier ancêtre commun des Cyanobacteria (*i.e.* avant la diversification de ce groupe) (Figure 77A, triangle gris). Ce THG a également dû se produire avant la diversification des eucaryotes (toujours pour expliquer que les Srx cyanobactériennes émergent en tant que groupe-frère des Srx eucaryotes, et non pas en leur sein).

En parallèle, les mutations conduisant au passage d'une Prx robuste à une Prx sensible se sont produites chez l'ancêtre des eucaryotes, puis la Prx sensible a été transférée à l'ancêtre des Cyanobacteria (Figure 77A, carré gris).

Hypothèse 1 ou hypothèse 2 ?

Les données actuelles ne permettent pas de déduire le sens du transfert entre l'ancêtre des Eucarya et celui des Cyanobacteria. Néanmoins, étant donné que les cyanobactéries ne présentent qu'une seule copie de Prx (à savoir, la version sensible), si la seconde hypothèse est exacte, alors elle implique que la Prx bactérienne typique ait été remplacée dans ce groupe par la version sensible, ce qui semble rendre la première hypothèse plus parcimonieuse.

Quel que soit le sens du transfert, la Srx et la 2-Cys Prx sensible ont été conservées au cours de l'évolution et étaient présentes chez l'ancêtre des Cyanobacteria (Figure 77B, triangles et carrés blancs) et chez l'ancêtre des Eucarya (Figure 77C, triangles et carrés noirs).

3.2.3.2.3 Origine des deux copies chez les Plantae

Après l'endosymbiose photosynthétique primaire (Figure 77D), les Plantae avaient deux copies de Srx et deux copies de Prx (la copie eucaryote "ancestrale" et la copie nouvellement acquise depuis les Cyanobacteria). Au cours de leur diversification, elles ont retenu l'une ou l'autre selon les lignées (Figure 77E) :

- Le glaucophyte *Cyanophora paradoxa* aurait gardé la version eucaryote des deux protéines;
- L'algue verte *Volvox carteri*, elle, aurait gardé la version "eucaryote" de Srx tandis qu'elle aurait gardé la version chloroplastique de Prx;
- L'algue verte *Chlamydomonas reinhardtii* aurait gardé la version "eucaryote" de Srx, et les deux copies de Prx. Ceci est appuyé par le fait que la séquence qui groupe avec les 2-Cys Prx des autres eucaryotes n'est pas annotée comme chloroplastique dans la fiche GenPept, tandis que celle qui groupe avec les Prx cyanobactériennes l'est;
- Dans les autres lignées de plantes, les copies eucaryotes de Srx et Prx auraient été remplacées par celles provenant des cyanobactéries. Cette hypothèse est soutenue par la présence d'un peptide d'adressage au chloroplaste dans quasiment toutes les séquences de ce groupe (Tableau 10 et Tableau 11), et par la mise en évidence de ces enzymes dans le chloroplaste chez plusieurs espèces de Plantae (Baier and Dietz, 1997; Rey et al., 2007). De plus, il est également important de noter que toutes les espèces qui ont une Srx prédite comme étant chloroplastique ont également une Prx prédite comme étant chloroplastique;
- Enfin, plusieurs lignées auraient perdu les deux copies (par exemple l'algue verte *Ostreococcus*).

Ce scénario implique de nombreuses pertes. Cependant, même sans tenir compte de ce qui s'est passé chez les plantes, Srx semble avoir été perdue à maintes reprises dans toutes les lignées eucaryotes. Par exemple, chez les champignons, Srx n'est présente que dans quelques groupes comme les Saccharomycotina, les Schizosaccharomycota et les Basidiomycota. La même situation se retrouve chez les cyanobactéries, puisque tous les sous-groupes de cyanobactéries comprennent des espèces qui ont perdu Srx.

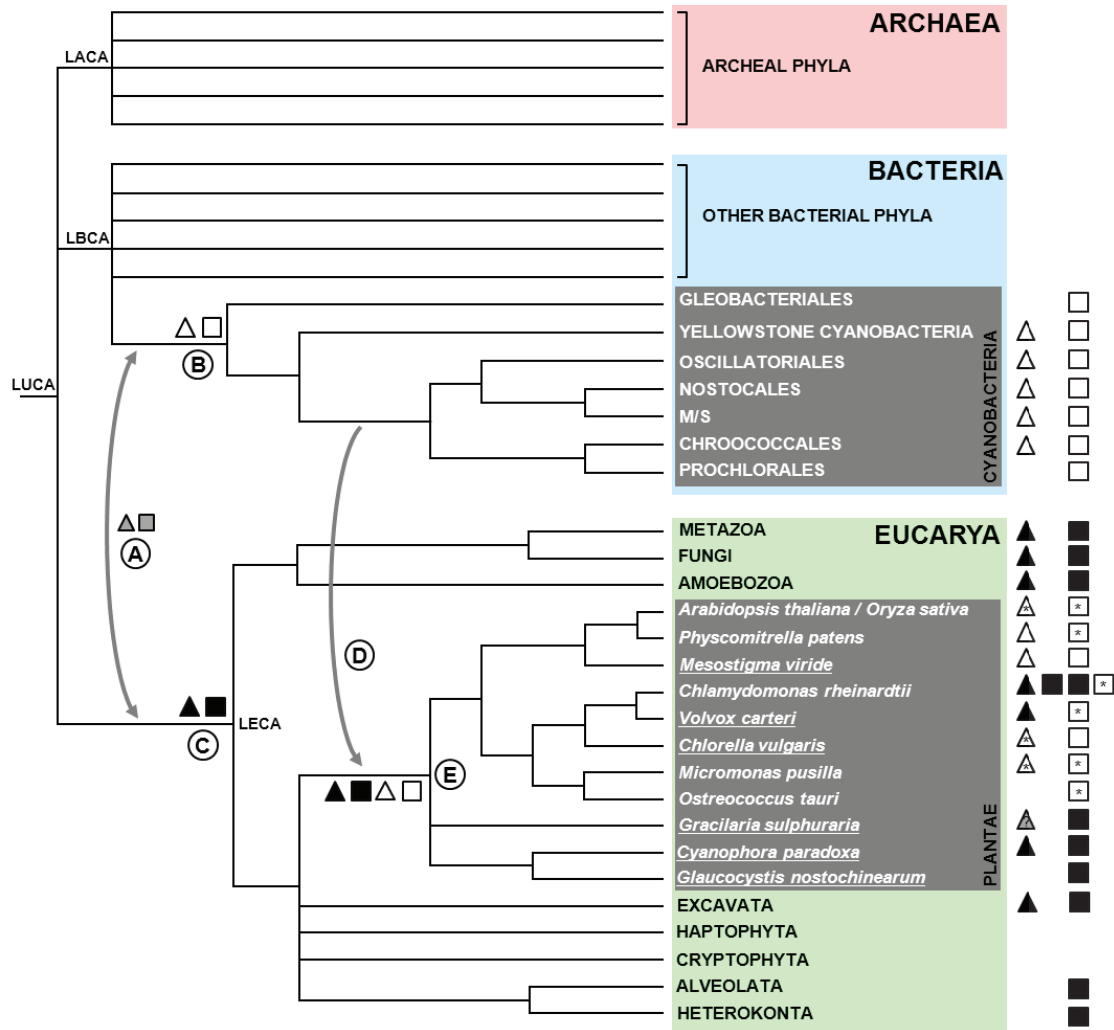


Figure 77. Scénario possible de l'origine et l'évolution des séquences de Srx et de 2-Cys Prx sensibles. L'arbre présente les trois domaines de la vie cellulaire (Archaea, Eucarya et Bacteria). Les séquences de Srx (triangle gris) et de 2-Cys Prx sensible (carré gris) sont apparues avant l'ancêtre des Cyanobacteria ou des Eucarya et ont été transférées horizontalement à l'autre lignée avant sa diversification (A). Srx et Prx ont été conservées au cours de l'évolution et étaient présentes chez l'ancêtre des Cyanobacteria (B, triangle et carré blancs) et chez l'ancêtre des Eucarya (C, triangle et carré noirs). Les séquences de Srx et 2-Cys Prx cyanobactériennes ont ensuite été acquises secondairement par les plantes via l'endosymbiose chloroplastique (D). Après cet événement, des copies eucaryotes ou cyanobactériennes de Srx et 2-Cys Prx ont été perdues au cours de l'évolution et de la diversification des lignées de plantes actuelles (E). Les séquences de Srx et de Prx portant un CTP à leur extrémité N-terminale sont marqués d'un astérisque (*). La séquence de Srx chez *G.sulphuraria* est indiquée en gris car elle est trop partielle pour en déterminer l'origine. Les espèces dont le génome complet n'est pas encore disponible sont soulignées. LECA, LACA, LBCA et LUCA désignent respectivement : le dernier ancêtre commun des Eucarya, des Archaea, des Bacteria et le dernier ancêtre commun universel.

3.3 Conclusion et perspectives

L'histoire évolutive de Srx est nécessairement étroitement liée à celle des 2-Cys Prx sensibles, puisque que la fonction de Srx est entièrement dépendante de la présence d'une 2-Cys Prx sensible. Les données actuelles montrent clairement que les 2-Cys Prx sensibles et les Srx ont une histoire évolutive similaire, ce qui indique que les deux partenaires ont co-évolué (Figure 77). Les données suggèrent un THG entre le dernier ancêtre commun des Eucarya et le dernier ancêtre commun des Cyanobacteria, sans toutefois pouvoir en préciser le sens. Ce transfert a donc eu lieu avant la diversification de chacun de ces deux groupes et il est intéressant de le replacer dans le cadre du débat sur l'origine des Eucarya. En effet, si les hypothèses 2D s'avèrent exactes (*i.e.* les Eucarya descendent d'une lignée d'archée), elles impliquent que les Archaea étaient déjà diversifiées (puisque les Eucarya avaient déjà émergé), alors que les Cyanobacteria ne l'étaient pas encore.

Enfin, la question se pose de pourquoi certaines cyanobactéries ont besoin d'une sulfirédoxine et pas d'autres. L'absence de Srx dans les autres espèces de cyanobactéries s'expliquerait alors soit par l'existence d'autres protéines capables d'assurer le même rôle que Srx, soit par la disparition du cycle de suroxydation/rétronréduction des 2-Cys Prx dans ces lignées. Maintenant que le rôle biochimique de Srx a été mis en évidence chez *Anabaena*, il reste à comprendre son rôle cellulaire précis dans la réponse au stress. Il sera également nécessaire d'étudier les mécanismes de perception du H₂O₂ et ceux impliqués dans la régulation des réponses appropriées, à la fois chez les cyanobactéries possédant une sulfirédoxine et chez les autres.

Pour conclure, de très récents papiers suggèrent un rôle jusque-là inconnu pour la 2-Cys Prx. Jusqu'à présent, les modèles de rythmes circadiens étaient tous basés sur des boucles de rétroaction de transcription-traduction. Cependant, il a été découvert dans ces dernières années qu'il existe des mécanismes de contrôle des oscillations circadiennes qui sont transcription-indépendantes, à la fois chez l'homme (O'Neill and Reddy, 2011), chez *O. tauri* (une algue verte) (O'Neill et al., 2011) et chez *Synechococcus elongatus* (une cyanobactérie) (Lakin-Thomas, 2006; Tomita et al., 2005). De façon très intéressante, il vient d'être montré, chez l'homme (O'Neill and Reddy, 2011) et chez *O. tauri* (O'Neill et al., 2011), que les 2-Cys Prx jouent un rôle dans ces mécanismes de contrôle transcription-indépendants. Il reste donc à confirmer qu'elle joue le même rôle chez *S. elongatus*. Cette découverte ouvre de nouvelles pistes, notamment chez les organismes qui présentent un rythme circadien et qui n'expriment pourtant aucun gène classiquement connu pour être impliqué dans ce cycle (par exemple, la levure et *C. elegans*) mais qui expriment la 2-Cys Prx.

AXE 2:

PHYLOGENIE DES EUCARYA

Le travail présenté ici est en cours, et certains résultats sont donc encore préliminaires.

1 Introduction

Comme détaillé dans l'introduction (cf. p.47), les relations de parenté entre les grands groupes eucaryotes restent encore mal connues. Déterminer ces relations reste l'une des questions les plus importantes et les plus controversées au sein des évolutionnistes.

Cette irrésolution peut avoir plusieurs origines :

- La possible radiation à l'origine des lignées majeures d'eucaryotes, c'est-à-dire une diversification trop rapide pour avoir été enregistrée au niveau des séquences (cf. p.29), certains parlant même de 'big-bang' (Philippe et al., 2000a);
- Les artefacts méthodologiques (biais de composition, vitesses d'évolutions hétérogènes) (cf. p.28) ;
- Les transferts latéraux de gènes entre eucaryotes (cf. p.30) ;
- La saturation moléculaire, phénomène particulièrement important lorsque l'on s'intéresse à des événements anciens (cf. p.30).

Une approche pour essayer d'améliorer la résolution de la phylogénie des eucaryotes a été proposée à la fin des années 1990 et consiste à unir l'information phylogénétique provenant de plusieurs gènes (parfois plusieurs centaines). L'idée est que, combiner le signal faible porté par chaque marqueur individuel qui reflète la véritable histoire des organismes pourrait en faire émerger un signal fort compensant les effets de la saturation et de la radiation. Par ailleurs, cette approche pourrait permettre d'atténuer les biais pouvant être présents dans chacun des gènes (à condition que les biais ne soient pas systématiquement dans le même sens), afin de palier au problème de TGH et des différences de vitesse d'évolution. Enfin, ces analyses doivent être couplées à l'utilisation de méthodes et de modèles d'évolution adaptés aux données.

Cependant, l'application de ces approches a été limitée par la quantité de données génomiques disponibles. Cette barrière a été progressivement levée grâce à la multiplication des projets de séquençage d'EST et de génomes complets, et c'est au début des années 2000 qu'ont été réalisées les premières concaténations massives de gènes.

Jusqu'à présent, la plupart des analyses consacrées à la reconstruction de phylogénies à grande échelle (*i.e.* à l'échelle d'un ou plusieurs domaines du vivant) a été fondée sur l'analyse des protéines de type informationnel (*i.e.* impliquées dans la transmission et l'expression de l'information génétique) (ex : protéines ribosomales, facteurs d'élongation, facteurs d'initiation, exosome, RNAPol, DNAPol, ...). Ce choix peut sembler intuitif puisque la plupart des gènes codant pour ces protéines présentent plusieurs avantages :

- Ils font généralement partie de familles monogéniques permettant de limiter le risque de paralogie cachée ;
- Ce sont habituellement des gènes qui évoluent lentement ;
- Ils ont une origine ancienne et sont bien conservés au sein de l'ensemble des organismes ;
- Ils sont rarement échangés par transferts horizontaux.

Toutes ces caractéristiques font de ces gènes une ressource de choix pour étudier l'évolution ancienne dans tous les domaines de la vie. Les études phylogénétiques du domaine Eucarya n'ont pas échappé à cette règle (Tableau 12). Cependant, les gènes impliqués dans des processus informationnels sont en nombre relativement limité. La conséquence en est que la quasi-totalité des études consacrées à la phylogénie des eucaryotes repose sur des jeux de données extrêmement similaires, voire identiques (Tableau 12). Ceci implique également qu'il

n'existe pas de point de comparaison permettant d'éventuellement remettre en cause les résultats inférés.

Auteurs	Nombre de protéines concaténées	Nombre d'espèces	Nombre de positions	Type de données
(Baptiste, Brinkmann et al. 2002)	123	23	25000	2/3 gènes informationnels (~60 protéines ribosomales)
(Philippe, Snell et al. 2004)	129	36	30339	Même gènes que Baptiste (+7), enrichi en espèces
(Rodriguez-Ezpeleta, Brinkmann et al. 2005)	143	34	30113	~70 protéines ribosomales
(Burki and Pawlowski 2006)	85	37	13258	~1/3 gènes informationnels
(Patron, Inagaki et al. 2007)	102	38	18425	Protéines de Rodriguez-Ezpeleta 2005, enrichi en espèces
(Rodriguez-Ezpeleta, Brinkmann et al. 2007)	143	64	31604	Protéines de Rodriguez-Ezpeleta 2005, enrichi en espèces
(Burki, Shalchian-Tabrizi et al. 2007)	123	49	29908	2/3 sont des protéines de Rodriguez-Ezpeleta 2005
(Burki, Shalchian-Tabrizi et al. 2008)	135	65	31921	Mêmes protéines qu'en 2007 (+12); enrichi en espèces
(Hampl, Hug et al. 2009)	143	48	35584	Mêmes protéines que Rodriguez-Ezpeleta 2007, Espèces différentes
(Minge, Silberman et al. 2009)	78	37	17234	Même protéines que Burki 2008, enrichi en espèces
(Baurain, Brinkmann et al. 2010)	108	57	19933	Même protéines que Rodriguez-Ezpeleta 2005, enrichi en espèces

Tableau 12. Détail des données utilisées pour la reconstruction de la phylogénie des eucaryotes dans les analyses récentes basées sur des concaténations massives de protéines.

A l'opposé, la plupart des protéines impliquées dans les fonctions de ménage (protéines opérationnelles) est considérée comme étant moins bien adaptée à l'étude de l'évolution ancienne :

- Elles évoluent plus rapidement que les protéines composant les systèmes informationnels;
- Elles sont moins bien conservées (*i.e.* elles sont plus souvent perdues);
- Elles appartiennent fréquemment à des familles multigéniques et complexes qui ont connu de nombreux événements de duplication et de pertes de gènes au cours de leur histoire évolutive, ce qui requiert des analyses préliminaires fastidieuses afin de délimiter les sous-familles d'orthologues;
- Elles sont souvent de grande taille (*i.e.* plusieurs centaines d'acides aminés), ce qui rend difficile l'utilisation des banques d'EST. En effet, la taille de ces EST dépasse rarement les 500 nucléotides, ce qui ne permet pas d'obtenir la totalité de l'ARN messenger correspondant à la protéine d'intérêt sans passer par une étape d'assemblage de plusieurs EST. De plus, ces données partielles ne permettent pas toujours d'identifier à quelle sous-famille appartiennent ces séquences (dans le cas de familles multigéniques).

Par conséquent, bien que les protéines opérationnelles soient plus nombreuses que les informationnelles, et souvent de plus grande taille, elles sont rarement utilisées (à part quelques exceptions) pour retracer l'évolution ancienne des eucaryotes.

Cependant, malgré leurs qualités intrinsèques, les jeux de données typiques constitués de protéines informationnelles se sont révélés insuffisants pour inférer de façon robuste l'ensemble des nœuds profonds de la phylogénie des eucaryotes (Burki et al., 2007; Parfrey et al., 2010 830). Bien sûr, des progrès sont attendus avec l'amélioration des méthodes et des modèles, et l'augmentation l'échantillonnage taxonomique. Il est également indispensable d'augmenter le nombre de données analysées en identifiant d'autres marqueurs protéiques utilisables pour inférer les relations de parenté au sein des eucaryotes. Ceux-ci, s'ils existent, se situent donc nécessairement parmi les protéines opérationnelles. Par ailleurs, l'utilisation de ces dernières permettra de confronter l'histoire évolutive qu'elles portent avec celle portée par les protéines informationnelles.

Le projet présenté ici constitue le deuxième axe majeur de ma thèse, dont l'objectif était d'identifier de nouveaux marqueurs originaux pour l'étude de la phylogénie des eucaryotes. Les protéines constitutives des SME sont une source potentiellement intéressante pour rechercher ces nouveaux marqueurs.

Les analyses phylogénomiques des systèmes que j'ai réalisées auparavant (*i.e.* analyse du midbody et de l'APC/C) m'ont permis de générer environ 200 jeux de protéines opérationnelles dont une centaine correspond à des protéines inférées comme étant anciennes et déjà présentes chez LECA. Ces protéines sont donc susceptibles de contenir un signal phylogénétique utilisable pour reconstruire les relations de parenté entre eucaryotes. Les analyses préliminaires rigoureuses nécessaires à distinguer orthologues et paralogues ont été réalisées lors de l'analyse phylogénomique des SME dont elles font partie et j'ai donc souhaité étudier l'information phylogénétique portée par cet ensemble conséquent de protéines opérationnelles.

2 Méthodes

2.1 Jeux de données analysés

J'ai reconstruit six types de supermatrices (après élimination manuelle des régions où l'homologie entre les sites est douteuse) :

1. La supermatrice APC/C (**SM APC**) a été construite par concaténation des composants, co-activateurs et cibles de l'APC/C inférés comme étant présents chez LECA (**62 espèces et 24 protéines, soit 3115 acides aminés**).
2. La supermatrice M75-67 (**SM M75-67**) a été construite par concaténation des composants du midbody inférés comme étant présents chez LECA. Parmi ceux-ci, certains ont été éliminés car leur répartition taxonomique peut laisser suspecter des transferts horizontaux entre eucaryotes, ou encore parce qu'ils évoluent très rapidement. La concaténation SM M75-67 est donc constituée de 75 protéines. Les jeux de données de l'analyse du midbody ayant été constitués en juin 2008, j'ai réalisé une mise à jour en février 2010 en y intégrant de nombreuses espèces eucaryotes séquencées entre-temps. Le nombre d'espèces est ainsi passé de 32 à 67 (Figure 78). J'ai utilisé à la fois des données protéiques et nucléotidiques provenant de génomes eucaryotes complets ou en cours de séquençage disponibles au NCBI, au JGI (<http://genome.jgi-psf.org/>), à la TBestDB (<http://www.bch.umontreal.ca/pepdb/pep.html>), à la banque d'EST de *Porphyra yezoensis* (<http://est.kazusa.or.jp/en/plant/porphyra/EST/blast.html>), à la banque dédiée à *Galdieria sulphuraria* (<http://genomics.msu.edu/cgi-bin/galdieria/blast.cgi>) et à celle dédiée au projet génome de *Cyanidioschyzon merolae* (<http://merolae.biol.s.u-tokyo.ac.jp/blast/blast.html>). D'autres espèces (qui sont souvent présentes dans les jeux de données informationnels) auraient été intéressantes à intégrer lors de la mise à jour mais n'ont pas pu l'être. En effet, comme mentionné précédemment, les protéines opérationnelles sont généralement de grande taille, et les banques d'EST de certaines espèces contiennent trop peu de données pour pouvoir récupérer l'ensemble des séquences me permettant de "reconstituer" les protéines du midbody. C'est par exemple le cas pour les excavés du genre *Malawimonas* ou le rhizaire *Cercomonas sp.*
Au total, la matrice contient donc **67 espèces et 75 protéines, soit 19244 acides aminés**.
3. La supermatrice M35-67 (**SM M35-67**) : des supermatrices plus petites ont également été construites, par retrait successif des marqueurs contenant le plus de données

manquantes (*i.e.* présents chez le moins d'espèces). Je présenterai notamment les résultats obtenus pour la supermatrice M35-67, qui correspond à la concaténation des **35 protéines** du midbody contenant le moins de données manquantes (soit **9617 positions**). Les jeux de données individuels correspondant à ces 35 protéines contiennent au maximum un tiers d'espèces manquantes.

4. La supermatrice RE (**SM RE**) : J'ai également réalisé une mise à jour des jeux de données utilisés lors de l'analyse publiée par Rodriguez-Ezpeleta en 2007 (155 jeux de données protéiques fournis par Hervé Philippe, communication personnelle). J'y ferai référence en tant que "SM RE" (pour Rodriguez-Ezpeleta). Comme évoqué précédemment, ces jeux de données sont de type informationnel, et il était intéressant de comparer la phylogénie obtenue sur la base de ces données en appliquant les mêmes méthodes qu'à nos propres jeux de données. J'ai donc souhaité que l'échantillonnage taxonomique soit comparable à celui que j'avais dans les jeux de données issus de l'analyse du midbody. Pour cela, j'ai retiré certaines espèces (Tableau 3) et j'en ai ajouté d'autres (Tableau 15). Au final, la matrice RE contient 56 espèces, et non 67 comme celle du midbody. En effet, par souci de temps, lors de la mise à jour des données RE, j'ai décidé de ne pas considérer certains organismes qui sont présents dans les données issues du midbody. C'est le cas par exemple des espèces pour lesquelles un autre représentant du même genre était déjà présent (ex : je n'ai ajouté qu'*Ostreococcus lucimarinus* et non *O. tauri*) ou encore lorsque le groupe était déjà bien représenté (par exemple, au sein des animaux, je n'ai pas jugé utile d'ajouter *Apis mellifera* ou *Danio rerio*). En revanche, lorsque l'espèce faisait déjà partie de la sélection d'espèce initiale de Rodriguez-Ezpeleta, mais était absente dans certains jeux de données, j'ai réalisé des vérifications systématiques de ces absences. La matrice RE contient donc **56 espèces et 155 protéines soit 29769 positions**.
5. Supermatrice M75-56 ("**SM M75-56**") : afin de pouvoir entreprendre des comparaisons rigoureuses entre les données de la supermatrice RE et les données "midbody", j'ai construit une supermatrice sur les **75 protéines (soit 19244 positions)** du midbody, mais cette fois, avec un échantillonnage taxonomique limité aux **56 espèces** présentes dans la supermatrice RE.
6. Enfin la concaténation des deux matrices RE et M75-56 a aboutit à une supermatrice de **56 espèces et 230 protéines (soit 49013 positions)**, que j'appellerai "**SM REM**" (pour Rodriguez-Ezpeleta + Midbody).

Les caractéristiques des différentes matrices présentées dans ce chapitre sont résumées dans le tableau ci-dessous :

Supermatrice	Nombre d'espèces	Nombre de protéines	Nombre de positions
APC	62	24	3115
M75-67	67	75	19244
M35-67	67	35	19244
RE	56	155	29769
M75-56	56	75	19244
REM	56	230	49013

Tableau 13. Caractéristiques des différentes supermatrices utilisées pour reconstruire la phylogénie des eucaryotes.

Quelle que soit l'origine des données, les phylogénies individuelles de chaque protéine ont été examinées afin d'éliminer les séquences résultant de duplications spécifiques à une lignée eucaryote (ex : chez les Viridiplantae), sans qu'il soit pour autant aisé de distinguer les différents groupes de paralogues. Dans ce cas, l'ensemble des séquences de la lignée a été écarté avant la concaténation afin de ne pas risquer d'introduire un signal biaisé.

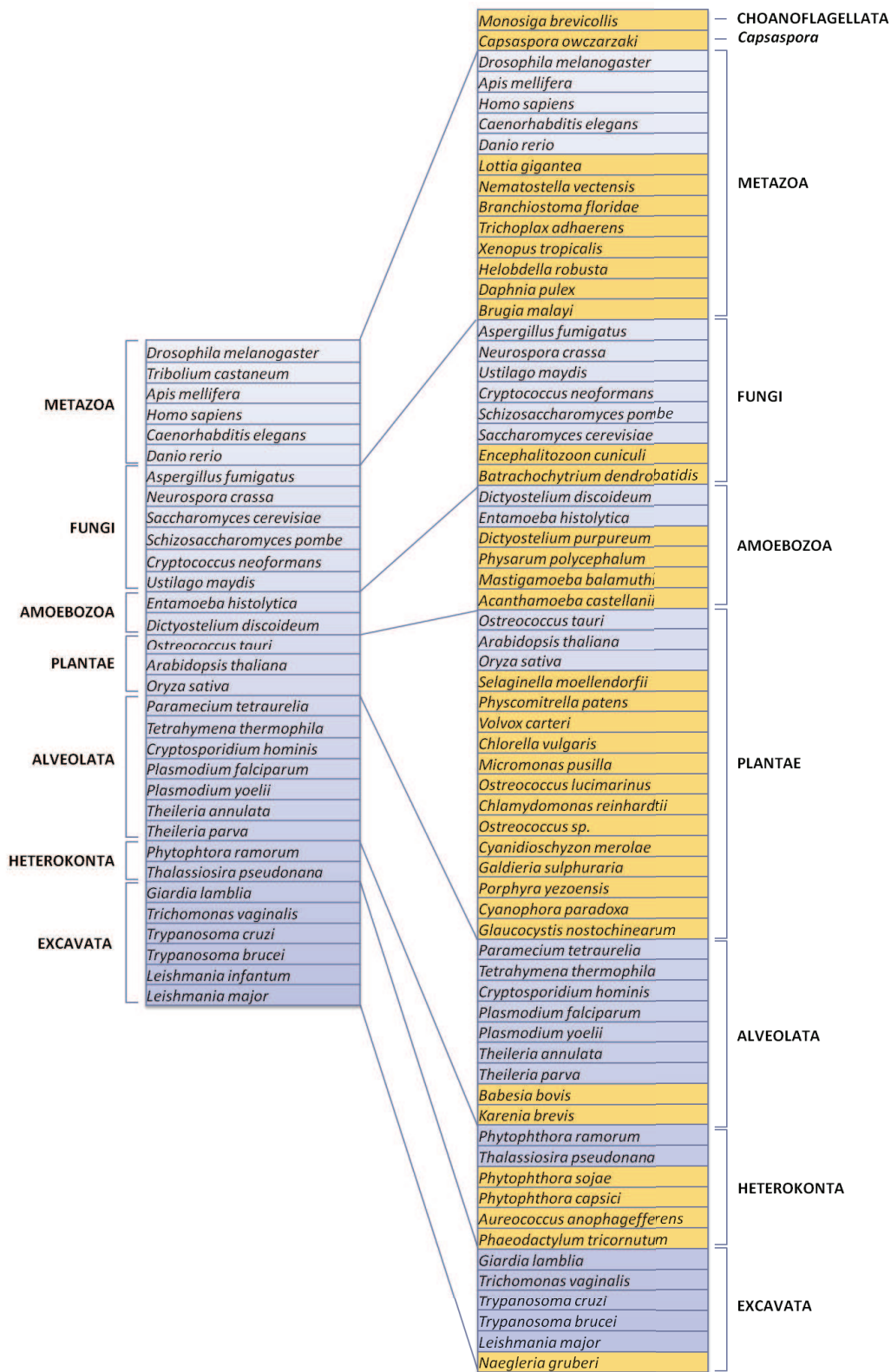


Figure 78. Liste des espèces incluses lors de la construction des jeux de données du midbody en juin 2008 (32 espèces, à gauche) et en février 2010 (67 espèces, à droite). Les espèces nouvellement incluses lors de la mise à jour sont indiqués en jaune.

Alexandrium tamarense	Euglena gracilis	Perkinsus marinus
Bigelowiella natans	Gracilaria changii/Chondrus crispus	Phanerochaete chrysosporium
Blastocladiella emersonii	Hartmannella vermiformis	Pinus taeda
Blastocystis hominis	Histiona aroides	Reclinomonas americana
Bombyx mori	Hydra magnipapillata	Rhizopus oryzae
Cercomonas longicauda	Ixodes scapularis	Sawyeria marylandensis
Ciona savignyi	Jakoba bahamiensis	Seculamonas ecuadoriensis
Cryptosporidium parvum	Jakoba libera	Stachyamoeba lipophora
Dunaliella salina	Malawimonas californiana	Sterkiella histriomuscorum
Ectocarpus siliculosus	Malawimonas jakobiformis	Strongylocentrotus purpuratus
Eimeria tenella	Monosiga ovata	Toxoplasma gondii

Tableau 14. Liste des espèces présentes dans le jeu de données initial de Rodriguez-Ezpeleta 2007 non conservées dans la supermatrice RE. Ces espèces n'ont pas été utilisées pour mettre à jour les jeux de données issus de l'analyse du midbody, soit parce que les représentants du groupe auquel elles appartiennent sont assez nombreux, soit parce que les données de séquençage EST disponibles ne sont pas assez nombreuses.

Acanthamoeba castellanii	Micromonas pusilla
Apis mellifera	Monosiga brevicollis
Arabidopsis thaliana	Naegleria gruberi
Aspergillus fumigatus	Nematostella vectensis
Aureococcus anophagefferens	Neurospora crassa
Babesia bovis	Oryza sativa
Batrachochytrium dendrobatidis	Ostreococcus lucimarinus
Branchiostoma floridae	Ostreococcus sp.
Brugia malayi	Ostreococcus tauri
Caenorhabditis elegans	Paramecium tetraurelia
Capsaspora owczarzewski	Phaeodactylum tricoratum
Chlamydomonas reinhardtii	Physarum polycephalum
Chlorella vulgaris	Physcomitrella patens
Cryptococcus neoformans	Phytophthora capsici
Cryptosporidium hominis	Phytophthora ramorum
Cyanidioschyzon merolae	Phytophthora sojae
Cyanophora paradoxa	Plasmodium falciparum
Danio rerio	Plasmodium yoelii
Daphnia pulex	Porphyra yezoensis
Dictyostelium discoideum	Saccharomyces cerevisiae
Dictyostelium purpureum	Schizosaccharomyces pombe
Drosophila melanogaster	Selaginella moellendorffii
Emiliania huxleyi	Tetrahymena thermophila
Encephalitozoon cuniculi	Thalassiosira pseudonana
Entamoeba histolytica	Theileria annulata
Galdieria sulphuraria	Theileria parva
Giardia lamblia	Trichomonas vaginalis
Glaucozystis nostochinearum	Trichoplax adhaerens
Helobdella robusta	Trypanosoma brucei
Homo sapiens	Trypanosoma cruzi
Karenia brevis	Ustilago maydis
Leishmania major	Volvox carteri
Lottia gigantea	Xenopus tropicalis
Mastigamoeba balamuthi	

Tableau 15. Liste d'espèces utilisée pour la mise à jour des jeux de données de Rodriguez-Ezpeleta. Les 67 espèces listées sont celles présentes dans la supermatrice midbody. Les espèces en bleu étaient initialement présentes dans le jeu de données de Rodriguez-Ezpeleta mais les absences de certains orthologues ont tout de même été vérifiées. Les espèces en jaune n'étaient pas présentes dans ce jeu de données et ont été ajoutées lors de la mise à jour. Celles en gris n'étaient pas présentes non plus dans le jeu de données initial (à l'exception de *Theileria parva* et *Trypanosoma cruzi*) et n'ont pas été insérées lors des mises à jour, soit parce que les représentants du groupe auquel elles appartiennent sont assez nombreux (ex : *Apis mellifera*, *Xenopus tropicalis* et *Danio rerio*), soit parce que d'autres représentants du même genre sont déjà présents (ex : *Ostreococcus tauri*, *Dictyostelium purpureum*, etc.).

2.2 Analyses phylogénétiques

2.2.1 Méthodes de reconstruction

Pour chacune des concaténations, les arbres phylogénétiques ont été calculés en utilisant des méthodes de maximum de vraisemblance (ML, pour Maximum Likelihood) en utilisant Treefinder (Jobb et al., 2004) et PhyML (Guindon and Gascuel, 2003). Les arbres reconstruits par PhyML se sont montrés mieux résolus que ceux reconstruits par Treefinder. Pour cette raison je ne présenterai que les premiers. J'ai également utilisé une méthode bayésienne, *via* le programme Phylobayes (Lartillot et al., 2009) (calculs lancés sur deux chaînes indépendantes pendant au moins 10000 cycles, avec échantillonnage d'un arbre sur 10; élimination des 300 premiers arbres, et test de la convergence des deux chaînes sur les arbres restants; puis calcul de l'arbre consensus à 50%).

2.2.2 Support statistique

La robustesse des branches des arbres reconstruits en maximum de vraisemblance a été estimée, dans un premier temps, par SH-like (pour PhyML) ou par LR-ELW (pour Treefinder). Ces mesures constituent une approximation du bootstrap permettant une économie de temps de calcul non négligeable lors des premières phases de l'analyse. Dans un second temps, les arbres présentant une topologie intéressante ont été recalculés en utilisant la procédure de bootstrap non paramétrique (100 réplicats de l'alignement d'origine) implémentée dans PhyML ou Treefinder. Il faut préciser qu'à ce jour certains calculs de bootstrap sont toujours en cours.

2.2.3 Recodage

Pour tenter de limiter l'influence des biais de composition, j'ai recodé les supermatrices en 4 états de caractère sur la base des propriétés physico-chimiques des acides aminés (Dayhoff et al., 1978). Un tel recodage permet en théorie de réduire le risque d'artefact d'attraction des longues branches en diminuant le nombre d'homoplasies ainsi que certains biais compositionnels (Susko and Roger, 2007). En effet, le recodage, en une seule lettre, de plusieurs acides aminés aux propriétés biochimiques similaires, peut réduire le signal potentiellement saturé provoqué par les substitutions répétées entre ces acides aminés. De plus, il peut également éliminer les éventuels biais de composition (par exemple, les fréquences relatives de l'arginine et la lysine au sein des protéines sont très sensibles à la teneur en GC du génome (Susko and Roger, 2007)).

Les alignements ont donc ainsi été recodés en 4 catégories. Dans ce recodage, les acides aminés aromatiques (FWY) et hydrophobes (MILV) ont été regroupés en un seul état, les acides aminés basiques (HKR) dans un autre, les acides aminés acides (DENQ) dans un troisième état, et le quatrième état regroupe les autres acides aminés (AGPST) à l'exception de la cystéine, qui a été codée comme des données manquantes.

2.2.4 Modèles d'évolution

Pour les reconstructions en ML, le modèle d'évolution LG a été utilisé pour les concaténations protéiques, et le modèle GTR pour les concaténations après recodage. Dans tous les cas, une loi gamma a été appliquée afin de prendre en compte l'hétérogénéité des taux d'évolution entre les sites (4 classes distinctes de sites et estimation du paramètre alpha). Pour les reconstructions par Phylobayes, j'ai utilisé le modèle CAT + dirichlet.

2.2.5 Saturation et Vitesse d'évolution des marqueurs

La saturation interne des différents jeux de données a été estimée par comparaison de la distance évolutive apparente (*p*-distance) avec la distance évolutive inférée sous le modèle d'évolution choisi. La vitesse d'évolution des concaténations prises deux à deux a aussi été estimée *via* la comparaison des distances évolutives relatives inférées pour chaque paire d'espèces et estimée à partir des arbres.

3 Résultats

Je présenterai dans un premier temps les topologies générales obtenues à partir de chaque supermatrice, avant de discuter de certains points précis de la phylogénie des eucaryotes.

3.1 Les concaténations "midbody"

3.1.1 Effet du recodage

Ici sont comparés deux arbres reconstruits par PhyML à partir de la supermatrice M75-56 : l'un basé sur l'alignement recodé (Figure 80), l'autre sur l'alignement non recodé (56 espèces, 75 protéines, 19244 positions) (Figure 79). Les arbres présentent des topologies assez similaires. Néanmoins, le recodage a un effet positif sur ce jeu de données puisqu'il permet de replacer la microsporidie au sein des Fungi, et de retrouver la monophylie des Excavata, ainsi que le regroupement des Heterokonta avec les Alveolata.

Par ailleurs, le support aux nœuds n'est pas affaibli, ce qui semble en accord avec la conclusion de Susko et Roger sur le fait que le recodage n'aboutit pas à une perte importante de l'information (Susko and Roger, 2007).

J'ai également observé un effet positif du recodage sur les reconstructions à partir d'autres jeux de données (M75-67, M35-67, RE et REM), et, par la suite, je présenterai ces arbres plutôt que ceux basés sur les alignements "non recodés".

En revanche, dans le cas de la fusion APC, le recodage a donné de mauvais résultats. Par exemple, le placement des microsporidies au sein des champignons est retrouvé à partir de l'alignement non recodé alors qu'il ne l'est pas sur la base de l'alignement recodé. Il est possible que cette méthode élimine une trop grande partie du signal contenu dans la matrice APC qui est relativement petite (~3000 positions) comparée aux autres matrices présentées dans cette étude.

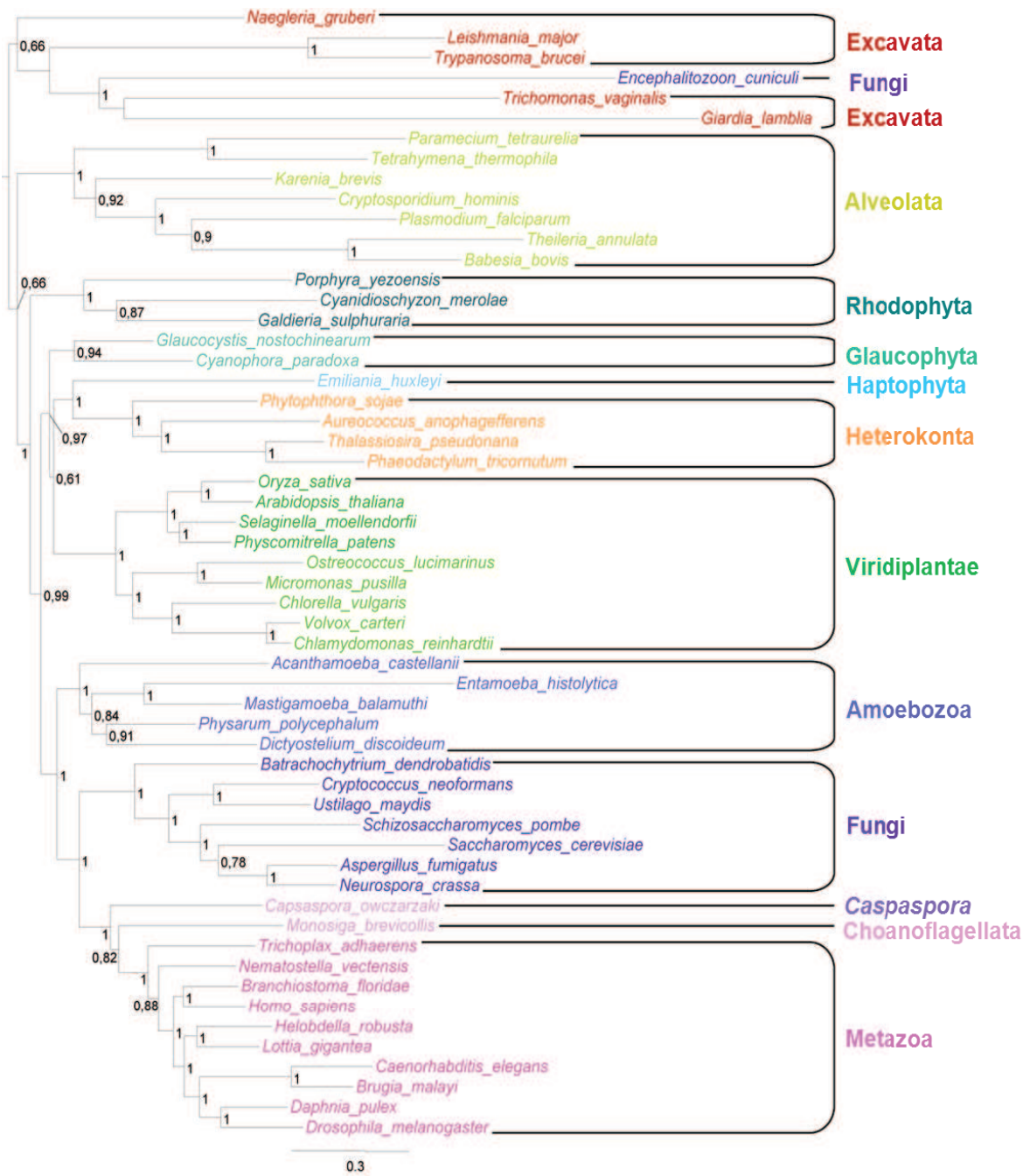


Figure 79. Arbre en Maximum de Vraisemblance calculé sur SM M75-56 (56 espèces, 19244 positions, PhyML). Les valeurs aux nœuds correspondent aux supports SH-like. Seules les valeurs supérieures à 0.50 sont indiquées. La barre d'échelle indique le nombre moyen de substitutions par site.

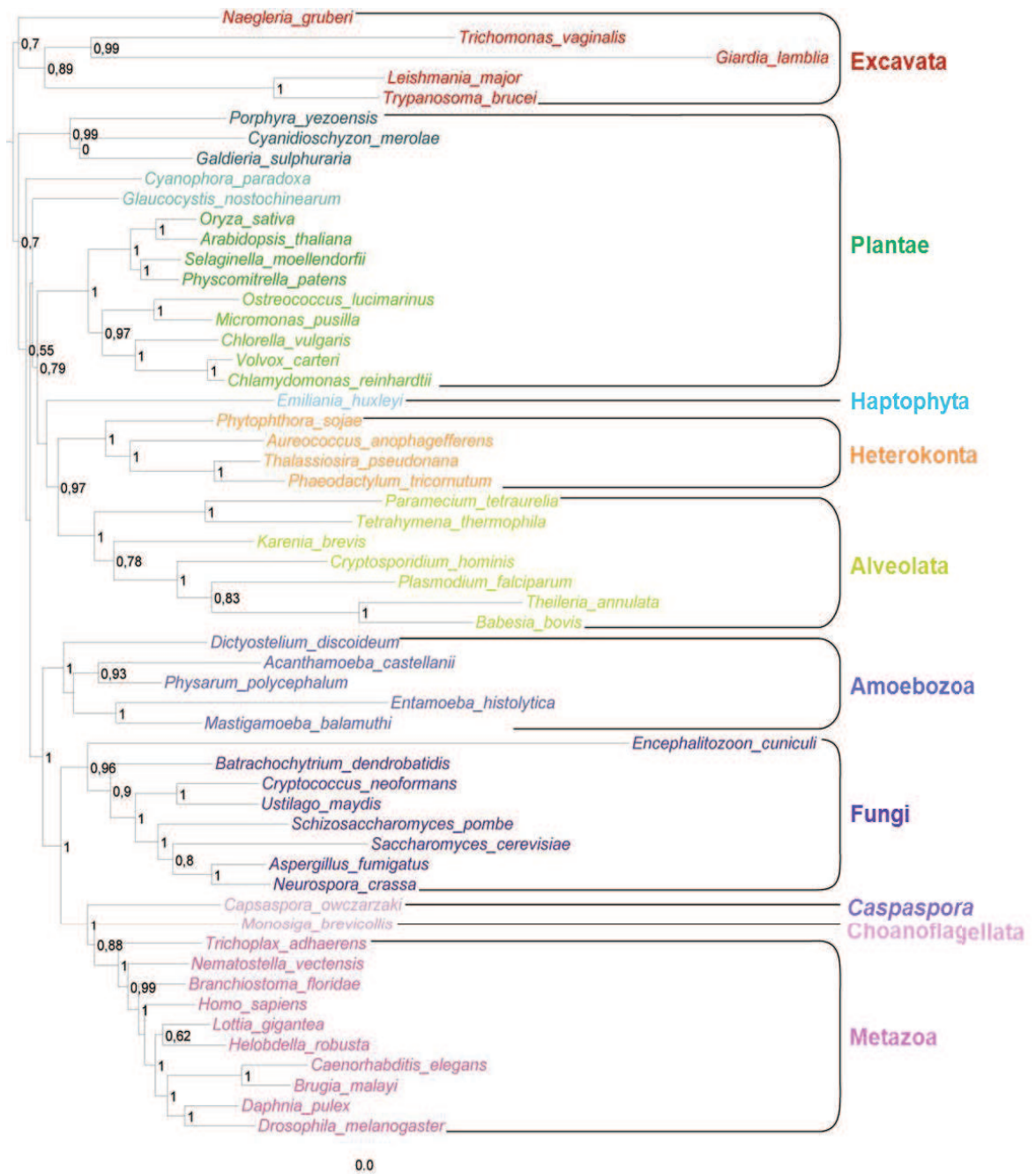


Figure 80. Arbre en Maximum de Vraisemblance calculé sur SM M75-56 recodé (56 espèces, 19244 positions, PhyML). Les valeurs aux nœuds correspondent aux valeurs de bootstrap. La barre d'échelle indique le nombre moyen de substitutions par site.

3.1.2 La supermatrice M75-67 (concaténation de 75 protéines)

L'arbre reconstruit en maximum de vraisemblance (*via* PhyML) à partir de la supermatrice contenant 75 protéines du midbody (19244 positions recodées) avec l'échantillonnage taxonomique le plus large permet de retrouver la monophylie des grands phylums (Figure 81). Ceux-ci obtiennent le plus souvent un soutien maximal, même s'il ne s'agit pour l'instant que de support SH-like qui est probablement surestimé : Alveolata (SH-like = 1.0), Heterokonta (SH-like = 1.0), Viridiplantae (SH-like = 1.0), Excavata (SH-like = 0.86), Rhodophyta (SH-like = 0.99), Metazoa (SH-like = 1), Fungi (SH-like = 0.95), Amoebozoa (SH-like = 1.0). En revanche, les relations entre les groupes sont très faiblement soutenues, à l'exception de la séparation entre Unikontes et Bikontes (SH-like = 1.0).

Ces résultats indiquent que ces protéines opérationnelles contiennent un signal phylogénétique intéressant, et qu'elles constituent donc de nouveaux marqueurs exploitables pour étudier la phylogénie des eucaryotes.

3.1.3 La supermatrice M35-67 (concaténation de 35 protéines)

L'objectif ici était de tester l'influence des données manquantes en réduisant le nombre de marqueurs, mais en conservant ceux qui contiennent le moins de données manquantes, c'est-à-dire les marqueurs présentant au maximum un tiers d'espèces manquantes.

J'ai donc concaténé uniquement les 35 protéines du midbody (supermatrice "M35-67") contenant le moins de données manquantes (9617 positions recodées analysées avec Phyml) (Figure 82). On retrouve quasiment la même topologie que lorsque les 75 protéines sont utilisées mais il est intéressant de remarquer qu'on retrouve la monophylie des Plantae (*i.e.* Viridiplantae + Rhodophyta + Glaucophyta), ce qui n'est pas le cas avec la supermatrice M75-67. Néanmoins, le support statistique pour ce regroupement est extrêmement faible : VB < 50%.

Outre, l'impact des données manquantes, il est possible que les protéines retirées (*i.e.* contenant le plus de données manquantes) soient également celles qui évoluent le plus vite, et donc les plus soumises à saturation. De façon surprenante, lorsque les distances évolutives inférées à partir des deux jeux de données sont comparées, il est notable que le jeu de données M35-67 évolue plus rapidement que M75-67 (Figure 83). Il est possible que la grande quantité de données manquantes dans la matrice M75-67 ait pour conséquence une sous-estimation des vitesses d'évolution inférées par la méthode de maximum de vraisemblance, mais ce point reste à éclaircir.

Enfin, les valeurs de bootstrap associées aux nœuds profonds sont extrêmement faibles bien que tous les grands groupes eucaryotes soient retrouvés comme étant monophylétiques.



Figure 81. Arbre en Maximum de Vraisemblance calculé sur SM M75-67 (67 espèces, 19244 positions, alignement recodé, PhyML). Les valeurs aux nœuds correspondent aux supports SH-like. Seules les valeurs supérieures à 0.50 sont indiquées. La barre d'échelle indique le nombre moyen de substitutions par site.

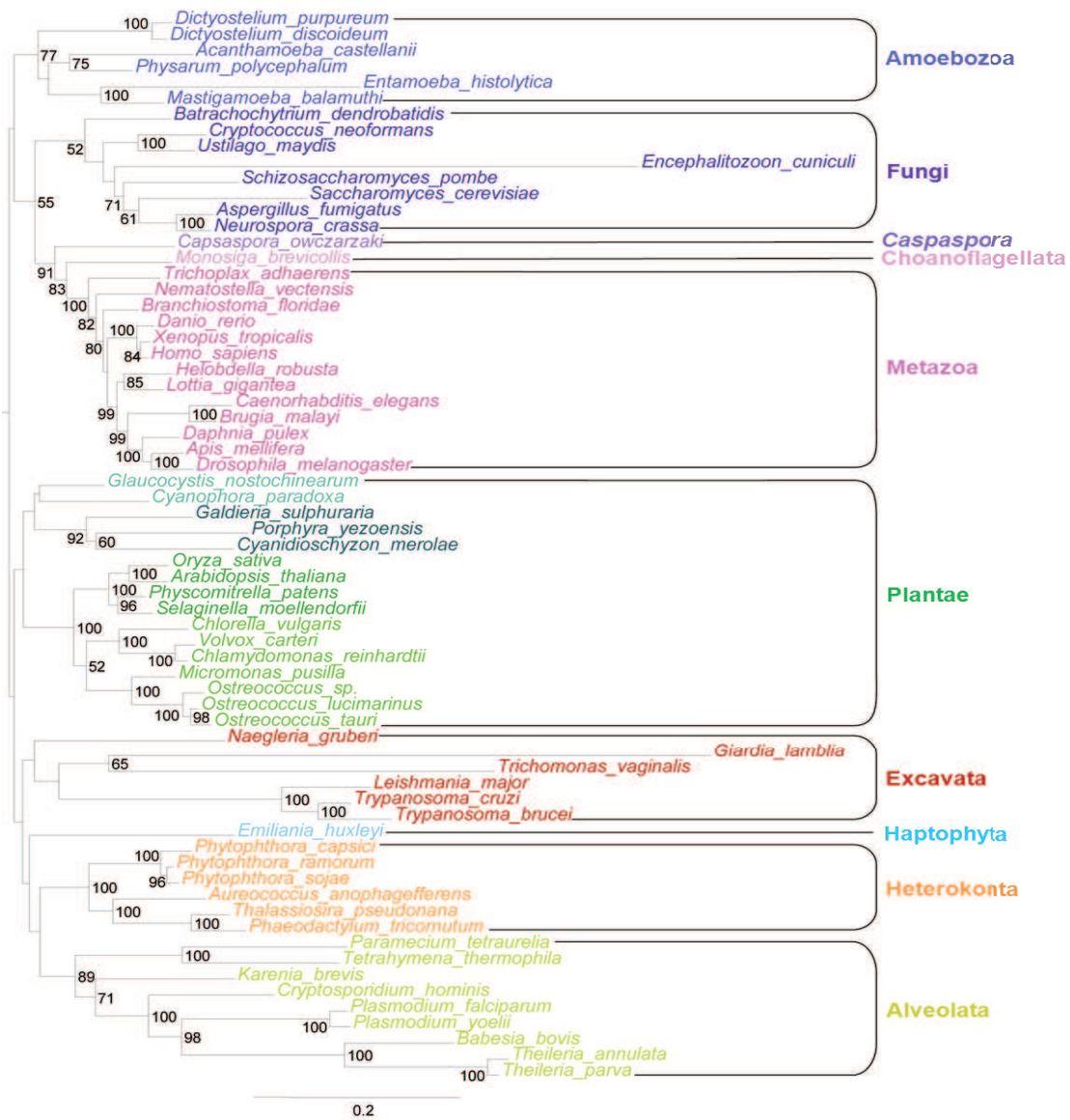


Figure 82. Arbre en Maximum de Vraisemblance calculé sur SM M35-67 (67 espèces, 9617 positions, alignement recodé, PhyML). Les valeurs aux nœuds correspondent aux valeurs de bootstrap. Seules les valeurs supérieures à 50% sont indiquées. La barre d'échelle indique le nombre moyen de substitutions par site.

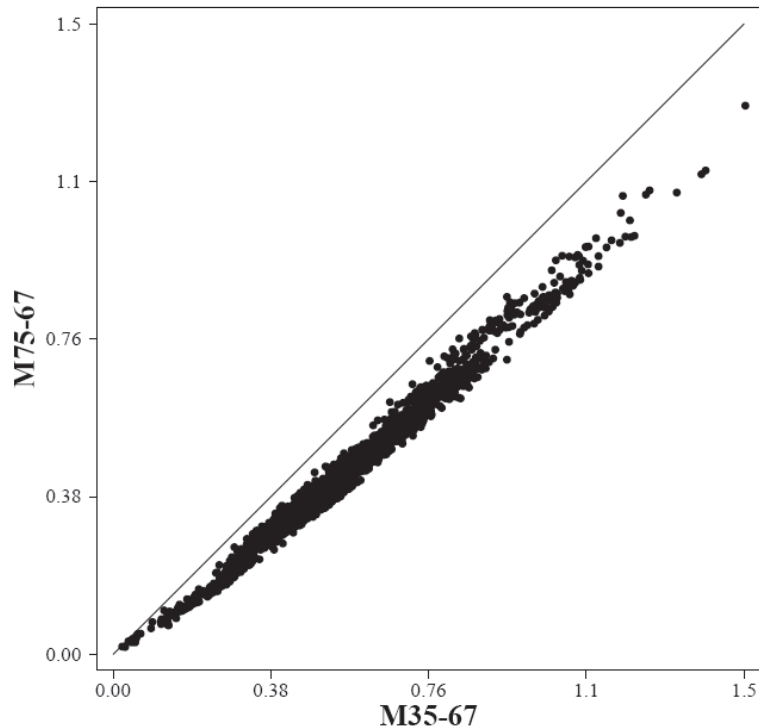


Figure 83. Comparaison des distances évolutives (nombre de substitutions/site) déduites des arbres ML basés sur les supermatrices M35-67 et M75-67. Les coordonnées de chaque point représentent la distance évolutive séparant une paire d'espèces dans l'un et l'autre des arbres reconstruits à partir de ces deux matrices. Toutes les paires d'espèces sont représentées. De façon surprenante, les distances évolutives observées sont plus faibles dans le cas de M75-67, ce qui suggère une vitesse d'évolution moindre des protéines constituant la matrice, comparativement à M35-67.

3.2 Concaténation "Rodriguez-Ezpeleta" mise à jour

Comme mentionné précédemment, j'ai mis à jour les jeux de données utilisés par Rodriguez-Ezpeleta et collaborateurs pour leur publication de 2007 (Rodriguez-Ezpeleta et al., 2007a). L'arbre reconstruit en ML (par PhyML, Figure 85) sur la base de cette supermatrice (SM RE) a une topologie très semblable à celle retrouvée par les auteurs (Figure 84), tout en étant construit sur un échantillonnage taxonomique différent comprenant des lignées qui n'étaient précédemment pas représentées (ex : la microsporidie, *Capsaspora*) (Tableau 15) et d'autres qui étaient présentes dans l'étude originale mais ne le sont plus dans la matrice RE (Tableau 14). A ce propos, il est intéressant de remarquer que l'on replace bien la microsporidie (*Encephalitozoon cuniculi*) avec les Fungi (VB = 56%). En revanche, il est étonnant de remarquer qu'on ne retrouve pas la monophylie des Unikontes. Il est possible que les Excavata (dont certains représentants sont portés par de longues branches) soient attirés par la longue branche d'*E.cuniculi*. Par ailleurs, il faut noter que le soutien pour les nœuds profonds est généralement faible. Le point le plus important est que cette topologie est également congruente avec celle de l'arbre M75-56 (Figure 86, identique à la Figure 80, mais portant les valeurs de bootstrap), à la différence que ce dernier ne permet pas de retrouver la monophylie des Plantae (qui est retrouvée dans l'arbre RE avec un support faible, VB = 50%), mais permet de retrouver celles des Unikontes (VB = 72%).

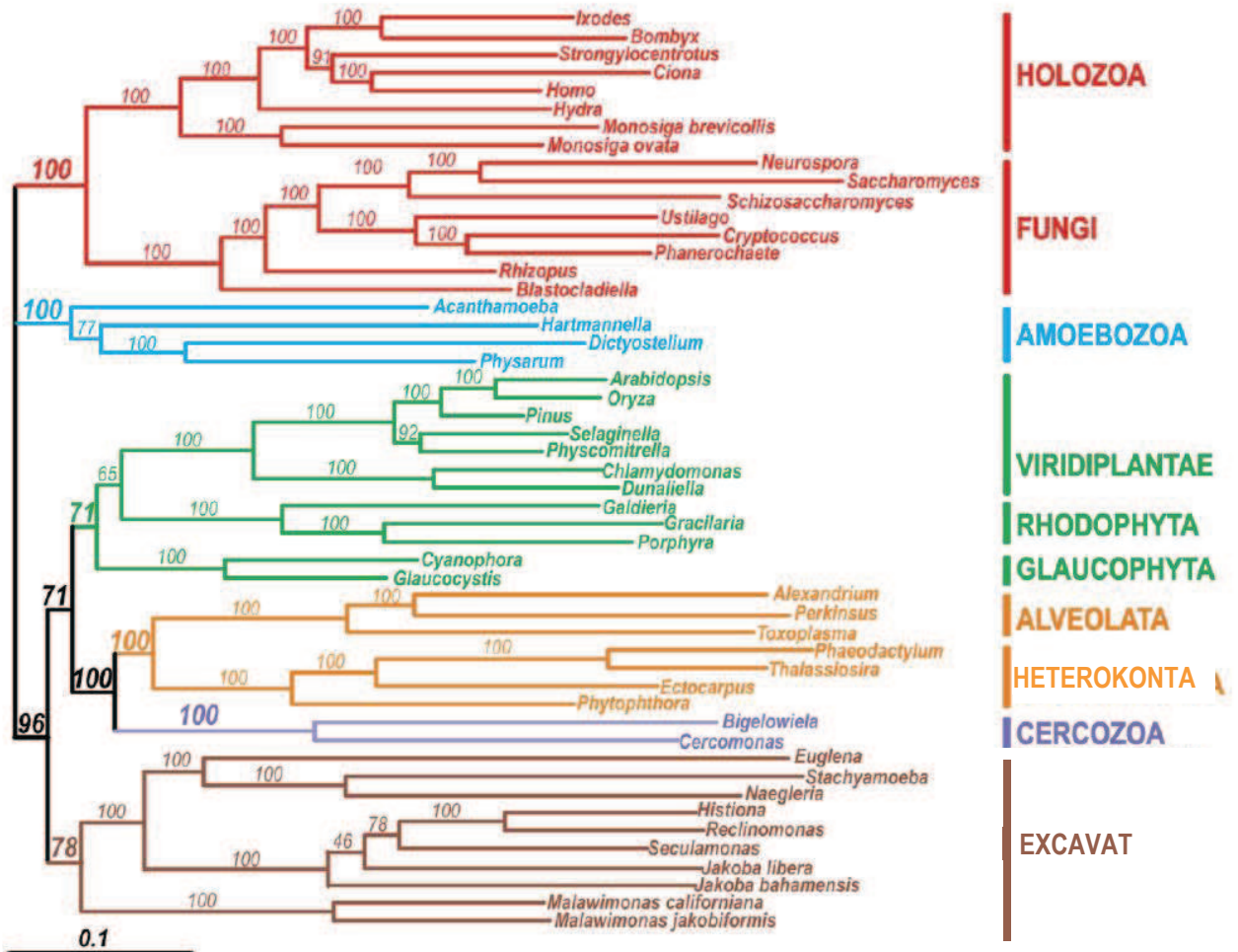


Figure 84. Arbre en Maximum de Vraisemblance tiré de (Rodriguez-Ezpeleta et al., 2007a). L'arbre comprend 51 espèces et est basé sur 143 protéines concaténées (31 604 positions). Les chiffres indiquent les valeurs de bootstrap. La barre d'échelle représente le nombre estimé de substitutions d'acides aminés par site.

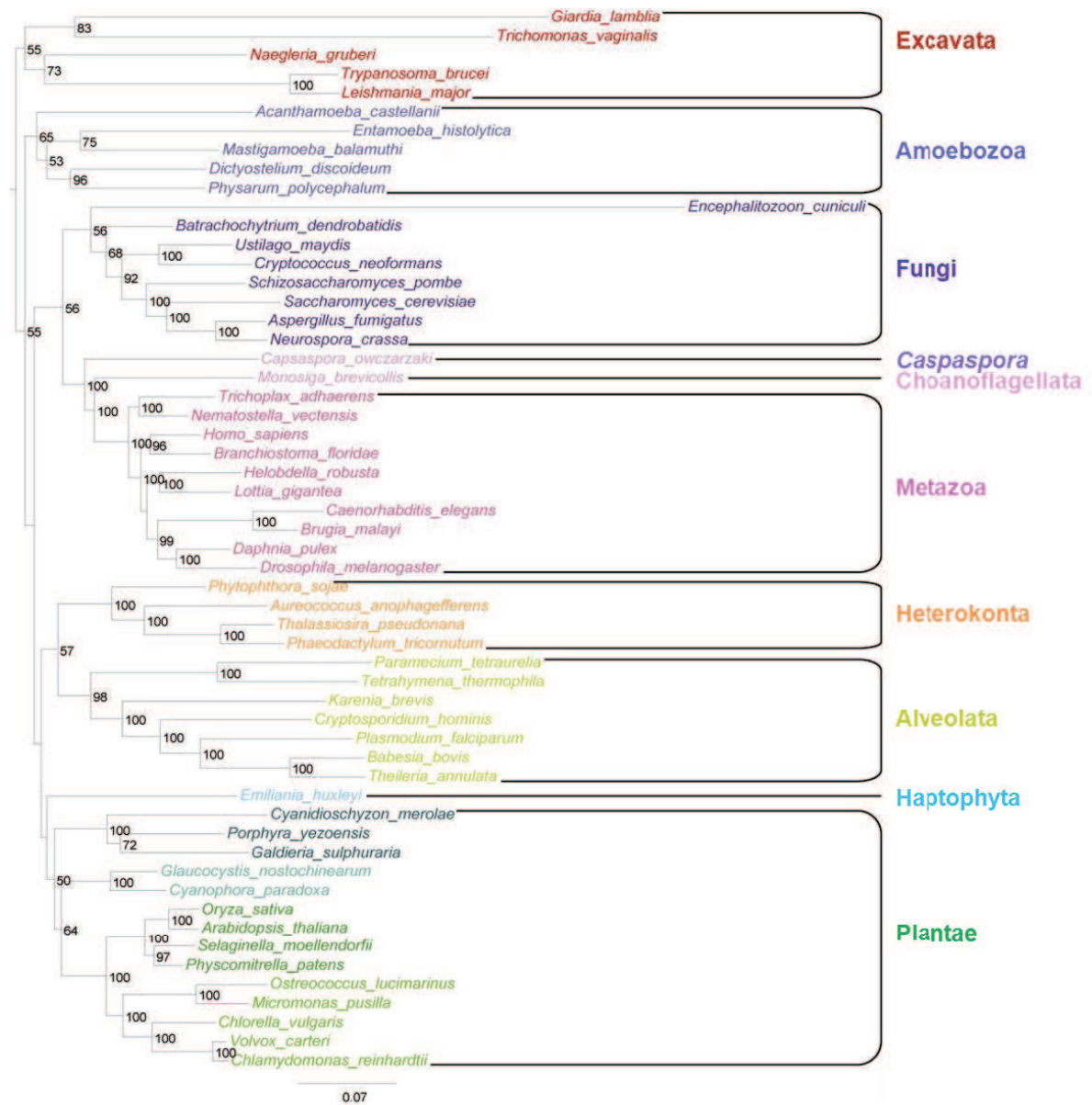


Figure 85. Arbre en Maximum de Vraisemblance calculé sur la SM RE (56 espèces, 29769 positions, alignement recodé, PhyML). Les valeurs aux nœuds correspondent aux valeurs de bootstrap. Seules les valeurs supérieures à 50% sont indiquées. La barre d'échelle indique le nombre moyen de substitutions par site.

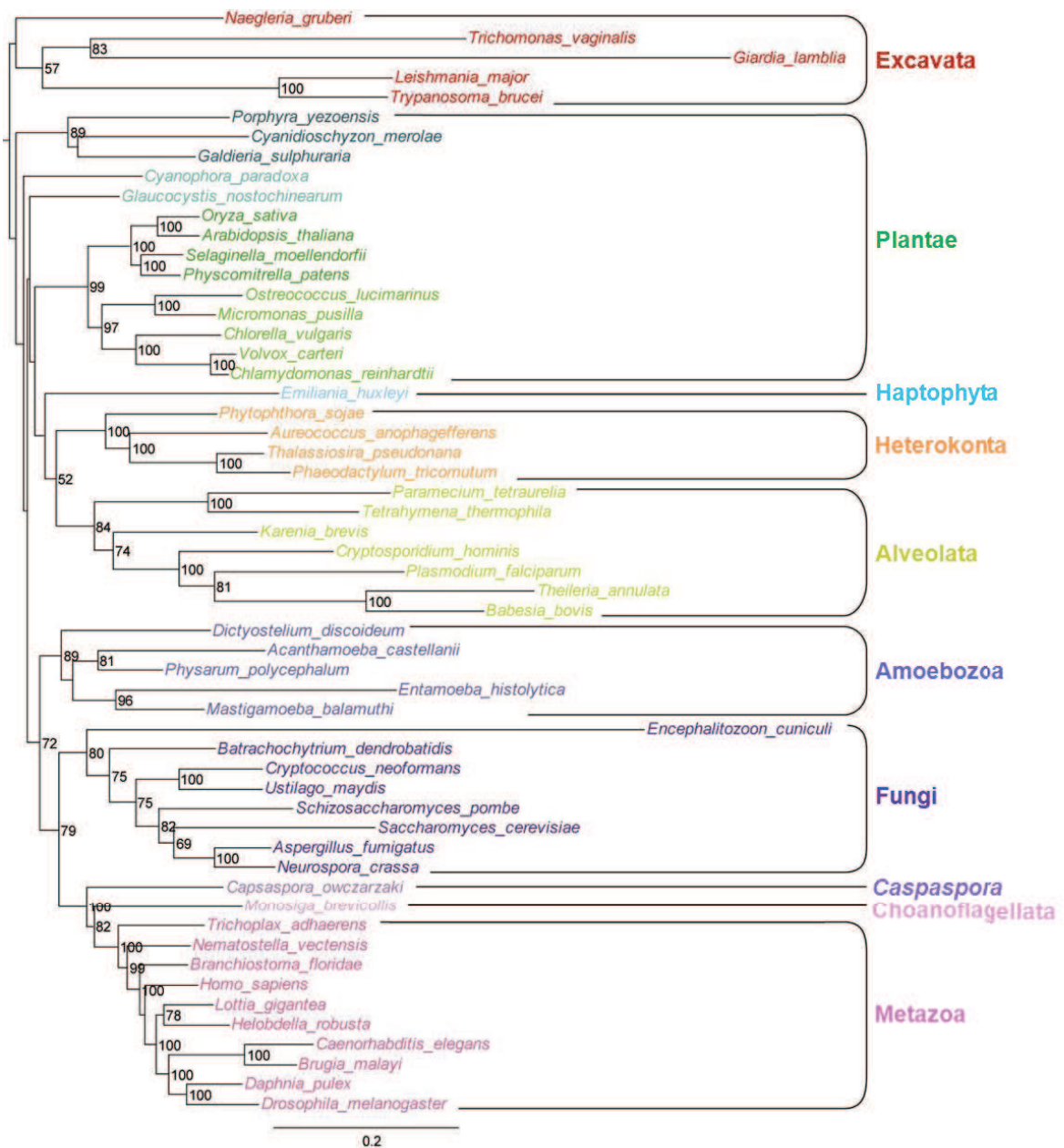


Figure 86. Arbre en Maximum de Vraisemblance calculé sur SM M75-56 (56 espèces, 19244 positions, alignement recodé, PhyML). Les valeurs aux nœuds correspondent aux valeurs de bootstrap. Seules les valeurs supérieures à 50% sont indiquées. La barre d'échelle indique le nombre moyen de substitutions par site.

Comparaison des vitesses d'évolution des matrices RE et M75-56

J'ai souhaité comparer les distances évolutives inférées à partir des deux jeux de données (RE et M75-56). Comme attendu, le jeu de données M75-56 évolue globalement plus vite (~1,5 fois plus rapidement) que le jeu de données RE, indiquant une saturation potentiellement plus importante des données M75-56.

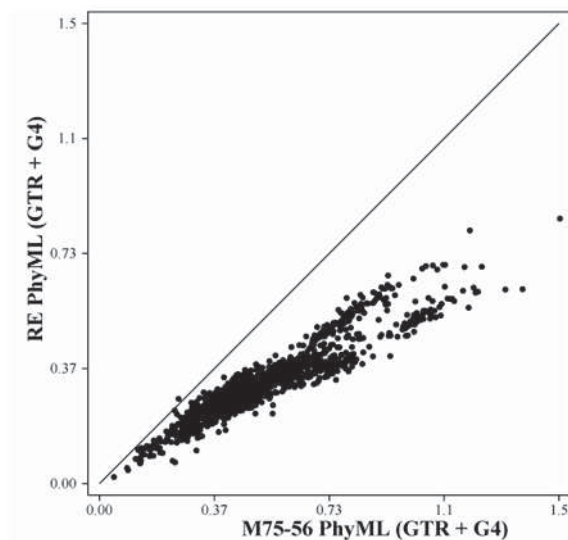


Figure 87. Comparaison des vitesses d'évolution des supermatrices M75-56 et RE. Les coordonnées de chaque point représentent la distance évolutive séparant une paire d'espèces dans l'un et l'autre des arbres reconstruits à partir de ces deux matrices. Toutes les paires d'espèces sont représentées.

Comparaison du nombre de données manquantes

Un aspect intéressant à considérer lorsque l'on travaille à partir de supermatrices est la quantité de données manquantes qu'elles contiennent, même s'il a été argumenté qu'elles n'ont pas un fort impact sur la reconstruction phylogénétique en ML (Philippe et al., 2004). J'ai donc comparé la quantité de données manquantes (en pourcentage du nombre total de positions) pour chaque espèce dans les deux matrices M75-56 et RE (Figure 88).

Comme attendu par le fait que M75-56 est constituée de protéines opérationnelles, cette matrice contient plus de données manquantes que la matrice RE (38% en moyenne contre 24%). Néanmoins, il est intéressant de remarquer que, dans la matrice M75-56, seule la moitié des espèces présentent plus de 30% de données manquantes, et qu'un tiers seulement présente plus de 50% de données manquantes. Ces absences conséquentes dans certaines espèces peuvent s'expliquer de deux façons : soit il s'agit d'espèces pour lesquelles les données génomiques sont encore partielles, et donc pour lesquelles je n'ai pas pu récupérer l'ensemble des homologues, soit ces absences sont le reflet de pertes réelles, comme observées dans certaines lignées lors de l'analyse des SME (ex : Alveolata).

Enfin, bien que la quantité de données manquantes soit plus importante dans la matrice M75-56 que la matrice RE, cela ne semble pas affecter de façon dramatique la topologie, puisque les arbres inférés depuis ces deux matrices sont majoritairement congruents, en accord avec les analyses de Philippe *et al.* (Philippe et al., 2004).

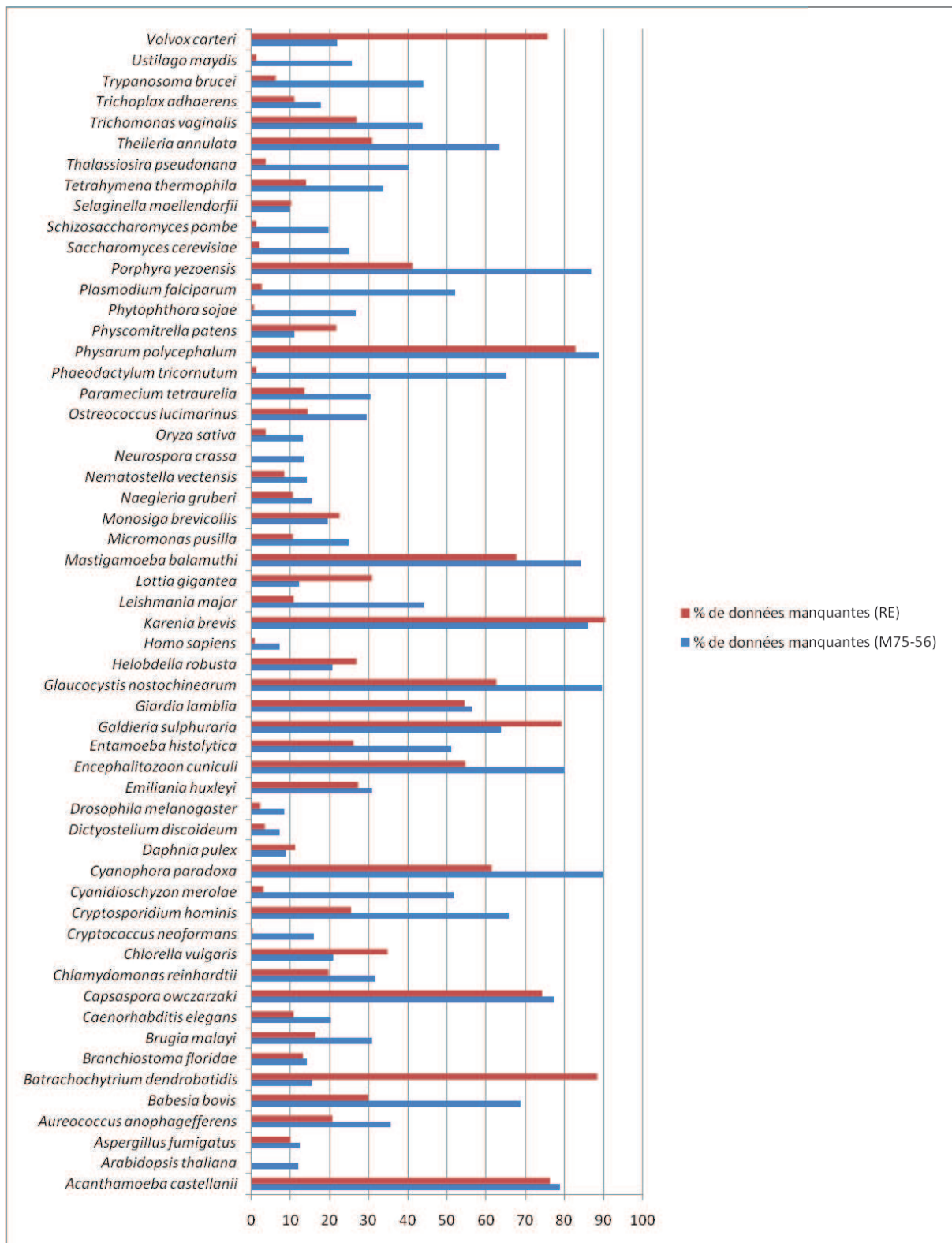


Figure 88. Pourcentage de données manquantes pour chaque espèce, dans les matrices M75-56 et RE. Les valeurs sont données en pourcentage du nombre total de positions de l'alignement. Il est à noter qu'une erreur a été détectée au niveau de la matrice RE, dans laquelle seule une partie des homologues de *Batrachochytrium dendrobatidis* (Fungi) a été incluse. Ce problème est en cours de résolution.

3.3 Concaténations de 230 protéines (M75-56 + RE)

Les analyses des supermatrices RE et M75-56 montrent qu'elles contiennent un signal phylogénétique majoritairement congruent. Néanmoins, les arbres inférés à partir de ces deux matrices montrent une incapacité à retrouver les relations de parenté entre les grands groupes. Dans le but d'augmenter encore le signal phylogénétique, j'ai donc entrepris de concaténer ces deux supermatrices afin d'obtenir un très grand alignement (56 espèces, 230 protéines, 49013 positions) (supermatrice "REM"). Cette matrice constitue à ce jour et à ma connaissance, la plus grande concaténation utilisée pour l'étude de la phylogénie des eucaryotes. Il n'est pas surprenant d'observer que les données contenues dans cette supermatrice évoluent légèrement plus rapidement que celles contenues dans la matrice RE (Figure 89), mais moins rapidement que celles de la matrice M75-65.

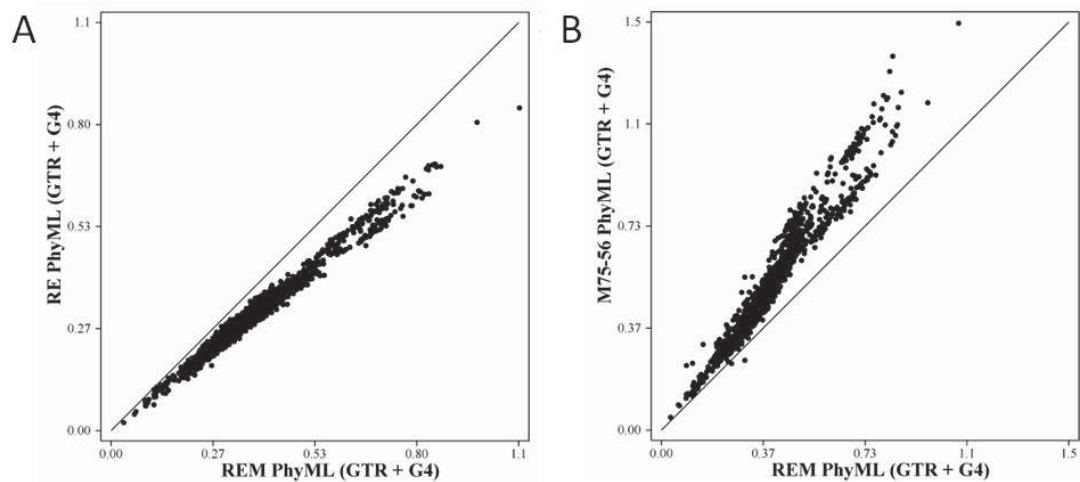


Figure 89. Comparaison des vitesses d'évolution des supermatrices REM/RE et REM/M75-56. Les coordonnées de chaque point représentent la distance évolutive séparant une paire d'espèces dans l'un et l'autre des arbres reconstruits à partir de ces deux matrices. Toutes les paires d'espèces sont représentées.

La topologie obtenue à partir de la matrice REM (Figure 90) est quasiment identique à celle obtenue à partir de la SM RE (Figure 85), à la différence (i) qu'on retrouve la monophylie des Unikonta dans l'arbre REM (VB = 87%), qu'on ne retrouve pas dans l'arbre RE, et (ii) que le placement d'*Emiliana huxleyi* (Haptophyta) varie.

Mis à part ces deux points, la différence essentielle concerne les valeurs de bootstrap qui se révèlent bien supérieures au niveau des nœuds profonds, ce qui confirme le gain de signal phylogénétique : monophylie des Alveolata (VB = 100% au lieu de 96%), monophylie des Heterokonta (VB = 100% dans les deux cas), regroupement Alveolata + Heterokonta (VB = 70% au lieu de 56%), monophylie des Excavata (VB= 81% au lieu de 55%), monophylie des Metazoa (VB = 100% dans les deux cas), monophylie des Fungi (y compris la microsporidie) (VB = 91% au lieu de 56%), et monophylie des Amoebozoa (VB = 96% au lieu de 75%). La seule exception concerne les Plantae, dont la monophylie reste, de manière surprenante, très peu soutenue (VB = 45% et 50%).

Enfin, l'ensemble de ces arbres ont été reconstruits en ML, et des améliorations sont à attendre *via* l'obtention des arbres inférés par Phylobayes sous le modèle CAT (qui permet de prendre en compte l'hétérogénéité substitutionnelle des sites de l'alignement). En effet, il a été montré que cette méthode couplée à ce modèle est particulièrement performante pour l'étude de la phylogénie des eucaryotes sur la base de grandes concaténations (Lartillot et al., 2007; Lartillot and Philippe, 2004; Rodriguez-Ezpeleta et al., 2007b).

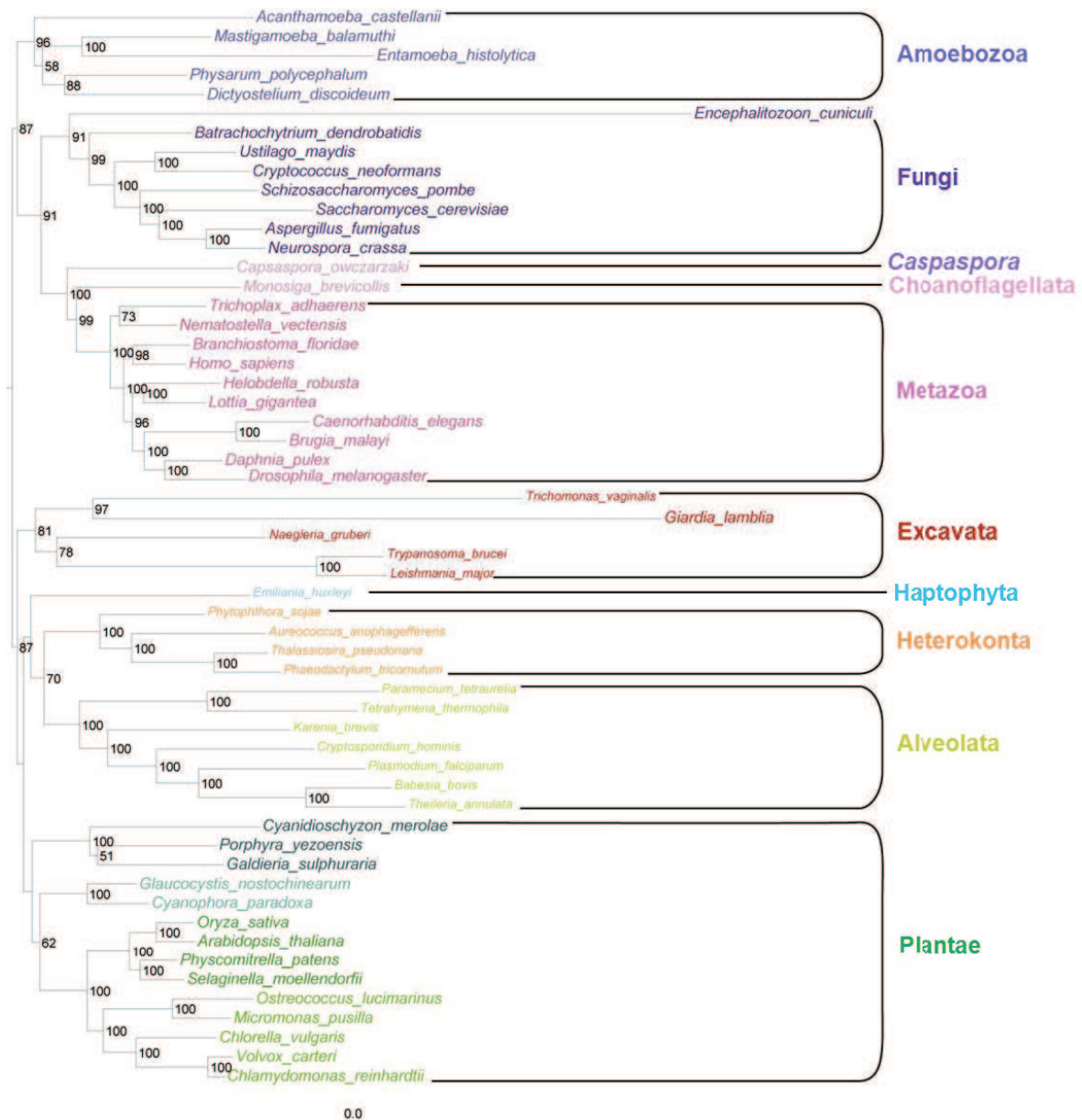


Figure 90. Arbre en Maximum de Vraisemblance calculé sur la matrice REM (56 espèces, 49013 positions, alignement recodé, PhyML). Les valeurs aux nœuds correspondent aux valeurs de bootstrap. Seules les valeurs supérieures à 50% sont indiquées. La barre d'échelle indique le nombre moyen de substitutions par site.

3.4 La supermatrice APC

Les concaténations basées sur les protéines du midbody ont montré que celles-ci contiennent un signal phylogénétique intéressant pour l'étude des relations de parenté au sein des eucaryotes. La question peut se poser de savoir ces marqueurs constituent une exception au sein des protéines opérationnelles ou si d'autres parmi celles-ci possèdent le même potentiel.

Pour tenter de répondre à cette question, j'ai concaténé les alignements des composants, co-activateurs et cibles de l'APC/C inférés comme étant présents chez LECA (62 espèces, 24 protéines, 3115 acides aminés). Nous avons dû retirer les séquences partielles et/ou à évolution très rapide de *Enterocytozoon bieneusi* (microsporidie), *Plasmodium hominis* (Apicomplexa) et *Blastocystis hominis* (Heterokonta) et *Trichomonas vaginalis* (Excavata). En effet, les trois premières contiennent >60% de données manquantes et *Trichomonas* présente de nombreuses séquences dupliquées et très divergentes. Lorsque ces espèces sont incluses, elles ont tendance à se regrouper ensemble, en dépit de leur position phylogénétique réelle, mettant en évidence un artefact d'attraction des longues branches (ALB) (Figure 91). Ces ALB sont également visibles au sein même des grands groupes, comme illustré par la position de *Caenorhabditis elegans* et *Brugia malayi* qui émergent très profondément au sein des animaux (Figure 92).

Lorsque ces espèces qui posent problème ont été retirées de l'analyse, l'arbre inféré par Phylobayes (Figure 92) permet de retrouver la monophylie de chacun des phylums eucaryotes. Ceux-ci sont le plus souvent statistiquement bien soutenus : Excavata (PP = 0,98; VB = 97%), Heterokonta (PP = 0,99; VB = 100%), Metazoa (PP = 1,0; VB = 100%), Fungi (PP = 0,87; VB = 98%), Choanoflagellata (PP = 0,99; VB = 100%), Viridiplantae (PP = 0,99; VB = 99%) et Alveolata (PP = 0,95 et VB = 78%). La phylogénie au sein de ces groupes est également en accord avec les analyses phylogénétiques précédemment publiées (Minge et al., 2009 880) (Burki et al., 2008 885), et celles inférées à partir des supermatrices REM, RE et M75-56.

En conclusion, malgré sa taille assez modeste comparée à celle des autres concaténations présentées précédemment, cette matrice contient un signal intéressant pour étudier les relations de parenté au sein des eucaryotes. De plus, cette analyse confirme que les protéines opérationnelles sont une source de données prometteuse mais trop souvent négligée pour identifier des marqueurs permettant d'étudier la phylogénie ancienne des eucaryotes.

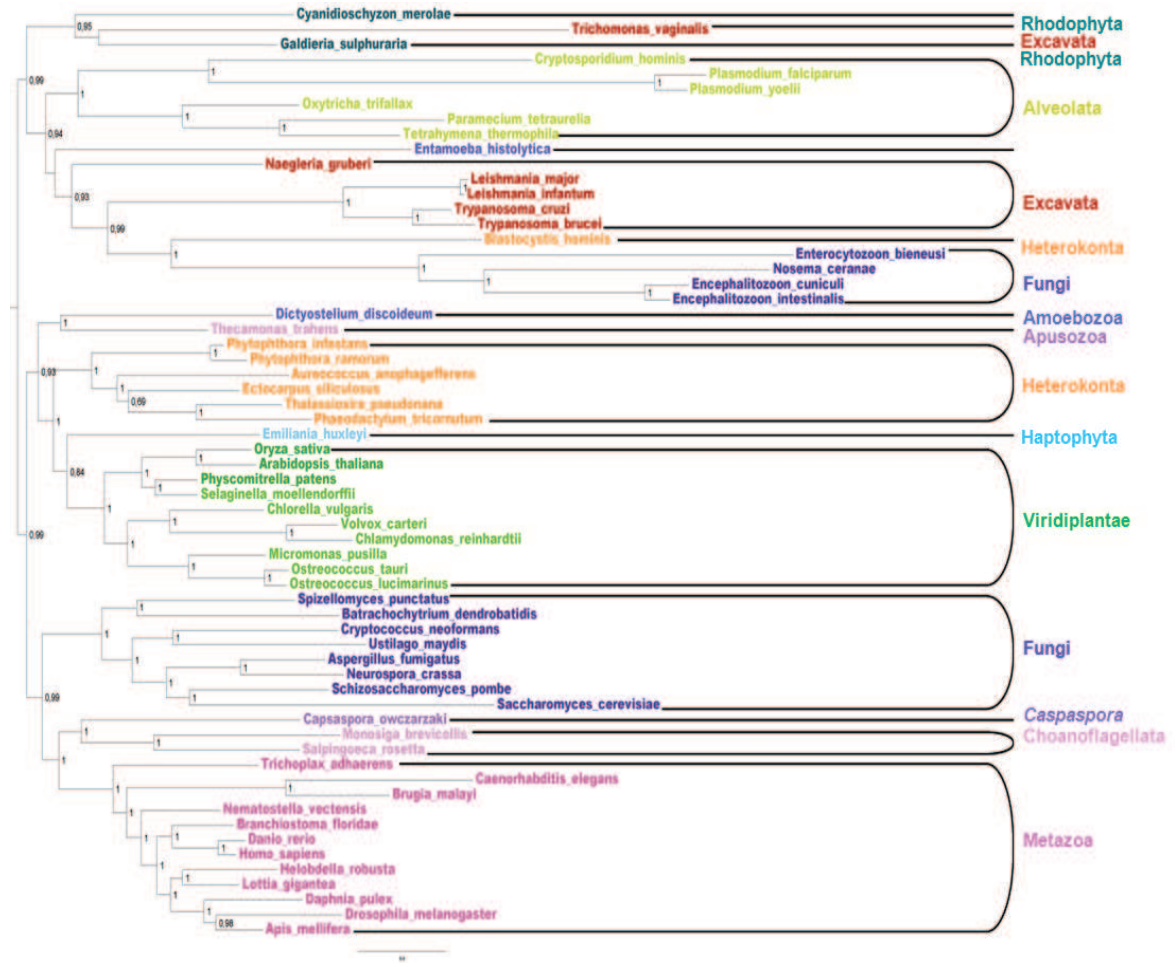


Figure 91. Arbre bayésien comprenant l'ensemble des 62 génomes étudiés (3115 positions, Phylobayes). Les valeurs aux nœuds correspondent aux probabilités postérieures. Seules les valeurs supérieures à 0.50 sont indiquées. La barre d'échelle indique le nombre moyen de substitutions par site.



Figure 92. Arbre bayésien basée sur la supermatrice APC (3115 positions d'acides aminés, 56 espèces, Phylobayes). Les valeurs aux nœuds correspondent aux probabilités postérieures. Les points indiquent les valeurs de bootstrap de 100% (noir) ou > 95% (gris). La barre d'échelle indique le nombre moyen de substitutions par site.

4 Points de débats

Je vais maintenant discuter plus en détail les résultats obtenus à la lueur des données de la littérature. Le soutien obtenu pour divers regroupements dans chacun des arbres présentés dans ce chapitre est récapitulé dans le tableau ci-dessous :

	Metazoa	Fungi	Amoebozoa	Plantae	Alveolata	Heterokonta
APC (Phylobayes, PP)	0.95	0.87	NA (un seul représentant)	-	0.95	1.0
M75-67 (recodé, PhyML, SH-like)	1.0	0.95	1.0	-	1.0	1.0
M35-67 (recodé, PhyML, VB)	100	52	77	9	89	100
RE (recodé, PhyML, VB)	100	56	65	50	98	100
M75-56 (recodé, PhyML, VB)	100	80	89	-	84	100
REM (recodé, PhyML, VB)	100	91	96	45	100	100

	Excavata	Unikonta	Opisthokonta	Ecdysozoa	Capsaspora à la base des Holozoa	Heterokonta + Alveolata
APC (Phylobayes, PP)	0.98	0.81	0.91	0.76	-	-
M75-67 (recodé, PhyML, SH-like)	0.54	1.0	1.0	1.0	0.88	1.0
M35-67 (recodé, PhyML, VB)	9	36	55	99	83	45
RE (recodé, PhyML, VB)	55	-	56	99	100	57
M75-56 (recodé, PhyML, VB)	19	72	79	100	82	52
REM (recodé, PhyML, VB)	81	87	91	100	99	70

Tableau 16. Soutiens obtenus pour différents regroupements d'eucaryotes, en fonction de la matrice utilisée. Chaque ligne correspond à un arbre présenté dans ce chapitre. Pour chaque arbre, sont mentionnés : le nom de la matrice, si les données sont recodées ou non, le logiciel utilisé pour la reconstruction ainsi que le type de support statistique.

4.1 La monophylie des Plantes

Comme je l'ai discuté dans l'introduction, la monophylie des Plantae est rarement remise en question, mais elle peut s'avérer difficile à retrouver selon le jeu de données analysé (Hampl et al., 2009; Minge et al., 2009; Nozaki et al., 2007; Nozaki et al., 2003; Parfrey et al., 2010), ou bien être faiblement soutenue (Burki et al., 2007). C'est également ce que j'ai pu observer puisque, dans les arbres construits à partir des supermatrices APC (Figure 92), M75-67 (Figure 81), et M75-56 (Figure 86), la monophylie de ce groupe n'est pas retrouvée, alors qu'elle l'est dans les arbres M35-67 (Figure 82), RE (Figure 85) et REM (Figure 90). Néanmoins, même dans ces derniers cas, le soutien statistique est très faible (VB = 9%, 50% et 45%, respectivement). La difficulté à retrouver cette monophylie réside probablement en partie dans le fait que les représentants des Rhodophyta présents dans nos matrices sont peu nombreux, et sont des espèces qui évoluent globalement assez rapidement. De plus, elles sont très souvent absentes des jeux de données issus de l'analyse du midbody.

En parallèle, les relations de parenté entre les trois lignées de plantes restent très débattues (Deschamps and Moreira, 2009; Reyes-Prieto and Bhattacharya, 2007). Dans l'ensemble des arbres que j'ai reconstruits, le soutien obtenu pour l'émergence basale de l'une ou l'autre des lignées de plantes est trop faible pour répondre à cette question. Deschamps *et al.* ont montré en 2009 que la lignée de plante qui émerge en position basale est dépendante de l'échantillonnage des séquences utilisées comme groupe extérieur. Ainsi, les Rhodophytes seraient retrouvées comme étant basales lorsque des lignées portant des plastes secondaires rouges sont intégrées au groupe extérieur, vraisemblablement à cause des transferts de gènes ayant eu lieu de ces plastes vers le noyau de leurs hôtes.

Dans cette optique, j'ai reconstruit des arbres (méthode bayésienne *via* Phylobayes) n'incluant que les Plantae et des eucaryotes ne possédant pas de plaste (ou n'ayant pas d'ancêtre ayant possédé un plaste). J'ai fait deux sélections d'espèces pour le groupe extérieur : l'une ne contenant que des organismes unikontes présentant des branches plutôt courtes (Figure 93), et l'autre en y ajoutant les Excavata, et des organismes présentant de très longues branches (ex : *T.vaginalis*, *G.lambliia*, *E.cuniculi*) (Figure 94). Dans les deux cas, les arbres montrent une émergence précoce des Glaucophyta, en accord avec les résultats de Deschamps 2009 sur les gènes nucléaires (VB = 97% et PP = 0.96), Reyes-Prieto 2007 (VB = 100% et PP = 1.0), Rodríguez-Ezpeleta 2005 (VB = 38% et 78%), mais en désaccord avec ceux de Burki 2007 (Rhodophyta à la base, VB = 70% et PP = 1.0), Patron 2007 (Rhodophyta à la base, VB = 68%) et Deschamps 2009 (Viridiplantae à la base, VB = 100% et PP = 1.0), sur la base des gènes chloroplastiques transférés au noyau.

Le soutien que nous obtenons pour l'émergence basale des Glaucophyta est quasiment maximal (PP = 0.99 dans les deux cas). Néanmoins, la longueur de la branche séparant les Glaucophyta des autres Plantae est très petite, ce qui suggère que la quantité de signal phylogénétique disponible pour résoudre cette région de l'arbre est très faible, et qu'il faut donc prendre ces résultats avec précaution.

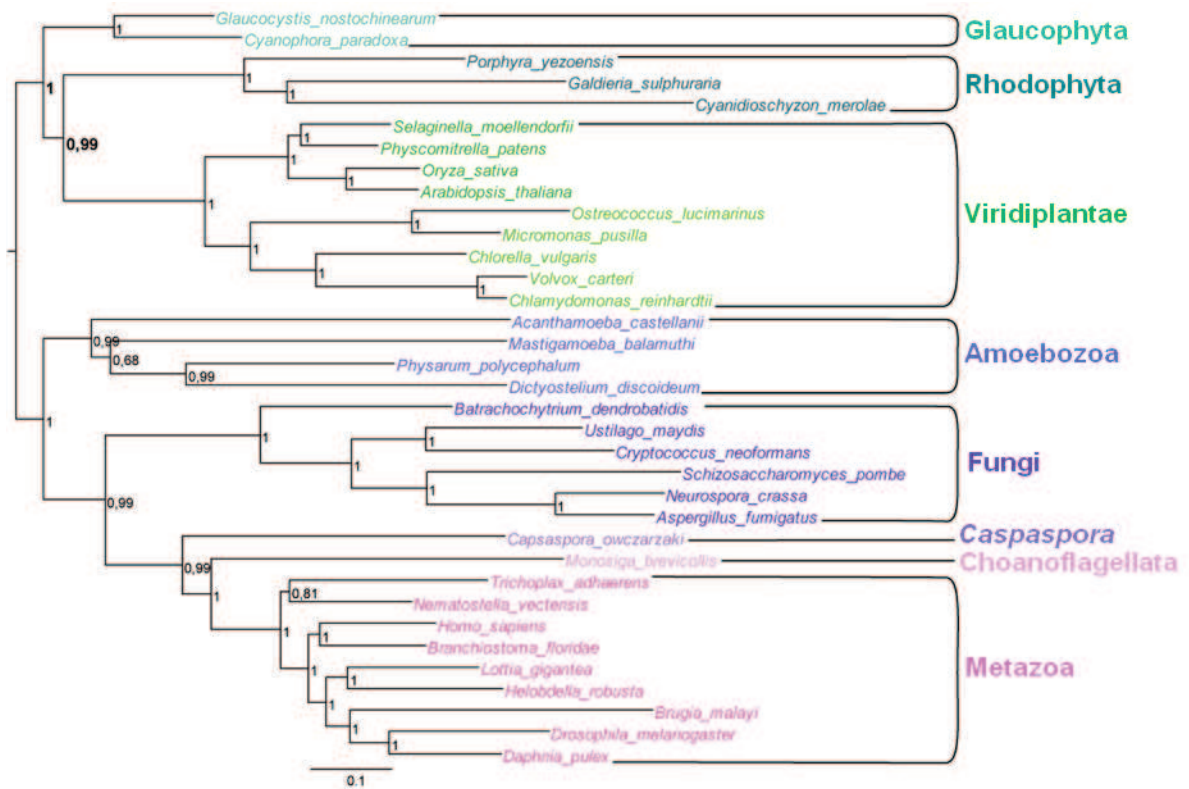


Figure 93. Phylogénie des Plantae. Arbre bayésien (35 espèces, 49013 positions, Phylobayes). Les données ont été extraites de la supermatrice REM. Le groupe extérieur est représenté par une sélection d'espèces unikontes ne présentant pas de branches longues.

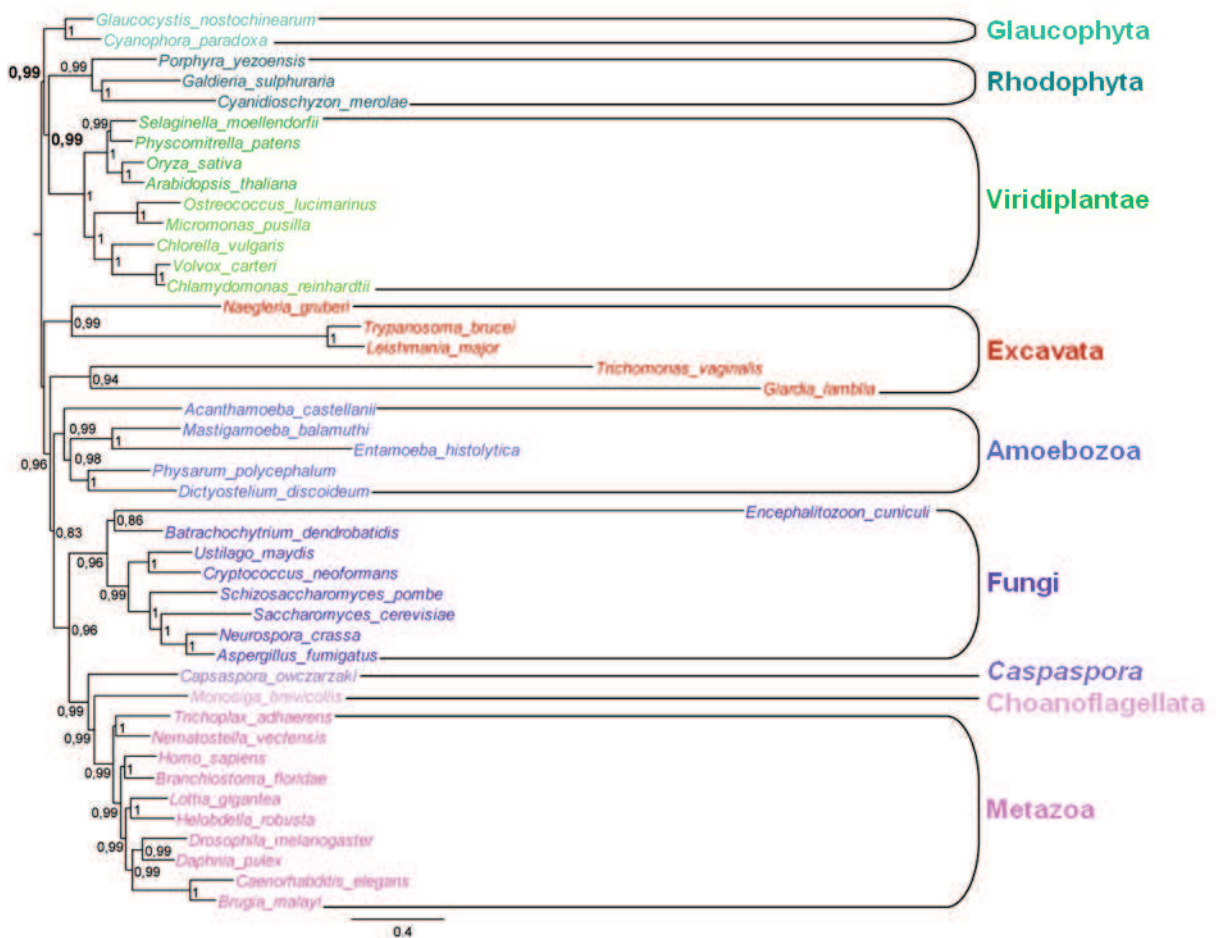


Figure 94. Phylogénie des Plantae. Arbre bayésien (44 espèces, 49013 positions, Phylobayes). Le groupe extérieur est représenté par une sélection d'eucaryotes ne possédant pas de plastes, dont certains présentent de très longues branches (*T.vaginalis*, *G.lamblii*, *E.cuniculi*). L'ordre d'émergence des trois lignées de plantes n'est pas modifié par la présence de ces longues branches.

4.2 Placement des microsporidies au sein des fungi

Le résultat le plus remarquable est probablement le support important retrouvé pour la monophylie des champignons, y compris les microsporidies (PP = 0.87 et VB = 95% dans l'arbre APC, et VB = 80%, 99% et 100% dans les arbres M75-56, RE et REM respectivement). En effet, aucune des études phylogénomiques massives à l'échelle des eucaryotes citées dans le Tableau 2 n'a intégré la microsporidie, et les résultats présentés ici sont donc parmi les rares qui permettent de replacer correctement ce parasite au sein des Fungi (Thomarat et al., 2004). Cependant, la position des microsporidies à l'intérieur des Fungi n'est pas claire, et il serait intéressant de réaliser des analyses plus ciblées sur les Unikonta en augmentant l'échantillonnage taxonomique, et notamment en incluant les séquences provenant d'autres microsporidies.

4.3 La Monophylie des Excavata

La monophylie des Excavata est toujours très débattue, et les phylogénies moléculaires ont rarement permis de retrouver ce groupe, ou bien avec un support modéré, ou encore avec un échantillonnage taxonomique limité aux *Trypanosoma*/*Leishmania*, qui sont des organismes très proches (Tableau 17). Un deuxième résultat très remarquable est donc que les différentes matrices que j'ai construites ont permis de retrouver la monophylie de ce groupe quasiment systématiquement, avec des supports parfois très élevés : PP = 0.98 et VB = 95% pour l'arbre APC, et VB = 19%, 55% et 81% pour les arbres M75-56, RE et REM. A ma connaissance, un support aussi important que ce dernier (81%) n'a jamais été retrouvé avec l'échantillonnage taxonomique que j'ai utilisé.

Auteurs	Monophylie des Excavata
(Baptiste, Brinkmann et al. 2002)	Non retrouvée
(Philippe, Snell et al. 2004)	VB = 100% Seules espèces: <i>Trypanosoma</i> et <i>Leishmania</i>
(Rodriguez-Ezpeleta, Brinkmann et al. 2005)	Excavata non représentés
(Burki and Pawlowski 2006)	VB = 93% <i>Naegleria</i> absent
(Patron, Inagaki et al. 2007)	Non représentés
(Rodriguez-Ezpeleta, Brinkmann et al. 2007)	VB = 100% <i>Giardia</i> et <i>Trichomonas</i> absents
(Burki, Shalchian-Tabrizi et al. 2007)	VB = 61% et PP = 1.0 <i>Giardia</i> et <i>Trichomonas</i> absents
(Burki, Shalchian-Tabrizi et al. 2008)	VB = 100% et PP = 1.0 <i>Giardia</i> et <i>Trichomonas</i> absents
(Hampl, Hug et al. 2009)	VB = 54%
(Minge, Silberman et al. 2009)	VB = 63 % Aucune espèce en commun avec nos arbres
(Baurain, Brinkmann et al. 2010)	Non représentés

Tableau 17. Support pour la monophylie des Excavata dans les analyses phylogénomiques récentes. Les valeurs en vert correspondent aux analyses qui contiennent au moins les mêmes espèces d'Excavata que les nôtres.

4.4 Le Placement d'*Emiliana*

Le placement d'*Emiliana huxleyi* (et des Haptophyta en général) représente un défi de taille. Sa position est très dépendante de la supermatrice utilisée, et le soutien obtenu est généralement faible. Je la retrouve alternativement groupe-frère des :

- Viridiplantae (arbre APC : PP = 0.78; arbre M75-67 : SH-like = 0.3)
- Plantae (arbre RE : VB = 28%)
- Heterokonta + Alveolata (arbre M75-56 : VB = 52% ; arbre M35-67 : VB = 45%; arbre REM : VB = 28%)

Dans la bibliographie, cette inconsistance est également régulièrement reportée (Tableau 18).

Auteurs	Groupe frère retrouvé pour les Haptophyta [+Cryptophyta éventuellement]
(Bapteste, Brinkmann et al. 2002)	Haptophyta absents
(Philippe, Snell et al. 2004)	Haptophyta absents
(Rodriguez-Ezpeleta, Brinkmann et al. 2005)	Haptophyta absents
(Burki and Pawlowski 2006)	Haptophyta absents
(Patron, Inagaki et al. 2007)	Plantae (VB = 94%)
(Rodriguez-Ezpeleta, Brinkmann et al. 2007)	Haptophyta absents
(Burki, Shalchian-Tabrizi et al. 2007)	Alveolata+Heterokonta+Rhizaria (VB < 50% et PP < 0.8)
(Burki, Shalchian-Tabrizi et al. 2008)	Plantae (VB = 80% et PP = 1.0)
(Hampl, Hug et al. 2009)	Rhodophyta (VB = 59%)
(Minge, Silberman et al. 2009)	Glaucophyta (VB = 57% et PP = 0.89)
(Baurain, Brinkmann et al. 2010)	Alveolata+Heterokonta (VB = 40%)

Tableau 18. Position des Haptophyta dans les analyses phylogénomiques massives récentes. Les analyses les plus anciennes (antérieures à 2007) ne contenait pas de représentant des Haptophyta.

En dépit du très grand nombre de positions analysées, mes analyses ne permettent donc pas de proposer un placement fiable pour cette Haptophyta.

Il paraît indispensable de disposer de plus de données, et notamment d'augmenter l'échantillonnage taxonomique au sein de ce groupe afin d'espérer pouvoir inférer correctement sa position au sein des eucaryotes. Il serait également intéressant de pouvoir ajouter des séquences de Cryptophyta qui sont considérés comme étant le groupe-frère des Haptophyta (même si ce point fait encore débat).

4.5 Relations au sein des Holozoa

Les relations entre Metazoa, Choanoflagellata et *Capsaspora owczarzaki* (tous trois formant les Holozoa avec quelques autres lignées) sont encore assez floues. Le regroupement des choanoflagellés et de *Capsaspora owczarzaki* est soutenu par les analyses de Ruiz-Trillo 2006 et Minge 2009 (Minge et al., 2009; Ruiz-Trillo et al., 2006), mais pas par celles de Shalchian-Tabrizi et Ruiz-Trillo en 2008 (Ruiz-Trillo et al., 2008; Shalchian-Tabrizi et al., 2008) qui, elles, soutiennent l'émergence de *Capsaspora* à la base du groupe composé des Choanoflagellata et des Metazoa.

Dans l'arbre APC, la première hypothèse est modérément soutenue (PP = 0.88 pour le regroupement Choanoflagellés et *Capsaspora*). En revanche, les autres arbres soutiennent plutôt la seconde hypothèse avec des supports modérés à forts : VB = 82%, 99% et 100% pour les arbres M75-56, RE et REM, respectivement. Cette seconde hypothèse est également fortement soutenue dans les arbres "réduits" qui sont racinés par les Plantae (Figure 93 et Figure 94) : PP = 1.00 et 0.99.

4.6 Ecdysozoa ou Coelomata ?

Les animaux possédant une cavité corporelle entièrement doublée par des dérivés mésodermiques (c'est-à-dire le coelome), tels que les chordés et les arthropodes, ont été classiquement regroupés dans le clade Coelomata, tandis que les animaux sans cavité corporelle (Acoelomata) ou avec une cavité partiellement doublée par des dérivés mésodermiques (Pseudocoelomata), tels que les nématodes, ont été considérés comme des lignées distinctes (Raff, 1996). Ce point de vue a été remis en cause par les analyses de l'ARNr SSU qui regroupent des animaux qui muent (par exemple, les nématodes et les arthropodes) dans le clade des Ecdysozoa, modifiant ainsi la vision de l'époque sur l'évolution animale (Aguinaldo et al. 1997).

Depuis la proposition de cette hypothèse, la plupart des analyses phylogénomiques massives (Bourlat et al., 2008; Dunn et al., 2008; Lartillot et al., 2007; Philippe et al., 2009; Philippe et al., 2005; Schierwater et al., 2009), ainsi que des analyses d'insertions/délétions (Belinky et al., 2010), des études de la structure secondaire des gènes mitochondriaux (Schierwater et al., 2009), des analyses d'indels (Irimia et al. 2007) ainsi que des études morphologiques (Glennier et al., 2004; Schierwater et al., 2009) ont apporté des arguments en faveur de l'existence des Ecdysozoa.

A l'inverse, quelques auteurs prétendent que la topologie Coelomata est la bonne, sur la base d'analyses phylogénomiques (Blair et al., 2002), d'analyses d'indels (Wolf et al., 2004) ou par des approches de parcimonie pour interpréter des changements d'acides aminés (Rogozin et al., 2007a, b).

L'ensemble des arbres reconstruits et présentés dans ce manuscrit est en faveur de l'hypothèse Ecdysozoa avec de très forts supports statistiques : support SH-like = 1.0 pour l'arbre M75-56, VB = 99% et 100% pour les arbres RE et REM, et PP = 1.0 et 0.99 dans les arbres réduits racinés sur les Plantae.

4.7 Monophylie des Amoebozoa

En accord avec la proposition initiale de Baptiste *et al.* de 2002 et avec les analyses phylogénomiques récentes qui tendent à soutenir la monophylie des Amoebozoa, l'ensemble de nos arbres confirment cette hypothèse avec un soutien modéré à fort : VB = 89%, 65% et 96% pour les arbres M75-56 (Figure 86), RE (Figure 85) et REM (Figure 90).

4.8 Monophylie des Unikonta

Il est intéressant de remarquer que les Unikonta sont relativement bien soutenus, à l'exception de l'arbre RE où on ne les retrouve pas : PP = 0.81 pour l'arbre APC, SH-like = 1.0 pour l'arbre M75-56, VB = 36%, 72% et 87% pour les arbres M35-67, M75-67 et REM.

4.9 L'hypothèse chromalvéolés

Il est à noter que l'hypothèse chromalvéolés n'a pas pu être testée sur la base des concaténations présentées ici puisque les données EST disponibles pour les Rhizaria sont trop restreintes pour inclure des représentants de ce groupe dans les jeux de données issus de l'analyse du midbody et de l'APC/C.

5 Conclusion

L'objectif de ce volet de ma thèse était l'identification de nouveaux marqueurs pouvant être utilisés pour étudier l'histoire évolutive ancienne des eucaryotes. J'ai choisi de les rechercher parmi les protéines opérationnelles. Pris dans leur ensemble, et bien que ce travail soit encore en cours, les analyses présentées dans ce chapitre ont montré que, en accord avec ce qui est couramment admis :

- les protéines opérationnelles évoluent plus rapidement que les protéines informationnelles (du moins celles étudiées ici) ;
- Elles sont moins bien conservées (*i.e.* elles sont plus souvent perdues), induisant l'assemblage de supermatrices contenant plus de données manquantes ;
- Elles appartiennent fréquemment à des familles importantes et complexes qui ont connu de nombreux événements de duplications de gènes et de pertes au cours de leur histoire évolutive, rendant fastidieuses les analyses préliminaires de délimitation des familles d'orthologues ;
- Elles sont souvent de grande taille (*i.e.* plusieurs centaines d'acides aminés), ce qui rend difficile l'utilisation des banques d'EST, mais permet l'assemblage de jeux de données de grande taille, même à partir d'un nombre relativement modeste de protéines.

Malgré tous ces inconvénients, mes analyses ont montré que certaines protéines de type opérationnel contiennent une information phylogénétique pouvant être exploitée pour inférer une partie de l'histoire évolutive ancienne des eucaryotes. J'ai ainsi pu montrer que l'on obtient un signal phylogénétique congruent avec celui obtenu à partir des protéines informationnelles et que l'inclusion de ces deux types de données dans les analyses phylogénomiques massives permet d'augmenter significativement le soutien obtenu pour certaines régions de l'arbre, y compris pour certaines considérées comme difficiles à retrouver (ex : Excavata, microsporidies au sein des Fungi, etc.)

Bien entendu, comme précisé en introduction, ces calculs sont encore préliminaires pour certains, et il sera nécessaire d'affiner les conclusions lorsque l'ensemble des reconstructions seront disponibles en maximum de vraisemblance (avec bootstrap) et en bayésien (*via* l'utilisation prometteuse de Phylobayes et du modèle CAT).

Le jeu de données REM assemblé lors de ce travail représente la plus grande concaténation à ce jour, et constitue ainsi une référence qui pourra être enrichie par l'inclusion d'autres espèces au fur et à mesure qu'elles deviendront disponibles.

Il serait également intéressant de poursuivre l'exploration des protéines opérationnelles car, compte-tenu de leur abondance dans les génomes eucaryotes, ces protéines constituent une source de données prometteuse qui devra, à l'avenir, être exploitée.

AXE 3:

PHYLOGENIE DES ARCHAEA ET RELATIONS AVEC LES EUCARYA

1 Introduction

Comme abordé en introduction (cf. p.16), sur la base de l'existence de nombreux caractères et mécanismes moléculaires communs aux archées et aux eucaryotes, il est couramment admis qu'Archaea et Eucarya partagent un ancêtre commun exclusif. Néanmoins, la nature de ce lien fait encore débat : les Eucarya et les Archaea représentent-ils deux lignées sœurs distinctes qui descendent d'un ancêtre commun exclusif (hypothèse 3D) ? Ou bien les Eucarya descendent-ils d'une lignée d'archée "véritable" (*i.e.* possédant toutes les caractéristiques des archées modernes) (hypothèse 2D) ? Dans le cadre de l'hypothèse 2D, plusieurs scénarios ont été envisagés, et plusieurs lignées d'archées ont été proposées comme étant spécifiquement apparentées aux eucaryotes.

Plusieurs analyses phylogénomiques récentes ont fourni une phylogénie des Archaea (Figure 95) indiquant une division claire entre les deux phylums majeurs que sont les Crenarchaeota et les Euryarchaeota (et qui sont reconnus depuis les années 1990), mais suggèrent aussi l'existence d'autres phylums : les Korarchaeota (anaérobies hyperthermophiles) (Barns et al., 1996; Elkins et al., 2008) et les Thaumarchaeota (mésophiles et thermophiles retrouvées dans des milieux très divers et jouant un rôle clé dans la nitrification et donc le cycle de l'azote) (Brochier-Armanet et al., 2008; Schleper and Nicol, 2010). D'autres phylums ont également été proposés, tels que les Nanoarchaeota (Huber et al., 2002), pour lesquelles il a été montré qu'elles font en fait partie des Euryarchaeota (Brochier et al., 2005). Plus récemment, il a également été proposé que '*Candidatus* Caldiarchaeum subterraneum' (un génome composite obtenu à partir d'une banque métagénomique provenant d'eaux géothermales d'une mine d'or souterraine) soit en fait le représentant non cultivé d'un nouveau phylum, que les auteurs ont proposé d'appeler 'Aigarchaeota' (Nunoura et al., 2010).

Dans ce contexte, j'ai eu l'opportunité de collaborer à l'analyse de données génomiques (fosmides) de Thaumarchaeota issues d'une banque métagénomique construite à partir d'une source chaude terrestre du Kamtchatka. Les données issues de cette analyse sont importantes car elles couvrent la diversité taxonomique de deux grands groupes d'archées non cultivées pour lesquels aucune donnée génomique n'était disponible. L'objectif de ce projet était donc de replacer ces lignées dans l'arbre des Archaea, mais également d'observer l'influence de ces données sur les relations inférées entre Archaea et Eucarya.

Ce travail a été réalisé en collaboration avec Laila Reigstad, du Centre de Géobiologie de l'université de Bergen (Norvège), ainsi qu'avec Christa Schleper et Anja Spang du Département de Génétique et Ecologie de l'Université de Vienne (Autriche). Il a donné lieu à une publication en cours de rédaction.

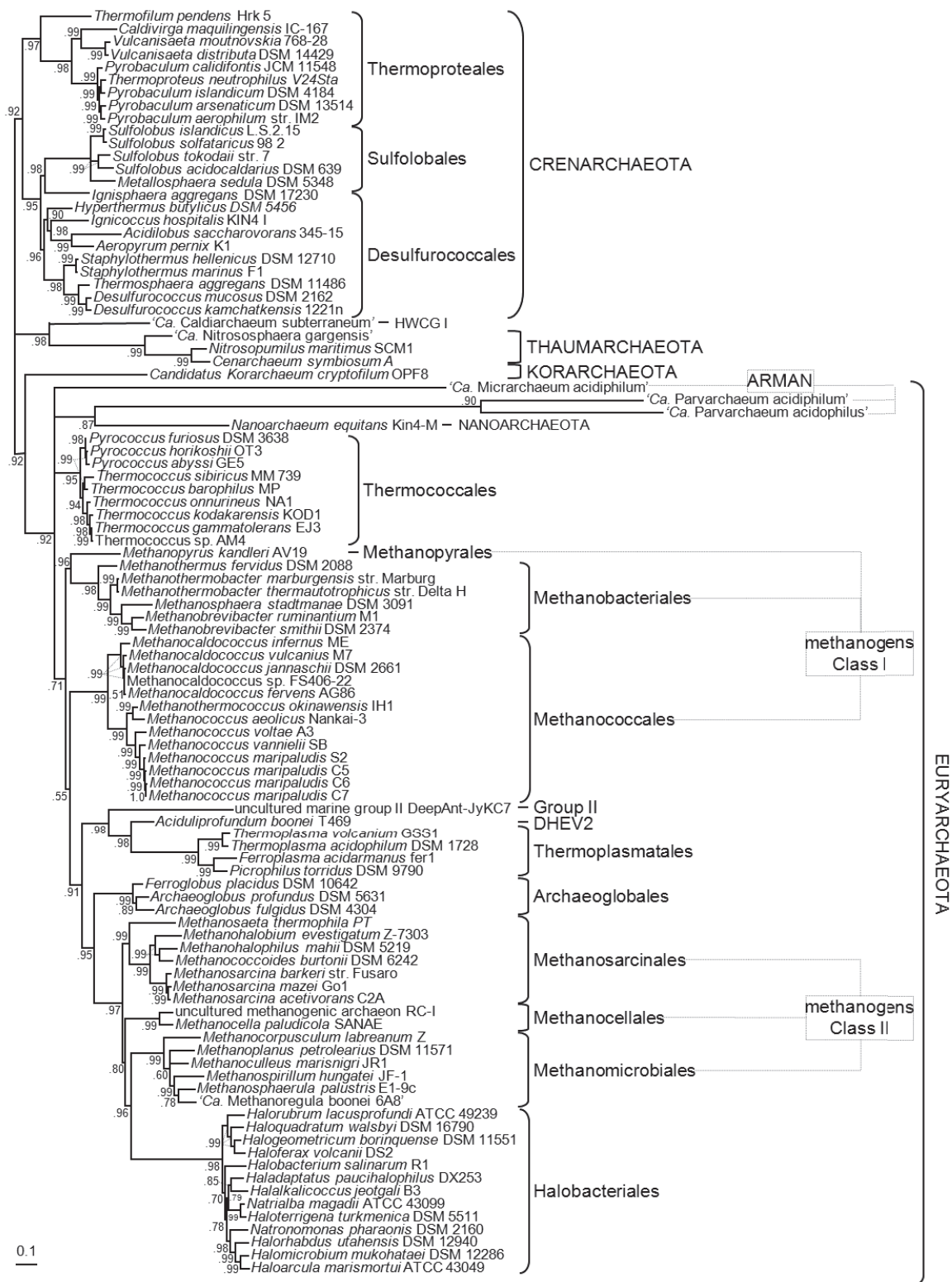


Figure 95. Phylogénie de référence des Archaea. Arbre bayésien reconstruit avec Phylobayes sur la base d'une concaténation de 57 protéines ribosomales (recodage en Dayhoff 6, modèle CAT + dirichlet (57 protéines présentes dans au moins 89/99 génomes, soit 5838 positions) (Céline Brochier-Armanet, communication personnelle).

2 Description du site d'échantillonnage et diversité biologique

La péninsule du Kamtchatka, située à l'extrême Est de la Russie, représente l'arc volcanique le plus actif du monde en termes de production de magma et d'éruptions explosives. La plupart des études microbiennes provenant des sources chaudes du Kamtchatka s'est concentrée sur l'isolement et la culture de souches thermophiles et acidophiles (Islam et al., 2008; Kozina et al., 2010; Kublanov et al., 2009; Wagner et al., 2008). Ces dernières années, la possibilité d'étudier les communautés microbiennes sans avoir besoin de les cultiver (*i.e.* grâce aux analyses métagénomiques) a permis d'améliorer la compréhension de la biodiversité et des fonctions biochimiques présentes dans ces sources (Mardanov et al., 2010; Perevalova et al., 2008; Reigstad et al., 2010; Reigstad et al., 2008).

Dans l'étude présentée ici, nous avons étudié une source chaude terrestre (nommée Kam37) des thermes de Uzon Caldera, qui sont situés dans la ceinture volcanique orientale du Kamtchatka. La température des filaments prélevés était de 85°C, tandis que le courant où ils ont été prélevés avait une température de 91°C et un pH de 5,5. A partir de ces prélèvements, deux banques de clones d'ARNr SSU ont été construites, une à partir d'amorces "universelles", l'autre à partir d'amorces archées-spécifiques. Afin de décrire les communautés microbiennes présentes dans l'échantillon Kam37, respectivement 108 et 149 clones issus de ces deux banques ont été séquencés. En parallèle, une importante banque de fosmides (27 366 fosmides uniques) contenant de grands inserts d'ADN (~40 kb) provenant de Kam37 a été construite.

2.1 Diversité microbienne

Parmi les clones d'ARNr SSU, la majorité est, sans surprise, représentée par les bactéries, et plus précisément par des séquences proches des Aquificales (Tableau 19), en accord avec les précédentes études sur la domination des Aquificae dans de nombreuses sources terrestres hyperthermophiles (Hugenholtz et al., 1998; Reigstad et al., 2010; Reysenbach et al., 2000; Spear et al., 2005). Les lignées bactériennes mineures sont apparentées à des Caldisericia, des delta-, bêta- et alpha-protéobactéries, des Firmicutes et des Thermodesulfobacteria. Au sein des Archaea, la majorité des séquences proviennent des Euryarchaeota (plus précisément des Thermoplasmatales) et des Crenarchaeota (Thermoproteales, Sulfolobales, Sulfohococcales et "Miscellaneous Crenarchaeota Group" (MCG)). Fait intéressant, il a aussi été détecté un certain nombre de clones appartenant tous à des groupes de Thaumarchaeota non cultivées, suggérant que les Thaumarchaeota sont abondantes, même si elles ne sont pas dominantes, et que la banque de fosmides pourrait contenir des fragments de génomes provenant de ces groupes de thaumarchées non cultivées.

Library	Number of sequences	Domain	Phylum/Class	Order/Family/Genus
"Universal" prokaryotic SSU rRNA	28	Bacteria	Aquificae	Sulfurihydrogenibium
	27		Aquificae	Hydrogenobaculum
	4		Aquificae	Hydrogenobacter
	4		Thermodesulfobacteria	Caldimicrobium
	3		Alphaproteobacteria	Rhizobiales/Methylocystaceae
	2		Betaproteobacteria	Burkholderiales/Thiomonas
	2		Deltaproteobacteria	Desulfurellales
	2		Clostridia	Thermoanaerobacteriales
	1		Caldiserica	Caldisericales
	21		Archaea	Euryarchaeota
	2	Thaumarchaeota		Hyperthermophilic Thaumarchaeal Cluster 1 (HTC1)
	3	Thaumarchaeota		Hyperthermophilic Thaumarchaeal Cluster 2 (HTC2)
	2	Crenarchaeota		Sulfobaceae
	2	Crenarchaeota		Desulfurococcaceae/Sulfophobococcus
2	Crenarchaeota	Thermofiliaceae/Thermofilum		
2	Crenarchaeota	Miscellaneous Crenarchaeota group		
1	Euryarchaeota	Candidatus Parvarchaeum		
Archaeal SSU rRNA	58	Archaea	Euryarchaeota	Thermoplasmatales
	42		Crenarchaeota	Thermoproteaceae/Pyrobaculum
	25		Crenarchaeota	Miscellaneous Crenarchaeota group
	4		Thaumarchaeota	Hyperthermophilic Thaumarchaeal Cluster 1 (HTC1)
	7		Thaumarchaeota	Hyperthermophilic Thaumarchaeal Cluster 2 (HTC2)
	2		Crenarchaeota	Thermofiliaceae/Thermofilum
	7		Crenarchaeota	Desulfurococcaceae/Thermofermentum
	3		Crenarchaeota	Desulfurococcaceae/Staphylothermus
	1		Crenarchaeota	Desulfurococcaceae/Caldisphaera

Tableau 19. : Affiliation taxonomique des 257 séquences d'ARNr SSU de l'échantillon Kam37 issues des deux banques de clones.

2.2 Isolation de grands contigs de Thaumarchaeota

L'utilisation des mêmes amorces a révélé 37 fosmidés contenant un gène ARNr SSU d'Archaea dont neuf pourraient être affiliés aux Thaumarchaeota. Afin d'élargir les connaissances sur les capacités fonctionnelles et la phylogénie des Thaumarchaeota hyperthermophiles, six (1M19, 2C9, 1C18, 34P11, 1C23 et 1N15) de ces neuf fosmidés de thaumarchées ont été sélectionnés pour un séquençage complet. Un des fosmidés (1C23) n'a pas pu être complètement assemblé et est représenté par deux contigs distincts.

L'annotation des six fosmidés a révélé une organisation compacte des gènes avec de courtes régions intergéniques. Les analyses phylogénétiques basées sur les séquences d'ARNr SSU ont montré que les fosmidés appartiennent à deux clusters : les fosmidés 1C23, 1C18, 34P11 et 1N15 se regroupent dans le cluster HTC1 (Hyperthermophilic Thaumarchaeota Cluster 1) et les fosmidés 2C9 et 1M19 appartiennent au cluster HTC2 (Hyperthermophilic Thaumarchaeota Cluster 2) (cf. résultats p.188). Les trois fosmidés 1C18, 34P11 et 1N15 (HTC1) montrent une forte synténie (Figure 96). En revanche, les deux fragments de 1C23 (HTC1), ainsi que 2C9 et 1M19 (HTC2) montrent une faible conservation des gènes. La mauvaise conservation de la synténie dans la région de l'opéron ARNr a déjà été observée sur des fosmidés de thaumarchées, suggérant que des événements de recombinaison ont fréquemment eu lieu dans ce phylum (Lopez-Garcia et al., 2004; Martin-Cuadrado et al., 2008). De plus, les deux fosmidés 1M19 et 2C9 montrent une teneur globale en G + C (62% et 67%, respectivement) nettement plus élevée que les autres fosmidés (44-51%) (Tableau 20). Ce résultat est surprenant car en dehors de *C. symbiosum* qui possède une teneur exceptionnellement élevée en G + C (57,7%), la plupart des séquences génomiques disponibles de Thaumarchaeota ont un contenu en G + C inférieur à 43%.

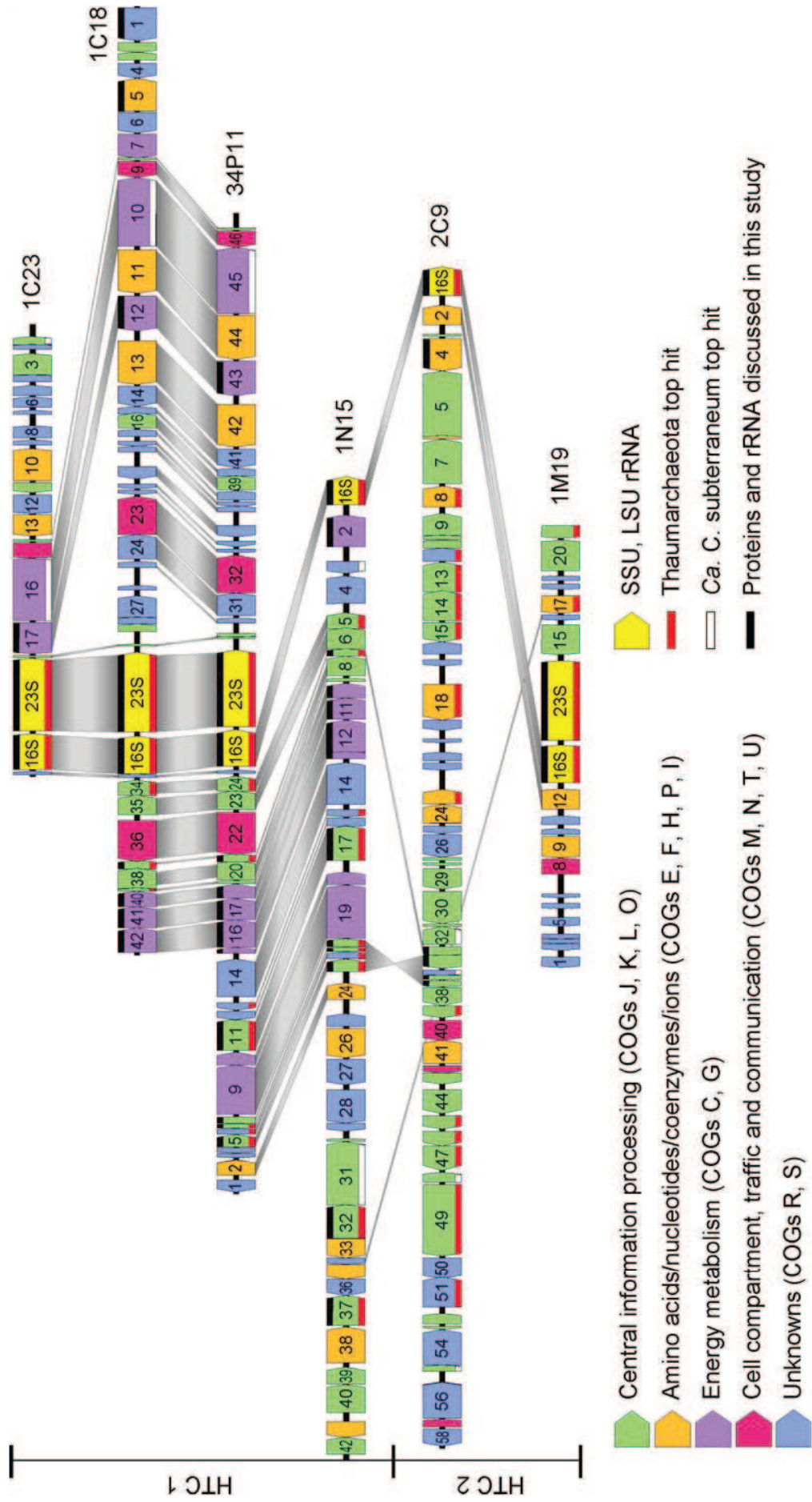


Figure 96 (page précédente). Synténie des six fosmides de Thaumarchaeota 1C23, 1C18, 34P11, 1N15, 2C9 et 1M19. Les couleurs des cadres de lecture ouverts (ORF) se réfèrent aux classifications fonctionnelles COG. **Vert** : Traitement de l'information génétique (J : traduction, biogénèse et structure du ribosome; K : Transcription; L : Réplication de l'ADN, recombinaison et réparation; O : Modification post-traductionnelle, turn-over des protéines, chaperonnes). **Orange** : Acides aminés/nucléotides/coenzymes/ions (E : Transport et métabolisme des acides aminés; F : Transport et métabolisme des nucléotides, H : Métabolisme des coenzymes, I : Métabolisme des lipides, P : Transport et métabolisme des ions inorganiques). **Violet** : Métabolisme énergétique (C : Production et conversion d'énergie; G : Métabolisme et transport des glucides). **Rouge** : Compartimentation, trafic et communication cellulaire (M : biogénèse de l'enveloppe cellulaire, membrane externe; N : Motilité cellulaire et sécrétion; T : Mécanismes de transduction du signal; U : Trafic intracellulaire, sécrétion, et transport vésiculaire). **Bleu** : Inconnu (R : prédiction de fonction très générale; S : Fonction inconnue). Les ORF marqués d'un trait rouge en bas ont comme meilleur hit par Blast une séquence de Thaumarchaeota. Les ORF marqués d'un trait blanc en bas ont comme meilleur hit par Blast une séquence de *Ca. C.subterraneum*. Les ORF marqués d'un trait noir en haut sont ceux qui seront discutés dans l'article. Les ORF codant pour des protéines similaires sur différents fosmides sont reliés par des traits gris (> 40% d'identité en acides aminés, E-value <10⁻⁵).

Fosmids	1N15	1C18	34P11	1C23	2C9	1M19
Size (bp)	36379	36885	39723	19222	44176	17060
Total fosmid G+C content (%)	46	51	51	46	67	62
SSU rRNA G+C content (%)	62.2	63.6	63.6	62.0	65.7	63.3
Number of ORFs	42	42	47	21	58	21
Average ORF length (bp)	789	745	687	691	721	450
ORFs with assigned function	30	23	25	11	38	9
Conserved hypothetical proteins	4	4	5	2	9	1
Hypothetical proteins	4	3	5	2	1	7
Proteins with no hits in any public database	1	7	8	2	3	2
RNA (tRNA and rRNA)	3	5	5	4	7	2

Tableau 20. Description des six fosmides de Thaumarchaeota hyperthermophiles.

3 Matériel et méthodes

3.1 Construction des jeux de données

3.1.1 Données nucléotidiques

189 séquences d'ARNr SSU de Thaumarchaeota ont été extraites de la base de données RDP (<http://rdp.cme.msu.edu/>; version 10, mise à jour 22; 30 août 2010 (Cole et al, 2008)). Ce jeu de données a été enrichi avec 1753 séquences d'ARNr SSU d'archées de plus de 1200 nucléotides disponibles à la RDP. D'autres séquences d'ARNr SSU provenant de représentants des Thaumarchaeota, Korarchaeota et '*Candidatus (Ca.) Caldiarchaeum subterraneum*' ont été récupérées dans GenBank à l'aide du logiciel BLASTn.

Toutes les séquences récupérées ont été alignées avec les séquences d'ARNr SSU thaumarchées issues des banques de clones et de fosmidés de cette étude (15 et 9 séquences, respectivement) à l'aide de Clustalw2 (paramètres par défaut (Larkin et al., 2007)). L'alignement résultant a été inspecté visuellement et raffiné manuellement si nécessaire avec ED. Une phylogénie préliminaire a été reconstruite en utilisant une méthode de distance (neighbor-joining). Sur la base de l'arbre résultant, nous avons sélectionné un sous-ensemble de 62 séquences reflétant la diversité des Thaumarchaeota et 28 séquences représentant les autres phylums d'archées. Les 90 séquences correspondantes ont été réalignées et l'alignement résultant raffiné en utilisant la même procédure. Les régions dont l'homologie est douteuse ont été éliminées permettant la conservation de 1000 positions pour les analyses phylogénétiques.

Nous avons construit un deuxième jeu de données basé sur la concaténation des séquences d'ARNr des grandes (LSU) et petites (SSU) sous-unités du ribosome, provenant de génomes complets d'archées ou de fosmidés d'archées disponibles dans Genbank, mais aussi des génomes non publiés des thaumarchées '*Ca. Nitrososphaera gargensis*' et '*Ca. Nitrososphaera viennensis*' (A. Spang, T. Rattei, C. Schleper, données non publiées), et de nos quatre fosmidés thaumarchéens (1M19, 1C23, 1C18 et 34P11) qui contiennent les deux gènes. Les séquences d'ARNr SSU et LSU ont été alignées séparément comme décrit ci-dessus et concaténées. Un total de 3462 positions a été conservé pour les analyses phylogénétiques.

3.1.2 Données protéiques

En utilisant la même approche, j'ai construit quatre supermatrices protéiques :

- La première concaténation (F1a, 653 positions) contient cinq protéines informationnelles : L31e, L39e et S19e, RPR2 et EF-1 α provenant des fosmidés 2C9, 1N15 et 34P11, des génomes complets d'Archaea disponibles au NCBI, ainsi que de '*Ca. Nitrososphaera gargensis*' et '*Ca. Nitrososphaera viennensis*'. L31e et L39e sont deux protéines ribosomales de la grande sous-unité, tandis que S19e fait partie de la petite sous-unité. La protéine RPR2 est une sous-unité de la ribonucléase P, un complexe protéique capable de générer des ARNt matures à partir de leurs précurseurs. EF-1 α est une sous-unité du facteur d'élongation-1 qui est impliqué dans la biosynthèse des protéines.
- La deuxième fusion (F2a, 1193 positions) contient une topo-isomérase primase (Toprim) et le facteur de libération du polypeptide aRF1 en plus des cinq protéines utilisés pour la fusion F1a. Cela a permis la construction d'un alignement plus long, mais l'inclusion d'une seule séquence de fosmide (1N15).

- Pour chaque concaténation, j'ai construit un deuxième alignement comprenant aussi des séquences eucaryotes en plus de la séquence Archaea (F1ae, 618 positions et F2ae, 1139 positions).

3.2 Reconstructions phylogénétiques

Pour les six jeux de données (ARNr SSU, ARNr SSU / LSU, F1a, F1ae, F2a et F2ae), des arbres phylogénétiques ont été calculés par maximum de vraisemblance (ML) et par des méthodes bayésiennes, en utilisant Treefinder (Jobb et al., 2004) et *via* MrBayes (Ronquist and Huelsenbeck, 2003), respectivement. Pour les analyses ML, les modèles d'évolution utilisés ont été choisis en fonction de l'outil de sélection de modèles disponible dans Treefinder. La robustesse des branches a été estimée par la méthode de bootstrap non paramétrique de Treefinder en appliquant les mêmes modèles d'évolution que pour la recherche de l'arbre ML. Pour les analyses bayésiennes nous avons utilisé le modèle GTR (pour les séquences d'ARNr) ou un modèle mixte (pour les concaténations de protéines). Dans les deux cas nous avons inclus une loi gamma (quatre catégories de sites et estimation du paramètre alpha). MrBayes a été exécuté sur quatre chaînes indépendantes pendant 1 million de générations et les arbres ont été échantillonnés toutes les 100 générations. Pour construire l'arbre consensus, les 1500 premiers arbres ont été rejetés ("burn-in"). Pour les quatre concaténations protéiques, nous avons également effectué des analyses phylogénétiques utilisant PhyloBayes 3.2 (Lartillot et al., 2009) avec le modèle CAT. Deux chaînes indépendantes ont été exécutées pendant au moins 10000 cycles, avec échantillonnage d'un arbre sur dix. Les 300 premiers arbres ont été rejetés ("burn-in") et les arbres restants de chaque chaîne ont été utilisés pour tester la convergence et calculer l'arbre consensus à 50%. Les probabilités postérieures calculées par MrBayes seront dénommées PP_{MB} alors que celles calculées avec PhyloBayes sont appelées PP_{PB} .

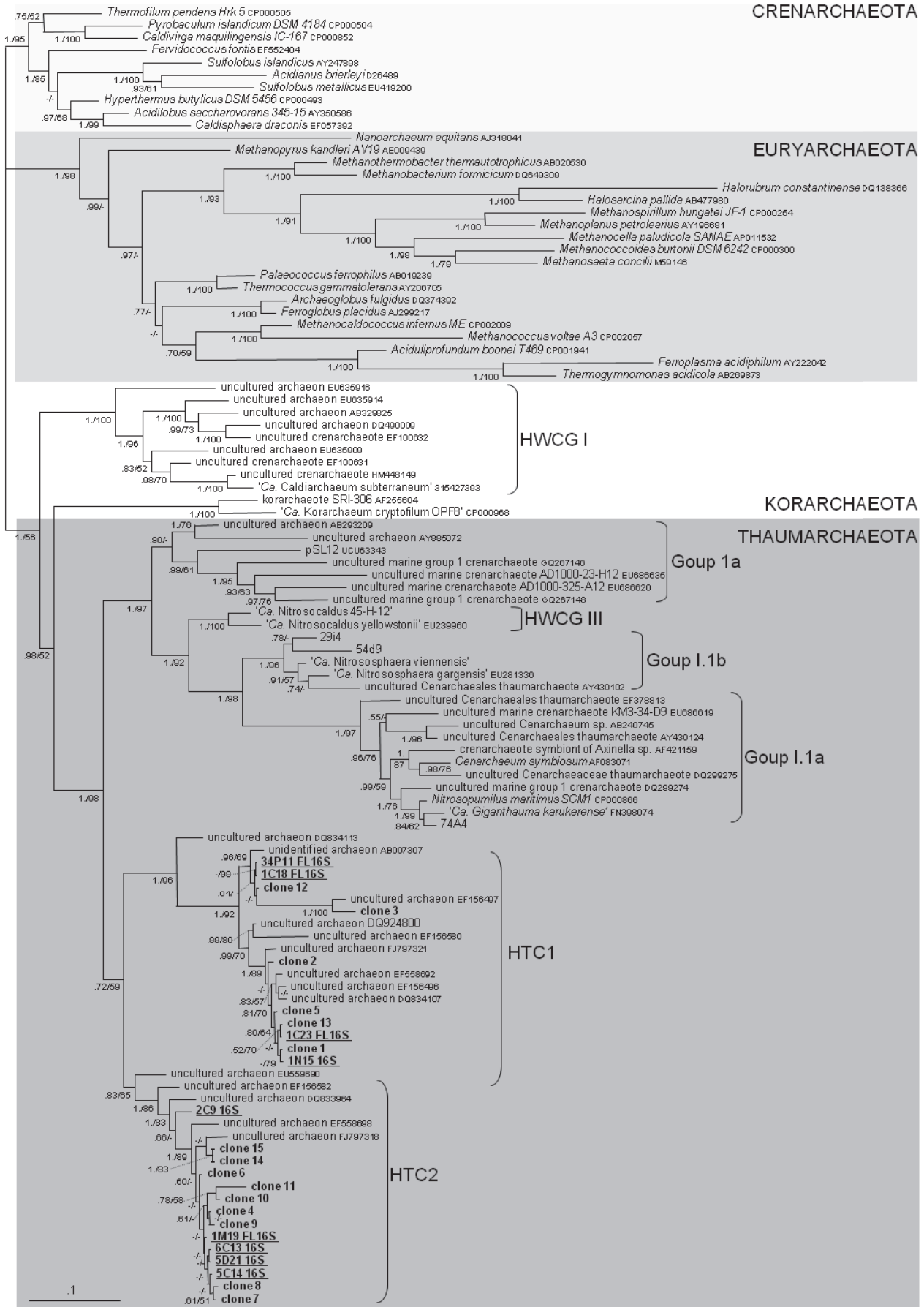
4 Résultats

4.1 Relations phylogénétiques au sein des séquences de Thaumarchaeota

L'analyse phylogénétique bayésienne des 16 séquences d'ARNr SSU de thaumarchées (Kam37 clone de 1 à 15, et 35) a révélé que ces clones appartiennent à deux clusters distincts et bien soutenus ($PP_{MB} = 1.00$ pour les deux groupes, et valeurs de bootstrap (VB) = 92% et 86%) (Figure 97). Ces deux clusters émergent avant les autres lignées de thaumarchées, c'est-à-dire avant les groupes I.1a (contenant *Caenarchaeum symbiosum* et *Nitrosopumilus maritimus*), I.1b (contenant '*Ca. Nitrosophaera gargensis*' et '*Ca. N. viennensis*'), 1a et le groupe HWCG III ("Hot Water Crenarchaeotic Group", contenant '*Ca. Nitrosocaldus yellowstonii*') ($PP_{MB} = 1.00$ et VB = 97%) (Figure 97). Sur la base de cet arbre, les deux groupes contenant nos clones forment deux groupes-frères, mais le soutien statistique en faveur de cette relation n'est pas significatif ($PP_{MB} = 0,72$ et VB = 59%). Les deux groupes contenant nos clones seront dénommés HTC1 ("Hyperthermophilic Thaumarchaeal Cluster 1") et HTC2 ("Hyperthermophilic Thaumarchaeal Cluster 2"), parce qu'ils ne contiennent que des séquences d'ARNr SSU d'archées non cultivées d'habitats de type thermes.

Au final, l'analyse phylogénétique des séquences d'ARNr SSU suggère que HTC1 et HTC2 représentent deux nouvelles lignées majeures de Thaumarchaeota.

Figure 97 (page suivante). Phylogénie non racinée des ARNr SSU de Thaumarchaeota. Arbre reconstruit avec MrBayes sur séquences représentatives de la diversité des Archaea (1000 positions). Les séquences issues de la présente étude sont en gras. Les séquences issues des fosmidés sont soulignées. La barre d'échelle représente le nombre moyen de substitutions par site. Les nombres aux nœuds représentent les probabilités postérieures inférées par MrBayes et les valeurs de bootstrap calculées par Treefinder. Seules les valeurs supérieures à 0.5 et 50%, respectivement, sont indiquées. HTC = "Hyperthermophilic Thaumarchaeal Cluster", HWCG = "Hot Water Crenarchaeotic Group".



4.2 Position phylogénétique des Thaumarchaeota au sein des Archaea

Afin d'explorer plus précisément le placement des séquences de fosmides au sein des archées, nous avons effectué des analyses phylogénétiques à la fois sur la base des ARNr SSU et LSU, mais aussi des marqueurs protéiques. Tout d'abord, nous avons analysé les séquences d'ARNr SSU et LSU concaténées pour les quatre fosmides qui contiennent les deux gènes (fosmides 1C18, 34P11, 1C23 et 1M19) (Figure 98). Les arbres sont en accord avec l'arbre reconstruit sur l'ARNr SSU seul.

Les fosmides 34P11 et 1C18 apparaissent très étroitement liés ($PP_{MB} = 1.00$ et $VB = 99\%$) et regroupés avec 1C23 ($PP_{MB} = 1.00$ et $VB = 100\%$) (Figure 98). Fait intéressant, alors que le soutien pour le groupement de HTC1 et HTC2 est faible dans l'arbre ARNr SSU ($PP_{MB} = 0.72$ = $VB = 59\%$) (Figure 97), il est important lorsque les ARNr SSU et LSU sont analysés simultanément ($PP_{MB} = 0.99$ et $VB = 74\%$) (Figure 98). Enfin, tout comme dans l'arbre ARNr SSU, un support très fort est retrouvé en faveur du branchement profond de HTC1 et HTC2 au sein des Thaumarchaeota ($PP_{MB} = 1.00$ et $VB = 100\%$). La principale différence entre les deux analyses repose sur le regroupement des Thaumarchaeota avec '*Ca. Caldarchaeum subterraneum*' dans l'arbre ARNr SSU + LSU ($PP_{MB} = 1.00$ et $VB = 99\%$), alors que dans l'arbre SSU, les Thaumarchaeota apparaissent plus étroitement liées aux Korarchaeota, mais avec un support statistique plus faible ($PP_{MB} = 0.98$ et $VB = 52\%$ seulement).

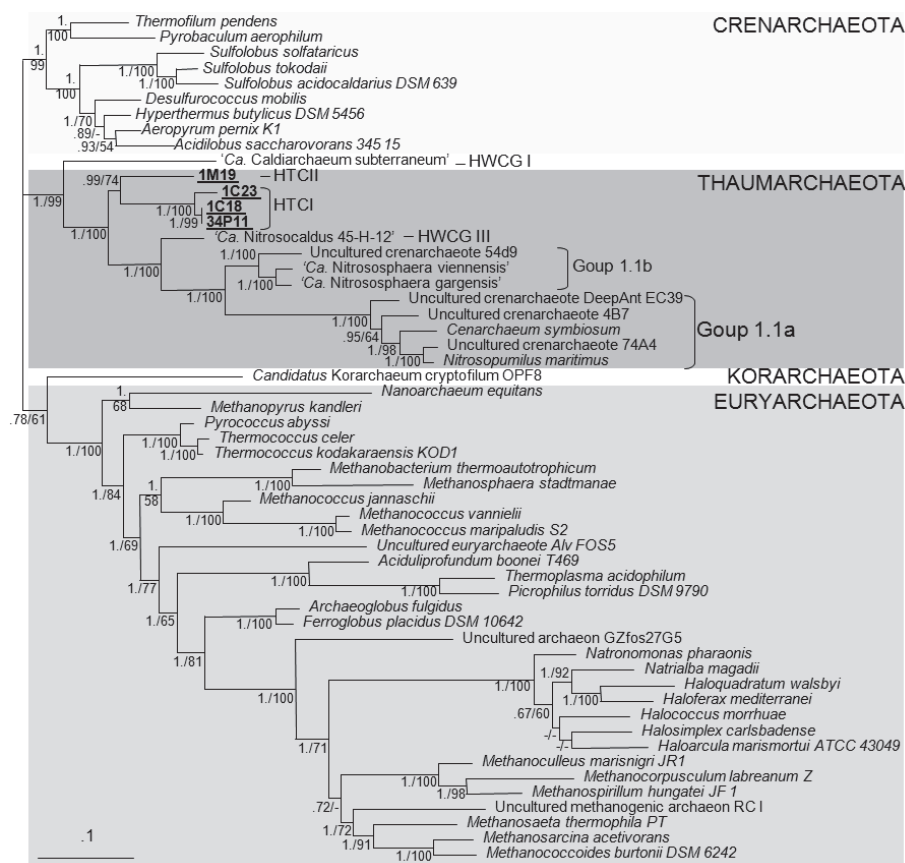


Figure 98. Arbre bayésien non raciné basé sur la concaténation des séquences d'ARNr SSU et LSU (55 séquences, 3462 positions). Arbre inféré par MrBayes. Les noms des séquences de fosmides issues de cette étude sont soulignés. La barre d'échelle représente le nombre moyen de substitutions par site. Les nombres aux nœuds représentent respectivement les probabilités postérieures inférées par MrBayes et les valeurs de bootstrap calculées par TreeFinder. Seules les valeurs supérieures à 0,50 et 50%, respectivement sont indiquées. HTC = "hyperthermophilic Thaumarchaeal Cluster"; HWCG = "Hot Water Crenarchaeotic Group".

Les relations entre les groupes de thaumarchées ainsi que le regroupement des Thaumarchaeota avec "*Ca. Caldiarchaeum subterraneum*" ont été confirmés par les analyses phylogénétiques en maximum de vraisemblance et en bayésien de la concaténation de cinq protéines informationnelles (F1a, 653 positions) (trois protéines ribosomales (L31e, L39e et S19e), la RNase P et le facteur d'élongation EF-1 α codés par trois fosmides : 2C9 (HTC2), et 1N15 et 34P11 (HTC1)) (Figure 99A). Plus précisément, l'arbre F1a soutient la monophylie des HTC1 ($PP_{MB} = 1.00$, VB = 100% et $PP_{PB} = 1,00$), le groupement de HTC1 et HTC2 ($PP_{MB} = 1.0$, VB = 50% et $PP_{PB} = 0.95$); leur émergence basale dans les Thaumarchaeota ($PP_{MB} = 0.56$ et $PP_{PB} = 0.81$) et le regroupement des Thaumarchaeota avec "*Ca. Caldiarchaeum subterraneum*" ($PP_{MB} = 1.0$, VB = 50% et $PP_{PB} = 1.00$) (Figure 99A).

L'ajout de deux autres protéines, l'ADN primase Toprim et le facteur aRF1, à notre analyse (F2a, 1193 positions) (Figure 99B), confirme ces regroupements, avec un support plus important.

Le groupement observé entre les Thaumarchaeota et '*Ca. Caldiarchaeum subterraneum*' est en accord avec les résultats publiés sur le génome complet de cette archée non cultivée (Nunoura et al., 2010). Dans cet article, il était proposé que cette lignée puisse représenter un nouveau phylum d'Archaea provisoirement appelé 'Aigarchaeota'. L'argument principal a été la découverte de gènes de type eucaryote impliqués dans une voie d'ubiquitinylation et qui n'existent pas chez les Thaumarchaeota. Cependant, à ce moment-là, les seuls génomes complets disponibles au sein des Thaumarchaeota étaient ceux du groupe I.1a, une lignée de thaumarchées marines très dérivées. Bien que les groupes HTC1 et HTC2 représentent des lignées distinctes de '*Ca. Caldiarchaeum subterraneum*', il n'est pas possible d'exclure que ce dernier soit en fait une lignée basale de Thaumarchaeota plutôt qu'un représentant d'un nouveau phylum. Seules de nouvelles études génomiques et microbiologiques de '*Ca. Caldiarchaeum subterraneum*' (et d'autres membres du groupe HWCG I), ainsi que des représentants de HTC1 et HTC2 aideront à décider si ces 'Aigarchaeota' doivent être considérées comme un nouveau phylum d'archées ou comme un ordre au sein des Thaumarchaeota.

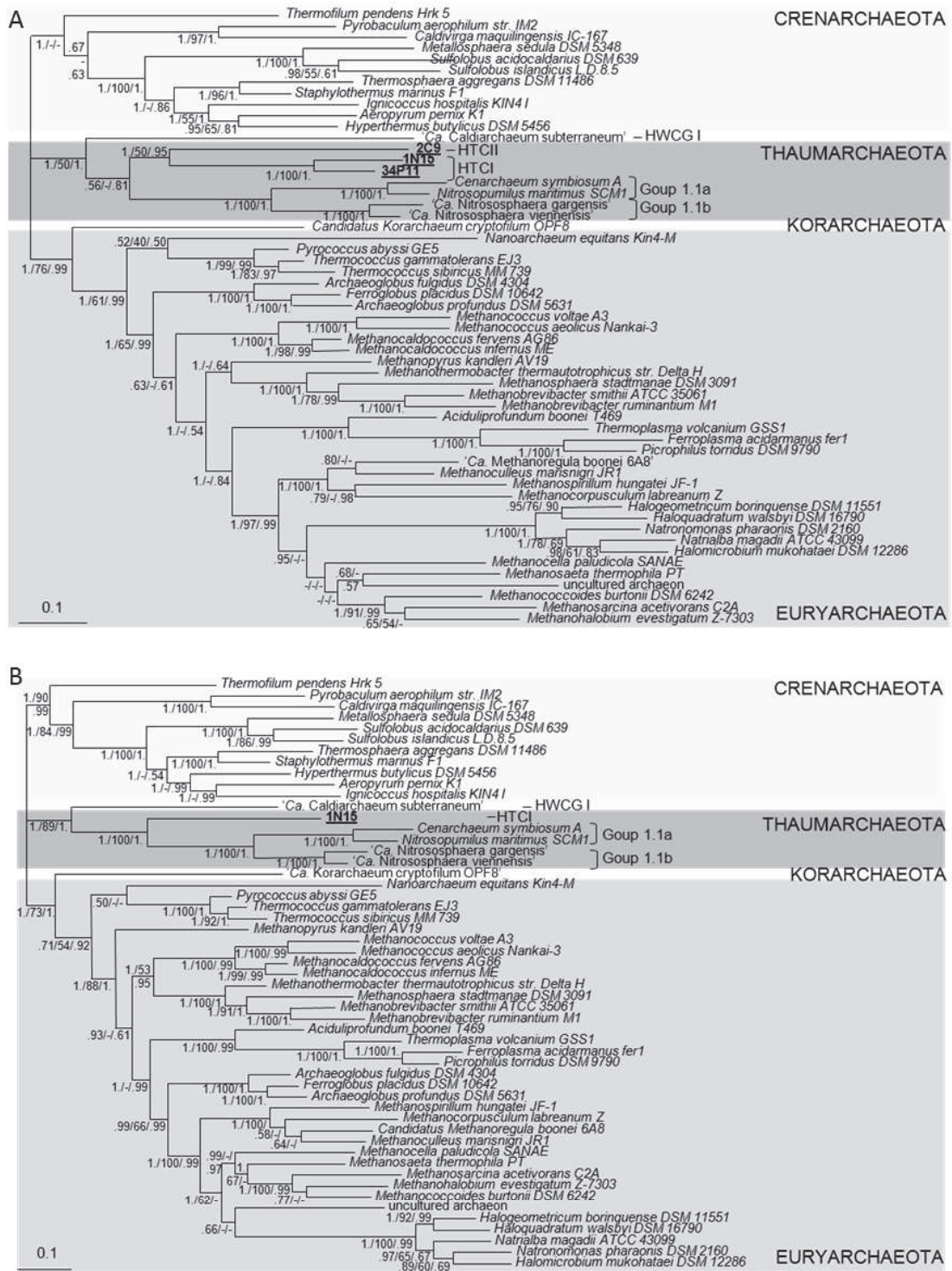


Figure 99. Arbres bayésiens non racinés basés sur (A) la concaténation F1a (cinq protéines informationnelles, 55 séquences, 653 positions) et (B) la concaténation F2a (sept protéines informationnelles, 53 séquences, 1193 positions). Arbre inféré par MrBayes. La barre d'échelle représente le nombre moyen de substitution par site. Les nombres aux nœuds représentent respectivement les probabilités postérieures inférées par MrBayes, les valeurs de bootstrap calculées par TreeFinder et les probabilités postérieures inférées par Phylobayes. Seules les valeurs supérieures à 0,50, 50% et 0,5, respectivement, sont indiquées. HTC = "hyperthermophilic Thaumarchaeal Cluster"; HWCG = "Hot Water Crenarchaeotic Group".

4.3 Relations entre Eucarya et Archaea

Après inclusion d'homologues eucaryotes dans les concaténations F1a et F2a, nos analyses par maximum de vraisemblance et par des approches bayésiennes nous permettent d'observer une affinité fortement soutenue entre Eucarya et Thaumarchaeota (y compris '*Ca. Caldiarchaeum subterraneum*') : $PP_{MB} = 1.00$ et $PP_{PB} = 0.98$ dans l'arbre F1ae (Figure 100A), et $PP_{MB} = 1.00$, VB = 84%, et $PP_{PB} = 99$ dans l'arbre F2ae (Figure 100B). La plupart des études sur la phylogénie des Archaea utilisent des séquences eucaryotes comme groupe extérieur, soit parce que les homologues bactériens n'existent pas, soit parce que les séquences eucaryotes sont plus proches de celles des Archaea, limitant ainsi les artefacts de reconstruction d'arbres. Toutefois, si les Eucarya descendent des Archaea (*i.e.* hypothèses 2D), alors les séquences eucaryotes ne peuvent pas être utilisées comme groupe extérieur valide. Tant que la nature des relations de parentés entre Eucarya et Archaea n'est pas éclaircie, il nous faut envisager deux interprétations des phylogénies obtenues.

4.3.1 Hypothèse 3D

Si l'hypothèse 3D s'avère correcte, alors les séquences eucaryotes peuvent en effet être utilisées comme groupe extérieur. Dans ce cas, nos analyses phylogénétiques soutiennent fortement l'émergence précoce des Thaumarchaeota et de '*Ca. Caldiarchaeum subterraneum*'.

Les premières analyses phylogénétiques basées sur les séquences d'ARNr SSU avaient permis de proposer que les Thaumarchaeota et les Crenarchaeota forment des groupes frères (DeLong, 1992). En conséquence, les Thaumarchaeota ont, à l'époque, été dénommées Crenarchaeota mésophiles, même si l'existence de lignées thermophiles était postulée. Cependant, les analyses phylogénétiques plus récentes basées sur des données protéiques ont soutenu l'émergence plus précoce de cette lignée dans l'arbre des Archaea, c'est-à-dire avant la spéciation des Euryarchaeota et des Crenarchaeota (Brochier-Armanet et al., 2008; Spang et al., 2010; Walker et al., 2010). Toutefois, le soutien statistique en faveur de cette position s'est amenuisé avec l'ajout de nouveaux représentants de Thaumarchaeota suggérant qu'une conclusion définitive quant à la position phylogénétique de ce phylum au sein des Archaea ne viendra qu'après davantage d'analyses et de données.

Puisque toutes les analyses précédentes ont été basées sur les protéines de quelques génomes appartenant à deux lignées de thaumarchées étroitement liées (Groupes I.1a et I.1b), les artefacts de reconstruction d'arbres ou de biais méthodologiques ne pouvaient être exclus. Ici, nous avons eu pour la première fois l'occasion de tester ces hypothèses avec des représentants de plusieurs lignées éloignées. Nos résultats semblent confirmer les analyses précédentes et suggèrent que les Thaumarchaeota pourraient en effet représenter la lignée la plus basale d'Archaea. Toutefois, ce résultat devra être confirmé lorsque des données supplémentaires sur les Thaumarchaeota et les groupe HWCG I seront disponibles.

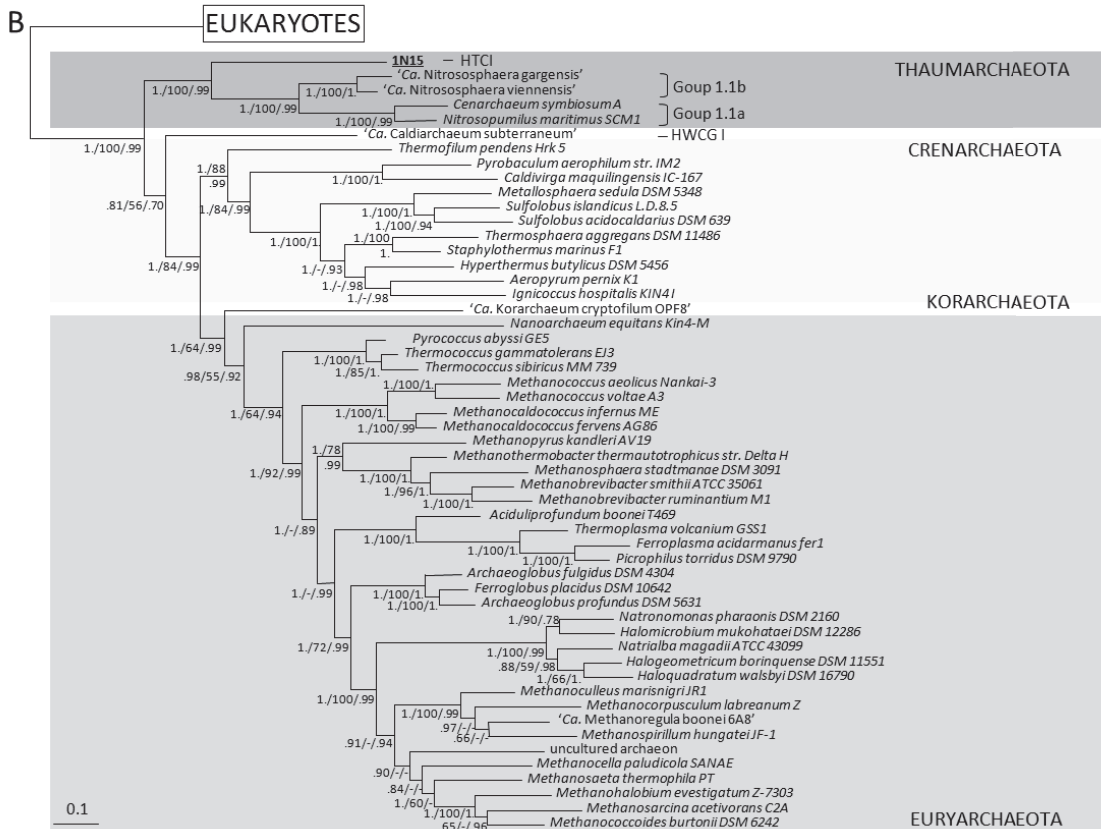
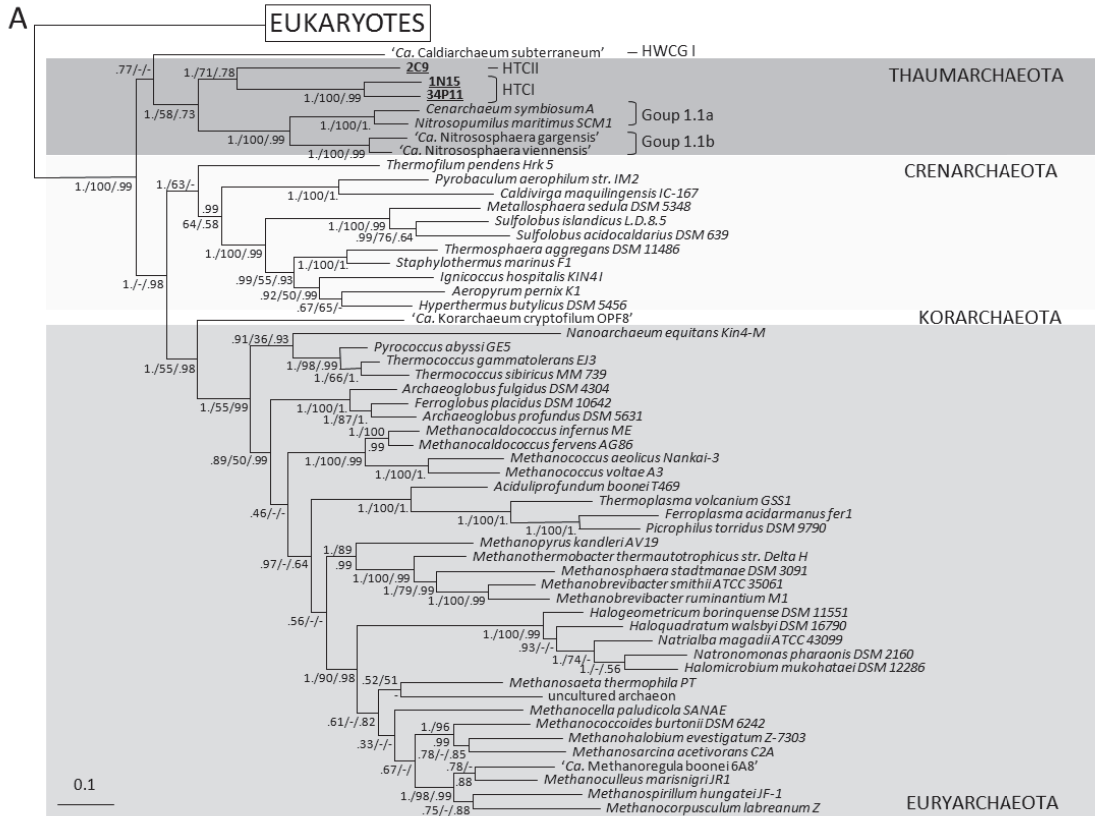
4.3.2 Hypothèse 2D

Dans le cadre de l'hypothèse 2D, nos résultats soutiennent fortement le regroupement des Eucarya et des Thaumarchaeota, en accord avec les résultats récents de Kelly *et al.* sur la base d'analyses phylogénomiques (Kelly et al., 2011). L'hypothèse d'une origine thaumarchée des Eucarya, bien que séduisante à plusieurs égards, a été discutée de façon critique par Patrick Forterre (Forterre, 2011).

A défaut de nous permettre de valider un scénario particulier, et même s'il s'agit de résultats préliminaires basés sur quelques protéines seulement, les résultats obtenus affaiblissent le support pour certains scénarios 2D, tels que :

- **l'hypothèse hydrogène**, qui postule que les eucaryotes soient issus de la fusion d' α -protéobactéries productrices d'hydrogène et d'euryarchées méthanogènes consommatrices d'hydrogène (Martin and Muller, 1998);
- **L'hypothèse de Rivera et Lake**, selon laquelle l'archée-hôte appartient au groupe des Crenarchaeota (Rivera and Lake, 1992);
- **Les hypothèses de Searcy** ("Sulfur syntrophy hypothesis") et **de Lynn Margulis**, qui font tous les deux l'hypothèse que l'hôte aurait été une archée semblable aux membres du genre *Thermoplasma*, une lignée d'Euryarchaeota ((Margulis et al., 2006; Searcy, 2003));
- **L'hypothèse syntrophique**, faisant intervenir une archée méthanogène, c'est-à-dire une euryarchée (Lopez-Garcia and Moreira, 2006; Moreira and Lopez-Garcia, 1998).

Figure 100 (page suivante). Arbres bayésiens basés sur (A) la concaténation F1ae (cinq protéines informationnelles, 66 séquences dont 11 séquences eucaryotes, 618 positions) et (B) la concaténation F2ae (sept protéines informationnelles, 64 séquences dont 11 séquences eucaryotes, 1139 positions). Arbres inférés par MrBayes. La barre d'échelle représente le nombre moyen de substitutions par site. Les nombres aux nœuds représentent respectivement les probabilités postérieures inférées par MrBayes, les valeurs de bootstrap calculées par TreeFinder et les probabilités postérieures inférées par Phylobayes. Seules les valeurs supérieures à 0,50 et 50% sont indiquées. HTC = "hyperthermophilic Thaumarchaeal Cluster"; HWCG = "Hot Water Crenarchaeotic Group".



5 Conclusion

Cette analyse métagénomique a permis le séquençage de fosmides appartenant à deux lignées majeures d'Archaea pour lesquels aucune donnée autre que des séquences d'ARNr SSU n'était disponible jusqu'à présent. Les phylogénies obtenues sur la base de marqueurs protéiques sont en accord avec la phylogénie de référence des Archaea obtenue à partir de la concaténation de 57 protéines ribosomales (Figure 95), suggérant que nous avons mis en évidence de nouveaux marqueurs utilisables pour reconstruire la phylogénie des Archaea. Les phylogénies obtenues nous ont également permis de tester les relations entre Eucarya et les différents phylums d'Archaea et d'affaiblir certaines hypothèses 2D sur l'origine des eucaryotes. Ces résultats sont basés sur seulement quelques protéines et sont donc à considérer avec précaution. Néanmoins, ils ouvrent des pistes intéressantes sur la manière dont les relations de parenté au sein des Archaea et entre Archaea et Eucarya peuvent être étudiées.

CONCLUSION

Mon travail de doctorat s'inscrit au cœur du débat sur l'origine des eucaryotes et leur évolution, et a bénéficié d'une conjecture favorable grâce :

- A la démocratisation de nouvelles méthodes de reconstruction (telles que les méthodes bayésiennes) et au développement de nouveaux modèles d'évolution, toujours plus réalistes (tels que les modèles LG, CAT, ...);
- Aux capacités informatiques permettant de traiter des jeux de données de grande taille;
- A la caractérisation des SME impliquées dans la division cellulaire par des méthodes de protéomique à haut débit;
- Aux données génomiques disponibles de plus en plus nombreuses.

L'objectif de mon projet de recherche était double : apporter des éléments permettant d'affiner notre connaissance de LECA, et améliorer notre compréhension de l'évolution des eucaryotes. Pour réaliser ce projet, j'ai utilisé une approche originale couplant à la fois l'étude phylogénomique de systèmes biologiques eucaryotes et l'étude des relations de parenté entre organismes.

Le premier axe de cette approche a consisté en l'étude phylogénomique de deux structures impliquées dans certaines étapes du cycle cellulaire. Ces analyses ont constitué un travail innovant puisque jusque-là, aucune étude à grande échelle évolutive n'avait été menée sur ces structures. Par conséquent, le déroulement du cycle cellulaire chez LECA était jusque-là totalement inconnu. Grâce à l'étude de l'évolution de la cytocinèse, j'ai pu inférer qu'il existait vraisemblablement, chez cet ancêtre, un système complexe (comptant possiblement ~90 composants), issu d'une innovation eucaryote, impliqué dans les dernières étapes de la division cellulaire, et à partir de laquelle ont dérivé les différentes structures ayant ce rôle dans les eucaryotes actuels. En parallèle, par le biais de l'analyse phylogénomique du cyclosome, j'ai pu montrer que la plupart des composants de l'APC/C, ses adaptateurs et ses principales cibles étaient déjà présents chez LECA et qu'un système similaire à celui qui régule le cycle cellulaire des eucaryotes actuels existait déjà probablement dans cet organisme ancestral.

Ces analyses apportent donc des informations précieuses sur l'histoire ancienne et sur l'origine des eucaryotes et indiquent que les deux systèmes étudiés étaient déjà très élaborés chez LECA. Ces résultats permettent de compléter le portrait-robot de LECA déjà esquissé par plusieurs autres études (Figure 101). Au vu de l'ensemble de ces analyses, il apparaît que LECA était une cellule complexe présentant déjà de très nombreuses caractéristiques des eucaryotes modernes, dont beaucoup représentent des innovations spécifiques des eucaryotes.

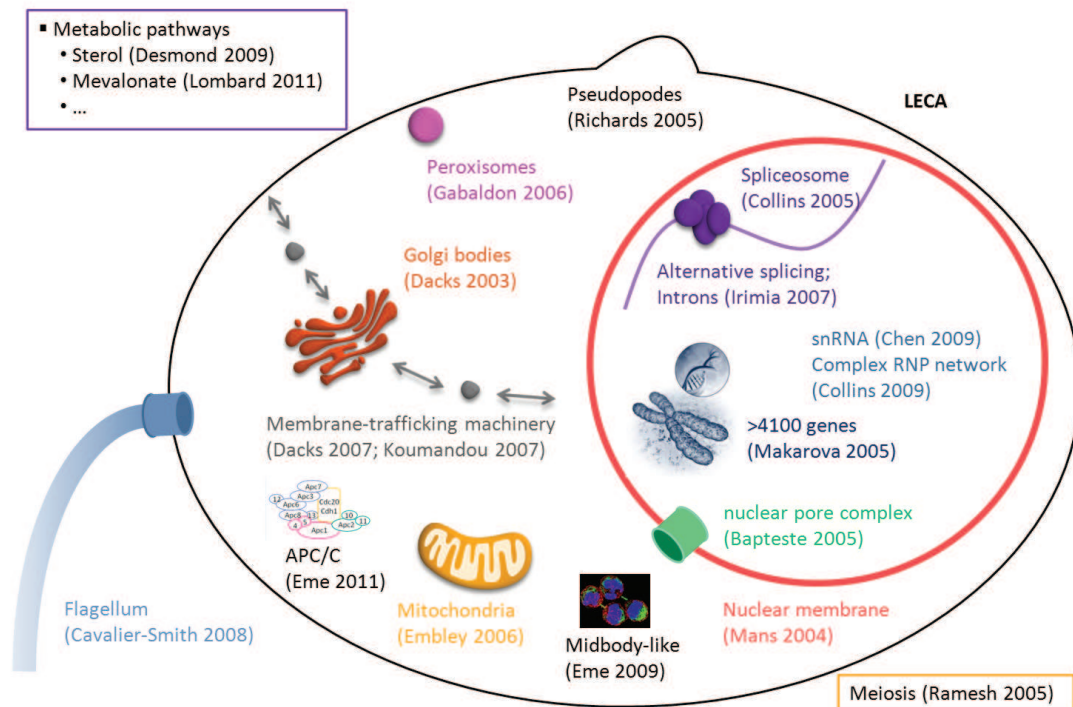


Figure 101. Représentation schématique de LECA et des systèmes eucaryotes-spécifiques déjà présents chez cet ancêtre.

Les nombreuses spécificités des eucaryotes qui se sont mises en place entre l'émergence de ce domaine et la diversification donnant naissance aux eucaryotes actuels implique une quantité d'évolution importante. Dans le cadre de l'hypothèse 2D, l'absence de ces intermédiaires peut s'expliquer si l'origine des eucaryotes est très récente, auquel cas l'apparition de tous ces caractères s'est faite sur une très courte période. Mais il faut alors expliquer comment ces spécificités sont apparues si rapidement à partir de caractères archéens et d'un génome de taille modeste (~1000 gènes). Dans le cadre d'une hypothèse 3D, la contrainte temporelle se relâche. De plus, il est possible d'envisager qu'un certain nombre de gènes impliqués dans ces systèmes étaient déjà présents chez l'ancêtre commun des Archaea et des Eucarya. Mais dans ce cas, se pose toujours le problème de l'absence d'intermédiaires et la question de la nature du lien entre Archaea et Eucarya reste toujours ouverte.

En parallèle, ces analyses pourraient également constituer une avancée de la connaissance de certaines parties du cycle cellulaire dans les espèces vivant actuellement. En effet, bien que très étudié, ce cycle cellulaire reste encore mal caractérisé dans de nombreuses lignées eucaryotes. Dans ce contexte, mes analyses représentent un travail inédit puisqu'aucune analyse à grande échelle des protéines impliquées dans les diverses parties du cycle n'avait été réalisée jusque-là. Ce travail permet donc de dévoiler certains aspects de la division cellulaire dans plusieurs lignées de protistes pour lesquels les caractérisations expérimentales manquent, offrant ainsi des candidats de choix pour les travaux expérimentaux, point particulièrement important en ce qui concerne les lignées pathogènes.

La suite logique à donner à ces analyses sera d'explorer l'histoire évolutive des systèmes impliqués dans les autres parties du cycle cellulaire. Dans cette optique, j'ai entrepris l'étude phylogénomique de l'appareil mitotique.

Le deuxième axe de ma thèse a consisté en la recherche de nouveaux marqueurs phylogénétiques pour l'étude de la phylogénie des eucaryotes, et en particulier des nœuds profonds de celle-ci. Jusque-là, les relations de parenté entre eucaryotes ont toujours été investigués à partir du même ensemble de protéines informationnelles. Malgré leur intense utilisation au cours de la dernière décennie, ces données n'ont pas suffi à résoudre l'ensemble des nœuds de l'arbre. Lors de mes analyses, et en dépit ce qui avait été présumé jusque-là, j'ai pu montrer que, même si les protéines opérationnelles évoluent plus vite que les protéines informationnelles, et sont moins bien conservées (*i.e.* plus souvent perdues), au moins une partie d'entre elles porte un signal phylogénétique de qualité et exploitable pour inférer les relations de parenté entre lignées eucaryotes. De fait, ce travail a indiqué que l'utilisation de ces protéines constitue un point de comparaison pour les relations inférées par les protéines informationnelles, et également, que l'utilisation combinée de ces deux types de données est une piste indispensable à suivre à l'avenir pour mieux résoudre certaines parties de l'arbre des eucaryotes (ex : la monophylie des Excavata, les relations au sein des Plantae, etc.). Ainsi, j'ai pu mettre en évidence de nouveaux marqueurs protéiques opérationnels qui pourront, tout comme les protéines informationnelles, être utilisés comme référence, et être mis à jour au fur et à mesure que de nouvelles données génomiques deviendront disponibles. Ces résultats ouvrent le champ des possibles quant à l'emploi d'autres marqueurs encore inexploités qui sont potentiellement nombreux à pouvoir être utilisés. Il est donc nécessaire de poursuivre l'exploration d'autres SME.

En conclusion, ce travail a permis de montrer le potentiel des approches évolutives intégrées pour l'étude de l'évolution des eucaryotes par la combinaison des analyses des systèmes biologiques et de la phylogénie des organismes.

BIBLIOGRAPHIE

- Adachi, J., and Hasegawa, M. (1995). Improved dating of the human/chimpanzee separation in the mitochondrial DNA tree: heterogeneity among amino acid sites. *J Mol Evol* 40, 622-628.
- Adl, S.M., Simpson, A.G., Farmer, M.A., Andersen, R.A., Anderson, O.R., Barta, J.R., Bowser, S.S., Brugerolle, G., Fensome, R.A., Fredericq, S., *et al.* (2005). The new higher level classification of eukaryotes with emphasis on the taxonomy of protists. *J Eukaryot Microbiol* 52, 399-451.
- Andersen, J.S., Lyon, C.E., Fox, A.H., Leung, A.K., Lam, Y.W., Steen, H., Mann, M., and Lamond, A.I. (2002). Directed proteomic analysis of the human nucleolus. *Curr Biol* 12, 1-11.
- Andersson, J.O. (2005). Lateral gene transfer in eukaryotes. *Cell Mol Life Sci* 62, 1182-1197.
- Angenstein, F., Evans, A.M., Ling, S.C., Settlage, R.E., Ficarro, S., Carrero-Martinez, F.A., Shabanowitz, J., Hunt, D.F., and Greenough, W.T. (2005). Proteomic characterization of messenger ribonucleoprotein complexes bound to nontranslated or translated poly(A) mRNAs in the rat cerebral cortex. *J Biol Chem* 280, 6496-6503.
- Aravind, L., Iyer, L.M., and Koonin, E.V. (2006). Comparative genomics and structural biology of the molecular innovations of eukaryotes. *Curr Opin Struct Biol* 16, 409-419.
- Archibald, J.M. (2005). Jumping genes and shrinking genomes--probing the evolution of eukaryotic photosynthesis with genomics. *IUBMB Life* 57, 539-547.
- Archibald, J.M. (2009). The puzzle of plastid evolution. *Curr Biol* 19, R81-88.
- Archibald, J.M., Longet, D., Pawlowski, J., and Keeling, P.J. (2003). A novel polyubiquitin structure in Cercozoa and Foraminifera: evidence for a new eukaryotic supergroup. *Mol Biol Evol* 20, 62-66.
- Arkowitz, R.A., and Iglesias, P.A. (2008). Basic principles of polarity establishment and maintenance. Conference on Mechanisms of Cell Polarity. *EMBO Rep* 9, 847-852.
- Bachvaroff, T.R., Sanchez Puerta, M.V., and Delwiche, C.F. (2005). Chlorophyll c-containing plastid relationships based on analyses of a multigene data set with all four chromalveolate lineages. *Mol Biol Evol* 22, 1772-1782.
- Baier, M., and Dietz, K.J. (1996). Primary structure and expression of plant homologues of animal and fungal thioredoxin-dependent peroxide reductases and bacterial alkyl hydroperoxide reductases. *Plant Mol Biol* 31, 553-564.
- Baier, M., and Dietz, K.J. (1997). The plant 2-Cys peroxiredoxin BAS1 is a nuclear-encoded chloroplast protein: its expressional regulation, phylogenetic origin, and implications for its specific physiological function in plants. *Plant J* 12, 179-190.
- Baldauf, S.L., and Palmer, J.D. (1993). Animals and Fungi Are Each Others Closest Relatives - Congruent Evidence from Multiple Proteins. *Proc Natl Acad Sci U S A* 90, 11558-11562.
- Baldauf, S.L., Roger, A.J., Wenk-Siefert, I., and Doolittle, W.F. (2000). A kingdom-level phylogeny of eukaryotes based on combined protein data. *Science* 290, 972-977.
- Bannai, H., Tamada, Y., Maruyama, O., Nakai, K., and Miyano, S. (2002). Extensive feature detection of N-terminal protein sorting signals. *Bioinformatics* 18, 298-305.
- Baptiste, E., Brinkmann, H., Lee, J.A., Moore, D.V., Sensen, C.W., Gordon, P., Durufle, L., Gaasterland, T., Lopez, P., Muller, M., *et al.* (2002). The analysis of 100 genes supports the grouping of three highly divergent amoebae: Dictyostelium, Entamoeba, and Mastigamoeba. *Proc Natl Acad Sci U S A* 99, 1414-1419.
- Baptiste, E., Charlebois, R.L., MacLeod, D., and Brochier, C. (2005). The two tempos of nuclear pore complex evolution: highly adapting proteins in an ancient frozen structure. *Genome Biol* 6, R85.
- Baptiste, E., and Philippe, H. (2002). The potential value of indels as phylogenetic markers: position of trichomonads as a case study. *Mol Biol Evol* 19, 972-977.
- Barns, S.M., Delwiche, C.F., Palmer, J.D., and Pace, N.R. (1996). Perspectives on archaeal diversity, thermophily and monophyly from environmental rRNA sequences. *Proc Natl Acad Sci U S A* 93, 9188-9193.
- Barranco-Medina, S., Lazaro, J.J., and Dietz, K.J. (2009). The oligomeric conformation of peroxiredoxins links redox state to function. *FEBS Lett* 583, 1809-1816.
- Basu, M.K., and Koonin, E.V. (2005). Evolution of eukaryotic cysteine sulfinic acid reductase, sulfiredoxin (Srx), from bacterial chromosome partitioning protein ParB. *Cell Cycle* 4, 947-952.

- Baurain, D., Brinkmann, H., Petersen, J., Rodriguez-Espeleta, N., Stechmann, A., Demoulin, V., Roger, A.J., Burger, G., Lang, B.F., and Philippe, H. (2010). Phylogenomic evidence for separate acquisition of plastids in cryptophytes, haptophytes, and stramenopiles. *Mol Biol Evol* 27, 1698-1709.
- Becker, B., and Marin, B. (2009). Streptophyte algae and the origin of embryophytes. *Ann Bot* 103, 999-1004.
- Belinky, F., Cohen, O., and Huchon, D. (2010). Large-scale parsimony analysis of metazoan indels in protein-coding genes. *Mol Biol Evol* 27, 441-451.
- Bell, S.D., and Jackson, S.P. (2001). Mechanism and regulation of transcription in archaea. *Curr Opin Microbiol* 4, 208-213.
- Biteau, B., Labarre, J., and Toledano, M.B. (2003). ATP-dependent reduction of cysteine-sulphinic acid by *S. cerevisiae* sulphiredoxin. *Nature* 425, 980-984.
- Blair, J.E., Ikeo, K., Gojobori, T., and Hedges, S.B. (2002). The evolutionary position of nematodes. *BMC Evol Biol* 2, 7.
- Bodyl, A. (2005). Do plastid-related characters support the chromalveolate hypothesis. *J Phycol* 41, 712-719.
- Bourlat, S.J., Nielsen, C., Economou, A.D., and Telford, M.J. (2008). Testing the new animal phylogeny: a phylum level molecular analysis of the animal kingdom. *Mol Phylogenet Evol* 49, 23-31.
- Breglia, S.A., Slamovits, C.H., and Leander, B.S. (2007). Phylogeny of phagotrophic euglenids (Euglenozoa) as inferred from hsp90 gene sequences. *J Eukaryot Microbiol* 54, 86-92.
- Brehelin, C., Meyer, E.H., de Souris, J.P., Bonnard, G., and Meyer, Y. (2003). Resemblance and dissemblance of Arabidopsis type II peroxiredoxins: similar sequences for divergent gene expression, protein localization, and activity. *Plant Physiol* 132, 2045-2057.
- Broadwater, S.T. (1992). Cytoskeleton and mitotic spindle in red algae. In *Cytoskeleton of the red algae*, D. Menzel, ed. (Boca Raton, CRC press), pp. 93-112.
- Brochier-Armanet, C., Boussau, B., Gribaldo, S., and Forterre, P. (2008). Mesophilic Crenarchaeota: proposal for a third archaeal phylum, the Thaumarchaeota. *Nat Rev Microbiol* 6, 245-252.
- Brochier, C., Forterre, P., and Gribaldo, S. (2005). An emerging phylogenetic core of Archaea: phylogenies of transcription and translation machineries converge following addition of new genome sequences. *BMC Evol Biol* 5, 36.
- Brown, J.R., and Doolittle, W.F. (1997). Archaea and the prokaryote-to-eukaryote transition. *Microbiol Mol Biol Rev* 61, 456-502.
- Budanov, A.V., Sablina, A.A., Feinstein, E., Koonin, E.V., and Chumakov, P.M. (2004). Regeneration of peroxiredoxins by p53-regulated sestrins, homologs of bacterial AhpD. *Science* 304, 596-600.
- Budin, K., and Philippe, H. (1998). New insights into the phylogeny of eukaryotes based on ciliate Hsp70 sequences. *Mol Biol Evol* 15, 943-956.
- Burki, F., Inagaki, Y., Brate, J., Archibald, J.M., Keeling, P.J., Cavalier-Smith, T., Sakaguchi, M., Hashimoto, T., Horak, A., Kumar, S., *et al.* (2009). Large-scale phylogenomic analyses reveal that two enigmatic protist lineages, telonemia and centroheliozoa, are related to photosynthetic chromalveolates. *Genome Biol Evol* 1, 231-238.
- Burki, F., Shalchian-Tabrizi, K., Minge, M., Skjaeveland, A., Nikolaev, S.I., Jakobsen, K.S., and Pawlowski, J. (2007). Phylogenomics reshuffles the eukaryotic supergroups. *PLoS One* 2, e790.
- Burki, F., Shalchian-Tabrizi, K., and Pawlowski, J. (2008). Phylogenomics reveals a new 'megagroup' including most photosynthetic eukaryotes. *Biol Lett* 4, 366-369.
- Capron, A., Okresz, L., and Genschik, P. (2003a). First glance at the plant APC/C, a highly conserved ubiquitin-protein ligase. *Trends Plant Sci* 8, 83-89.
- Capron, A., Serralbo, O., Fulop, K., Frugier, F., Parmentier, Y., Dong, A., Lecureuil, A., Guerche, P., Kondorosi, E., Scheres, B., *et al.* (2003b). The Arabidopsis anaphase-promoting complex or cyclosome: molecular and genetic characterization of the APC2 subunit. *Plant Cell* 15, 2370-2382.
- Cavalier-Smith, T. (1987a). The origin of fungi and pseudofungi. In *Evolutionary biology of the fungi*, A.D.M. Rayner, C.M. Brasier, and C. Moore, eds. (Cambridge University Press), pp. 339-353.
- Cavalier-Smith, T. (1987b). The simultaneous symbiotic origin of mitochondria, chloroplasts, and microbodies. *Ann N Y Acad Sci* 503, 55-71.

- Cavalier-Smith, T. (1999). Principles of protein and lipid targeting in secondary symbiogenesis: euglenoid, dinoflagellate, and sporozoan plastid origins and the eukaryote family tree. *J Eukaryot Microbiol* *46*, 347-366.
- Cavalier-Smith, T. (2002). The phagotrophic origin of eukaryotes and phylogenetic classification of Protozoa. *Int J Syst Evol Microbiol* *52*, 297-354.
- Cavalier-Smith, T. (2003). Genomic reduction and evolution of novel genetic membranes and protein-targeting machinery in eukaryote-eukaryote chimaeras (meta-algae). *Philos Trans R Soc Lond B Biol Sci* *358*, 109-133; discussion 133-104.
- Cavalier-Smith, T., Allsopp, M.T., and Chao, E.E. (1994). Chimeric conundra: are nucleomorphs and chromists monophyletic or polyphyletic? *Proc Natl Acad Sci U S A* *91*, 11368-11372.
- Cavalier-Smith, T., and Chao, E.E. (2006). Phylogeny and megasystematics of phagotrophic heterokonts (kingdom Chromista). *J Mol Evol* *62*, 388-420.
- Cha, M.K., Yun, C.H., and Kim, I.H. (2000). Interaction of human thiol-specific antioxidant protein 1 with erythrocyte plasma membrane. *Biochemistry* *39*, 6944-6950.
- Chapman, R.L., Borkhsenius, O., Brown, R.C., Henk, M.C., and Waters, D.A. (2001). Phragmoplast-mediated cytokinesis in *Trentepohlia*: results of TEM and immunofluorescence cytochemistry. *Int J Syst Evol Microbiol* *51*, 759-765.
- Chapman, R.L., and Henk, C. (1986). Phragmoplast in cytokinesis of *Cephaleuros parasiticus* (Chlorophyta) and a comparison with *C. virescens* vegetative cells. *J Phycol* *22*, 83-88.
- Chatton, E. (1937). *Titres et Travaux Scientifiques (1906–1937)*.
- Chuang, M.H., Wu, M.S., Lo, W.L., Lin, J.T., Wong, C.H., and Chiou, S.H. (2006). The antioxidant protein alkylhydroperoxide reductase of *Helicobacter pylori* switches from a peroxide reductase to a molecular chaperone function. *Proc Natl Acad Sci U S A* *103*, 2552-2557.
- Clute, P., and Pines, J. (1999). Temporal and spatial control of cyclin B1 destruction in metaphase. *Nat Cell Biol* *1*, 82-87.
- Collins, L.J., and Chen, X.S. (2009). Ancestral RNA: the RNA biology of the eukaryotic ancestor. *RNA Biol* *6*, 495-502.
- Cox, C.J., Foster, P.G., Hirt, R.P., Harris, S.R., and Embley, T.M. (2008). The archaeobacterial origin of eukaryotes. *Proc Natl Acad Sci U S A* *105*, 20356-20361.
- Csuros, M., and Miklos, I. (2009). Streamlining and large ancestral genomes in Archaea inferred with a phylogenetic birth-and-death model. *Mol Biol Evol* *26*, 2087-2095.
- Cubonova, L., Sandman, K., Hallam, S.J., Delong, E.F., and Reeve, J.N. (2005). Histones in crenarchaea. *J Bacteriol* *187*, 5482-5485.
- da Fonseca, P.C., Kong, E.H., Zhang, Z., Schreiber, A., Williams, M.A., Morris, E.P., and Barford, D. (2011). Structures of APC/C(Cdh1) with substrates identify Cdh1 and Apc10 as the D-box co-receptor. *Nature* *470*, 274-278.
- Dacks, J.B., Davis, L.A., Sjogren, A.M., Andersson, J.O., Roger, A.J., and Doolittle, W.F. (2003). Evidence for Golgi bodies in proposed 'Golgi-lacking' lineages. *Proc Biol Sci* *270 Suppl 2*, S168-171.
- Dacks, J.B., and Doolittle, W.F. (2001). Reconstructing/deconstructing the earliest eukaryotes: how comparative genomics can help. *Cell* *107*, 419-425.
- Dacks, J.B., and Field, M.C. (2007). Evolution of the eukaryotic membrane-trafficking system: origin, tempo and mode. *J Cell Sci* *120*, 2977-2985.
- Dacks, J.B., Peden, A.A., and Field, M.C. (2009). Evolution of specificity in the eukaryotic endomembrane system. *Int J Biochem Cell Biol* *41*, 330-340.
- Dagan, T., and Martin, W. (2007). Testing hypotheses without considering predictions. *Bioessays* *29*, 500-503.
- Dayhoff, M.O., Schwartz, R.M., and Orcutt, B.C. (1978). A model of evolutionary change in proteins. In: Dayhoff MO, editor. *Atlas of protein sequence and structure. Atlas of protein sequence and structure* *5*, 345–352.
- de Crecy-Lagard, V., Brochier-Armanet, C., Urbonavicius, J., Fernandez, B., Phillips, G., Lyons, B., Noma, A., Alvarez, S., Droogmans, L., Armengaud, J., *et al.* (2010). Biosynthesis of wyosine derivatives in tRNA: an ancient and highly diverse pathway in Archaea. *Mol Biol Evol* *27*, 2062-2077.
- De Duve, C. (1969). Evolution of the peroxisome. *Ann N Y Acad Sci* *168*, 369-381.
- de Duve, C. (2007). The origin of eukaryotes: a reappraisal. *Nat Rev Genet* *8*, 395-403.

- DeLong, E.F. (1992). Archaea in coastal marine environments. *Proc Natl Acad Sci U S A* *89*, 5685-5689.
- Delsuc, F., Brinkmann, H., and Philippe, H. (2005). Phylogenomics and the reconstruction of the tree of life. *Nat Rev Genet* *6*, 361-375.
- Deschamps, P., and Moreira, D. (2009). Signal conflicts in the phylogeny of the primary photosynthetic eukaryotes. *Mol Biol Evol* *26*, 2745-2753.
- Deshaies, R.J., and Joazeiro, C.A. (2009). RING domain E3 ubiquitin ligases. *Annu Rev Biochem* *78*, 399-434.
- Desmond, E., and Gribaldo, S. (2009). Phylogenomics of sterol synthesis: insights into the origin, evolution, and diversity of a key eukaryotic feature. *Genome Biol Evol* *1*, 364-381.
- Dietz, K.J., Horling, F., Konig, J., and Baier, M. (2002). The function of the chloroplast 2-cysteine peroxiredoxin in peroxide detoxification and its regulation. *J Exp Bot* *53*, 1321-1329.
- Diffley, J.F. (2004). Regulation of early events in chromosome replication. *Curr Biol* *14*, R778-786.
- Dreger, M., Bengtsson, L., Schoneberg, T., Otto, H., and Hucho, F. (2001). Nuclear envelope proteomics: novel integral membrane proteins of the inner nuclear membrane. *Proc Natl Acad Sci U S A* *98*, 11943-11948.
- Dunn, C.W., Hejnal, A., Matus, D.Q., Pang, K., Browne, W.E., Smith, S.A., Seaver, E., Rouse, G.W., Obst, M., Edgecombe, G.D., *et al.* (2008). Broad phylogenomic sampling improves resolution of the animal tree of life. *Nature* *452*, 745-749.
- Edgar, R.C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* *32*, 1792-1797.
- Edgell, D.R., and Doolittle, W.F. (1997). Archaea and the origin(s) of DNA replication proteins. *Cell* *89*, 995-998.
- Eggert, U.S., Mitchison, T.J., and Field, C.M. (2006). Animal cytokinesis: from parts list to mechanisms. *Annu Rev Biochem* *75*, 543-566.
- Eisen, J.A. (1998). Phylogenomics: improving functional predictions for uncharacterized genes by evolutionary analysis. *Genome Res* *8*, 163-167.
- Eisen, J.A., Coyne, R.S., Wu, M., Wu, D., Thiagarajan, M., Wortman, J.R., Badger, J.H., Ren, Q., Amedeo, P., Jones, K.M., *et al.* (2006). Macronuclear genome sequence of the ciliate *Tetrahymena thermophila*, a model eukaryote. *PLoS Biol* *4*, e286.
- Eisen, J.A., and Fraser, C.M. (2003). Phylogenomics: intersection of evolution and genomics. *Science* *300*, 1706-1707.
- Eisen, J.A., and Hanawalt, P.C. (1999). A phylogenomic study of DNA repair genes, proteins, and processes. *Mutat Res* *435*, 171-213.
- Elkins, J.G., Podar, M., Graham, D.E., Makarova, K.S., Wolf, Y., Randau, L., Hedlund, B.P., Brochier-Armanet, C., Kunin, V., Anderson, I., *et al.* (2008). A korarchaeal genome reveals insights into the evolution of the Archaea. *Proc Natl Acad Sci U S A* *105*, 8102-8107.
- Eloy, N.B., Coppens, F., Beemster, G.T., Hemerly, A.S., and Ferreira, P.C. (2006). The Arabidopsis anaphase promoting complex (APC): regulation through subunit availability in plant tissues. *Cell Cycle* *5*, 1957-1965.
- Emanuelsson, O., Brunak, S., von Heijne, G., and Nielsen, H. (2007). Locating proteins in the cell using TargetP, SignalP and related tools. *Nat Protoc* *2*, 953-971.
- Emanuelsson, O., Nielsen, H., and von Heijne, G. (1999). ChloroP, a neural network-based method for predicting chloroplast transit peptides and their cleavage sites. *Protein Sci* *8*, 978-984.
- Embley, T.M. (2006). Multiple secondary origins of the anaerobic lifestyle in eukaryotes. *Philos Trans R Soc Lond B Biol Sci* *361*, 1055-1067.
- Errico, A., Claudiani, P., D'Addio, M., and Rugarli, E.I. (2004). Spastin interacts with the centrosomal protein NA14, and is enriched in the spindle pole, the midbody and the distal axon. *Hum Mol Genet* *13*, 2121-2132.
- Esser, C., Ahmadinejad, N., Wiegand, C., Rotte, C., Sebastiani, F., Gelius-Dietrich, G., Henze, K., Kretschmann, E., Richly, E., Leister, D., *et al.* (2004). A genome phylogeny for mitochondria among alpha-proteobacteria and a predominantly eubacterial ancestry of yeast nuclear genes. *Mol Biol Evol* *21*, 1643-1660.
- Fast, N.M., Kissinger, J.C., Roos, D.S., and Keeling, P.J. (2001). Nuclear-encoded, plastid-targeted genes suggest a single common origin for apicomplexan and dinoflagellate plastids. *Mol Biol Evol* *18*, 418-426.
- Fensome, R.A., Saldarriaga, J.F., and Taylor, F.J.R. (1999). Dinoflagellate phylogeny revisited: reconciling morphological and molecular based phylogenies. *Grana* *38*, 66-80.

- Field, M.C., and Dacks, J.B. (2009). First and last ancestors: reconstructing evolution of the endomembrane system with ESCRTs, vesicle coat proteins, and nuclear pore complexes. *Curr Opin Cell Biol* 21, 4-13.
- Flemming, W. (1891). Neue Beitrage zur Kenntnis der Zelle II. *Arch Mikros Anat* 37, 685.
- Foe, I., and Toczyski, D. (2011). Structural biology: a new look for the APC. *Nature* 470, 182-183.
- Forterre, P. (2011). A new fusion hypothesis for the origin of Eukarya: better than previous ones, but probably also wrong. *Res Microbiol* 162, 77-91.
- Forterre, P., Gribaldo, S., and Brochier, C. (2005). [Luca: the last universal common ancestor]. *Med Sci (Paris)* 21, 860-865.
- Foster, P.G., Cox, C.J., and Embley, T.M. (2009). The primary divisions of life: a phylogenomic approach employing composition-heterogeneous methods. *Philos Trans R Soc Lond B Biol Sci* 364, 2197-2207.
- Frankel, J. (2000). Cell biology of Tetrahymena thermophila. *Methods Cell Biol* 62, 27-125.
- Fulop, K., Tarayre, S., Kelemen, Z., Horvath, G., Kevei, Z., Nikovics, K., Bako, L., Brown, S., Kondorosi, A., and Kondorosi, E. (2005). Arabidopsis anaphase-promoting complexes: multiple activators and wide range of substrates might keep APC perpetually busy. *Cell Cycle* 4, 1084-1092.
- Gabalton, T., and Huynen, M.A. (2003). Reconstruction of the proto-mitochondrial metabolism. *Science* 301, 609.
- Gabalton, T., Snel, B., van Zimmeren, F., Hemrika, W., Tabak, H., and Huynen, M.A. (2006). Origin and evolution of the peroxisomal proteome. *Biol Direct* 1, 8.
- Gajadhar, A.A., Marquardt, W.C., Hall, R., Gunderson, J., Ariztia-Carmona, E.V., and Sogin, M.L. (1991). Ribosomal RNA sequences of Sarcocystis muris, Theileria annulata and Cryptocodium cohnii reveal evolutionary relationships among apicomplexans, dinoflagellates, and ciliates. *Mol Biochem Parasitol* 45, 147-154.
- Gaspin, C., Cavaille, J., Erauso, G., and Bachellerie, J.P. (2000). Archaeal homologs of eukaryotic methylation guide small nucleolar RNAs: lessons from the Pyrococcus genomes. *J Mol Biol* 297, 895-906.
- Gassmann, R., Henzing, A.J., and Earnshaw, W.C. (2005). Novel components of human mitotic chromosomes identified by proteomic analysis of the chromosome scaffold fraction. *Chromosoma* 113, 385-397.
- Glenner, H., Hansen, A.J., Sorensen, M.V., Ronquist, F., Huelsenbeck, J.P., and Willerslev, E. (2004). Bayesian inference of the metazoan phylogeny; a combined molecular and morphological approach. *Curr Biol* 14, 1644-1649.
- Glotzer, M. (2001). Animal cell cytokinesis. *Annu Rev Cell Dev Biol* 17, 351-386.
- Gogarten, J.P., Kibak, H., Dittrich, P., Taiz, L., Bowman, E.J., Bowman, B.J., Manolson, M.F., Poole, R.J., Date, T., Oshima, T., et al. (1989). Evolution of the vacuolar H⁺-ATPase: implications for the origin of eukaryotes. *Proc Natl Acad Sci U S A* 86, 6661-6665.
- Golding, G.B., and Gupta, R.S. (1995). Protein-based phylogenies support a chimeric origin for the eukaryotic genome. *Mol Biol Evol* 12, 1-6.
- Gorr, I.H., Boos, D., and Stemmann, O. (2005). Mutual inhibition of separase and Cdk1 by two-step complex formation. *Mol Cell* 19, 135-141.
- Gould, S.B., Waller, R.F., and McFadden, G.I. (2008). Plastid evolution. *Annu Rev Plant Biol* 59, 491-517.
- Gray, M.W., Burger, G., and Lang, B.F. (1999). Mitochondrial evolution. *Science* 283, 1476-1481.
- Gray, M.W., and Doolittle, W.F. (1982). Has the endosymbiont hypothesis been proven? *Microbiol Rev* 46, 1-42.
- Gribaldo, S., and Brochier-Armanet, C. (2006). The origin and evolution of Archaea: a state of the art. *Philos Trans R Soc Lond B Biol Sci* 361, 1007-1022.
- Gribaldo, S., and Brochier, C. (2009). Phylogeny of prokaryotes: does it exist and why should we care? *Res Microbiol* 160, 513-521.
- Gribaldo, S., and Cammarano, P. (1998). The root of the universal tree of life inferred from anciently duplicated genes encoding components of the protein-targeting machinery. *J Mol Evol* 47, 508-516.
- Gribaldo, S., Poole, A.M., Daubin, V., Forterre, P., and Brochier-Armanet, C. (2010). The origin of eukaryotes and their relationship with the Archaea: are we at a phylogenomic impasse? *Nat Rev Microbiol* 8, 743-752.
- Grolig, F. (1992). The cytoskeleton of the Zygnemataceae. . In *In The Cytoskeleton of the Algae*, D. Menzel, ed. (Boca Raton, CRC Press), pp. 165-194. .

- Guertin, D.A., Trautmann, S., and McCollum, D. (2002). Cytokinesis in eukaryotes. *Microbiol Mol Biol Rev* 66, 155-178.
- Guindon, S., and Gascuel, O. (2003). A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst Biol* 52, 696-704.
- Guizetti, J., and Gerlich, D.W. (2010). Cytokinetic abscission in animal cells. *Semin Cell Dev Biol* 21, 909-916.
- Gupta, R.S., and Golding, G.B. (1996). The origin of the eukaryotic cell. *Trends Biochem Sci* 21, 166-171.
- Gustafson, D.E., Jr., Stoecker, D.K., Johnson, M.D., Van Heukelem, W.F., and Sneider, K. (2000). Cryptophyte algae are robbed of their organelles by the marine ciliate *Mesodinium rubrum*. *Nature* 405, 1049-1052.
- Hackett, J.D., Yoon, H.S., Li, S., Reyes-Prieto, A., Rummele, S.E., and Bhattacharya, D. (2007). Phylogenomic analysis supports the monophyly of cryptophytes and haptophytes and the association of rhizaria with chromalveolates. *Mol Biol Evol* 24, 1702-1713.
- Hackett, J.D., Yoon, H.S., Soares, M.B., Bonaldo, M.F., Casavant, T.L., Scheetz, T.E., Nosenko, T., and Bhattacharya, D. (2004). Migration of the plastid genome to the nucleus in a peridinin dinoflagellate. *Curr Biol* 14, 213-218.
- Hall, A., Sankaran, B., Poole, L.B., and Karplus, P.A. (2009). Structural changes common to catalysis in the Tpx peroxiredoxin subfamily. *J Mol Biol* 393, 867-881.
- Hampl, V., Hug, L., Leigh, J.W., Dacks, J.B., Lang, B.F., Simpson, A.G., and Roger, A.J. (2009). Phylogenomic analyses support the monophyly of Excavata and resolve relationships among eukaryotic "supergroups". *Proc Natl Acad Sci U S A* 106, 3859-3864.
- Harper, J.T., Waanders, E., and Keeling, P.J. (2005). On the monophyly of chromalveolates using a six-protein phylogeny of eukaryotes. *Int J Syst Evol Microbiol* 55, 487-496.
- Harper, J.W., Burton, J.L., and Solomon, M.J. (2002). The anaphase-promoting complex: it's not just for mitosis any more. *Genes Dev* 16, 2179-2206.
- Hartung, S., and Hopfner, K.P. (2009). Lessons from structural and biochemical studies on the archaeal exosome. *Biochem Soc Trans* 37, 83-87.
- Hedtke, S.M., Townsend, T.M., and Hillis, D.M. (2006). Resolution of phylogenetic conflict in large data sets by increased taxon sampling. *Syst Biol* 55, 522-529.
- Helmchen, T.A., Bhattacharya, D., and Melkonian, M. (1995). Analyses of ribosomal RNA sequences from glaucocystophyte cyanelles provide new insights into the evolutionary relationships of plastids. *J Mol Evol* 41, 203-210.
- Horiike, T., Hamada, K., Kanaya, S., and Shinozawa, T. (2001). Origin of eukaryotic cell nuclei by symbiosis of Archaea in Bacteria is revealed by homology-hit analysis. *Nat Cell Biol* 3, 210-214.
- Horiike, T., Hamada, K., Miyata, D., and Shinozawa, T. (2004). The origin of eukaryotes is suggested as the symbiosis of pyrococcus into gamma-proteobacteria by phylogenetic tree based on gene content. *J Mol Evol* 59, 606-619.
- Horton, P., Park, K.J., Obayashi, T., Fujita, N., Harada, H., Adams-Collier, C.J., and Nakai, K. (2007). WoLF PSORT: protein localization predictor. *Nucleic Acids Res* 35, W585-587.
- Huber, H., Hohn, M.J., Rachel, R., Fuchs, T., Wimmer, V.C., and Stetter, K.O. (2002). A new phylum of Archaea represented by a nanosized hyperthermophilic symbiont. *Nature* 417, 63-67.
- Hugenholtz, P., Pitulle, C., Hershberger, K.L., and Pace, N.R. (1998). Novel division level bacterial diversity in a Yellowstone hot spring. *J Bacteriol* 180, 366-376.
- Hughes, T., and Rusten, T.E. (2007). Origin and evolution of self-consumption: autophagy. *Adv Exp Med Biol* 607, 111-118.
- Hutchins, J.R., Toyoda, Y., Hegemann, B., Poser, I., Heriche, J.K., Sykora, M.M., Augsburg, M., Hudecz, O., Buschhorn, B.A., Bulkescher, J., *et al.* (2010). Systematic analysis of human protein complexes identifies chromosome segregation proteins. *Science* 328, 593-599.
- Inagaki, Y., Dacks, J.B., Doolittle, W.F., Watanabe, K.I., and Ohama, T. (2000). Evolutionary relationship between dinoflagellates bearing obligate diatom endosymbionts: insight into tertiary endosymbiosis. *Int J Syst Evol Microbiol* 50 Pt 6, 2075-2081.
- Islam, T., Jensen, S., Reigstad, L.J., Larsen, O., and Birkeland, N.K. (2008). Methane oxidation at 55 degrees C and pH 2 by a thermoacidophilic bacterium belonging to the Verrucomicrobia phylum. *Proc Natl Acad Sci U S A* 105, 300-304.

- Iwabe, N., Kuma, K., Hasegawa, M., Osawa, S., and Miyata, T. (1989). Evolutionary relationship of archaeobacteria, eubacteria, and eukaryotes inferred from phylogenetic trees of duplicated genes. *Proc Natl Acad Sci U S A* *86*, 9355-9359.
- Jallepalli, P.V., Waizenegger, I.C., Bunz, F., Langer, S., Speicher, M.R., Peters, J.M., Kinzler, K.W., Vogelstein, B., and Lengauer, C. (2001). Securin is required for chromosomal stability in human cells. *Cell* *105*, 445-457.
- Jang, H.H., Kim, S.Y., Park, S.K., Jeon, H.S., Lee, Y.M., Jung, J.H., Lee, S.Y., Chae, H.B., Jung, Y.J., Lee, K.O., *et al.* (2006). Phosphorylation and concomitant structural changes in human 2-Cys peroxiredoxin isotype I differentially regulate its peroxidase and molecular chaperone functions. *FEBS Lett* *580*, 351-355.
- Jang, H.H., Lee, K.O., Chi, Y.H., Jung, B.G., Park, S.K., Park, J.H., Lee, J.R., Lee, S.S., Moon, J.C., Yun, J.W., *et al.* (2004). Two enzymes in one; two yeast peroxiredoxins display oxidative stress-dependent switching from a peroxidase to a molecular chaperone function. *Cell* *117*, 625-635.
- Jeffrey, P.D., Russo, A.A., Polyak, K., Gibbs, E., Hurwitz, J., Massague, J., and Pavletich, N.P. (1995). Mechanism of CDK activation revealed by the structure of a cyclinA-CDK2 complex. *Nature* *376*, 313-320.
- Jekely, G., and Arendt, D. (2006). Evolution of intraflagellar transport from coated vesicles and autogenous origin of the eukaryotic cilium. *Bioessays* *28*, 191-198.
- Jobb, G., von Haeseler, A., and Strimmer, K. (2004). TREEFINDER: a powerful graphical analysis environment for molecular phylogenetics. *BMC Evol Biol* *4*, 18.
- Jonsson, T.J., Johnson, L.C., and Lowther, W.T. (2008). Structure of the sulphiredoxin-peroxiredoxin complex reveals an essential repair embrace. *Nature* *451*, 98-101.
- Jurgens, G. (2005). Plant cytokinesis: fission by fusion. *Trends Cell Biol* *15*, 277-283.
- Jurica, M.S., and Moore, M.J. (2003). Pre-mRNA splicing: awash in a sea of proteins. *Mol Cell* *12*, 5-14.
- Katoh, K., Misawa, K., Kuma, K., and Miyata, T. (2002). MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res* *30*, 3059-3066.
- Keeling, P.J. (2004). Diversity and evolutionary history of plastids and their hosts. *Am J Bot* *91*, 1481-1493.
- Keeling, P.J. (2009). Chromalveolates and the evolution of plastids by secondary endosymbiosis. *J Eukaryot Microbiol* *56*, 1-8.
- Keeling, P.J., Burger, G., Durnford, D.G., Lang, B.F., Lee, R.W., Pearlman, R.E., Roger, A.J., and Gray, M.W. (2005). The tree of eukaryotes. *Trends Ecol Evol* *20*, 670-676.
- Keeling, P.J., and Palmer, J.D. (2008). Horizontal gene transfer in eukaryotic evolution. *Nat Rev Genet* *9*, 605-618.
- Kelly, S., Wickstead, B., and Gull, K. (2011). Archaeal phylogenomics provides evidence in support of a methanogenic origin of the Archaea and a thaumarchaeal origin for the eukaryotes. *Proc Biol Sci* *278*, 1009-1018.
- Kies, L. (1974). Elektronenmikroskopische Untersuchungen an *Paulinella chromatophora* Lauterborn, einer Thekamöbe mit blaugrünen Endosymbionten (Cyanellen). *Protoplasma* *80*, 69-89.
- Konig, J., Baier, M., Horling, F., Kahmann, U., Harris, G., Schurmann, P., and Dietz, K.J. (2002). The plant-specific function of 2-Cys peroxiredoxin-mediated detoxification of peroxides in the redox-hierarchy of photosynthetic electron flux. *Proc Natl Acad Sci U S A* *99*, 5738-5743.
- Koo, K.H., Lee, S., Jeong, S.Y., Kim, E.T., Kim, H.J., Kim, K., Song, K., and Chae, H.Z. (2002). Regulation of thioredoxin peroxidase activity by C-terminal truncation. *Arch Biochem Biophys* *397*, 312-318.
- Koonin, E.V. (2010). The origin and early evolution of eukaryotes in the light of phylogenomics. *Genome Biol* *11*, 209.
- Koumandou, V.L., Dacks, J.B., Coulson, R.M., and Field, M.C. (2007). Control systems for membrane fusion in the ancestral eukaryote; evolution of tethering complexes and SM proteins. *BMC Evol Biol* *7*, 29.
- Kozina, I.V., Kublanov, I.V., Kolganova, T.V., Chernyh, N.A., and Bonch-Osmolovskaya, E.A. (2010). *Caldanaerobacter uzonensis* sp. nov., an anaerobic, thermophilic, heterotrophic bacterium isolated from a hot spring. *Int J Syst Evol Microbiol* *60*, 1372-1375.
- Kraft, C., Herzog, F., Gieffers, C., Mechtler, K., Hagting, A., Pines, J., and Peters, J.M. (2003). Mitotic regulation of the human anaphase-promoting complex by phosphorylation. *EMBO J* *22*, 6598-6609.
- Kramer, E.R., Scheuringer, N., Podtelejnikov, A.V., Mann, M., and Peters, J.M. (2000). Mitotic regulation of the APC activator proteins CDC20 and CDH1. *Mol Biol Cell* *11*, 1555-1569.

- Kublanov, I.V., Bidjjeva, S., Mardanov, A.V., and Bonch-Osmolovskaya, E.A. (2009). *Desulfurococcus kamchatkensis* sp. nov., a novel hyperthermophilic protein-degrading archaeon isolated from a Kamchatka hot spring. *Int J Syst Evol Microbiol* 59, 1743-1747.
- Kumar, A., Agarwal, S., Heyman, J.A., Matson, S., Heidtman, M., Piccirillo, S., Umansky, L., Drawid, A., Jansen, R., Liu, Y., *et al.* (2002). Subcellular localization of the yeast proteome. *Genes Dev* 16, 707-719.
- Kumar, P., and Wang, C.C. (2005). Depletion of anaphase-promoting complex or cyclosome (APC/C) subunit homolog APC1 or CDC27 of *Trypanosoma brucei* arrests the procyclic form in metaphase but the bloodstream form in anaphase. *J Biol Chem* 280, 31783-31791.
- Kumsta, C., and Jakob, U. (2009). Redox-regulated chaperones. *Biochemistry* 48, 4666-4676.
- Kurland, C.G., Canback, B., and Berg, O.G. (2003). Horizontal gene transfer: a critical view. *Proc Natl Acad Sci U S A* 100, 9658-9662.
- LaCount, D.J., Barrett, B., and Donelson, J.E. (2002). *Trypanosoma brucei* FLA1 is required for flagellum attachment and cytokinesis. *J Biol Chem* 277, 17580-17588.
- Lake, J.A. (1988). Origin of the eukaryotic nucleus determined by rate-invariant analysis of rRNA sequences. *Nature* 331, 184-186.
- Lake, J.A., Henderson, E., Oakes, M., and Clark, M.W. (1984). Eocytes: a new ribosome structure indicates a kingdom with a close relationship to eukaryotes. *Proc Natl Acad Sci U S A* 81, 3786-3790.
- Lake, J.A., and Rivera, M.C. (1994). Was the nucleus the first endosymbiont? *Proc Natl Acad Sci U S A* 91, 2880-2881.
- Lakin-Thomas, P.L. (2006). New models for circadian systems in microorganisms. *FEMS Microbiol Lett* 259, 1-6.
- Larkin, M.A., Blackshields, G., Brown, N.P., Chenna, R., McGettigan, P.A., McWilliam, H., Valentin, F., Wallace, I.M., Wilm, A., Lopez, R., *et al.* (2007). Clustal W and Clustal X version 2.0. *Bioinformatics* 23, 2947-2948.
- Lartillot, N., Brinkmann, H., and Philippe, H. (2007). Suppression of long-branch attraction artefacts in the animal phylogeny using a site-heterogeneous model. *BMC Evol Biol* 7 *Suppl* 1, S4.
- Lartillot, N., Lepage, T., and Blanquart, S. (2009). PhyloBayes 3: a Bayesian software package for phylogenetic reconstruction and molecular dating. *Bioinformatics* 25, 2286-2288.
- Lartillot, N., and Philippe, H. (2004). A Bayesian mixture model for across-site heterogeneities in the amino-acid replacement process. *Mol Biol Evol* 21, 1095-1109.
- Latifi, A., Ruiz, M., and Zhang, C.C. (2009). Oxidative stress in cyanobacteria. *FEMS Microbiol Rev* 33, 258-278.
- Lauterborn, R. (1895). Protozoenstudien II. *Paulinella chromatophora* nov. gen., nov. spec., ein beschalter Rhizopode des Süßwassers mit blaugrünen chromatophorenartigen Einschlüssen. *Z Wiss Zool* 59, 537-544.
- Leander, B.S., Esson, H.J., and Breglia, S.A. (2007). Macroevolution of complex cytoskeletal systems in euglenids. *Bioessays* 29, 987-1000.
- Lecointre, G., and Le Guyader, H. *Classification phylogénétique du vivant* 3e édition., Editions Belin, 2001 edn.
- Lecompte, O., Ripp, R., Thierry, J.C., Moras, D., and Poch, O. (2002). Comparative analysis of ribosomal proteins in complete genomes: an example of reductive evolution at the domain scale. *Nucleic Acids Res* 30, 5382-5390.
- Lee, K.K., Prochasson, P., Florens, L., Swanson, S.K., Washburn, M.P., and Workman, J.L. (2004). Proteomic analysis of chromatin-modifying complexes in *Saccharomyces cerevisiae* identifies novel subunits. *Biochem Soc Trans* 32, 899-903.
- Leipe, D.D., Aravind, L., and Koonin, E.V. (1999). Did DNA replication evolve twice independently? *Nucleic Acids Res* 27, 3389-3401.
- Leung, K.F., Dacks, J.B., and Field, M.C. (2008). Evolution of the multivesicular body ESCRT machinery; retention across the eukaryotic lineage. *Traffic* 9, 1698-1716.
- Li, J.B., Gerdes, J.M., Haycraft, C.J., Fan, Y., Teslovich, T.M., May-Simera, H., Li, H., Blacque, O.E., Li, L., Leitch, C.C., *et al.* (2004). Comparative genomics identifies a flagellar and basal body proteome that includes the BBS5 human disease gene. *Cell* 117, 541-552.
- Lim, J.C., Choi, H.I., Park, Y.S., Nam, H.W., Woo, H.A., Kwon, K.S., Kim, Y.S., Rhee, S.G., Kim, K., and Chae, H.Z. (2008). Irreversible oxidation of the active-site cysteine of peroxiredoxin to cysteine sulfonic acid for enhanced molecular chaperone activity. *J Biol Chem* 283, 28873-28880.

- Lima Mde, F., Eloy, N.B., Pegoraro, C., Sagit, R., Rojas, C., Bretz, T., Vargas, L., Eloffsson, A., de Oliveira, A.C., Hemerly, A.S., *et al.* (2010). Genomic evolution and complexity of the Anaphase-promoting Complex (APC) in land plants. *BMC Plant Biol* 10, 254.
- Lindas, A.C., Karlsson, E.A., Lindgren, M.T., Ettema, T.J., and Bernander, R. (2008). A unique cell division machinery in the Archaea. *Proc Natl Acad Sci U S A* 105, 18942-18946.
- Lombard, J., and Moreira, D. (2011). Origins and early evolution of the mevalonate pathway of isoprenoid biosynthesis in the three domains of life. *Mol Biol Evol* 28, 87-99.
- Londei, P. (2005). Evolution of translational initiation: new insights from the archaea. *FEMS Microbiol Rev* 29, 185-200.
- Lopez-Bautista, J.M., Waters, D.A., and Chapman, R.L. (2003). Phragmoplastin, green algae and the evolution of cytokinesis. *Int J Syst Evol Microbiol* 53, 1715-1718.
- Lopez-Garcia, P., Brochier, C., Moreira, D., and Rodriguez-Valera, F. (2004). Comparative analysis of a genome fragment of an uncultivated mesopelagic crenarchaeote reveals multiple horizontal gene transfers. *Environ Microbiol* 6, 19-34.
- Lopez-Garcia, P., and Moreira, D. (2006). Selective forces for the origin of the eukaryotic nucleus. *Bioessays* 28, 525-533.
- Lowther, W.T., and Haynes, A.C. (2010). Reduction of Cysteine Sulfinic Acid in Eukaryotic, Typical 2-Cys Peroxiredoxins by Sulfiredoxin. *Antioxid Redox Signal*.
- Ludwig, M., and Gibbs, S.P. (1987). Are the nucleomorphs of cryptomonads and Chlorarachnion the vestigial nuclei of eukaryotic endosymbionts. *Ann N Y Acad Sci* 503, 198-211.
- Makarova, K.S., Wolf, Y.I., Mekhedov, S.L., Mirkin, B.G., and Koonin, E.V. (2005). Ancestral paralogs and pseudoparalogs and their role in the emergence of the eukaryotic cell. *Nucleic Acids Res* 33, 4626-4638.
- Manchado, E., Eguren, M., and Malumbres, M. (2010). The anaphase-promoting complex/cyclosome (APC/C): cell-cycle-dependent and -independent functions. *Biochem Soc Trans* 38, 65-71.
- Mans, B.J., Anantharaman, V., Aravind, L., and Koonin, E.V. (2004). Comparative genomics, evolution and origins of the nuclear envelope and nuclear pore complex. *Cell Cycle* 3, 1612-1637.
- Marchant, H.J., and Pickett-Heaps, J.D. (1973). Mitosis and cytokinesis in *Coleochaete scutata*. *J Phycol* 9, 461-471.
- Mardanov, A.V., Gumerov, V.M., Beletsky, A.V., Bonch-Osmolovskaya, E.A., Ravin, N.V., and Skryabin, K.G. (2010). Characteristic of biodiversity of thermophilic microbial community by parallel pyrosequencing method. *Dokl Biochem Biophys* 432, 110-113.
- Margulis, L., ed. (1970). *Origin of Eukaryotic Cells* (Yale University Press).
- Margulis, L. (1996). Archaeal-eubacterial mergers in the origin of Eukarya: phylogenetic classification of life. *Proc Natl Acad Sci U S A* 93, 1071-1076.
- Margulis, L., and Bermudes, D. (1985). Symbiosis as a mechanism of evolution: status of cell symbiosis theory. *Symbiosis* 1, 101-124.
- Margulis, L., Chapman, M., Guerrero, R., and Hall, J. (2006). The last eukaryotic common ancestor (LECA): acquisition of cytoskeletal motility from aerotolerant spirochetes in the Proterozoic Eon. *Proc Natl Acad Sci U S A* 103, 13080-13085.
- Margulis, L., Dolan, M.F., and Guerrero, R. (2000). The chimeric eukaryote: origin of the nucleus from the karyomastigont in amitochondriate protists. *Proc Natl Acad Sci U S A* 97, 6954-6959.
- Marin, B., Nowack, E.C., and Melkonian, M. (2005). A plastid in the making: evidence for a second primary endosymbiosis. *Protist* 156, 425-432.
- Martin-Cuadrado, A.B., Rodriguez-Valera, F., Moreira, D., Alba, J.C., Ivars-Martinez, E., Henn, M.R., Talla, E., and Lopez-Garcia, P. (2008). Hindsight in the relative abundance, metabolic potential and genome dynamics of uncultivated marine archaea from comparative metagenomic analyses of bathypelagic plankton of different oceanic regions. *ISME J* 2, 865-886.
- Martin, W. (2005). Archaeobacteria (Archaea) and the origin of the eukaryotic nucleus. *Curr Opin Microbiol* 8, 630-637.
- Martin, W., Hoffmeister, M., Rotte, C., and Henze, K. (2001). An overview of endosymbiotic models for the origins of eukaryotes, their ATP-producing organelles (mitochondria and hydrogenosomes), and their heterotrophic lifestyle. *Biol Chem* 382, 1521-1539.

- Martin, W., and Muller, M. (1998). The hydrogen hypothesis for the first eukaryote. *Nature* 392, 37-41.
- Martin, W., Rujan, T., Richly, E., Hansen, A., Cornelsen, S., Lins, T., Leister, D., Stoebe, B., Hasegawa, M., and Penny, D. (2002). Evolutionary analysis of *Arabidopsis*, cyanobacterial, and chloroplast genomes reveals plastid phylogeny and thousands of cyanobacterial genes in the nucleus. *Proc Natl Acad Sci U S A* 99, 12246-12251.
- Martin, W., Stoebe, B., Goremykin, V., Hapsmann, S., Hasegawa, M., and Kowallik, K.V. (1998). Gene transfer to the nucleus and the evolution of chloroplasts. *Nature* 393, 162-165.
- Matyskiela, M.E., Rodrigo-Brenni, M.C., and Morgan, D.O. (2009). Mechanisms of ubiquitin transfer by the anaphase-promoting complex. *J Biol* 8, 92.
- McFadden, G. (2001). Primary and secondary endosymbiosis and the origin of plastids. *J Phycol* 34, 951-959.
- Mereschkowsky, C. (1905). Über Natur und Ursprung der Chromatophoren im Pflanzenreiche. *Biol Centralbl* 25, 593-604.
- Minge, M.A., Silberman, J.D., Orr, R.J., Cavalier-Smith, T., Shalchian-Tabrizi, K., Burki, F., Skjaeveland, A., and Jakobsen, K.S. (2009). Evolutionary position of breviate amoebae and the primary eukaryote divergence. *Proc Biol Sci* 276, 597-604.
- Mitchell, D.R. (2007). The evolution of eukaryotic cilia and flagella as motile and sensory organelles. *Adv Exp Med Biol* 607, 130-140.
- Mondal, S., Burgute, B., Rieger, D., Muller, R., Rivero, F., Faix, J., Schleicher, M., and Noegel, A.A. (2010). Regulation of the actin cytoskeleton by an interaction of IQGAP related protein GAPA with filamin and cortexillin I. *PLoS One* 5, e15440.
- Moon, J.C., Hah, Y.S., Kim, W.Y., Jung, B.G., Jang, H.H., Lee, J.R., Kim, S.Y., Lee, Y.M., Jeon, M.G., Kim, C.W., *et al.* (2005). Oxidative stress-dependent structural and functional switching of a human 2-Cys peroxiredoxin isotype II that enhances HeLa cell resistance to H₂O₂-induced cell death. *J Biol Chem* 280, 28775-28784.
- Moreira, D., Le Guyader, H., and Philippe, H. (2000). The origin of red algae and the evolution of chloroplasts. *Nature* 405, 69-72.
- Moreira, D., and Lopez-Garcia, P. (1998). Symbiosis between methanogenic archaea and delta-proteobacteria as the origin of eukaryotes: the syntrophic hypothesis. *J Mol Evol* 47, 517-530.
- Moreira, D., von der Heyden, S., Bass, D., Lopez-Garcia, P., Chao, E., and Cavalier-Smith, T. (2007). Global eukaryote phylogeny: Combined small- and large-subunit ribosomal DNA trees support monophyly of Rhizaria, Retaria and Excavata. *Mol Phylogenet Evol* 44, 255-266.
- Mu, Z.M., Yin, X.Y., and Prochownik, E.V. (2002). Pag, a putative tumor suppressor, interacts with the Myc Box II domain of c-Myc and selectively alters its biological function and target gene expression. *J Biol Chem* 277, 43175-43184.
- Nakayama, T., and Ishida, K. (2009). Another acquisition of a primary photosynthetic organelle is underway in *Paulinella chromatophora*. *Curr Biol* 19, R284-285.
- Nasmyth, K., and Haering, C.H. (2009). Cohesin: its roles and mechanisms. *Annu Rev Genet* 43, 525-558.
- Nelissen, B., Van de Peer, Y., Wilmotte, A., and De Wachter, R. (1995). An early origin of plastids within the cyanobacterial divergence is suggested by evolutionary trees based on complete 16S rRNA sequences. *Mol Biol Evol* 12, 1166-1173.
- Neujahr, R., Albrecht, R., Kohler, J., Matzner, M., Schwartz, J.M., Westphal, M., and Gerisch, G. (1998). Microtubule-mediated centrosome motility and the positioning of cleavage furrows in multinucleate myosin II-null cells. *J Cell Sci* 111 (Pt 9), 1227-1240.
- Nikolaev, S.I., Berney, C., Fahrni, J.F., Bolivar, I., Polet, S., Mylnikov, A.P., Aleshin, V.V., Petrov, N.B., and Pawlowski, J. (2004). The twilight of Heliozoa and rise of Rhizaria, an emerging supergroup of amoeboid eukaryotes. *Proc Natl Acad Sci U S A* 101, 8066-8071.
- Notredame, C., Higgins, D.G., and Heringa, J. (2000). T-Coffee: A novel method for fast and accurate multiple sequence alignment. *J Mol Biol* 302, 205-217.
- Nowack, E.C., Melkonian, M., and Glockner, G. (2008). Chromatophore genome sequence of *Paulinella* sheds light on acquisition of photosynthesis by eukaryotes. *Curr Biol* 18, 410-418.
- Nozaki, H., Iseki, M., Hasegawa, M., Misawa, K., Nakada, T., Sasaki, N., and Watanabe, M. (2007). Phylogeny of primary photosynthetic eukaryotes as deduced from slowly evolving nuclear genes. *Mol Biol Evol* 24, 1592-1595.

- Nozaki, H., Matsuzaki, M., Takahara, M., Misumi, O., Kuroiwa, H., Hasegawa, M., Shin-i, T., Kohara, Y., Ogasawara, N., and Kuroiwa, T. (2003). The phylogenetic position of red algae revealed by multiple nuclear genes from mitochondria-containing eukaryotes and an alternative hypothesis on the origin of plastids. *J Mol Evol* 56, 485-497.
- Nunoura, T., Takaki, Y., Kakuta, J., Nishi, S., Sugahara, J., Kazama, H., Chee, G.J., Hattori, M., Kanai, A., Atomi, H., *et al.* (2010). Insights into the evolution of Archaea and eukaryotic protein modifier systems revealed by the genome of a novel archaeal group. *Nucleic Acids Res.*
- O'Neill, J.S., and Reddy, A.B. (2011). Circadian clocks in human red blood cells. *Nature* 469, 498-503.
- O'Neill, J.S., van Ooijen, G., Dixon, L.E., Troein, C., Corellou, F., Bouget, F.Y., Reddy, A.B., and Millar, A.J. (2011). Circadian rhythms persist without transcription in a eukaryote. *Nature* 469, 554-558.
- Obuse, C., Yang, H., Nozaki, N., Goto, S., Okazaki, T., and Yoda, K. (2004). Proteomics analysis of the centromere complex from HeLa interphase cells: UV-damaged DNA binding protein 1 (DDB-1) is a component of the CEN-complex, while BMI-1 is transiently co-localized with the centromeric region in interphase. *Genes Cells* 9, 105-120.
- Ohi, M.D., Feoktistova, A., Ren, L., Yip, C., Cheng, Y., Chen, J.S., Yoon, H.J., Wall, J.S., Huang, Z., Penczek, P.A., *et al.* (2007). Structural organization of the anaphase-promoting complex bound to the mitotic activator Slp1. *Mol Cell* 28, 871-885.
- Omer, A.D., Lowe, T.M., Russell, A.G., Ebhardt, H., Eddy, S.R., and Dennis, P.P. (2000). Homologs of small nucleolar RNAs in Archaea. *Science* 288, 517-522.
- Onn, I., Heidinger-Pauli, J.M., Guacci, V., Unal, E., and Koshland, D.E. (2008). Sister chromatid cohesion: a simple concept with a complex reality. *Annu Rev Cell Dev Biol* 24, 105-129.
- Otegui, M.S., Verbrugghe, K.J., and Skop, A.R. (2005). Midbodies and phragmoplasts: analogous structures involved in cytokinesis. *Trends Cell Biol* 15, 404-413.
- Palmer, J.D. (2003). The symbiotic birth and spread of plastids: how many times and whodunit? *J Phycol* 39, 4-12.
- Parfrey, L.W., Barbero, E., Lasser, E., Dunthorn, M., Bhattacharya, D., Patterson, D.J., and Katz, L.A. (2006). Evaluating support for the current classification of eukaryotic diversity. *PLoS Genet* 2, e220.
- Parfrey, L.W., Grant, J., Tekle, Y.I., Lasek-Nesselquist, E., Morrison, H.G., Sogin, M.L., Patterson, D.J., and Katz, L.A. (2010). Broadly sampled multigene analyses yield a well-resolved eukaryotic tree of life. *Syst Biol* 59, 518-533.
- Park, S.G., Cha, M.K., Jeong, W., and Kim, I.H. (2000). Distinct physiological functions of thiol peroxidase isoenzymes in *Saccharomyces cerevisiae*. *J Biol Chem* 275, 5723-5732.
- Parsonage, D., Youngblood, D.S., Sarma, G.N., Wood, Z.A., Karplus, P.A., and Poole, L.B. (2005). Analysis of the link between enzymatic activity and oligomeric state in AhpC, a bacterial peroxiredoxin. *Biochemistry* 44, 10583-10592.
- Passmore, L.A., Booth, C.R., Venien-Bryan, C., Ludtke, S.J., Fioretto, C., Johnson, L.N., Chiu, W., and Barford, D. (2005). Structural analysis of the anaphase-promoting complex reveals multiple active sites and insights into polyubiquitylation. *Mol Cell* 20, 855-866.
- Passmore, L.A., McCormack, E.A., Au, S.W., Paul, A., Willison, K.R., Harper, J.W., and Barford, D. (2003). Doc1 mediates the activity of the anaphase-promoting complex by contributing to substrate recognition. *EMBO J* 22, 786-796.
- Patron, N.J., Inagaki, Y., and Keeling, P.J. (2007). Multiple gene phylogenies support the monophyly of cryptomonad and haptophyte host lineages. *Curr Biol* 17, 887-891.
- Patron, N.J., Rogers, M.B., and Keeling, P.J. (2004). Gene replacement of fructose-1,6-bisphosphate aldolase supports the hypothesis of a single photosynthetic ancestor of chromalveolates. *Eukaryot Cell* 3, 1169-1175.
- Patterson, D.J. (1999). The Diversity of Eukaryotes. *Am Nat* 154, S96-S124.
- Patterson, D.J., and Sogin, M.L., eds. (1992). *Eukaryote origins and protistan diversity.*, Hartman H.
- Matsuno, K. edn (World Scientific: Singapore).
- Pawlowski, J., Bolivar, I., Fahrni, J.F., Cavalier-Smith, T., and Gouy, M. (1996). Early origin of foraminifera suggested by SSU rRNA gene sequences. *Mol Biol Evol* 13, 445-450.
- Pawlowski, J., Bolivar, I., Fahrni, J.F., De Vargas, C., and Bowser, S.S. (1999). Molecular evidence that *Reticulomyxa filosa* is a freshwater naked foraminifer. *J Eukaryot Microbiol* 46, 612-617.

- Pazour, G.J. (2004). Comparative genomics: prediction of the ciliary and basal body proteome. *Curr Biol* *14*, R575-577.
- Perevalova, A.A., Kolganova, T.V., Birkeland, N.K., Schleper, C., Bonch-Osmolovskaya, E.A., and Lebedinsky, A.V. (2008). Distribution of Crenarchaeota representatives in terrestrial hot springs of Russia and Iceland. *Appl Environ Microbiol* *74*, 7620-7628.
- Peters, J.M. (2006). The anaphase promoting complex/cyclosome: a machine designed to destroy. *Nat Rev Mol Cell Biol* *7*, 644-656.
- Petersen, J., Teich, R., Brinkmann, H., and Cerff, R. (2006). A "green" phosphoribulokinase in complex algae with red plastids: evidence for a single secondary endosymbiosis leading to haptophytes, cryptophytes, heterokonts, and dinoflagellates. *J Mol Evol* *62*, 143-157.
- Petroski, M.D., and Deshaies, R.J. (2005). Function and regulation of cullin-RING ubiquitin ligases. *Nat Rev Mol Cell Biol* *6*, 9-20.
- Pfleghaar, K., Heubes, S., Cox, J., Stemmann, O., and Speicher, M.R. (2005). Securin is not required for chromosomal stability in human cells. *PLoS Biol* *3*, e416.
- Philippe, H. (1993). MUST, a computer package of Management Utilities for Sequences and Trees. *Nucleic Acids Res* *21*, 5264-5272.
- Philippe, H. (2000). Opinion: long branch attraction and protist phylogeny. *Protist* *151*, 307-316.
- Philippe, H., Derelle, R., Lopez, P., Pick, K., Borchiellini, C., Boury-Esnault, N., Vacelet, J., Renard, E., Houliston, E., Queinnee, E., *et al.* (2009). Phylogenomics revives traditional views on deep animal relationships. *Curr Biol* *19*, 706-712.
- Philippe, H., and Douzery, E. (1994). The pitfalls of molecular phylogeny based on four species as illustrated by the Cetacea/Artiodactyla relationships. *Journal of Mammalian Evolution* *2*, 133-152.
- Philippe, H., and Germot, A. (2000). Phylogeny of eukaryotes based on ribosomal RNA: long-branch attraction and models of sequence evolution. *Mol Biol Evol* *17*, 830-834.
- Philippe, H., Germot, A., and Moreira, D. (2000a). The new phylogeny of eukaryotes. *Curr Opin Genet Dev* *10*, 596-601.
- Philippe, H., Lartillot, N., and Brinkmann, H. (2005). Multigene analyses of bilaterian animals corroborate the monophyly of Ecdysozoa, Lophotrochozoa, and Protostomia. *Mol Biol Evol* *22*, 1246-1253.
- Philippe, H., Lopez, P., Brinkmann, H., Budin, K., Germot, A., Laurent, J., Moreira, D., Muller, M., and Le Guyader, H. (2000b). Early-branching or fast-evolving eukaryotes? An answer based on slowly evolving positions. *Proc Biol Sci* *267*, 1213-1221.
- Philippe, H., Snell, E.A., Baptiste, E., Lopez, P., Holland, P.W., and Casane, D. (2004). Phylogenomics of eukaryotes: impact of missing data on large alignments. *Mol Biol Evol* *21*, 1740-1752.
- Pickart, C.M. (2001). Mechanisms underlying ubiquitination. *Annu Rev Biochem* *70*, 503-533.
- Pickett-Heaps, J.D. (1967). Ultrastructure and differentiation in *Chara* (fibrosa). *Aust J Biol Sci* *20*, 883-894.
- Poole, A.M., and Neumann, N. (2011). Reconciling an archaeal origin of eukaryotes with engulfment: a biologically plausible update of the Eocyte hypothesis. *Res Microbiol* *162*, 71-76.
- Poole, A.M., and Penny, D. (2007). Evaluating hypotheses for the origin of eukaryotes. *Bioessays* *29*, 74-84.
- Poole, L.B., Reynolds, C.M., Wood, Z.A., Karplus, P.A., Ellis, H.R., and Li Calzi, M. (2000). AhpF and other NADH:peroxiredoxin oxidoreductases, homologues of low Mr thioredoxin reductase. *Eur J Biochem* *267*, 6126-6133.
- Rabilloud, T., Heller, M., Rigobello, M.P., Bindoli, A., Aebersold, R., and Lunardi, J. (2001). The mitochondrial antioxidant defence system and its response to oxidative stress. *Proteomics* *1*, 1105-1110.
- Raff, M.C. (1996). Size control: the regulation of cell numbers in animal development. *Cell* *86*, 173-175.
- Ragan, M.A., Goggin, C.L., Cawthorn, R.J., Cerenius, L., Jamieson, A.V., Plourde, S.M., Rand, T.G., Soderhall, K., and Gutell, R.R. (1996). A novel clade of protistan parasites near the animal-fungal divergence. *Proc Natl Acad Sci U S A* *93*, 11907-11912.
- Ramesh, M.A., Malik, S.B., and Logsdon, J.M., Jr. (2005). A phylogenomic inventory of meiotic genes; evidence for sex in *Giardia* and an early eukaryotic origin of meiosis. *Curr Biol* *15*, 185-191.

- Rappsilber, J., Ryder, U., Lamond, A.I., and Mann, M. (2002). Large-scale proteomic analysis of the human spliceosome. *Genome Res* 12, 1231-1245.
- Reeve, J.N., Sandman, K., and Daniels, C.J. (1997). Archaeal histones, nucleosomes, and transcription initiation. *Cell* 89, 999-1002.
- Reigstad, L.J., Jorgensen, S.L., and Schleper, C. (2010). Diversity and abundance of Korarchaeota in terrestrial hot springs of Iceland and Kamchatka. *ISME J* 4, 346-356.
- Reigstad, L.J., Richter, A., Daims, H., Urich, T., Schwark, L., and Schleper, C. (2008). Nitrification in terrestrial hot springs of Iceland and Kamchatka. *FEMS Microbiol Ecol* 64, 167-174.
- Rey, P., Becuwe, N., Barrault, M.B., Rumeau, D., Havaux, M., Biteau, B., and Toledano, M.B. (2007). The *Arabidopsis thaliana* sulfiredoxin is a plastidic cysteine-sulfinic acid reductase involved in the photooxidative stress response. *Plant J* 49, 505-514.
- Reyes-Prieto, A., and Bhattacharya, D. (2007). Phylogeny of Calvin cycle enzymes supports Plantae monophyly. *Mol Phylogenet Evol* 45, 384-391.
- Reyes-Prieto, A., Weber, A.P., and Bhattacharya, D. (2007). The origin and establishment of the plastid in algae and plants. *Annu Rev Genet* 41, 147-168.
- Reysenbach, A.L., Ehringer, M., and Hershberger, K. (2000). Microbial diversity at 83 degrees C in Calcite Springs, Yellowstone National Park: another environment where the Aquificales and "Korarchaeota" coexist. *Extremophiles* 4, 61-67.
- Richards, T.A., and Cavalier-Smith, T. (2005). Myosin domain evolution and the primary divergence of eukaryotes. *Nature* 436, 1113-1118.
- Richards, T.A., Dacks, J.B., Campbell, S.A., Blanchard, J.L., Foster, P.G., McLeod, R., and Roberts, C.W. (2006). Evolutionary origins of the eukaryotic shikimate pathway: gene fusions, horizontal gene transfer, and endosymbiotic replacements. *Eukaryot Cell* 5, 1517-1531.
- Rivera, M.C., and Lake, J.A. (1992). Evidence that eukaryotes and eocyte prokaryotes are immediate relatives. *Science* 257, 74-76.
- Rivera, M.C., and Lake, J.A. (2004). The ring of life provides evidence for a genome fusion origin of eukaryotes. *Nature* 431, 152-155.
- Rodriguez-Ezpeleta, N., Brinkmann, H., Burey, S.C., Roure, B., Burger, G., Loffelhardt, W., Bohnert, H.J., Philippe, H., and Lang, B.F. (2005). Monophyly of primary photosynthetic eukaryotes: green plants, red algae, and glaucophytes. *Curr Biol* 15, 1325-1330.
- Rodriguez-Ezpeleta, N., Brinkmann, H., Burger, G., Roger, A.J., Gray, M.W., Philippe, H., and Lang, B.F. (2007a). Toward resolving the eukaryotic tree: the phylogenetic positions of jakobids and cercozoans. *Curr Biol* 17, 1420-1425.
- Rodriguez-Ezpeleta, N., Brinkmann, H., Roure, B., Lartillot, N., Lang, B.F., and Philippe, H. (2007b). Detecting and overcoming systematic errors in genome-scale phylogenies. *Syst Biol* 56, 389-399.
- Roger, A.J., Sandblom, O., Doolittle, W.F., and Philippe, H. (1999). An evaluation of elongation factor 1 alpha as a phylogenetic marker for eukaryotes. *Mol Biol Evol* 16, 218-233.
- Roger, A.J., and Simpson, A.G. (2009). Evolution: revisiting the root of the eukaryote tree. *Curr Biol* 19, R165-167.
- Rogozin, I.B., Wolf, Y.I., Carmel, L., and Koonin, E.V. (2007a). Analysis of rare amino acid replacements supports the Coelomata clade. *Mol Biol Evol* 24, 2594-2597.
- Rogozin, I.B., Wolf, Y.I., Carmel, L., and Koonin, E.V. (2007b). Ecdysozoan clade rejected by genome-wide analysis of rare amino acid replacements. *Mol Biol Evol* 24, 1080-1090.
- Ronquist, F., and Huelsenbeck, J.P. (2003). MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* 19, 1572-1574.
- Rout, M.P., Aitchison, J.D., Suprapto, A., Hjertaas, K., Zhao, Y., and Chait, B.T. (2000). The yeast nuclear pore complex: composition, architecture, and transport mechanism. *J Cell Biol* 148, 635-651.
- Ruiz-Trillo, I., Lane, C.E., Archibald, J.M., and Roger, A.J. (2006). Insights into the evolutionary origin and genome architecture of the unicellular opisthokonts *Capsaspora owczarzaki* and *Sphaeroforma arctica*. *J Eukaryot Microbiol* 53, 379-384.
- Ruiz-Trillo, I., Roger, A.J., Burger, G., Gray, M.W., and Lang, B.F. (2008). A phylogenomic investigation into the origin of metazoa. *Mol Biol Evol* 25, 664-672.

- Sagan, L. (1967). On the origin of mitosing cells. *J Theor Biol* *14*, 255-274.
- Samson, R.Y., Obita, T., Freund, S.M., Williams, R.L., and Bell, S.D. (2008). A role for the ESCRT system in cell division in archaea. *Science* *322*, 1710-1713.
- Sanchez-Puerta, M.V., Bachvaroff, T.R., and Delwiche, C.F. (2007). Sorting wheat from chaff in multi-gene analyses of chlorophyll c-containing plastids. *Mol Phylogenet Evol* *44*, 885-897.
- Satir, P., Guerra, C., and Bell, A.J. (2007). Evolution and persistence of the cilium. *Cell Motil Cytoskeleton* *64*, 906-913.
- Sauer, G., Korner, R., Hanisch, A., Ries, A., Nigg, E.A., and Sillje, H.H. (2005). Proteome analysis of the human mitotic spindle. *Mol Cell Proteomics* *4*, 35-43.
- Sawitzky, H., and Grolig, F. (1995). Phragmoplast of the green alga *Spirogyra* is functionally distinct from the higher plant phragmoplast. *J Cell Biol* *130*, 1359-1371.
- Schein, A.I., Kissinger, J.C., and Ungar, L.H. (2001). Chloroplast transit peptide prediction: a peek inside the black box. *Nucleic Acids Res* *29*, E82.
- Scherl, A., Coute, Y., Deon, C., Calle, A., Kindbeiter, K., Sanchez, J.C., Greco, A., Hochstrasser, D., and Diaz, J.J. (2002). Functional proteomic analysis of human nucleolus. *Mol Biol Cell* *13*, 4100-4109.
- Schierwater, B., Eitel, M., Jakob, W., Osigus, H.J., Hadrys, H., Dellaporta, S.L., Kolokotronis, S.O., and Desalle, R. (2009). Concatenated analysis sheds light on early metazoan evolution and fuels a modern "urmetazoon" hypothesis. *PLoS Biol* *7*, e20.
- Schleper, C., Jurgens, G., and Jonuscheit, M. (2005). Genomic studies of uncultivated archaea. *Nat Rev Microbiol* *3*, 479-488.
- Schleper, C., and Nicol, G.W. (2010). Ammonia-oxidising archaea--physiology, ecology and evolution. *Adv Microb Physiol* *57*, 1-41.
- Schreiber, A., Stengel, F., Zhang, Z., Enchev, R.I., Kong, E.H., Morris, E.P., Robinson, C.V., da Fonseca, P.C., and Barford, D. (2011). Structural basis for the subunit assembly of the anaphase-promoting complex. *Nature* *470*, 227-232.
- Schroder, E., Littlechild, J.A., Lebedev, A.A., Errington, N., Vagin, A.A., and Isupov, M.N. (2000). Crystal structure of decameric 2-Cys peroxiredoxin from human erythrocytes at 1.7 Å resolution. *Structure* *8*, 605-615.
- Schwab, M., Lutum, A.S., and Seufert, W. (1997). Yeast Hct1 is a regulator of Clb2 cyclin proteolysis. *Cell* *90*, 683-693.
- Schwab, M., Neutzner, M., Mocker, D., and Seufert, W. (2001). Yeast Hct1 recognizes the mitotic cyclin Clb2 and other substrates of the ubiquitin ligase APC. *EMBO J* *20*, 5165-5175.
- Schwickart, M., Havlis, J., Habermann, B., Bogdanova, A., Camasses, A., Oelschlaegel, T., Shevchenko, A., and Zachariae, W. (2004). Swm1/Apc13 is an evolutionarily conserved subunit of the anaphase-promoting complex stabilizing the association of Cdc16 and Cdc27. *Mol Cell Biol* *24*, 3562-3576.
- Searcy, D.G. (2003). Metabolic integration during the evolutionary origin of mitochondria. *Cell Res* *13*, 229-238.
- Seidl, M.F., and Schultz, J. (2009). Evolutionary flexibility of protein complexes. *BMC Evol Biol* *9*, 155.
- Selvapandiyani, A., Kumar, P., Morris, J.C., Salisbury, J.L., Wang, C.C., and Nakhasi, H.L. (2007). Centrin1 is required for organelle segregation and cytokinesis in *Trypanosoma brucei*. *Mol Biol Cell* *18*, 3290-3301.
- Seo, M.S., Kang, S.W., Kim, K., Baines, I.C., Lee, T.H., and Rhee, S.G. (2000). Identification of a new type of mammalian peroxiredoxin that forms an intramolecular disulfide as a reaction intermediate. *J Biol Chem* *275*, 20346-20354.
- Serralbo, O., Perez-Perez, J.M., Heidstra, R., and Scheres, B. (2006). Non-cell-autonomous rescue of anaphase-promoting complex function revealed by mosaic analysis of HOBBIT, an *Arabidopsis* CDC27 homolog. *Proc Natl Acad Sci U S A* *103*, 13250-13255.
- Shalchian-Tabrizi, K., Minge, M.A., Espelund, M., Orr, R., Ruden, T., Jakobsen, K.S., and Cavalier-Smith, T. (2008). Multigene phylogeny of choanozoa and the origin of animals. *PLoS One* *3*, e2098.
- Shirayama, M., Toth, A., Galova, M., and Nasmyth, K. (1999). APC(Cdc20) promotes exit from mitosis by destroying the anaphase inhibitor Pds1 and cyclin Clb5. *Nature* *402*, 203-207.

- Silverman-Gavrila, R.V., and Wilde, A. (2006). Ran is required before metaphase for spindle assembly and chromosome alignment and after metaphase for chromosome segregation and spindle midbody organization. *Mol Biol Cell* **17**, 2069-2080.
- Simpson, A.G. (2003). Cytoskeletal organization, phylogenetic affinities and systematics in the contentious taxon Excavata (Eukaryota). *Int J Syst Evol Microbiol* **53**, 1759-1777.
- Simpson, A.G., Inagaki, Y., and Roger, A.J. (2006). Comprehensive multigene phylogenies of excavate protists reveal the evolutionary positions of "primitive" eukaryotes. *Mol Biol Evol* **23**, 615-625.
- Simpson, A.G., and Patterson, D.J. (2001). On core jakobids and excavate taxa: the ultrastructure of *Jakoba incarcerata*. *J Eukaryot Microbiol* **48**, 480-492.
- Simpson, A.G., and Roger, A.J. (2004). The real 'kingdoms' of eukaryotes. *Curr Biol* **14**, R693-696.
- Simpson, A.G., Roger, A.J., Silberman, J.D., Leipe, D.D., Edgcomb, V.P., Jermini, L.S., Patterson, D.J., and Sogin, M.L. (2002). Evolutionary history of "early-diverging" eukaryotes: the excavate taxon *Carpedidomonas* is a close relative of *Giardia*. *Mol Biol Evol* **19**, 1782-1791.
- Skop, A.R., Liu, H., Yates, J., 3rd, Meyer, B.J., and Heald, R. (2004). Dissection of the mammalian midbody proteome reveals conserved cytokinesis mechanisms. *Science* **305**, 61-66.
- Sogin, M.L. (1991). Early evolution and the origin of eukaryotes. *Curr Opin Genet Dev* **1**, 457-463.
- Sontheimer, E.J. (2005). Assembly and function of RNA silencing complexes. *Nat Rev Mol Cell Biol* **6**, 127-138.
- Spang, A., Hatzepichler, R., Brochier-Armanet, C., Rattei, T., Tischler, P., Spieck, E., Streit, W., Stahl, D.A., Wagner, M., and Schleper, C. (2010). Distinct gene set in two different lineages of ammonia-oxidizing archaea supports the phylum Thaumarchaeota. *Trends Microbiol* **18**, 331-340.
- Spear, J.R., Walker, J.J., McCollom, T.M., and Pace, N.R. (2005). Hydrogen and bioenergetics in the Yellowstone geothermal ecosystem. *Proc Natl Acad Sci U S A* **102**, 2555-2560.
- Stanier, R.Y. (1961). La place des bactéries dans le monde vivant. *Ann Inst Pasteur* **3**, 297-312.
- Stechmann, A., and Cavalier-Smith, T. (2002). Rooting the eukaryote tree by using a derived gene fusion. *Science* **297**, 89-91.
- Stechmann, A., and Cavalier-Smith, T. (2003a). Phylogenetic analysis of eukaryotes using heat-shock protein Hsp90. *J Mol Evol* **57**, 408-419.
- Stechmann, A., and Cavalier-Smith, T. (2003b). The root of the eukaryote tree pinpointed. *Curr Biol* **13**, R665-666.
- Steenkamp, E.T., Wright, J., and Baldauf, S.L. (2006). The protistan origins of animals and fungi. *Mol Biol Evol* **23**, 93-106.
- Stemmann, O., Zou, H., Gerber, S.A., Gygi, S.P., and Kirschner, M.W. (2001). Dual inhibition of sister chromatid separation at metaphase. *Cell* **107**, 715-726.
- Stiller, J.W., and Hall, B.D. (1999). Long-branch attraction and the rDNA model of early eukaryotic evolution. *Mol Biol Evol* **16**, 1270-1279.
- Stoecker, D.K. (1999). Mixotrophy among dinoflagellates. *J Eukaryot Microbiol* **46**, 397-401.
- Striepen, B., Jordan, C.N., Reiff, S., and van Dooren, G.G. (2007). Building the perfect parasite: cell division in apicomplexa. *PLoS Pathog* **3**, e78.
- Susko, E., and Roger, A.J. (2007). On reduced amino acid alphabets for phylogenetic inference. *Mol Biol Evol* **24**, 2139-2150.
- Tang, G. (2005). siRNA and miRNA: an insight into RISCs. *Trends Biochem Sci* **30**, 106-114.
- Taylor, F.J.R. (1987). *The Biology of Dinoflagellates*. (Blackwell Scientific Publications).
- Tengs, T., Dahlberg, O.J., Shalchian-Tabrizi, K., Klaveness, D., Rudi, K., Delwiche, C.F., and Jakobsen, K.S. (2000). Phylogenetic analyses indicate that the 19'Hexanoyloxy-fucoxanthin-containing dinoflagellates have tertiary plastids of haptophyte origin. *Mol Biol Evol* **17**, 718-729.
- Thomarat, F., Vivares, C.P., and Gouy, M. (2004). Phylogenetic analysis of the complete genome sequence of *Encephalitozoon cuniculi* supports the fungal origin of microsporidia and reveals a high frequency of fast-evolving genes. *J Mol Evol* **59**, 780-791.
- Thompson, J.D., Gibson, T.J., and Higgins, D.G. (2002). Multiple sequence alignment using ClustalW and ClustalX. *Curr Protoc Bioinformatics* *Chapter 2*, Unit 2.3.

- Thornton, B.R., and Toczyski, D.P. (2003). Securin and B-cyclin/CDK are the only essential targets of the APC. *Nat Cell Biol* 5, 1090-1094.
- Thornton, B.R., and Toczyski, D.P. (2006). Precise destruction: an emerging picture of the APC. *Genes Dev* 20, 3069-3078.
- Tomita, J., Nakajima, M., Kondo, T., and Iwasaki, H. (2005). No transcription-translation feedback in circadian rhythm of KaiC phosphorylation. *Science* 307, 251-254.
- Tourasse, N.J., and Gouy, M. (1999). Accounting for evolutionary rate variation among sequence sites consistently changes universal phylogenies deduced from rRNA and protein-coding genes. *Mol Phylogenet Evol* 13, 159-168.
- Van de Peer, Y., and De Wachter, R. (1997). Evolutionary relationships among eukaryotic crown taxa taking into account site-to-site variation in 18S rRNA. *J Mol Evol* 45, 619-630.
- Verma, D.P. (2001). Cytokinesis and Building of the Cell Plate in Plants. *Annu Rev Plant Physiol Plant Mol Biol* 52, 751-784.
- Visintin, R., Prinz, S., and Amon, A. (1997). CDC20 and CDH1: a family of substrate-specific activators of APC-dependent proteolysis. *Science* 278, 460-463.
- Vodermaier, H.C., Gieffers, C., Maurer-Stroh, S., Eisenhaber, F., and Peters, J.M. (2003). TPR subunits of the anaphase-promoting complex mediate binding to the activator protein CDH1. *Curr Biol* 13, 1459-1468.
- von Dohlen, C.D., Kohler, S., Alsop, S.T., and McManus, W.R. (2001). Mealybug beta-proteobacterial endosymbionts contain gamma-proteobacterial symbionts. *Nature* 412, 433-436.
- von Schubert, C., Xue, G., Schmuckli-Maurer, J., Woods, K.L., Nigg, E.A., and Dobbelaere, D.A. (2010). The transforming parasite *Theileria* co-opts host cell mitotic and central spindles to persist in continuously dividing cells. *PLoS Biol* 8.
- Wagner, E., Luche, S., Penna, L., Chevallet, M., Van Dorsselaer, A., Leize-Wagner, E., and Rabilloud, T. (2002). A method for detection of overoxidation of cysteines: peroxiredoxins are oxidized in vivo at the active-site cysteine during oxidative stress. *Biochem J* 366, 777-785.
- Wagner, I.D., Zhao, W., Zhang, C.L., Romanek, C.S., Rohde, M., and Wiegell, J. (2008). *Thermoanaerobacter uzonensis* sp. nov., an anaerobic thermophilic bacterium isolated from a hot spring within the Uzon Caldera, Kamchatka, Far East Russia. *Int J Syst Evol Microbiol* 58, 2565-2573.
- Wainright, P.O., Hinkle, G., Sogin, M.L., and Stickel, S.K. (1993). Monophyletic origins of the metazoa: an evolutionary link with fungi. *Science* 260, 340-342.
- Walker, C.B., de la Torre, J.R., Klotz, M.G., Urakawa, H., Pinel, N., Arp, D.J., Brochier-Armanet, C., Chain, P.S., Chan, P.P., Gollabgir, A., *et al.* (2010). *Nitrosopumilus maritimus* genome reveals unique mechanisms for nitrification and autotrophy in globally distributed marine crenarchaea. *Proc Natl Acad Sci U S A* 107, 8818-8823.
- Walther, A., and Wendland, J. (2003). Septation and cytokinesis in fungi. *Fungal Genet Biol* 40, 187-196.
- Wang, Z., Yu, R., and Melmed, S. (2001). Mice lacking pituitary tumor transforming gene show testicular and splenic hypoplasia, thymic hyperplasia, thrombocytopenia, aberrant cell cycle progression, and premature centromere division. *Mol Endocrinol* 15, 1870-1879.
- Werner, F. (2008). Structural evolution of multisubunit RNA polymerases. *Trends Microbiol* 16, 247-250.
- Wilson, E.B., ed. (1925). *The Cell in Development and Heredity*, 3rd edition edn (New York, The Macmillan Company.).
- Wirth, K.G., Ricci, R., Gimenez-Abian, J.F., Taghybeeglu, S., Kudo, N.R., Jochum, W., Vasseur-Cognet, M., and Nasmyth, K. (2004). Loss of the anaphase-promoting complex in quiescent cells causes unscheduled hepatocyte proliferation. *Genes Dev* 18, 88-98.
- Woese, C.R. (1987). Bacterial evolution. *Microbiol Rev* 51, 221-271.
- Woese, C.R., and Fox, G.E. (1977). Phylogenetic structure of the prokaryotic domain: the primary kingdoms. *Proc Natl Acad Sci U S A* 74, 5088-5090.
- Woese, C.R., Kandler, O., and Wheelis, M.L. (1990). Towards a natural system of organisms: proposal for the domains Archaea, Bacteria, and Eucarya. *Proc Natl Acad Sci U S A* 87, 4576-4579.
- Wolf, Y.I., Rogozin, I.B., and Koonin, E.V. (2004). Coelomata and not Ecdysozoa: evidence from genome-wide phylogenetic analysis. *Genome Res* 14, 29-36.

- Wonsey, D.R., Zeller, K.I., and Dang, C.V. (2002). The c-Myc target gene PRDX3 is required for mitochondrial homeostasis and neoplastic transformation. *Proc Natl Acad Sci U S A* *99*, 6649-6654.
- Woo, H.A., Bae, S.H., Park, S., and Rhee, S.G. (2009). Sestrin 2 is not a reductase for cysteine sulfinic acid of peroxiredoxins. *Antioxid Redox Signal* *11*, 739-745.
- Wood, Z.A., Poole, L.B., Hantgan, R.R., and Karplus, P.A. (2002). Dimers to doughnuts: redox-sensitive oligomerization of 2-cysteine peroxiredoxins. *Biochemistry* *41*, 5493-5504.
- Wood, Z.A., Poole, L.B., and Karplus, P.A. (2003a). Peroxiredoxin evolution and the regulation of hydrogen peroxide signaling. *Science* *300*, 650-653.
- Wood, Z.A., Schroder, E., Robin Harris, J., and Poole, L.B. (2003b). Structure, mechanism and regulation of peroxiredoxins. *Trends Biochem Sci* *28*, 32-40.
- Yamamoto, A., Guacci, V., and Koshland, D. (1996). Pds1p is required for faithful execution of anaphase in the yeast, *Saccharomyces cerevisiae*. *J Cell Biol* *133*, 85-97.
- Yang, Z. (1993). Maximum-likelihood estimation of phylogeny from DNA sequences when substitution rates differ over sites. *Mol Biol Evol* *10*, 1396-1401.
- Yoon, H.J., Feoktistova, A., Wolfe, B.A., Jennings, J.L., Link, A.J., and Gould, K.L. (2002a). Proteomics analysis identifies new components of the fission and budding yeast anaphase-promoting complexes. *Curr Biol* *12*, 2048-2054.
- Yoon, H.S., Grant, J., Tekle, Y.I., Wu, M., Chaon, B.C., Cole, J.C., Logsdon, J.M., Jr., Patterson, D.J., Bhattacharya, D., and Katz, L.A. (2008). Broadly sampled multigene trees of eukaryotes. *BMC Evol Biol* *8*, 14.
- Yoon, H.S., Hackett, J.D., Pinto, G., and Bhattacharya, D. (2002b). The single, ancient origin of chromist plastids. *Proc Natl Acad Sci U S A* *99*, 15507-15512.
- Yoon, H.S., Reyes-Prieto, A., Melkonian, M., and Bhattacharya, D. (2006). Minimal plastid genome evolution in the *Paulinella* endosymbiont. *Curr Biol* *16*, R670-672.
- Zablen, L.B., Kissil, M.S., Woese, C.R., and Buetow, D.E. (1975). Phylogenetic origin of the chloroplast and prokaryotic nature of its ribosomal RNA. *Proc Natl Acad Sci U S A* *72*, 2418-2422.
- Zachariae, W., and Nasmyth, K. (1996). TPR proteins required for anaphase progression mediate ubiquitination of mitotic B-type cyclins in yeast. *Mol Biol Cell* *7*, 791-801.
- Zachariae, W., Shin, T.H., Galova, M., Obermaier, B., and Nasmyth, K. (1996). Identification of subunits of the anaphase-promoting complex of *Saccharomyces cerevisiae*. *Science* *274*, 1201-1204.
- Zhaxybayeva, O., and Gogarten, J.P. (2004). Cladogenesis, coalescence and the evolution of the three domains of life. *Trends Genet* *20*, 182-187.
- Zhou, Z., Licklider, L.J., Gygi, S.P., and Reed, R. (2002). Comprehensive proteomic analysis of the human spliceosome. *Nature* *419*, 182-185.
- Zhu, C., Bossy-Wetzel, E., and Jiang, W. (2005). Recruitment of MKLP1 to the spindle midzone/midbody by INCENP is essential for midbody formation and completion of cytokinesis in human cells. *Biochem J* *389*, 373-381.
- Zillig, W., Klenk, H.P., Palm, P., Puhler, G., Gropp, F., Garrett, R.A., and Leffers, H. (1989). The phylogenetic relations of DNA-dependent RNA polymerases of archaeobacteria, eukaryotes, and eubacteria. *Can J Microbiol* *35*, 73-80.
- Zou, H., McGarry, T.J., Bernal, T., and Kirschner, M.W. (1999). Identification of a vertebrate sister-chromatid separation inhibitor involved in transformation and tumorigenesis. *Science* *285*, 418-422.
- Zuckerkindl, E., and Pauling, L. (1965). Molecules as documents of evolutionary history. *J Theor Biol* *8*, 357-366.
- Zwickl, D.J., and Hillis, D.M. (2002). Increased taxon sampling greatly reduces phylogenetic error. *Syst Biol* *51*, 588-598.

ARTICLE 1 :

A Complex Cell Division Machinery Was Present in
the Last Common Ancestor of Eukaryotes

A Complex Cell Division Machinery Was Present in the Last Common Ancestor of Eukaryotes

Laura Eme¹, David Moreira², Emmanuel Talla^{1,3}, Céline Brochier-Armanet^{1,4*}

1 Laboratoire de Chimie Bacterienne – UPR CNRS 9043, Marseille, France, **2** Unité d'Ecologie, Systématique et Evolution – UMR CNRS 8079, Université Paris-Sud, Orsay, France, **3** Université de la Méditerranée, Marseille, France, **4** Université de Provence, Marseille, France

Abstract

Background: The midbody is a transient complex structure containing proteins involved in cytokinesis. Up to now, it has been described only in Metazoa. Other eukaryotes present a variety of structures implied in the last steps of cell division, such as the septum in fungi or the phragmoplast in plants. However, it is unclear whether these structures are homologous (derive from a common ancestral structure) or analogous (have distinct evolutionary origins). Recently, the proteome of the hamster midbody has been characterized and 160 proteins identified.

Methodology/Principal Findings: Using phylogenomic approaches, we show here that nearly all of these 160 proteins (95%) are conserved across metazoan lineages. More surprisingly, we show that a large part of the mammalian midbody components (91 proteins) were already present in the last common ancestor of all eukaryotes (LECA) and were most likely involved in the construction of a complex multi-protein assemblage acting in cell division.

Conclusions/Significance: Our results indicate that the midbodies of non-mammalian metazoa are likely very similar to the mammalian one and that the ancestor of Metazoa possessed a nearly modern midbody. Moreover, our analyses support the hypothesis that the midbody and the structures involved in cytokinesis in other eukaryotes derive from a large and complex structure present in LECA, likely involved in cytokinesis. This is an additional argument in favour of the idea of a complex ancestor for all contemporary eukaryotes.

Citation: Eme L, Moreira D, Talla E, Brochier-Armanet C (2009) A Complex Cell Division Machinery Was Present in the Last Common Ancestor of Eukaryotes. PLoS ONE 4(4): e5021. doi:10.1371/journal.pone.0005021

Editor: Jason E. Stajich, University of California, Berkeley, United States of America

Received: January 21, 2009; **Accepted:** March 2, 2009; **Published:** April 7, 2009

Copyright: © 2009 Eme et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Funding: LE is the recipient of a grant from the Ministère de l'Enseignement Supérieur et de la Recherche. CBA is supported by an Action Thématique et Incitative sur Programme (ATIP) of the French Centre National de la Recherche Scientifique, CNRS. This work was supported by a grant from the French Agence Nationale de la Recherche (ANR JC05_44674). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

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

* E-mail: celine.brochier@ifr88.cnrs-mrs.fr

Introduction

The presence of large multiprotein complexes has been often considered as one of the main distinctive differences between eukaryotic and prokaryotic cells. These Eukaryotic Multiprotein Complexes (EMC), such as the nuclear pore complex associated to nuclear membranes [1], the spliceosome [2], the telomeric complex [3], and the nucleolus [4], are involved in various cellular processes or structures. Surprisingly, even if similar cellular processes exist in prokaryotes, these appear to be fulfilled by simpler and often non homologous systems [5,6]. This suggests that the appearance of EMC was a eukaryotic-specific evolutionary trend. Despite their importance to understand eukaryotic evolution, the evolutionary origin of many EMC remains elusive. In fact, most studies of EMC have focused on the functional characterization of some of their components rather than on evolutionary considerations.

Recently, thanks to the development of massive proteomic methods, a number of EMC, including the telomeric complex [7], the flagellum and the associated basal body [8,9], or the centromere complex [10], have been characterized leading to the identification of most of their components. This, in parallel with the increasing availability of genomic data for diverse

eukaryotic and prokaryotic species, has opened the possibility to carry out phylogenomic studies of these biological systems. This kind of analysis consists in identifying the homologues of each component involved in a biological system or cellular process in all lineages of the studied taxonomic group [11,12]. The phylogenetic analysis of each component then allows determining its evolutionary origin as well as its subsequent evolutionary history (duplication events, losses, horizontal gene transfer (HGT)). The interpretation of functional data from experimental studies (e.g., mutant defects, interactome data, expression data) or from *in silico* analyses (e.g., functional domain and motif searches, genomic context analysis) in the light of the evolutionary framework obtained may help to infer functional predictions for organisms for which no such functional data are available, as well as to propose targets for experimental analyses. The recent availability of both proteomic and genomic data in eukaryotes has provided the necessary starting point for the first evolutionary studies of EMC [13,14,15]. This has allowed addressing questions such as when a particular EMC appeared and what was its ancestral composition, and to retrace its history (e.g., component acquisitions and losses) in the different lineages.

Among the different EMC known, those responsible of cytokinesis, the last step of the cell division, are particularly

complex. First described in 1891 [16], the midbody is a transient organelle-like structure found in animal cells that forms a bridge between the two daughter cells at the end of the cytokinesis [17,18,19,20]. It consists of a compact and dense matrix of proteins of 1–1.5 μm in diameter, tightly bundled by the cleavage furrow [21]. It is composed mainly of microtubules derived from the spindle midzone that are coated by microtubule-associated proteins that maintain the structure [17]. Golgi-derived vesicles, possibly involved in the transport of proteins and membranes, as well as proteins with secretory and membrane-traffic activities, have also been reported along and around the midbody microtubules [22]. A midbody association with endoplasmic reticulum structures has also been described in some animals, as *Drosophila melanogaster* and *Caenorhabditis elegans* (see [17] and references therein). At the end of the cytokinesis, the midbody is severed, leading to the separation of the daughter cells. Even if no clear function has been ascribed to the midbody, mutation or silencing of most of its components is associated to defects in cleavage furrow formation or completion and to defects in germline cytokinesis [22].

Up to now, the midbody has only been described in metazoans. Despite similarities at the mechanistic level, Metazoa, Fungi and other eukaryotes, present a variety of structures involved in the last steps of cell division. Consequently, it is unclear whether these structures are homologous (derived from a common ancestral structure) or analogous (have distinct evolutionary origins) [17,23]. For example, the amoebozoan *Dictyostelium discoideum* produces cytoplasmic bridges that present some similarities with the animal midbody (see [20,24,25]), whereas Fungi exhibit a chitin-rich septum at the division site [26]. The most distinct cytokinesis structure analysed in detail, the phragmoplast, is found in land plants. It is a dense structure likely derived from the mitotic spindle by reassembling of microtubules, which become responsible of the targeted delivery of membrane vesicles for the formation of the division plate [19,23,27,28]. Although it contains a certain amount of actin with an unclear role, the plant phragmoplast appears not to involve an actinomyosin contractile ring, in contrast with the animal midbody [28], and the division plate grows centrifugally. Both the midbody and the phragmoplast contain a number of microtubule-associated proteins and kinesin motor molecules that stabilize the structure [23]. Interestingly, in red algae, which are close relatives to the green plants (Figure S1), the presence of an actin contractile ring and furrowing have been reported [29], suggesting that the plant phragmoplast is a very derived structure. Accordingly, various situations have been described in green algae [27,30]. In most green algae, cytokinesis involves furrowing from the cell cortex, but centrifugal division by means of a phragmoplast also exists (see [30] and [31] and references therein). Intermediate situations have also been described as in *Spirogyra*, in which cytokinesis is initiated by an actin-based cleavage furrow and completed by a phragmoplast-like array of microtubules. However, in contrast with land plants, that phragmoplast-like structure progresses towards the centre of the cell division plan and does not grow centrifugally [32].

A recent proteomic analysis has identified 160 different proteins in the midbody from Chinese hamster ovary cells [22]. As expected, most of them are involved in secretory and membrane traffic activities (53 proteins), are actin- or microtubule-associated proteins (46 and 18 proteins, respectively), and protein kinases (18 proteins). The remaining 25 proteins display a variety of functions [22] (Table S1). Using these proteins as a starting point, we have carried out a phylogenomic analysis of the mammalian midbody and we show that a midbody very similar in protein composition to the mammalian one was most likely present in the last common

ancestor of Metazoa. We also show that a large fraction of the mammalian midbody components are ancient and were already present in the last eukaryotic common ancestor (LECA), even if they appear to have originated after the prokaryotes/eukaryotes separation. Moreover, taking into account available functional data and based on functional domains present in these proteins, it is possible to infer that in non-metazoan eukaryotes these proteins have similar molecular functions and are most likely also involved in the late step of cytokinesis. All this indicates that a complex cellular structure responsible of cell division was already present in LECA, from which likely all of the present-day eukaryotic cytokinesis structures derive.

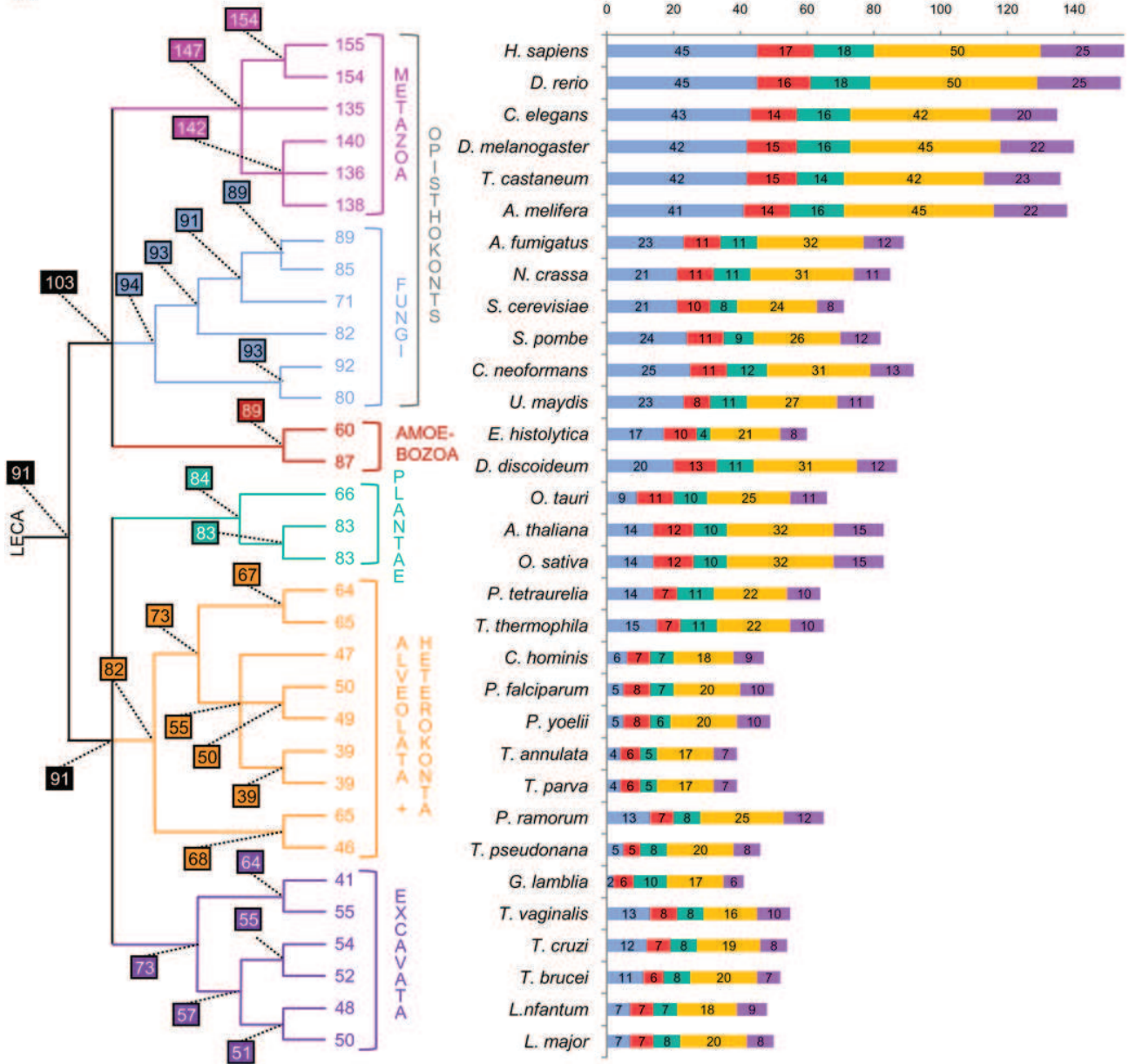
Results and Discussion

To start our analysis, we first retrieved from public databases all the homologues of each of the 160 proteins identified in the hamster midbody [22] (Table S1). To distinguish between orthologues and paralogues among all the homologues retrieved, we performed a phylogenetic analysis of each protein. This allowed an accurate identification of orthologues, which is an essential step of phylogenomic analyses because orthologues are more likely to share the same function than paralogues [11]. At this step, we discarded two proteins: CDK4 (kin6), because of the lack of phylogenetic signal to differentiate orthologues from paralogues, and Ankyrin B/Ank2 (act33), due to the presence of ankyrin repeats that distorted the multiple sequence alignment. Phylogenetic analyses of Dynamin 2, Dynamin-like I and Dynamin-like DLP1 (sec19, sec20 and sec21) and of Calnexin and Calreticulin (sec10 and sec12) showed that they resulted from two and one recent duplication events, respectively, which occurred in vertebrates and were thus gathered in two datasets (sec19/20/21 and sec10/12). 155 protein datasets were thus finally kept for further analyses. Our phylogenetic analyses showed a high number of duplications that are vertebrate-specific (more than one third of the midbody components had paralogues that were vertebrate-specific), in agreement with the proposal that the vertebrate lineage evolved through two successive rounds of complete genome duplication [33]. However, with the exception of the sec19/sec20/sec21 and sec10/sec12 cases mentioned above, the paralogues resulting from these duplication events were not part of the hamster midbody proteome. This suggests that functional divergences occurred after these duplication events.

A modern-like midbody in the ancestor of Metazoa

Based on the phylogenies of the 155 datasets retained, we identified as many as 140 and 135 orthologues in distant invertebrate species such as *Drosophila melanogaster* and *Caenorhabditis elegans*, respectively. This supported the hypothesis that the midbodies of these animals are very similar to the mammalian ones. In agreement with this result, a recent RNA interference-based inactivation analysis has showed that in *C. elegans*, ~100 homologues of hamster midbody components are involved in cleavage furrow formation or completion, or in germ line cytokinesis [22]. This confirmed that the composition and function of the midbody are extremely well conserved in the whole animal phylum. We also inferred that 154 and 147 proteins were present in the ancestor of Vertebrata and in the ancestor of Metazoa, representing 99% and 95% of the hamster midbody components, respectively (Figure 1A). This implied that only eight components were recruited in the lineage leading from the ancestor of Metazoa to mammals and suggested that the ancestor of living Metazoa had already a midbody with modern characteristics.

A



B

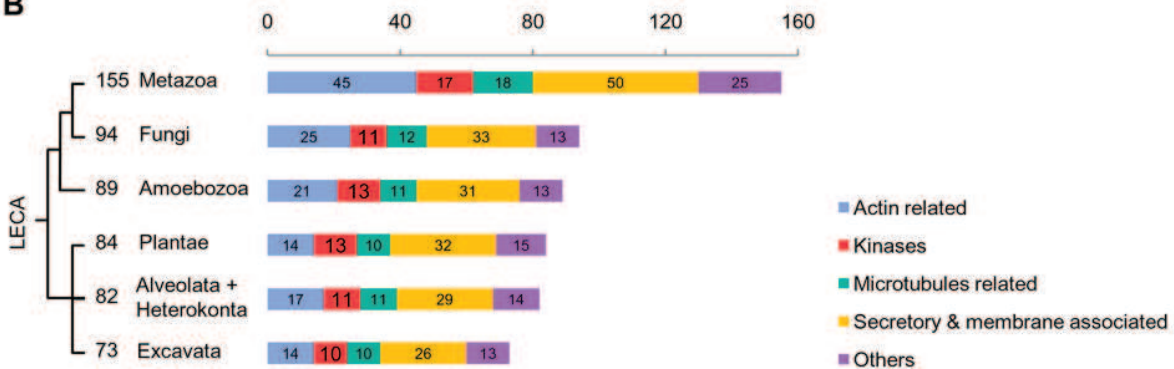


Figure 1. Distribution of the orthologues of the hamster midbody among eukaryotic lineages. Number of orthologues of the mammalian-midbody components present in each of the 32 eukaryotic species (A) and in the ancestor of each of the six main eukaryotic phyla (B). Colours represent the functional categories defined by Skop, et al. [22]. The phylogenetic relationships between the lineages analysed are indicated on the left. Numbers at nodes indicate the number of orthologues inferred to be present in the corresponding ancestors.
doi:10.1371/journal.pone.0005021.g001

Most midbody components were present in LECA

In non-metazoan eukaryotes, cytological studies have revealed a great diversity of structures involved in late cytokinesis, such as the plant phragmoplast and the fungal division septum [19,26,34]. However, cell biology data alone did not allow inferring whether these structures are homologous to the metazoan midbody (i.e., derive from an ancestral structure) or just analogous (i.e., have independent evolutionary origins) [17]. Thus, two competing hypotheses could be proposed for the origin of the midbody (Figure S1): (i) it appeared recently in the metazoan lineage after its separation from the fungal lineage or (ii) it derived from a more ancient structure that was present in the ancestor of Opisthokonta (i.e. ancestor of Metazoa and Fungi) or even in the Last Eukaryotic Common Ancestor (LECA). In the first case, the midbody would be expected to have no counterpart in non-metazoan species, whereas, in the second case, the ancestral structure from which the midbody derives could also be at the origin of the structures involved in cytokinesis in the other eukaryotic lineages. Although no proteomic data are available for those structures, phylogeny-based inference of the number of orthologues in non-metazoan lineages can be useful to discriminate whether the midbody is a metazoan-specific innovation (few orthologues would be expected to be conserved in non-metazoan lineages) or a descendant from a more ancient structure (a large set of orthologues should be present in various eukaryotic lineages). Accordingly, we analysed the orthologues found in 32 complete (or almost complete) genome sequences representative of the eukaryotic diversity, including species from most major eukaryotic lineages (Figure S1). At this step, the use of complete genome data was essential since it allowed the identification of true gene absences in the lineages studied.

The phylogenetic analysis of each of the 155 midbody proteins showed that only nine of them were shared by all the 32 eukaryotic species: actin (*act1*), casein kinase II (*kin3*), alpha tubulin (*mic1*), beta tubulin (*mic2*), G protein beta 2 (*oth9*), clathrin heavy chain (*sec14*), COPI alpha subunit (*sec16*), dynamin (*sec19/sec20/sec21*), and *sec23* (*sec44*). This appeared to support the hypothesis of a recent metazoan origin of the midbody. However, this extremely reduced number of shared components could also reflect a poor representation of midbody components only in a few particular species or lineages. To clarify this point, we analysed the number of orthologues shared by an increasing number of eukaryotic species (i.e. how many orthologues are found in at least two eukaryotic species, three eukaryotic species, and so on). As the number of species increased, the number of shared midbody components decreased in a very regular way until an almost complete extinction (Figure 2A). This reflected regular and independent losses that occurred in all the eukaryotic species. To minimize the impact of such species-specific losses, we carried out a similar analysis but considering the number of components shared by an increasing number of phyla (Figure 2B).

In sharp contrast with the previous single-species analysis, the phylum-based comparison allowed inferring that 57 proteins were shared by the six eukaryotic phyla, i.e., six times more than the nine components shared by all 32 eukaryotic species. Although some very divergent homologues might have escaped detection, this confirmed the hypothesis that the 32 eukaryotic genomes

analysed shared only a very small number of components because of independent losses that occurred at the species level. Accordingly, the number of midbody proteins inferred to have existed in the ancestors of each phylum was significantly higher than that observed in single species (Figure 1A and 1B). For example, we inferred 82 components in the ancestor of Alveolata and Heterokonta compared to a maximum of 65 detected in *Tetrahymena thermophila* and in *Phytophthora ramorum*. Similarly, we inferred that 94 orthologues were present in the ancestor of Fungi while only 71 were detected in *Saccharomyces cerevisiae* (Figure 1A and 1B). Surprisingly, the number of orthologues of midbody components found in each species was highly variable, even in species belonging to the same phylum. For example, within the Amoebozoa, we identified 87 orthologues in *D. discoideum* and only 60 in its relative *Entamoeba histolytica* (Figure 1A). Similarly, within ascomycetous fungi, *Aspergillus fumigatus* contained 89 orthologues, 18 more than *S. cerevisiae*, suggesting recent losses in the latter. This suggested different tempos of component loss across species, which may be relatively high for some of them, especially in parasites such as the apicomplexan *Theileria parva* or the diplomonad *Giardia lamblia* (with only 39 and 41 orthologues identified, whereas 55 and 73 are inferred in the ancestor of Apicomplexa and in the ancestor of Excavata, respectively, Figure 1A). Surprisingly, despite the fact that we used the metazoan midbody as reference, the number of orthologues detected did not correlate with the phylogenetic distance of each phylum to the Metazoa. For example, Metazoa did not share more orthologues with the closely related Fungi than they did with the much more distant land plants (an average of 83 in both cases).

Using the number of orthologues deduced for the ancestor of each eukaryotic phylum, we inferred that genes coding for at least 91 components (59%) of the mammalian midbody were already present in LECA. This implied that 56 additional midbody components originated later, in the branch leading from LECA to the ancestor of Metazoa (which likely contained 147 proteins, see above and Figure 1A). Therefore, the metazoan midbody appeared to have a dual nature, resulting from the combination of ancient (already present in LECA) and more recent components. The 91 LECA orthologues were well distributed among the five functional categories defined by Skop et al. [22] (Figure 3). In fact, we inferred that LECA possessed 82% of the kinases, 61% of the microtubule-associated proteins, 38% of the actin-associated proteins, 66% of the secretory and membrane-trafficking proteins and 64% of the proteins with other functions (Figure 3). Similar trends were observed for present-day species. The only exception concerned the actin-associated proteins (Figure 1A), for which dramatic losses have occurred in several unrelated lineages, such as Apicomplexa or the diplomonad *G. lamblia* (with only 4–6 and 2 out of 45 components detected, respectively, Figure 1A). This may reflect a less important role of actin in the cytokinesis of these eukaryotes. As a matter of fact, microtubules instead of actin filaments have been shown to drive daughter cell budding and cell division in Apicomplexa [35] whereas in the case of *G. lamblia*, a remarkable acceleration of the evolutionary rate in its actin sequence was observed (Figure S2). These modifications involving actin may have led to an evolutionary convergence in these lineages through the loss (or the replacement) of most actin-associated proteins.

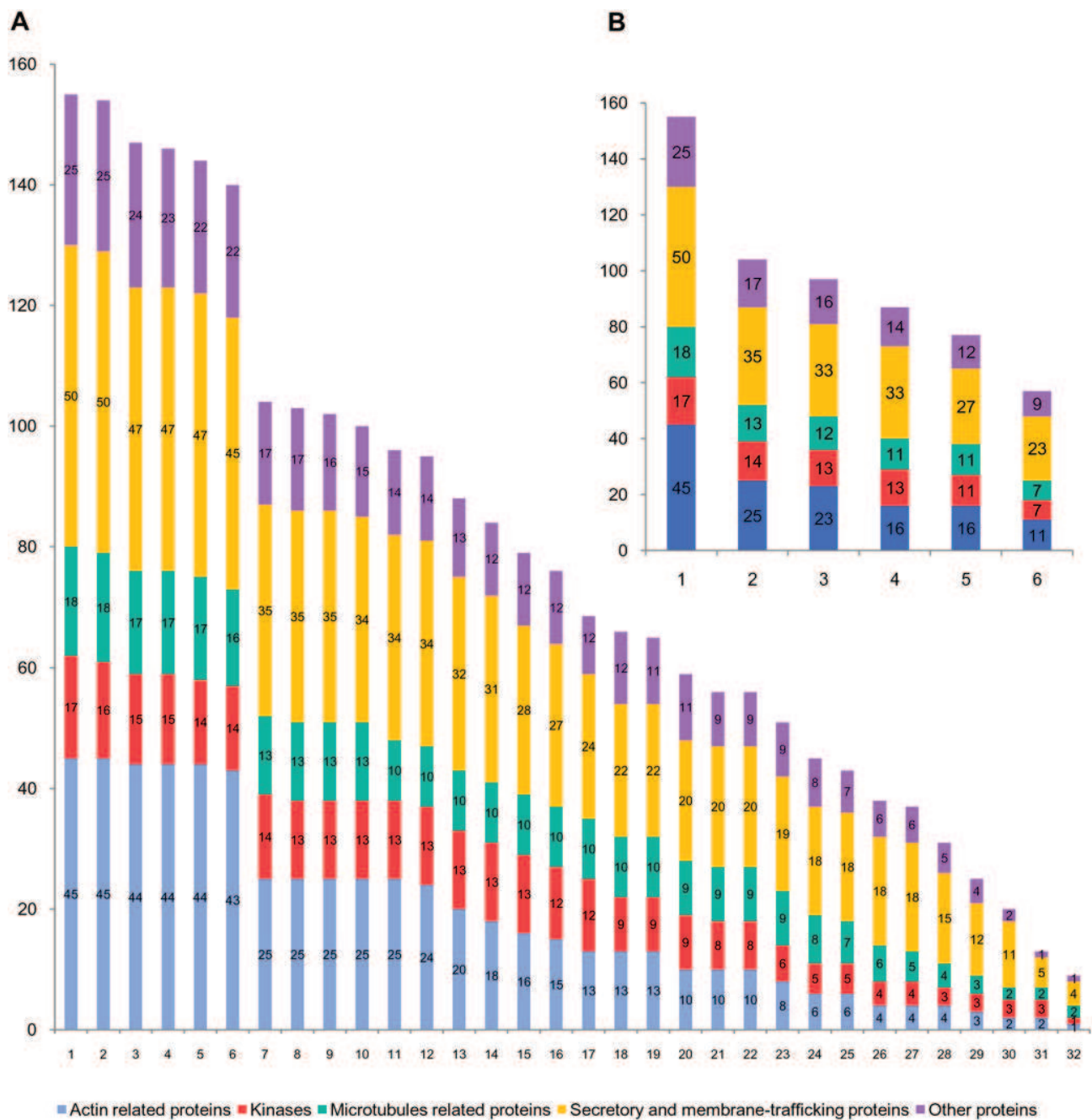


Figure 2. Shared orthologues of the hamster midbody among eukaryotic lineages. Number of midbody components shared (A) by an increasing number of single eukaryotic species (from at least 1 to 32) and (B) by an increasing number of eukaryotic phyla (from 1 to 6). Colours represent functional categories as in Figure 1. doi:10.1371/journal.pone.0005021.g002

Functional conservation in the midbody orthologues

To get more insight in the functional characterisation of the orthologues of non-metazoan lineages and those inferred in LECA, we studied their functional domain organisation in representatives of each eukaryotic phylum (Figure 4). We found that 66% up to 86% (average of 79%) of the orthologues in these non-metazoan species had functional domain compositions identical to those of their mammalian counterparts (Figure 4 and Tables S2 and S3). Although available functional data are scarce for non-metazoan lineages, this striking conservation across

very distant eukaryotic lineages strongly suggested that the non-metazoan orthologues have the same molecular functions than the mammalian ones and, therefore, may have similar cellular functions. To test this hypothesis, we carried out a survey of the *Saccharomyces* Genome Database (SGD, <http://www.yeastgenome.org/>) that showed that 31 of the 71 yeast orthologues (44%) have been characterised as involved in cytokinesis or cell division, and 55 (77%) have been found to interact with at least one of the other 71 (Table S4). These observations strongly suggested that, as in Metazoa, most of the yeast orthologues act together and are

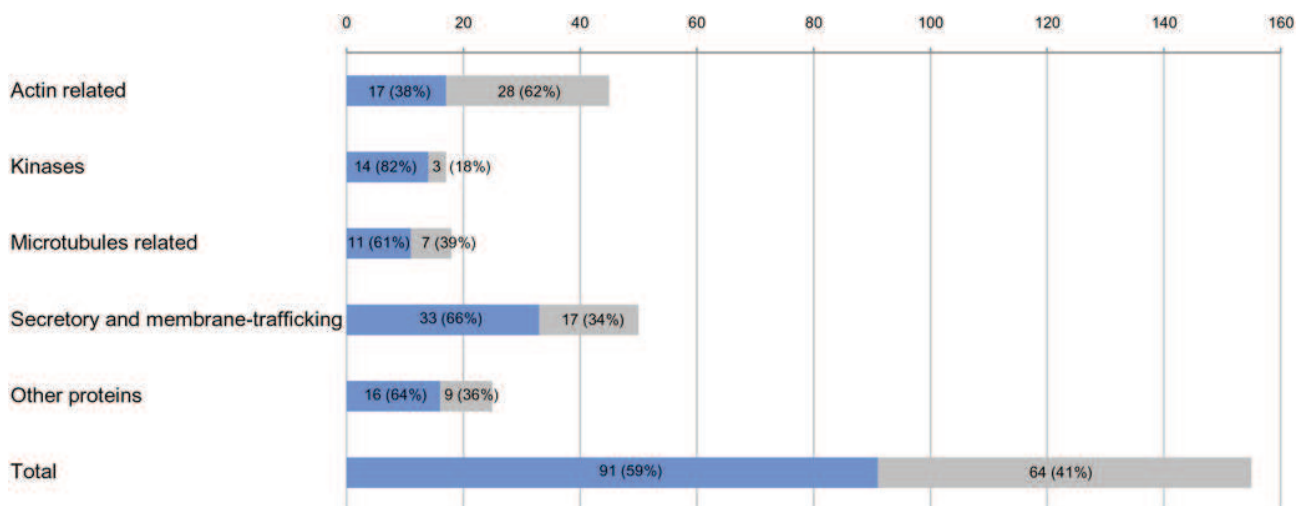


Figure 3. Evolutionary origin of the midbody components. Components inferred to have been present in the Last Eukaryotic Common Ancestor (LECA) are represented by blue bars, whereas those having a more recent origin (i.e. that appeared in the lineage leading from LECA to Mammalia) are represented by grey bars. Details for each functional class are provided. doi:10.1371/journal.pone.0005021.g003

involved in cell division, a characteristic likely inherited from the last common ancestor of Fungi and Metazoa. In Amoebozoa, functional data are too scarce to make this kind of analysis. However, among the high number of orthologues detected in the amoeba *D. discoideum* (at least 87), 70 have the same domain organisation of their metazoan orthologues, which correlated well with the description in this species of a cytokinesis structure that presents clear similarities with the animal midbody [20,24]. This suggests that this structure, the septum of Fungi and the metazoan midbody most likely derive from an ancestral structure already present in the ancestor of Unikonts.

These results are in agreement with a recent comparative study of 24 functionally characterised components of the midbody of various animals, the division septum of two fungi and the phragmoplast of plants, which suggested that most of them are homologous in these three distant groups, tracing back their origin to LECA [17]. Among those 24 components, we detected ten in the proteome of the hamster midbody (Table S5), and confirmed that at least four of them were true orthologues in the three groups (animals, fungi and plants), and eight were orthologues at least in two of them. For example, kin1 (called AIM-1 in mammals and AIR-2 in *C. elegans*) is essential for normal central spindle and cleavage furrow formation, and for the ultimate separation of the two daughter cells in Metazoa [36,37]. Its orthologues AtAurora 1 and 2 of *A. thaliana* and Ipl1p of *S. cerevisiae* have been shown to be associated to the phragmoplast and necessary to ensure appropriate positioning and assembly of the new cell wall in plants [38], and important for chromosome segregation in yeast [39], respectively. An additional evolutionary link between the midbody and the plant phragmoplast comes from the fact that the phragmoplast grows predominantly by the fusion of Golgi-derived vesicles [27] and it has been shown that 38 mammalian midbody components are Golgi-associated proteins [22].

In summary, our results support that (i) nearly 60% of the present-day mammalian midbody components were present in LECA and were well distributed in all the five functional categories, (ii) an average of 79% of the orthologues identified in a variety of unikonts and bikonts displayed a high conservation of functional domain composition, indicating a similar molecular

function, (iii) most of these orthologues interact in yeast and are likely involved in cytokinesis, and (iv) a number of these orthologues are part of the plant phragmoplast. All these four lines of evidence suggested that these structures derive from a much more ancient one dating back to LECA. This ancestral structure may have been composed of at least 90 interacting components and was already involved in cytokinesis. It is important to note here that the estimation of that ancestral number of components was likely underestimated since our analysis was based on the only proteome available (the hamster midbody), preventing the detection of all the components that have been lost specifically in the lineage leading from LECA to Mammalia. The characterisation of proteomes of equivalent structures in other eukaryotic lineages (especially in bikonts, such as the plant phragmoplast) will allow confirming the relationship between these structures, but also having a better estimation of the number of ancestral components and a clearer picture of the complexity of the ancestral structure that was present in LECA. Finally, the future availability of additional data from non-metazoan lineages will highlight the gains or replacements that occurred in each lineage. For example, it is expected that similar gains as the 56 new proteins recruited in the lineage leading from LECA to Metazoa also occurred in other lineages. The identification of these gains may help to explain the important differences observed in the structures involved in late cytokinesis in the different eukaryotic groups.

Testing the hypothesis of a prokaryotic origin

Since a complex multi-protein structure involved in cytokinesis most likely existed in LECA, it would be interesting to test if it derived from an even older structure that would have been present in prokaryotic ancestors. To address this question, we searched for homologues of the 155 hamster midbody components in all available complete prokaryotic genomes (730 genomes). We detected homologues for only nine: Actin (act1), ADP-Ribosylation factor-like 1 (sec2), Flotillin 1 (sec26), Gmx33/Golph3 (sec29), NSF (sec36), Rack 1 (sec42), alpha-tubulin (mic1), beta-tubulin (mic2), Novel/CGI-49 (oth19). Nevertheless, phylogenetic analyses revealed that four of them (Flotillin 1, Gmx33/Golph3, and alpha and beta-tubulin) likely resulted from

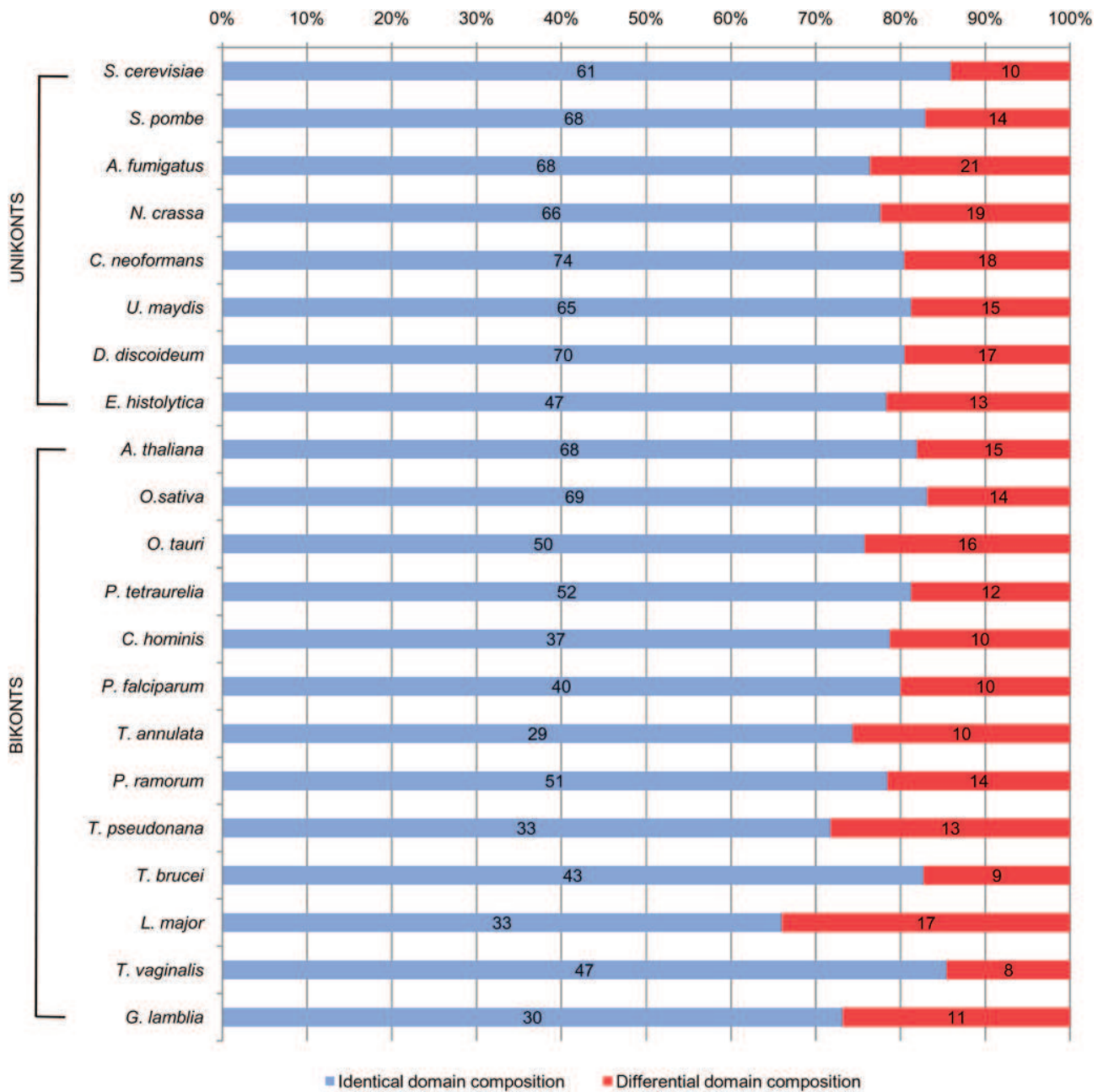


Figure 4. Domain conservation of the orthologues of the hamster midbody among eukaryotic lineages. Conservation of functional domains composition among mammalian components, and their orthologues in fungi (*A. fumigatus*, *Neurospora crassa*, *Cryptococcus neoformans*, *Ustilago maydis*, *Schizosaccharomyces pombe* and *S. cerevisiae*), amoebozoia (*D. discoideum* and *E. histolytica*), green plants and algae (*A. thaliana*, *Oryza sativa* and *Ostreococcus tauri*), alveolates (*Paramecium tetraurelia*, *Cryptosporidium hominis*, *Plasmodium falciparum*, and *Theileria annulata*), heterokonts (*P. ramorum* and *T. pseudonana*) and excavates (*T. brucei*, *Leishmania major*, *G. lamblia* and *Trichomonas vaginalis*). Orthologues in these 21 non-animal species having identical functional domains composition to their mammalian counterparts are represented by blue bars, whereas those having a different functional domain composition are represented by red bars.
doi:10.1371/journal.pone.0005021.g004

horizontal gene transfer between eukaryotes and prokaryotes (Figures S3, S4, S5, S6). Therefore, only five of the midbody components inferred to have existed in LECA were probably present also in prokaryotic ancestors. Although this was a minimal set since, once again, very divergent prokaryotic homologues might have escaped detection, this small number indicated that nearly all the proteins present in LECA appeared after the emergence of the eukaryotic lineage.

Such a small number of orthologues detected in prokaryotes does not mean that all midbody components completely originated *de novo* in eukaryotes. Indeed, the analysis of the protein domain composition of the hamster midbody proteins showed that only half of them were composed of domains that are specific to eukaryotes (i.e. domains that originated after the divergence of eukaryotes and prokaryotes) (Figure 5A). Interestingly, this proportion did not change when the analysis was restricted to

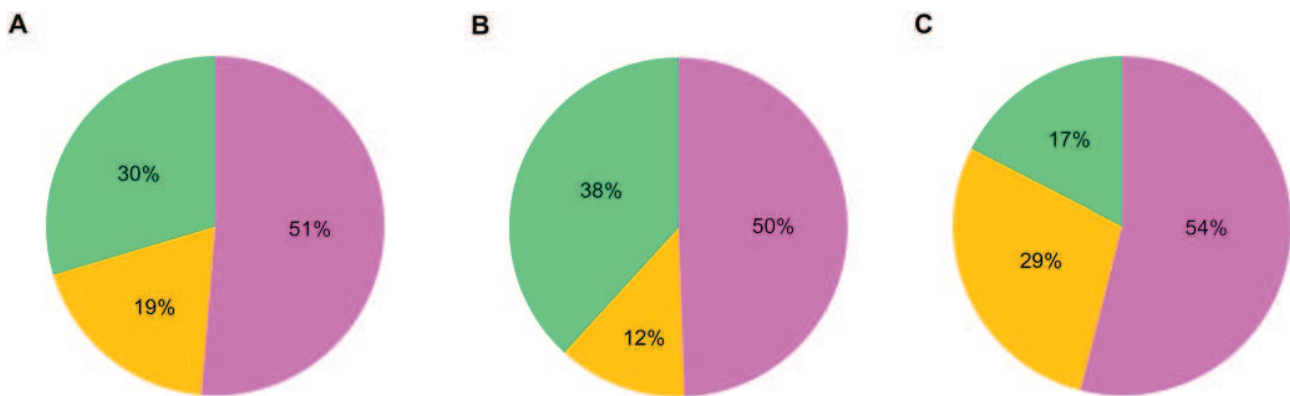


Figure 5. Functional domain composition of the mammalian midbody proteins. Pink sections represent proteins composed only of eukaryotic-specific functional domains (i.e. they have no homologues in prokaryotes, meaning that they probably appeared after the divergence of eukaryotes and prokaryotes); green sections represent proteins exclusively composed of domains shared by prokaryotes and eukaryotes (i.e. suggesting that these domains may have a more ancient origin since they may have arisen before eukaryotic divergence); yellow sections represent proteins composed by a combination of both types of domains. (A) Domain composition of 152 components of the mammalian midbody having detectable functional domains. The three remaining proteins have no functional domain detected above threshold. (B) Domain composition of 88 components inferred to be present in LECA having detectable functional domains. (C) Domain composition of the 64 Unikont-specific components. doi:10.1371/journal.pone.0005021.g005

the 91 components already present in LECA or to the 64 components of more recent origin (i.e., those that originated in the lineage leading from LECA to the present-day mammals) (50% and 54%, respectively, Figure 5B and 5C). By contrast, the proportion of components composed exclusively of domains that are also found in prokaryotes (i.e., ancient domains probably present in the ancestor of eukaryotes and prokaryotes) gone down from 38% to 17% (Figure 5B and 5C), whereas the proportion of components composed of a combination of both types of domains showed an opposite trend, going up from 12% to 29% (Figure 5B and 5C, respectively). This implied that only half of the hamster midbody proteins were actually *de novo* eukaryotic innovations whereas the other half derived from tinkering of more ancient protein domains, sometimes in combination with recent ones.

Conclusion

Although still scarce, some phylogenomic studies of EMC involved in major eukaryotic processes have suggested that LECA was already a highly complex and modern-like eukaryotic cell. In fact, it likely had a nuclear pore complex [13,14] and, therefore, a nuclear membrane [14], a spliceosome thought to be similar in complexity to contemporary ones [15], and the basic architecture of the vesicle trafficking system [40,41]. The idea of a LECA with complex intracellular structures is in agreement with the estimation of the number of genes present in this ancestral organism (more than 4100 [42]), suggesting that it had a rather large genome. Our analysis provided additional evidence for the hypothesis of a nearly-modern eukaryotic LECA, because we showed that it likely possessed a large multiprotein complex involved in the last steps of cytokinesis, from which the various eukaryotic present-day related structures derived. The future availability of genomic and proteomic data for diverse eukaryotic lineages will allow generalizing phylogenomic studies to other eukaryotic cellular structures. These analyses are expected to provide a more precise picture of LECA, as well as of the evolution of the different eukaryotic lineages. Finally, besides evolutionary inferences, an important aspect of this kind of work is the identification of components of cellular complexes in eukaryotic lineages for which data are inexistent or sparse, thus providing candidates for experimental tests in these groups.

Materials and Methods

Data set construction

Homologues of each of the 160 proteins identified in the hamster midbody proteome were retrieved from the *nr* database at the NCBI using the BLASTp program version 2.2.18 [43]. Additional BLASTp searches were performed using various seeds in order to retrieve divergent homologues that escaped detection in the first search. This multiple-step BLASTp approach was used to ensure that each gene family investigated was exhaustively sampled. Homologues from *Thalassiosira pseudonana* and *Phytophthora ramorum* were retrieved from the eukaryotic ongoing genome projects at the NCBI using the tBLASTn [43]. For each of the 160 assembled datasets, multiple alignments were done with ClustalW version 1.83 [44], T-COFFEE version 4.45 [45] and MUSCLE 3.6 [46]. The best alignment (in terms of maximal length and minimal number of gaps) was kept for further analyses. All the alignments were edited and manually refined using the ED program from the MUST package [47]. Regions where homology between sites was doubtful were manually removed from the datasets before phylogenetic analyses using the program NET from the MUST package. All datasets are available on request from LE or CB.

Phylogenetic analyses

For each dataset, Maximum Likelihood (ML) phylogenetic trees were computed with PHYML using the JTT model and a gamma correction to take into account the heterogeneity of evolutionary rates across sites (4 discrete classes of sites, an estimated alpha parameter and an estimated proportion of invariable sites) [48]. The robustness of each branch was estimated by the non-parametric bootstrap procedure implemented in PHYML (100 replicates of the original dataset and the same parameters). We carried out a first round of phylogenetic analyses to identify the orthologues of each mammalian midbody component. Based on these phylogenies, we selected the orthologues present in 32 eukaryotes for which complete (or nearly complete) genome sequences were available and representative of the diversity of this domain (Figure S1). Then, a second round of phylogenetic analyses was performed on these orthologues.

Inference of the number of components present at each node of the eukaryotic phylogeny

We applied a simple presence/absence parsimony criterion minimising the occurrence of horizontal gene transfers between eukaryotic lineages for inference. This implies that the presence of an orthologue of a mammalian midbody component in one representative of a non-mammalian lineage was interpreted as the existence of the component in their last common ancestor. The use of more stringent criteria (such as the presence of orthologues in representatives of at least two or three non mammalian-lineage) did not significantly change the number of components inferred at each node (not shown).

Analysis of functional domain conservation

The analysis of functional domains was carried out using the HMMER package [49]. First, we identified the functional domain composition of all the eukaryotic orthologues of the components of the mammalian midbody by performing hmmpfam searches against a local database of HMM profiles (pfam database [50] – version 21.0). HMM profiles having e-values lower than 0.1 were considered as significant if the corresponding domains did not overlap in the protein sequences.

Search for functional domains shared with prokaryotes

The hmmpfam program was used to extract each significant HMM profile corresponding to functional domains previously identified in mammalian midbody components. A survey for sequences matching these profiles on a local database of protein sequences from the 730 complete prokaryotic genomes available in June 2008 at the NCBI was done using the hmmpfam program.

Supporting Information

Figure S1 Phylogenetic tree of Eukaryotes. A consensus phylogeny of eukaryotes showing the phylogenetic position of the 32 eukaryotic representatives used in our study. The Last Eukaryotic Common Ancestor is indicated by a grey dot, whereas red stars indicate alternative positions for the origin of the midbody: red star "1" indicates a recent origin of the midbody (i.e. outbreak in the metazoan lineage), red stars "2" and "3" point to two possibilities for an ancient origin of the midbody (i.e. emergence before the ancestor of opisthokonts or before LECA).

Found at: doi:10.1371/journal.pone.0005021.s001 (3.04 MB TIF)

Figure S2 Phylogeny of actin. Maximum Likelihood (ML) tree of the actin homologues present in the 32 eukaryotic lineages studied (360 positions analysed). Numbers at nodes represent Bootstrap Values (for clarity only those greater than 50% are shown). The scale bar represents the average number of substitutions per site.

Found at: doi:10.1371/journal.pone.0005021.s002 (4.68 MB TIF)

Figure S3 Phylogeny of flotillin 1. ML tree of the flotillin 1 (sec26) homologues present in the 32 eukaryotic lineages studied and in prokaryotes (186 positions analysed). Numbers at nodes represent Bootstrap Values (for clarity only those greater than 50% are shown). The scale bar represents the average number of substitutions per site.

Found at: doi:10.1371/journal.pone.0005021.s003 (2.85 MB TIF)

Figure S4 Phylogeny of gmx33/golph3. ML tree of the gmx33/golph3 (sec29) homologues present in the 32 eukaryotic lineages studied and in prokaryotes (105 positions analysed). Numbers at nodes represent Bootstrap Values (for clarity only those greater

than 50% are shown). The scale bar represents the average number of substitutions per site.

Found at: doi:10.1371/journal.pone.0005021.s004 (2.24 MB TIF)

Figure S5 Phylogeny of alpha-tubulin. ML tree of the alpha-tubulin homologues present in the 32 eukaryotic lineages and in prokaryotes (318 positions analysed). Numbers at nodes represent Bootstrap Values (for clarity only those greater than 50% are shown). The scale bar represents the average number of substitutions per site.

Found at: doi:10.1371/journal.pone.0005021.s005 (4.43 MB TIF)

Figure S6 Phylogeny of beta-tubulin. ML tree of the beta-tubulin homologues present in the 32 eukaryotic lineages and in prokaryotes (345 positions analysed). Numbers at nodes represent Bootstrap Values (for clarity only those greater than 50% are shown). The scale bar represents the average number of substitutions per site.

Found at: doi:10.1371/journal.pone.0005021.s006 (3.98 MB TIF)

Table S1 Proteome of the mammalian midbody. The proteins are classified according to the five functional categories defined by Skop et al. (actin associated proteins, kinases, microtubules associated proteins, secretory and membrane trafficking associated proteins, and other) [22]. The name and the Genbank accession number of each protein are provided. The components inferred to have been present in the ancestor of Metazoa, Fungi, Amoebozoa, Plantae, Alveolata, Heterokonta, Excavata, as well as those present in LECA are indicated by an "x". As act33 and kin6 were discarded from our phylogenetic analyses, their presence in the different ancestors was not determined (ND).

Found at: doi:10.1371/journal.pone.0005021.s007 (0.13 MB PDF)

Table S2 Functional domain composition of the mammalian midbody proteins. This table shows the domain composition of each of the 155 components of the hamster midbody. The name and the PFAM accession number of each domain are provided.

Found at: doi:10.1371/journal.pone.0005021.s008 (0.06 MB PDF)

Table S3 Comparison of the functional domain composition of mammalian midbody proteins with their non metazoan eukaryotic orthologues. Comparison with the domain composition of the orthologues found in *Aspergillus fumigatus*, *Neurospora crassa*, *Cryptococcus neoformans*, *Ustilago maydis*, *Saccharomyces cerevisiae*, *Schizosaccharomyces pombe*, *Dictyostelium discoideum*, *Entamoeba histolytica*, *Arabidopsis thaliana*, *Oryza sativa*, *Ostreococcus tauri*, *Paramecium tetraurelia*, *Cryptosporidium hominis*, *Plasmodium falciparum*, *Theileria annulata*, *Phytophthora Ramorum*, *Thalassiosira pseudonana*, *Trypanosoma brucei*, *Leishmania major*, *Giardia lamblia* and *Trichomonas vaginalis*. Asterisks indicate orthologues having the same domain composition as their mammalian counterparts; dollars designate orthologues displaying at least one difference with their mammalian counterparts; whereas "no" indicate that no orthologues are available.

Found at: doi:10.1371/journal.pone.0005021.s009 (0.27 MB PDF)

Table S4 Physical interactions between orthologues of the mammalian midbody components found in *Saccharomyces cerevisiae*. Interaction data was taken from a survey of the *Saccharomyces Genome Database* (SGD, <http://www.yeastgenome.org/>). For each of the 71 yeast orthologues, the standard names of the orthologues with which it interacts are indicated. Components marked by an asterisk are annotated as involved in cytokinesis, or in a biological process related to cell division.

Found at: doi:10.1371/journal.pone.0005021.s010 (0.04 MB PDF)

Table S5 Components of the midbody, fungal septum division and phragmoplast studied by Otegui et al. and their relations of orthology. Grey columns correspond to proteins studied by Otegui et al. in mammals, *Saccharomyces cerevisiae*, *Schizosaccharomyces pombe* and plants [17]. Orthologues of these components present in the hamster midbody proteome are indicated by a cross. Dollars indicate paralogues of components of the hamster midbody.

Found at: doi:10.1371/journal.pone.0005021.s011 (0.04 MB PDF)

References

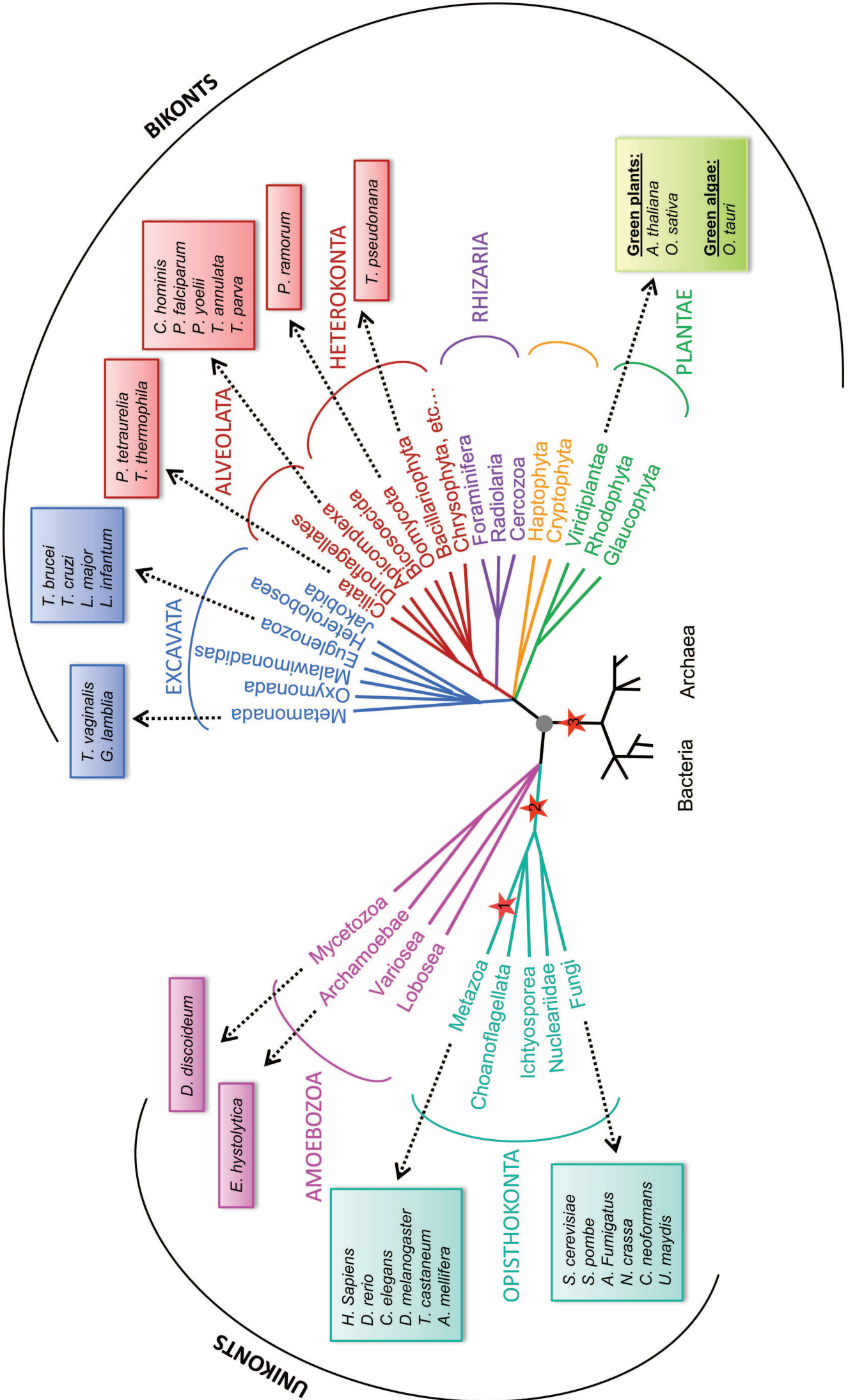
1. Franke WW (1970) On the universality of nuclear pore complex structure. *Z Zellforsch Mikrosk Anat* 105: 405–429.
2. Lamond AI (1993) The spliceosome. *Bioessays* 15: 595–603.
3. Kanoh J, Ishikawa F (2003) Composition and conservation of the telomeric complex. *Cell Mol Life Sci* 60: 2295–2302.
4. Staub E, Fizev P, Rosenthal A, Hinzmann B (2004) Insights into the evolution of the nucleolus by an analysis of its protein domain repertoire. *Bioessays* 26: 567–581.
5. Struhl K (1999) Fundamentally different logic of gene regulation in eukaryotes and prokaryotes. *Cell* 98: 1–4.
6. Fogel MA, Waldor MK (2006) A dynamic, mitotic-like mechanism for bacterial chromosome segregation. *Genes Dev* 20: 3269–3282.
7. Keith WN, Bilsland A, Evans TR, Glasspool RM (2002) Telomerase-directed molecular therapeutics. *Expert Rev Mol Med* 4: 1–25.
8. Li JB, Gerdes JM, Haycraft CJ, Fan Y, Teslovich TM, et al. (2004) Comparative genomics identifies a flagellar and basal body proteome that includes the BBS5 human disease gene. *Cell* 117: 541–552.
9. Pazour GJ (2004) Comparative genomics: prediction of the ciliary and basal body proteome. *Curr Biol* 14: R575–577.
10. Ohuse C, Yang H, Nozaki N, Goto S, Okazaki T, et al. (2004) Proteomics analysis of the centromere complex from HeLa interphase cells: UV-damaged DNA binding protein 1 (DDB-1) is a component of the CEN-complex, while BMI-1 is transiently co-localized with the centromeric region in interphase. *Genes Cells* 9: 105–120.
11. Eisen JA, Fraser CM (2003) Phylogenomics: intersection of evolution and genomics. *Science* 300: 1706–1707.
12. Eisen JA, Hanawalt PC (1999) A phylogenomic study of DNA repair genes, proteins, and processes. *Mutat Res* 435: 171–213.
13. Baptiste E, Charlebois RL, MacLeod D, Brochier C (2005) The two tempos of nuclear pore complex evolution: highly adapting proteins in an ancient frozen structure. *Genome Biol* 6: R85.
14. Mans BJ, Anantharaman V, Aravind L, Koonin EV (2004) Comparative genomics, evolution and origins of the nuclear envelope and nuclear pore complex. *Cell Cycle* 3: 1612–1637.
15. Collins L, Penny D (2005) Complex spliceosomal organization ancestral to extant eukaryotes. *Mol Biol Evol* 22: 1053–1066.
16. Flemming W (1891) Neue Beitrage zur Kenntnis der Zelle II. *Arch Mikros Anat* 37: 685.
17. Otegui MS, Verbrugge KJ, Skop AR (2005) Midbodies and phragmoplasts: analogous structures involved in cytokinesis. *Trends Cell Biol* 15: 404–413.
18. Eggert US, Mitchison TJ, Field CM (2006) Animal cytokinesis: from parts list to mechanisms. *Annu Rev Biochem* 75: 543–566.
19. Guertin DA, Trautmann S, McCollum D (2002) Cytokinesis in eukaryotes. *Microbiol Mol Biol Rev* 66: 155–178.
20. Glotzer M (2001) Animal cell cytokinesis. *Annu Rev Cell Dev Biol* 17: 351–386.
21. Mullins JM, McIntosh JR (1982) Isolation and initial characterization of the mammalian midbody. *J Cell Biol* 94: 654–661.
22. Skop AR, Liu H, Yates J 3rd, Meyer BJ, Heald R (2004) Dissection of the mammalian midbody proteome reveals conserved cytokinesis mechanisms. *Science* 305: 61–66.
23. Jurgens G (2005) Plant cytokinesis: fission by fusion. *Trends Cell Biol* 15: 277–283.
24. Neujahr R, Albrecht R, Kohler J, Matzner M, Schwartz JM, et al. (1998) Microtubule-mediated centrosome motility and the positioning of cleavage furrows in multinucleate myosin II-null cells. *J Cell Sci* 111(Pt 9): 1227–1240.
25. Robinson DN, Spudich JA (2000) Towards a molecular understanding of cytokinesis. *Trends Cell Biol* 10: 228–237.
26. Walther A, Wendland J (2003) Septation and cytokinesis in fungi. *Fungal Genet Biol* 40: 187–196.
27. Verma DP (2001) Cytokinesis and Building of the Cell Plate in Plants. *Annu Rev Plant Physiol Plant Mol Biol* 52: 751–784.
28. Jurgens G (2005) Cytokinesis in higher plants. *Annu Rev Plant Biol* 56: 281–299.
29. Broadwater ST (1992) Cytoskeleton and mitotic spindle in red algae. In: Menzel D, ed. *Cytoskeleton of the red algae*. Boca Raton Fla: CRC press. pp 93–112.
30. Lopez-Bautista JM, Waters DA, Chapman RL (2003) Phragmoplastin, green algae and the evolution of cytokinesis. *Int J Syst Evol Microbiol* 53: 1715–1718.
31. Pickett-Heaps JD, Gunning BES, Brown RC, Lemmon BE, Cleary AL (1999) The cytoplasm concept in dividing cell plant cells: cytoplasmic domains and the evolution of spatially organized cell division. *AM J Bot* 86: 153–172.
32. Sawitzky H, Grolig F (1995) Phragmoplast of the green alga *Spirogyra* is functionally distinct from the higher plant phragmoplast. *J Cell Biol* 130: 1359–1371.
33. Ohno S (1970) *Evolution by gene duplication*. Berlin: Springer.
34. Field C, Li R, Oegema K (1999) Cytokinesis in eukaryotes: a mechanistic comparison. *Curr Opin Cell Biol* 11: 68–80.
35. Striepen B, Jordan CN, Reiff S, van Dooren GG (2007) Building the perfect parasite: cell division in apicomplexa. *PLoS Pathog* 3: e78. doi:10.1371/journal.ppat.0030078
36. Bischoff JR, Plowman GD (1999) The Aurora/Ipl1p kinase family: regulators of chromosome segregation and cytokinesis. *Trends Cell Biol* 9: 454–459.
37. Terada Y (2001) Role of chromosomal passenger complex in chromosome segregation and cytokinesis. *Cell Struct Funct* 26: 653–657.
38. Demidov D, Van Damme D, Geelen D, Blattner FR, Houben A (2005) Identification and dynamics of two classes of aurora-like kinases in Arabidopsis and other plants. *Plant Cell* 17: 836–848.
39. Ke YW, Dou Z, Zhang J, Yao XB (2003) Function and regulation of Aurora/Ipl1p kinase family in cell division. *Cell Res* 13: 69–81.
40. Dacks JB, Field MC (2007) Evolution of the eukaryotic membrane-trafficking system: origin, tempo and mode. *J Cell Sci* 120: 2977–2985.
41. Koumandou VL, Dacks JB, Coulson RM, Field MC (2007) Control systems for membrane fusion in the ancestral eukaryote; evolution of tethering complexes and SM proteins. *BMC Evol Biol* 7: 29.
42. Makarova KS, Wolf YI, Mekhedov SL, Mirkin BG, Koonin EV (2005) Ancestral paralogs and pseudoparalogs and their role in the emergence of the eukaryotic cell. *Nucleic Acids Res* 33: 4626–4638.
43. Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, et al. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* 25: 3389–3402.
44. Thompson JD, Higgins DG, Gibson TJ (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res* 22: 4673–4680.
45. Notredame C, Higgins DG, Heringa J (2000) T-Coffee: A novel method for fast and accurate multiple sequence alignment. *J Mol Biol* 302: 205–217.
46. Edgar RC (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* 32: 1792–1797.
47. Philippe H (1993) MUST, a computer package of Management Utilities for Sequences and Trees. *Nucleic Acids Res* 21: 5264–5272.
48. Guindon S, Gascuel O (2003) A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst Biol* 52: 696–704.
49. Durbin R (1999) *Biological Sequence Analysis*. Cambridge, UK: Cambridge University Press.
50. Finn RD, Misty J, Schuster-Bockler B, Griffiths-Jones S, Hollich V, et al. (2006) Pfam: clans, web tools and services. *Nucleic Acids Res* 34: D247–251.

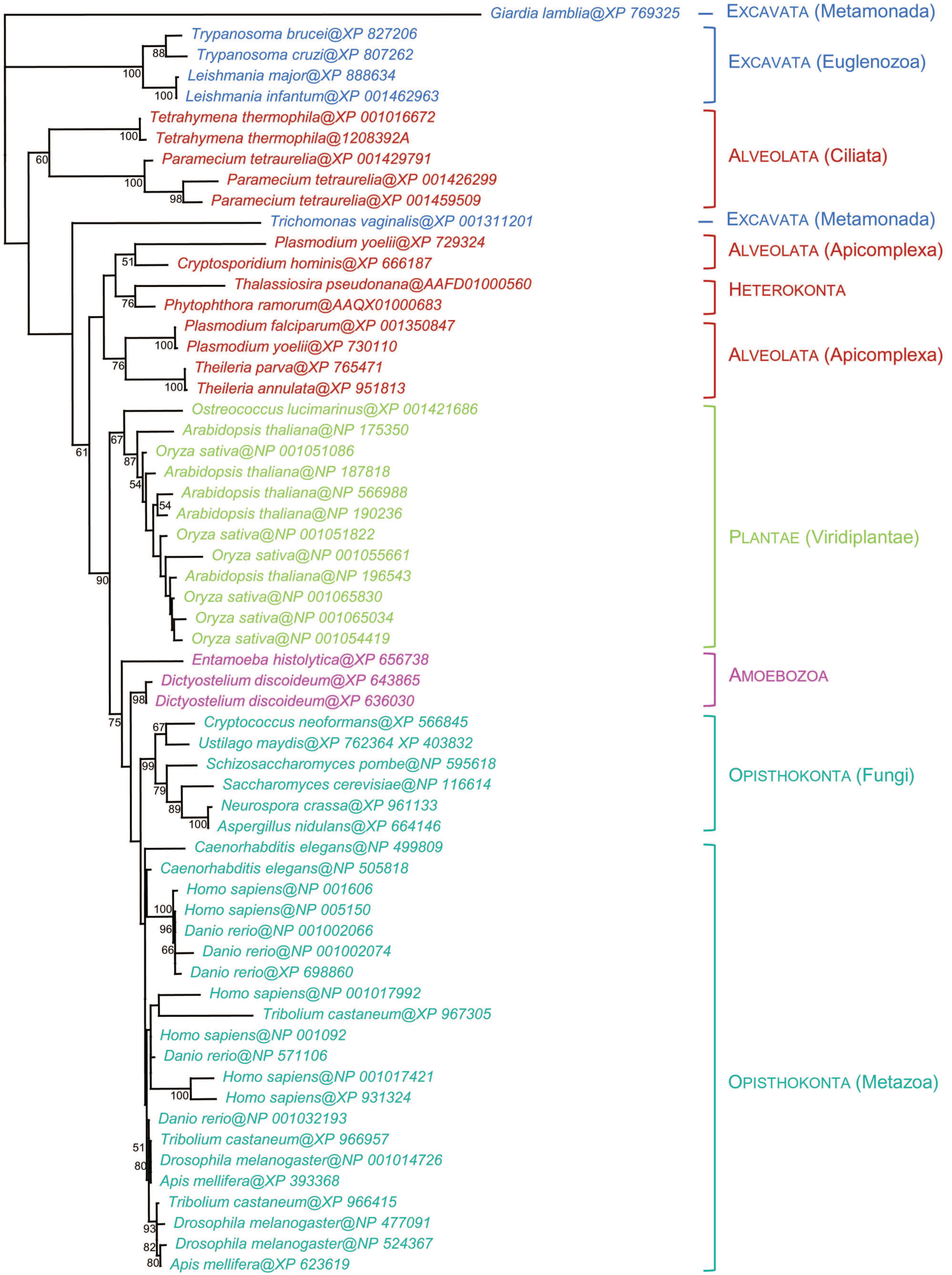
Acknowledgments

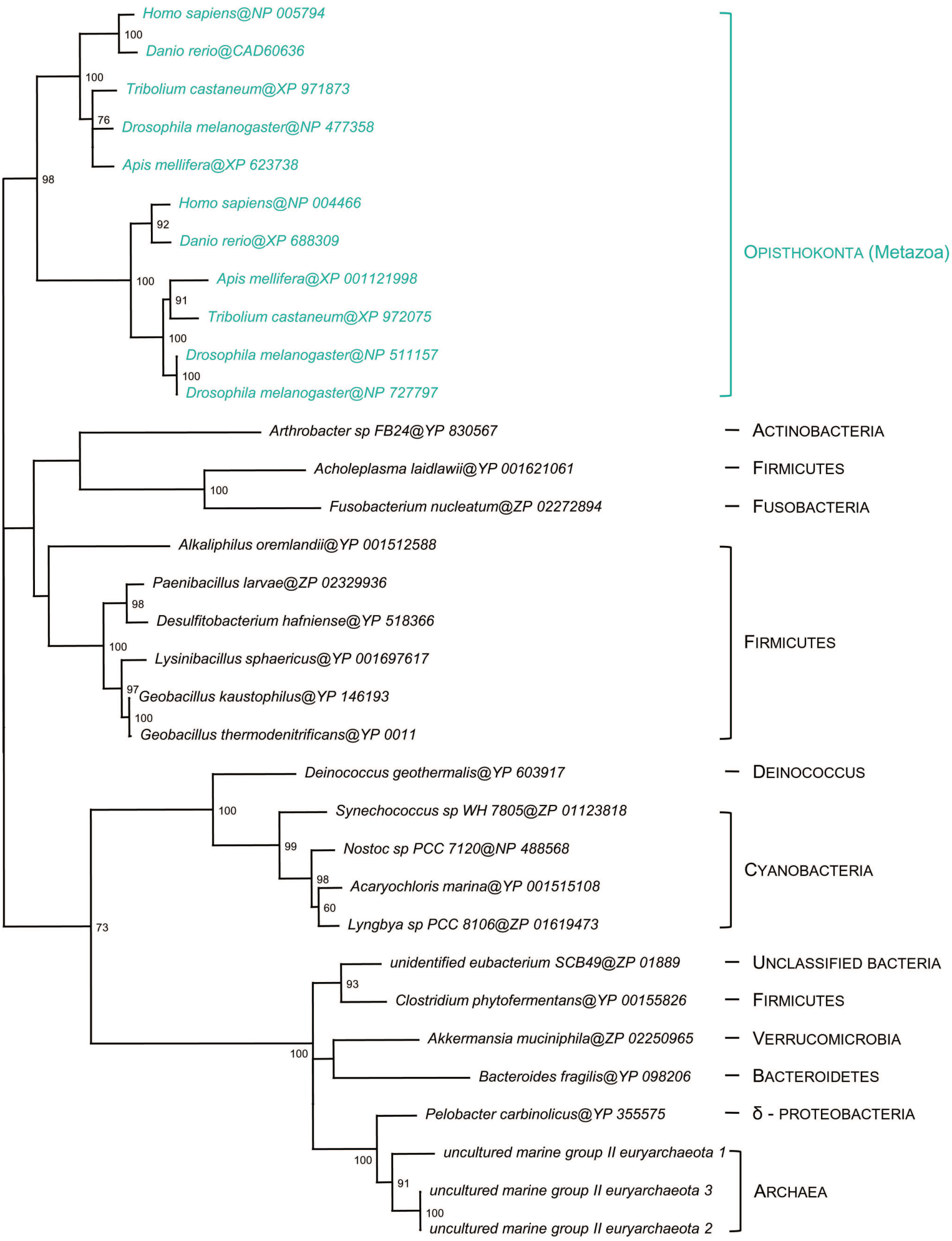
We wish to thank Purificación López-García and Simonetta Gribaldo for their critical reading of the manuscript and stimulating discussions.

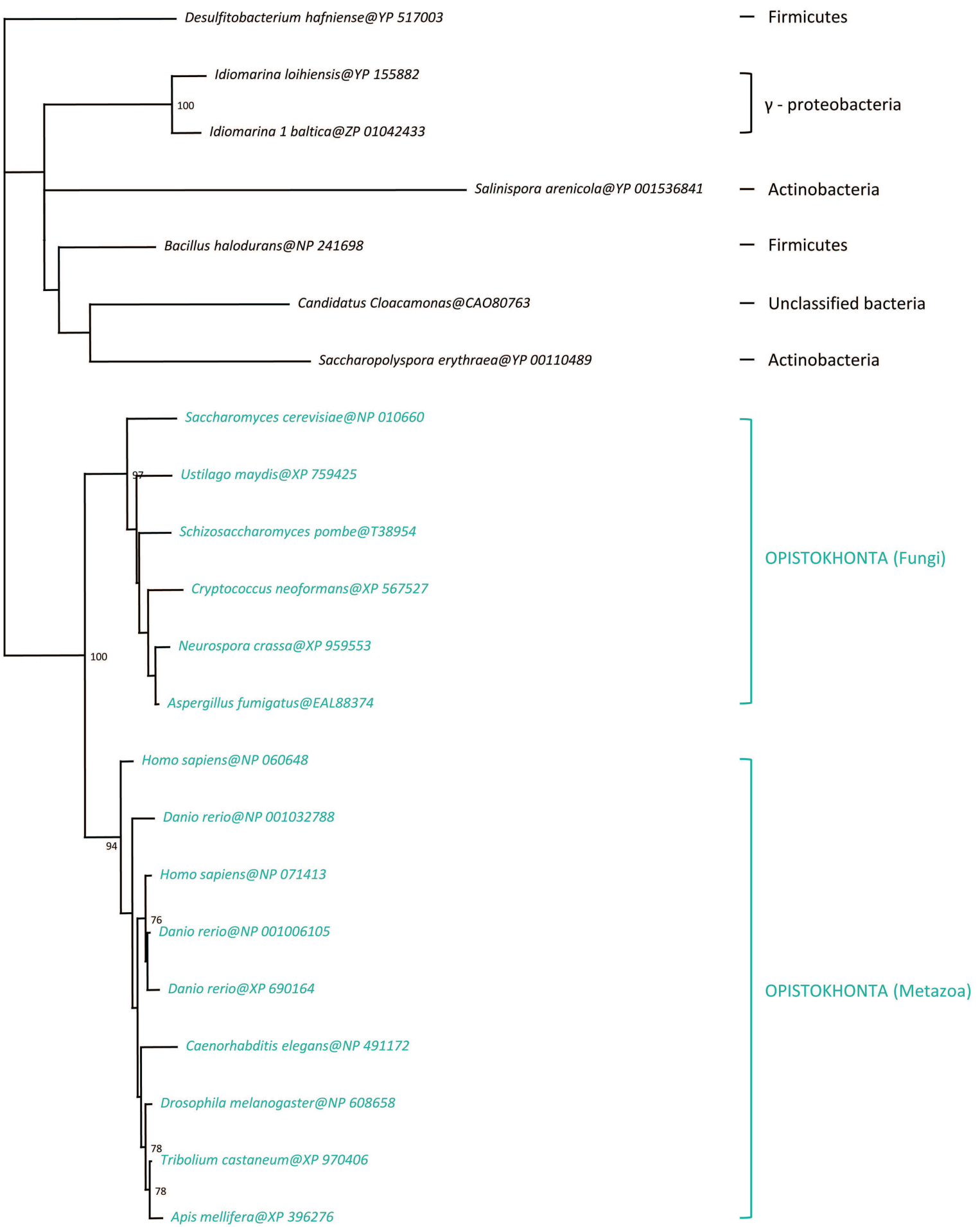
Author Contributions

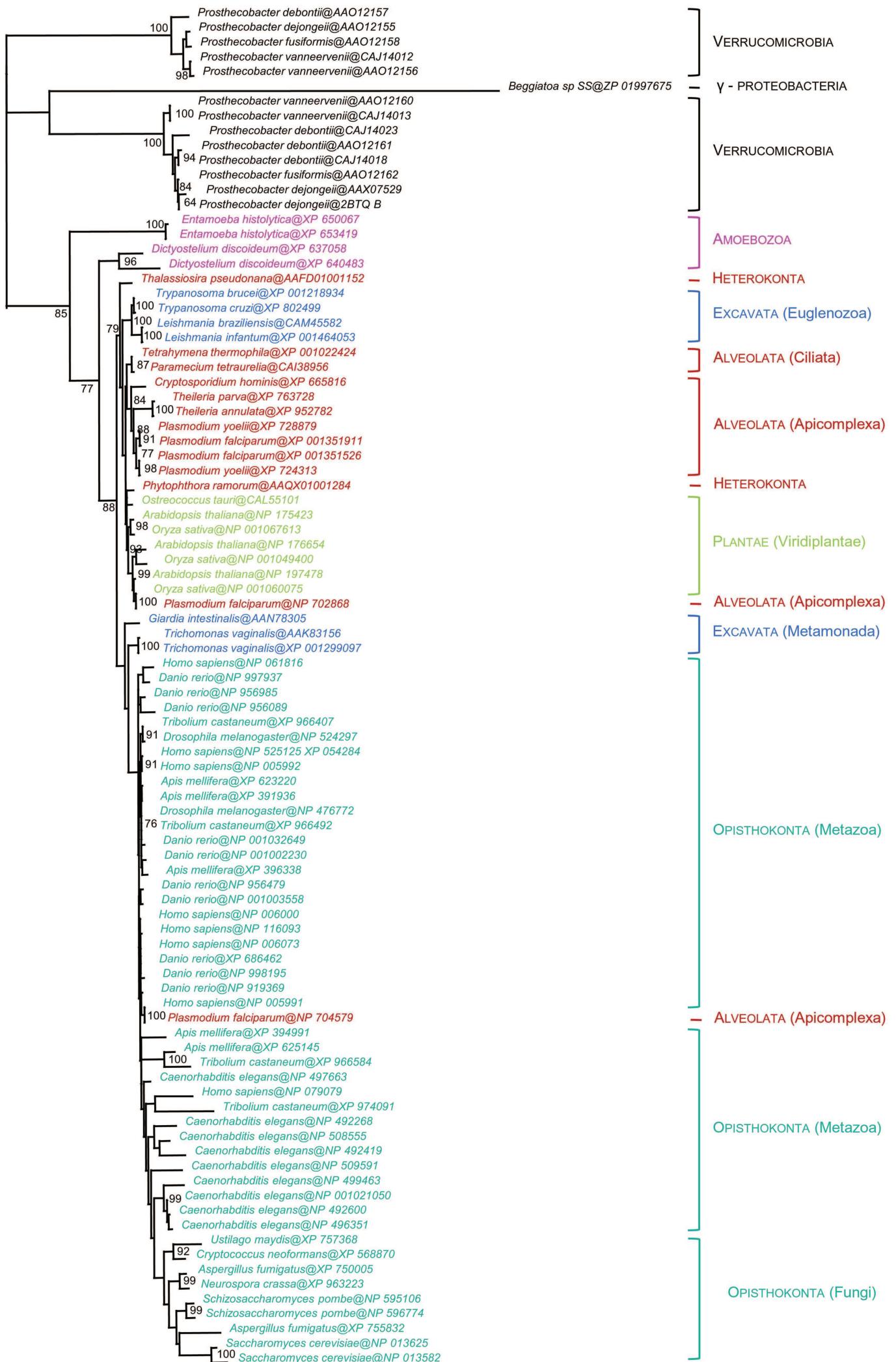
Conceived and designed the experiments: LE DM ET CBA. Performed the experiments: LE. Analyzed the data: LE DM ET CBA. Contributed reagents/materials/analysis tools: LE CBA. Wrote the paper: LE DM CBA.

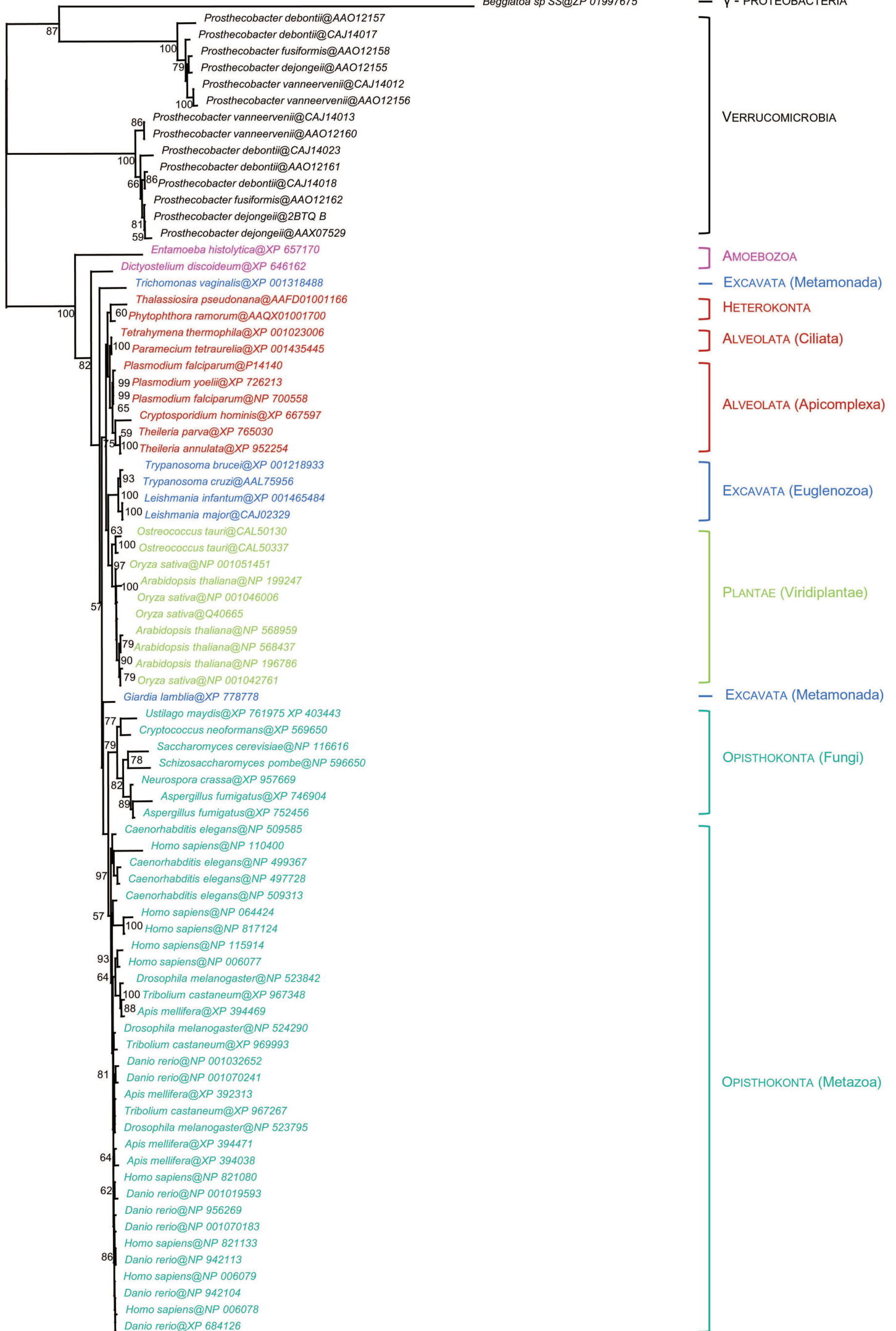












Supplementary Information Table S2. Functional domain composition of the mammalian midbody proteins.

Short name				Mammalian functional domains		
act1	Actin (PF00022.9)					
act2	Spectrin (PF00435.11)	SH3_1 (PF00018.18)	efhand (PF00036.22)	efhand_Ca_insen (PF08726.1)	SH3_2 (PF07653.7)	
act3	Actin (PF00022.9)					
act4	WD40 (PF00400.22)					
act5	Actin (PF00022.9)					
act6	G-patch (PF01585.13)	TFP11 (PF08697.1)				
act7	Cofilin_ADF (PF00241.10)					
act8	Myosin_head (PF00063.11)	IQ (PF00612.17)	Myosin_TH1 (PF06017.4)		SH3_1 (PF00018.18)	
act9	Myosin_N (PF02736.9)	Myosin_head (PF00063.11)	IQ (PF00612.17)	Myosin_tail_1 (PF01576.9)	Filament (PF00038.11)	
act10	WW (PF00397.16)	IQ (PF00612.17)	RasGAP (PF00616.9)		RasGAP_C (PF03836.5)	
act11	efhand (PF00036.22)					
act12	F_actin_cap_B (PF01115.8)					
act13	LIM (PF00412.12)					
act14	Filament_head (PF04732.4)	Filament (PF00038.11)				
act15	HEAT (PF02985.12)	Arm (PF00514.13)				
act16	CH (PF00307.21)	Spectrin (PF00435.11)	efhand (PF00036.22)			DUF1690 (PF07956.2)
act17	Band_41 (PF00373.8)	ERM (PF00769.9)				
act18	Cofilin_ADF (PF00241.10)					
act19	P21-Arc (PF04062.5)	ERM (PF00769.9)				
act20	Band_41 (PF00373.8)	Talin_middle (PF09141.1)				
act21	Band_41 (PF00373.8)		VBS (PF08913.1)			
act22	Septin (PF00735.9)					
act23	efhand (PF00036.22)	CH (PF00307.21)				
act24	BRCT (PF00533.16)	RhoGEF (PF00621.11)	PH (PF00169.19)			
act25	Band_41 (PF00373.8)	FA (PF08736.1)	SAB (PF04382.3)			
act26	CH (PF00307.21)	Filamin (PF00630.9)				
act27	CH (PF00307.21)	Spectrin (PF00435.11)	PH (PF00169.19)			
act28	CH (PF00307.21)	Spectrin (PF00435.11)	efhand (PF00036.22)	efhand_Ca_insen (PF08726.1)		
act29	P34-Arc (PF04045.5)					
act30	CH (PF00307.21)	Calponin (PF00402.8)				
act31	Dif_GBD (PF06371.4)	Dif_FH3 (PF06367.7)	FH2 (PF02181.13)			
act32	Ras (PF00071.12)	Air (PF00025.11)				
act34	Band_41 (PF00373.8)	ERM (PF00769.9)				
act35	Vinculin (PF01044.9)	MCPsignal (PF00015.11)	ArgJ (PF01960.9)			
act36	DUF1899 (PF08953.1)	WD40 (PF00400.22)	DUF1900 (PF08954.1)			
act37	Septin (PF00735.9)	MMR_HSR1 (PF01926.13)				
act38	Profilin (PF00235.9)					
act39	TBPP (PF07106.3)					
act40	Extensin_2 (PF04554.4)	CH (PF00307.21)				
act41	PDZ (PF00595.14)	SH3_2 (PF07653.7)				
act42	efhand (PF00036.22)					
act43	PDZ (PF00595.14)	LIM (PF00412.12)				
act44	Actin (PF00022.9)					
act45	BAR (PF03114.8)	RhoGAP (PF00620.17)				
act46	BTB (PF00651.21)	BACK (PF07707.5)	Kelch_1 (PF01344.15)			
kin1	Kinase (PF00069.15)					
kin2	Kinase (PF00069.15)					
kin3	CK_II_beta (PF01214.9)					
kin4	Kinase (PF00069.15)					
kin5	CKS (PF01111.9)					
kin7	Kinase (PF00069.15)	Pkinase_C (PF00433.14)	ATG16 (PF08614.1)	C1_1 (PF00130.12)	PH (PF00169.19)	CNH (PF00780.12)
kin8	No domain detected above threshold					
kin9	Kinase (PF00069.15)	HEAT (PF02985.12)				
kin10	Nucleoplasmin (PF03066.5)					
kin11	Kinase (PF00069.15)	POLO_box (PF00659.9)				
kin12	Metallophos (PF00149.18)					
kin13	HEAT (PF02985.12)					
kin14	Metallophos (PF00149.18)					
kin15	LRR_1 (PF00560.23)					
kin16	LRR_1 (PF00560.23)					

sec192021	Dynamlin_N (PF00350.13)	Dynamlin_M (PF01031.10)	GED (PF02212.8)			
sec22	HEAT (PF02985.12)					
sec23	Dynamlin_N (PF00350.13)	efhand (PF00036.22)				
sec24	BAR (PF03114.8)	SH3_1 (PF00018.18)		SH3_2 (PF07653.7)		
sec25	HATPase_c (PF02518.16)	HSP90 (PF00183.9)				
sec26	Band_7 (PF01145.15)	Floctilin (PF03149.6)				
sec27	PDZ (PF00595.14)					
sec28	C2 (PF00168.20)					
sec29	GPP34 (PF05719.2)					
sec30	GRASP55_65 (PF04495.4)					
sec31	G-alpha (PF00503.10)	Arf (PF00025.11)				
sec32	Spectrin (PF00435.11)	RhoGEF (PF00621.11)				
sec33	C2 (PF00168.20)	Fer1 (PF08151.2)		FerB (PF08150.2)		
sec34	Myosin_head (PF00063.11)	IQ (PF00612.17)				
sec35	NIPSNAP (PF07978.3)					
sec36	CDC48_N (PF02359.8)	CDC48_2 (PF02933.7)				
sec37	efhand_like (PF09279.1)	PI (PF00388.9)			PLC-beta_C (PF08703.1)	
sec38	GDI (PF00996.9)					
sec39	Ras (PF00071.12)					
sec40	Kinesin (PF00225.13)					
sec41	Ras (PF00071.12)					
sec42	WD40 (PF00400.22)					
sec43	WD40 (PF00400.22)					
sec44	zf-Sec23_Sec24 (PF04810.5)	Sec23_trunk (PF04811.5)				
sec45	WD40 (PF00400.22)	Extensin_2 (PF04554.4)		Sec23_BS (PF08033.2)	Sec23_helical (PF04815.5)	Gelsolin (PF00626.12)
sec46	Vps52 (PF04129.3)					
sec47	PX (PF00787.14)	Vps5 (PF09325.1)				
sec48	dsm (PF00035.15)					
sec49	<i>No domain detected above threshold</i>					
sec50	Torsin (PF06309.2)					
sec51	Vps35 (PF03635.7)					
sec52	Motile_Sperm (PF00635.16)					
sec53	Patched (PF02460.9)					

Supplementary Information Table S3.

Comparison with the domain composition of the orthologues found in various eukaryotes

Short name	<i>S. cerevisiae</i>	<i>S. pombe</i>	<i>Cryptococcus neoformans</i>	<i>Ustilago maydis</i>	<i>Dictyostelium discoideum</i>	<i>Entamoeba histolytica</i>	<i>A. thaliana</i>	<i>O. sativa</i>	<i>Ostreococcus tauri</i>	<i>Paramecium tetraurelia</i>	<i>Cryptosporidium hominis</i>	<i>Plasmodium falciparum</i>	<i>Theileria annulata</i>	<i>Phytophthora ramorum</i>	<i>Thalassiosira pseudonana</i>	<i>Trypanosoma brucei</i>	<i>Leishmania major</i>	<i>Trichomonas vaginalis</i>	<i>Giardia lamblia</i>	<i>Aspergillus fumigatus</i>	<i>Neurospora crassa</i>		
act1	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	
act2	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	
act3	*	*	*	*	*	*	*	*	*	*	no	no	no	*	no	*	*	*	no	*	*	*	
act4	*	*	*	*	*	*	*	*	no	*	*	no	no	*	no	*	no	*	no	*	*	*	
act5	*	*	*	*	*	*	*	*	*	*	no	no	no	*	no	*	*	*	no	*	*	*	
act6	no	*	*	*	*	no	*	*	*	no	no	no	no	*	\$	no	no	no	no	no	*	no	
act7	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	no	*	*	*
act8	\$	\$	\$	\$	\$	\$	\$	\$	\$	\$	\$	\$	\$	no	\$	\$	\$	no	no	\$	\$	\$	
act9	\$	\$	\$	\$	\$	\$	no	no	no	no	no	no	no	no	no	no	no	no	no	no	\$	\$	
act10	*	*	\$	\$	\$	\$	no	no	no	no	no	no	no	no	no	no	no	no	no	no	\$	\$	
act11	*	no	*	*	*	*	no	no	no	no	no	no	no	no	no	no	no	no	no	no	*	*	
act12	no	*	*	*	*	*	*	*	no	*	no	*	no	*	no	no	no	*	no	*	*	*	
act13	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	
act14	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	
act15	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	
act16	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	
act17	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	
act18	no	\$	\$	\$	no	\$	no	no	no	no	no	no	no	no	no	no	no	no	no	no	\$	\$	
act19	*	*	*	*	*	*	*	*	no	*	no	no	no	*	no	*	no	*	no	*	*	*	
act20	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	
act21	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	
act22	*	\$	*	*	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	*	*	
act23	\$	*	\$	*	*	no	\$	\$	no	\$	no	no	no	\$	no	no	no	*	no	\$	\$	\$	
act24	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	
act25	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	
act26	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	
act27	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	
act28	no	*	*	\$	\$	\$	no	no	no	no	no	no	no	\$	no	no	no	\$	no	\$	\$	\$	
act29	*	*	*	*	*	*	*	*	*	*	no	no	no	\$	no	*	no	*	no	*	*	*	
act30	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	
act31	*	\$	\$	\$	\$	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	\$	\$	
act32	\$	*	*	*	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	\$	*	
act34	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	
act35	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	
act36	*	*	\$	*	*	*	no	no	no	*	*	\$	\$	*	*	*	*	*	no	*	*	*	
act37	*	*	\$	*	no	no	no	no	no	*	no	no	no	no	no	no	no	no	no	no	*	*	
act38	*	*	*	*	*	*	*	*	no	*	no	no	no	no	no	*	*	*	no	*	*	*	
act39	*	*	*	no	no	no	*	*	*	*	no	no	no	*	no	*	no	*	*	no	no	no	

act40	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no
act41	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no
act42	*	*	*	*	*	*	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no
act43	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no
act44	*	*	*	no	*	no	*	*	*	no	*	no	no	no	no	no	no	no	no	*	no
act45	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no
act46	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no
kin1	*	\$	*	*	*	no	*	*	*	*	no	no	no	*	*	no	no	*	no	*	\$
kin2	*	*	*	*	*	*	\$	\$	*	no	*	*	\$	*	*	*	*	*	*	*	*
kin3	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*
kin4	*	\$	*	no	*	*	*	*	*	no	*	*	*	no	no	no	no	no	no	*	*
kin5	no	*	*	*	\$	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*
kin7	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no
kin8	no	no	no	no	no	no	no	no	no	no	*	no	no	no	no	no	no	no	no	no	no
kin9	\$	\$	\$	\$	*	\$	*	*	\$	no	\$	\$	\$	no	no	no	no	no	no	\$	\$
kin10	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no
kin11	*	*	*	*	*	*	no	no	*	*	no	no	no	*	\$	*	*	*	\$	\$	*
kin12	*	*	*	no	*	no	*	*	no	*	*	*	*	no	no	no	no	no	*	*	*
kin13	*	*	*	*	*	*	*	*	*	no	no	no	no	no	no	no	no	no	no	\$	\$
kin14	*	*	*	no	*	*	*	*	*	no	*	no	*	no	*	*	*	*	*	*	*
kin15	no	no	no	no	\$	*	*	*	no	no	no	no	no	no	\$	\$	\$	no	no	no	no
kin16	no	no	no	no	\$	*	*	*	\$	no	\$	\$	no	no	no	no	\$	no	no	no	no
kin17	*	*	*	\$	\$	no	*	*	\$	no	no	no	no	no	no	no	no	*	no	*	*
kin18	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no
mic1	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*
mic2	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*
mic3	\$	\$	\$	\$	*	\$	\$	\$	*	*	\$	\$	no	*	\$	no	no	\$	*	\$	\$
mic4	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no
mic5	no	no	*	\$	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	\$	\$
mic6	no	no	\$	no	\$	no	no	no	no	no	no	no	no	no	no	no	no	no	no	*	*
mic7	\$	\$	\$	\$	*	no	no	no	no	\$	no	\$	\$	\$	no	\$	\$	*	no	\$	\$
mic8	*	*	*	*	*	no	*	*	*	*	*	*	*	\$	\$	no	no	*	*	\$	*
mic9	*	*	*	*	*	no	*	*	*	*	no	no	no	*	*	*	*	*	*	*	*
mic10	\$	\$	\$	\$	no	no	\$	\$	*	*	no	no	no	no	no	*	\$	no	\$	\$	\$
mic11	*	*	*	*	*	*	*	*	*	*	no	no	no	no	no	*	\$	no	*	*	*
mic12	no	*	\$	\$	*	no	*	*	*	*	*	*	no	*	\$	*	\$	*	*	\$	\$
mic13	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no
mic14	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no
mic15	no	no	no	no	*	no	*	*	*	*	*	*	*	*	*	*	*	*	\$	no	no
mic16	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no
mic17	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no
mic18	no	no	*	\$	*	no	*	*	\$	\$	no	no	no	no	no	no	no	no	no	no	no
oth1	*	*	*	*	*	\$	*	*	*	\$	*	no	*	*	*	*	*	*	\$	*	*
oth2	no	no	no	no	no	\$	*	*	no	*	\$	*	\$	\$	\$	no	no	\$	no	no	no
oth3	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no
oth4	*	*	*	*	\$	no	\$	*	*	*	*	no	no	no	no	*	*	*	no	*	\$
oth5	*	*	*	*	*	*	*	*	\$	*	no	*	no	*	*	*	*	*	\$	*	*
oth6	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no
oth7	no	*	*	no	*	*	*	*	*	*	*	*	*	*	*	\$	*	*	no	*	*

oth8	no	no	no	no	*	no	*	*	no	no	*	*	no	*	no	no	no	no	no	no	no
oth9	*	*	*	*	*	*	*	*	*	\$	*	1	*	*	*	*	\$	*	\$	*	*
oth10	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no
oth11	*	*	*	*	*	*	\$	\$	no	no	no	no	no	*	*	no	no	*	no	*	*
oth12	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no
oth13	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no
oth14	no	*	*	*	*	*	*	*	*	*	no	*	*	*	*	*	*	no	no	*	*
oth15	no	no	no	no	*	no	no	no	no	no	no	no	\$	no	no	no	*	no	no	no	
oth16	*	*	*	*	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	*	*
oth17	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no
oth18	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no
oth19	no	no	*	*	no	no	*	*	*	no	*	*	no	*	no	no	*	no	no	*	*
oth20	*	*	*	*	*	*	*	*	*	*	*	*	*	no	no	no	\$	*	*	*	
oth21	no	no	no	no	*	no	*	*	no	no	no	no	no	no	no	no	no	no	no	no	no
oth22	no	*	\$	\$	\$	no	*	*	\$	\$	*	\$	*	*	\$	no	no	no	\$	\$	\$
oth23	\$	*	\$	no	no	no	\$	\$	\$	\$	no	\$	no	no	no	\$	\$	\$	\$	\$	no
oth24	no	*	*	*	no	no	\$	\$	\$	no	no	no	no	no	no	no	no	no	no	no	no
oth25	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no
sec1	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no
sec2	*	*	*	*	*	*	*	*	*	*	*	*	*	no	no	*	*	no	*	*	*
sec3	no	no	*	*	*	no	*	*	*	no	no	no	no	*	*	no	no	no	*	*	*
sec4	no	no	*	*	*	no	*	*	*	no	no	no	no	*	*	no	no	no	*	*	*
sec5	no	no	*	*	*	no	*	*	no	no	no	no	no	*	*	no	no	no	*	*	*
sec6	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no
sec7	no	no	*	*	*	*	*	*	no	*	*	*	*	*	no	*	*	*	*	*	no
sec8	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no
sec9	no	*	\$	\$	*	no	*	*	*	no	no	no	no	no	no	*	\$	no	no	*	*
sec1012	no	*	*	no	*	*	*	*	\$	no	no	no	no	*	no	no	no	*	no	*	*
sec11	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no
sec13	*	*	*	*	*	\$	\$	\$	\$	\$	no	no	no	\$	\$	no	no	\$	*	*	*
sec14	*	*	*	*	*	*	*	*	\$	\$	\$	\$	\$	*	\$	*	\$	\$	\$	\$	*
sec15	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no
sec16	*	*	*	*	*	\$	*	*	*	*	*	*	*	\$	*	*	*	*	*	*	*
sec17	\$	\$	*	\$	no	no	\$	\$	\$	no	no	no	no	\$	\$	no	no	no	no	*	*
sec18	no	no	no	no	*	no	*	*	no	*	no	no	no	*	no	no	\$	*	no	no	no
sec192021	*	*	*	*	*	*	*	*	\$	*	\$	*	*	*	\$	*	*	*	*	*	*
sec22	*	\$	no	no	\$	no	*	*	\$	no	no	no	no	no	no	no	no	no	no	*	*
sec23	no	no	no	no	*	\$	*	*	no	no	no	\$	no	*	\$	no	no	no	no	no	no
sec24	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no
sec25	*	*	*	*	*	*	*	*	*	*	*	*	no	no	no	*	*	*	*	*	*
sec26	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no
sec27	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no
sec28	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no
sec29	*	*	*	*	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	*	*
sec30	*	*	*	*	*	no	no	no	no	*	*	*	*	*	*	no	no	no	no	*	*
sec31	*	*	*	no	*	*	\$	*	no	no	no	no	no	no	no	no	no	no	*	*	
sec32	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no
sec33	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no
sec34	*	*	\$	\$	\$	\$	*	*	\$	*	\$	\$	\$	*	\$	\$	\$	no	no	\$	\$

Mammals		<i>S. pombe</i>		<i>S. cerevisiae</i>		Plants	
From Otegui 2005	Present in the mammalian midbody proteome	From Otegui 2005	Orthologues of mammalian midbody components	From Otegui 2005	Orthologues of mammalian midbody components	From Otegui 2005	Orthologues of mammalian midbody components
AuroraB/IMK-1	Aurora kinase E (kin1)	Akk1p	x	Ipf1p	x	At-Aurora 1, 2	x
EB1	EB1 (mcb3)	Mal3p	x	Blm1p	x	AEBT (A, B, C)	x
KIFC1/CHO2, KIFC2/KIFC3	KIFC1/KAR3P HOMOLOG (mcb11) paralogs of KIFC1/KAR3P (mcb11)	Pkl1p Klp2p	x x	Kar3p	x	AKY1/AKata, AKC2/AKc2B, AK3/AKAc, AK4/AKc4D, AK5, AKCBP	no, x x, x x, no
BlmC	BIMC-KINESIN/VEG5 (mcb3)	Cur7p	x	Klp1p Cln6p	x \$	TKRP125, DcKRP12b-2	x x
PLK1-4	POLO-LIKE KINASE (kin11)	Pkc1p	x	Cdc5p	x	AMg24400	\$
Dynamitin II	Dynamitin (sec192/221)	Dnm1p	x	Dnm1p	x	DRP24b, ADG1P1/Phragmoplastin, ABL1A, ABL1E	\$ \$ \$ \$
ch-TOG/XMAP215 homolog	TOG/XMAP215 HOMOLOG (mcb18)	Dis1p Alp14p	\$ \$	Stu2p	\$	MORI1/GEH1	x
KIF4	KIF4 (mcb10)	-		-		FRAY1/ASc4/7820, ASc4/6170, ASc6/6930	x x x
CLIP-170/Reslin	CLIP-170 (mcb4)	Tlp1p		Blk1p		MGJIP-170	
MKLP1/CHO1, RabK6/MKLP2	MKLP1 (mcb16)	-		-		AT1g20060 (AT20H2.17)	
p34cdc2/CDK1	paralogues of CDC2 KINASE (kin4)	Cdc2p	\$	Cdc28p	\$	CDC2	\$
Fornin/mb1a		Bni1p		Cdc12p		AFH5	\$
TD-60		Plm1p		-		At1g19880	
Symaxin 2		-		-		KNOLLE/ASYP111	
Borealin/CDC8		-		-		-	
TACC1		-		-		-	
Cdc14a		Cdc14p, Clp1p/Fip1p		Cdc14p		-	
PRC1		Ase1p		Ase1p		MAP-65	
MgcRacGAP		Rgt7p		Rgt1p		AMg24500	
INCENP		Pic1p		Slt15p		-	
Survivin		Bir1p/Cut17p		Bir1p		-	

ARTICLE 2 :

The phylogenomic analysis of the ACP/C and its targets points to complex and modern-like control of the cell cycle in LECA

1 **The phylogenomic analysis of the Anaphase Promoting Complex and its**
2 **targets points to complex and modern-like control of the cell cycle in the**
3 **last common ancestor of eukaryotes**

4

5 Laura Eme¹, Aurélie Trilles¹, David Moreira², Céline Brochier-Armanet^{1,§}

6

7 ¹ Aix-Marseille Université, Laboratoire de Chimie Bactérienne – UPR CNRS 9043,
8 Marseilles, France

9 ²Unité d'Ecologie, Systématique et Evolution – UMR CNRS 8079, Université Paris-
10 Sud, Orsay, France

11

12 [§]Corresponding author

13

14 Email addresses:

15 LE: laura.eme@ifr88.cnrs-mrs.fr

16 AT: aurelie.trilles@etu.univ-provence.fr

17 CBA: celine.brochier@univ-provence.fr

18 DM: david.moreira@u-psud.fr

19

20 Keywords: Anaphase Promoting Complex; Cohesin Complex; Phylogeny;
21 Eukaryotes; LECA; Evolution

22

23 Running title: Phylogenomic analysis of the APC/C

24

25 **Abstract**

26 **Background**

27 The Anaphase Promoting Complex or Cyclosome (APC/C) is the largest member of
28 the ubiquitin ligase [E3] family. It plays a crucial role in the control of the cell cycle
29 and cell proliferation by mediating the proteolysis of key components by the
30 proteasome. APC/C is made of a dozen subunits that assemble into a large complex
31 of ~1.5 MDa, interacting with various cofactors and targets.

32 **Results**

33 Using comparative genomic and phylogenetic approaches, we showed that 24 out of
34 37 known APC/C subunits, adaptors/co-activators and main targets, were already
35 present in the Last Eukaryotic Common Ancestor (LECA) and were well conserved to
36 a few exceptions in all present-day eukaryotic lineages. The phylogenetic analysis of
37 the 24 components inferred to be present in LECA showed that they contain a
38 reliable phylogenetic signal to reconstruct the phylogeny of the domain Eucarya.

39 **Conclusions**

40 Taken together our analyses indicated that LECA had a complex and highly
41 controlled modern-like cell cycle. Moreover, we showed that, despite what is
42 generally assumed, proteins involved in housekeeping cellular functions may be a
43 good complement to informational genes to study the phylogeny of eukaryotes.

44

45

46 **Background**

47 The anaphase-promoting complex or cyclosome (APC/C) has been recently
48 characterized as a member of the ubiquitin ligase family (also referred as E3) (for a
49 recent review see [1]). E3s mediate in the transfer of one or several ubiquitin
50 monomers on a protein substrate in a two-step reaction involving at least three
51 partners. First, an ubiquitin-activating enzyme (E1) activates and transfers ubiquitin
52 to an ubiquitin-conjugating enzyme (E2). Next, E3 mediates the transfer of ubiquitin
53 from E2 to a lysine residue of the target protein. Both steps require ATP. Most E3s
54 are able to polyubiquitinate proteins by adding new ubiquitin monomers to the first
55 attached one [2]. Polyubiquitinated proteins are targeted to the 26S proteasome for
56 degradation, whereas mono-ubiquitinated proteins can be altered in their function or
57 subcellular location by proteins containing ubiquitin-binding domains [2]. E3s are
58 divided in several families according to the presence of signature motifs. Among
59 them E3s containing a HECT (Homologous to E6-AP Carboxyl Terminus) domain
60 receive ubiquitin from E2 before attaching it on the substrate, whereas E3s harboring
61 a RING (Really Interesting New Gene) finger-domain mediate the transfer of ubiquitin
62 directly from E2 to the substrate [2, 3]. RING finger E3s form the largest family and
63 may also contain a subunit with a Cullin domain [1]. Among them, the APC/C is
64 atypically large (1.5 MDa) and complex, being composed of one or several copies of
65 at least a dozen subunits and of various adaptors/co-activators (Figure 1) [4]. The
66 function of the APC/C has been extensively studied in animals and yeast, where it
67 was shown to have a critical role in cell cycle progression through the tight control of
68 degradation of key proteins (e.g. mitotic cyclins, anaphase inhibitors, spindle
69 proteins, regulatory kinases, inhibitors of DNA replication, etc.) [5, 6].

70 Electron microscopy observations, *in vitro* assays, genetic experiments and structural
71 studies have shed light on the composition, structural organization, assembly and

72 molecular activity of the APC/C (Table 1 and Figure 1) [1, 4, 7]. The APC/C core is
73 divided in three functional parts: i) the structural complex that serves as scaffold,
74 made of Apc1, Apc4 and Apc5 subunits; ii) the catalytic arm that houses the E2-
75 binding site, made of Apc10, and of the Apc2 and Apc11 proteins that contain the
76 Cullin and RING finger domains, respectively; and iii) the TPR arm that allows
77 positioning both the E2 and the substrate in order to promote the ubiquitin transfer.
78 This second arm is composed of four subunits (Apc3, Apc6, Apc7, and Apc8)
79 containing tetratricopeptide repeats (TPR) and of several accessory proteins (Apc9,
80 Apc12, Apc13, and Apc15, the last two being essential in *Schizosaccharomyces*
81 *pombe* [8]). It has been proposed that the TPR repeats interact with the isoleucine-
82 and arginine-rich motifs found in the C-terminal regions of adaptors/co-activator (see
83 [1] and references therein for additional details). The TPR arm may also contain
84 Apc16, a subunit recently reported by the MitoCheck consortium [9]. Finally, the
85 location of Apc14, a yeast essential subunit [8] remains undetermined.

86 The APC/C activity and specificity are modulated by several adaptors/co-activators
87 (Cdc20, Cdh1, Ama1, Mfr1, Rap and Cortex) (Table 1 and Figure 1) [7, 10]. These
88 are paralogous proteins containing WD-repeats that mediate the interaction between
89 the APC/C and the D-, KEN-, A- or O-boxes present on target substrates (see [11]
90 and references therein). Among those adaptors, Cdc20 and Cdh1 are the most
91 important, being directly involved in the activation and substrate selectivity of the
92 APC/C at different stages of the cell cycle (reviewed in [1, 6, 7]). The interaction of
93 the APC/C and either Cdc20 (in early-mid-mitosis) or Cdh1 (during late-mitosis and
94 G1-S transition) is strongly dependent of the high or low activity of Cdks (Cyclin
95 dependent kinases) [6]. Briefly, Cdc20 activates the APC/C during early mitosis once
96 the chromosomes are properly attached and bi-oriented at the metaphase plate
97 during a process known as the spindle assembly checkpoint [12]. The APC/C^{-Cdc20}
98 targets securins and cyclins B1 towards destruction by the proteasome. The
99 degradation of these two proteins promotes the activation of separases, which then

100 cleave the cohesin complex (CC) leading to the separation of sister chromatids and
101 the initiation of the anaphase [6, 12, 13]. During anaphase, the APC/C^{-Cdh1} targets
102 Polo-like kinase 1, Aurora kinases, mitotic cyclins and Cdc20 towards degradation
103 leading to the exit of mitosis. The APC/C^{-Cdh1} remains active during the G1/S phase
104 ensuring the degradation by the 26S proteasome of several inhibitors of DNA
105 replication, thus allowing the synthesis of DNA [7]. At the end of the S phase, the
106 increase of the activity of Cdks inhibits the interaction between Cdh1 and the APC/C
107 complex, precluding new rounds of DNA synthesis [1]. By contrast, other APC/C
108 activators seem to have more restricted roles: Ama1 is required for sporulation and
109 during the anaphase of meiosis I in budding yeast [7, 14]; Mfr1 acts at the end of
110 meiosis II in *S. pombe* [10]; Cortex encodes a putative *Drosophila melanogaster*
111 female meiosis-specific co-activator of the APC/C prior to the metaphase I arrest [15]
112 and, finally, Rap (retina aberrant in pattern) mediates the degradation of cyclins
113 during the development of eye imaginal discs in *D. melanogaster* [16].

114 If most of the APC/C studies have been carried out in yeast and animals, recent
115 experiments with the land plant *Arabidopsis thaliana* have allowed the identification
116 of 12 transcribed genes that are homologous to vertebrate and yeast APC/C subunits
117 and of eight Cdh1/Cdc20 homologues (Table 1) [17-20]. By contrast, very little
118 information is available for representatives of the other major eukaryotic lineages.
119 The only exception concerns the kinetoplastid species *Trypanosoma brucei*, shown
120 to encode seven APC/C subunit homologues in its genome (Table 1) [21]. The
121 apparent conservation of components of the APC/C in these few distantly related
122 eukaryotes opens the question of the origin and evolution of this atypically large and
123 complex ubiquitin-ligase. To tackle this issue, we carried out an in-depth analysis of
124 16 APC/C subunits and six adaptors/co-activators (Table 1) in all eukaryotic lineages
125 for which representatives with complete (or nearly complete) genome sequences are
126 available (Figure 2). We also included in our study several major direct or indirect
127 targets of APC/C, namely the separase, the securin, cyclins A and B, Cdks-1 and -2

128 and the nine components of the cohesin complex (Table 2). The phylogenomic
129 analysis of these proteins supports that most of them were present in the last
130 eukaryotic common ancestor (LECA), indicating that this organism likely possessed a
131 highly controlled cell cycle that may have been very similar to that of present-day
132 eukaryotes. Finally our analyses indicate that APC/C components and targets carry a
133 *bona fide* phylogenetic signal that can be used to trace back the evolutionary history
134 of the eukaryotic domain.

135

136 **Results and Discussion**

137

138 *Most APC/C components and main targets were present in LECA*

139 We used a phylogenomic approach [22] in order to study the origin and evolution of
140 APC/C and its main targets in eukaryotes. The first step consisted in the survey of
141 complete (or nearly complete) genomes available in public databases to retrieve
142 homologues of each component of this system. Working on complete genomes
143 ensures the rigorous inference of the presence or absence of homologues in each
144 genome. Then, phylogenetic analyses allowed inferring the origin and the
145 subsequent evolution of each component.

146 We searched for orthologues in 65 taxa representing the eukaryotic diversity (all
147 major groups were represented except for Rhizaria and Cryptophyta). More
148 precisely, our taxonomic sampling covered (i) Holozoa represented by Metazoa, and
149 their close relatives the Choanoflagellata and Capsaspora, (ii) Fungi, with three main
150 lineages: Dikarya, Chytridiomycota and Microsporidia and (iii) Apusozoa. Whereas
151 the position of Apusozoa remains uncertain, Metazoa, their unicellular allies and
152 Fungi represent the opisthokont lineage that together with Amoebozoa, form one of
153 the two major divisions of eukaryotes, the Unikonta [23]. The other major division, the
154 Bikonta, was represented in our study by genomes from Excavata, Heterokonta,

155 Alveolata and Ciliata (two main groups Alveolata), Haptophyta, and Viridiplantae and
156 Rhodophyta (two main Plantae lineages) (Figure 2).

157 At this step, it is interesting to notice that except for adaptors/co-activators and a few
158 other exceptions, we identified at most only one homologue of each APC/C
159 component and main target coding genes in each genome (Tables S1 and S2). In
160 addition, some of them were found only in very restricted sets of species. For
161 example, orthologues of Apc14 and of two subunits of the TPR arm (Apc9 and
162 Apc15) were present only in a few ascomycetes (Tables 1 and S1) suggesting that
163 they are recent innovations that emerged after the first diversification of Fungi (Figure
164 2). The TPR arm protein Apc16 may be even more recent because it was present
165 only in the two Gnathostomata representatives within metazoa (Table 1 and Figure
166 2). By contrast, based on ML and Bayesian phylogenetic analyses, we identified
167 orthologues of the other 12 APC/C subunits and of two adaptors/co-activators in at
168 least two bikont and two unikont major lineages indicating that they were likely
169 present in LECA (Table 1 and Figure 2). Accordingly their absence in any present-
170 day genome should be interpreted as secondary losses. This is for instance the case
171 of the six subunits composing the catalytic arm and the structural complex, for which
172 we detected homologues in all opisthokonts but choanoflagellates, which ancestrally
173 missed Apc11 and Apc5, suggesting ancient losses in this lineage. Similarly,
174 whereas the four main subunits of the TPR arm (Apc3, Apc6, Apc7 and Apc8) were
175 inferred to be present in the ancestor of opisthokonts, Apc7 was missing in
176 ascomycete and basidiomycete representatives suggesting a secondary loss in the
177 ancestor of these two fungal groups after their separation from chytrids. Regarding
178 the other proteins associated to the TPR arm, beside the case of Apc9, Apc15 and
179 Apc16 already mentioned, homologues of Apc12 and Apc13 were poorly represented
180 in opisthokonts. More precisely they were missing in choanoflagellates, *Capsaspora*,
181 most fungi and some animals. However, the presence of Apc12 and Apc13
182 orthologues also in some bikont lineages indicated that these subunits were present

183 in LECA and thus in the ancestor of opisthokonts. Accordingly, their poor taxonomic
184 distribution in this eukaryotic lineage results from convergent secondary losses
185 (Figure 2). Finally, orthologues of the two adaptors Cdc20 and Cdh1 were present in
186 all opisthokonts. Among them, the case of Microsporidia deserved attention. Indeed,
187 whereas only one APC/C subunit (Apc10, located in the catalytic arm) had previously
188 been reported in *Encephalitozoon cuniculi* [24], we additionally found orthologues of
189 one component of the structural complex (Apc1), three of the TPR arm (Apc3, Apc6
190 and Apc8) and of two adaptors/co-activators (Cdh1 and Cdc20) in genomes of four
191 representatives of this group of highly derived parasitic anaerobic fungi. The
192 conservation of at least one component of each functional part of the APC/C
193 suggested that a minimalist version of the APC/C might exist in Microsporidia. More
194 drastic losses were observed in the anaerobic parasite *Entamoeba histolytica* where
195 the absence of all but four components contrasted with the conservation of all 14
196 subunits and adaptors/co-activators inferred to be present in LECA in the second
197 amoebozoan studied (the free-living *Dictyostelium discoideum*) (Table 1 and Figure
198 2). Such massive losses were also observed for the excavate *Giardia intestinalis*.
199 However, in contrast with Microsporidia, the more reduced set of components and,
200 more precisely, the absence of all proteins composing the TPR arm appeared less
201 compatible with a fully operational APC/C system in these two anaerobic parasites.
202 In bikonts, orthologues of the 12 components inferred to be present in LECA were
203 also inferred to be present in the ancestors of Plantae and Heterokonta (Figure 2).
204 However, in red algae, the haptophyte *Emiliania huxleyi*, ciliates and most excavates,
205 a slightly more restricted set of proteins was observed (Table 1). Notably, none of
206 them harboured the Apc5 subunit of the structural complex, along with two
207 components of the TPR arm (Apc12 and Apc13), whereas we detected Apc4 only in
208 the ciliate *Tetrahymena thermophila* and the haptophyte *E. huxleyi*, and Apc7 in the
209 excavate *Trichomonas vaginalis* (Table 1). This suggested that the great similarity
210 observed in the distribution of APC/C components in these four major eukaryotic

211 lineages was likely due to convergent losses occurred during their evolution (Figure
212 2). The situation was completely different in Apicomplexa, the second main lineage of
213 alveolates in our dataset together with ciliates. While the three ciliates have retained
214 a large number of components inferred to be present in LECA (Table 1), most of
215 them have been lost in Apicomplexa (Table 1). More specifically, all APC/C proteins
216 were missing in *Babesia bovis* and *Theileria annulata*, whereas the remaining four
217 apicomplexa harboured only four or five of them. Surprisingly, the components that
218 were present in those organisms were diverse depending on the species (Table 1).
219 For instance, Apc10, Apc11 and Apc1 were found in *Toxoplasma gondii*, whereas the
220 two *Plasmodium* species contained orthologues of Apc10, Apc11, Apc3 and the
221 anaerobic *Cryptosporidium hominis* harboured Apc1, Apc11, Apc3, Apc6 and Apc8
222 (Table S1). The presence of nearly all components in ciliates indicated that massive
223 and differential losses occurred secondarily in Apicomplexa (Figure 2). As mentioned
224 above, such massive losses were also observed in the excavate *G. intestinalis* and in
225 the amoebozoan *E. histolytica* (Table S1). However, it is important to note that when
226 orthologues existed in these lineages, they showed highly divergent sequences
227 compared to those found in other eukaryotic lineages (not shown). It is thus possible
228 that orthologues of some components might have escaped detection (despite the use
229 of PsiBlast) because of their extreme degree of sequence divergence. In any case,
230 the possible massive losses or the high divergence of APC/C components both
231 suggested that important changes have occurred relatively recently in these parasitic
232 lineages, maybe as a consequence of their atypical cell division mechanism. This is
233 notably the case of *Theileria* that, acting like a disguised chromosome during host
234 cell division, inserts itself into both daughter cells by co-opting parts of the host cell
235 division machinery, in particular the host cell's microtubules to be pulled towards the
236 opposing ends of the dividing cell [25].

237 An interesting evolutionary pattern emerged from our analyses concerning the
238 APC/C adaptors/co-activators. Their phylogenies supported that only the two

239 paralogues Cdc20 and Cdh1 were present in LECA and conserved in nearly all
240 eukaryotic lineages (even in Microsporidia), whereas all the remaining co-activators
241 resulted from independent duplications that occurred recently in different eukaryotic
242 lineages (Figures S1 and S2). For instance, Rap and Cortex resulted from
243 duplications of Cdh1 and Cdc20, respectively, which occurred in *D. melanogaster*,
244 whereas Ama1 and Mfr1 derived from duplications of Cdc20 in Ascomycota and
245 Cdh1 in *S. pombe*, respectively. Within plants, our analyses confirmed the presence
246 of multiple Cdc20 and Cdh1 copies in *A. thaliana*, but also in other land plants (*Oryza*
247 *sativa*, *Selaginella moellendorffii* and *Physcomitrella patens*) (Table S1) [19].
248 Phylogenetic analyses suggested that the numerous homologues observed in these
249 plants arose from independent duplications (Figures S1 and S2). In *A. thaliana*, the
250 various Cdc20 and Cdh1 paralogues have been shown to be differently expressed
251 through the cell cycle and depending on cell types or tissues [19], suggesting that
252 subfunctionalization events occurred after the duplications. More intriguing was the
253 huge expansion of the repertory of adaptor/co-activators observed in the two ciliates
254 *T. thermophila* and *Paramecium tetraurelia*, for which we identified eight and ten
255 copies of Cdh1, respectively. Among excavates, *Leishmania* and *Trypanosoma*
256 genomes encoded only one adaptor/co-activator affiliated to the Cdc20 subfamily
257 (Table 1), whereas the genome of *Naegleria* (as in most eukaryotes) encoded one
258 Cdc20 and one Cdh1 copies. The genome of *Trichomonas* contained three
259 homologues but due to their great divergence we were unable to classify them as
260 Cdc20 or Cdh1 without ambiguity (Table S1). Finally, in *G. intestinalis* as in some
261 apicomplexa (*Babesia*, *Theileria* and *Toxoplasma*), we failed to detect any
262 adaptor/co-activator, reinforcing the hypothesis that their APC/C proteins have
263 experienced a very divergent and fast evolution.

264 Regarding the main APC/C targets, our phylogenetic analyses were not conclusive in
265 the case of cyclins A and B and Cdks-1 and -2 to determine whether they were found

266 in LECA or not because these proteins belonged to very large multigenic families
267 with complex evolutionary histories precluding a precise inference of their
268 evolutionary origin (not shown). For the remaining targets, our analyses allowed
269 inferring that the separase and the nine subunits composing the CC (i.e. the core
270 complex, associated proteins, the loading complex and the proteins involved in the
271 cohesion establishment) were present in LECA (Tables 2 and S2) have been
272 conserved in most eukaryotic lineages (Table 2 and Figure 2). The taxonomic
273 distribution of Smc1, Smc3, Scc1, Scc3 and Psd5 homologues was globally in
274 agreement with a previous study focus the analysis of 29 genes involved in meiosis
275 in eukaryotes [26]. In contrast, Scc4 (a protein composing the loading complex) and
276 Wpl1/Rad61 (a protein that forms together with Psd5 a complex involved in CC
277 maintenance [12, 13]) are present only in Viridiplantae and in Opisthokonta
278 suggesting convergent losses in other lineages (Table 2 and Figure 2). However, it
279 can also be speculated that these proteins are not under strong selective pressure,
280 as attested by the fast evolutionary rate of Wpl1/Rad61 found in
281 Saccharomycetaceae that are shorter and highly divergent compared to those of
282 metazoa (647 versus 1190 amino acids, 14% identity and 29% similarity between
283 yeast (NP_010297) and human (NP_055860) sequences). So it would even be
284 possible that they have been replaced by non-homologous proteins in other lineages.
285 The only targets of APC/C that were not inferred to be present in LECA are securins
286 that were found only in Metazoa and Fungi. However, even though they fulfil the
287 same function (preventing the separation of the two sister chromatids) through the
288 binding of separases that are homologous in metazoa and fungi, fungal securins are
289 not homologous to those from metazoa [27], suggesting again a non-homologous
290 replacement in one of these two groups (Table 1 and Figure 2).

291

292 *Functional data pointed to a nearly modern APC/C controlling the cell cycle in LECA*

293 Our phylogenomic analysis of the APC/C, its main adaptors/co-activators and targets
294 supported the hypothesis that most of the corresponding genes were already present
295 in LECA. It was therefore tempting to conclude that a complex similar to the one
296 involved in present-day regulation of the eukaryotic cell cycle was already present in
297 this ancestral organism. Supporting this hypothesis, nearly all the proteins under
298 study possessed conserved functional domains similar to those present in
299 components that were experimentally proved to be part of the APC/C or targeted by
300 this ubiquitin ligase (Tables S3 and S4). This indicated that all the identified
301 orthologues of APC/C components and targets had similar molecular functions.
302 Moreover, experimental data from plants and excavates have shown that most
303 components identified by our analysis were part of or targeted by the APC/C [18-21,
304 28-35]. This strongly suggested that they inherited their function from the ancestral
305 proteins present in LECA, and therefore that a nearly modern APC/C and control of
306 the cell cycle existed in LECA. This also suggested that, in lineages for which no
307 functional data were available, the orthologues that we have identified were likely
308 involved in the control of the cell cycle and, therefore, constituted interesting targets
309 for experimental work.

310 Although the origin of most APC/C components and targets could be traced back to
311 LECA and most of them have been conserved throughout the evolution of
312 eukaryotes, some component gains could also be observed. Most of them resulted
313 from gene duplications of adaptors/co-activators that occurred independently in
314 different lineages. This was in agreement with recent reports of new, and often
315 specific, activities of the APC/C in some eukaryotes (e.g. neuronal activity [36] or
316 early eye development [16], or in the progression of the endocycle in flowering plant
317 endosperm [18, 20]). This suggested that most of the APC/C evolution since LECA
318 has concerned the acquisition of new regulatory functions by increasing the
319 repertoire of adaptors/co-activators, even if we can not rule out the possibility that
320 adaptors/co-activators present in single copies in some lineages were (and still are)

321 able to interact with a larger spectrum of targets than their multiple-copies
322 counterparts. In addition to the classical activators/co-activators, a recent
323 interactomic study in *A. thaliana* suggested the presence of three novel APC/C
324 interactors specific to land plants that were not homologous to the Cdc20/Cdh1
325 family [34]. However, although they interacted with the APC/C, their biological
326 function (e.g. adaptor/co-activators, inhibitors, targets, etc.) has still not been
327 established. Nevertheless, this supported that lineage-specific innovations are
328 expected to be discovered when biological data on a broader diversity of eukaryotes
329 becomes available.

330 In contrast, we also observed that convergent events of streamlining occurred
331 secondarily in various lineages, like Apicomplexa, *G. lamblia* and *E. histolytica*. The
332 reasons explaining those massive loss events are unclear, though we could not
333 discard that, at least for a number of cases, they might be linked to an extreme
334 acceleration of their evolutionary rate beyond detection. This hypothesis was
335 supported by the high evolutionary rates exhibited by the few components still
336 harboured by Apicomplexa. However, the detection of APC/C subunits and targets in
337 fast evolving organisms, like *E. cuniculi* and *T. vaginalis*, rather suggested that most
338 of the missing components in *G. lamblia*, *E. histolytica* and Apicomplexa, reflected
339 true losses. In that case we could wonder whether the lost components have been
340 replaced by non homologous proteins that fulfil the same role or whether these
341 parasites recruit the APC/C components from their hosts. In both cases,
342 experimental investigations in these parasitic lineages will be useful to elucidate the
343 nature of their APC/C or even to discover putative divergent systems involved in the
344 control of the cell cycle that may provide interesting medical drug targets. Likewise, a
345 previous phylogenomic study of proteins involved in late cytokinesis revealed a
346 similar pattern of reductive evolution in these lineages [37]. Indeed, *E. histolytica*, *G.*
347 *lamblia* and Apicomplexa have undergone massive losses of proteins of the
348 cytokinesis machinery, including conserved ancient ones inferred to have been

349 present in LECA [37]. This information combined to our present analysis suggests
350 that major changes have occurred in various steps of the cell cycle in these parasitic
351 eukaryotes.

352

353 *Most components of the APC/C and targets are eukaryotic innovations*

354 Despite our extensive survey of public sequence databases, we did not identify any
355 homologue of the APC/C components in prokaryotes with two exceptions. This
356 indicated that this large E3 complex and its main targets are eukaryotic innovations
357 that emerged after the separation of this domain from prokaryotes but prior to its
358 diversification into the present-day eukaryotic lineages. The two exceptions, Smc1
359 and Smc3, are two paralogous proteins that are part of the core complex of the
360 cohesin complex (Table 1). According to previous reports and to their critical role in
361 higher-order chromosome organization and dynamics [38, 39], we identified
362 homologues of Smc1 and Smc3 in nearly all archaeal and bacterial lineages (not
363 shown). The lack of APC/C prokaryotic homologues was surprising because distant
364 homologues harbouring structures similar to eukaryotic ubiquitin, E1 and E2 exist in
365 prokaryotes [40] and because a *bona fide* homologue of the eukaryotic proteasome
366 has been described in Archaea and Actinobacteria [41, 42]. Moreover, it was recently
367 reported that homologues of the eukaryotic ubiquitination pathway are encoded in the
368 genome of the archaeon '*Candidatus Caldiarchaeum subterraneum*' [43], a relative of
369 the recently proposed phylum Thaumarchaeota [44]. This system is composed of a
370 cluster of four genes coding for the ubiquitin, E1-like and E2-like enzymes and a
371 small Zn RING finger protein [43]. The first three proteins are much more similar to
372 their eukaryotic counterparts than to the very distant homologues usually found in
373 prokaryotes. While no *bona fide* homologue of E3 enzymes has been identified in this
374 archaeon, it was proposed that the fourth protein might mediate the ligation of
375 ubiquitin as in eukaryotes [43]. Despite this recent discovery, our analyses together

376 with data from the literature suggested that the vast majority of APC/C components
377 and main targets were eukaryotic innovations that may have played a role in the
378 emergence and evolution of the complex cell cycle observed in this domain.

379

380 *APC/C components and main targets can be used to infer the phylogeny of*
381 *eukaryotes.*

382 Proteins inferred to have been present in LECA are valuable material to reconstruct
383 the characteristics of this ancestral organism (e.g. [37, 45-49]. In addition, they can
384 preserve a phylogenetic signal useful to infer the evolutionary history of eukaryotes.
385 Until now, most analyses dedicated to the reconstruction of the eukaryotic phylogeny
386 were based on the analysis of components of informational systems (those involved
387 in the transmission and expression of the genetic information). This was so because
388 most of the genes coding for these proteins present the advantage of being (i) part of
389 monogenic gene families allowing the easy identification of orthologous proteins, (ii)
390 slowly evolving, (iii) ancient and well conserved among life domains, and (iv) rarely
391 exchanged by horizontal gene transfers. Accordingly, they represent first choice
392 material to investigate ancient evolution in all domains of life [50-53]. Phylogenetic
393 studies of the eukaryotic domain did not escape this rule and most of them have
394 been largely based on the phylogenetic analysis of informational proteins (e.g., [51,
395 54-57]). By contrast, most proteins involved in housekeeping functions (operational
396 proteins) are considered to evolve faster than those of informational systems, and
397 thus to be less suitable to study ancient evolution. Moreover, they are often part of
398 large and complex protein families that have experienced numerous gene duplication
399 and loss events during their evolutionary history meaning that distinguishing between
400 orthologues and paralogues is difficult and requires fastidious preliminary analyses.
401 Consequently, although these proteins are more numerous than informational ones
402 and often of larger size, they have rarely been used to infer the ancient evolution of
403 eukaryotes. Nevertheless, typical datasets based on informational proteins have

404 been shown to be insufficient to robustly infer all the deep nodes of the global
405 eukaryotic phylogenies [58, 59]. Increasing the protein sampling is therefore
406 becoming as necessary as increasing the taxonomic sampling in order to fully
407 resolve the phylogeny of eukaryotes, meaning that new useful protein markers have
408 to be found among the conserved operational proteins.

409 Our phylogenetic analyses of the APC/C subunits and main targets showed that with
410 a few exceptions they are ancient proteins well conserved throughout the diversity of
411 present day eukaryotic lineages. Accordingly, they are potential suitable markers to
412 reconstruct global eukaryotic phylogenies. The maximum likelihood and Bayesian
413 phylogenetic analyses of individual components showed that they have retained
414 ancient phylogenetic signal despite the fact that some basal nodes of the inferred
415 phylogenies showed a poor resolution (a typical result in single-marker analyses)
416 (Data S1). To increase resolution, we built a supermatrix (3115 positions) by
417 concatenating the individual alignments of the 24 APC/C components and main
418 targets inferred to be present in LECA. To minimise potential long-branch attraction
419 (LBA) tree reconstruction artefacts [60], we discarded at this step the extremely fast
420 evolving and/or very partial sequences of *Enterocytozoon bieneusi* (Microsporidia),
421 *Plasmodium* spp. (Apicomplexa) and *B. hominis* (Heterokonta), all them having >60%
422 of missing data. We also discarded *T. vaginalis* (Excavata) because this species
423 contained multiple (and often divergent) copies of several APC/C components and
424 main targets. The resulting Bayesian and maximum likelihood trees recovered the
425 monophyly of each eukaryotic phyla and most often supported by high Bayesian
426 posterior probabilities (PP) and maximum likelihood bootstrap values (BV):
427 Discicristata (PP=0.98; BV=97%), Heterokonta (PP=1.0; BV=100%), Metazoa
428 (PP=1.0; BV=100%), Fungi (PP=0.87; BV=98%), Choanoflagellata (PP=1.0;
429 BV=100%), Viridiplantae (PP=1.0; BV=99%) and Alveolata (PP=0.95 and BV=78%)
430 (Figure 3). The most remarkable result was the good support retrieved for the
431 monophyly of Fungi, including the extremely fast evolving Microsporidia. The APC/C

432 proteins are therefore among the rare ones [61] that support the correct phylogenetic
433 placement of these parasites as members of the Fungi.

434 Regarding the relationships among the main eukaryotic groups, the sister-grouping of
435 the two choanoflagellates and *C. owczarzaki* was supported both by the Bayesian
436 (PP=0.88) and maximum likelihood (BV=87%) trees; these two groups being closely
437 related to Metazoa (namely, a support for the Holozoa with PP=0.97 and BV=100%).
438 This was in agreement with results from [62, 63], but not with those supporting the
439 emergence of *Capsaspora* at the base of the group composed of Choanoflagellata
440 and Metazoa [64, 65]. The phylogenetic signal contained in APC/C components and
441 main targets also supported the dichotomy between Opisthokonta (PP=0.91;
442 BV=100%) and of Unikonta, including the apusozoan *Thecamonas trahens*, albeit
443 with a weaker support (PP=0.81; BV=78%). Similarly, most of the relationships
444 among bikont lineages were poorly resolved (Figure 3). In particular, the monophyly
445 of Plantae was not recovered: green plants and algae appeared closely related to the
446 haptophyte *E. huxleyi* (PP=0.78; BV=81%), whereas the phylogenetic position of red
447 algae was not resolved. This agreed with other studies showing the difficulty to
448 retrieve the monophyly of Plantae, even with much larger datasets (e.g., [66]).
449 Finally, heterokonts formed the sister group of green plants and haptophytes
450 (PP=0.89; BV=100%).

451 Taken together, our phylogenetic analyses showed that despite the relative small
452 size of the supermatrix constructed with the APC/C components and main targets
453 (3115 positions), it contained an interesting phylogenetic signal that can be useful to
454 infer part of the ancient evolutionary history of eukaryotes. ~~As mentioned before, we
455 retrieved homologues of these proteins from complete or nearly complete genome
456 sequences in order to avoid possible problems due to the undesired inclusion of non-
457 orthologous sequences and to limit the amount of missing data in the supermatrix.~~
458 However, the large size of most of the APC/C subunits and main targets made
459 difficult the use of the abundant body of sequence data coming from the analysis of

460 expressed sequence tags (ESTs) and limited our taxonomic sampling. Nevertheless,
461 our dataset can now be employed as a reference to enrich it with assembled EST
462 sequences as well as with those from upcoming complete genome sequences from
463 various protist species. This limitation will likely apply to other operational proteins
464 candidates to be used as markers for the reconstruction of global eukaryotic
465 phylogenies but, as already discussed, their inclusion in multi-marker analyses
466 becomes crucial to infer several deep nodes that the traditional supermatrices of
467 informational proteins have failed to resolve.

468

469 **Conclusions**

470 Our analyses of the APC/C and its main targets showed that this complex system
471 was very likely present in LECA and has been conserved, to a few exceptions, all
472 along the diversification of the eukaryotic domain. This study provided first insights
473 into the mechanisms responsible of the control of the cell cycle in LECA, suggesting
474 that it was tightly regulated like in present-day eukaryotes. Finally we showed that the
475 components of the APC/C and its main targets can be good phylogenetic markers to
476 complement those used so far (mostly proteins involved in transcription and
477 translation). Indeed, the latter have proven not to be sufficient to fully resolve the
478 phylogeny of eukaryotes, making necessary to identify new complementary markers.
479 This will certainly be a difficult task that will require many analyses but we think that
480 the phylogenomic study of conserved cellular systems is a promising approach to
481 tackle this issue.

482

483 **Methods**

484 *Dataset assembly*

485 We used the 37 APC/C components and main targets identified in four opisthokont
486 species (*Homo sapiens*, *D. melanogaster*, *S. cerevisiae* and *S. pombe* [7, 11]), the

487 plant *A. thaliana* [17, 19] and the kinetoplastid *T. brucei* [21] (Table 1) to survey
488 public sequence databases. We identified homologues of these proteins using
489 BLASTp and PSI-BLAST (with default parameters) [67] in a subset of complete or
490 ongoing genomes representative of eukaryotic diversity available at the NCBI (Figure
491 2). To increase the taxonomic sampling, homologues of *Monosiga brevicollis*,
492 *Salpingoeca rosetta*, *Lottia gigantea*, *Nematostella vectensis*, *Helobdella robusta*,
493 *Daphnia pulex*, *Capsaspora owczarzaki*, *Batrachochytrium dendrobatidis*,
494 *Spizellomyces punctatus*, *Thecamonas trahens*, *Naegleria gruberi*, *Phaeodactylum*
495 *tricornutum*, *Aureococcus anophagefferens*, *Ostreococcus lucimarinus*,
496 *Physcomitrella patens*, *Chlorella vulgaris*, *Micromonas pusilla*, *Selaginella*
497 *moellendorffii* and *Emiliana huxleyi* were retrieved using the BLASTp and tBLASTn
498 programs from the JGI (<http://genome.jgi-psf.org/>), the Broad Institute
499 (<http://www.broadinstitute.org>) and the TBestDB database
500 (<http://tbestdb.bcm.umontreal.ca/>) [68]. In addition, homologues of two
501 representatives of Rhodophyta were retrieved from the *Galdieria sulphuraria* genome
502 project (<http://genomics.msu.edu/cgi-bin/galdieria/blast.cgi>) and the *Cyanidioschyzon*
503 *merolae* genome project (<http://merolae.biol.s.u-tokyo.ac.jp/blast/blast.html>). BLAST
504 outputs were examined by eye to identify homologues of each protein to avoid
505 applying an arbitrary cutoff on e-value or score. To ensure an exhaustive sampling of
506 homologues, we performed additional searches using as seeds homologues that
507 were identified at previous steps. The absence of any homologue in a given lineage
508 was systematically verified by hand using tBLASTn searches on the nucleotide
509 sequences of the corresponding complete genomes.

510 For each protein, the homologous sequences were gathered in a dataset and aligned
511 with MAFFT 6.833 [69]. All the resulting alignments were edited and manually
512 adjusted using the ED program from the MUST package [70]. Regions where
513 homology between sites was doubtful were manually removed from the alignments
514 before phylogenetic analyses. A supermatrix was built by concatenating the

515 alignments corresponding to the 24 APC/C components, adaptor/co-activators and
516 targets inferred to be present in LECA.

517

518 *Conserved functional domain search*

519 The identification of functional domains was carried out using the HMMER package
520 [71] and the HMM profiles of the Pfam database [72]. HMM profiles having e-values
521 lower than 0.1 were considered as significant.

522

523 *Phylogenetic analysis*

524 Phylogenetic trees were reconstructed for each single protein dataset, except for
525 Apc9, Apc14, Apc15, Apc16 and securin because of their very restricted number of
526 homologues. Maximum Likelihood (ML) phylogenetic trees were computed using
527 Treefinder [73] with the LG model [74] and a gamma correction to take into account
528 the heterogeneity of evolutionary rates across sites (4 discrete classes of sites and
529 estimated alpha parameter), as proposed by the model selection tool implemented in
530 Treefinder. The branch robustness of ML trees was estimated using the non
531 parametric bootstrap procedure (100 replicates of the original alignment)
532 implemented in Treefinder with the same parameters than for ML tree inference.
533 Bayesian trees were computed using MrBAYES 3.1.2 [75] with a mixed amino acid
534 substitution model and a gamma correction (four rate categories). The Markov chain
535 Monte Carlo search was run with 4 chains for 1,000,000 generations, with trees
536 being sampled every 100 generations (the first 2,500 trees were discarded as “burn-
537 in”). Individual Bayesian trees are provided as Data S1.

538 TreeFinder and PhyloBayes 3.2 [76] were used to perform ML and Bayesian
539 analyses on the concatenated dataset. Treefinder was run with the same parameters
540 than for the analysis of single datasets. Phylobayes was run using the CAT model
541 and a gamma correction with 4 rate categories [77]. For each dataset, two
542 independent chains were run for at least 10000 cycles, saving one tree in ten. The

543 first 300 trees were discarded as “burn-in”, and the remaining trees from each chain
544 in each dataset were used to test for convergence and compute the 50% majority
545 rule consensus tree.

546

547 *Inference of the origin of APC/C components and main targets*

548 The phylogenetic analysis of each individual component allowed distinguishing
549 orthologues that originated from speciation events from paralogues that resulted from
550 gene duplications. This point is important because inferring the evolutionary origin of
551 a protein requires the analysis of the evolutionary history of orthologues. To
552 determine the origin of each component and subunit, we used a parsimony criterion,
553 making the assumption that horizontal gene transfers between eukaryotes are rare.
554 Accordingly, the ancestral absence of a component in the ancestor of a taxonomic
555 group was inferred if no orthologues are found in any present-day representative of
556 the group (even in those that were not included in our taxonomic sampling). By
557 contrast, the presence of orthologues in representatives of two lineages was
558 interpreted as the existence of the corresponding component in their last common
559 ancestor. Similarly, the presence of a component in an ancestral taxon was also
560 inferred if orthologues are present in at least one representative of its offspring and in
561 one of its ancestors.

562

563 **Authors' contributions**

564 L.E., D.M. and C.B-A designed the analyses and wrote the paper. L.E. and A.T.
565 performed the analyses.

566

567 **Acknowledgements and Funding**

568 C.B-A. is supported by an Action Thématique et Incitative sur Programme (ATIP) of
569 the French Centre National de la Recherche Scientifique (CNRS). We would like to

570 thank the JGI institute, the TBestDB database, the *Galdieria sulphuraria* genome
571 project and the *Cyanidioschyzon merolae* genome project for sharing sequence data
572 from ongoing projects with the scientific community. We also thank Rym Agrebi and
573 Celine Petitjean for careful reading of the manuscript and stimulating discussions.

574

575 **References**

576

- 577 1. Thornton BR, Toczyski DP: **Precise destruction: an emerging picture of the APC.** *Genes Dev* 2006, **20**(22):3069-3078.
- 578 2. Fang S, Weissman AM: **A field guide to ubiquitylation.** *Cell Mol Life Sci* 2004,
579 **61**(13):1546-1561.
- 580 3. Pickart CM: **Mechanisms underlying ubiquitination.** *Annu Rev Biochem* 2001,
581 **70**:503-533.
- 582 4. Schreiber A, Stengel F, Zhang Z, Enchev RI, Kong EH, Morris EP, Robinson CV, da
583 Fonseca PC, Barford D: **Structural basis for the subunit assembly of the**
584 **anaphase-promoting complex.** *Nature* 2011, **470**(7333):227-232.
- 585 5. Reed SI: **Ratchets and clocks: the cell cycle, ubiquitylation and protein**
586 **turnover.** *Nat Rev Mol Cell Biol* 2003, **4**(11):855-864.
- 587 6. Machado E, Eguren M, Malumbres M: **The anaphase-promoting**
588 **complex/cyclosome (APC/C): cell-cycle-dependent and -independent functions.**
589 *Biochem Soc Trans* 2010, **38**(Pt 1):65-71.
- 590 7. Peters JM: **The anaphase promoting complex/cyclosome: a machine designed**
591 **to destroy.** *Nat Rev Mol Cell Biol* 2006, **7**(9):644-656.
- 592 8. Yoon HJ, Feoktistova A, Wolfe BA, Jennings JL, Link AJ, Gould KL: **Proteomics**
593 **analysis identifies new components of the fission and budding yeast anaphase-**
594 **promoting complexes.** *Curr Biol* 2002, **12**(23):2048-2054.
- 595 9. Hutchins JR, Toyoda Y, Hegemann B, Poser I, Heriche JK, Sykora MM, Augsburg M,
596 Hudecz O, Buschhorn BA, Bulkescher J *et al*: **Systematic analysis of human**
597 **protein complexes identifies chromosome segregation proteins.** *Science* 2010,
598 **328**(5978):593-599.
- 599 10. Acquaviva C, Pines J: **The anaphase-promoting complex/cyclosome: APC/C.** *J*
600 *Cell Sci* 2006, **119**(Pt 12):2401-2404.
- 601 11. Matyskiela ME, Rodrigo-Brenni MC, Morgan DO: **Mechanisms of ubiquitin transfer**
602 **by the anaphase-promoting complex.** *J Biol* 2009, **8**(10):92.
- 603 12. Nasmyth K, Haering CH: **Cohesin: its roles and mechanisms.** *Annu Rev Genet*
604 2009, **43**:525-558.
- 605 13. Onn I, Heidingen-Pauli JM, Guacci V, Unal E, Koshland DE: **Sister chromatid**
606 **cohesion: a simple concept with a complex reality.** *Annu Rev Cell Dev Biol* 2008,
607 **24**:105-129.
- 608 14. Oelschlaegel T, Schwickart M, Matos J, Bogdanova A, Camasses A, Havlis J,
609 Shevchenko A, Zachariae W: **The yeast APC/C subunit Mnd2 prevents premature**
610 **sister chromatid separation triggered by the meiosis-specific APC/C-Ama1.** *Cell*
611 2005, **120**(6):773-788.
- 612 15. Pesin JA, Orr-Weaver TL: **Developmental role and regulation of cortex, a**
613 **meiosis-specific anaphase-promoting complex/cyclosome activator.** *PLoS*
614 *Genet* 2007, **3**(11):e202.
- 615 16. Pimentel AC, Venkatesh TR: **rap gene encodes Fizzy-related protein (Fzr) and**
616 **regulates cell proliferation and pattern formation in the developing Drosophila**
617 **eye-antennal disc.** *Dev Biol* 2005, **285**(2):436-446.
- 618

- 619 17. Capron A, Okresz L, Genschik P: **First glance at the plant APC/C, a highly**
620 **conserved ubiquitin-protein ligase.** *Trends Plant Sci* 2003, **8**(2):83-89.
- 621 18. Eloy NB, Coppens F, Beemster GT, Hemerly AS, Ferreira PC: **The Arabidopsis**
622 **anaphase promoting complex (APC): regulation through subunit availability in**
623 **plant tissues.** *Cell Cycle* 2006, **5**(17):1957-1965.
- 624 19. Lima Mde F, Eloy NB, Pegoraro C, Sagit R, Rojas C, Bretz T, Vargas L, Elofsson A,
625 de Oliveira AC, Hemerly AS *et al*: **Genomic evolution and complexity of the**
626 **Anaphase-promoting Complex (APC) in land plants.** *BMC Plant Biol* 2010,
627 **10**:254.
- 628 20. Fulop K, Tarayre S, Kelemen Z, Horvath G, Kevei Z, Nikovics K, Bako L, Brown S,
629 Kondorosi A, Kondorosi E: **Arabidopsis anaphase-promoting complexes: multiple**
630 **activators and wide range of substrates might keep APC perpetually busy.** *Cell*
631 *Cycle* 2005, **4**(8):1084-1092.
- 632 21. Kumar P, Wang CC: **Depletion of anaphase-promoting complex or cyclosome**
633 **(APC/C) subunit homolog APC1 or CDC27 of Trypanosoma brucei arrests the**
634 **procyclic form in metaphase but the bloodstream form in anaphase.** *J Biol Chem*
635 2005, **280**(36):31783-31791.
- 636 22. Eisen JA, Fraser CM: **Phylogenomics: intersection of evolution and genomics.**
637 *Science* 2003, **300**(5626):1706-1707.
- 638 23. Stechmann A, Cavalier-Smith T: **The root of the eukaryote tree pinpointed.** *Curr*
639 *Biol* 2003, **13**(17):R665-666.
- 640 24. Seidl MF, Schultz J: **Evolutionary flexibility of protein complexes.** *BMC Evol Biol*
641 2009, **9**:155.
- 642 25. von Schubert C, Xue G, Schmuckli-Maurer J, Woods KL, Nigg EA, Dobbelaere DA:
643 **The transforming parasite Theileria co-opts host cell mitotic and central**
644 **spindles to persist in continuously dividing cells.** *PLoS Biol* 2010, **8**(9).
- 645 26. Malik SB, Pightling AW, Stefaniak LM, Schurko AM, Logsdon JM, Jr.: **An expanded**
646 **inventory of conserved meiotic genes provides evidence for sex in**
647 **Trichomonas vaginalis.** *PLoS One* 2008, **3**(8):e2879.
- 648 27. Zou H, McGarry TJ, Bernal T, Kirschner MW: **Identification of a vertebrate sister-**
649 **chromatid separation inhibitor involved in transformation and tumorigenesis.**
650 *Science* 1999, **285**(5426):418-422.
- 651 28. Lam WS, Yang X, Makaroff CA: **Characterization of Arabidopsis thaliana SMC1**
652 **and SMC3: evidence that AtSMC3 may function beyond chromosome cohesion.**
653 *J Cell Sci* 2005, **118**(Pt 14):3037-3048.
- 654 29. Li W, Yang X, Lin Z, Timofejeva L, Xiao R, Makaroff CA, Ma H: **The AtRAD51C gene**
655 **is required for normal meiotic chromosome synapsis and double-stranded**
656 **break repair in Arabidopsis.** *Plant Physiol* 2005, **138**(2):965-976.
- 657 30. Liu Cm CM, McElver J, Tzafrir I, Joosen R, Wittich P, Patton D, Van Lammeren AA,
658 Meinke D: **Condensin and cohesin knockouts in Arabidopsis exhibit a titan seed**
659 **phenotype.** *Plant J* 2002, **29**(4):405-415.
- 660 31. Bai X, Peirson BN, Dong F, Xue C, Makaroff CA: **Isolation and characterization of**
661 **SYN1, a RAD21-like gene essential for meiosis in Arabidopsis.** *Plant Cell* 1999,
662 **11**(3):417-430.
- 663 32. Bhatt AM, Lister C, Page T, Fransz P, Findlay K, Jones GH, Dickinson HG, Dean C:
664 **The DIF1 gene of Arabidopsis is required for meiotic chromosome segregation**
665 **and belongs to the REC8/RAD21 cohesin gene family.** *Plant J* 1999, **19**(4):463-
666 472.
- 667 33. Cai X, Dong F, Edelmann RE, Makaroff CA: **The Arabidopsis SYN1 cohesin**
668 **protein is required for sister chromatid arm cohesion and homologous**
669 **chromosome pairing.** *J Cell Sci* 2003, **116**(Pt 14):2999-3007.
- 670 34. Van Leene J, Hollunder J, Eeckhout D, Persiau G, Van De Slijke E, Stals H, Van
671 Isterdael G, Verkest A, Neiryneck S, Buffel Y *et al*: **Targeted interactomics reveals a**
672 **complex core cell cycle machinery in Arabidopsis thaliana.** *Mol Syst Biol* 2010,
673 **6**:397.
- 674 35. Sebastian J, Ravi M, Andreuzza S, Panoli AP, Marimuthu MP, Siddiqi I: **The plant**
675 **adherin AtSCC2 is required for embryogenesis and sister-chromatid cohesion**
676 **during meiosis in Arabidopsis.** *Plant J* 2009, **59**(1):1-13.
- 677 36. Harper JW, Burton JL, Solomon MJ: **The anaphase-promoting complex: it's not**
678 **just for mitosis any more.** *Genes Dev* 2002, **16**(17):2179-2206.

- 679 37. Eme L, Moreira D, Talla E, Brochier-Armanet C: **A complex cell division machinery**
680 **was present in the last common ancestor of eukaryotes.** *PLoS One* 2009,
681 4(4):e5021.
- 682 38. Cobbe N, Heck MM: **The evolution of SMC proteins: phylogenetic analysis and**
683 **structural implications.** *Mol Biol Evol* 2004, 21(2):332-347.
- 684 39. Soppa J: **Prokaryotic structural maintenance of chromosomes (SMC) proteins:**
685 **distribution, phylogeny, and comparison with MukBs and additional prokaryotic**
686 **and eukaryotic coiled-coil proteins.** *Gene* 2001, 278(1-2):253-264.
- 687 40. Iyer LM, Burroughs AM, Aravind L: **The prokaryotic antecedents of the ubiquitin-**
688 **signaling system and the early evolution of ubiquitin-like beta-grasp domains.**
689 *Genome Biol* 2006, 7(7):R60.
- 690 41. De Mot R, Nagy I, Walz J, Baumeister W: **Proteasomes and other self-**
691 **compartmentalizing proteases in prokaryotes.** *Trends Microbiol* 1999, 7(2):88-92.
- 692 42. Darwin KH: **Prokaryotic ubiquitin-like protein (Pup), proteasomes and**
693 **pathogenesis.** *Nat Rev Microbiol* 2009, 7(7):485-491.
- 694 43. Nunoura T, Takaki Y, Kakuta J, Nishi S, Sugahara J, Kazama H, Chee GJ, Hattori M,
695 Kanai A, Atomi H *et al*: **Insights into the evolution of Archaea and eukaryotic**
696 **protein modifier systems revealed by the genome of a novel archaeal group.**
697 *Nucleic Acids Res* 2010.
- 698 44. Brochier-Armanet C, Boussau B, Gribaldo S, Forterre P: **Mesophilic Crenarchaeota:**
699 **proposal for a third archaeal phylum, the Thaumarchaeota.** *Nat Rev Microbiol*
700 2008, 6(3):245-252.
- 701 45. Baptiste E, Charlebois RL, MacLeod D, Brochier C: **The two tempos of nuclear**
702 **pore complex evolution: highly adapting proteins in an ancient frozen**
703 **structure.** *Genome Biol* 2005, 6(10):R85.
- 704 46. Desmond E, Brochier-Armanet C, Forterre P, Gribaldo S: **On the last common**
705 **ancestor and early evolution of eukaryotes: reconstructing the history of**
706 **mitochondrial ribosomes.** *Res Microbiol* 2011, 162(1):53-70.
- 707 47. Desmond E, Gribaldo S: **Phylogenomics of sterol synthesis: insights into the**
708 **origin, evolution, and diversity of a key eukaryotic feature.** *Genome Biol Evol*
709 2009, 2009:364-381.
- 710 48. Koonin EV: **Preview. The incredible expanding ancestor of eukaryotes.** *Cell* 2010,
711 140(5):606-608.
- 712 49. Fritz-Laylin LK, Prochnik SE, Ginger ML, Dacks JB, Carpenter ML, Field MC, Kuo A,
713 Paredes A, Chapman J, Pham J *et al*: **The genome of Naegleria gruberi**
714 **illuminates early eukaryotic versatility.** *Cell* 2010, 140(5):631-642.
- 715 50. Brown JR, Douady CJ, Italia MJ, Marshall WE, Stanhope MJ: **Universal trees based**
716 **on large combined protein sequence data sets.** *Nat Genet* 2001, 28(3):281-285.
- 717 51. Baptiste E, Brinkmann H, Lee JA, Moore DV, Sensen CW, Gordon P, Durufle L,
718 Gaasterland T, Lopez P, Muller M *et al*: **The analysis of 100 genes supports the**
719 **grouping of three highly divergent amoebae: Dictyostelium, Entamoeba, and**
720 **Mastigamoeba.** *Proc Natl Acad Sci U S A* 2002, 99(3):1414-1419.
- 721 52. Brochier C, Forterre P, Gribaldo S: **An emerging phylogenetic core of Archaea:**
722 **phylogenies of transcription and translation machineries converge following**
723 **addition of new genome sequences.** *BMC Evol Biol* 2005, 5(1):36.
- 724 53. Brochier C, Baptiste E, Moreira D, Philippe H: **Eubacterial phylogeny based on**
725 **translational apparatus proteins.** *Trends Genet* 2002, 18(1):1-5.
- 726 54. Rodriguez-Ezpeleta N, Brinkmann H, Burey SC, Roure B, Burger G, Löffelhardt W,
727 Bohnert HJ, Philippe H, Lang BF: **Monophyly of primary photosynthetic**
728 **eukaryotes: green plants, red algae, and glaucophytes.** *Curr Biol* 2005,
729 15(14):1325-1330.
- 730 55. Burki F, Inagaki Y, Brate J, Archibald JM, Keeling PJ, Cavalier-Smith T, Sakaguchi M,
731 Hashimoto T, Horak A, Kumar S *et al*: **Large-scale phylogenomic analyses reveal**
732 **that two enigmatic protist lineages, telonemia and centroheliozoa, are related to**
733 **photosynthetic chromalveolates.** *Genome Biol Evol* 2009, 1:231-238.
- 734 56. Burki F, Shalchian-Tabrizi K, Pawlowski J: **Phylogenomics reveals a new**
735 **'megagroup' including most photosynthetic eukaryotes.** *Biol Lett* 2008, 4(4):366-
736 369.
- 737 57. Hampl V, Hug L, Leigh JW, Dacks JB, Lang BF, Simpson AG, Roger AJ:
738 **Phylogenomic analyses support the monophyly of Excavata and resolve**

739 **relationships among eukaryotic "supergroups"**. *Proc Natl Acad Sci U S A* 2009,
740 **106**(10):3859-3864.

741 58. Delsuc F, Brinkmann H, Philippe H: **Phylogenomics and the reconstruction of the**
742 **tree of life**. *Nat Rev Genet* 2005, **6**(5):361-375.

743 59. Rodriguez-Ezpeleta N, Brinkmann H, Roure B, Lartillot N, Lang BF, Philippe H:
744 **Detecting and overcoming systematic errors in genome-scale phylogenies**. *Syst*
745 *Biol* 2007, **56**(3):389-399.

746 60. Felsenstein J: **Cases in which parsimony or compatibility methods will be**
747 **positively misleading**. *Syst Zool* 1978, **27**:401-410.

748 61. Thomarat F, Vivares CP, Gouy M: **Phylogenetic analysis of the complete genome**
749 **sequence of Encephalitozoon cuniculi supports the fungal origin of**
750 **microsporidia and reveals a high frequency of fast-evolving genes**. *J Mol Evol*
751 2004, **59**(6):780-791.

752 62. Ruiz-Trillo I, Lane CE, Archibald JM, Roger AJ: **Insights into the evolutionary**
753 **origin and genome architecture of the unicellular opisthokonts Capsaspora**
754 **owczarzaki and Sphaeroforma arctica**. *J Eukaryot Microbiol* 2006, **53**(5):379-384.

755 63. Minge MA, Silberman JD, Orr RJ, Cavalier-Smith T, Shalchian-Tabrizi K, Burki F,
756 Skjaveland A, Jakobsen KS: **Evolutionary position of breviate amoebae and the**
757 **primary eukaryote divergence**. *Proc Biol Sci* 2009, **276**(1657):597-604.

758 64. Shalchian-Tabrizi K, Minge MA, Espelund M, Orr R, Ruden T, Jakobsen KS, Cavalier-
759 Smith T: **Multigene phylogeny of choanozoa and the origin of animals**. *PLoS One*
760 2008, **3**(5):e2098.

761 65. Ruiz-Trillo I, Roger AJ, Burger G, Gray MW, Lang BF: **A phylogenomic**
762 **investigation into the origin of metazoa**. *Mol Biol Evol* 2008, **25**(4):664-672.

763 66. Baurain D, Brinkmann H, Petersen J, Rodriguez-Ezpeleta N, Stechmann A, Demoulin
764 V, Roger AJ, Burger G, Lang BF, Philippe H: **Phylogenomic evidence for separate**
765 **acquisition of plastids in cryptophytes, haptophytes, and stramenopiles**. *Mol*
766 *Biol Evol* 2010, **27**(7):1698-1709.

767 67. Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ:
768 **Gapped BLAST and PSI-BLAST: a new generation of protein database search**
769 **programs**. *Nucleic Acids Res* 1997, **25**(17):3389-3402.

770 68. O'Brien EA, Koski LB, Zhang Y, Yang L, Wang E, Gray MW, Burger G, Lang BF:
771 **TBestDB: a taxonomically broad database of expressed sequence tags (ESTs)**.
772 *Nucleic Acids Res* 2007, **35**(Database issue):D445-451.

773 69. Katoh K, Misawa K, Kuma K, Miyata T: **MAFFT: a novel method for rapid multiple**
774 **sequence alignment based on fast Fourier transform**. *Nucleic Acids Res* 2002,
775 **30**(14):3059-3066.

776 70. Philippe H: **MUST, a computer package of Management Utilities for Sequences**
777 **and Trees**. *Nucleic Acids Res* 1993, **21**(22):5264-5272.

778 71. Durbin R, Eddy SR, Krogh A, Mitchison G: **Biological Sequence Analysis**.
779 Cambridge: University Press Cambridge; 1999.

780 72. Finn RD, Mistry J, Schuster-Bockler B, Griffiths-Jones S, Hollich V, Lassmann T,
781 Moxon S, Marshall M, Khanna A, Durbin R *et al*: **Pfam: clans, web tools and**
782 **services**. *Nucleic Acids Res* 2006, **34**(Database issue):D247-251.

783 73. Jobb G, von Haeseler A, Strimmer K: **TREEFINDER: a powerful graphical analysis**
784 **environment for molecular phylogenetics**. *BMC Evol Biol* 2004, **4**:18.

785 74. Le SQ, Gascuel O: **An improved general amino acid replacement matrix**. *Mol Biol*
786 *Evol* 2008, **25**(7):1307-1320.

787 75. Ronquist F, Huelsenbeck JP: **MrBayes 3: Bayesian phylogenetic inference under**
788 **mixed models**. *Bioinformatics* 2003, **19**(12):1572-1574.

789 76. Lartillot N, Lepage T, Blanquart S: **PhyloBayes 3: a Bayesian software package**
790 **for phylogenetic reconstruction and molecular dating**. *Bioinformatics* 2009,
791 **25**(17):2286-2288.

792 77. Lartillot N, Philippe H: **A Bayesian mixture model for across-site heterogeneities**
793 **in the amino-acid replacement process**. *Mol Biol Evol* 2004, **21**(6):1095-1109.

794
795

796

797

798

800 **Tables**801 **Table 1.** Taxonomic distribution of homologues of APC/C subunits and adaptors/co-activators in major lineages of eukaryotes.

	Catalytic arm			Structural complex					TPR arm subunits and associated proteins					Mitotic adaptors			Meiotic adaptors			
	Apc1 0	Apc1 1	Apc 2	Apc 1	Apc 4	Apc 5	Apc 6	Apc 3	Apc 7	Apc 2	Apc1 3	Apc1 6	Apc 9	Apc1 5	Apc1 4	Cdh 0	Cdh 1	Ama 1	Ra p	Corte x
Holozoa/	100																			
Choanoflagellata	100		100	100	≥50		100	100	100	≥50						100	100			
Holozoa/																				
Metazoa	100	100	100	100	100	100	100	100	100	100	≥50	≥50	≥50	<50		100	100		<5	<50
Holozoa/																				
Capasporia	100	100	100	100	100	100	100	100	100	100						100	100			
Fungi/																				
Dikarya	100	100	100	100	100	100	100	100	100	100	≥50	≥50	≥50	<50	<50	100	100			≥50
Fungi/																				
Microsporidia	100	<50		100	100	100	100	100	100	100						100	100			
Fungi/																				
Chytridiomycota	≥50	100	100	100	≥50	≥50	100	100	100	100						100	100			
Apusozoa	100	100	100	100	100	100	100	100	100	100						100	100			
Amoebozoa	100	100	≥50	100	≥50	≥50	≥50	≥50	≥50	≥50						≥50	100			
Excavata	≥50	≥50	≥50	≥50	≥50	≥50	≥50	≥50	≥50	≥50						≥50	<50			
Alveolata/																				
Ciliata	100	100	100	100	100	100	100	100	100	100						≥50	100			
Alveolata/																				
Apicomplexa	≥50	≥50	≥50	<50	<50	<50	<50	<50	<50	<50										
Heterokonta	100	≥50	≥50	50	<50	≥50	100	≥50	≥50	≥50						100	≥50			
Plantae/																				
Viridiplantae	100	100	100	100	100	100	100	100	100	100						100	≥50			
Plantae/																				
Rhodophyta	100	100	100	100	100	100	100	100	100	100						100	100			
Haptophyta	100	100	100	100	100	100	100	100	100	100						100	100			

100, ≥50, and <50 indicate that at least one orthologue has been detected respectively in all, at least 50%, and less than 50% of the studied genomes for a given lineage.

804

805 **Table 2.** Taxonomic distribution of homologues of major APC/C targets in major lineages of eukaryotes.

	Core complex			Loading complex		Cohesion establishment		Maintenance		Dissolution	
	Smc1	Smc3	Scc1	Scc3	Scc2	Scc4	Eco1	Pds5	Wpl1/Rad61	Separase	Securin
Choanoflagellata	100	100	100	100	100	100	Half	100	Half	100	100
Metazoa	>50	>50	>50	100	100	100	100	100	100	>50	<50
Capsaspora	100	100	100	100	100	100	100	100	100	100	100
Fungi/Dikarya	100	100	100	100	100	>50	>50	100	100	100	<50
Fungi/Microsporidia	>50	>50	>50	>50						>50	
Fungi/Chytridiomycota	100	100	100	100	100	50	100	100	100	100	100
Apusozoa	100	100	100	100	100		100			100	
Amoebozoa	100	100	100	>50	>50		>50	>50		100	
Excavata	100	100	>50	>50	<50		>50	>50		100	
Alveolata/Ciliata	100	<50	<50	<50						>50	
Alveolata/Alveolomaxilla	100	>50	<50	>50						>50	
Heterokonta	100	>50	>50	>50	>50		>50	>50		>50	
Plantae/Viridiplantae	100	100	100	100	>50	>50	>50		>50	>50	
Plantae/Rhodophyta	100	100	100	100	>50		100	>50		100	
Haptophyta	100	100	100	100	100	100	100	100		100	

806 100, >50, and <50 indicate that at least one orthologue has been detected respectively in all, at least 50%, and less than 50% of the studied genomes for a given lineage.

807

808

809

810

811

812 **Figures**

813 **Figure 1.** Schematic structure of the APC/C.

814 The APC/C is divided in three parts: the structural complex composed of three
815 subunits (in pink), the TPR arm made of nine proteins (in blue) and a second arm
816 (the catalytic arm) made of three proteins (in green) interacting with the E2 protein (in
817 purple) and the substrate (in grey). The activity and specificity of the APC/C is
818 modulated by various adaptor/co-activators (in orange). The location of Apc14 (in
819 black) remains uncertain. The presence of RING finger and Cullin conserved
820 functional domains, and of TPR and WD repeats is indicated.

821 **Figure 2.** Evolutionary history of APC/C subunits, adaptors/co-activators and main
822 targets. We used a reference phylogenetic tree rooted in-between Unikonta and
823 Bikonta [23] showing the relationships between the eleven major eukaryotic lineages
824 for which complete (or nearly complete) genome sequences are available. Holozoa
825 are represented by Capsaspora, Choanoflagellata and Metazoa (in pink); Fungi by
826 Dikarya, Chytridiomycota and Microsporidia lineages (in dark purple); Alveolata by
827 Apicomplexa and Ciliata (in brown) and Plantae by Viridiplantae and Rhodophyta (in
828 green). The completion status for each genome sequence is indicated: "C":
829 complete," A": draft assembly, "P": ongoing (accordingly the loss of some
830 components inferred in the corresponding lineages should be confirmed when
831 complete genome sequences are available). LECA (grey circle) represents the Last
832 Eukaryotic Common Ancestor. Based on phylogenetic analyses, we inferred the time
833 of appearance or loss of orthologues of APC/C subunits (dark blue), activators (light
834 blue) and main targets (pink). For each lineage, protein gains and losses are
835 indicated by green and red arrows, respectively.

836

837 **Figure 3.** Bayesian phylogenetic tree of eukaryotes. The tree was inferred by
838 PhyloBayes based on a supermatrix made of the 24 APC/C components,
839 adaptors/co-activators and main targets inferred to be present in the last common
840 ancestor of eukaryotes (3115 amino-acid positions, 56 species). Numbers at nodes
841 are Bayesian posterior probabilities, filled circles indicate maximum likelihood
842 bootstrap values of 100% (black) or >95% (grey). The scale bar indicates the
843 average number of substitutions per site.

844

845 **Additional files**

846 **Table S1.** Table showing the taxonomic distribution of homologues of APC/C
847 subunits and activators in eukaryotes.

848 **Table S2.** Table showing the taxonomic distribution of homologues of APC/C main
849 targets in eukaryotes.

850 **Table S3.** Table showing the conserved functional domains present in homologues
851 of APC/C subunits and activators.

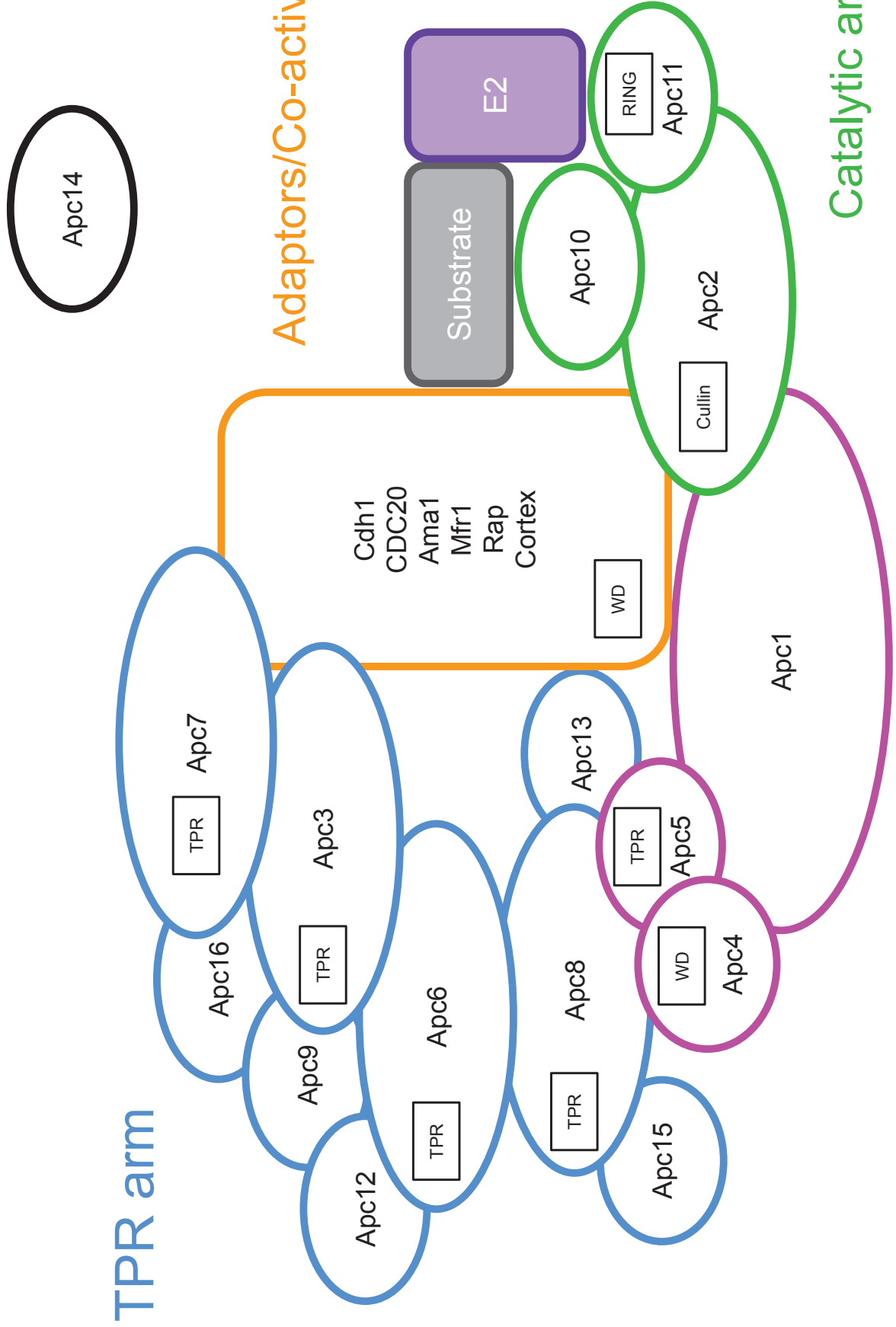
852 **Table S4.** Table showing the conserved functional domains present in homologues
853 of APC/C main targets.

854 **Figure S1.** Bayesian phylogenetic tree of the adaptor/co-activator Cdh1 inferred with
855 MrBayes. Numbers at nodes correspond to posterior probabilities (PP) estimated by
856 MrBayes and bootstrap values (BV) estimated by TreeFinder (only PP and BV
857 greater than 0.5 and 50%, respectively, are shown). The scale bar indicates the
858 average number of substitutions per site.

859 **Figure S2.** Bayesian phylogenetic tree of the adaptor/co-activator Cdc20 inferred
860 with MrBayes. Numbers at nodes correspond to posterior probabilities (PP) estimated
861 by MrBayes and bootstrap values (BV) estimated by TreeFinder (only PP and BV
862 greater than 0.5 and 50%, respectively, are shown).. The scale bar indicates the
863 average number of substitutions per site.

864 **Data S1.** Bayesian phylogenetic trees inferred MrBayes for each APC/C subunit,
865 activator and main target (except Apc9, Apc14, Apc15 and Apc16 due to a too
866 restricted number of homologues).

867



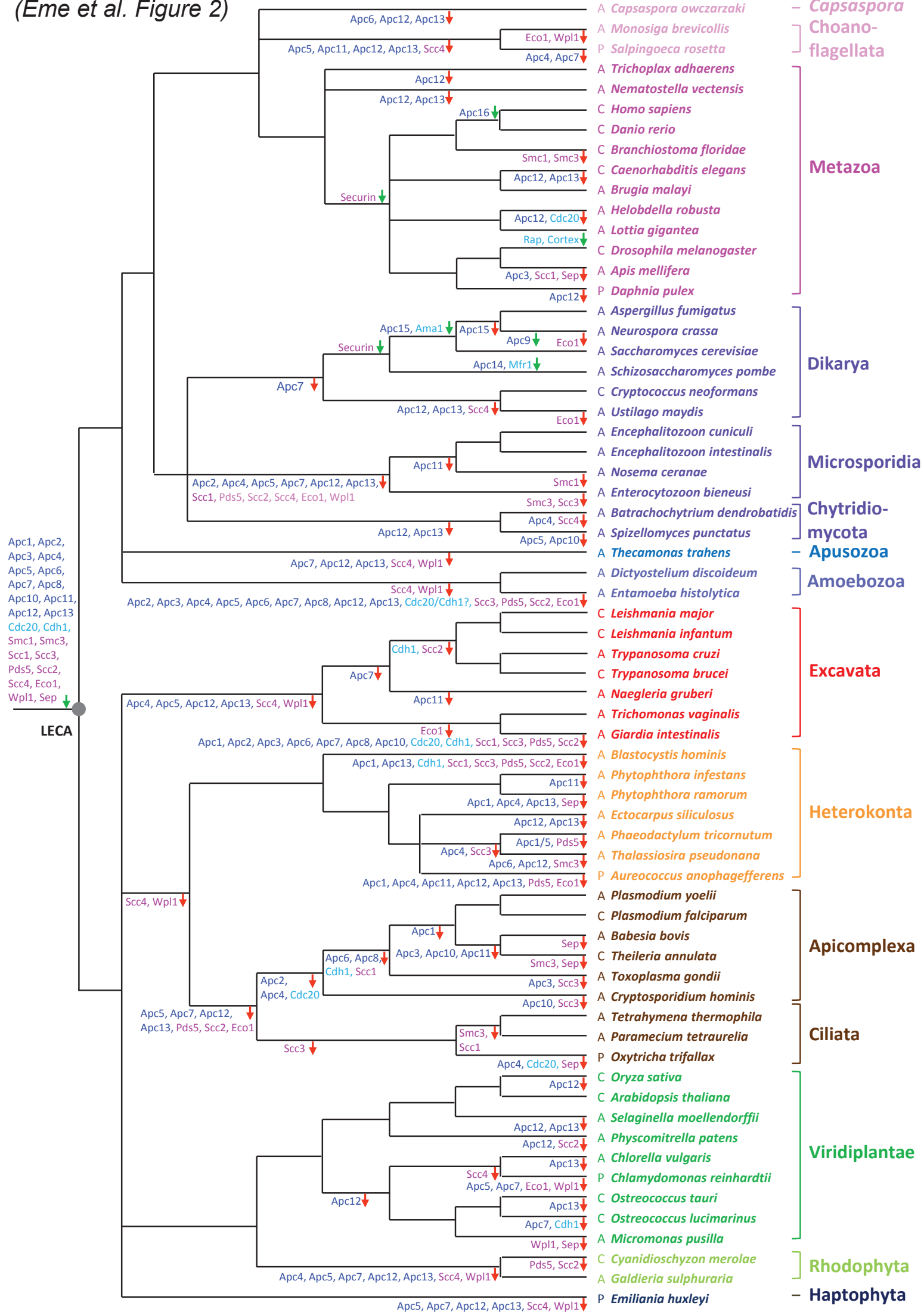
Structural complex

Catalytic arm

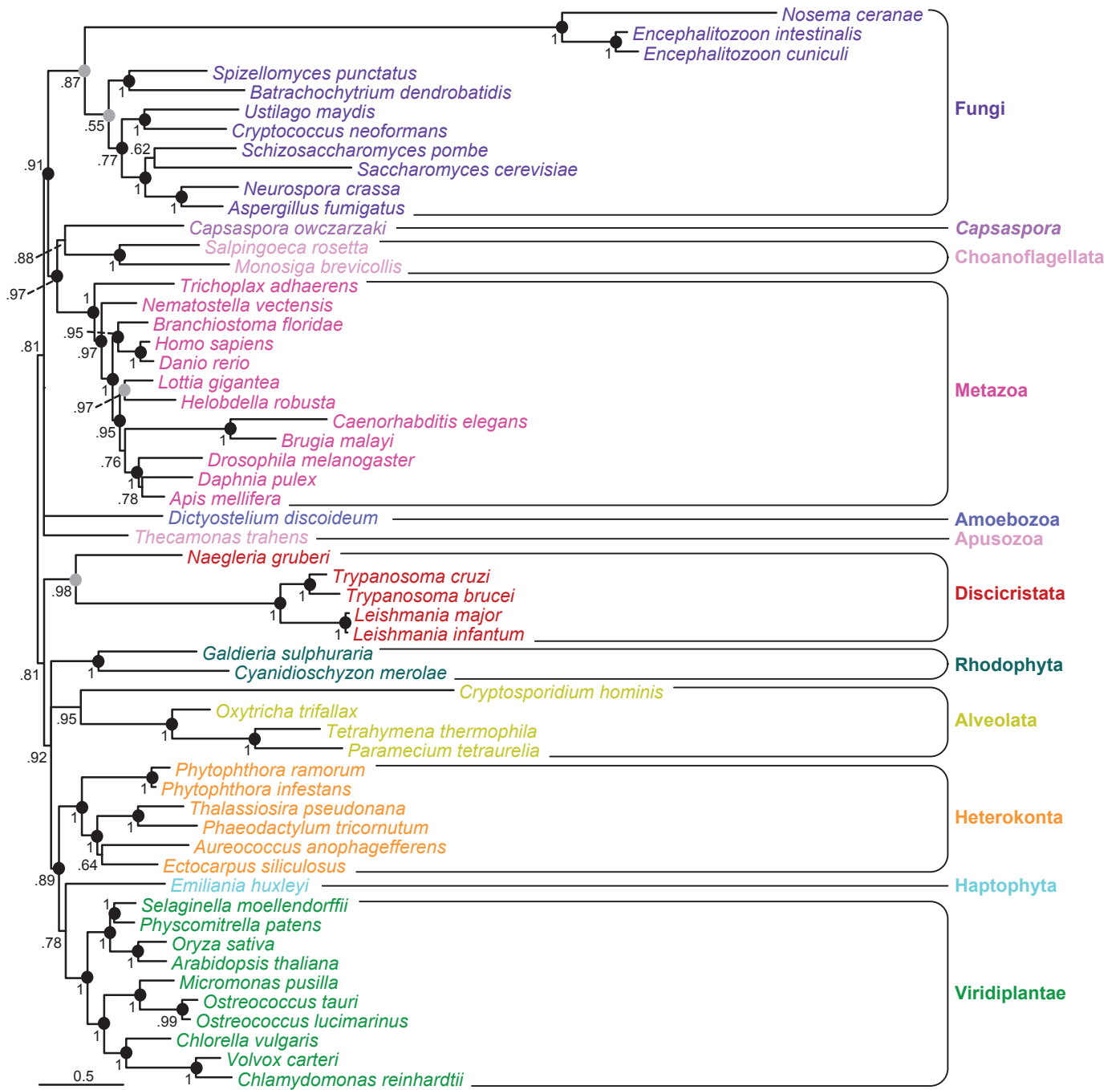
Adaptors/Co-activators

TPR arm

(Eme et al. Figure 2)



(Eme et al. Figure 3)



Metazoa, Choanoflagellata, Apusozoa,
 Capsaspora, Fungi, Amoebozoa, Haptophyta,
 Plantae (Rhodophytes + Viridiplantae),
 Heterokonta, Alveolata (Ciliata + Apicomplexa)
 Exavata



Metazoa, Choanoflagellata, Apusozoa,
 Capsaspora, Fungi, Amoebozoa, Haptophyta,
 Plantae (Rhodophytes + Viridiplantae),
 Heterokonta, Alveolata (Ciliata + Apicomplexa)
 Exavata



0.1

Supplementary Table S3. Domain composition of homologues of APC/C subunits and adaptors/co-activators. Domain presence was determined according to the Pfam database. A "*" indicates the absence of conserved domain in the corresponding protein. Empty cells indicate that no homologue of the protein has been detected in the corresponding genome. Profile accession numbers: APC10 (PF03256.7), zfc3HC4 (PF00097.15), APC2 (PF08672.1), Cullin (PF01851.12), APC8 (PF04049.4) and WD40 (PF00400.22). Profiles corresponding to TPR domains are either TPR_1 (PF00515.18), TPR_2 (PF07719.7) or TPR_4 (PF07721.5).

	Phylum	Species	Core subunits										Co-activators																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																													
			Catalytic arm			Structural complex			TPR arm and associated proteins				Mitosis co-activators			Meiotic co-activator																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																										
			APC10	APC11	APC2	APC1	APC2	APC3	APC4	APC5	APC6	APC7	APC8	APC9	APC10	APC11	APC12	APC13	APC14	APC15	APC16	APC17	APC18	APC19	APC20	APC21	APC22	APC23	APC24	APC25	APC26	APC27	APC28	APC29	APC30	APC31	APC32	APC33	APC34	APC35	APC36	APC37	APC38	APC39	APC40	APC41	APC42	APC43	APC44	APC45	APC46	APC47	APC48	APC49	APC50	APC51	APC52	APC53	APC54	APC55	APC56	APC57	APC58	APC59	APC60	APC61	APC62	APC63	APC64	APC65	APC66	APC67	APC68	APC69	APC70	APC71	APC72	APC73	APC74	APC75	APC76	APC77	APC78	APC79	APC80	APC81	APC82	APC83	APC84	APC85	APC86	APC87	APC88	APC89	APC90	APC91	APC92	APC93	APC94	APC95	APC96	APC97	APC98	APC99	APC100	APC101	APC102	APC103	APC104	APC105	APC106	APC107	APC108	APC109	APC110	APC111	APC112	APC113	APC114	APC115	APC116	APC117	APC118	APC119	APC120	APC121	APC122	APC123	APC124	APC125	APC126	APC127	APC128	APC129	APC130	APC131	APC132	APC133	APC134	APC135	APC136	APC137	APC138	APC139	APC140	APC141	APC142	APC143	APC144	APC145	APC146	APC147	APC148	APC149	APC150	APC151	APC152	APC153	APC154	APC155	APC156	APC157	APC158	APC159	APC160	APC161	APC162	APC163	APC164	APC165	APC166	APC167	APC168	APC169	APC170	APC171	APC172	APC173	APC174	APC175	APC176	APC177	APC178	APC179	APC180	APC181	APC182	APC183	APC184	APC185	APC186	APC187	APC188	APC189	APC190	APC191	APC192	APC193	APC194	APC195	APC196	APC197	APC198	APC199	APC200	APC201	APC202	APC203	APC204	APC205	APC206	APC207	APC208	APC209	APC210	APC211	APC212	APC213	APC214	APC215	APC216	APC217	APC218	APC219	APC220	APC221	APC222	APC223	APC224	APC225	APC226	APC227	APC228	APC229	APC230	APC231	APC232	APC233	APC234	APC235	APC236	APC237	APC238	APC239	APC240	APC241	APC242	APC243	APC244	APC245	APC246	APC247	APC248	APC249	APC250	APC251	APC252	APC253	APC254	APC255	APC256	APC257	APC258	APC259	APC260	APC261	APC262	APC263	APC264	APC265	APC266	APC267	APC268	APC269	APC270	APC271	APC272	APC273	APC274	APC275	APC276	APC277	APC278	APC279	APC280	APC281	APC282	APC283	APC284	APC285	APC286	APC287	APC288	APC289	APC290	APC291	APC292	APC293	APC294	APC295	APC296	APC297	APC298	APC299	APC300	APC301	APC302	APC303	APC304	APC305	APC306	APC307	APC308	APC309	APC310	APC311	APC312	APC313	APC314	APC315	APC316	APC317	APC318	APC319	APC320	APC321	APC322	APC323	APC324	APC325	APC326	APC327	APC328	APC329	APC330	APC331	APC332	APC333	APC334	APC335	APC336	APC337	APC338	APC339	APC340	APC341	APC342	APC343	APC344	APC345	APC346	APC347	APC348	APC349	APC350	APC351	APC352	APC353	APC354	APC355	APC356	APC357	APC358	APC359	APC360	APC361	APC362	APC363	APC364	APC365	APC366	APC367	APC368	APC369	APC370	APC371	APC372	APC373	APC374	APC375	APC376	APC377	APC378	APC379	APC380	APC381	APC382	APC383	APC384	APC385	APC386	APC387	APC388	APC389	APC390	APC391	APC392	APC393	APC394	APC395	APC396	APC397	APC398	APC399	APC400	APC401	APC402	APC403	APC404	APC405	APC406	APC407	APC408	APC409	APC410	APC411	APC412	APC413	APC414	APC415	APC416	APC417	APC418	APC419	APC420	APC421	APC422	APC423	APC424	APC425	APC426	APC427	APC428	APC429	APC430	APC431	APC432	APC433	APC434	APC435	APC436	APC437	APC438	APC439	APC440	APC441	APC442	APC443	APC444	APC445	APC446	APC447	APC448	APC449	APC450	APC451	APC452	APC453	APC454	APC455	APC456	APC457	APC458	APC459	APC460	APC461	APC462	APC463	APC464	APC465	APC466	APC467	APC468	APC469	APC470	APC471	APC472	APC473	APC474	APC475	APC476	APC477	APC478	APC479	APC480	APC481	APC482	APC483	APC484	APC485	APC486	APC487	APC488	APC489	APC490	APC491	APC492	APC493	APC494	APC495	APC496	APC497	APC498	APC499	APC500	APC501	APC502	APC503	APC504	APC505	APC506	APC507	APC508	APC509	APC510	APC511	APC512	APC513	APC514	APC515	APC516	APC517	APC518	APC519	APC520	APC521	APC522	APC523	APC524	APC525	APC526	APC527	APC528	APC529	APC530	APC531	APC532	APC533	APC534	APC535	APC536	APC537	APC538	APC539	APC540	APC541	APC542	APC543	APC544	APC545	APC546	APC547	APC548	APC549	APC550	APC551	APC552	APC553	APC554	APC555	APC556	APC557	APC558	APC559	APC560	APC561	APC562	APC563	APC564	APC565	APC566	APC567	APC568	APC569	APC570	APC571	APC572	APC573	APC574	APC575	APC576	APC577	APC578	APC579	APC580	APC581	APC582	APC583	APC584	APC585	APC586	APC587	APC588	APC589	APC590	APC591	APC592	APC593	APC594	APC595	APC596	APC597	APC598	APC599	APC600	APC601	APC602	APC603	APC604	APC605	APC606	APC607	APC608	APC609	APC610	APC611	APC612	APC613	APC614	APC615	APC616	APC617	APC618	APC619	APC620	APC621	APC622	APC623	APC624	APC625	APC626	APC627	APC628	APC629	APC630	APC631	APC632	APC633	APC634	APC635	APC636	APC637	APC638	APC639	APC640	APC641	APC642	APC643	APC644	APC645	APC646	APC647	APC648	APC649	APC650	APC651	APC652	APC653	APC654	APC655	APC656	APC657	APC658	APC659	APC660	APC661	APC662	APC663	APC664	APC665	APC666	APC667	APC668	APC669	APC670	APC671	APC672	APC673	APC674	APC675	APC676	APC677	APC678	APC679	APC680	APC681	APC682	APC683	APC684	APC685	APC686	APC687	APC688	APC689	APC690	APC691	APC692	APC693	APC694	APC695	APC696	APC697	APC698	APC699	APC700	APC701	APC702	APC703	APC704	APC705	APC706	APC707	APC708	APC709	APC710	APC711	APC712	APC713	APC714	APC715	APC716	APC717	APC718	APC719	APC720	APC721	APC722	APC723	APC724	APC725	APC726	APC727	APC728	APC729	APC730	APC731	APC732	APC733	APC734	APC735	APC736	APC737	APC738	APC739	APC740	APC741	APC742	APC743	APC744	APC745	APC746	APC747	APC748	APC749	APC750	APC751	APC752	APC753	APC754	APC755	APC756	APC757	APC758	APC759	APC760	APC761	APC762	APC763	APC764	APC765	APC766	APC767	APC768	APC769	APC770	APC771	APC772	APC773	APC774	APC775	APC776	APC777	APC778	APC779	APC780	APC781	APC782	APC783	APC784	APC785	APC786	APC787	APC788	APC789	APC790	APC791	APC792	APC793	APC794	APC795	APC796	APC797	APC798	APC799	APC800	APC801	APC802	APC803	APC804	APC805	APC806	APC807	APC808	APC809	APC810	APC811	APC812	APC813	APC814	APC815	APC816	APC817	APC818	APC819	APC820	APC821	APC822	APC823	APC824	APC825	APC826	APC827	APC828	APC829	APC830	APC831	APC832	APC833	APC834	APC835	APC836	APC837	APC838	APC839	APC840	APC841	APC842	APC843	APC844	APC845	APC846	APC847	APC848	APC849	APC850	APC851	APC852	APC853	APC854	APC855	APC856	APC857	APC858	APC859	APC860	APC861	APC862	APC863	APC864	APC865	APC866	APC867	APC868	APC869	APC870	APC871	APC872	APC873	APC874	APC875	APC876	APC877	APC878	APC879	APC880	APC881	APC882	APC883	APC884	APC885	APC886	APC887	APC888	APC889	APC890	APC891	APC892	APC893	APC894	APC895	APC896	APC897	APC898	APC899	APC900	APC901	APC902	APC903	APC904	APC905	APC906	APC907	APC908	APC909	APC910	APC911	APC912	APC913	APC914	APC915	APC916	APC917	APC918	APC919	APC920	APC921	APC922	APC923	APC924	APC925	APC926	APC927	APC928	APC929	APC930	APC931	APC932	APC933	APC934	APC935	APC936	APC937	APC938	APC939	APC940	APC941	APC942	APC943	APC944	APC945	APC946	APC947	APC948	APC949	APC950	APC951	APC952	APC953	APC954	APC955	APC956	APC957	APC958	APC959	APC960	APC961	APC962	APC963	APC964	APC965	APC966	APC967	APC968	APC969	APC970	APC971	APC972	APC973	APC974	APC975	APC976	APC977	APC978	APC979	APC980	APC981	APC982	APC983	APC984	APC985	APC986	APC987	APC988	APC989	APC990	APC991	APC992	APC993	APC994	APC995	APC996	APC997
Choanoflagellata		<i>Monosiga brevicollis</i>	APC10	APC11	APC2	APC1	APC2	APC3	APC4	APC5	APC6	APC7	APC8	APC9	APC10	APC11	APC12	APC13	APC14	APC15	APC16	APC17	APC18	APC19	APC20	APC21	APC22	APC23	APC24	APC25	APC26	APC27	APC28	APC29	APC30	APC31	APC32	APC33	APC34	APC35	APC36	APC37	APC38	APC39	APC40	APC41	APC42	APC43	APC44	APC45	APC46	APC47	APC48	APC49	APC50	APC51	APC52	APC53	APC54	APC55	APC56	APC57	APC58	APC59	APC60	APC61	APC62	APC63	APC64	APC65	APC66	APC67	APC68	APC69	APC70	APC71	APC72	APC73	APC74	APC75	APC76	APC77	APC78	APC79	APC80	APC81	APC82	APC83	APC84	APC85	APC86	APC87	APC88	APC89	APC90	APC91	APC92	APC93	APC94	APC95	APC96	APC97																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																				

Supplementary Table S4. Domain composition of homologues of APC/C main targets. Domain presence was determined according to the Pfam database. A "*" indicates the absence of conserved domain in the corresponding protein. Empty cells indicate that no homologue of the protein has been detected in the corresponding genome. Profile accession numbers: SMC_hinge (PF06470.4), SMC_N (PF02463.9), Rad21_Rec8 (PF04824.7), Rad21_Rec8_N (PF04825.4), HEAT (PF02985.12), STAG (PF07814.3), WAPL (PF07814.3), Peptidase_C50 (PF03568.7) and Securin (PF04856.4).

		Core complex						Loading complex		Cohesion establishment		Maintenance		Dissolution	
		Smc1	Smc3	Smc1	Smc3	Smc3	Smc3	Sec2	Sec4	Eco1	Pds5	Wpl1	Separase	Securin	
Opisthokonta	Choanoflagellata	<i>Monosiga brevicollis</i>	SMC_hinge/SMC_N		Rad21_Rec8/Rad21_Rec8_N			HEAT			HEAT		Peptidase_C50		
		<i>Salpingoeca rosetta</i>	SMC_hinge/SMC_N		Rad21_Rec8/Rad21_Rec8_N			HEAT			HEAT		Peptidase_C50		
		<i>Trichoplax adhaerens</i>			HEAT/STAG			HEAT	TPR		HEAT				
		<i>Homo sapiens</i>			HEAT			HEAT	TPR		HEAT			Securin	
		<i>Danio rerio</i>			HEAT			HEAT						Securin	
		<i>Branchiostoma floridae</i>			STAG			HEAT						Peptidase_C50	
		<i>Drosophila melanogaster</i>			STAG			HEAT						Peptidase_C50	
		<i>Lotia gigantea</i>			HEAT			HEAT						Peptidase_C50	
		<i>Apis mellifera</i>		SMC_N		HEAT		HEAT						Peptidase_C50	
		<i>Caenorhabditis elegans</i>		SMC_hinge/SMC_N		HEAT		HEAT						Peptidase_C50	
Opisthokonta	Capssporia	<i>Capssporia owczarzakii</i>	SMC_hinge/SMC_N		Rad21_Rec8/Rad21_Rec8_N			HEAT			HEAT		Peptidase_C50		
		<i>Cryptococcus neoformans</i>	SMC_hinge/SMC_N		Rad21_Rec8/Rad21_Rec8_N			HEAT			HEAT		Peptidase_C50		
		<i>Ustilago maydis</i>	SMC_hinge/SMC_N		Rad21_Rec8/Rad21_Rec8_N			HEAT			HEAT		Peptidase_C50		
		<i>Aspergillus fumigatus</i>			STAG			HEAT			HEAT		Peptidase_C50		
		<i>Schizosaccharomyces pombe</i>			STAG			HEAT			HEAT		Peptidase_C50		
		<i>Saccharomyces cerevisiae</i>			STAG			HEAT			HEAT		Peptidase_C50		
		<i>Neurospora crassa</i>			STAG			HEAT			HEAT		Peptidase_C50		
		<i>Encephalitozoon cuniculi</i>			HEAT			HEAT			HEAT		Peptidase_C50		
		<i>Enterocytozoon bieneusi H348</i>			HEAT			HEAT			HEAT		Peptidase_C50		
		<i>Nosema ceranae BRL01</i>			HEAT			HEAT			HEAT		Peptidase_C50		
Fungi		<i>Encephalitozoon intestinalis ATCC 50506</i>	SMC_hinge/SMC_N		Rad21_Rec8/Rad21_Rec8_N			HEAT			HEAT		Peptidase_C50		
		<i>Batrachochytrium dendrobatidis</i>	SMC_hinge/SMC_N		Rad21_Rec8/Rad21_Rec8_N			HEAT			HEAT		Peptidase_C50		
		<i>Spizellomyces punctatus</i>	SMC_hinge/SMC_N		Rad21_Rec8/Rad21_Rec8_N			HEAT			HEAT		Peptidase_C50		
		<i>Thecamonas trahens</i>	SMC_hinge/SMC_N		Rad21_Rec8/Rad21_Rec8_N			HEAT			HEAT		Peptidase_C50		
		<i>Dicystostelium discoideum</i>	SMC_hinge/SMC_N		Rad21_Rec8/Rad21_Rec8_N			HEAT			HEAT		Peptidase_C50		
		<i>Entamoeba histolytica</i>	SMC_hinge/SMC_N		Rad21_Rec8/Rad21_Rec8_N			HEAT			HEAT		Peptidase_C50		
		<i>Giardia intestinalis</i>	SMC_hinge/SMC_N		Rad21_Rec8/Rad21_Rec8_N			HEAT			HEAT		Peptidase_C50		
		<i>Trichomonas vaginalis</i>	SMC_hinge/SMC_N		Rad21_Rec8/Rad21_Rec8_N			HEAT			HEAT		Peptidase_C50		
		<i>Leishmania major</i>	SMC_hinge/SMC_N		Rad21_Rec8/Rad21_Rec8_N			HEAT			HEAT		Peptidase_C50		
		<i>Leishmania infantum</i>	SMC_hinge/SMC_N		Rad21_Rec8/Rad21_Rec8_N			HEAT			HEAT		Peptidase_C50		
Excavata		<i>Trypanosoma cruzi</i>	SMC_hinge/SMC_N		Rad21_Rec8/Rad21_Rec8_N			HEAT			HEAT		Peptidase_C50		
		<i>Trypanosoma brucei</i>	SMC_hinge/SMC_N		Rad21_Rec8/Rad21_Rec8_N			HEAT			HEAT		Peptidase_C50		
		<i>Naegleria gruberi</i>	SMC_hinge/SMC_N		Rad21_Rec8/Rad21_Rec8_N			HEAT			HEAT		Peptidase_C50		
		<i>Tetrahymena thermophila</i>	SMC_hinge/SMC_N		Rad21_Rec8/Rad21_Rec8_N			HEAT			HEAT		Peptidase_C50		
		<i>Paramecium tetraurelia</i>	SMC_hinge/SMC_N		Rad21_Rec8/Rad21_Rec8_N			HEAT			HEAT		Peptidase_C50		
		<i>Oxytricha trifallax</i>	SMC_hinge/SMC_N		Rad21_Rec8/Rad21_Rec8_N			HEAT			HEAT		Peptidase_C50		
		<i>Plasmodium yoelii</i>	SMC_hinge/SMC_N		Rad21_Rec8/Rad21_Rec8_N			HEAT			HEAT		Peptidase_C50		
		<i>Plasmodium falciparum</i>	SMC_hinge/SMC_N		Rad21_Rec8/Rad21_Rec8_N			HEAT			HEAT		Peptidase_C50		
		<i>Cryptosporidium hominis</i>	SMC_hinge/SMC_N		Rad21_Rec8/Rad21_Rec8_N			HEAT			HEAT		Peptidase_C50		
		<i>Babesia bovis</i>	SMC_hinge/SMC_N		Rad21_Rec8/Rad21_Rec8_N			HEAT			HEAT		Peptidase_C50		
Alveolata		<i>Theileria annulata</i>	SMC_hinge/SMC_N		Rad21_Rec8/Rad21_Rec8_N			HEAT			HEAT		Peptidase_C50		
		<i>Toxoplasma gondii</i>	SMC_hinge/SMC_N		Rad21_Rec8/Rad21_Rec8_N			HEAT			HEAT		Peptidase_C50		
		<i>Blastocystis hominis</i>	SMC_hinge/SMC_N		Rad21_Rec8/Rad21_Rec8_N			HEAT			HEAT		Peptidase_C50		
		<i>Ectocarpus siliculosus</i>	SMC_hinge/SMC_N		Rad21_Rec8/Rad21_Rec8_N			HEAT			HEAT		Peptidase_C50		
		<i>Phytophthora ramorum</i>	SMC_hinge/SMC_N		Rad21_Rec8/Rad21_Rec8_N			HEAT			HEAT		Peptidase_C50		
		<i>Phytophthora infestans</i>	SMC_hinge/SMC_N		Rad21_Rec8/Rad21_Rec8_N			HEAT			HEAT		Peptidase_C50		
		<i>Thalassiosira pseudonana</i>	SMC_hinge/SMC_N		Rad21_Rec8/Rad21_Rec8_N			HEAT			HEAT		Peptidase_C50		
		<i>Phaeodactylum tricornutum</i>	SMC_hinge/SMC_N		Rad21_Rec8/Rad21_Rec8_N			HEAT			HEAT		Peptidase_C50		
		<i>Aureococcus anophagefferens</i>	SMC_hinge/SMC_N		Rad21_Rec8/Rad21_Rec8_N			HEAT			HEAT		Peptidase_C50		
		<i>Oryza sativa</i>	SMC_hinge/SMC_N		Rad21_Rec8/Rad21_Rec8_N			HEAT			HEAT		Peptidase_C50		
Heterokonta		<i>Arabidopsis thaliana</i>	SMC_hinge/SMC_N		Rad21_Rec8/Rad21_Rec8_N			HEAT			HEAT		Peptidase_C50		
		<i>Chlamydomonas reinhardtii</i>	SMC_hinge/SMC_N		Rad21_Rec8/Rad21_Rec8_N			HEAT			HEAT		Peptidase_C50		
		<i>Ostreococcus tauri</i>	SMC_hinge/SMC_N		Rad21_Rec8/Rad21_Rec8_N			HEAT			HEAT		Peptidase_C50		
		<i>Ostreococcus lucimarinus</i>	SMC_hinge/SMC_N		Rad21_Rec8/Rad21_Rec8_N			HEAT			HEAT		Peptidase_C50		
		<i>Physcomitrella patens</i>	SMC_hinge/SMC_N		Rad21_Rec8/Rad21_Rec8_N			HEAT			HEAT		Peptidase_C50		
		<i>Chlorella vulgaris</i>	SMC_hinge/SMC_N		Rad21_Rec8/Rad21_Rec8_N			HEAT			HEAT		Peptidase_C50		
		<i>Micromonas pusilla</i>	SMC_hinge/SMC_N		Rad21_Rec8/Rad21_Rec8_N			HEAT			HEAT		Peptidase_C50		
		<i>Selaginella moellendorffii</i>	SMC_hinge/SMC_N		Rad21_Rec8/Rad21_Rec8_N			HEAT			HEAT		Peptidase_C50		
		<i>Cyanidioschyzon meroiiae</i>	SMC_hinge/SMC_N		Rad21_Rec8/Rad21_Rec8_N			HEAT			HEAT		Peptidase_C50		
		<i>Galdieria sulphuraria</i>	SMC_hinge/SMC_N		Rad21_Rec8/Rad21_Rec8_N			HEAT			HEAT		Peptidase_C50		
Plantae		<i>Emiliania huxleyi</i>	SMC_hinge/SMC_N		Rad21_Rec8/Rad21_Rec8_N			HEAT			HEAT		Peptidase_C50		

ARTICLE 3 :

A eukaryotic-like sulfiredoxin involved in oxidative stress responses and in the reduction of the sulfinic form of 2 Cys-peroxiredoxin in the cyanobacterium *Anabaena* PCC 7120

A eukaryotic-like sulfiredoxin involved in oxidative stress responses and in the reduction of the sulfinic form of 2-Cys peroxiredoxin in the cyanobacterium *Anabaena* PCC 7120

Céline Boileau, Laura Eme, Céline Brochier-Armanet, Annick Janicki, Cheng-Cai Zhang and Amel Latifi
Aix-Marseille University and Laboratoire de Chimie Bactérienne, IMM-CNRS, 31 Chemin Joseph Aiguier, F-13402 Marseille cedex 20, France

Summary

Author for correspondence:
Amel Latifi
Tel: +33 4 91164188
Email: latifi@ifr88.cnrs-mrs.fr

Received: 2 February 2011
Accepted: 17 April 2011

New Phytologist (2011)
doi: 10.1111/j.1469-8137.2011.03774.x

Key words: *Anabaena*, cyanobacteria, oxidative stress, peroxiredoxin, phylogeny, sulfiredoxin.

- The overoxidation of 2-Cys peroxiredoxins (Prxs) into a sulfinic form was thought to be an irreversible protein inactivation process until sulfiredoxins (Srxs) were discovered. These are enzymes occurring among eukaryotes, which are able to reduce sulfinylated Prxs. Although Prxs are present in the three domains of life, their reduction by Srxs has been described only in eukaryotes so far.
- Here it was established that the cyanobacterium *Anabaena* PCC 7120 has a Srx homologue (SrxA), which is able to specifically reduce the sulfinic form of the 2-Cys Prx (PrxA) both *in vivo* and *in vitro*. A mutant lacking the *srxA* gene was found to be more sensitive than the wild type to oxidative stress.
- Sulfiredoxin homologues are restricted to the cyanobacterial and eukaryotic genomes sequenced so far. The present phylogenetic analysis of Srx and 2-Cys Prx sequences showed a pattern of coevolution of the enzyme and its substrate that must have involved an ancient gene transfer between ancestors of Cyanobacteria and Eukaryotes, followed by a more recent transfer from Cyanobacteria to Plantae through the chloroplastic endosymbiosis.
- This is the first functional characterization of a Srx enzyme in a prokaryotic organism.

Introduction

Peroxiredoxins (Prxs) are ubiquitous thiol-specific peroxidases that catalyse the reduction of H₂O₂, alkyl hydroperoxides and peroxyxynitrite, using thioredoxin and other thiol-containing reducing agents as electron donors (Dietz, 2007; Poole, 2007). The first step in ROS (reactive oxygen species) reduction by Prxs involves an *N*-terminal cysteine residue in which the thiol side-chain is oxidized into sulfenic acid (SOH). In the second step, the sulfenic acid is re-reduced to a thiol before the next catalytic cycle begins. Among Prxs, typical 2-Cys Prxs are obligate homodimers having a well-known action mechanism (Konig *et al.*, 2003; Guimaraes *et al.*, 2005; Hall *et al.*, 2009). In eukaryotes, the sulfenic form of the 2-Cys Prxs can undergo a further oxidation step, yielding a sulfinic acid form (SOOH). The reduction of the sulfinic form and the subsequent recycling of the active form of 2-Cys Prxs are catalysed by a specific reductase named sulfiredoxin (Srx). After being first characterized in yeast (Biteau *et al.*,

2003), Srxs were identified in several eukaryotes, including plants and humans (Chang *et al.*, 2004; Liu *et al.*, 2006; Rey *et al.*, 2007). The mechanism underlying the Srx-catalyzed reduction of overoxidized Prxs has been thoroughly described (Biteau *et al.*, 2003; Jeong *et al.*, 2006; Jonsson *et al.*, 2008a,b): it involves a conserved cysteine residue and requires ATP hydrolysis, Mg²⁺, and a thiol serving as an electron donor. The overoxidation of 2-Cys Prxs and their reduction by sulfiredoxins were assumed for a long time to be eukaryotic features coselected in the course of evolution, giving Prxs a regulatory function in eukaryotes (Biteau *et al.*, 2003; Wood *et al.*, 2003; Jacob *et al.*, 2004; Bozonet *et al.*, 2005; Jonsson & Lowther, 2007). In prokaryotes, Prxs are thought to be required only for ROS detoxification purposes (Wood *et al.*, 2003). It has therefore been assumed that bacteria do not contain any active Srxs (Biteau *et al.*, 2003; Findlay *et al.*, 2005; Jonsson & Lowther, 2007). However, it was recently reported that 2-Cys Prxs from the cyanobacteria *Synechocystis* PCC 6803 and *Anabaena* PCC 7120 are

sensitive to overoxidation (Pascual *et al.*, 2010). In addition, the fact that overoxidation of the *Anabaena* 2-Cys Prx into a SOOH form can be reversed *in vivo* suggests the presence of Srx activity in this bacterium (Pascual *et al.*, 2010). However, to our knowledge, no such enzyme activity has ever been characterized so far in cyanobacteria. The results obtained here show that the cyanobacterium *Anabaena* PCC 7120 actually possesses a Srx enzyme (SrxA) which is homologous to the eukaryotic Srxs, and that it is able to reduce the SOOH of the 2-Cys Prx (PrxA) both *in vivo* and *in vitro*. Our data also show that Srx occurs in most cyanobacteria. Phylogenetic analyses showed that the enzyme (Srx) and its substrate (2-Cys Prx) have coevolved with and that an early horizontal gene transfer (HGT) of both genes must have occurred between the ancestors of present-day eukaryotes and cyanobacteria. This early transfer was followed later by a second HGT from cyanobacteria to the ancestor of the Plantae via the primary chloroplastic endosymbiosis, which explains the close similarities observed between the cyanobacterial and most plant sequences.

Materials and Methods

Bacterial strains, plasmids and growth conditions

Anabaena sp. PCC 7120 was grown in BG11 medium at 30°C in air under continuous illumination (40 $\mu\text{mol m}^{-2} \text{s}^{-1}$). To analyse the effects of high light intensities, *Anabaena* strains were grown under 150 $\mu\text{mol m}^{-2} \text{s}^{-1}$. Cyanobacterial growth was monitored by measuring the absorbance at 750 nm (OD 750).

The *srxA* knock-out mutant (strain $\Delta srxA$) was generated by homologous recombination, substituting the spectinomycin cassette for the *Anabaena srxA* gene (asl4146). A PCR fragment containing the spectinomycin resistance conferring cassette flanked by 500 bp upstream and downstream *srxA* was generated using the following primers: *srxA* Δ 1 primer 5'-TACCTAAAACCAACAACCTCTG-3', *srxA* Δ 2 primer 5'-ACTCGAGAGGGATCCTCGAAT TCAGGAGTAAACTCAACTAGTTCCTT-3', *srxA* Δ 3 primer 5'-GAATTCGAGGATCCCTCT CGAGTGTAGAGAATAGGAAAAGGCAGG-3' and *srxA* Δ 4 primer 5'-AAAATAGAACGTGCTGTTCTCA-3'. After a subcloning step in the Bluescript SK vector, this fragment was cloned into the *SacI* site of the conjugal suicide plasmid pRL271. The recombinant plasmid obtained was transferred to *Anabaena* using the previously described conjugation method (Cai & Wolk, 1990). *Anabaena* recombinant clones were selected on BG11 supplemented with spectinomycin at 20 $\mu\text{g ml}^{-1}$. Double recombinants were selected on the basis of their ability to grow in the presence of sucrose. Complete segregation of the mutation was confirmed by PCR. Three independent conjugations were performed and clones obtained from each experiment were used in all subsequent studies on the mutant.

Expression and purification of recombinant proteins

A DNA fragment corresponding to the entire coding region of alr4641 was amplified by performing PCR using the PrxA forward primer 5'-CATATGTCCATCACCTA-3' (*NdeI* site underlined) and the PrxA reverse primer 5'-CTCGAGTTACACAGCAGCGAAGTA-3' (*SacI* site underlined). The PCR product was cloned into the pET28 vector (Novagen, Darnstadt, Germany). A clone confirmed by DNA sequencing was transformed into the BL21DE3 *Escherichia coli* strain (Novagen). The recombinant clones obtained were grown in kanamycin-supplemented medium to an optical density (OD) of 0.3–0.4 and protein expression was induced by adding 1 mM isopropyl- β -D-thiogalactoside (IPTG) for 4 h. The recombinant proteins were purified using Hitrap columns as recommended by Healthcare. Imidazol was removed from purified proteins using PD10 columns (Healthcare, Orsay, France). Proteins were concentrated on Vivaspin columns and used for subsequent analyses.

For SrxA production purposes, a DNA fragment corresponding to the entire coding region of asl4146 was amplified by performing PCR using the SrxA forward primer 5'-GAATTCATGGTTAGGGTACAAGAAATT-3' (*EcoRI* site underlined) and the SrxA reverse primer 5'-AAGCTTCTACGCTAAGTGCATCTTC-3' (*HindIII* site underlined). The PCR product was cloned into the pMAL-c2X plasmid (BioLabs, Charbonnières-les-Bains, France). JM109 *E. coli* strain (Promega) was transformed with the pMAL-c2X-*srxA* plasmid. The recombinant clones obtained were grown in ampicillin-supplemented medium up to an OD of 0.3–0.4 and protein expression was induced by adding 1 mM IPTG for 4 h. Cells were collected by centrifugation and lysed by sonication in a buffer containing Tris-HCl (pH 7.4) and protease inhibitor cocktail (Roche, Every, France). The recombinant proteins were first purified using affinity column (amylose resin, NEB) according to the manufacturer's instructions, and then further purified using PD10 columns (Healthcare). The maltose binding protein tag (MBP) was removed in line with the manufacturer's instructions (Healthcare). The cysteine residue at position 57 of SrxA was changed into a serine residue by performing PCR using the megaprimer strategy (Wu *et al.*, 2005). The primers used for this purpose were: SrxA forward, SrxA reverse and SrxA mut primer 5'-ATATCTATGGCTACCAG-3'.

Sodium dodecyl sulfate–polyacrylamide gel electrophoresis (SDS-PAGE) and immunoblot analysis

PrxA and SrxA proteins were fractionated by SDS-PAGE (12.5% gel) and stained using the SeeBand procedure (Euromedex, Souffelweyrshim, France). When SDS-PAGE was performed under nonreducing conditions, 2-mercaptoethanol was omitted from the loading buffer. For

immunoblot analysis, the proteins were transferred onto nitrocellulose membranes for revelation with polyclonal antibodies. Immune complexes were detected with anti-rabbit peroxidase-conjugated secondary antibodies (Promega) and enhanced chemiluminescence reagents (Pierce, Illklich, France). Antibodies raised against overoxidized human PrxI (α -Prx-SO₃; Abcam, Paris, France) were diluted to 1 : 5000. Antibodies specific to Arabidopsis 2-Cys Prx were diluted to 1 : 5000. Antibodies raised against the ferredoxin-NADP⁺ reductase (FNR) from *Anabaena* PCC 7119 were diluted to 1 : 5000 (Razquin *et al.*, 1996). Antibodies specific for His-tag (Invitrogen) were diluted to 1 : 5000.

Reactive oxygen species detection

The ROS generated by *Anabaena* cells were analysed after performing reactions with 2,7-DCFH-DA (Molecular Probes, Illklich, France) as described elsewhere (Jang *et al.*, 2004). The fluorescence microscope Nikon Eclipse E800 (Champigny sur Marne, France) was used to detect the fluorescence intensity emitted between 515 and 550 nm.

Pigment determination

Chlorophyll *a* was determined spectroscopically in methanol extracts. For this purpose, 100 μ l of *Anabaena* cell cultures were mixed with 1 ml of methanol and vortexed. Cell debris were pelleted by centrifugation, and the absorbance of the supernatant was measured at 665 nm (OD 655). Chlorophyll concentration was calculated in terms of the formula: $\mu\text{g ml}^{-1} = 13.43 \times \text{OD}655 \times \text{dilution factor}$.

Phycocyanin content was calculated from the difference spectrum of cell cultures determined before and after a 8-min heat treatment at 75°C (Collier & Grossman, 1992).

Preparation of sulfinic PrxA

Recombinant PrxA (5 $\mu\text{g } \mu\text{l}^{-1}$) was incubated in a 50 μ l reaction mixture containing 20 mM dithiothreitol (DTT), 3 mM H₂O₂, 50 mM Tris-HCl (pH 6.8) and 100 mM KCl. Oxidation was initiated by adding H₂O₂ to the reaction mixture and continued for 30 min at 30°C. The addition of 3 mM H₂O₂ to the reaction mixture was repeated three times.

SrxA activity assays

Sulfinic PrxA (15 ng μl^{-1}) was reduced by adding SrxA (15 ng μl^{-1}) to a 200- μ l reaction mixture containing 50 mM Tris-HCl (pH 6.8), and 100 mM KCl, 10 mM DTT, 0.5 mM MgCl₂; 1 mM ATP was added to the reaction mixture when required. After a 30-min period of incubation at 30°C, the proteins were precipitated with 10% trichloroacetic acid. Proteins were then subjected to immunoblot analysis with specific antibodies to overoxidized

human PrxI (α -Prx-SO₃; Ambion), or with antibodies to *Arabidopsis* 2-Cys Prx. Before *in vivo* reduction, mid-log *Anabaena* strains were treated with methyl viologen 50 μM for 60 min. The cell cultures were then washed three times in BG11 medium and reincubated for 1 h in BG11 medium to enable them to recover from the stress. The redox state of PrxA was analysed as explained earlier.

Dataset construction

Prokaryotic homologues of *Anabaena* sp. PCC 7120 SrxA and PrxA were retrieved in February 2010 using the BLASTP program (Altschul *et al.*, 1997) (default parameters) from a local database composed of the 1083 complete prokaryotic (1002 bacterial and 81 archaeal) genomes available at the NCBI (<ftp.ncbi.nih.gov>). BLAST outputs were checked manually to identify homologues (no arbitrary cut-off *e*-value was used). To obtain a broad sampling of eukaryotic sequences, homologues of Srx and Prx were retrieved using the BLASTP and TBLASTN programs from the *nr* database at the NCBI, from the JGI (<http://genome.jgi-psf.org/>) and TBestDB databases (<http://www.bch.umontreal.ca/pepdb/pep.html>). In addition, homologues of three representatives of Rhodophyta were retrieved from the *Porphyra yezoensis* expressed sequences tag (EST) dataset (<http://est.kazusa.or.jp/en/plant/porphyra/EST/blast.html>), the *Galdieria sulphuraria* genome project (<http://genomics.msu.edu/cgi-bin/galdieria/blast.cgi>) and the *Cyanidioschyzon merolae* Genome Project (<http://merolae.biol.s.u-tokyo.ac.jp/blast/blast.html>). The Srx and Prx homologues were finally assembled in two datasets. The corresponding sequences were aligned using CLUSTALW 2.0.12 (Larkin *et al.*, 2007). The resulting alignments were then inspected visually and refined manually using the ED program from the MUST package (Philippe, 1993). Before phylogenetic analyses, regions of doubtful homology were removed manually from the alignments using NET from the MUST package (datasets are available on request from L. E. or C. B-A.).

The cellular location of each eukaryotic Srx and 2-Cys Prx from photosynthetic organisms was investigated using information available in the Genpept entries combined with bioinformatic predictions of chloroplast transit peptides (CTP) obtained using various software programs, such as CHLOROP (Emanuelsson *et al.*, 1999), TARGETP (Emanuelsson *et al.*, 2007), IPSORT (Bannai *et al.*, 2002), WOLF PSORT (Horton *et al.*, 2007) and PCLR, Chloroplast Localization Prediction release 0.9 (Schein *et al.*, 2001).

Phylogenetic analyses

Maximum likelihood (ML) phylogenetic trees were computed with PHYML 3.0 (Guindon & Gascuel, 2003) using the Le and Gascuel (LG) model (Le & Gascuel, 2008) and a gamma correction to take into account the heterogeneity

of the evolutionary rates across sites (using four discrete classes of sites and an estimated alpha parameter). Branch robustness of the resulting trees was estimated using the nonparametric bootstrap procedure implemented in PHYLIP using 100 replicates of the original dataset and the same parameters as for tree reconstruction. Additional ML phylogenetic analyses were performed using TREEFINDER and the same evolutionary model and parameters as for PHYLIP (Jobb *et al.*, 2004). Bayesian trees were computed using MRBAYES 3.1.2 (Ronquist & Huelsenbeck, 2003) with a mixed amino acid substitution model and a gamma correction (using four discrete classes of sites and an estimated alpha parameter). The Markov chain Monte Carlo search was run with four chains on 1 000 000 generations, and the trees were sampled every 100 generations (the first 2500 trees were discarded as 'burning' trees).

Results

srxA and *prxA* genes are induced in response to oxidative stress

The ORF Asl4146 (SrxA) present in *Anabaena* genome (<http://www.kazusa.or.jp/cyano/>) encodes a protein that is homologous to eukaryotic Srxs (Kaneko *et al.*, 2001). SrxA is an 87-residue protein with predicted molecular mass of 9.7 kDa and a predicted isoelectric point of 9.39. It shows higher levels of identity with proteins found in the land plant *Arabidopsis thaliana* (59%) than with yeast (42%) and human homologues (40%). The transcription of *srxA* gene is induced when *Anabaena* cells are challenged by oxidative conditions such as methyl viologen treatment (Fig. 1a) and strong light stress (Fig. 1b). The expression of the *prxA* gene encoding the *Anabaena* 2-Cys Prx was also induced under the same conditions (Fig. 1). It is worth noting here that the results obtained in Fig. 1 could also indicate that a possible degradation of *srxA* and *prxA* mRNAs is inhibited upon oxidative stress.

A mutant ($\Delta srxA$) from which the *srxA* gene was deleted was constructed, as explained in the Materials and Methods section. A PCR analysis indicated that the mutant was completely segregated (data not shown). Semi-quantitative reverse-transcription polymerase chain reaction (RT-PCR) analyses performed with RNAs extracted from this mutant showed that, contrary to what was observed for the wild type, the *srxA* transcripts were not detected at all when mutant cells were grown under oxidant conditions (Fig. 2a). This finding further confirmed that the *srxA* gene was completely inactivated in the mutant. When cells were exposed to high light intensities or hydrogen peroxide stress, the growth rate of the $\Delta srxA$ mutant did not differ significantly from that of the wild-type strain. However, when the strains were concomitantly subjected to strong light and hydrogen peroxide stresses, the mutant showed a much slower rate of growth

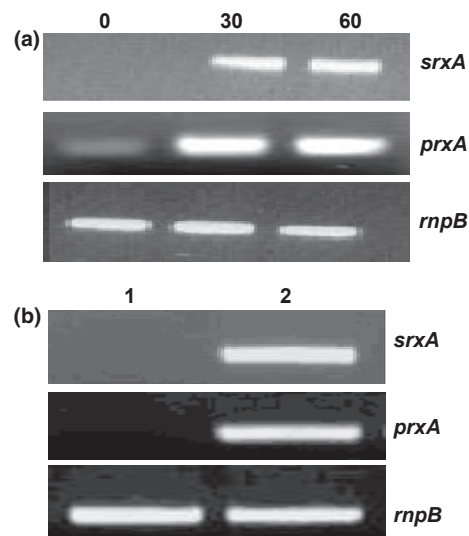


Fig. 1 *srxA*, encoding a prokaryotic sulfiredoxin, is induced under oxidant conditions. (a,b) Reverse-transcription polymerase chain reaction (RT-PCR) analysis of *srxA* and *prxA* gene transcription. RNA was collected from cells grown in BG11 medium (a, column 0) or in BG11 incubated for 30 min (a, column 30), or 1 h (a, column 60) with 50 μ M of methyl viologen (a), or 1 h after transferring the cell culture from normal lighting conditions (b, column 1) to high light (b, column 2). Samples were collected during the exponential phase of the PCR. All RT-PCR experiments were performed in triplicate, and similar results were consistently obtained. Expression of the *rnpB* gene was used as the control assay.

than the wild type strain (Fig. 2b). Deletion of the *srxA* gene led to the strong accumulation of ROS (Fig. 3), which indicates that the $\Delta srxA$ mutant is less able than the wild type to reduce ROS. To further characterize the phenotype of the $\Delta srxA$ mutant, the light-harvesting pigments chlorophyll *a* and phycocyanin were quantified, because the mutant was found to have a less green appearance than the wild-type strain. Photobleaching has been reported to occur in cyanobacteria in response to high light stress (Andreeva *et al.*, 2007; Jeanjean *et al.*, 2008). When exposed to strong illumination conditions, the mutant cells were found to contain 75% less Chl*a* and 60% less phycocyanin than the wild-type cells (Table 1).

SrxA reduces sulfinic PrxA *in vitro*

To investigate the activity of SrxA, the *srxA* coding region was overexpressed in *E. coli* as a translation fusion protein with the MBP. After an amylose-affinity chromatography purification step, the recombinant protein recovered in the soluble fraction was found to be relatively homogeneous (Fig. 4a). His-tagged PrxA was expressed in *E. coli* and purified by performing Nickel-affinity chromatography. Under reducing conditions, the monomeric form of the recombinant protein, which was the predominant form observed, had the expected size of 25.5 kDa (which corresponds to PrxA at 25 kDa and the His-tag at 0.5 kDa) (Fig. 4b).

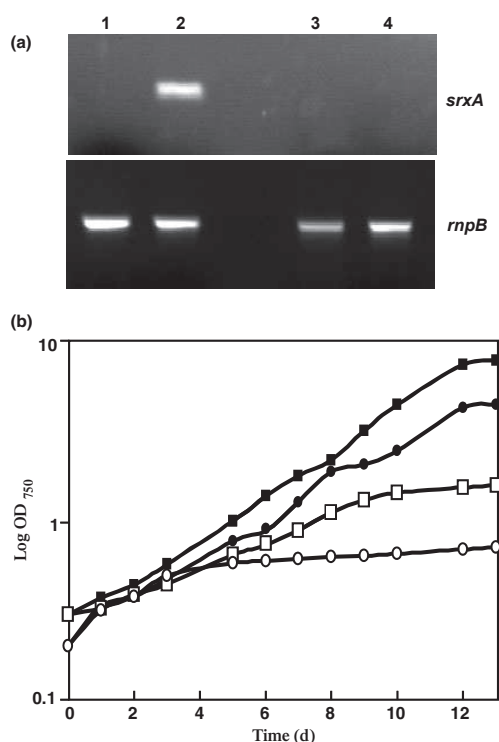


Fig. 2 Characterization of the $\Delta srxA$ mutant. (a) Reverse-transcription polymerase chain reaction (RT-PCR) analysis of *srxA* gene transcription in the $\Delta srxA$ mutant. RNA from the wild-type strain (columns 1 and 2) or from the $\Delta srxA$ mutant (columns 3 and 4) was collected from cells grown in BG11 medium (columns 1 and 3) or in BG11 incubated for 1 h with 50 μM of methyl viologen (columns 2 and 4). Samples were collected during the exponential phase of the PCR. All RT-PCR experiments were performed in triplicate, and similar results were consistently obtained. Expression of the *rnpB* gene was used as the control assay. (b) Representative experiment showing growth of the wild-type (squares) and $\Delta srxA$ (circles) mutant strains under normal lighting in BG11 without hydrogen peroxide (H_2O_2 , closed symbols), or at high light in BG11 medium with 50 μM of H_2O_2 (open symbols).

To examine the ability of SrxA to catalyse the reduction of the sulfinylated form of 2 Cys-Prx, PrxA was subjected to overoxidation, as described in the Materials and Methods section. Overoxidized PrxA is denoted Ox-PrxA. The sulfinylation of PrxA was assessed using antibodies (α -Prx-SO₃) directed against the sulfonic form of the human 2-Cys Prx. These antibodies recognize purified Ox-PrxA (Fig. 4c). These antibodies were therefore used to monitor the possible reduction of Ox-PrxA by SrxA in the presence of ATP and Mg²⁺. Under these conditions, the band revealed by α -Prx-SO₃ disappeared completely after a 30-min period of incubation at 30°C, which indicates that SrxA is able to reduce the sulfinic form of PrxA. As in the case of the eukaryotic Srxs, the action of SrxA depends on the presence of ATP: only the overoxidized form (Ox-PrxA) was detected when ATP was omitted from the reaction mixture (Fig. 4c). All the eukaryotic Srx enzymes described so far contain a conserved cysteine residue, which is required for Prx-reduction.

In *Anabaena* Srx, the conserved cysteine is Cys⁵⁷. Replacing SrxA Cys⁵⁷ by a serine residue totally abolished the reductase activity (Fig. 4d), which provides further evidence that the catalytic mechanism involving SrxA is similar to that involving eukaryotic Srxs.

SrxA reduces the sulfinic form of PrxA *in vivo*

In order to determine whether SrxA is active *in vivo*, PrxA oxido-reduction was compared between the wild strain and the $\Delta srxA$ mutant. In the wild type, PrxA was converted into a sulfinic form (Ox-PrxA) after a 30-min period of incubation with methyl viologen. Ox-PrxA was converted into its reduced form once the stress had been removed (Fig. 5a). However, in the $\Delta srxA$ mutant, the SOOH form of PrxA accumulated and was not converted into the reduced form (Fig. 5b), which clearly indicates that SrxA is involved in PrxA-Cys-SOOH recycling processes *in vivo*. The evidence obtained here that *Anabaena* contains an active Srx challenges the belief that these enzymes play an important role in eukaryotes alone, and shows that they are involved in oxidative stress responses in cyanobacteria.

2-Cys Prxs are more widely distributed than sulfiredoxins

To investigate the evolution of 2-Cys Prxs and Srxs, an extensive phylogenetic analysis of these proteins was performed. A survey of sequences databases yielded 1329 Prxs harboring the two cysteine residues characteristic of 2-Cys Prxs (at positions 56 and 178 in the *Anabaena* PrxA sequence). 2-Cys Prxs are present in the complete genomes of many prokaryotes (including all genomes of the Cyanobacteria) and in those of most major lineages of Eucarya for which genomic data are available (Metazoa, Fungi, Amoebozoa, Excavata, Alveolata, Heterokonta and Plantae (i.e. Glaucophytes, Rhodophyta, green algae and land plants; Fig. 6). Phylogenetic analyses showed that the 2-Cys Prx sequences can be divided into two groups: one containing most of the bacterial and archaeal sequences, and the other consisting of sequences from Cyanobacteria, Eucarya and a few other bacteria. Interestingly, only sequences from the latter group harbored the two specific motifs (GG(LIV)G and (YF)(FWL)) previously reported to be a signature of these enzymes' ability to undergo overoxidation (Wood *et al.*, 2003). Accordingly, these sequences will be referred subsequently as sensitive 2-Cys Prxs. These motifs are located at positions 99–102 and 199–200 in the *Anabaena* PrxA sequence, respectively. By contrast, the taxonomic distribution of Srxs is much narrower than that of Prxs. Only 109 Srx homologues belonging to Cyanobacteria and to some eukaryotic lineages (i.e. Metazoa, Fungi, Amoebozoa and land plants) were detected (Fig. 6), in agreement with previous data (Basu & Koonin, 2005).

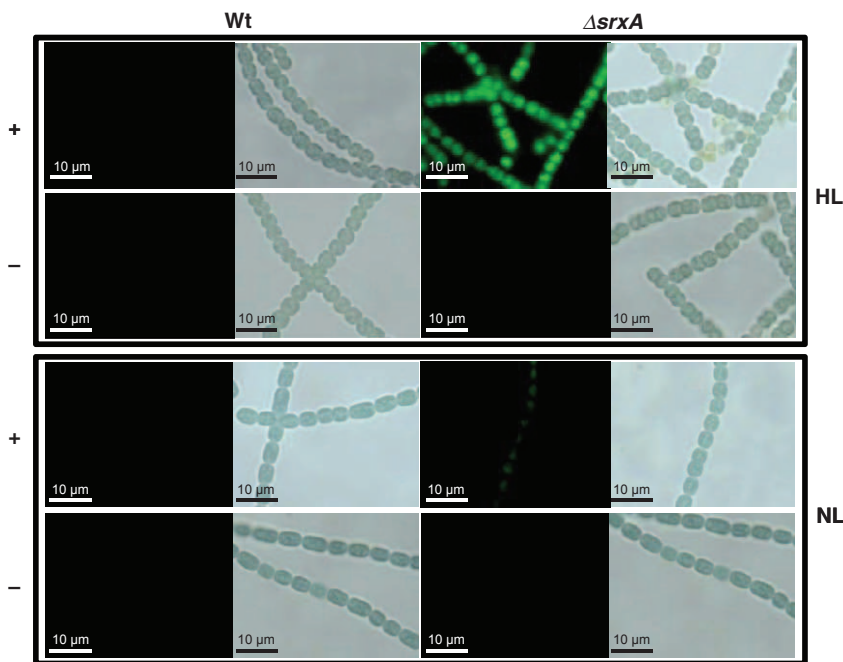


Fig. 3 SrxA contributes to oxidative stress response. Fluorescence micrographs (right columns), and light transmission micrographs (left columns), of *Anabaena* PCC 7120 wild-type strain (Wt) or $\Delta srxA$ mutant ($\Delta srxA$). Mid-log grown cells were incubated under either under normal (NL) or high light conditions (HL) for 24 h. They were then treated (+) or not (–) with the reactive oxygen species (ROS)-detecting probe CM-H₂DCFDA. This experiment was performed with three $\Delta srxA$ mutant strains obtained from three independent conjugation events, and similar results were obtained.

Table 1 Light-dependent decrease in the chlorophyll a (Chla) and phycocyanin contents of *Anabaena* wild-type (Wt) and $\Delta srxA$ mutant cells

Cell type and growth conditions	Amount of Chla ¹	Amount of Phycocyanin ²
Wt, NL	5.5 ± 0.32	250 ± 12
$\Delta srxA$, NL	4.9 ± 0.28	224 ± 15
Wt, HL	4.4 ± 0.19	183 ± 11
$\Delta srxA$, HL	1.1 ± 0.25	70 ± 9

Mid-log grown cells were incubated for a further 48 h under either normal (NL: 40 $\mu\text{mol m}^{-2} \text{s}^{-1}$) or higher light intensities (HL: 150 $\mu\text{mol m}^{-2} \text{s}^{-1}$).

¹Expressed as micrograms per OD₇₅₀ unit of cell suspension.

²Expressed as $(\Delta\text{OD}_{620} - \Delta\text{OD}_{750}) \times 10\,000 / \text{OD}_{750}$. ΔOD_{620} and ΔOD_{750} are the differences between the OD values recorded at the respective wavelengths before and after incubating the cell suspensions at 75°C (Collier & Grossman, 1992).

Values are means of four independent determinations (± SD).

However, based on the present genomic data, Srxs can now be said to occur in green and red algae, glaucophytes and representatives of the large protist group of Excavates (Fig. 6). However, no Srx homologues were detected in Chromalveolata (i.e. Alveolata, Heterokonta, Haptophyta or Cryptophyta).

In order to obtain further insights into the evolutionary history of these two proteins, an accurate phylogenetic analysis was performed on a subset of Srx and sensitive 2-Cys Prx sequences reflecting the taxonomic diversities of these enzymes. The resulting maximum likelihood and Bayesian trees showed similar topologies; both Srxs and sensitive 2-Cys Prxs trees show two well distinct clusters: the first

group contains most of the eukaryotic sequences (including some Plantae sequences and a few bacterial sequences in the case of Prxs, probably resulting from recent horizontal gene transfer (HGT) events), whereas the second gathers all the cyanobacterial sequences and other Plantae sequences (BV = 87% and PP = 1.0 in Srx, and BV = 97% and PP = 1.0 in Prx, Fig. 6). Therefore, Plantae representatives can be said to differ from other eukaryotes in that they are the only lineage which harbors Srxs and sensitive 2-Cys Prx sequences from both clusters.

The relationships between eukaryotic Srxs and between sensitive 2-Cys Prxs sequences are poorly resolved (BV < 50% and PP < 0.5, Fig. 6). However, the fact that these enzymes are present in most eukaryotic lineages suggests that Srxs and sensitive 2-Cys Prxs have existed for a long time in this domain, and that they were probably present in the last common ancestor of eukaryotes (LECA, closed squares and triangles in Fig. 7c). This indicates that the few plants harboring sequences of this kind (such as the glaucocystophyte *Cyanophora paradoxa* and two green algae, *Chlamydomonas reinhardtii* and *Volvox carteri*, Fig. 6) have inherited them from LECA (closed squares and triangles, Fig. 7). Upon carefully examining the sequences forming the second cluster, it was observed that both 2-Cys Prxs and Srxs are present in most cyanobacterial lineages, especially in deep branching ones such as *Synechococcus* sp. JA-2-3Ba2-13 (Fig. 6). This suggests that sensitive 2-Cys Prxs and Srxs may have existed in cyanobacteria for a long time and that they may have been present in their last common ancestor (open squares and triangles in Fig. 7b). It is worth noting that most Plantae Srxs and sensitive 2-Cys Prxs were found to be closely related to cyanobacterial sequences and

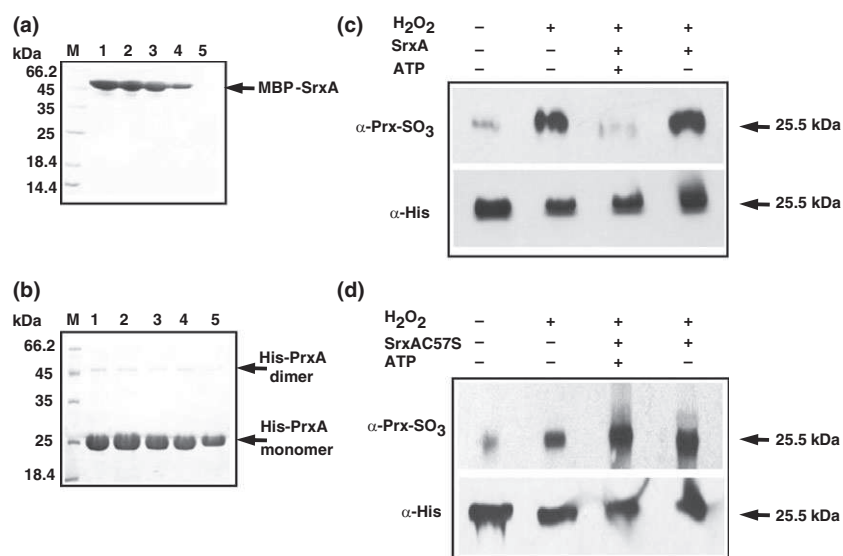


Fig. 4 SrxA reduces Ox-PrxA *in vitro* (a,b) Purification of recombinant SrxA and PrxA proteins. SrxA was purified as a maltose-binding protein (MBP)-tagged protein and PrxA as a His-tagged fusion protein. After an affinity chromatography purification step, MBP-SrxA (a) and His-PrxA (b) were further purified using PD10 columns. Proteins eluted showed good levels of homogeneity in the 4th (a) and 5th (b) elution fractions. Proteins were separated by performing sodium dodecyl sulfate–polyacrylamide gel electrophoresis (SDS–PAGE) (12.5% acrylamide gel) and stained with SeeBand. Lanes: M, molecular size markers; 1–5, PD10 elution fractions. (c,d) Western blot analysis of the sulfenic form of PrxA. Overoxidized PrxA was obtained as specified in the Materials and Methods section. *In vitro* reduction assays were performed with SrxA (c) or SrxAC57S (d) on the Ox-PrxA thus obtained. Untreated PrxA protein was used as the control substance (–H₂O₂ lane). Where indicated, ATP was omitted from the reaction mixtures. Western blotting performed with His-probe antibodies (Invitrogen) served to check the amounts of protein used in each assay.

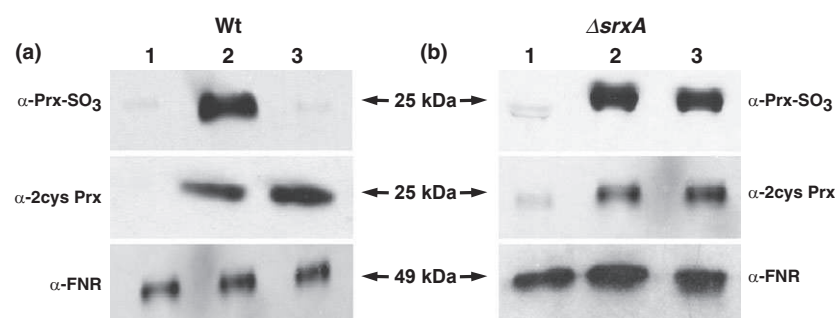


Fig. 5 SrxA recycles Ox-PrxA *in vivo* Western blot analysis of PrxA in the wild-type *Anabaena* strain (a) and in the *ΔsrxA* mutant (b). Cells were grown in BG11 medium up to the mid-log growth phase (1) and incubated with 50 μM methyl viologen for 1 h (2). Methyl viologen treated cells were left to recover from the stress as explained in the Materials and Methods section (3). Samples of 50 μg of proteins were used in each assay. The antibodies directed against the 2-Cys Prx from *Arabidopsis thaliana* served to detect the PrxA protein, and anti-ferredoxin-NADP⁺ reductase (FNR) antibodies served to control the amount of protein used in each assay.

to be only distantly related to sequences found in other eukaryotes (Fig. 6). This indicates that these Plantae sequences have a different evolutionary origin. The most parsimonious hypothesis accounting for the close relationship between Plantae and cyanobacterial sequences is that a HGT has occurred between those two lineages.

Discussion

The present *in vivo* and *in vitro* analysis of *Anabaena* 2-Cys Prx (PrxA) oxidation processes showed the presence in this organism of a sulfiredoxin (SrxA), which is similar to those

present in eukaryotes. SrxA reduces Ox-PrxA via a similar mechanism to that occurring in eukaryotes and thus helps *Anabaena* to cope with oxidative stress. It was therefore concluded that Srxs are not restricted to eukaryotes alone, and that SrxA is a prokaryotic member of the Srxs enzyme family. As some other cyanobacterial strains have sulfiredoxin homologs (triangles in Fig. 7), it is tempting to suggest that sulfinylation and reduction of sensitive 2-Cys Prxs may also occur in these organisms. Transcription of the *srxA* gene can be strongly induced under various conditions resulting in oxidative damage. It is therefore rather surprising that the lack of this gene significantly affects the bacterial growth

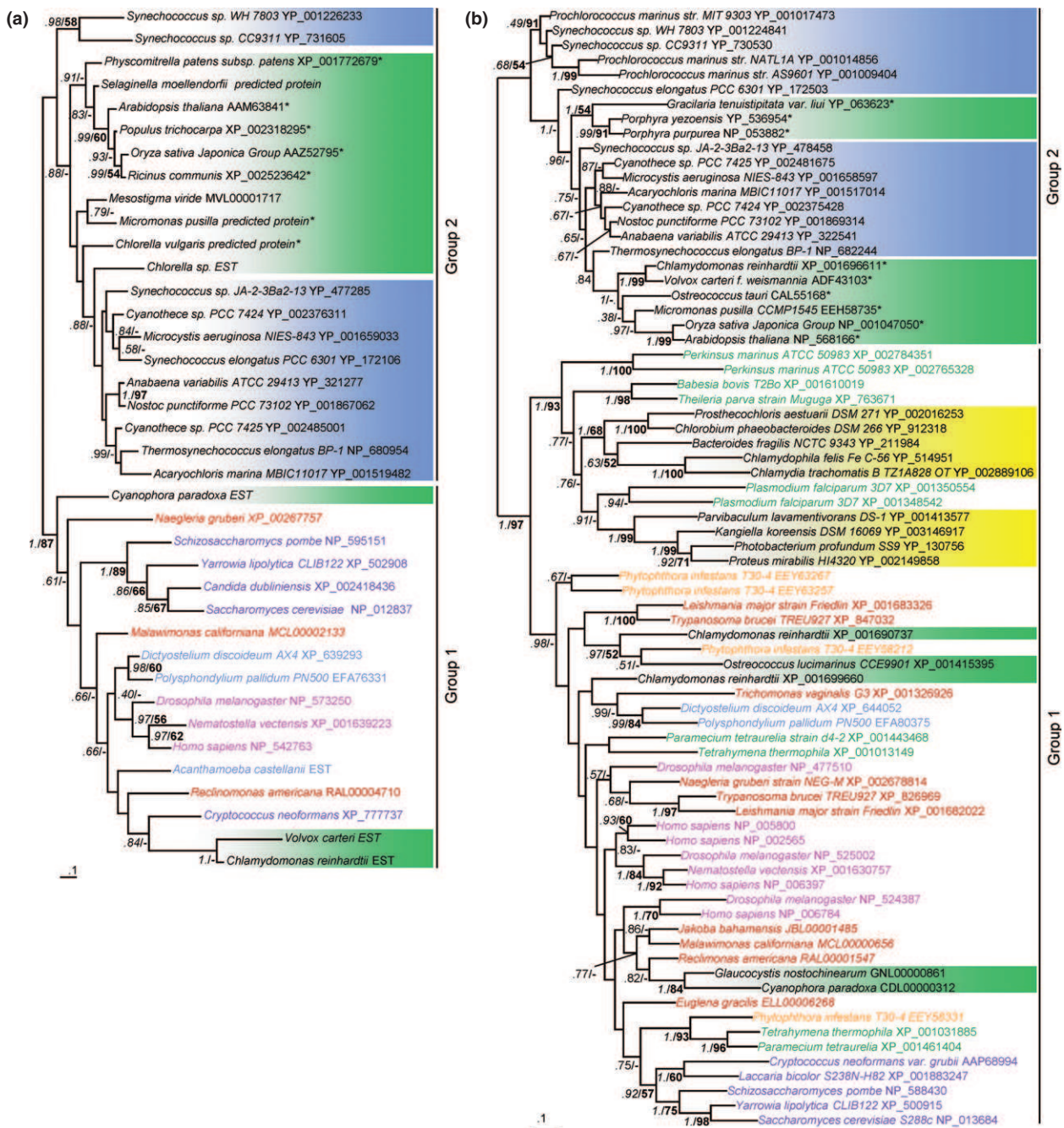


Fig. 6 SrX and PrxA are closely related to plant enzymes. Rooted Bayesian trees of SrX (74 positions, 38 sequences) (a) and sensitive 2-Cys Prx (167 positions, 76 sequences) (b) homologues. The SrX tree was rooted using a distant homologue, namely the chromosome partitioning ParB protein as an outgroup (Basu & Koonin, 2005). The sensitive 2-Cys Prx phylogeny was rooted using nonsensitive 2-Cys Prx homologues as an outgroup (not shown). Green, blue and yellow boxes give Plantae, cyanobacterial and other bacterial sequences, respectively. Other eukaryotic groups are presented in fuchsia (Metazoa), purple (Fungi), blue (Amoebozoa), dark red (Excavata), dark green (Alveolata) and orange (Heterokonta) fonts. Plant sequences harboring a chloroplastic target peptide are indicated by an asterisk. Numbers at nodes are posterior probabilities (PP)/bootstrap values (BV). For the sake of clarity, only PP and BV values > 0.5% and 50%, respectively are indicated. The scale bars give the average number of substitutions per site.

rates only under accumulated stresses. Similar results have been obtained with the SrX mutant of *A. thaliana*, the growth of which was impaired only under high light and low temperature conditions (Rey *et al.*, 2007). The SrXs are

important players in the regulation of Prxs, but their role becomes essential only under severe oxidative conditions.

Our sequence database survey showed that both sensitive 2-Cys Prxs and SrXs have a similar evolutionary history,

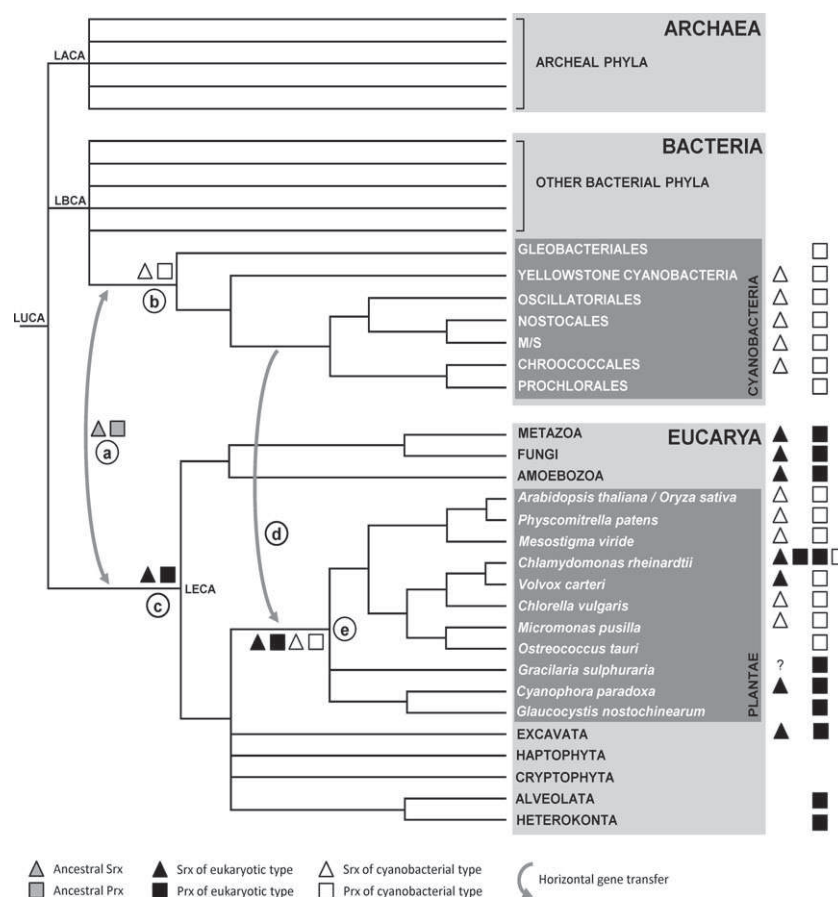


Fig. 7 Evolution of Srxs and sensitive 2-Cys Prxs. Possible scenario of the origin and evolutionary history of Srx and sensitive 2-Cys Prx sequences. The tree gives the three domains of cell life (Archaea, Eucarya and Bacteria). LECA, LACA, LBCA and LUCA are, respectively: the Last Eukaryotic, Archaeal, Bacterial and Universal Common Ancestor. The Srx (tinted triangle) and sensitive 2-Cys Prx (tinted square) sequences originated from the ancestor of either Cyanobacteria or Eucarya and were horizontally transferred to the other lineage before their diversification (a). Srx and sensitive 2-Cys Prx sequences were conserved during evolution and were present in the ancestor of Cyanobacteria (b, open triangle and square) and in the ancestor of Eucarya (c, closed triangle and square). The cyanobacterial Srx and 2-Cys Prx sequences were acquired secondarily by Plantae via the chloroplastic endosymbiosis (d). After this event, either eukaryotic or cyanobacterial copies of Srx and 2-Cys Prx were lost during the evolution and diversification of the present-day Plantae lineages (e). The late acquisition of sensitive 2-Cys-Prxs is not indicated (see text for details).

which means that they have coevolved. This result is not surprising as these enzymes are functionally linked. It is worth mentioning that all the genomes containing a Srx homologue also encode a sensitive 2-Cys Prx homologue (Fig. 7). However, the opposite is not true because some sensitive 2-Cys Prx coding genes were found to exist in genomes devoid of *srx* homologues (such as *Synechocystis* PCC 6803, and Chromalveolata, Fig. 6 and Table S1). This suggests either that some other hitherto uncharacterized proteins may reduce sulfinic Prxs, or that the presence of the overoxidation motif in a 2-Cys Prx does not necessarily mean that this protein is prone to sulfinylation. The fact that the 2-Cys Prx of *Synechocystis* PCC 6803, which lacks Srx, is much less sensitive to overoxidation than the 2-Cys Prx of *Anabaena-PrxA* (Pascual *et al.*, 2010) argues in favor of the second possibility.

The taxonomic distribution of 2-Cys Prxs and Srxs observed here is rather puzzling, and raises questions about the origin of these enzymes. It has been previously suggested that Cyanobacteria may have acquired their Srxs via an HGT event from plants (Basu & Koonin, 2005). Since the latter study, many more genome sequences have become available, and the evolutionary history of these proteins was therefore reassessed here. Our data indicate that both Srx and sensitive 2-Cys Prx were probably present in the last common ancestor of Eucaryotes (Fig. 7c), as well as in the last common ancestor of present-day Cyanobacteria (Fig. 6b). This strongly suggests that an ancient HGT event may have occurred between these two lineages before their diversification (tinted square and triangle, Fig. 7a). In addition to this event, a second HGT event may have occurred between cyanobacteria and plants, which would explain the presence of a

cyanobacterial copy of these enzymes in plants (Fig. 7d): the presence of two distinct versions of 2-Cys Prxs and Srxs in plants does not support the hypothesis that an HGT may have occurred from plants to cyanobacteria, but rather suggests that one of the two Srx and 2-Cys Prx copies may be of chloroplastic origin. The presence of a chloroplastid target peptide in plant sequences closely related to cyanobacterial Srx and sensitive 2-Cys Prx, and the fact that these enzymes were found to be localized in the chloroplast in several species (Baier & Dietz, 1997; Rey *et al.*, 2007), are consistent with this evolutionary scenario. The diversity of the Srx and 2-Cys Prx repertoire observed in present-day plants therefore results from the subsequent loss of one of the two copies inherited from either Cyanobacteria or LECA (either closed or open squares and triangles, Fig. 7), or both.

Why do some cyanobacteria require a Srx and not others? The answer to this question will shed light on the strategies developed by these organisms to cope with oxidative stress. If having a Srx serves the same purpose in both cyanobacteria and eukaryotes, it might be possible to conclude that in cyanobacteria strains carrying a Srx, the peroxide signaling processes may involve a 2-Cys Prx. Therefore, studying the peroxide perception mechanisms and those involved in the regulation of the appropriate responses between cyanobacteria with and without sulfiredoxin is the next challenge that needs to be met in this field.

Acknowledgements

The authors thank P. Rey for providing antibodies directed against the *Arabidopsis* 2-Cys Prx and Jessica Blanc for revising the English manuscript. C.B-A. was supported in this project by an 'Action Thematique et Incitative sur Programme' (ATIP CNRS). This work was supported by an 'Environnement et santé' grant (AFSSE).

References

- Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Research* **25**: 3389–3402.
- Andreeva A, Abarova S, Stoitchkova K, Picorel R, Velitchkova M. 2007. Selective photobleaching of chlorophylls and carotenoids in photosystem I particles under high-light treatment. *Photochemistry and Photobiology* **83**: 1301–1307.
- Baier M, Dietz KJ. 1997. The plant 2-Cys peroxiredoxin BAS1 is a nuclear-encoded chloroplast protein: its expressional regulation, phylogenetic origin, and implications for its specific physiological function in plants. *Plant Journal* **12**: 179–190.
- Bannai H, Tamada Y, Maruyama O, Nakai K, Miyano S. 2002. Extensive feature detection of N-terminal protein sorting signals. *Bioinformatics* **18**: 298–305.
- Basu MK, Koonin EV. 2005. Evolution of eukaryotic cysteine sulfonic acid reductase, sulfiredoxin (Srx), from bacterial chromosome partitioning protein ParB. *Cell Cycle* **4**: 947–952.
- Biteau B, Labarre J, Toledano MB. 2003. ATP-dependent reduction of cysteine–sulphinic acid by *S. cerevisiae* sulphiredoxin. *Nature* **425**: 980–984.
- Bozonet SM, Findlay VJ, Day AM, Cameron J, Veal EA, Morgan BA. 2005. Oxidation of a eukaryotic 2-Cys peroxiredoxin is a molecular switch controlling the transcriptional response to increasing levels of hydrogen peroxide. *Journal of Biological Chemistry* **280**: 23319–23327.
- Cai YP, Wolk CP. 1990. Use of a conditionally lethal gene in *Anabaena* sp. strain PCC 7120 to select for double recombinants and to entrap insertion sequences. *Journal of Bacteriology* **172**: 3138–3145.
- Chang TS, Jeong W, Woo HA, Lee SM, Park S, Rhee SG. 2004. Characterization of mammalian sulfiredoxin and its reactivation of overoxidized peroxiredoxin through reduction of cysteine sulfonic acid in the active site to cysteine. *Journal of Biological Chemistry* **279**: 50994–51001.
- Collier JL, Grossman AR. 1992. Chlorosis induced by nutrient deprivation in *Synechococcus* sp. strain PCC 7942: not all bleaching is the same. *Journal of Bacteriology* **174**: 4718–4726.
- Dietz KJ. 2007. The dual function of plant peroxiredoxins in antioxidant defence and redox signaling. *Sub-cellular Biochemistry* **44**: 267–294.
- Emanuelsson O, Brunak S, von Heijne G, Nielsen H. 2007. Locating proteins in the cell using TargetP, SignalP and related tools. *Nature Protocols* **2**: 953–971.
- Emanuelsson O, Nielsen H, von Heijne G. 1999. ChloroP, a neural network-based method for predicting chloroplast transit peptides and their cleavage sites. *Protein Science* **8**: 978–984.
- Findlay VJ, Tapiero H, Townsend DM. 2005. Sulfiredoxin: a potential therapeutic agent? *Biomedicine and Pharmacotherapy* **59**: 374–379.
- Guimaraes BG, Souchon H, Honore N, Saint-Joanis B, Brosch R, Shepard W, Cole ST, Alzari PM. 2005. Structure and mechanism of the alkyl hydroperoxidase AhpC, a key element of the *Mycobacterium tuberculosis* defense system against oxidative stress. *Journal of Biological Chemistry* **280**: 25735–25742.
- Guindon S, Gascuel O. 2003. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Systematic Biology* **52**: 696–704.
- Hall A, Karplus PA, Poole LB. 2009. Typical 2-Cys peroxiredoxins – structures, mechanisms and functions. *FEBS Journal* **276**: 2469–2477.
- Horton P, Park KJ, Obayashi T, Fujita N, Harada H, Adams-Collier CJ, Nakai K. 2007. WoLF PSORT: protein localization predictor. *Nucleic Acids Research* **35**(Web Server issue): W585–W587.
- Jacob C, Holme AL, Fry FH. 2004. The sulfonic acid switch in proteins. *Organic & Biomolecular Chemistry* **2**: 1953–1956.
- Jang HH, Lee KO, Chi YH, Jung BG, Park SK, Park JH, Lee JR, Lee SS, Moon JC, Yun JW *et al.* 2004. Two enzymes in one; two yeast peroxiredoxins display oxidative stress-dependent switching from a peroxidase to a molecular chaperone function. *Cell* **117**: 625–635.
- Jeanjean R, Latifi A, Matthijs HC, Havaux M. 2008. The Psae subunit of photosystem I prevents light-induced formation of reduced oxygen species in the cyanobacterium *Synechocystis* sp. PCC 6803. *Biochimica et Biophysica Acta* **1777**: 308–316.
- Jeong W, Park SJ, Chang TS, Lee DY, Rhee SG. 2006. Molecular mechanism of the reduction of cysteine sulfonic acid of peroxiredoxin to cysteine by mammalian sulfiredoxin. *Journal of Biological Chemistry* **281**: 14400–14407.
- Jobb G, von Haeseler A, Strimmer K. 2004. TREEFINDER: a powerful graphical analysis environment for molecular phylogenetics. *BMC Evolutionary Biology* **4**: 18.
- Jonsson TJ, Lowther WT. 2007. The peroxiredoxin repair proteins. *Subcellular Biochemistry* **44**: 115–141.
- Jonsson TJ, Murray MS, Johnson LC, Lowther WT. 2008a. Reduction of cysteine sulfonic acid in peroxiredoxin by sulfiredoxin proceeds directly through a sulfonic phosphoryl ester intermediate. *Journal of Biological Chemistry* **283**: 23846–23851.

- Jonsson TJ, Tsang AW, Lowther WT, Furdul CM. 2008b. Identification of intact protein thiosulfinate intermediate in the reduction of cysteine sulfinic acid in peroxiredoxin by human sulfiredoxin. *Journal of Biological Chemistry* **283**: 22890–22894.
- Kaneko T, Nakamura Y, Wolk CP, Kuritz T, Sasamoto S, Watanabe A, Iriguchi M, Ishikawa A, Kawashima K, Kimura T *et al.* 2001. Complete genomic sequence of the filamentous nitrogen-fixing cyanobacterium *Anabaena* sp. strain PCC 7120. *DNA Research* **8**: 205–213; 227–253.
- Konig J, Lotte K, Plessow R, Brockhinke A, Baier M, Dietz KJ. 2003. Reaction mechanism of plant 2-Cys peroxiredoxin. Role of the C terminus and the quaternary structure. *Journal of Biological Chemistry* **278**: 24409–24420.
- Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, Valentin F, Wallace IM, Wilm A, Lopez R *et al.* 2007. Clustal W and Clustal X version 2.0. *Bioinformatics* **23**: 2947–2948.
- Le SQ, Gascuel O. 2008. An improved general amino acid replacement matrix. *Molecular Biology and Evolution* **25**: 1307–1320.
- Liu XP, Liu XY, Zhang J, Xia ZL, Liu X, Qin HJ, Wang DW. 2006. Molecular and functional characterization of sulfiredoxin homologs from higher plants. *Cell Research* **16**: 287–296.
- Pascual MB, Mata-Cabana A, Florencio FJ, Lindahl M, Cejudo FJ. 2010. Overoxidation of 2-Cys peroxiredoxin in prokaryotes: cyanobacterial 2-Cys peroxiredoxins sensitive to oxidative stress. *Journal of Biological Chemistry* **285**: 34485–34492.
- Philippe H. 1993. MUST, a computer package of Management Utilities for Sequences and Trees. *Nucleic Acids Research* **21**: 5264–5272.
- Poole LB. 2007. The catalytic mechanism of peroxiredoxins. *Sub-cellular Biochemistry* **44**: 61–81.
- Razquin P, Fillat MF, Schmitz S, Stricker O, Böhme H, Gomez-Moreno C, Peleato ML. 1996. Expression of ferredoxin-NADP⁺ reductase in heterocysts from *Anabaena* sp. *Biochemical Journal* **15**: 157–160.
- Rey P, Becuwe N, Barrault MB, Rumeau D, Havaux M, Biteau B, Toledano MB. 2007. The *Arabidopsis thaliana* sulfiredoxin is a plastidic cysteine-sulfinic acid reductase involved in the photooxidative stress response. *Plant Journal* **49**: 505–514.
- Ronquist F, Huelsenbeck JP. 2003. MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* **19**: 1572–1574.
- Schein AI, Kissinger JC, Ungar LH. 2001. Chloroplast transit peptide prediction: a peek inside the black box. *Nucleic Acids Research* **29**: E82.
- Wood ZA, Poole LB, Karplus PA. 2003. Peroxiredoxin evolution and the regulation of hydrogen peroxide signaling. *Science* **300**: 650–653.
- Wu W, Jia Z, Liu P, Xie Z, Wei Q. 2005. A novel PCR strategy for high-efficiency, automated site-directed mutagenesis. *Nucleic Acids Research* **33**: e110.

Supporting Information

Additional supporting information may be found in the online version of this article.

Table S1 The taxonomic distribution of sulfiredoxin (Srx) and sensitive 2-Cys peroxiredoxins (Prx) homologues in all cyanobacterial complete genomes available in February 2010 and in a subset of complete genomes representative of main eukaryotic lineages

Please note: Wiley-Blackwell are not responsible for the content or functionality of any supporting information supplied by the authors. Any queries (other than missing material) should be directed to the *New Phytologist* Central Office.

RESUMÉ

Retracer l'histoire évolutive des eucaryotes est une question majeure en évolution et fait l'objet de nombreux débats. Le développement de techniques à haut débit, en particulier en protéomique et en génomique, a permis d'obtenir de nombreuses données pouvant être exploitées lors d'analyses évolutives. Dans ce contexte, les structures multiprotéiques eucaryotes (SME) constituent des objets d'intérêt. En effet, ces gros complexes protéiques sont impliqués dans de nombreux processus fondamentaux de la cellule eucaryote, et n'ont pas d'homologues chez les procaryotes (même si les fonctions dans lesquelles ils sont impliqués peuvent exister). Ils ont donc certainement joué un rôle prépondérant dans l'eucaryogénèse. L'analyse phylogénomique de deux SME impliquées dans la division cellulaire (le midbody et l'APC/C) montre que ces systèmes ont une origine évolutive ancienne et étaient déjà présents chez le dernier ancêtre commun des eucaryotes (LECA), tout en étant issus d'innovations eucaryotes. Ceci implique que l'émergence de ces deux SME s'est faite après la divergence de la lignée eucaryote et avant la diversification ayant donné naissance aux lignées actuelles. Au-delà de ces considérations évolutives, l'analyse de ces SME ouvre des pistes sur certains aspects de la biologie de ces systèmes. En effet, si ces systèmes ont été globalement bien conservés au cours de la diversification des eucaryotes, leur analyse révèle une grande plasticité de composition dans certaines lignées de protistes. Ceci suggère des changements récents concernant certaines étapes du cycle cellulaire de ces organismes qu'il serait intéressant d'explorer expérimentalement.

En parallèle, ce travail a montré que, bien qu'étant des protéines opérationnelles, les composants de ces SME portent un signal phylogénétique exploitable pour inférer les relations de parentés entre lignées eucaryotes. La construction de supermatrices à partir de ces protéines a permis l'inférence de phylogénies de qualité, même si non totalement résolues, dans lesquelles, par exemple, la monophylie des Excavata ou encore le placement des microsporidies au sein des Fungi est bien supporté. La combinaison de ces données avec celles issues d'analyses basées sur des protéines informationnelles montrent des avancées significatives concernant la résolution des arbres inférés. Ces résultats ouvrent le champ des possibles quant à la recherche d'autres marqueurs encore inexploités parmi les protéines opérationnelles. L'intégration de ces nouveaux marqueurs associée à l'augmentation de l'échantillonnage taxonomique représente une piste prometteuse pour l'avenir.

Ce travail illustre l'intérêt de généraliser les approches évolutives intégrées des systèmes biologiques pour l'étude de l'évolution et de la phylogénie des eucaryotes.

Tracing back the evolutionary history of eukaryotes is one of the major issues in the field of evolution and is hotly debated. The development of high throughput techniques, especially in proteomics and genomics has yielded extensive data that can be used in evolutionary analyses. In this context, eukaryotic multiprotein structures (EMS) are objects of interest. Indeed, these large protein complexes are involved in many fundamental processes of eukaryotic cells, and have no homologues in prokaryotes (even if the functions in which they are involved may exist) and therefore have certainly played a major role in the eukaryogenesis. The phylogenomic analysis of two EMS involved in cell division (the midbody and the APC/C) shows that these systems have an ancient evolutionary origin and were already present in the last common ancestor of eukaryotes (LECA), while resulting from eukaryotic innovations. This implies that the emergence of these two EMS occurred after the divergence of the eukaryotic lineage and before the diversification that gave rise to the current lineages. Beyond these evolutionary questions, analyses of these EMS uncover some biological aspects of these systems. Indeed, if these systems were generally well conserved during the diversification of eukaryotes, their analysis shows a high plasticity of composition in some protist lineages. This suggests that recent changes regarding certain phases of these organisms cell cycle which would be interesting to explore experimentally.

Concomitantly, this work showed that, although being operational protein, components of these EMS carry a phylogenetic signal usable for inferring phylogenetic relationships among eukaryotic lineages. Construction of supermatrixes from these proteins led to the inference of phylogenies of high quality, even if not fully resolved, in which, for example, the monophyly of Excavata or the placement of Microsporidia within Fungi is well supported. Combining these data with those from analyses based on informational proteins show significant progress on the resolution of inferred trees. These results open the field of possibilities to find other markers among the untapped proteins operational. The integration of these new markers associated with increased taxonomic sampling represents a promising approach.

This work illustrates the interest of generalizing integrated evolutionary approaches of biological systems for studying the evolution and phylogeny of eukaryotes.