Phylogeny and evolutionary perspective of Opisthokonta protists

Guifré Torruella i Cortés

Programa de Doctorat de Genètica
Departament de Genètica
Facultat de Biologia
Universitat de Barcelona

Phylogeny and evolutionary perspective of Opisthokonta protists

Filogènia i perspectiva evolutiva dels protists opistoconts

Memòria presentada per Guifré Torruella i Cortés per tal d'optar al títol de Doctor per la Universitat de Barcelona

Dr. Iñaki Ruiz-Trillo	Dra. Marta Riutort León	Guifré Torruella i Cortés
Director de la tesi	Tutora de la tesi	Autor

Barcelona, setembre de 2014

Cover image: Opisthokonta tree of life

SYLLABUS

Figures and tables

1. INTRODUCTION

My excrement being so thin, I was at divers times persuaded to examine it; and each time I kept in mind what food I had eaten, and what drink I had drunk, and what I found afterwards. I have sometimes seen animalcules a-moving prettily…

A van Leeuwenhoek 1681.

1.1 A historical perspective to protist taxonomy

Animals and fungi, along with plants, are two groups of multicellular eukaryotes extensively studied since the origin of our species. This is not surprising since we depend on them to survive; they provide us with food, clothes, transport, rituals, medicines, etc. It is not difficult, then, to assume that our dependency on these living things lead us to an intrinsic fascination to understand such organisms and their origins. Although I do not pretend to exhaustively review history, the following paragraphs provide the context, fundaments and motivation for the work here presented.

Intrinsic to the human condition is also the need to classify any kind of objects with the aim to structure and communicate knowledge. Natural organisms have always been classified into artificial groups based on observable similarities, according to each time's cosmogony (Marks 2008). Early human populations needed to discern between edible or poisonous plants, preys or predators. Later on, ancient writings from Chinese or Egyptians already provided practical information for medicinal plants (Manktelow 2010), while Greeks and Romans started more systematic, although somehow arbitrary, classifications. For example, Aristotle's *Scala naturae*, a.k.a. the Great Chain of Being (a hierarchical classification of organisms from inferior/simple to superior/complex) was religiously established in Europe during medieval ages (**Figure 1 left**) until the 18th century (although some reminiscent thinking is still present nowadays). It was not until Linné's *Systema Naturae* that biological classification (taxonomy or systematics) was formally proposed. During the 19th century, there were great advances in comparative anatomy and palaeontology (i.e., the works of Cuvier and Owen). These lead to revolutionary changes of paradigm from a fixist taxonomy to the transmutation of species (Lamarck), to the evolution by natural selection as the mechanism for species origin from common ancestry (Darwin, Wallace). The homology concept, used before for structural or functional similarity between taxa, changed to include the sense of common ancestry. And so organisms were not classified anymore into distinct arbitrary criteria, but only with the purpose to reconstruct their evolutionary history or phylogeny. As new species were proposed to radiate over time from common ancestors, networks and tree-like diagrams were popularized (Ragan 2009) to represent such biological changes between extant living species (**Figure 1 right**).

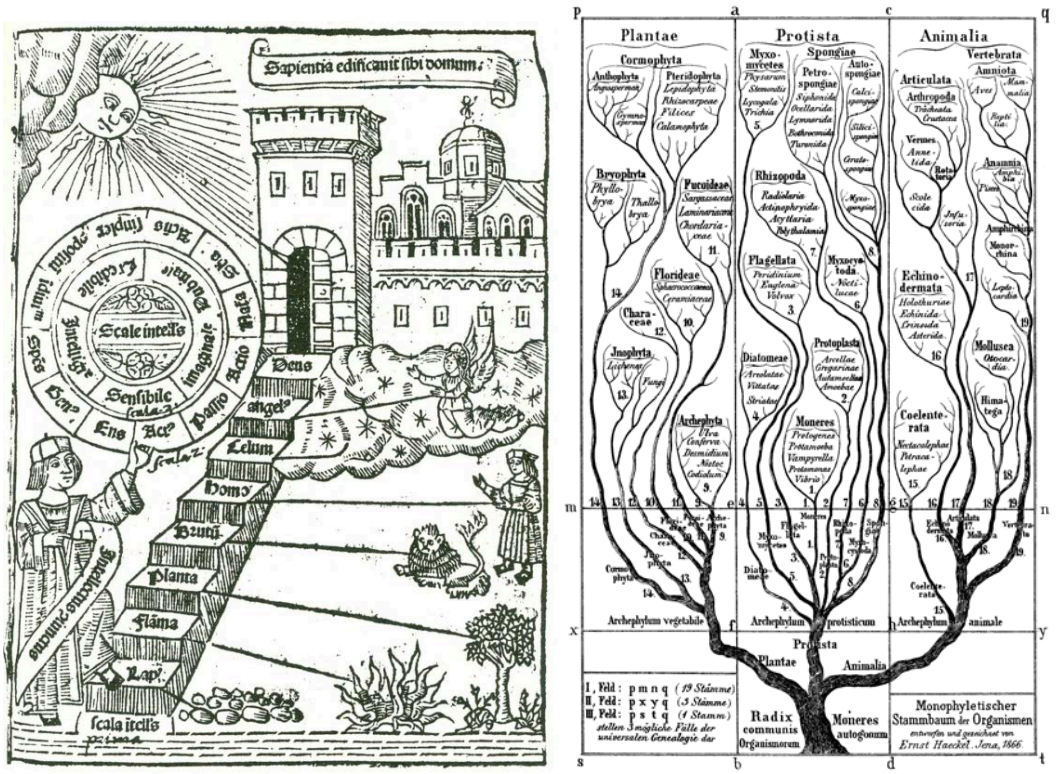


Figure 1. Left: hierarchical classification of nature in *Liber de ascensu et decensu intellectus* of Ramon Llull (written 1304, first published 1512), from (Ragan 2009). Right: *Monophyletischer Stambaum der Organismen* from *Generelle Morphologie der Organismen* of Ernst Haeckel (1866).

Thanks to the invention of microscopes (Leeuwenhoek) and the discovery of the cell (Hooke) in the 17th century, protistology was already identifying several lineages of unicellular eukaryotes. In parallel to the development of the evolutionary thinking, during the 18th and 19th centuries many naturalists used optic microscopes to describe novel eukaryotes. For example, during that period, Henry James Clark, William Saville Kent and LS Cienkowski, among many others, described distinct protist species, including some relevant to this work, such as choanoflagellates, nucleariids or apusomonads (**sections 1.2.1**, **1.2.5** and **1.3** respectively). Descriptions were limited to cell morphology, feeding modes and lifestyle. And first correlations between unicellular and multicellular organisms were established, such between choanoflagellates and sponge choanocytes. Haeckel (**Figure 1 right**) included protistan organisms as a distinct branch from animals and plants, but connected all three at the root, indicating a single protistan kingdom, an idea that persisted for long.

In the 20th century several major discoveries on superficially distinct research fields (such as palaeontology, comparative anatomy, electronic microscopy, genetics, molecular biology, and statistics) lead to the current modern synthesis of evolution (Mayr 1982), and are the main fundamental topics for this thesis. An important implication of the new synthesis is that molecular evolution provides huge numbers of homologous characters that change over time (see Zuckerkandl and Pauling's work on molecular clocks or

Kimura's on neutral evolution). These characters can be used in computational phylogeny to infer evolutionary relationships between species (see for example Willi Hennig's work on phylogeny as hypothetical tests or Fitch's on orthology and paralogy concepts). Moreover, the development of transmission electron microscopy (TEM) provided a new powerful technique in protistology (Patterson 1999) that allowed the study of inner cell morphology, including nuclei, membranes, organelles, vesicles or cilia. These provided additional characters to classify, and thus, different protists groups were established (see (Parfrey et al. 2006) for a historical review on protist classifications). However, morphology-based phylogeny has serious limitations – especially in protistology – from the amount of comparable characters, to the validity of them as phylogenetic markers. The major problem is that morphological character states are not always homologs, but analogous due to reversion or convergent evolution (see **section 1.4** for further details).

With the invention/popularization of molecular and sequencing techniques it came a more systematic approach to phylogeny, using single marker trees and specially the small ribosomal subunit (SSU rDNA) (Woese and Fox 1977). Molecular phylogenies provide significant advantages in comparison to morphology-based phylogenies. More characters (nucleotidic or amino acidic positions) can be easily obtained; and objectively compared in terms of orthology, between properly aligned sequences. Molecular phylogenies reshaped protistan classification, as they appeared paraphyletic and distinctly related with multicellular lineages (Cavalier-Smith 1993; Baldauf et al. 2000) As a result of these and other studies, several eukaryotic supergroups were established: Opisthokonta, Amoebozoa, Excavata, Archaeplastida, Rhizaria and Chromalveolata (Simpson and Roger 2004). But SSU rDNA has no resolution to resolve relationships between supergroups, so such deep speciation events started to be elucidated with the development of high-throughput sequencing and phylogenomics in the past 10 years (Bapteste et al. 2002; Ruiz-Trillo et al. 2004; Burki et al. 2007; Rodríguez-Ezpeleta, Brinkmann, Burger, et al. 2007) (see **section 1.4** for methodological details).

1.2 Who are the opisthokonts?

Opisthokonta is a supergroup of eukaryotes first proposed by Thomas Cavalier-Smith in a symposium of the British Mycological Society in 1986 (published in the book Evolutionary Biology of the Fungi Ed. Cambridge University Press, 1987). There, he proposed that Fungi and Metazoa shared a more recent common ancestor than each one to plants. Choanoflagellate protists were also included within the Opisthokonta clade (**Figure 2**), whose major morphological synapomorphy was the single emerging flagellum that is located at the posterior end of the cell. This classification proposal was soon confirmed by SSU rDNA phylogenetic analyses, and later on corroborated by protein-coding gene trees. Also, some molecular synapomorphies were proposed (Huang et al. 2005; Elias 2008; Shadwick and Ruiz-Trillo 2012), being a 12 amino acid insertion in the translation elongation factor 1α (EF-1α) gene (Baldauf and Palmer 1993; Steenkamp et al. 2006) the best documented so far (see also **sections 1.4.1** and **3.1 R1** for more on molecular synapomorphies).

The list of Opisthokonta protists have been progressively increased during the past couple of decades (reviewed in Paps and Ruiz-Trillo 2010 and expanded in **section 1.2**). One of the new opisthokonta lineage was the Ichthyosporea (Cavalier-Smith 1998) (**section 1.2.3**) – a.k.a. Mesomycetozoa (Herr et al. 1999) or DRIPs (Ragan et al. 1996) – , which were confidently placed as sister group to Choanoflagellata (**section 1.2.1**) and Metazoa (coining the Holozoa clade) using mitochondrial genes (Lang et al. 2002). These fungi-like species had been traditionally included within Fungi along with some current stramenopile groups (oomycetes, hyphochytrids and thraustochytrids), due to morphological and lifestyle similarities (i.e., cell wall and osmotrophy). Another example is the enigmatic free living fungi-like *Corallochytrium limacisporum*, a species originally classified as a thraustochytrid (Raghu-Kumar 1987), but molecular phylogenies placed it within the Opisthokonta (Cavalier-Smith and Paula Allsopp 1996). There are also filose amoebae called *Nuclearia* (**section 1.2.5**) (Patterson 1984), which were positioned as sister group to Fungi (Medina et al. 2003). Interestingly, one of them was at some point placed outside the group but still within Opisthokonta and renamed as *Capsaspora owczarzaki* (Hertel et al. 2002). This amoeba resulted closer to animals using ribosomal and actin gene trees (Ruiz-Trillo et al. 2004) or multigene analyses (Ruiz-Trillo et al. 2008). Later positioned with *Ministeria vibrans* forming a clade named Filasterea (**section 1.2.2**) (Shalchian-Tabrizi et al. 2008). Recently, *Fonticula alba* was positioned as sister to *Nuclearia* at the root of the Holomycota=Nucletmyceta clade (Fungi and its related protists) (Liu 2009, Brown et al. 2009). Last year, the Aphelida (Karpov et al. 2013) and the *Rozella* (=Cryptomycota) (James and Berbee 2012) groups were proposed to cluster with Microsporidia in a clade called Opisthosporidia (**section 1.2.6**) (Karpov, Mamkaeva,

Aleoshin, et al. 2014) sister to Chytridiomycota (**section 1.2.7**) and Fungi (**Figure 2**). What follows is a more detailed description of the different groups.

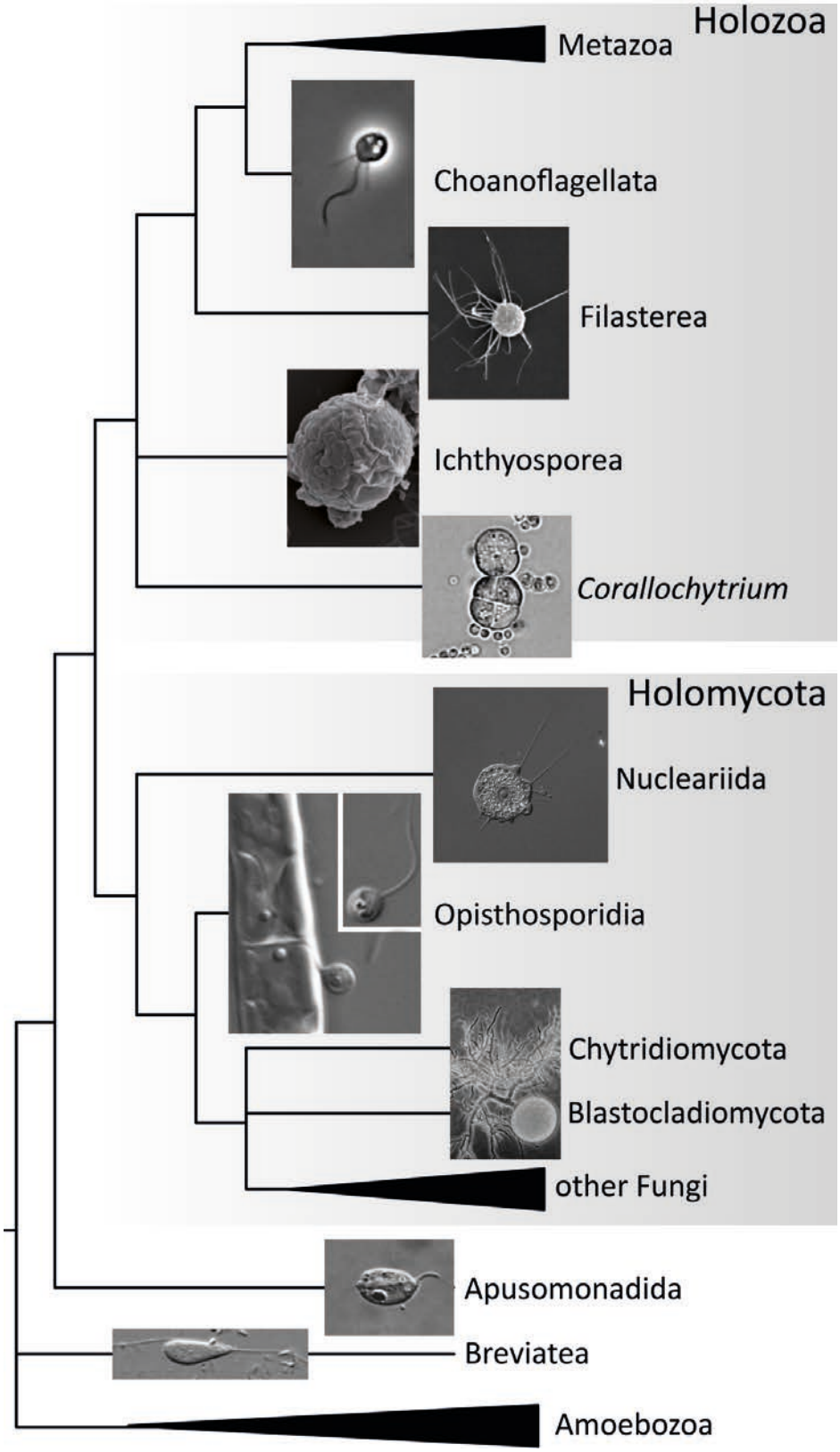


Figure 2. Overview of the Opisthokonta clade. Notice the clear division between Holozoa (Lang et al. 2002) and Holomycota (Liu et al. 2009), as well as different uncertainties as politomies.

1.2.1 Choanomonada

Choanomonada (a.k.a. Choanoflagellata, Choanoflagellida or simply choanoflagellates) is a monophyletic group of heterotrophic nanoflagellates (3-10 μm in cell size). They were already described in the mid- 19th century by James Clark and Saville Kent, back then classified with other unrelated protists as Infusoria. The choanoflagellates are free-living organisms ubiquitously and abundantly distributed across the whole water column, even in abyssal plains (Nitsche et al. 2007). They are mainly marine, but also found in fresh-water and even anoxic brackish habitats (Wylezich et al. 2012). There are approximately 250 species described so far, (Carr et al. 2008), and environmental data show additional groups (del Campo and Massana 2011; del Campo and Ruiz-Trillo 2013). Traditionally, Choanoflagellates have been classified using morphological characters (Leadbeater et al. 2009), but recent molecular phylogenetic studies have established clear internal groups (Carr et al. 2008; Nitsche et al. 2011) (**Figure 3**). Cell morphology is highly conserved among the different choanoflagellates and consists of a spherical to ovoid cell with a single apical flagellum surrounded by a funnel-shaped actin-supported microvilli collar (Karpov and Leadbeater 1998) (**Figure 4**). Choanoflagellates are suspension filter feeders (specialized bacterivores) that use the flagellar beating to create water currents to bring prey to the outer surface of the collar (Pettitt et al. 2002), and then these particles are phagocytized using both microvilli and pseudopodia (Dayel and King 2014). To achieve feeding efficiency, locomotory force must be resisted, and so most choanoflagellates attach to substrates either directly or by a peduncular stalk during most of the life cycle, being benthonic and sedentary; but see Stephanoecidae. Other characteristics include a conserved intracellular organization (i.e., organelle number and disposition) (Karpov and Leadbeater 1997). Less conserved are the periplastic structures they can posses (such as the thecae or loricae), although all have glycocalyx surrounding the cell membrane. Choanoflagellates are currently classified into two monophyletic groups, the Acanthoecida and the Craspedida (Nitsche et al. 2011).

Acanthoecida is a monophyletic group that comprises choanoflagellates with a particular siliceous lorica surrounding the cell (**Figure 4 D**), whose variety of shapes and formation patterns have been widely used to classify them (Leadbeater 2008). The Acanthoecida are divided into the Stephanoecidae (tectiformes) and the Acanthoecidae (nudiforms). The Stephanoecidae comprise half of the described choanoflagellates, with marine and brackish species – only one fresh-water species described so far (Paul 2012) – which have the most complex of all periplast structures (Leadbeater et al. 2009). The lorica for the offspring is prebuild before cell division, then the daughter cell occupies and completes the lorica immediately, without the typical swimmer dispersion observed in other groups. The lorica provides rigidity and a volume-to-weight ratio (up to 80 μm) that,

together with a physically neutral flagellar beating propulsive force, allows the cells to live in pelagic environments, becoming resistant to sink (Leadbeater 2008). Instead, the Acanthoecidae is represented by less than ten marine or brackish sedentary species, usually associated with biofilms, surrounded by a lorica with helical costae during the vegetative stage. After the cell division, the naked daughter cell swims for dispersion and settles again to recreate the lorica from scratch (hence the name nudiform). This lorica is simpler than the one from tectiformes and thus cannot be much bigger than the cell size.

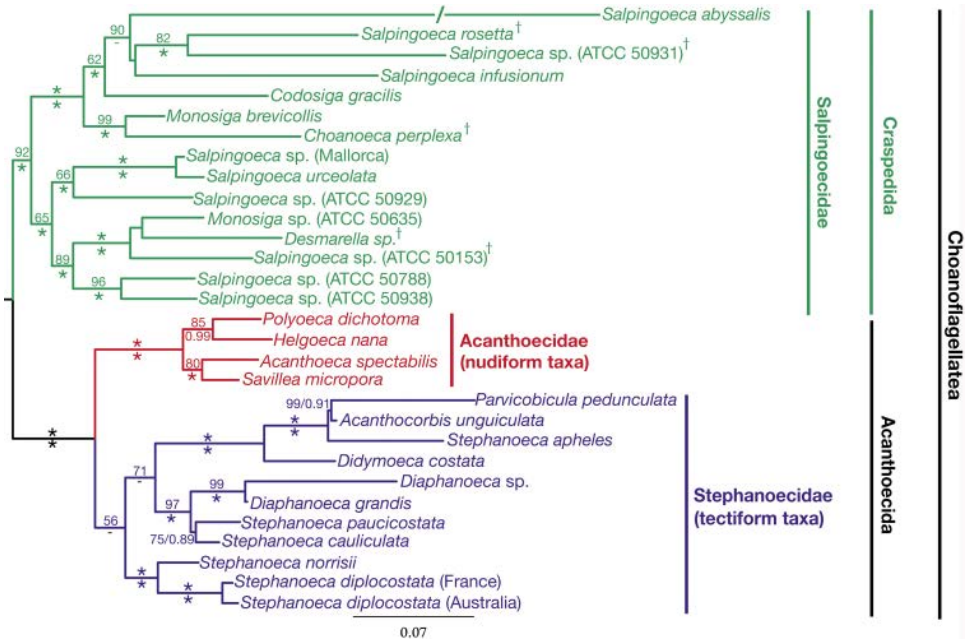


Figure 3. The internal phylogeny of the Choanomonada (Nitsche et al. 2011).

Craspedida was traditionally divided into two groups according to its morphology: codosigids or salpingoecids. However, molecular data did not corroborate this division (Carr et al. 2008). Both codosigids and salpingoecids have a vegetative sedentary stage anchored to the substrate, directly or through a periplastic stalk. Codosigids (such as *Monosiga brevicollis*) have no other extracellular structure, and thus when the daughter cell appears by lateral cell division, it directly swims (using the flagellum) until it settles on the substrate (see **Figure 4 A to C**). On the other hand, salpingoecids (e.g., *Salpingoeca rosetta*) have a flask-like organic theca that daughter cells must leave for dispersion. Such described simple life cycles are probably the consequence of insufficient study or due to observations in culture conditions, as other cell stages have been recently described: e.g., dormant cysts (Stoupin et al. 2012) and anisogamous gametes (Levin and King 2013; Umen and Heitman 2013). The colonial stage is being thoroughly studied in cultures of *S. rosetta* (Fairclough et al. 2010; Alegado et al. 2012). Clusters of cells are surrounded by an extracellular matrix (ECM) and bound by complex intercellular bridges as a consequence of incomplete cytokinesis (Dayel et al. 2011). Apparently, this cluster of cells increase feeding efficiency (more hydrodynamic water flow) and let the organism feed in

the water column (Roper et al. 2013); a probable source of selective pressure to fix this trait, which probably was already present in the last common ancestor of both choanoflagellates and animals. Choanoflagellates are very similar both in morphology and feeding mode to the choanocytes, a specific cell type of sponges. However, their homology is still contentious (Nielsen 2008; Mah and Leys 2014). The genome of two Craspedida choanoflagellates have been fully sequenced, specifically those of the single-cell *M. brevicollis* (King et al. 2008) and the colonial *S. rosetta* (Nichols et al. 2012).

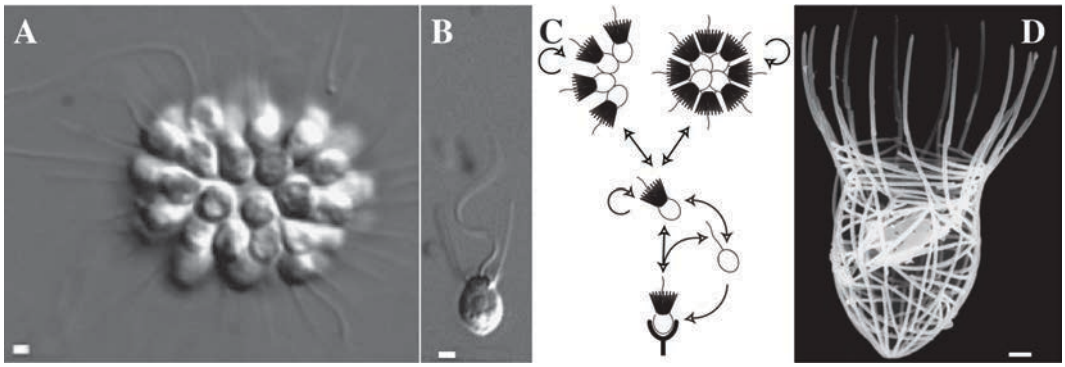


Figure 4. Choanomonada morphology and cell cycle. A) Craspedid *Salpingoeca rosetta* colony. B) *S. rosetta* swimmer. A and B are DIC light microscopy pictures, modified from (Dayel et al. 2011). C) *S. rosetta* life cycle (Fairclough et al. 2013). D) Scanning electron microscopy image of the Acanthoecid *Acanthocorbis unguiculata* lorica (Leadbeater et al. 2008). All scale bars = 1 *μ*m.

1.2.2 Filasterea

Filasterea (Shalchian-Tabrizi et al. 2008) is a group with only two known genera: *Capsaspora* (Hertel et al. 2002) and *Ministeria* (Patterson et al. 1993). Both are small (3-5 *μ*m) naked filose amoebae, with a pretty similar morphology (rounded with filopodia) but contentious conservation of lifestyle (see below). Surprisingly, environmental studies have failed so far to describe a wider diversity for Filasterea (del Campo and Ruiz-Trillo 2013; personal communication del Campo & Ruiz-Trillo), and only 2 or 3 cultivable strains are available.

Ministeria amoebae have been reported in coastal marine water samples for which two species have been described. *Ministeria vibrans* (**Figure 5 A**), described as being suspended by a flagellum-like stalk attached to the substrate (hence the name), has two cultivable strains (Cavalier-Smith and Chao 2003), although on shows no stalk. The other species is *M. marisola* and was described without a stalk either, but no culture was ever available.

Capsaspora owczarzaki was described 30 years ago as an amoeba-like "symbiont" from the fresh-water mollusc *Biomphalaria glabrata* (Stibbs and Owczarzak 1979; Owczarzak et al. 1980). The amoebae were obtained from the pericardial explants and mantle swabs of snails originally sampled in Puerto Rico. As mentioned, the original morphological description situated *Capsaspora* among the *Nucleariа* (**section 1.2.5**) (Owczarzak et al. 1980), but molecular phylogenies placed it somewhere closer to animals (Zettler et al. 2001; Hertel et al. 2002; Medina et al. 2003) and finally it was clearly shown to be within the Holozoa (Ruiz-Trillo et al. 2004). The *Capsaspora* life cycle have been recently studied in fine detail (Sebé-Pedrós et al. 2013) (**Figure 5 D**). They found that under initial culture conditions, the filose amoeba crawls attached at the substrate, with active replication until the end of the exponential growth phase. Then cells start to detach retracting the branching filopodia and encyst. During this phase, cells suffer no division and have a genetic expression profile of typical dormant cells. Alternatively, amoebae can actively aggregate to each other by unknown factors, even secreting an unstructured extracellular material that seems to prevent direct cell-cell contact.

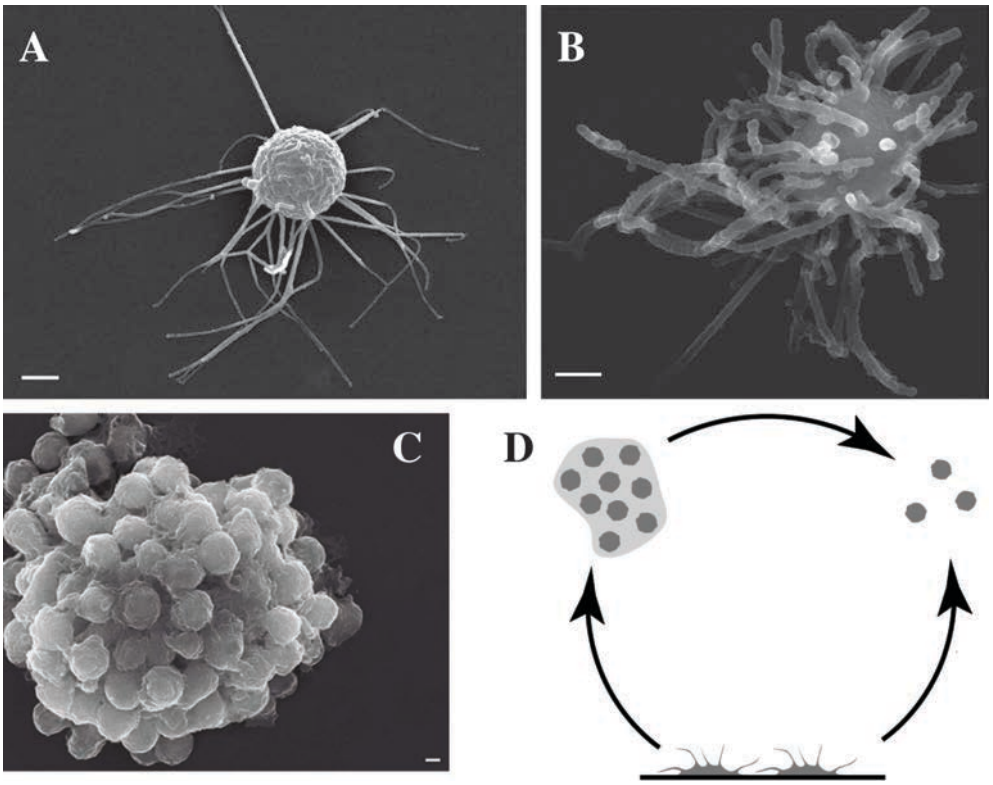


Figure 5. Morphology of Filasterea. A) *Ministeria vibrans*. B) *Capsaspora owczarzaki* filose amoeba. C) *C. owczarzaki* aggregative colony. D) *C. owczarzaki* life cycle. All SEM images have scale bars = 1 μm; courtesy of Dr. Sebé-Pedrós.

1.2.3 Ichthyosporea

Ichthyosporea (a.k.a. Mesomycetozoea or DRIPs) is a group of ecologically and morphologically diverse osmotrophic/saprotrophic fungi-like organisms (see reviews by Mendoza et al. 2002; Glockling et al. 2013). This feeding behaviour differentiates them for other mentioned holozoans, which are all phagotrophs. Ichthyosporea also have a morphology and a life cycle similar to non-dikaryotic fungi (see **section 1.2.7**); they produce large multinucleate spherical cells (from a dozen to a hundreds of microns), surrounded by thick cell walls (some chitin-like positive), traditionally called coenocytic endospores; or in the protistology nomenclature: schizonts (Kocan 2013) (**Figure 7 O**). These structures are proposed to produce the offspring by cellularization, which are small limax-shaped amoebae (i.e., no apparent pseudopodia), flagellates or simply other spherical cells (less than 10 μm), which can be called merozites, the products of schizogony. Some other observed cellular forms are hyphae-like (non-septate hyaline tubules) and plasmodia (**Figure 7 H**).

Ichthyosporea were first defined as a cluster of four parasites of aquatic animals (*Dermocystidium*, *Ichthyophonus hoferi*, *Psorospermium haeckelii* and the "rosette agent" a.k.a. *Sphaerothecum destruens*) based on SSU rDNA phylogenetic trees (Ragan et al. 1996). These protists were known previously, even *Psorospermium* was known by Haeckel – hence the name– , but they were classified among other protists or fungi based solely on morphological and life cycle features. Ichthyosporea as a taxonomic name was formally established due to the fact that (back then) the group was only composed of fish parasites (Cavalier-Smith 1998). When the agent of rhinosporidiosis in mammals and birds, *Rhinosporidium seeberi* was included (R. Herr et al. 1999), the name was amended to Mesomycetozoea (Mendoza et al. 2002) (within Mesomycetozoa: all three non-metazoan holozoa protist groups; a synonym of Choanozoa (Cavalier-Smith 1993)). Currently, Ichthyosporea is divided into two clear monophyletic groups of over 40 taxa (Glockling et al. 2013). The two groups differ in morphology and probably in diversity, but their main distinctive characteristic is the dispersal cell stage: a flagellum in the Dermocystida, a naked crawling amoeba in the Ichthyophonida.

So far, all species described come from inside or outside the body of fresh water (like fishes and amphibians), marine (such fishes and arthropods), or terrestrial animals (mammals, birds, insects (Lord et al. 2012)). Most of them attach to gut surfaces or exoskeletons. In most cases, especially among the Ichthyophonida, it is unknown whether they are strict parasites, commensals or have been found by chance inside the animals. Recent environmental studies from water and sediment samples have shown cryptic diversity in Ichthyophonida (similarly to Choanoflagellates), including two unknown clusters of fresh-water and marine sequences. This suggests that some of these

sequences may come from free-living organisms, or at least from organisms that have an external life stage (Marshall and Berbee 2011; del Campo and Ruiz-Trillo 2013).

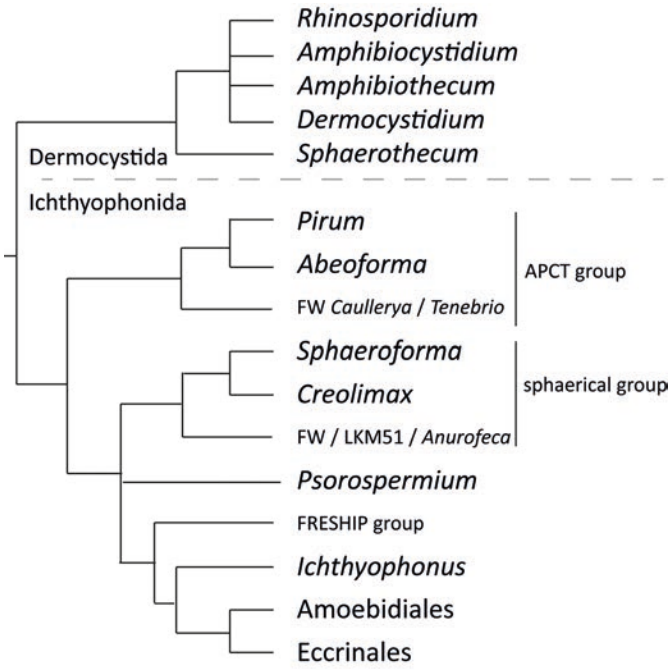


Figure 6. Internal phylogeny of Ichthyosporea, the current consensus (del Campo and Ruiz-Trillo 2013; Glockling et al. 2013).

Dermocystida (a.k.a. Rhinosporideacae) is composed by few world wide distributed strict parasite species of vertebrates, most of them with a described flagellated stage (**Figure 7 C**). The difficulties of isolation and the lack of monoxenic cultures have prevented so far a good knowledge on their full life cycles. Moreover, their phylogenetic relationships and taxonomy are not yet well established because many morphological observations are not complemented with molecular studies (Feist et al. 2004; Raffel et al. 2008).

In agreement with most SSU rDNA trees, the first branching lineage would be *Sphaerothecum destruens*, a.k.a. rosette agent (**Figures 7 A to D**) (Arkush et al. 2003), which potentially represent the ancestral Dermocystida bauplan (see **section 4.2.1** and **Figure 18**). *S. destruens* is a generalist fish parasite that causes high mortality due to chronic intracellular infections (Andreou et al. 2012). Multiple isolations and *in vitro* cultures in salmonid cells have been achieved and now studies of life cycle and pathogenicity can be performed (Paley et al. 2012). Although the infection process is not known, the flagellated stage of this species is achieved in contact with fresh-water, which has been hypothesised to happen in nature when host fishes return from the sea to spawn (**Figure 7 D**). This would represent the transmission stage (Arkush et al. 2003). TEM ultrastructure has shown small 2 μm cells with dense intracellular content (no central vacuole), a nucleus with homogeneous content (no apparent nucleoli) and a three-layer extracellular wall. No amoeba stage have been ever reported (Arkush et al. 2003). The other Dermocystida species fairly studied is *Rhinosporidium seeberi* (**Figures F and G**)

that can parasite humans. This species vegetative stage is a massive schizont that produce thousands of unflagellated offspring (Pereira et al. 2005; Vilela and Mendoza 2012). Between *S. destruens* and *Rhinosporidium*, it seems to exist a not-so clear range of species/morphotypes, such *Dermocystidium percae* (Pekkarinen et al. 2003) or *Amphibiocystidium* (Pascolini et al. 2003), which goes from smaller and flagellated to bigger and unflagellated (**Figure 7 E**).

Ichthyophonida (a.k.a. Ichthyophonae) is the division of Ichthyosporea with more described species and phylotypes. The variability of this group in terms of morphology, cell cycle and host organism is huge. Moreover, the internal phylogeny is not well resolved, although a few groups are highly supported (Glockling et al. 2013), One group (here called **APCT**) contains four recently described species and 16 environmental sequences, eight of them forming a marine cryptic lineage (MAIP1; (Lohr et al. 2010)). *Abeorforma whisleri* and *Pirum gemmata* (Marshall and Berbee 2011) come from the gut of diverse marine lophotrocozoan animals, and are easily culturable in axenic conditions (see **section 3.3 R3**). Both have typical spherical (**Figure 7 H**), vacuolated and multinucleated schizonts with thick cell walls (only *A. whisleri* has extracellular calyx) that will maturate (cellularization) and release the offspring. Non-motile amoebae have been detected in *P. gemmata*, but usually the offspring is already a small schizont. In the case of *A. whisleri* three different kinds of "amoebae" that can asexually produce more amoebae have been reported (Marshall and Berbee 2011), presenting lobose pseudopodia, multinucleate hyphae-like and even plasmodial structures (**Figure 7 I**). Although few is known from the other species described of this clade (which are *Tenebrio molitor* symbiont or TMS (Lord et al. 2012) and *Caullerya mesnii* (Lohr et al. 2010)), both found inside arthropod tissues and both with quite different morphologies than *P. gemmata* and *A. whisleri* (schizonts are not sphaerical and they do not have amoeboid stages) (**Figure 7**).

The crayfish parasite ***Psorospermium haeckelii*** (Vogt and Rug 1995) constitute a clearly distinct clade in Ichthyophonida, its SSU rDNA branches suggest that this morphotype (Bangyeekhun 2001) might be derived. They have been described as big (50 to 200 μm) elongated ovoid schizont forms (**Figure 7 J**), with a dense extracellular multi-layered organic shell with many spherical polynucleated cells within (mature stage) that later release motile amoebae (Henttonen et al. 1997).

The **Ichthyophonida/Eccrinales/Amoebidiales** clade (**Figure 6**) is composed by species with more complex and diverse cell forms than any other Ichthyophonida, including polarized tubular multinucleate cells resembling to the coenocytic thalli of chytrid or kickxellomycetes (formerly zygomycetes) fungi (**Figures 7 K to O**). Indeed, the

Eccrinales and Amoebidiales, have traditionally been classified within the Trichomycetes fungi (Lichtwardt et al. 2001). The branching order within this group of ichthyophonids is not clear, but they seem to be sister group to a cryptic group of fresh-water environmental sequences (clade FRESHIP 1) (del Campo and Ruiz-Trillo 2013).

Amoebidium (6 described species) and *Paramoebidium* (13 species) appear as sister group in SSU rDNA phylogenies. Both are found as ecto- or endocommensals of fresh-water arthropods. Their most apparent form is as coenocytic unbranched polarized thalli with holdfast at one end: structures that attach to all kinds of substrates, but naturally to arthropod guts (**Figure 7 K and L**). When these multinucleate tubular cells mature they can produce either smaller multinucleate cigar-shaped cells or amoeboid cells. These limax-like amoebae crawl away and encyst, and then they progressively grow in size and number of nuclei forming a new schizont (although timing, size and morphology of schizonts differ between distinct species). The complete cycle is known thanks to Whisler's exhaustive work on *Amoebidium parasiticum*, who studied its ultrastructure during the 60s.

The Eccrinales, with more than 50 species described (Lichtwardt et al. 2001), are found inside arthropod guts all around the world, in fresh-water, marine and terrestrial habitats (Cafaro 2005). Eccrinales show a diverse and complex polarized maturation with no amoeboid stage described, but septate tips that will release the offspring (**Figure 7 M**). Their complexity and specificity for arthropod gut cavities, along with SSU rDNA phylogenies, lead Cafaro to propose that Eccrinales are specialized Amoebidiales.

Finally, within this large group, there are the *Ichthyophonus*, strict parasites that cause disease in fishes distributed worldwide. The morphotype called *I. hoferi* is the only one carefully studied (Spanggaard 1995) and with available monoxenic lab cultures. Only the asexual growth have been observed, which develops as a spherical multinucleate walled schizont that at low pH transforms into a polarized plasmodial (coenocyte hyphae-like) shape similar to Eccrinales or Amoebidiales but branched (**Figures 7 N and O**). Interestingly, studies on two distinct *Ichthyophonus sp*. strains differ from previous reports, such as the fact that no hyphae-like structures but big multinucleate amoeboids were seen (Kocan et al. 2013). This suggests that additional studies are needed in order to fully understand their whole life cycle, which seems to be more complex than in any other ichthyophonid group.

The remaining Ichthyophonida group has been recently proposed as the **"spherical" group** (Glockling et al. 2013) that includes two monophyletic clades. One is formed by fresh-water environmental samples, including the planktonic sequence LKM51 (Hannen et al. 1999), and *Anurofeca richardsi* (Baker et al. 1999) which has only been found in anural (amphibian) larvae. The other group is exclusively marine and has been well studied by Marshall and Berbee in a series of very informational papers. This group contains

Creolimax fragrantissima (Marshall et al. 2008), isolated from the gut of a peanut worm (**Figures Q and R**). Its life cycle has been well described in culture conditions (**Figure 7 S**) (Suga and Ruiz-Trillo 2013): it starts as a small walled spherical cell that grows as a multinucleate schizont, with peripherical nuclei and a big central vacuole until cellularization occurs. Then the offspring cells fill the whole sphere and are released as amoebae after breaking or decomposing the cell wall. The lobose limax-shaped amoebae (~10 x ~5 μm) disperse by crawling; they settle, become round and the cycle closes. The genus *Sphaeroforma* (Jøstensen et al. 2002) contains four genetically distinct species (Marshall and Berbee 2013). All show a simple life cycle in which a small walled spherical cell grows until maturation (cellularization) and releases a new offspring of encapsulated cells to the environment (**Figure 7 P**). The morphological variations are common, and both plasmodial (as in *Abeoforma* (Marshall and Berbee 2011) and *Ichthyophonus*) and amoeboid stages have been reported in *Sphaeroforma tapetis* (formerly known as *Pseudoperkinsus* (Figueras et al. 2000)). A population genetic study in *S. tapetis* (Marshall and Berbee 2010) showed prominent haploid cell populations with asexual life cycle, although presence of recombination was consistent with facultative sexuality. Moreover, host specificity was excluded once the same haplotypes were found in different invertebrate species (Marshall and Berbee 2013).

To summarize, Ichthyosporea are extremely plastic organisms. There are more morphotypes within a species (such as in *A. whisleri* or *S. tapetis*) than between some of them. This means the border between species is difficult to establish and only population genetic studies will properly distinguish the biodiversity this group. Non host-specificity of certain lineages, such in *Sphaerothecum destruens* or *Sphaeroforma*, and cryptic diversity shown by environmental studies, suggest that not all lineages might be strict parasites; some may be casual commensals from nutrient rich environments. This correlates between phylogenetically early branching lineages and the more derived ones, such as *Rhinosporidium*, Amoebidiales and Eccrinales that, in contrast, seem to be strictly host-dependent.

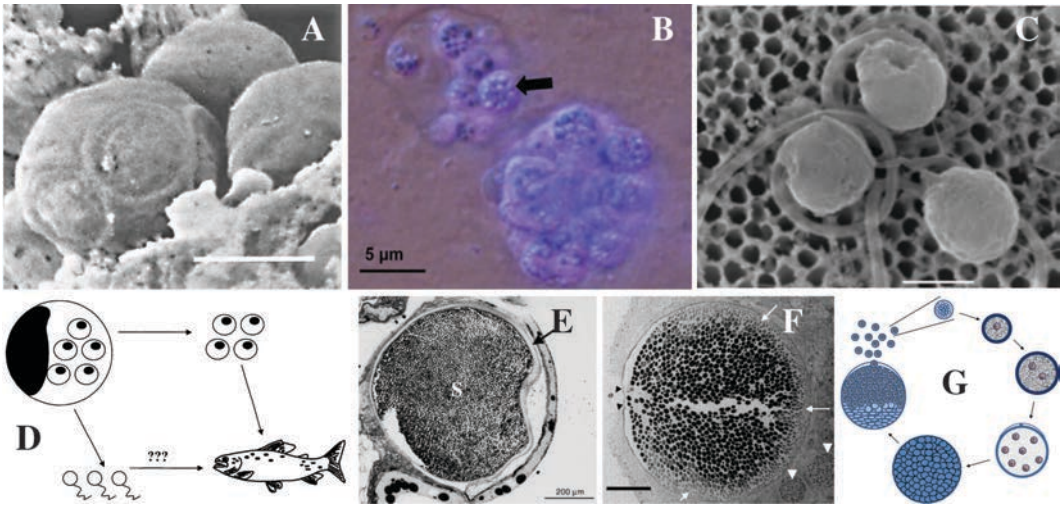


Figure 7. Ichthyosporea biodiversity. Panels from A to G represent distinct Dermocystida species. A) SEM *Sphaerothecum destruens* (bar = 2 µm). B) *S. destruens* rosette schizonts (arrow) within liver cells (bar = 5 µm). C) *S. destruens* flagellated stage (bar = 1 µm). D) Putative *S destruens* life cycle. E) *Dermocystidium sp.* walled schizont (arrow) (bar = 200 µm). F) *Rhinosporidium seeberi* shcizont (bar = 100 µm). G) *R. seeberi* life cycle.

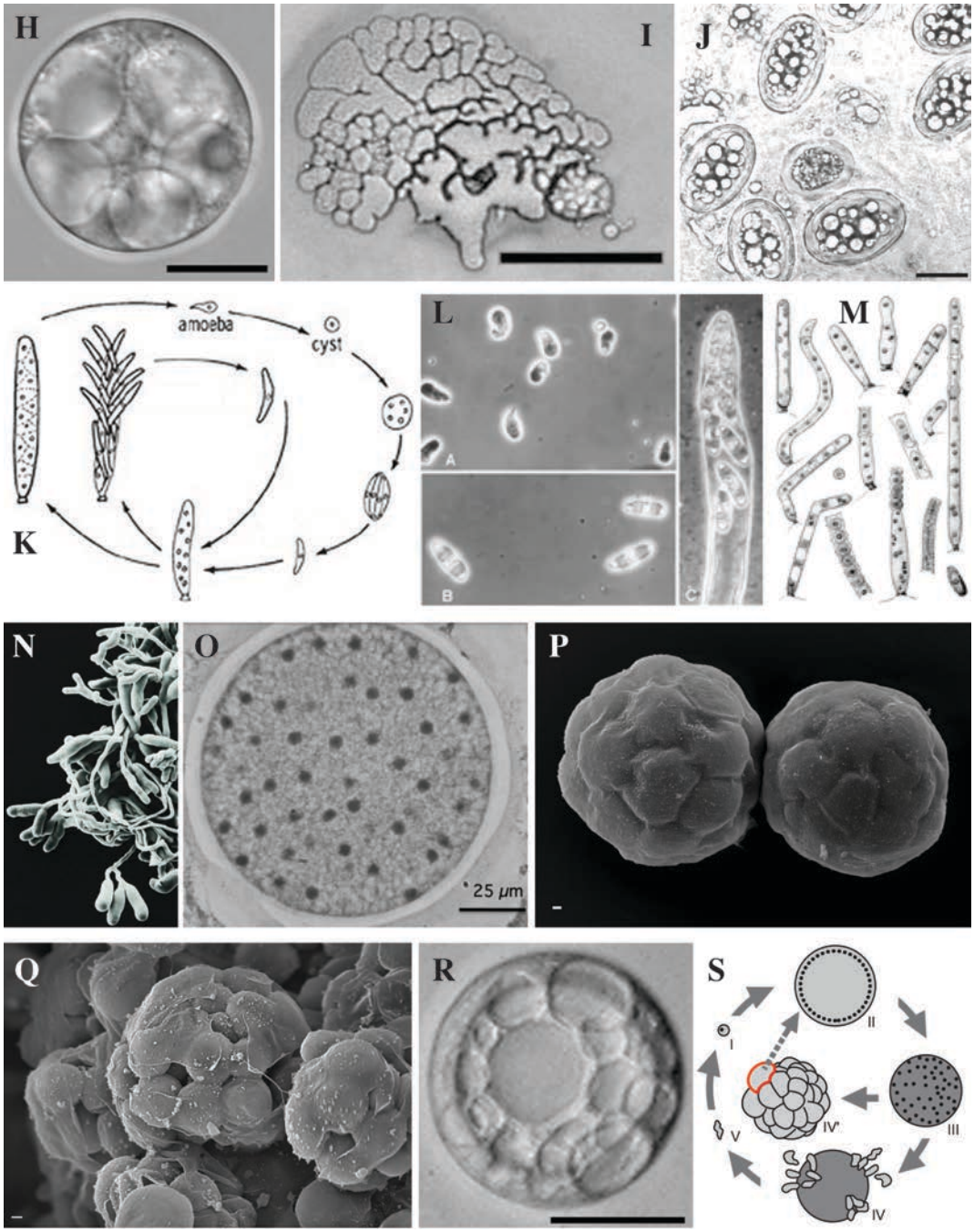


Figure 7 cont. Ichthyosporea biodiversity. Panels from H to S represent distinct Ichthyophonida species. H) *Abeoforma whisleri* schizont (bar = 20 µm). I) *Abeoforma whisleri* plasmodium (bar = 100 µm). J) *Psorospermium haeckelii* (bar = 50 µm). K) Amoebidiales life cycle. L) *Amoebidium australiense* merozoite, amoeba and schizont stages. M) Distinct Eccrinales forms of schizont thalli and cystic merozoites. N) *Ichthyophonus hoferi* merozoite thallus (x200). O) *Ichthyophonus hoferi* schizont (bar = 25 µm). P) *Sphaeroforma arctica* (bar = 1 µm). Q) *Creolimax fragrantissima* (bar = 1 µm). R) *C. fragrantissima* (bar = 1 µm). S) *C. fragrantissima* life cycle. A, C are scanning electron microscopy. B is a Normarski differential interference microscopy image. E and F are transmission electron microscopy. A and D adapted from (Arkush et al. 2003); B and C from (Paley et al. 2012); E from (Feist et al. 2004); F and G from (Vilela and Mendoza 2012); H and I from (Marshall and Berbee 2011); J from (Vogt and Rug 1995); K, L, and M from (Lichtwardt et al. 2001); N from (Spanggaard 1995); O from (Kocan 2013); P and Q courtesy of Dr. Sebé-Pedrós ; R from (Marshall et al. 2008); and S from (Suga and Ruiz-Trillo 2013).

1.2.4 Corallochytrium

Corallochytrium limacisporum is a small (4.5-20 μm) spherical walled free living saprotroph, first found in an Indian marine coral reef lagoon and classified as a Thraustochytrid based on its morphology (Raghu-Kumar et al. 1987). Its life cycle starts with a small uninucleated cell with a thin wall that divides binary, and the offspring keep attached during some rounds of cell division (up to 32 cells together) (**Figure 8**). Amoeboid limax-like cells can be observed crawling on the substrate until settlement, becoming then spherical and walled. Later on, this organisms was reclassified in its own group Corallochytrea using SSU rDNA within Opisthokonta, as sister group to Choanoflagellates (Cavalier-Smith and Paula Allsopp 1996); a topology systematically recovered using this marker in several publications (for example (Pereira et al. 2005)). In the famous six-kingdom system of life (Cavalier-Smith 1998), Cavalier-Smith suggested that *Corallochytrium* acquired osmotrophic lifestyle and lost its flagellum independently from fungi or any other Opisthokonta (but see **section 3.3 R3**). Trees based on few protein coding genes found Corallochytrium as sister group to *Amoebidium parasiticum* (Ichthyophonida) (Steenkamp et al. 2006). The same year, additional strains from the same Indian coral reef lagoon were isolated and grown in culture (Sumathi et al. 2006). The authors suspected fungal affinities of *Corallochytrium* and found the presence of the key enzyme (AAR) of the AAA lysine pathway, typical of fungi, so they proposed that the position of *Corallochytrium* within Holozoa should be revised (although they did not use the proper taxonomic sampling, see **section 3.1 R1** and **4.1**). Other phylogenetic studies using similar markers (some housekeeping genes and ribosomal subunits) noted the instability of this organism (branching in basically in almost all possible positions within Holozoa). In one study *Corallochytrium* was recovered sister to choanoflagellates (Ruiz-Trillo et al. 2006), then as sister to *Amoebidium* and *Ichthyophonus* (Carr et al. 2008), in another as the first holozoan splitting lineage (Brown et al. 2009), and finally sister to Filozoa (Paps et al. 2013). Thus, the phylogenetic position of *Corallochytrium* before remained highly contentious (**section 3.3 R3**).

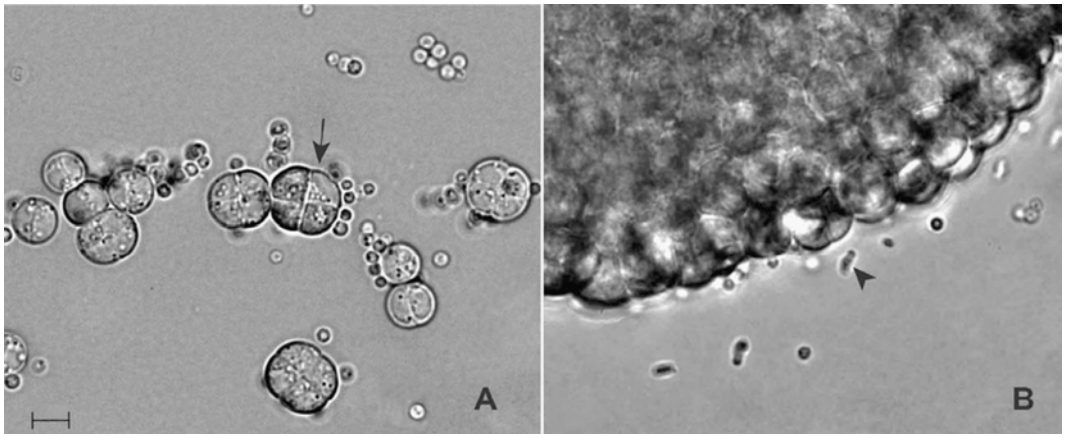


Figure 8. *Corallochytrium limacisporum*. A) clonal colonies, and B) amoeboid dispersal stages, from (Sumathi et al. 2006). Scale bar = 5 μm.

1.2.5 Nucleariids

A group of poorly known free-living naked filose amoebae composes the first splitting lineage in the Holomycota clade (**Figure 2**). Known since the 19th century, thanks to their abundance in fresh-water habitats and their relative big cell size (from 10-25 μm up to 60 μm). Classified within the group Filosea during the 20th century due to their morphology (spherical amoebae with long radiating filopodia without flagella), lifestyle (free-living predatory phagotrophs of filamentous cyanobacteria) and ultrastructure (discoidal mitochondrial cristae). Within nucleariids *sensu* Cienkowski 1865 (Patterson 1999), filose amoebae such *Pompholyxophrys* (Patterson 1985), *Pinaciophora* (Cavalier-Smith and Chao 2012), *Rabdiophrys*, or *Vampyrellidium* (Bass et al. 2009; Hess et al. 2012) shared taxonomic group with the *Nuclearia* genus. Currently, thanks to SSU rDNA molecular phylogenies, these species are classified among different rhizarian groups (Moreira et al. 2007; Adl et al. 2012), while the members of the genus *Nuclearia* (Patterson 1984) cluster within the Opisthokonta (Zettler et al. 2001; Hertel et al. 2002; Medina et al. 2003). *Nuclearia* contains 9 well-described species, mostly from European fresh-water or brackish habitats, but there are morphological reports from Antarctica (Tong et al. 1997). They share some common morphological traits: spherical or flattened protoplasm with radiating thin hyaline filopodia, usually with a central prominent nucleus with variable nucleoli, cytoplasm with no apparent microtubules and presence of a contractile vacuole (Mikrjukov and Mylnikov 2001). However, each species have some particularities, some are plurinucleated (such as *N. delicatula* (Blanc-Brude et al. 1955; Cann 1986)), some have branching filopodia (**Figure 9 D**) (like *N. moebiusi* (Patterson 1983)), have cystic stages (*N. simplex* (Mignot and Savoie 1979; Patterson 1984)), while others have some kind of extracellular matrix (like in *N. rubra* (Patterson 1984)). For a summary of morphological characters see (Yoshida et al. 2009).

Only two new *Nuclearia* species have been lately described: *N. pattersoni* (**Figure 9 C**), isolated from the gills of a fresh-water fish (Dyková et al. 2003), and *N. thermophila* (**Figure 9 A**) isolated from a hot spring in Japan (Yoshida et al. 2009). Also, only few environmental sequences have been positioned with the group (Zettler et al. 2002; Couradeau et al. 2011; del Campo and Ruiz-Trillo 2013). Recently, *Fonticula alba* (Worley et al. 1979) has been positioned as sister group to *Nuclearia* (Brown et al. 2009). *F. alba* is a small filose amoeba (5-10 μm) (**Figure 9 E and F**), that feeds from bacteria and forms aggregative multicellular fruiting bodies using a stalk formed with golgi derived extracellular matrix (Deasey and Olive 1981).

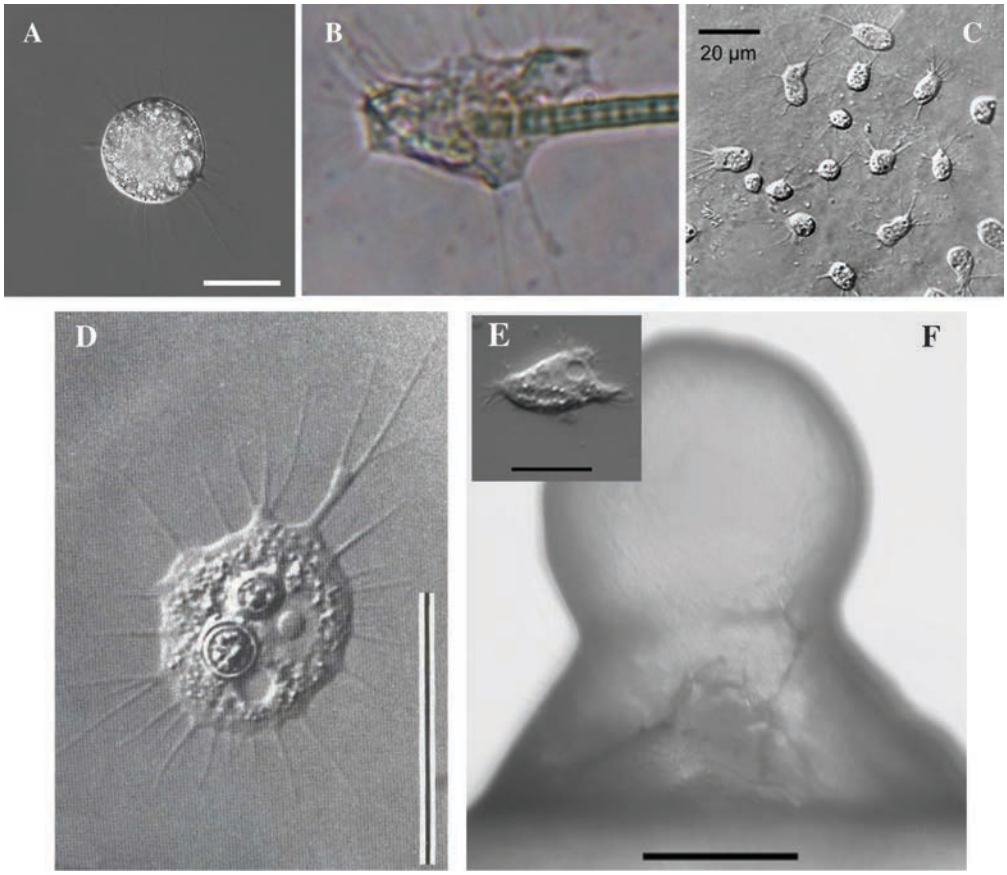


Figure 9. Nucleariids morphology. A) *Nuclearia thermophila* filose amoeba (bar = 25 µm) (Yoshida et al. 2009); B) *Nucleari aff. delicatula* feeding from *Phormidium sp.* (Cyanobacteria); C) *Nuclearia pattersonii* (Dyková et al. 2003); D) *Nuclearia moebiusi* (bar = 20 µm) (Patterson 1984); E) *Fonticula alba* trophic amoeba (bar = 10 µm); and F) *F. alba* fruiting body (bar = 500 µm); both from (Brown et al. 2009).

1.2.6 Opisthosporidia

The Opisthosporidia (Karpov et al. 2014) is a group recently proposed based solely on ribosomal trees, which comprises Cryptomycota (a.k.a. Rozellida or Rozellomycota) (James and Berbee 2012), Aphelida (Karpov, Mamkaeva, Aleoshin, et al. 2014) and the Microsporidia (Vávra and Lukeš 2013). They are all intracellular parasites with a naked amoeboid vegetative stage, a cystic stage and a specialized apparatus for penetration into host cell.

Aphelida is represented by at least three cultured species, all parasites of both marine and fresh-water algae, as well as some environmental sequences (Karpov et al. 2014). Previously proposed as *incertae sedis* opisthokonts (Pinevich et al. 1997) or as putative members of the Ichthyosporea (Adl et al. 2005) due to their cell morphology, the presence of flat mitochondrial cristae, parasitic lifestyle and partial SSU rDNA sequences. Last year full ribosomal sequences were obtained and they were positioned as sister group to Fungi clustering with *Rozella* and Microsporidia (Karpov et al. 2013; Letcher et al. 2013). Although similar life cycle, the morphology is quite variant between the different described species. The life cycle starts as a dispersal naked flagellated or amoeboflagellated stage that attaches to the host wall where it retracts the flagellum and encyst. The cyst wall provides the resistance to the enlargement of the posterior vacuole that pushes the parasitic protoplasm into the host wall through an infection tube. Inside, the parasitic amoeba phagocytise the host protoplasm using pseudopods and then develops as a multinucleated plasmodium. Once the host is consumed the plasmodium can encyst using the host cell wall or cellularize to release the offspring (see **Figures 10 D to G**).

Cryptomycota (Jones, Richards, et al. 2011) (a.k.a. Rozellida (Lara et al. 2010) or Rozellomycota (Corsaro et al. 2014)) is a large group that includes many diversified environmental sequences (Jones, Forn, et al. 2011), among them the clade LKM11, the first environmental sequences to be proposed as sister group to the filamentous fungi (Lara et al. 2010). There is only one characterized genus and one culturable species (*Rozella allomycis*), whose genome has been recently sequenced (James and Berbee 2012; James et al. 2013). *Rozella* was traditionally classified as a chytrid fungi but early SSU rDNA phylogenies already pointed to its further deep position in the fungal tree (James, Letcher, et al. 2006). Cryptomycota share many similarities with aphelids (**Figures A to C**), such as the intracellular amoeba or the dispersal flagellated cell types, as well as the parasitic life cycle. The main difference is the host range, fungi-like organisms in the case of *Rozella*, and its lack of amoeboid dispersal stage (Karpov et al. 2014).

Finally, **Microsporidia** represents an extremely derived group of intracellular parasites. They principally parasite animals, for which they seem to have deep strong phylogenetic association (Smith 2009). They are called “energy parasites” as they use ATP from the

host like parasitic organelles, and have many simplified characters such as reduced genome, ribosomes, and endomembrane system. They even lack dictyosomes, peroxisomes and canonical mitochondria. Such characteristics, along with the special structure polar tube they all use to penetrate the host cytoplasm, made these organisms with a uniform life cycle across different lineages a clear monophyletic group, but with no easy classification among other eukaryotes (Corradi and Keeling 2009). Currently, several independent phylogenetic studies have established them as sister to the filamentous fungi (Baldauf et al. 2000; Liu et al. 2006; Capella-Gutiérrez et al. 2012).

The Microsporidia produce intracellular sporonts, the offspring surrounded by a chitinous cell wall for dispersion, probably homologous to filamentous fungi. Some molecular synapomorphies between Microsporidia and Fungi have been proposed (Vávra and Lukeš 2013), but they now need to be revalidated including molecular data from other Opisthosporidia groups. Analysis of the genome sequence of *Rozella* already revealed that it encodes four chitin synthases from division II (one homologous to the only one in Microsporidia) (see **section 3.3 R3**) and some genes acquired by horizontal gene transfer (HGT) from the prokaryote *Chlamydia*, which may be considered shared synapomorphies with Microsporidia. Also, mitochondrial gene content presents signal for reduction suggesting a progressive specialization similar to the one in Microsporidia, although the metabolism is typical of a phagotroph. In any case, the Microsporidia are not particularly informative to reconstruct Opisthokonta ancestral state, since they are considered extremely derived species that have retained very few ancestral characteristics.

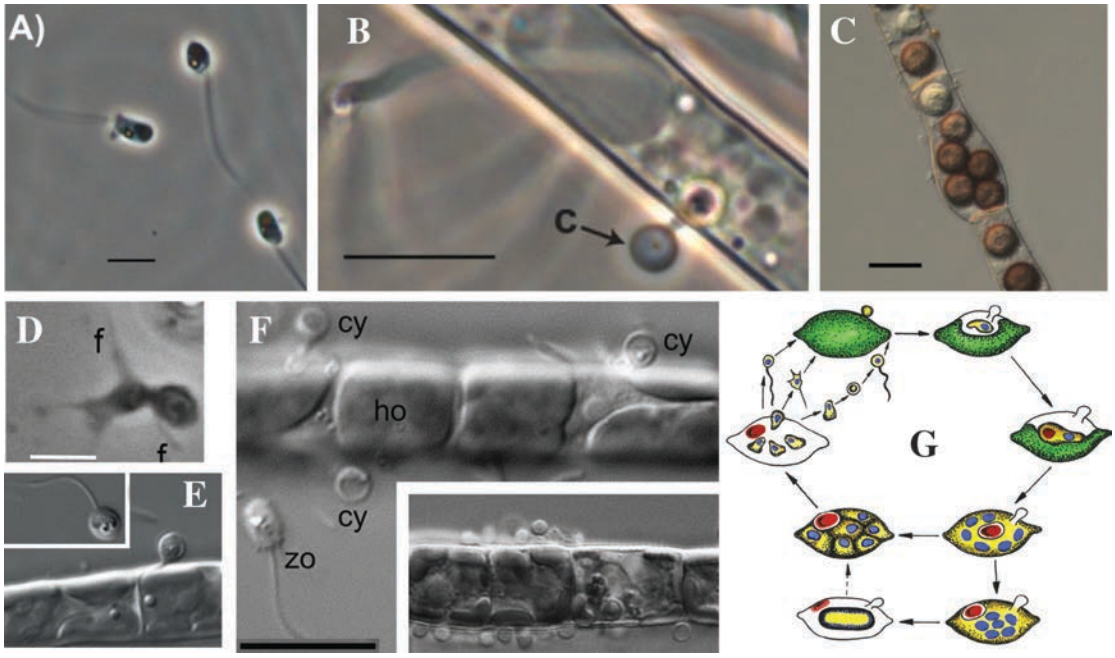


Figure 10. A to C represent respectively *Rozella allomycis* flagellated stage (bar = 5 μm), cyst and germ tubule infecting host (bar = 10 μm) and cysts within host cell wall once consumed (bar = 20 μm) (James and Berbee 2012). D to G represent respectively *Amoeboaphelidium protococcarum* filose amoeba (bar = 2 μm), *Aphelidium aff. melosirae* flagellated stage and infecting stage (bar = 10 μm), *Aphelidium sp.* flagellated stage (zo) before attachment to the host (ho) and cysts (cy) (bar = 15 μm) and life cycle (Karpov, Mamkaeva, Aleoshin, et al. 2014; Karpov, Mamkaeva, Benzerara, et al. 2014).

1.2.7 The early branching “fungi”

To understand Opisthokonta protists as a whole it is important to take into account the flagellated coenocytic filamentous fungi (**Figure 11**): the chytrid fungi (non-capitalized name in purpose, as it may be a paraphyletic group, **section R2** and **R3**). Filamentous fungi are the traditional Fungi kingdom: chytrids, zygomycetes and Dikarya. Nowadays it seems there are no synapomorphies for them (McLaughlin et al. 2009), since osmotrophy, chitin cell wall (**section 3.3 R3**), hypha or the specific lysine biosynthetic pathway (**section 3.1 R1**) are all found in other eukaryotic lineages. They all share, however, an osmotrophic/saprotrophic feeding mode and a chitinous cell wall during all its life cycle (except for flagellated stages in chytrids and the *incertae sedis Olpidium* (Sekimoto et al. 2011)). Filamentous fungi are still poorly studied and some lineages are exclusively represented by environmental samples (Richards et al. 2012; Manohar and Raghukumar 2013). Classification of animal phyla is still under no clear consensus, but in the case of Fungi the monophyly of certain groups is not even clear (Hibbett et al. 2007; McLaughlin et al. 2009; Ebersberger et al. 2012). Although most of the fungi diversity remains poorly understood, it is commonly accepted that the major transition from aquatic to soil or terrestrial environments took place after chytrids branched off from the rest of fungi. Recent molecular phylogenies show that the chytrid fungi may be paraphyletic, while Blastocladiomycetes (Porter et al. 2011) may be sister to zygomycetes (also paraphyletic) and Dikarya (Tanabe et al. 2005; James, Letcher, et al. 2006; James, Kauff, et al. 2006; Y. Liu et al. 2009). Both Chytridiomycota (the monophyletic taxonomic name) and Blastocladiomycota have cell cycles within aquatic habitats involving flagellated dispersal stages, as well as amoeboid forms with pseudopodia (Hoffman et al. 2008) (**Figure 11 A**); but no true hyphae or complex sexual reproduction as in other filamentous fungi, only anisogamous gametes (Pommerville 1982) (**Figure 11 B**). Therefore, some authors have proposed that these two groups should be considered protists (Margulis 1990). In fact, coenocytic filamentous fungi have many similarities with the Ichthyosporea (**section 1.2.3**). It might be important, therefore, to study and compare chytrid fungi with ichthyosporeans (and other fungi-like organisms such as oomycetes or labyrinthulids) to better understand the transition from phagotrophy to osmotrophy lifestyles (Richards and Talbot 2013); see **section 4.2.3**.

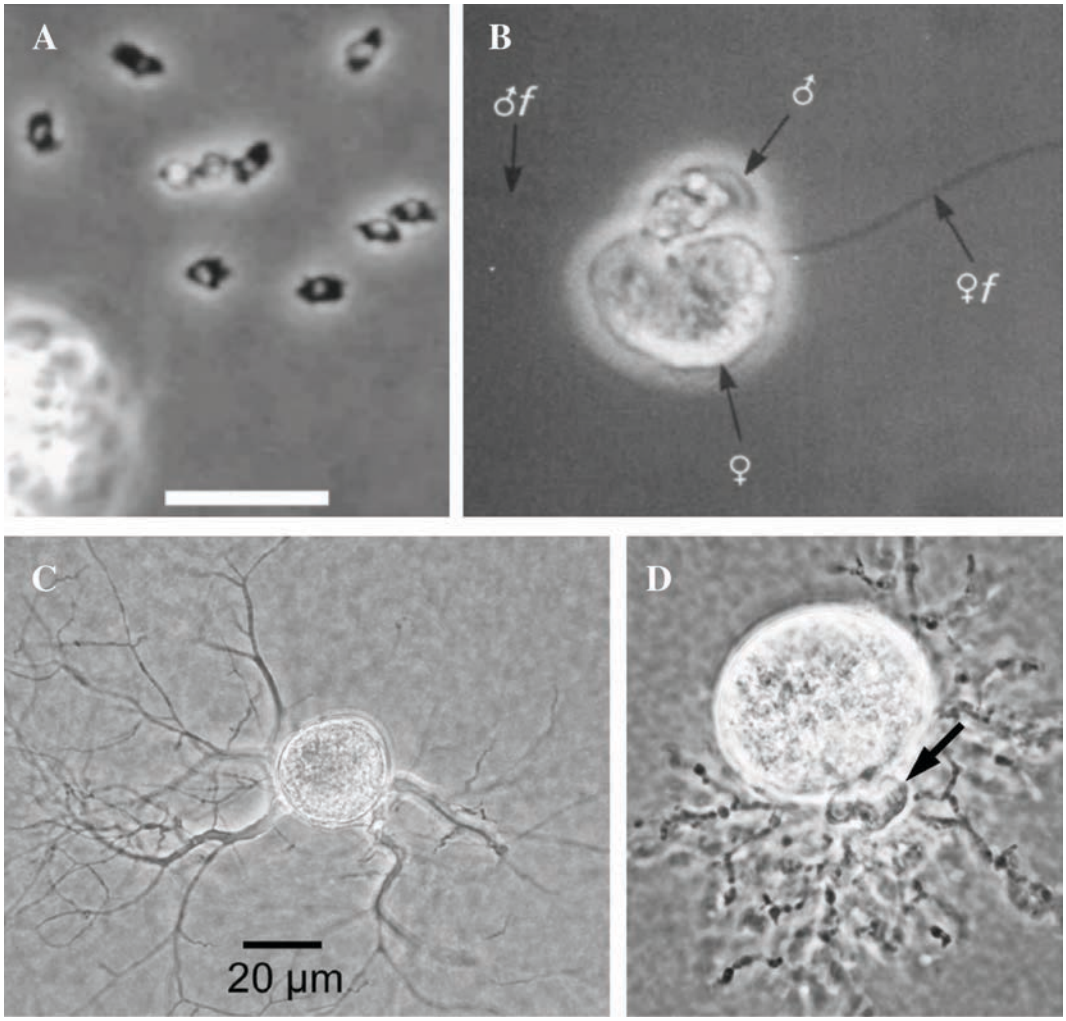


Figure 11. Chytrid morphology. A) *Physoderma sp.* (Blastocladiomycetes) amoebae (bar = 10 μm) (Hoffman et al. 2008); B) *Allomyces macrogynus* (Blastocladiomycetes) gamete mating (x1500) (Pommerville 1982); C) *Neokarlingia chitinophila* (JEL510) (Chytridiomycetes) coenocyte with rhizoids (Longcore and Simmons 2012) ; and D) *Dendrochytridium crassum* (Chytridiomycetes) “thallus” (Letcher et al. 2014).

1.3 The putative sister groups to the Opisthokonta: "Apusozoa" and Breviatea

The evolutionary history of the opisthokonts cannot be fully understood if we do not consider its extant closest living relatives. Before this PhD project, the sister group of Opisthokonta was considered to be the Amoebozoa (Smirnov et al. 2011; Glöckner and Noegel 2012) confirmed by molecular phylogeny (Bapteste et al. 2002). Both Opisthokonta and Amoebozoa formed the unikont clade (Cavalier-Smith 2002), since the ancestor was considered to be uniflagellated (while the remaining eukaryotes were termed "bikonts"). However, a few SSU rDNA and housekeeping gene-based phylogenies already pointed to the biflagellates Apusozoa as potential close relatives of Amoebozoa and Opisthokonta or, as sister group to Opisthokonta (Cavalier-Smith and Chao 1995; Kim et al. 2006) (see further in **section 1.4.1**). The Apusozoa include gliding bacterivores such as **Apusomonadida** (**Figure 12 A**) (Ekelund and Patterson 1997; Cavalier-Smith and Chao 2010), **Ancyromonadida** (**Figure 12 B**) (Heiss et al. 2010; Glücksman et al. 2013) and Mantamonadida (Glücksman et al. 2010). Although classified together based on morphology (Cavalier-Smith et al. 2008), Apusomonadida and Ancyromonadida show paraphyly in most SSU rDNA and housekeeping gene phylogenies (Cavalier-Smith et al. 2008; Cavalier-Smith and Chao 2010; Paps et al. 2013), see **section 3.3 R3**.

Another group of bacterivorous flagellates already proposed to be somehow related to Amoebozoa and Opisthokonta are the **Breviatea** (**Figure 12 C**) (Cavalier-Smith et al. 2004). These amoeboflagellates are found in hypoxic environments and contain mitochondria related organelles (MRO) (Stairs et al. 2014). Although originally classified as Amoebozoa, some phylogenies showed them related to the Apusomonadida (Walker et al. 2006; Katz et al. 2011). Recent EST based trees show them either as sister to Amoebozoa (Minge et al. 2009) (but no Apusomondadida included), as sister to Opisthokonta (Zhao et al. 2012), or close to Apusomonadida (Brown et al. 2013); see **section 3.3 R3**.

Due to this new data the sister group of Opisthokonta seem to be biflagellated amoeboid bacterivores, presenting most of the morphological characters present in distinct Opisthokonta groups: filopodia, cysts, flagellum, etc, even some molecular characters such the integrin machinery (Sebé-Pedrós et al. 2010; Brown et al. 2013), see **section 4.2.1**.

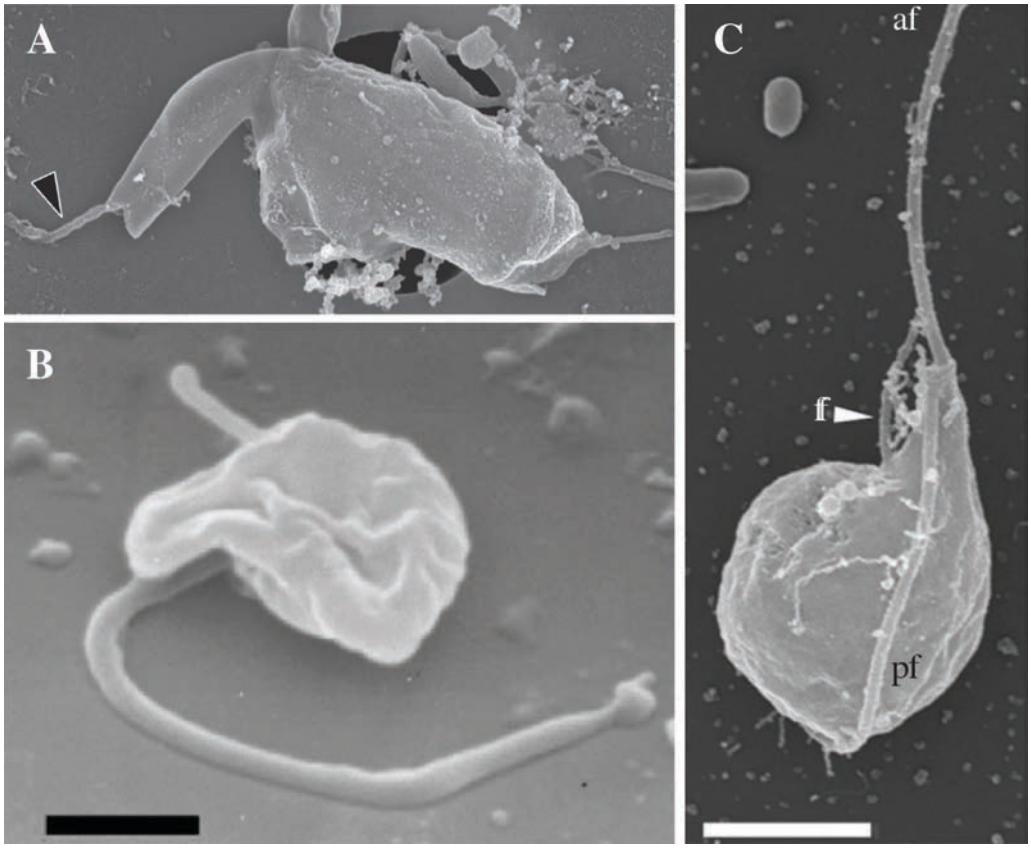


Figure 12. Biflagellates putatively closely related to Opisthokonts. SEM images. A) *Thecamonas trahens* (Apusomonadida) cell with ventral feeding groove (bottom of the Figure) and sleeve surrounding the shorter anterior flagellum (arrowhead) (Heiss et al. 2013); B) *Planomonas mylnikovi* (Ancyromonadida) (scale bar = 1 μm) (Cavalier-Smith et al. 2008); and C) *Pygsuia biforma* (Breviatea) gliding form with posterior (pf) and anterior (af) flagellum and a filopodia (f) (bar = 2.5 μm) (Brown et al. 2013).

1.4 Inferring deep phylogenies

Phylogenetic methods have significantly advanced in the past couple of decades. Thanks to the development of more complex phylogenetic methods, and better and cheaper high-throughput sequencing along with computational resources, we are now in the position to address the most complex questions in eukaryotic evolution. These questions include the root of eukaryotes (Derelle and Lang 2012; He et al. 2014), the order between extant living eukaryotic supergroups (Burki 2014); the position of *incertae sedis* lineages (Adl et al. 2012) such Ancyromonadida (Cavalier-Smith et al. 2014), Breviatea (Brown et al. 2013), *Collodictyon* (Zhao et al. 2013) or *Malawimonas* (Derelle, Torruella, Elias et al. in prep); as well as the internal relationships within supergroups. This is the case for the Opisthokonta tree (Ruiz-Trillo et al. 2008; Shalchian-Tabrizi et al. 2008; Ebersberger et al. 2012), the resolution of which is the main goal of this PhD project.

Although phylogenetics methods have improved a lot, they are not without problems. Thus, below I describe 1) the most important caveats of inferring complex deep phylogenetic questions using single gene trees or rare genomic changes (which serves as introduction to **section 3.1 R1**); and 2) how phylogenomics works and which are its drawbacks (introducing **sections 3.2 R2** and **3.3 R3)**.

1.4.1 Caveats in deep phylogenetic analyses: ultrastructure, single gene markers and rare genomic changes

As mentioned in **section 1.1**, in late 20th century ultrastructural characters were widely accepted to be valid markers to classify protistan lineages. Evolutionary protistologists considered character states such the presence/absence of mitochondria (e.g., the Archezoa hypothesis from Cavalier-Smith in 1987), the shape of mitochondrial cristae (tubular/flat), or the pseudopodia (lobose/filose) among other morphological patterns (Cavalier-Smith 1993; Cavalier-Smith 1998; Patterson 1999) as evolutionary steady. From these assumptions, groups such Archezoa or Filosea were proposed to refine the Protoctista/Protozoa clades (Parfrey et al. 2006), but progressively abandoned when molecular phylogenies proved these character states to be patchy distributed among protists, and much more variable than previously thought. As a clear example, most opisthokonts show flat mitochondrial cristae, but some of them present tubular shapes: *I. hoferi* (Spanggaard 1995), *Codosiga spp*. (Wylezich et al. 2012) or *Aphelidium aff. melosirae* (which indeed presents both shapes depending on the stage of the life cycle) (Karpov, Mamkaeva, Benzerara, et al. 2014). Thus, morphological characters should be taken with care, analyzed under a wide taxonomic sampling and exhaustively studied before considering them valid markers.

Small subunit ribosomal RNA (SSU rDNA) is the broadest used phylogenetic marker since it is present in all eukaryotic species and contains both conserved and variable regions. Moreover, its crucial function guarantees no complex evolutionary history (e.g., lateral gene transfer, domain shuffling, etc.). Indeed, SSU rDNA phylogenies have been very successful in unravelling phylogenetic relationships. However, this marker alone has not enough resolution to solve most deep nodes (see a recent example for the Opisthokonta in (del Campo and Ruiz-Trillo 2013) or others mentioned in **section 1.2)**. In general, studies show conflicting topologies with low statistical support. This lack of resolution is represented in a phylogeny by short branches between deep splits, and it is assumed to be an evidence for rapid speciation between lineages, whose molecular character states were not especially shaped in each lineage. The main problem for **single marker trees** is the relative amount of informative sites (the ones with true phylogenetic signal) compared to the amount of saturated sites (the ones with phylogenetic information blurred due to accumulation of multiple changes, which only provide random noise (see **Figure 13**)).

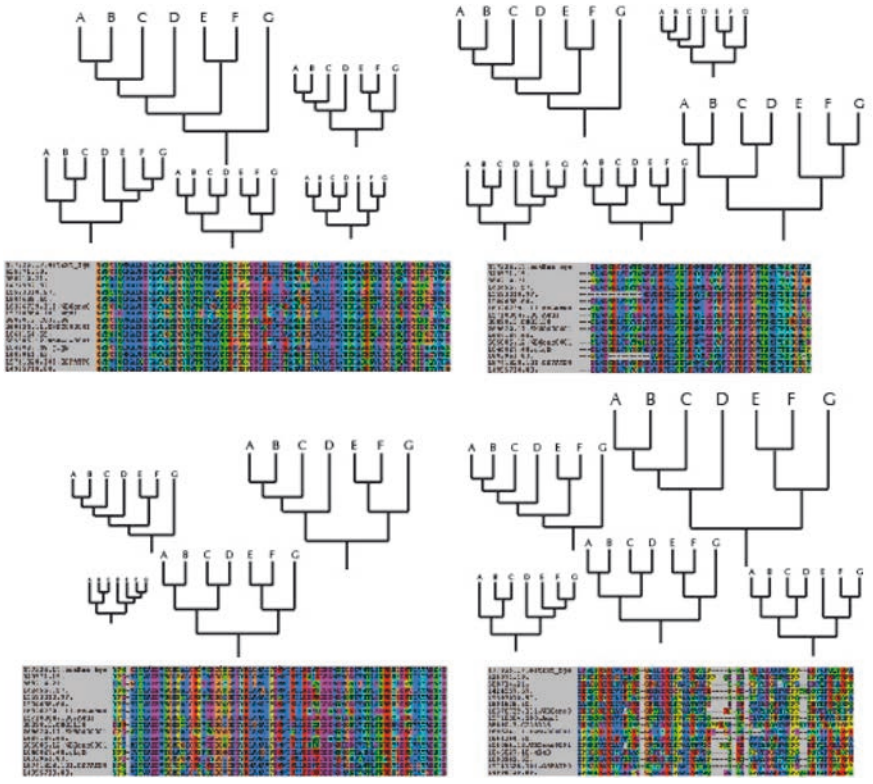


Figure 13. Four distinct molecular markers (individual gene or protein alignments) have a few positions and therefore are prone to stochastic errors, meaning several distinct topologies can be inferred with low supports.

Homoplasic sites are an example of saturated sites. They share the same character state for distinct taxa not because of ancestry, but because of convergence (e.g., A -> T -> A). This can create a false “impression” that two unrelated fast evolving species (whose sequences have evolved at a faster rate than others) are closely related. This what is

known as the Long Branch Attraction (LBA) artefact (Felsenstein 2004; Bergsten 2005). In the case of nucleotide-based alignments (such in SSU rDNA), this is a serious issue since there are only 4 possible character states and, thus, few changes are needed to create a LBA effect. Protein coding genes, instead, present 20 possible character states (Dayhoff et al. 1978) and therefore are less prone to such artifactual clusters. However, they provide fewer sites (each 3 nucleotides is a codon and a single amino acid) and are not more informative *per se*. Protein coding genes such as actin, tubulins or heat-shock proteins among others have been extensively used in phylogenetics, under the premises that they as well have "simple" evolutionary histories due to their fundamental cell functions. However, this may not be true for all of them. Horizontal gene transfer events in α-tubulin have been reported (Simpson et al. 2006); multiple paralogs of β-tubulin can be found in Opisthokonta protists (Steenkamp et al. 2006); and elongation factor-1α (EF-1α) has an ancient paralog EF-like with a patchy distribution in eukaryotes (Keeling and Inagaki 2004; Kamikawa et al. 2013). These are examples of orthology miss-assignment or paralogy problems, which happen when we compare sites that have not diverged from speciation, but from other mechanisms such as duplication, recombination, etc. They also state an important caveat in species phylogeny reconstruction, which is the fact that single genes can have distinct evolutionary histories than the species tree (also see **section 3.2 R2**).

Other ways to infer evolutionary relationships between distant lineages are the **rare genomic changes**. Qualitative molecular data (treated as binary data, in terms of presence/absence, similarly as for morphological characters) like mitochondrial gene order (Boore 2006), indels (Baldauf and Palmer 1993), gene fusions (Stechmann and Cavalier-Smith 2002; Richards and Cavalier-Smith 2005) or the presence/absence of certain molecules like key enzymes for metabolic pathways (Sumathi et al. 2006) or microRNAs (Philippe, Brinkmann, Copley, et al. 2011), are often assumed so infrequent to happen twice independently, that if two groups share them, they must be closer than others that do not present such character state. With the increase of genomic/transcriptomic data for more and more taxa, most of these putative molecular synapomorphies have been proven wrong or remain contentious (Leonard and Richards 2012; Shadwick and Ruiz-Trillo 2012; Thomson et al. 2014). As an example, the division of eukaryotes between biflagellate species (bikonts, e.g., Archaeplastida, Excavata, Apusomonadida, etc.) and ancestrally uniflagellates (unikonts, e.g., Amoebozoa and Opisthokonta) was based on the number and structure of the flagellar apparatus, and a dihydrofolate reductase (DHFR) and thymidylate synthase (TS) gene fusion present in biflagellated lineages but not in ancestrally uniflagellates (Stechmann and Cavalier-Smith 2002; Stechmann and Cavalier-Smith 2003). Such hypothesis was severely questioned once biflagellated species, particularly *Breviata anathema* (Breviatea) and

Apusomonadida were placed between or within unikonts in single gene molecular phylogenies (Minge et al. 2009) and (Kim et al. 2006), respectively (Roger and Simpson 2009; Rogozin et al. 2009) (see **sections 1.3**, **3.2**). The problem is that most rare genomic changes are not studied enough in terms of biological functions, for which they should be heavily constrained to be valid markers. Moreover, in practice, rare genomic changes provide cyclic arguments, as they need to be constantly tested by phylogeny and using the widest possible taxonomic sampling.

1.4.2 A pipeline for phylogenomics

Thanks to the large amount of genomic and transcriptomic data available, the phylogenomics approach has been received much attention lately. The idea is that multiple gene markers may help to recover the scarce phylogenetic signal, from the abundant random noise (Delsuc et al. 2005); see **Figure 14**. There are several distinct protocols or bioinformatic pipelines to perform phylogenomic analyses to recapitulate speciation events. These include concatenation of different markers known as the "supermatrix approach" (de Queiroz and Gatesy 2007; Roger et al. 2012); the supertree approach which generates several single marker trees and perform distinct types integration into a consensus topology (Bininda-Emonds 2004); phylogenetic networks (Huson et al. 2010); coalescent methods (L. Liu et al. 2009); or even recent alignment-free methods (Chan and Ragan 2013). The supermatrix methods basically follow the same steps as single marker phylogenies: multiple sequence alignment (MSA) (Katoh et al. 2002), trimming of spurious regions (Talavera and Castresana 2007; Capella-Gutiérrez et al. 2009; Criscuolo and Gribaldo 2010) and inferring phylogenies through Maximum Likelihood (Guindon and Gascuel 2003; Stamatakis 2006) or Bayesian methods (Ronquist and Huelsenbeck 2003; Lartillot and Philippe 2004) with distinct evolutionary models (Dayhoff, WAG, LG, GTR, etc.). These multiple markers need to be assembled and analyzed through at least a semi-automatic pipeline with the aim of standardization and reproducibility (Roure et al. 2007; Grant and Katz 2014; Struck 2014). But the most complex steps of the pipeline (i.e., to use a proper taxonomic sampling for the particular evolutionary question, to obtain a proper set of orthologous markers and deal with systematic errors) are difficult to automatize (Philippe, Brinkmann, Lavrov, et al. 2011) (**Figure 15**).

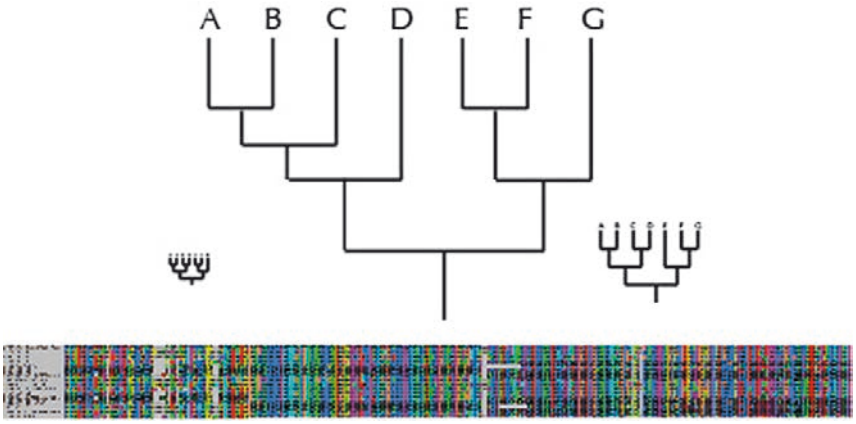


Figure 14. A concatenated dataset of multiple markers provides huge numbers of sites. This diminishes the sampling error or stochasticity, by increasing the signal-to-noise ratio. In theory, the increasing number of sites with directed signal heavily compensates the non-directed random noise. In this case, fewer possible topologies can be estimated.

1.4.2.1 Taxon sampling

For any phylogenetic analysis, taxon sampling is key. It is important to properly choose the representatives of the group of interest. Ideally, lineages with similar taxes of sequence evolution (i.e., similar rates of substitution and therefore branch lengths) should be used to minimize model violations (see next **section**). Moreover, a full representation of distinct lineages can help to break long branches. But for the same reason, it is also important to generously sample the outgroups (de la Torre-Bárcena et al. 2009). Outgroups should be the closest possible ones, since distant sequences provide higher levels of saturation and therefore LBA. In brief, one should always use the broadest taxonomic sampling possible, within the computational limits.

Another issue is the “matrix asymmetry” (Philippe et al. 2004; Ryberg and Matheny 2011), a problem when using several markers that are filled in by a portion of the taxonomic sampling (e.g., Sanger ESTs or poorly sequenced taxa). Some authors have proposed that asymmetrical matrices in simulated datasets can mislead topology reconstruction and exacerbate model violations (Hartmann and Vision 2008; Lemmon et al. 2009). Most real dataset analyses have shown, however, that the relative amount of missing data is not the most important factor in determining whether a correct phylogeny can be computed. Instead, the absolute amount of available informative data within an alignment is more important. In these cases, it is more beneficial to add a smaller number of characters for more taxa than a larger (e.g., more markers) for fewer taxa, as certain amount of missing data can be tolerated (Wiens 2006; Sanderson et al. 2010; Hinchliff and Roalson 2013; Roure et al. 2013).

1.4.2.2 Orthologous dataset

The next step in designing a phylogenomic study is the dataset itself. To infer species phylogeny from multiple protein markers is fundamental that single sequences are used

for each taxon, and that the relationship between these sequences is orthologous (i.e., descending from a speciation event and not due duplication, HGT, etc. (Gabaldón 2008)). This is the reason why several automatic orthology assignment methods have been developed (Sjölander et al. 2011), including the widely used orthoMCL (Chen et al. 2006). They mainly work by progressively clustering sequences by Blast and statistically establishing groups, but they are limited in discovering data contamination or hidden paralogy (Philippe, Brinkmann, Lavrov, et al. 2011). Therefore, manual curation with the aid of single marker trees is still necessary (see **section 3.2 R2**).

1.4.2.3 Systematic errors

Despite the massive amount of data in phylogenomic studies and in consequence the decrease of stochastic error, large-scale genome approaches have revealed significant incongruence among inference methods and datasets (Jeffroy et al. 2006). This is because the reliability of the tree is measured by standard statistical indices such as non-parametric bootstrap analysis (Felsenstein 1985) or Bayesian posterior probabilities (Rannala and Yang 1996). These statistical indices only assess sampling effects and provide an indication of tree reliability that is conditional on the data and the method. Thus, if the method does not correctly handle properties of the data, an incorrect tree can receive strong statistical support (Philippe et al. 2005; Salichos and Rokas 2013). Then, such incongruences between analyses come from systematic errors, which happen when datasets are not properly modeled by evolutionary models (Delsuc et al. 2005). For example, that sequences evolve under homogenous and stationary conditions (at the same constant rate and with base compositions similar to that of the ancestral sequence, respectively; see next paragraph). When these assumptions are violated, methods tend to converge towards a highly supported incorrect tree, artifactual non-phylogenetic signal is generated and it competes with the genuine phylogenetic signal. Usually, the more structured phylogenetic signal that leads to the same topology will be preferentially arise in contrast to the noise that normally leads to different topologies. However, when the phylogenetic signal is weak, such as for ancient phylogenetic relationships, the inference can be misled. Indeed, different sources of systematic error have been detected and analyzed, and some practices have been proposed to deal with them (Phillips et al. 2004; Delsuc et al. 2005; Philippe et al. 2005; Jeffroy et al. 2006; Rodríguez-Ezpeleta, Brinkmann, Roure, et al. 2007).

As mentioned above, most of these sources of errors come from the variability or heterogeneity in the rates of evolution, as models assume homogeneity, meaning that the pattern of substitutions is the same in all parts of the tree (between species, sites or time). However, it is known that the same molecule can change at different rates in different

lineages. Moreover, distinct regions within the molecules change faster or slower (functional domains are more constrained than interdomain regions, for example). This is currently addressed in all phylogenetic studies correcting for a gamma distribution (Yang 1993). Finally, the rates of evolution among sites are not constant through time. This is called heterotachy (Lopez et al. 2002) and it can be addressed using covarion (Wang et al. 2009; Zhou et al. 2010) and mixture models (Lartillot and Philippe 2004), although at a high computational cost. In practice, such model violations on rate heterogeneity mislead topology inference, but can also produce LBA. The other factors that clearly cause LBA is the already explained substitutional saturation and the compositional bias or compositional heterogeneity. In this case, distinct species with similar amino acidic frequencies can be clustered together (Foster 2004; Criscuolo and Gribaldo 2010; Nesnidal et al. 2010).

A part from the mentioned corrections to model violation, filtering the data is also a good strategy. Data can be filtered for taxa, genes or sites that are prone to highly violate model assumptions. For example, taxa with long branches can be removed from individual alignments without compromising matrix symmetry, or whole markers can be removed if their single-gene phylogenies show complex patterns of branch length for many taxa. One of the most used strategies is to remove the so-called fast evolving sites (Ruiz-Trillo 1999; Rodríguez-Ezpeleta, Brinkmann, Burger, et al. 2007; Minge et al. 2009). Basically sites are classified under arbitrary categories under the gamma distribution (Yang 1993) (or similar distributions (Susko et al. 2003)) and then removed from the dataset. The remaining sites are in theory less prone to violate models. Another way to minimize sources of systematic error such compositional bias are the recoding strategies, where amino acids are grouped into distinct classes (Phillips et al. 2004; Rodríguez-Ezpeleta, Brinkmann, Roure, et al. 2007; Susko and Roger 2007; Ruiz-Trillo et al. 2008), but this can diminish phylogenetic signal as well.

In summary, overcoming systematic errors in phylogenomics is a demanding task; multiple parallel analyses must be performed to contrast incongruences, and discern between true phylogenetic signal and artifacts.

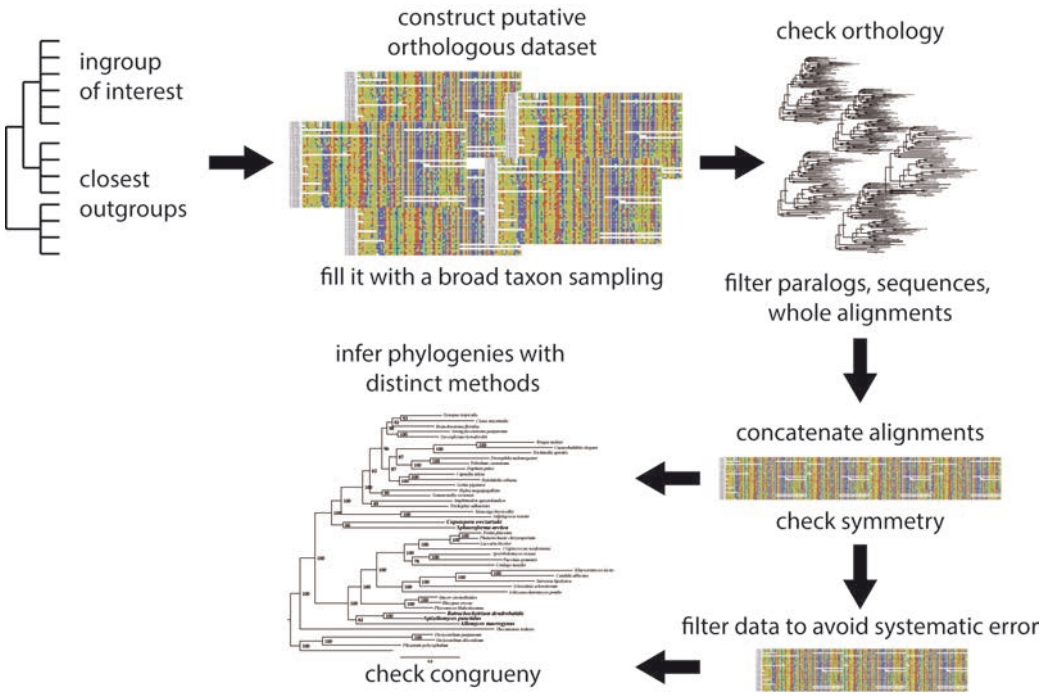


Figure 15. A schematic protocol for phylogenomics based on the supermatrix approach.

2. OBJECTIVES

From these remarks it will be seen that I look at the term species, as one arbitrarily given for the sake of convenience to a set of individuals closely resembling each other, and that it does not essentially differ from the term variety, which is given to less distinct and more fluctuating forms. The term variety, again, in comparison with mere individual differences, is also applied arbitrarily, and for mere convenience sake.

[...] I was much struck how entirely vague and arbitrary is the distinction between species and varieties.

Darwin 1859

The UNICORN project (Ruiz-Trillo 2007) consisting on sequencing the genomes of key taxa for the study of multicellularity in both animals and fungi, provided the framework to establish the following objectives:

1. To reconstruct phylogenetic relationships of different Opisthokonta taxa based on the whole-genome data from the UNICORN project, in order to provide a backbone to study the transition to multicellularity in both fungi and metazoans.

2. To increase the taxonomic sampling available for the Opisthokonta, especially within the Holozoa protists, by obtaining RNAseq data from additional lineages.

3. To re-evaluate phylogenetic relationships among all known Opisthokonta lineages with the wider possible taxonomic sampling.

4. To infer putative evolutionary transitions between lineages reconstructing ancestral states.

3. RESULTS

The view of evolution as chronic bloody competition [...] dissolves before a new view of continual cooperation. [...] Life did not take over the globe by combat, but by networking.

Marvellous microbes. Lynn Margulis & Dorion Sagan 2001.

In the animal world we have seen that the vast majority of species live in societies, and that they find in association the best arms for the struggle for life: understood, of course, in its wide Darwinian sense – not as a struggle for the sheer means of existence, but as a struggle against all natural conditions unfavourable to the species. The animal species, in which individual struggle has been reduced to its narrowest limits, and the practice of mutual aid has attained the greatest development, are invariably the most numerous, the most prosperous, and the most open to further progress. The mutual protection, which is obtained in this case, the possibility of attaining old age and of accumulating experience, the higher intellectual development, and the further growth of sociable habits, secure the maintenance of the species, its extension, and its further progressive evolution. The unsociable species, on the contrary, are doomed to decay.

Mutual aid, a factor of evolution. Pyotr Kropotkin 1902

INFORME DELS DIRECTORS SOBRE ELS ARTICLES PUBLICATS

Director: Dr. Iñaki Ruiz-Trillo

Tutora: Dra. Marta Riutort León

Els tres articles que conformen aquesta tesi doctoral (tots ells com a primer co-autor) han estat publicats o estan en vies de ser-ho en revistes d'alt impacte en l'àmbit de l'evolució, la genètica o la biologia; totes incloses a la base de dades PubMed. S'indiquen per cadascuna, quan són disponibles, l'índex d'impacte i la seva posició en rànquings de les corresponents disciplines. Els articles han estat en col·laboració amb altres grups d'investigació i membres del propi grup de recerca.

Article R1

Torruella G, Suga H, Riutort M, Peretó J, Ruiz-Trillo I. (2009). ***The Evolutionary History of Lysine Biosynthesis Pathways Within Eukaryotes.*** J Mol Evol. Sep;69(3):240-8.

Factor d'impacte (2009): 2.323

Posició dins l'àrea: Biochemistry & Molecular Biology = 167 de 283 (Q3)
Evolutionary Biology = 27 de 45 (Q3)
Genetics & Heredity = 81 de 146 (Q3)

El doctorand va participar activament en el disseny experimental i la realització de les anàlisis (cerca d'homologia i anotació de gens, alineaments i inferència filogenètica), així com en la discussió, la interpretació i la redacció de l'article.

Article R2

Torruella G, Derelle R, Paps J, Lang BF, Roger AJ, Shalchian-Tabrizi K, Ruiz-Trillo I. (2012). ***Phylogenetic relationships within the Opisthokonta based on phylogenomic analyses of conserved single copy protein domains.*** Mol Biol Evol. Feb;29(2):531-44.

Factor d'impacte (2012): 10.353

Posició dins l'àrea: Biochemistry & Molecular Biology = 14 de 290 (Q1)
Evolutionary Biology = 4 de 47 (Q1)
Genetics & Heredity = 8 de 161 (Q1)

El doctorand va participar activament en el disseny experimental i la realització de les anàlisis (cerca d'homologia, alineaments, inferència filogenètica i fonts d'error sistemàtic en filogenòmica), així com en la discussió, la interpretació i la redacció de l'article.

Article R3

Torruella G, de Mendoza A, Grau-Bové X, Donachie S, Pérez-Cordón G, Sitjà-Bobadilla A, Paley R, Manohar CS, Nichols K, Eme L, del Campo J, and Ruiz-Trillo I. **Phylotranscriptomics reveals ancient features in *Corallochytrium* and *Ministeria* (Holozoa, Opisthokonta)**.

A punt per enviar.

El doctorand va participar activament en la concepció, el disseny experimental i la redacció de l'article, del qual n'és co-autor. Va realitzar la totalitat de l'obtenció de dades (posada a punt de cultius, extracció d'ARN, muntatge *de novo*) i les anàlisis (cerca d'homologia, anàlisis filogenètiques i fonts d'error sistemàtic). Finalment, va contribuir significativament en la discussió i la interpretació dels resultats generats.

Iñaki Ruiz-Trillo Marta Riutort León

Results R1

The Evolutionary History of Lysine Biosynthesis Pathways Within Eukaryotes.

Resum article R1: L'història evolutiva de les vies de síntesi de la lisina als eucariotes.

La síntesi de la lisina segueix dues vies: la via del diaminopimelat (DAP) i la de l'alfa-aminodipat (AAA). La primera és present en eubacteris, plantes i algues, mentre que la segona via s'entenia quasi exclusiva de fongs. La troballa recent del gen aminoadipat reductasa (*AAR*), un dels gens principals de la via AAA, al protista marí *Corallochytrium limacisporum* fou per tant entesa com una sinapomorfia molecular entre ambdós, fongs i *C. limacisporum*. Per testar aquesta hipòtesi, es va realitzar una cerca més àmplia del gen *AAR* en eucariotes, i també es va analitzar la distribució del gen *LysA*, el gen principal de la via DAP. Demostrem que la història evolutiva dels dos gens, *AAR* i *LysA*, és molt més complexa del que es creia anteriorment. Més encara, el gen *AAR* és present en diversos opistoconts unicel·lulars, per tant refutem la teoria que la seva presència a *C. limacisporum* sigui indicatiu d'una sinapomorfia molecular entre *C. limacisporum* i fongs. El gen *AAR* sembla ser exclusiu d'excavats i uniconts, mentre que el gen *LysA* està present en diversos tàxons no directament emparentats dins de cadascun dels principals llinatges eucariotes, el que indica un paper per a diversos esdeveniments de transferència lateral de gens. Les nostres dades impliquen que el coanoflagel·lat *Monosiga Brevicollis* i el "coanozou" *Capsaspora owczarzaki* adquiriren les còpies de *LysA* d'un ancestre proteobacterià. En conjunt, aquestes observacions representen una nova evidència del paper que juga la transferència horitzontal de gens (sobretot en vies metabòliques) en la història evolutiva dels eucariotes.

J Mol Evol (2009) 69:240–248
DOI 10.1007/s00239-009-9266-x

The Evolutionary History of Lysine Biosynthesis Pathways Within Eukaryotes

Guifré Torruella · Hiroshi Suga · Marta Riutort · Juli Peretó · Iñaki Ruiz-Trillo

Received: 13 March 2009 / Accepted: 15 July 2009 / Published online: 11 August 2009

Abstract Lysine biosynthesis occurs in two ways: the diaminopimelate (DAP) pathway and the α-aminoadipate (AAA) pathway. The former is present in eubacteria, plants, and algae, whereas the latter was understood to be almost exclusive to fungi. The recent finding of the α-aminoadipate reductase (*AAR*) gene, one of the core genes of the AAA pathway, in the marine protist *Corallochytrium limacisporum* was, therefore, believed to be a molecular synapomorphy of fungi and *C. limacisporum*. To test this hypothesis, we undertook a broader search for the *AAR* gene in eukaryotes, and also analyzed the distribution of the *lysA* gene, a core gene of the DAP pathway. We show that the evolutionary history of both genes, *AAR* and *lysA*, is much more complex than previously believed. Furthermore, the *AAR* gene is present in several unicellular opisthokonts, thus rebutting the theory that its presence is a molecular synapomorphy between *C. limacisporum* and fungi. *AAR* gene seems to be exclusive of Excavata and Unikonts, whereas the *lysA* gene is present in several unrelated taxa within all major eukaryotic lineages, indicating a role for several lateral gene transfer (LGT) events. Our data imply that the choanoflagellate *Monosiga brevicollis* and the "choanozoan" *Capsaspora owczarzaki* acquired their *lysA* copies from a proteobacterial ancestor. Overall, these observations represent new evidence that the role of LGT in the evolutionary history of eukaryotes may have been more significant than previously thought.

Keywords Lysine biosynthesis · Molecular evolution · *Corallochytrium* · Opisthokonts · *AAR* gene · *lysA* gene · Lateral gene transfer

Introduction

Lysine is one of the ten essential amino acids which humans and animals need to acquire through their diet (Hutton et al. 2007). It is also the only known amino acid to have different biosynthetic pathways (Velasco et al. 2002; Xu et al. 2006) (Fig. 1). One way to synthesize lysine is via the diaminopimelate (DAP) pathway, which is understood to be present in bacteria, plants, and algae (Velasco et al. 2002; Hudson et al. 2005). The other is via the α-aminoadipate (AAA) pathway, present in fungi and *Euglena* (Miyazaki et al. 2004).

The AAA pathway was thought to be present only in fungi (and the excavate *Euglena*) (Vogel 1965; Xu et al. 2006), and was therefore proposed as a target for selective antifungal agents in order to control pathogenic yeasts and molds (Garrad and Bhattacharjee 1992; Xu et al. 2006). Although the AAA pathway was later also reported in

Electronic supplementary material The online version of this article (doi:10.1007/s00239-009-9266-x) contains supplementary material, which is available to authorized users.

G. Torruella · H. Suga · M. Riutort · I. Ruiz-Trillo
Departament de Genètica & Institut de Recerca en Biodiversitat (Irbio), Universitat de Barcelona, Barcelona, Spain

I. Ruiz-Trillo (✉)
Institució Catalana per a la Recerca i Estudis Avançats (ICREA), Parc Científic de Barcelona, Universitat de Barcelona, Edifici Hèlix; 1planta, Baldiri Reixac 15, 08028 Barcelona, Spain
e-mail: inaki.ruiz@icrea.es

J. Peretó
Departament de Bioquímica i Biologia Molecular, Institut Cavanilles de Biodiversitat i Biologia Evolutiva, Universitat de València, Valencia, Spain

J. Peretó
CIBER d'Epidemiologia i Salut Pública, Barcelona, Spain

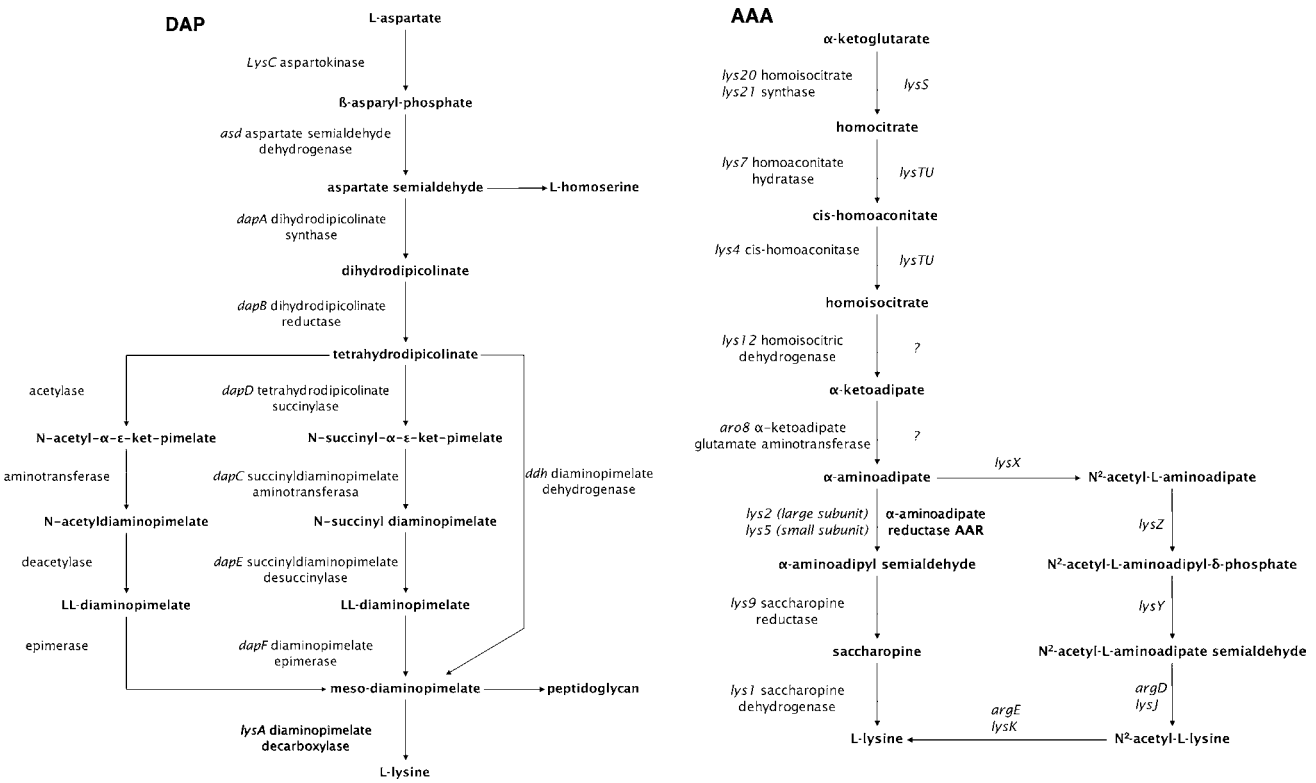


Fig. 1 A schematic representation of the AAA and DAP lysine biosynthesis pathways. Adapted from (Velasco et al. 2002). The two genes analyzed, *AAR* and *lysA*, are depicted

Archaea, such as *Thermoproteus, Sulfolobus,* and *Pyrococcus* (Velasco et al. 2002), as well as in the bacterium *Thermus thermophilus*, the pathway is not the same as the one in fungi (Xu et al. 2006). More recently, the α-aminoadipate reductase (*AAR*) gene, one of the core genes of the AAA pathway (see Fig. 1) was identified in the marine protist *Corallochytrium limacisporum* (Sumathi et al. 2006). *C. limacisporum* is a poorly understood unicellular opisthokont whose specific phylogenetic position within the "choanozoans" (i.e., unicellular opisthokonts closely related to Metazoa) remains unclear. Molecular phylogenies had shown *C. limacisporum* to be either the sister-group to Metazoa and fungi (Cavalier-Smith 2002), closely related to choanoflagellates (Mendoza et al. 2002; Cavalier-Smith and Chao 2003; Ruiz-Trillo et al. 2006), or as the sister-group to the ichthyosporean *Amoebidium parasiticum* (Steenkamp et al. 2006). Sumathi et al. (2006) proposed that the presence of the *AAR* gene in both *C. limacisporum* and fungi constituted a molecular synapomorphy and *C. limacisporum* was therefore more closely related to fungi than to Metazoa. However, the taxonomic sampling was rather low, and no data were presented regarding the presence of this gene in other unicellular opisthokonts.

Knowledge is also limited regarding the other lysine biosynthesis pathway, the DAP pathway (Velasco et al. 2002; Hudson et al. 2005). Although this pathway has been described in both bacteria and plants, it is not known whether it is present in other eukaryotic lineages. Once again, previous surveys of genes involved in the DAP pathway included a rather narrow taxon sampling of eukaryotes.

To test whether the *AAR* gene is a molecular synapomorphy of fungi and *C. limacisporum*, we performed a broader search for this gene in eukaryotes. In order to better understand the evolution of the DAP lysine biosynthesis pathway in eukaryotes, we also undertook a broader search for the *lysA* gene (see Fig. 1), a representative gene of the DAP pathway, in eukaryotes. Our data show that the evolutionary history of the *AAR* gene is actually much more intricate than previously thought. Furthermore, the presence of this gene is not a synapomorphy of fungi and *C. limacisporum*, and in fact it appears to be present in several other "choanozoans". Additionally, our search for the presence of *lysA* in eukaryotes shows that both *Monosiga brevicollis* and *Capsaspora owczarzaki* possess the *lysA* gene of the DAP pathway. Interestingly, our phylogenetic analysis suggests that the *lysA* gene of *M. brevicollis, C. owczarzaki* and *Naegleria gruberi* may have indeed been acquired from eubacteria through an interdomain lateral gene transfer (LGT) event.

Results

α-Aminoadipate Reductase (*AAR*) Gene

Screening of the *AAR* gene among eukaryotes showed that this gene is present in several other eukaryotes besides (the already reported) Fungi, *C. limacisporum* and *Euglena*. These include the excavate *Naegleria gruberi*, the amoebozoan *Acanthamoeba castellanii*, the ichthyosporean *Sphaeroforma arctica*, the choanoflagellate *M. brevicollis,* and the unicellular opisthokonts *Ministeria vibrans* and *C. owczarzaki* (see Figs. 2, 4; Table S1). However, the *AAR* gene seems to be absent from most other eukaryotes, being exclusive of Excavata and Unikonts. Phylogenetic analysis of the *AAR* gene showed all fungi grouping together and the "choanozoans" forming a well-resolved (more than 75% ML bootstrap support) clade (Fig. 2). *C. limacisporum* appears as the sister-group to the choanoflagellate *M. brevicollis,* although not strongly supported (91% bootstrap support in raxml). The *Monosiga* + *Corallochytrium* clade is the sister-group of a clade formed by *M. vibrans* and *C. owczarzaki*. Finally, the ichthyosporean *S. arctic*a appears as the first branch of the "choanozoan clade".

To gain further insights into the evolution of the AAA pathway in opisthokonts, we searched for the best-characterized genes from the AAA pathway in complete genome sequences from a broad sampling of eukaryotes. These include the only two choanozoan taxa for which the complete genome sequence is available to date (the choanoflagellate *M. brevicollis* (King et al. 2008), and *C. owczarzaki* whose genome project is under way (Ruiz-Trillo et al. 2007), as well as in deep-branching Metazoa for which there is a complete genome sequence, namely the poriferan *Amphimedon queenslandica* (Fahey et al. 2008), the cnidarian *Nematostella vectensis* (Putnam et al. 2007) and the placozoan *Trichoplax adhaerens* (Srivastava et al. 2008). *M. brevicollis* and *C. owczarzaki* seem to possess just a couple of representatives from all genes involved in

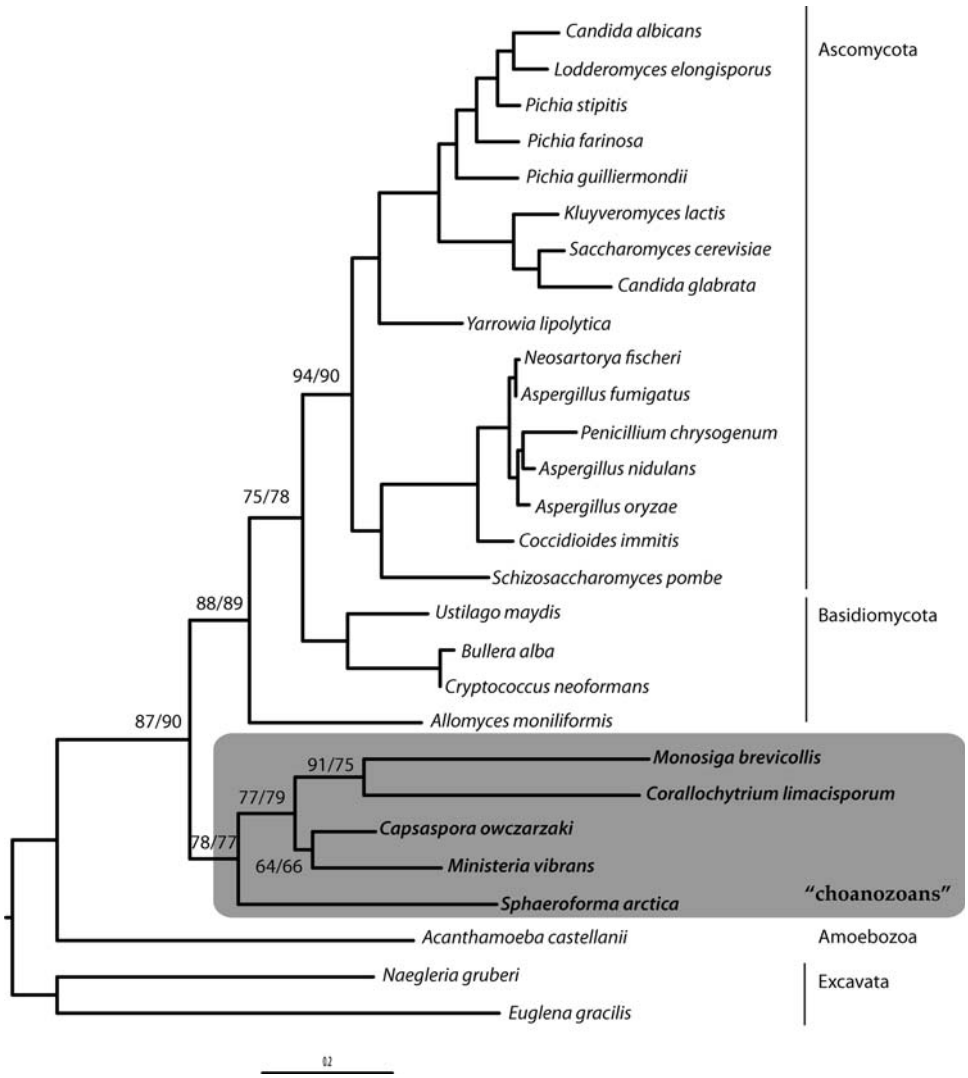


Fig. 2 Phylogeny of *AAR* gene sequences. The topology and branch lengths were obtained by maximum likelihood analysis performed in raxml. The tree is rooted by the excavates. Statistical support obtained by 100-bootstrap raxml replicates and 100-bootstrap phyml replicates is shown in relevant nodes

the AAR lysine biosynthesis pathway, in contrast to Fungi, *A. castellanii* or *N. gruberi* which have a more complete repertoire (see Table S2).

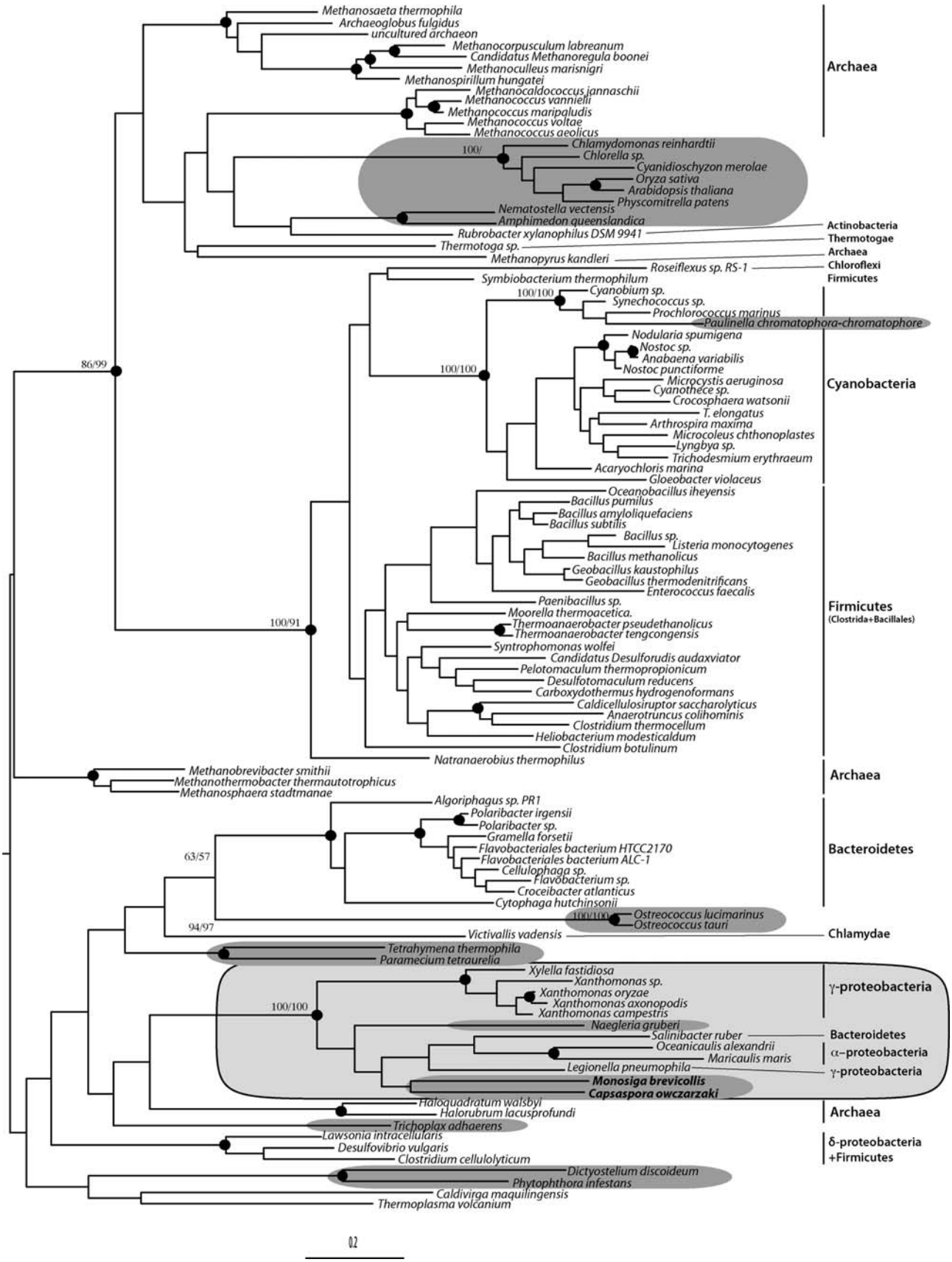
lysA Gene

Screening of the *lysA* gene from the DAP pathway in eukaryotes and in some prokaryotic lineages, showed that it is present not only in bacteria, plants, and algae, as previously reported (Velasco et al. 2002; Hudson et al. 2005), but also in some other unrelated eukaryotic lineages. In this regard, we also identified *lysA* sequences in the amoebozoan *Dictyostelium discoideum*, the stramenopile *Phytophthora infestans* and *Phaeodactylum tricornutum*, the excavate *N. gruberi*, the ciliates *Paramecium tetraurelia* and *Tetrahymena thermophyla,* the alveolate *Toxoplasma gondii,* the basal metazoans *A. queenslandica, T. adhaerens,* and *N. vectensis,* the choanoflagellate *M. brevicollis* and the "choanozoan" *C. owczarzaki* (see Figs. 3, 4; Tables S1, S2). The phylogenetic analysis of *lysA* shows several well-supported clades, and eukaryotes appear scattered throughout the tree. Interestingly, the *lys A* sequence from the chromatophore genome derived from *Paulinella chromatophora* clusters inside a well-supported (100% ML bootstrap support) cyanobacterial clade. Moreover, the choanoflagellate *M. brevicollis*, the "choanozoan" *C. owczarzaki*, and the excavate *N. gruberi* cluster inside a well-supported (100% ML bootstrap support) proteobacterial clade. Furthermore, all members of this clade share a unique 5′ extension in their *lysA* gene. As before, to gain further insights into the evolution of the DAP pathway we searched for all the well-characterized genes into the organisms for which the complete genome sequence is available. None of the taxa posses all the genes involved in the pathway, rather they have between 3 or 4 of those genes (see Table S2).

Discussion

Our data show that the evolutionary history of the *AAR* gene in eukaryotes is much more complex than previously thought. The *AAR* gene is not exclusive to Fungi, *Euglena* and *C. limacisporum,* as previously stated (Sumathi et al. 2006). Instead, the *AAR* gene is present in other eukaryotes, such as the excavate amoeba *N. gruberi*, the amoebozoan *A. castellanii*, and the "choanozoans" *S. arctica*, *M. brevicollis, C. owczarzaki* and *M. vibrans* (Figs. 2, 4). This contradicts the argument that *C. limacisporum* and Fungi may form a clade based on the shared presence of the *AAR* gene in both taxa (Sumathi et al. 2006). Our results agree with previous molecular phylogenies that show that *C. limacisporum* is a choanozoan (Ruiz-Trillo et al. 2006; Steenkamp et al. 2006; Cavalier-Smith and Chao 2003). What is clear, however, is that only a few eukaryotic lineages have the *AAR* gene. On the basis of existing knowledge, and even thought we have not sampled all bikont lineages (for example we have not searched into all *Plasmodium* strains) this gene seems to be present in just Excavata and the Unikonts, and then only in some of their lineages. The excavates *Euglena*, an euglenozoan, and *N. gruberi,* a heterolobosean, are indeed closely related and do share a common ancestor (Simpson et al. 2005); (Hampl et al. 2009), so it can hypothesized that the *AAR* gene is present in the Euglenozoa + Heterolobosea clade. Given the current data, there are two possible explanations for this patchy distribution. One is an eukaryote-to-eukaryote LGT event between an ancestral unikont and the common ancestor of the Euglenozoa and Heterolobosea, followed by differential loss in Metazoa, Amoebozoa and some fungi. However, the distribution of the *AAR* gene could just as easily be explained by its presence in the first eukaryotes, followed by differential loss in most of the lineages. Alternatively, *AAR* gene may have been present not in the first eukaryotes but in the common ancestor of Excavates and Unikonts and was consequently lost in several lineages within the excavates and the unikonts. However, this will mean that excavates and unikonts had a common ancestor, a theory that would have further implications on the root of the eukaryotes. Our data alone can not resolve among those explanations.

The presence of *lysA* from the DAP pathway in eukaryotes also shows a very patchy distribution (see Figs. 3, 4). The *lysA* gene is present in several distantly related eukaryotic lineages that in our tree do not appear in a single clade (Fig. 3). Again, this distribution may be explained by several LGT events. Although the topology of the *lysA* tree (Fig. 3) could be due to inconsistencies in the phylogenetic inference, there is not actual evidence of this (for example, extremely long-branches). Furthermore, there is increasing recognition of the prevalence of LGT in eukaryotes (see Keeling and Palmer 2008; Andersson 2005, but see Soria-Carrasco and Castresana 2008 for a cautionary tale). Even though loss of out-paralogs may as well account for the topology, we favor the hypothesis that the topology of the *lysA* gene tree is due to several LGT events. This is especially clear in the case of *M. brevicollis, N. gruberi,* and *C. owczarzaki*, where their homologs cluster strongly within proteobacteria. Moreover, the fact that the *lysA* of *M. brevicollis* does have some introns argues against a contamination artifact from bacterial samples. Indeed, there has already been a report of an LGT from proteobacteria to *M. brevicollis* (Foerstner et al. 2008), which suggest that LGT from proteobacteria to choanoflagellates may not be so uncommon. Furthermore, the *lysA* homologs of *N. gruberi, M. brevicollis*, *C. owczarzaki,* and proteobacteria share a


Methanosaeta thermophila
Archaeoglobus fulgidus
uncultured archaeon
Methanocorpusculum labreanum
Candidatus Methanoregula boonei
Methanoculleus marisnigri
Methanospirillum hungatei
Methanocaldococcus jannaschii
Methanococcus vannielii
Methanococcus maripaludis
Methanococcus voltae
Methanococcus aeolicus
Archaea
100/
Chlamydomonas reinhardtii
Chlorella sp.
Cyanidioschyzon merolae
Oryza sativa
Arabidopsis thaliana
Physcomitrella patens
Nematostella vectensis
Amphimedon queenslandica
Actinobacteria
Rubrobacter xylanophilus DSM 9941
Thermotogae
Thermotoga sp.
Archaea
Methanopyrus kandleri
Chloroflexi
Roseiflexus sp. RS-1
Firmicutes
Symbiobacterium thermophilum
100/100
Cyanobium sp.
Synechococcus sp.
Prochlorococcus marinus
Paulinella chromatophora-chromatophore
Nodularia spumigena
Nostoc sp.
Anabaena variabilis
Nostoc punctiforme
86/99
100/100
Microcystis aeruginosa
Cyanothece sp.
Crocosphaera watsonii
T. elongatus
Arthrospira maxima
Microcoleus chthonoplastes
Lyngbya sp.
Trichodesmium erythraeum
Acaryochloris marina
Gloeobacter violaceus
Cyanobacteria
Oceanobacillus iheyensis
Bacillus pumilus
Bacillus amyloliquefaciens
Bacillus subtilis
Bacillus sp.
Listeria monocytogenes
Bacillus methanolicus
Geobacillus kaustophilus
Geobacillus thermodenitrificans
Enterococcus faecalis
Paenibacillus sp.
Moorella thermoacetica.
100/91
Thermoanaerobacter pseudethanolicus
Thermoanaerobacter tengcongensis
Syntrophomonas wolfei
Candidatus Desulforudis audaxviator
Pelotomaculum thermopropionicum
Desulfotomaculum reducens
Carboxydothermus hydrogenoformans
Caldicellulosiruptor saccharolyticus
Anaerotruncus colihominis
Clostridium thermocellum
Heliobacterium modesticaldum
Clostridium botulinum
Natranaerobius thermophilus
Firmicutes
(Clostrida+Bacillales)
Methanobrevibacter smithii
Methanothermobacter thermautotrophicus
Methanosphaera stadtmanae
Archaea
Algoriphagus sp. PR1
Polaribacter irgensii
Polaribacter sp.
Gramella forsetii
Flavobacteriales bacterium HTCC2170
63/57
Flavobacteriales bacterium ALC-1
Cellulophaga sp.
Flavobacterium sp.
Croceibacter atlanticus
Cytophaga hutchinsonii
Bacteroidetes
100/100
Ostreococcus lucimarinus
Ostreococcus tauri
94/97
Victivallis vadensis
Chlamydae
Tetrahymena thermophila
Paramecium tetraurelia
Xylella fastidiosa
Xanthomonas sp.
Xanthomonas oryzae
Xanthomonas axonopodis
Xanthomonas campestris
γ-proteobacteria
100/100
Naegleria gruberi
Salinibacter ruber
Bacteroidetes
Oceanicaulis alexandrii
α–proteobacteria
Maricaulis maris
Legionella pneumophila
γ-proteobacteria
Monosiga brevicollis
Capsaspora owczarzaki
Haloquadratum walsbyi
Halorubrum lacusprofundi
Archaea
Trichoplax adhaerens
Lawsonia intracellularis
δ-proteobacteria
+Firmicutes
Desulfovibrio vulgaris
Clostridium cellulolyticum
Dictyostelium discoideum
Phytophthora infestans
Caldivirga maquilingensis
Thermoplasma volcanium
0.2

◀**Fig. 3** Phylogeny of the *lysA* gene sequences. The topology and branch lengths were obtained by maximum likelihood analysis performed in raxml. The tree is rooted using the midpoint rooting method. Statistical support obtained by 100-bootstrap raxml replicates and 100-bootstrap phyml replicates is shown in relevant nodes. *Black circles* show nodes with more than 85% raxml and phyml bootstrap support. Eukaryotic taxa are depicted with a *gray background*. The clade that comprises the *lysA* homologs that share a unique 5′ extension (see text) is represented with a *gray rectangular*

molecular synapomorphy, which is a 5′ extension of a round 500 amino acids. This extension is unique to these organisms, and it is not found in any other taxa. Similarly, the strongly supported sister-group relationship of *Dictyostelium* and *Phytophthora* may be due to a LGT event from an unknown donor, especially if it is taken into account that none of them possess any of the other enzymes from the DAP pathway in their genome. In fact, amoebozoans have been shown to have suffered several bacteria-to-eukaryote LGT events (Watkins and Gray 2006; Watkins and Gray 2008), and, interestingly, the genomic analysis of a *Phytophthora* species has revealed several proteins to be closely related to *Dictyostelium* (Tyler et al. 2006). Finally, the position of the chromatophore sequence of the filose thecamoeba *Paulinella chromatophora* among cyanobacteria, and especially related to *Prochlorococcus/Synechococcus*, is what one would expect since the chromotophores

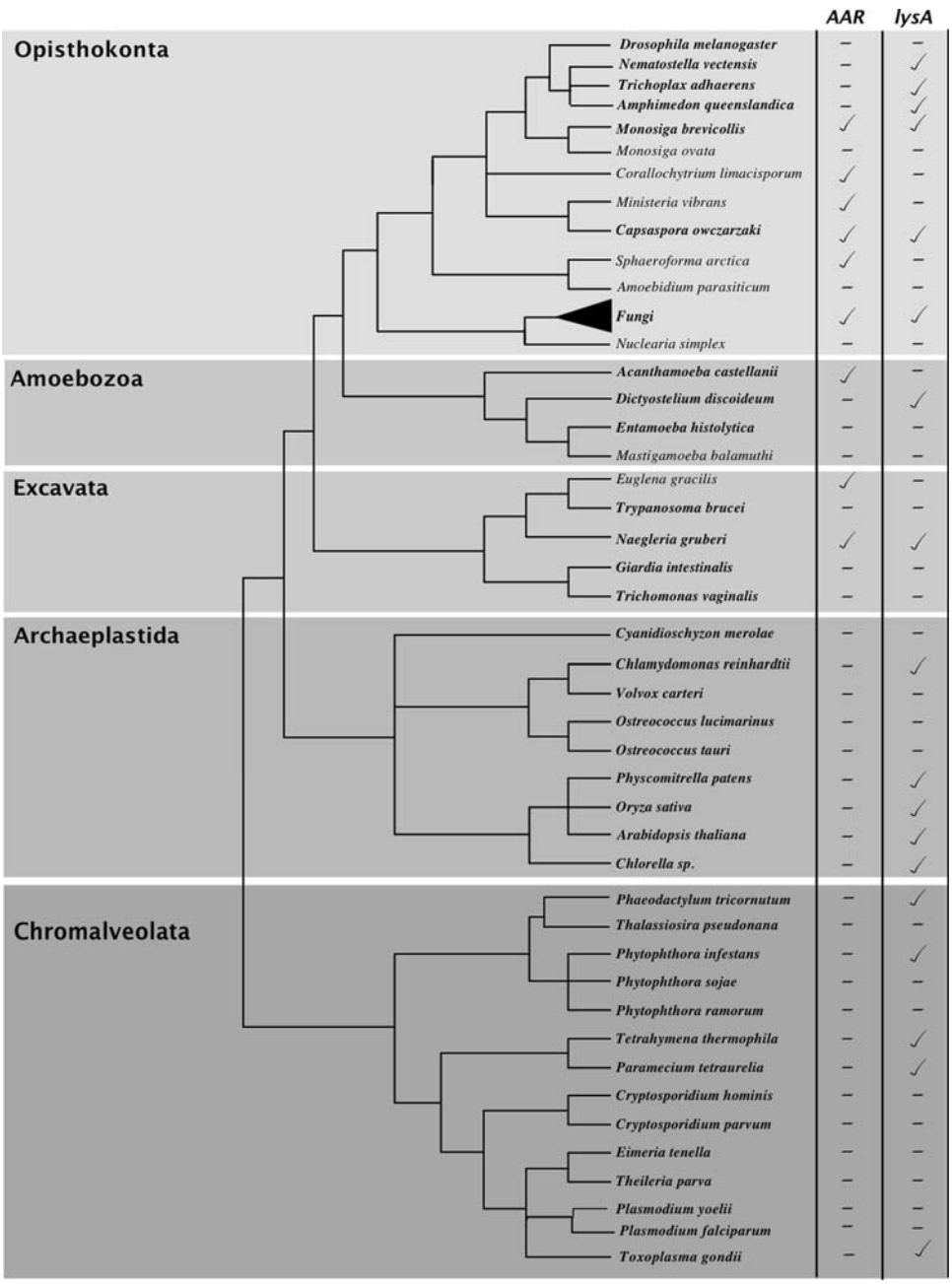


Fig. 4 A schematic representation of the presence of both *AAR* and *lysA* genes within the eukaryotes. The tree topology reflects the result of several recent phylogenomic studies (Ruiz-Trillo et al. 2008); (Shalchian-Tabrizi et al. 2008); (A Minge et al. 2008); (Burki et al. 2008); (Hampl et al. 2009). Taxa in *bold* indicate those for which the complete, or almost complete, genome sequence is available

of *Paulinella*, most likely come from a recent primary endosymbiosis from cyanobacteria of the *Prochlorococcus* type (Nowack et al. 2008); furthermore the biosynthesis of lysine seems to be one of the metabolic pathways apparently retained in the plastid of *P. chromatophora* (Nowack et al. 2008). This would represent a canonical case of metabolic acquisition by means of a symbiogenetic event. There are indeed many well-documented cases of essential amino acid biosynthesis acquisition through symbiotic associations, e.g., insects associated with bacteria during the last 200 million years (Zientz et al. 2004); (Moya et al. 2008).

Both *M. brevicollis* and *C. owczarzaki* possess the *AAR* gene and the *lysA* gene, which are, respectively, key genes of the AAA and DAP lysine biosynthesis pathways. This may suggest that choanozoans are able to synthesize lysine, from at least one of the pathways. However, the presence of those genes does not guarantee that the organism synthesizes lysine, since those genes may as well be used in other functional tasks. In this regard, it is worth noting that the *lysA* gene has been shown to play a role in the bacterial cell wall biosynthesis, it being the main precursor of peptidoglycans (Cirillo et al. 1994). However, the role played by the *lysA* or *AAR* gene in choanozoans goes beyond the scope of this manuscript.

The fact that metazoan relatives such as *M. brevicollis* and *C. owczarzaki* have both the *AAR* and the *lysA* genes seems to indicate that the common ancestor of both choanoflagellates and Metazoa was able to synthesize lysine, and that biosynthesis of lysine was specifically lost at the onset of (or within) the metazoan lineage. This was probably due to the fact that lysine was easily obtained from the environment. Metabolic pathways for essential amino acid biosyntheses are those with more enzymatic steps and, consequently, these amino acids are the most energetically expensive to synthesize (Berg et al. 2007). The corresponding long pathways are also more expensive to maintain in terms of energetic cost. Thus, when an organism adapts to a chemically rich environment, in which the more expensive amino acids are freely available, these anabolic pathways suffer a reductive evolution phenomenon and the living being becomes absolutely dependent on the supply of external amino acids. The process of gene loss is well-documented in symbiosis where some organisms permanently associate with different ones and, as a consequence, all redundant functions are affected by a syndrome of genome reduction (Moya et al. 2008). As stated above, further adaptations in metazoans include many cases of symbiotic association with bacteria retaining essential amino acid biosynthetic pathways. As a consequence, this syntrophic relationship and the re-acquisition of essential amino acid biosynthesis allow host adaptation to new ecological niches and the upgrading of deficient diets. Interestingly, basal metazoans (the placozoan *T. adhaerens*, the poriferan *A. queenslandica*, and the cnidarian *N. vectensis*) have several of the genes involved in the DAP lysine pathway, whereas none of them can be found in bilaterians. This may indicate that the pathway was lost in the transition from radial to bilaterian metazoans. Further work, including further sampling from metazoans, needs to be done to elucidate whether this is the case.

Finally, the present study provides an additional example of a LGT event involving eukaryotes. Although LGT in eukaryotes is not as relevant as in prokaryotes (Doolittle and Bapteste 2007; Bapteste and Boucher 2008; McInerney et al. 2008), it may nevertheless have played an important role in shaping the biochemical pathways of some eukaryotic lineages (see Keeling and Palmer 2008; Andersson 2005). Furthermore, our analysis demonstrates that assumptions made on the basis of relatively low taxon sampling (such as that of Sumathi et al. (2006) for the *AAR* gene), need to be re-evaluated when a broader taxon sampling becomes available. Similarly, our conclusions will certainly need to be tested again when the genome sequence data of different eukaryotic lineages becomes available.

Material and Methods

Database Searching

Both *AAR* and *lysA* sequences were obtained by performing blast searches (blastp and tblastn) against the Protein, Genome, and EST databases with the default parameters and an e-value threshold of e-05 at the NCBI (National Center for BiotechnologyInformation), TbestDB (http://amoebidia.bcm.umontreal.ca/pepdb/searches/welcome.php), and on-going genome projects at the JGI (Joint Genome Institute). The amino acid sequence of *AAR* from *Candida albicans* and the sequences of *lysA* from *Bacillus subtilis* were used as query. At a later stage, and to further check for out-paralogs, the sequence of the lysA from *Xanthomonas sp.* (which in our tree appears as far-related from that of *Bacillus subtilis*) was used as query to search into eukaryotic protein and genome databases. When searching against prokaryotic protein database, we did not incorporate all sequences, but specifically searched into some specific taxonomic groups to have at least some representatives. The taxonomic groups that we used in the NCBI organism search set were the following: Crenarchaeota, Euryarchaeota, Bacilli, Clostridia, Actinobacteria, Cyanobacteria, Chlorobi, Aquificae, Spirochaetes, Proteobacteria, and Chlamydae.

The sequences retrieved were then blasted against NCBI CDD (Conserved Domain Database). Only those that

retrieved an *AAR* or *lysA* homolog as a best hit in the reverse blast against NCBI were considered positives. Moreover, the affinity of the putative positives was further checked by gradually incorporating the sequences into a basic alignment of fungi *AAR* sequences and eubacteria *lysA* homologs. Only those sequences that could unambiguously be aligned were used in the phylogenetic analysis. The homolog of *AAR* from the unicellular opisthokont *Ministeria vibrans* was obtained by blasting against the Ministeria database of Dr. Kamran Shalchian-Tabrizi (University of Oslo).

Additionally we performed a search of all genes involved in both the AAA and DAP pathways that are well-characterized. Search was done using HMMER3.0a2 (Eddy 1998) and with PFAM profiles or profiles made from CDD alignments, against six-frame translations of several complete genome sequences of main representatives from a broad sampling of eukaryotic lineages (see Table S2 for a complete list and results). When genes were multi-domain, HMMER was performed in the most significant protein domain. E-value threshold of HMMER was e-08. All hits were then reverse-blasted against SWISSPROT, and only if the Blast search gave back the gene of interest (at a threshold value of e-08), the HMMER hits were considered positives. Results from *lysA* and *AAR* were carefully checked for false positives or negatives, going back to our blast results and by manually checking the sequences by incorporating them into the alignment.

Phylogenetic Analyses

Alignments of *AAR* and *lysA* genes were constructed using the Muscle (Edgar 2004) plug-in of the Geneious software (Biomatters Ltd, Auckland, New Zealand), before being manually inspected and edited. Only those positions that were unambiguously aligned were included in the final analyses, resulting in a total of 428 and 237 amino acid positions, respectively. The final protein alignments can be downloaded from the webpage http://www.multicellgenome.com.

Maximum likelihood phylogenetic trees were estimated by Raxml (Stamatakis 2006; Stamatakis et al. 2005) using a PROTCATWAG model of evolution and with a gamma distribution (8 categories) (WAG + Γ). We performed 50 inferences starting from distinct randomized maximum parsimony trees and adopted the one with the best likelihood value. Statistical support was obtained from 100-bootstrap replicates using the phyml program (Guindon and Gascuel 2003) following a WAG + Γ+ I model of evolution with 4 rate categories, and from 100-bootstrap replicates in raxml, using the PROTCATWAG model of evolution and with a gamma distribution (4 categories).

Acknowledgments Thanks to bioportal and to Andrew J. Roger for providing access to their computer resources. We thank Kamran Shalchian-Tabrizi for providing us with the sequence of *Ministeria vibrans* and Bernard Degnan for accession to the *Amphimedon* proteome data. We also thank Eric Bapteste, Tom Cavalier-Smith and Franz Lang for helpful insights. We also thank Roser Rotchés for her support. This work was supported by an ICREA contract and an ERC Starting Grant to IR-T, as well as a Grant BFU2006-06003 from MEC to JP.

References

Andersson JO (2005) Lateral gene transfer in eukaryotes. Cell Mol Life Sci 62:1182–1197

Bapteste E, Boucher Y (2008) Lateral gene transfer challenges principles of microbial systematics. Trends Microbiol 16:200–207

Berg JM, Tymoczko JL, Stryer L, Clarke ND (2007) Biochemistry, 6th edn. WH Freeman, New York

Burki F, Shalchian-Tabrizi K, Pawlowski J (2008) Phylogenomics reveals a new 'megagroup' including most photosynthetic eukaryotes. Biol Lett 4(4):366–369 Aug 23

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, Chao EE (2003) Phylogeny of choanozoa, apusozoa, and other protozoa and early eukaryote megaevolution. J Mol Evol 56:540–563

Cirillo JD, Weisbrod TR, Banerjee A, Bloom BR, Jacobs WRJ (1994) Genetic determination of the meso-diaminopimelate biosynthetic pathway of mycobacteria. J Bacteriol 176:4424–4429

Doolittle WF, Bapteste E (2007) Pattern pluralism and the Tree of Life hypothesis. Proc Natl Acad Sci USA 104:2043–2049

Eddy SR (1998) Profile hidden Markov models. Bioinformatics 14:755–763

Edgar RC (2004) MUSCLE: a multiple sequence alignment method with reduced time and space complexity. BMC Bioinformatics 5:113

Fahey B, Larroux C, Woodcroft BJ, Degnan BM (2008) Does the high gene density in the sponge NK homeobox gene cluster reflect limited regulatory capacity? Biol Bull 214:205–217

Foerstner KU, Doerks T, Muller J, Raes J, Bork P (2008) A nitrile hydratase in the eukaryote *Monosiga brevicollis*. PLoS ONE 3:e3976

Garrad RC, Bhattacharjee JK (1992) Lysine biosynthesis in selected pathogenic fungi: characterization of lysine auxotrophs and the cloned LYS1 gene of *Candida albicans*. J Bacteriol 174:7379–7384

Guindon S, Gascuel O (2003) A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. Syst Biol 52:696–704

Hampl V, Hug L, Leigh JW, Dacks JB, Lang BF, Simpson AG, Roger AJ (2009) Phylogenomic analyses support the monophyly of Excavata and resolve relationships among eukaryotic "supergroups". Proc Natl Acad Sci USA 106:3859–3864

Hudson AO, Bless C, Macedo P, Chatterjee SP, Singh BK, Gilvarg C, Leustek T (2005) Biosynthesis of lysine in plants: evidence for a variant of the known bacterial pathways. Biochim Biophys Acta 1721:27–36

Hutton CA, Perugini MA, Gerrard JA (2007) Inhibition of lysine biosynthesis: an evolving antibiotic strategy. Mol Biosyst 3:458–465

Keeling PJ, Palmer JD (2008) Horizontal gene transfer in eukaryotic evolution. Nat Rev Genet 9(8):605–618 Aug

King N, Westbrook MJ, Young SL, Kuo A, Abedin M, Chapman J, Fairclough S, Hellsten U, Isogai Y, Letunic I, Marr M, Pincus D, Putnam N, Rokas A, Wright KJ, Zuzow R, Dirks W, Good M, Goodstein D, Lemons D, Li W, Lyons JB, Morris A, Nichols S, Richter DJ, Salamov A, Sequencing JG, Bork P, Lim WA, Manning G, Miller WT, McGinnis W, Shapiro H, Tjian R, Grigoriev IV, Rokhsar D (2008) The genome of the choanoflagellate *Monosiga brevicollis* and the origin of metazoans. Nature 451:783–788

McInerney JO, Cotton JA, Pisani D (2008) The prokaryotic tree of life: past, present. and future? Trends Ecol Evol 23:276–281

Mendoza L, Taylor JW, Ajello L (2002) The class mesomycetozoea: a heterogeneous group of microorganisms at the animal-fungal boundary. Annu Rev Microbiol 56:315–344

Minge MA, Silberman JD, Orr RJ, Cavalier-Smith T, Shalchian-Tabrizi K, Burki F, Skjaeveland A, Jakobsen KS (2008) Evolutionary position of breviate amoebae and the primary eukaryote divergence. Proc Biol Sci 276:597–604

Miyazaki T, Miyazaki J, Yamane H, Nishiyama M (2004) alpha-Aminoadipate aminotransferase from an extremely thermophilic bacterium, *Thermus thermophilus*. Microbiology 150:2327–2334

Moya A, Pereto J, Gil R, Latorre A (2008) Learning how to live together: genomic insights into prokaryote-animal symbioses. Nat Rev Genet 9:218–229

Nowack EC, Melkonian M, Glockner G (2008) Chromatophore genome sequence of *Paulinella* sheds light on acquisition of photosynthesis by eukaryotes. Curr Biol 18:410–418

Putnam NH, Srivastava M, Hellsten U, Dirks B, Chapman J, Salamov A, Terry A, Shapiro H, Lindquist E, Kapitonov VV, Jurka J, Genikhovich G, Grigoriev IV, Lucas SM, Steele RE, Finnerty JR, Technau U, Martindale MQ, Rokhsar DS (2007) Sea anemone genome reveals ancestral eumetazoan gene repertoire and genomic organization. Science 317:86–94

Ruiz-Trillo I, Lane CE, Archibald JM, Roger AJ (2006) Insights into the evolutionary origin and genome architecture of the unicellular opisthokonts *Capsaspora owczarzaki* and *Sphaeroforma arctica*. J Eukaryot Microbiol 53(5):379–384

Ruiz-Trillo I, Burger G, Holland PW, King N, Lang BF, Roger AJ, Gray MW (2007) The origins of multicellularity: a multi-taxon genome initiative. Trends Genet 23:113–118

Ruiz-Trillo I, Roger AJ, Burger G, Gray MW, Lang BF (2008) A phylogenomic investigation into the origin of metazoa. Mol Biol Evol 25:664–672

Shalchian-Tabrizi K, Minge MA, Espelund M, Orr R, Ruden T, Jakobsen KS, Cavalier-Smith T (2008) Multigene phylogeny of choanozoa and the origin of animals. PLoS ONE 3:e2098

Simpson AG, Inagaki Y, Roger AJ (2005) Comprehensive Multi-Gene Phylogenies of Excavate Protists Reveal the Evolutionary Positions of 'Primitive' Eukaryotes. Mol Biol Evol 23:615–625

Soria-Carrasco V, Castresana J (2008) Estimation of phylogenetic inconsistencies in the three domains of life. Mol Biol Evol 25:2319–2329

Srivastava M, Begovic E, Chapman J, Putnam NH, Hellsten U, Kawashima T, Kuo A, Mitros T, Salamov A, Carpenter ML, Signorovitch AY, Moreno MA, Kamm K, Grimwood J, Schmutz J, Shapiro H, Grigoriev IV, Buss LW, Schierwater B, Dellaporta SL, Rokhsar DS (2008) The *Trichoplax* genome and the nature of placozoans. Nature 454:955–960

Stamatakis A (2006) RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. Bioinformatics 22:2688–2690

Stamatakis A, Ludwig T, Meier H (2005) RAxML-III: a fast program for maximum likelihood-based inference of large phylogenetic trees. Bioinformatics 21:456–463

Steenkamp ET, Wright J, Baldauf SL (2006) The protistan origins of animals and fungi. Mol Biol Evol 23:93–106

Sumathi JC, Raghukumar S, Kasbekar DP, Raghukumar C (2006) Molecular evidence of fungal signatures in the marine protist *Corallochytrium limacisporum* and its implications in the evolution of animals and fungi. Protist 157:363–376

Tyler BM, Tripathy S, Zhang X, Dehal P, Jiang RH, Aerts A, Arredondo FD, Baxter L, Bensasson D, Beynon JL, Chapman J, Damasceno CM, Dorrance AE, Dou D, Dickerman AW, Dubchak IL, Garbelotto M, Gijzen M, Gordon SG, Govers F, Grunwald NJ, Huang W, Ivors KL, Jones RW, Kamoun S, Krampis K, Lamour KH, Lee MK, McDonald WH, Medina M, Meijer HJ, Nordberg EK, Maclean DJ, Ospina-Giraldo MD, Morris PF, Phuntumart V, Putnam NH, Rash S, Rose JK, Sakihama Y, Salamov AA, Savidor A, Scheuring CF, Smith BM, Sobral BW, Terry A, Torto-Alalibo TA, Win J, Xu Z, Zhang H, Grigoriev IV, Rokhsar DS, Boore JL (2006) *Phytophthora* genome sequences uncover evolutionary origins and mechanisms of pathogenesis. Science 313:1261–1266

Velasco AM, Leguina JI, Lazcano A (2002) Molecular evolution of the lysine biosynthetic pathways. J Mol Evol 55:445–459

Vogel HJ (1965) Lysine biosynthesis and evolution. In: Bryson V (ed) Handbook of evolving genes and proteins, 5th edn. Academic Press, New York, pp 25–40

Watkins RF, Gray MW (2006) The frequency of eubacterium-to-eukaryote lateral gene transfers shows significant cross-taxa variation within amoebozoa. J Mol Evol 63:801–814

Watkins RF, Gray MW (2008) Sampling gene diversity across the supergroup Amoebozoa: large EST data sets from *Acanthamoeba castellanii*, *Hartmannella vermiformis, Physarum polycephalum, Hyperamoeba dachnaya and Hyperamoeba sp.*. Protist 159:269–281

Xu H, Andi B, Qian J, West AH, Cook PF (2006) The α-aminoadipate pathway for lysine biosynthesis in fungi. Cell Biochem and Biophys 46:43–64

Zientz E, Dandekar T, Gross R (2004) Metabolic interdependence of obligate intracellular bacteria and their insect hosts. Microbiol Mol Biol Rev 68:745–770

Results R2

Phylogenetic relationships within the Opisthokonta based on phylogenomic analyses of conserved single copy protein domains.

Resum article R2: Les relacions filogenètiques dins els opistoconts basades en anàlisis filogenòmiques de dominis proteics conservats de còpia única.

Moltes de les anàlisis filogenòmiques d'eucariotes publicades fins ara es basaven en alineaments de centenars o milers de gens. Freqüentment, en aquest tipus d'anàlisi, s'utilitzaven els models evolutius més realistes per reduir l'impacte d'errors sistemàtics. No obstant això, encara hi ha una controvèrsia notable sobre si les idiosincràsies en la dinàmica de les famílies gèniques (és a dir, duplicacions de gens i pèrdues) i les assignacions d'ortologia errònies sempre es tenen en compte correctament. En aquest treball, presentem una estratègia innovadora per superar els problemes d'assignació d'ortologia. En lloc d'identificar i eliminar gens amb problemes de paralogia, hem construït un conjunt de dades comprès únicament per dominis proteics conservats de còpia única que, a diferència de la majoria dels conjunts de dades filogenòmiques d'ús comú, han de ser menys propensos a problemes d'ortologia. Per avaluar la potència d'aquest enfocament, vam realitzar anàlisis de màxima versemblança i estadística bayesiana per inferir les relacions evolutives dins dels opistoconts (que inclou metazous, fongs i llinatges unicel·lulars emparentats). Vam utilitzar aquest enfocament per comprovar 1) si Filasterea i Ichthyosporea formen un clade monofilètic, 2) les interrelacions dels primers llinatges de metazous i 3) la relació entre els primers llinatges de fongs. També vam avaluar l'impacte d'alguns dels mètodes que es coneixen per minimitzar els errors sistemàtics, incloent la reducció de la distància entre els grups taxonòmics extern i intern, o l'ús del model evolutiu CAT. En general, les nostres anàlisis recolzen la hipòtesi dels Filozoa en què Ichthyosporea són el primer llinatge sorgit dels holozous seguit dels Filasterea, Choanoflagellata i Metazoa; i que els Blastocladiomycota apareixen com un llinatge separat dels Chytridiomycota. Aquests resultats representen proves independents d'hipòtesis filogenètiques plantejades en estudis anteriors, destacant la importància d'enfocar sofisticadament l'assignació d'ortologia i els mètodes d'anàlisis filogenòmiques.

Phylogenetic Relationships within the Opisthokonta Based on Phylogenomic Analyses of Conserved Single-Copy Protein Domains

Guifré Torruella,[1] Romain Derelle,†[,1] Jordi Paps,‡[,1] B. Franz Lang,[2] Andrew J. Roger,[3] Kamran Shalchian-Tabrizi,[4] and Iñaki Ruiz-Trillo*[,1,5]

[1]Departament de Genètica and Institut de Recerca en Biodiversitat (Irbio), Universitat de Barcelona, Barcelona, Spain

[2]Département de Biochimie, Robert-Cedergren Centre for Bioinformatics and Genomics, Université de Montréal, Montreal, Quebec, Canada

[3]Department of Biochemistry and Molecular Biology, Centre for Comparative Genomics and Evolutionary Bioinformatics, Faculty of Medicine, Dalhousie University, Halifax, Canada

[4]Microbial Evolution Research Group, Department of Biology, University of Oslo, Oslo, Norway

[5]Institució Catalana per a la Recerca i Estudis Avançats (ICREA), Barcelona, Spain

†Present address: Centre de regulació genòmica (CRG), Parc de recerca biomèdica de Barcelona (PRBB), Barcelona, Spain

‡Present address: Department of Zoology, University of Oxford, Oxford, United Kingdom

***Corresponding author:** E-mail: inaki.ruiz@icrea.es.

Associate editor: Manolo Gouy

Abstract

Many of the eukaryotic phylogenomic analyses published to date were based on alignments of hundreds to thousands of genes. Frequently, in such analyses, the most realistic evolutionary models currently available are often used to minimize the impact of systematic error. However, controversy remains over whether or not idiosyncratic gene family dynamics (i.e., gene duplications and losses) and incorrect orthology assignments are always appropriately taken into account. In this paper, we present an innovative strategy for overcoming orthology assignment problems. Rather than identifying and eliminating genes with paralogy problems, we have constructed a data set comprised exclusively of conserved single-copy protein domains that, unlike most of the commonly used phylogenomic data sets, should be less confounded by orthology miss-assignments. To evaluate the power of this approach, we performed maximum likelihood and Bayesian analyses to infer the evolutionary relationships within the opisthokonts (which includes Metazoa, Fungi, and related unicellular lineages). We used this approach to test 1) whether Filasterea and Ichthyosporea form a clade, 2) the interrelationships of early-branching metazoans, and 3) the relationships among early-branching fungi. We also assessed the impact of some methods that are known to minimize systematic error, including reducing the distance between the outgroup and ingroup taxa or using the CAT evolutionary model. Overall, our analyses support the Filozoa hypothesis in which Ichthyosporea are the first holozoan lineage to emerge followed by Filasterea, Choanoflagellata, and Metazoa. Blastocladiomycota appears as a lineage separate from Chytridiomycota, although this result is not strongly supported. These results represent independent tests of previous phylogenetic hypotheses, highlighting the importance of sophisticated approaches for orthology assignment in phylogenomic analyses.

Key words: *Capsaspora*, Filasterea, Filozoa, Holozoa, Ichthyosporea, multicellularity.

Introduction

A resolved phylogenetic tree that describes the relationships among organisms is the starting point for any research into the origins of fungi and multicellular animals (metazoans) from their unicellular ancestors (King 2004; Ruiz-Trillo et al. 2007; Rokas 2008; Shalchian-Tabrizi et al. 2008). Previous molecular studies have shown that Metazoa and Fungi share a common ancestor to the exclusion of plants, algae, and other eukaryotic lineages (Baldauf and Palmer 1993). The eukaryotic supergroup that comprises Fungi and Metazoa is known as the Opisthokonta (Cavalier-Smith 1987), and its monophyly has been confirmed by several molecular phylogenetic studies (Cavalier-Smith and Chao 1995; Lang et al. 2002; Medina et al. 2003; Ruiz-Trillo et al. 2004, 2006, 2008; Steenkamp et al. 2006); for a recent review, see Paps and Ruiz-Trillo (2010). Putative synapomorphies of Opisthokonta include a ~12 amino acid long insertion in the elongation factor 1-alpha gene (EF-1α) (Baldauf and Palmer 1993), a single posterior flagellum at least in one life-cycle stage (Patterson 1999; Cavalier-Smith and Chao 2003) and a haloarchaeal-type tyrosyl tRNA (Huang et al. 2005).

Molecular phylogenies have further shown that in unrooted trees of eukaryotes, Amoebozoa (Cavalier-Smith 1998; Adl et al. 2005) forms a clade with Opisthokonta (Minge et al. 2009). If the root falls outside of this clan (a

controversial point, see Roger and Simpson 2009), then the Amoebozoa plus Opisthokonta are a monophyletic group that has been named Unikonta (Cavalier-Smith 2002). In our analyses, Amoebozoa are assumed to be the nearest outgroup to the Opisthokonta. The Opisthokonta themselves are divided into two main lineages: the Holomycota (Liu et al. 2009), which was also named Nucletmyceta (Brown et al. 2009), containing Fungi and their unicellular relatives, such as the nucleariids and *Fonticula alba*; and the Holozoa (Lang et al. 2002; Brown et al. 2009), which includes Metazoa and their unicellular relatives, such as the Choanoflagellata, Filasterea (Shalchian-Tabrizi et al. 2008), and Ichthyosporea (Cavalier-Smith 1998; Mendoza et al. 2002). Several key groupings within the Holozoa and Holomycota remain contentious. Within Holomycota, the position of *F. alba* and nucleariids as the sister group to Fungi has been shown in multigene trees (Ruiz-Trillo et al. 2006; Steenkamp et al. 2006; Brown et al. 2009; Liu et al. 2009). However, further evidence is required to determine if the Blastocladiomycetes are part of the Chytridiomycota (Aleshin et al. 2007) or instead represent an independent phylum named the Blastocladiomycota (James et al. 2000, 2006; Brown et al. 2009; Liu et al. 2009).

Multigene and phylogenomic analyses corroborate the monophyly of Holozoa (Ruiz-Trillo et al. 2004, 2006, 2008; Steenkamp et al. 2006; Shalchian-Tabrizi et al. 2008; Brown et al. 2009; Liu et al. 2009). However, both the phylogenetic position of unicellular holozoans and the interrelationships among diploblastic metazoan lineages are still disputed. Filasterea (comprising *Capsaspora owczarzaki* and *Ministeria vibrans*) is either the sister group of Ichthyosporea (Ruiz-Trillo et al. 2008) or the sister group to a clade comprised of Metazoa and Choanoflagellata, which together form the Filozoa (Shalchian-Tabrizi et al. 2008). The solution to this issue is of major importance, as genomes of these lineages have been, or are being, sequenced (Ruiz-Trillo et al. 2007). Moreover, we still do not have a robust picture of the evolution and phylogeny of the earliest branching metazoan clades. Previous phylogenetic analyses have produced contradictory trees (Bridge et al. 1995; Cavalier-Smith and Chao 2003; Ender and Schierwater 2003; Dellaporta et al. 2006; Ruiz-Trillo et al. 2008; Srivastava et al. 2008; Schierwater et al. 2009; Philippe et al. 2009, 2011; Pick et al. 2010).

Phylogenomic analyses of the opisthokonts and the eukaryotes have mostly employed the supermatrix approach (for reviews, see Philippe et al. 2005; Jeffroy et al. 2006; de Queiroz and Gatesy 2007; Lartillot and Philippe 2008; Jenner and Littlewood 2008). The rationale is that large concatenated data sets eliminate the stochastic error caused by random noise (Philippe et al. 2005). However, different phylogenomic analyses can produce statistically supported incongruent trees (see, e.g., Ruiz-Trillo et al. 2008 vs. Shalchian-Tabrizi et al. 2008). These discrepancies may be due to differences in (or insufficient) taxon sampling or lack of realism in phylogenetic models that leads to systematic error. An alternative explanation is that orthologous genes may be incorrectly assigned and the use of paralogous genes confuses the phylogenetic signal and thus leads to incorrect trees. Two main approaches have been used to minimize paralogy problems. Some authors infer trees for each individual gene to check for possible orthology miss-assignments (Philippe et al. 2004, 2009; Brinkmann et al. 2005; Burki et al. 2007; Ruiz-Trillo et al. 2008; Shalchian-Tabrizi et al. 2008), whereas others have developed novel methods to improve the selection process of orthologous genes (Roure et al. 2007; Dunn et al. 2008; Hejnol et al. 2009). Here, we present a different strategy that may overcome orthology assignment problems: the identification and concatenation of conserved single-copy protein domains.

Proteins comprise distinct modular domains (Doolittle 1995) and often have complex evolutionary histories because of fusion, fission, shuffling, gain, and loss events (Caetano-Anolles G and Caetano-Anolles D 2003). Protein domains, in contrast, are discrete evolutionary units (Yang and Bourne 2009) that have been proposed to be a more stable "unit" of orthology than complete proteins (Gabaldon 2008), whose architecture between taxa may differ widely (Ponting and Russell 2002). As our principal goal is to generate a phylogeny that best represents the evolutionary histories of the taxa under investigation (Yang et al. 2005), single-copy protein domains (i.e., only one copy of the protein domain is found in each genome) may be more suitable and stable markers as their histories are less likely to be confused by recombination, fusion, fission, and duplication and loss (paralogy) dynamics. However, this approach needs complete or almost complete genome or transcriptome sequences from the taxa under examination.

We therefore constructed a phylogenomic data set comprising 93 conserved single-copy domains for the widest taxonomic sampling possible. Published and publicly available genome and expressed sequence tags (EST) data were used, as well as newly generated genome data from organisms sequenced by the UNICORN project (Ruiz-Trillo et al. 2007), such as the apusozoan *Thecamonas trahens*, the filasterean *C. owczarzaki*, the ichthyosporean *Sphaeroforma arctica*, the choanoflagellate *Salpingoeca rosetta*, and the fungi *Allomyces macrogynus* and *Spizellomyces punctatus*. In addition, we included EST data from another ongoing genome survey project of the filasterean *M. vibrans* (Shalchian-Tabrizi et al. 2008). Since this approach needs complete or almost complete genome sequences, some key taxa with few sequence data have not been included (i.e., *Nuclearia simplex* and *F. alba*). This new data set not only minimizes the problem of orthology assignment but also overlaps by less than 10% with the data used in other published phylogenomic investigations (Brinkmann et al. 2005; Burki et al. 2007; Ruiz-Trillo et al. 2008; Shalchian-Tabrizi et al. 2008; Philippe et al. 2009). Thus, it is an independent way to test previous phylogenetic hypotheses (supplementary tables S1 and S2 in supplementary file 1, Supplementary Material online).

To evaluate this novel data set, we performed maximum likelihood (ML) and Bayesian inference (BI) analyses to test previous hypotheses on 1) the relationship between

Filasterea and Ichthyosporea, 2) the phylogeny of early-branching metazoans, and 3) the phylogeny of early-branching fungi. We also assessed the impact of methods that are known to minimize phylogenetic systematic error (Delsuc et al. 2005). Several different data sets were considered each including the closest possible outgroup for the specific phylogenetic question under examination to check whether a closer outgroup affected our results (Philippe 1997; Lartillot and Philippe 2008). Finally, we also assessed the impact of recoding amino acids into functional categories (Woese et al. 1991; Rodriguez-Ezpeleta et al. 2007; Ruiz-Trillo et al. 2008), removing fast-evolving sites (Aguinaldo et al. 1997; Ruiz-Trillo et al. 1999), and excluding the taxa with the most missing data (Philippe 1997; Susko et al. 2005; Gatesy et al. 2007; de la Torre-Barcena et al. 2009).

Materials and Methods

EST and Genomic Data Sources

EST sequences from *Amoebidium parasiticum*, *Oscarella carmela*, *Oscarella lobularis*, *Blastocladiella emersonii*, *Acropora millepora*, *Acropora palmata*, *Monosiga ovata*, and *Clytia hemisphaerica* were extracted from the National Center for Biotechnology Information (NCBI). *Ministeria vibrans* ESTs were obtained in-house. Genome data were extracted from NCBI, the Joint Genome Institute (JGI) and the Broad Institute, as well as from the Baylor College of Medicine (BCM) for *Acanthamoeba castellanii*. *Amphimedon queenslandica* data were obtained from its genome database (www.metazome.net/amphimedon). *Capsaspora owczarzaki* and *S. punctatus* genome assemblies and annotations are available at the Broad Institute web site (http://www.broadinstitute.org/annotation/genome/multicellularity_project/MultiHome.html). In the case of *S. arctica*, *S. rosetta* (formerly known as *Proterospongia* sp.), *T. trahens* (formerly known as *Amastigomonas* sp., see Cavalier-Smith and Chao 2010), and *A. macrogynus*, the trace data were assembled in-house using the WGS assembler (http://sourceforge.net/apps/mediawiki/wgs-assembler/). The resulting contigs were translated using both Genomescan (Yeh et al. 2001) and Augustus (Stanke et al. 2006) to obtain independent databases of predicted protein sequences.

Selection of Single-Copy Protein Domains

For a schematic summary of sections 2 and 3, see figure 1. Taxonomic sampling started with several species for which a complete genome was available, including seven opisthokonts (*Mus musculus*, *Drosophila melanogaster*, *Nematostella vectensis*, *Monosiga brevicollis*, *Batrachochytrium dendrobatidis*, *Neurospora crassa*, *Schizosaccharomyces pombe*) and four other eukaryotes (*Dictyostelium discoideum*, *Ostreococcus lucimarinus*, *Leishmania infantum*, and *Plasmodium yoelii yoelii*). These taxa were also chosen in order to cover the largest possible diversity of eukaryotes and hence maximize the likelihood that the domains we selected were also single copy in other species. These 11 taxa were used as a "seed" sample to obtain an accurate data set of protein domains from PFAM (as available in

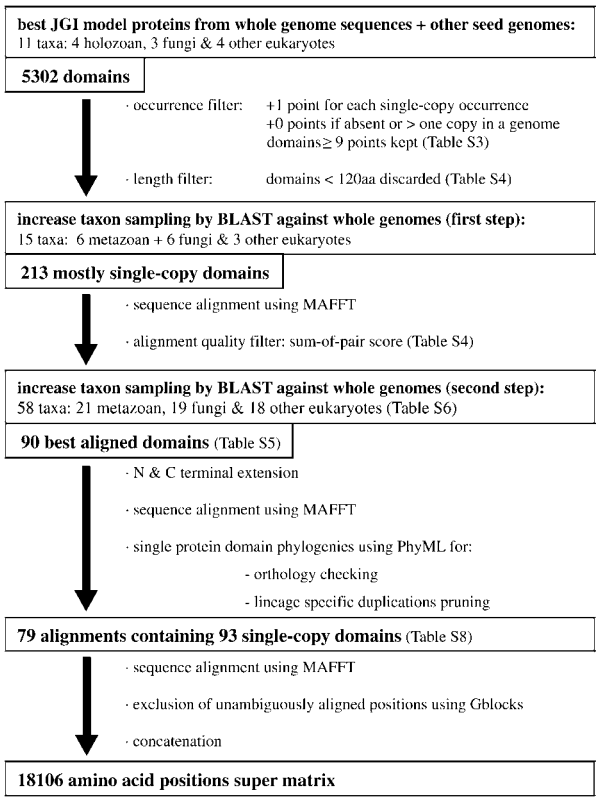


FIG. 1. Schematic pipeline of the supermatrix assembly as a summary of sections 2 and 3 of the Materials and Methods.

December 2008). The 5302 protein domains that were initially found were filtered using different methods. First, protein domains were indexed by the number of occurrences in the 11 seed taxa. Thus, for every domain, 1 point was given for each occurrence as a single copy in any of the 11 taxa, but no points were given if the protein domain was absent or present in more than one copy (see supplementary table S3 in supplementary file 1, Supplementary Material online). Furthermore, domains that were smaller than 120 amino acids were filtered using the "Domain model length" from the official PFAM database. To ensure that the domains were mostly single-copy within the widest taxonomic distribution, we only took into account domains with nine points or more, so that they were represented as single-copy in at least 9 of the 11 taxa. This resulted in a selection of 213 domains (supplementary table S4 in supplementary file 1, Supplementary Material online).

Once the "single-copy" domains had been chosen, the taxonomic sampling was enriched by representatives from the Opisthokonta. The new alignment included six Metazoa, six Fungi, and three other eukaryotes. We used BLAST to search for the corresponding proteins of the new taxa in the databases (for further details, see the supplementary file 1, Supplementary Material online). Each protein domain was then aligned using MAFFT (Katoh et al. 2002) with default parameters. The alignments were ranked using the "sum-of-pairs score," a metric that is used to predict the quality of an alignment (Ahola et al. 2008) (supplementary table S5 in supplementary file 1, Supplementary

Material online). To assure orthology, only the first 90 protein domains in this list were selected allowing us to select for the best alignments and thus the most conserved domains. Thereafter, to check whether some domains were usually associated within the same protein, PFAM analyses were carried out on the basis of *N. vectensis*, *M. brevicollis*, and *B. dendrobatidis* proteomes (see supplementary file 1, Supplementary Material online). This resulted in some individual alignments that included two different protein domains. By excluding these from consideration, the number of total alignments was reduced to 82. Finally, when possible, the alignments were further extended at the N or C termini to include conserved regions beyond the boundaries of the PFAM domain.

Data Curation

After the automatic steps described above, each of the 82 alignments was checked by eye. The final taxonomic sample included 58 taxa: 21 Metazoa, 19 Fungi, and 18 other eukaryotes (supplementary table S6 in supplementary file 1, Supplementary Material online). After the selection of the protein domain alignments, sequences for the additional taxa were obtained using tBLASTx for each of our protein domains against their proteomes. In addition, sequences that were not found in the proteomes were searched using tBLASTn against genomes.

Putative errors of protein prediction were detected in the alignments and corrected by performing new protein predictions based on genomic sequences using GENSCAN (Burge and Karlin 1997) and the ExPaSy Proteomics Server Translate tool (Gasteiger et al. 2003). At this point, three of the protein domain alignments were removed from the data set due to indels in some taxa that resulted in short conserved sequences. If possible, when a sequence was missing (or largely incomplete) for any given species, we added a sequence from a closely related species to the alignment (e.g., as in the case of the *Oscarella* and *Acropora* chimeric concatenated sequence; see supplementary file 1: table S7 for additional details, Supplementary Material online).

All alignments were again realigned with MAFFT, and single-gene phylogenetic trees were inferred using PhyML 3.0 (Guindon and Gascuel 2003) with the LG evolutionary model (Le and Gascuel 2008) and eight gamma rate categories. In the few instances in which species showed more than one copy of a marker due to lineage-specific duplications/diversifications (and they were grouping together in the single domain trees), only the sequence with the shortest branch to the outgroup was retained. Since these different copies represent recent lineage-specific domain duplications, they will not mislead the phylogenetic inference. Finally, after all single domain trees were checked for evidence of complex evolutionary histories, only one of the alignments was discarded for presenting a potential paralogy problem.

The remaining 78 protein domain alignments were realigned once more with MAFFT using default parameters. They were again checked by eye, trimmed by Gblocks 0.91 (Castresana 2000) with default parameters and finally concatenated. In the end, the supermatrix consisted of 78 independent alignments representing 93 conserved single-copy protein domains for 58 taxa. The matrix contained 18,106 amino acid positions (see supplementary table S8 in supplementary file 1, Supplementary Material online).

Saturation Test and Taxon Occupancy

The saturation plots were calculated for the original data sets 1 and 3, as well as the data sets derived from these by exclusion of the fastest evolving sites or recoding into functional categories. These consisted of scatter plots based on p-distances (pairwise observed distances) calculated with MEGA 4.1 (Kumar et al. 2008) plotted against the ML distances inferred from the Whelan and Goldman model of evolution (Whelan and Goldman 2001) including a gamma distribution of four rate categories (WAG + GAMMA) that were estimated with Tree-puzzle 5.2 (Schmidt et al. 2002). The resulting curve showed the degree of substitutional saturation qualitatively (see supplementary graphs S9 in supplementary file 2, Supplementary Material online).

Taxon occupancy (Hejnol et al. 2009; Sanderson et al. 2010), which is an approximate estimation of the degree of asymmetry within the matrix or the effect of the missing data, was calculated by summing the percentage of taxa present for each individual domain alignment and dividing the result by the total number of domain alignments (supplementary table S10 in supplementary file 2, Supplementary Material online).

The Outgroup Choice

The supermatrix contained a good sampling of Opisthokonta, Amoebozoa as well as some representatives of Stramenopiles and Viridiplantae (Cavalier-Smith 1998; Adl et al. 2005) to root the tree (Wheeler 1990; Huelsenbeck et al. 2002). In addition to the original alignment, three other supermatrices were created, each with reduced taxon sampling to test the impact of a closer outgroup on the topology recovered. The final data sets were as follows: 1) the original matrix with 58 taxa, Unikonta as the ingroup and Stramenopiles and Viridiplantae as the outgroup (Data set 1); 2) a data set with Amoebozoa and Apusozoa as the outgroup to Opisthokonta with a total of 52 taxa (Data set 2); 3) a data set with 36 taxa to specifically assess the branching order of Holozoa, in which a few representative Fungi were used as an outgroup (Data set 3); and 4) a 28 taxon data set to address the phylogeny of early divergent Fungi, using a few representative holozoan species as the outgroup (Data set 4).

Recoding into Functional Categories

To reduce the potential systematic error associated with compositional bias, a recoding approach was tested as described in Rodriguez-Ezpeleta et al. (2007) and Ruiz-Trillo et al. (2008). For each of the four data sets, the 20 amino acid characters were manually recoded into four chemically related categories to obtain the specific recoded data sets. The categories were based on the six Dayhoff groups (ASTGP, DNEQ, RKH, MVIL, FYW, and C [Dayhoff 1978]). However, there were two modifications to allow

for a general time reversible (GTR) matrix, as implemented in most programs: aromatic (FYW) and hydrophobic (MVIL) were combined in the same category, and the rare cysteines were coded as missing data.

Effects of Missing Data

To test the impact of missing data that resulted from the inclusion of EST data, two extra data sets were created for each of the four original ones. In the first extra data set, taxa with more than 45% of missing data were excluded (corresponding to the seven taxa whose sequences were derived from EST data). In the second data set, all taxa with more than 10% missing data were excluded (supplementary table S12 in supplementary file 2, Supplementary Material online).

Removal of Fast-Evolving Sites

The four data sets were analyzed by ML under the WAG + GAMMA model with the quartet puzzling algorithm (Strimmer and von Haeseler 1996) using Tree-puzzle 5.2 (Schmidt et al. 2002) to classify all sites into eight discrete categories of the gamma distribution. To reduce the computational burden and avoid problems with missing data, these analyses were performed by excluding taxa with more than 10% missing data (see the section above). Using the estimated site-rate categories from the quartet puzzling algorithm, two additional data sets were generated using the masking option on BioEdit (Hall 1999) to test the effect of progressive removal of fast-evolving sites. In one data set, positions classified as category 8 (the fastest-evolving sites) were excluded; in the other positions classified as categories 7 and 8 were excluded (see supplementary table S11 in supplementary file 3, Supplementary Material online).

Phylogenetic Analyses

ML trees were inferred using RAxML 7.2.8 parallel Pthreads-based version (Stamatakis 2006). The four main data sets were analyzed both with the WAG + GAMMA and the LG + GAMMA models. The data sets recoded into functional categories were analyzed using the generalized time reversible model (Lanave et al. 1984) (GTR + GAMMA). Runs started from four random trees, using subtree pruning and regrafting for branch swapping and the rapid hill-climbing algorithm (Stamatakis et al. 2007). ML analyses were also performed using 78 partitions for each of the genes in the main data sets. Rather than the fast-bootstrap feature of RAxML, the statistical support was assessed by performing 500 nonparametric bootstrap replicates with the same parameters as used in the initial tree search (see supplementary file 3, Supplementary Material online).

BI trees were constructed using the "Automatic stopping rule" with the default parameters as implemented in PhyloBayes 3.2 (Blanquart and Lartillot 2006, 2008) under the site-heterogeneous CAT evolutionary model (Lartillot and Philippe 2004). We chose the CAT-Poisson option, since it is a complex model that captures the diversity of site profiles in the data set without being overparameterized. To evaluate statistical support for splits, in addition to posterior probabilities, 100 jackknife pseudo replicates were obtained using SEQBOOT from the PHYLIP package (Felsenstein 1989) keeping 60% of positions. For each of these, a single Markov Chain Monte Carlo run was performed with 6,000 cycles using PhyloBayes. Thereafter, a burn-in of 3,000 cycles were applied to obtain a consensus tree for each chain (see main figs. 2–5).

Comparison of Topologies

Topologies estimated by ML can be compared using a number of statistical tests including the approximately unbiased (AU) test (Shimodaira 2002). Using the best ML tree from each data set, we prepared a series of specific node constraints with TreeView (Page 1996) to evaluate alternative topologies (e.g., we constrained Ichthyosporea to be the sister group to Filasterea and other Holozoa). The topologies were either based on previous studies or considered to be hypotheses of interest. First, RAxML was used to recalculate the optimal topology for each constraint and to calculate the site likelihoods for each tree. Then, the AU test was performed using CONSEL (Shimodaira and Hasegawa 2001) with the default scaling and replicate values (supplementary table S13 in supplementary file 2, Supplementary Material online).

Results

A New Data Set

To avoid orthology assignment problems, we constructed a novel supermatrix based on conserved single-copy domains (see the Materials and Methods and supplementary table S8 in supplementary file 1, Supplementary Material online). This data set represents an independent matrix for phylogenomic analysis since it shares less than 10% with other published eukaryotic phylogenomic data sets (see supplementary table S2 in supplementary file 1, Supplementary Material online). We performed ML and BI analyses to test previous hypotheses about the phylogenetic position of primary divergent lineages within Holozoa, Metazoa, and Fungi. A series of different data sets were analyzed to assess the effect of outgroup choice on the results for each phylogenetic question evaluated.

Data Set 1: Stramenopiles and Viridiplantae as Outgroup to Unikonta

The substitutional saturation plot for the original data set shows that the observed distances are, in general, proportional to the corrected distances without reaching a clear plateau, indicating that the data set does not display significant saturation (supplementary graphs S9*a*, *c*, and *e* in supplementary file 2, Supplementary Material online). Moreover, the average taxon occupancy is 90.78% of taxa per alignment (supplementary table S10 in supplementary file 2, Supplementary Material online), which is high relative to most published phylogenomic analyses.

The BI tree inferred with the CAT-Poisson model recovers the Amoebozoa, Apusozoa, and Opisthokonta as a clade (the Unikonta). In this analysis, the apusozoan *T. trahens* clearly

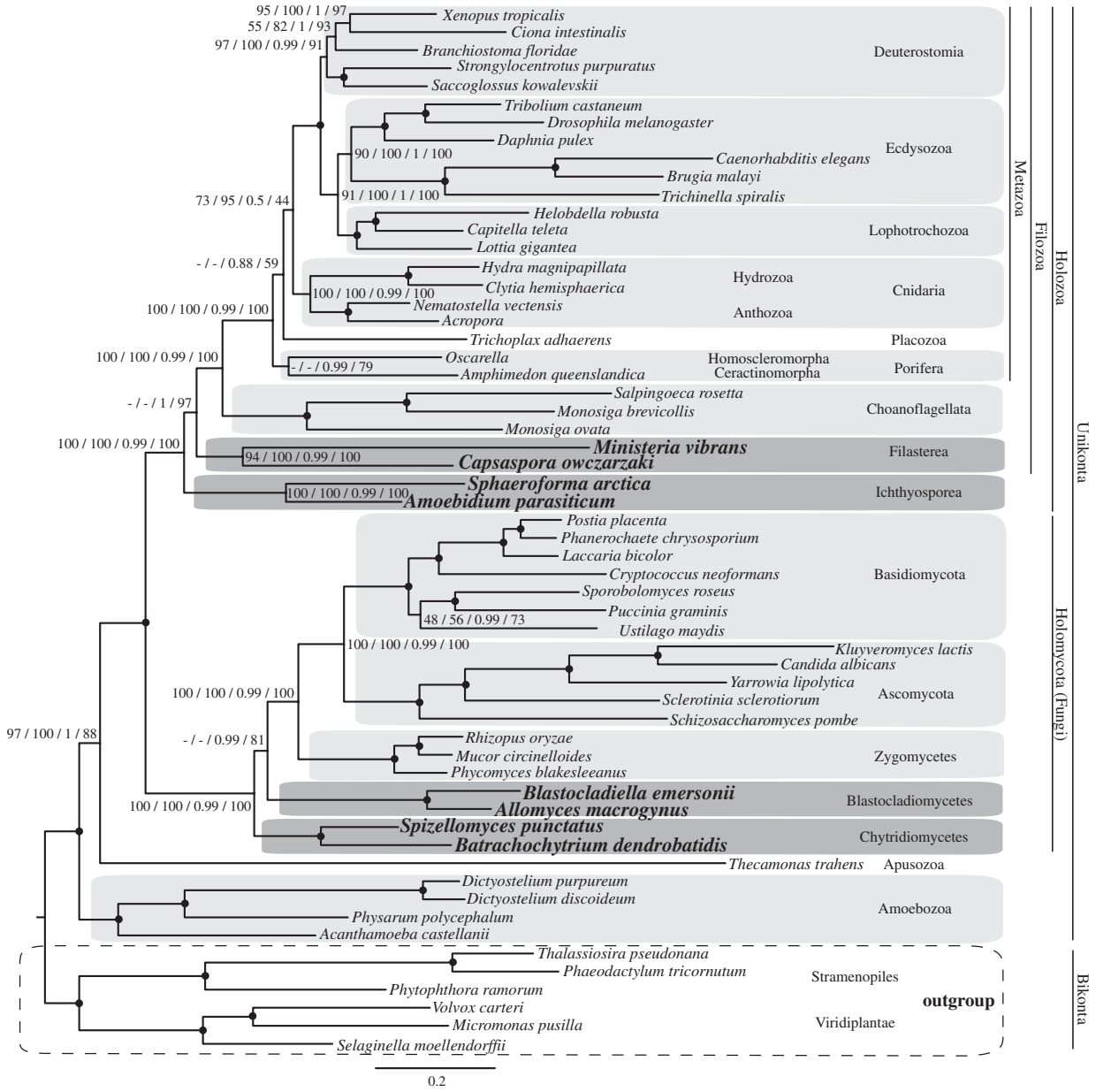


Fig. 2. Phylogenetic tree from BI for the data set 1. Stramenopiles and Viridiplantae are the outgroup. The numbers on branches indicate, from left to right, the following statistical support values: nonparametric bootstrap using WAG model (ML analysis with 500 replicates), nonparametric bootstrap using LG model (ML analysis with 500 replicates), posterior probability using CAT model (BI analysis), and delete 40% jackknife using CAT model (BI analysis). Nodes with maximum support values for all four analyzes (100 nonparametric bootstraps, 1 posterior probability and 100 jackknife runs) are depicted with black circles. Clades without support are marked with "-."

branches as the sister group to the opisthokonts (fig. 2), but given that there is only one representative for this group, a firm conclusion cannot be drawn at this time point with confidence. A discussion of phylogenomic results with broader species sampling and a mitochondrial protein set will be presented elsewhere.

Within Opisthokonta, Fungi appear as the sister group to a monophyletic Holozoa. Within Fungi, the Chytridiomycetes emerge as the earliest-branching fungal lineage, to the exclusion of the clade formed by the Blastocladiomycetes and the remaining fungi with a posterior probability support value of PP = 0.99 and a jackknife value of JV = 81%. Within the Holozoa, the Filasterea emerge as the sister group to a Choanoflagellata and Metazoa group, with Ichthyosporea as the first-branching holozoan lineage (PP = 1; JV = 97%) supporting the Filozoa hypothesis (Ruiz-Trillo et al. 2008; Shalchian-Tabrizi et al. 2008; Liu et al. 2009). Within the Metazoa, sponges are recovered as the sister group to all the remaining animal phyla (PP = 0.88; JV = 59%), whereas *Trichoplax adhaerens* emerges as an independent lineage that forms a sister group to Cnidaria plus Bilateria.

Remarkably, the ML analyses under the WAG + GAMMA or LG + GAMMA models recovered several topologies that

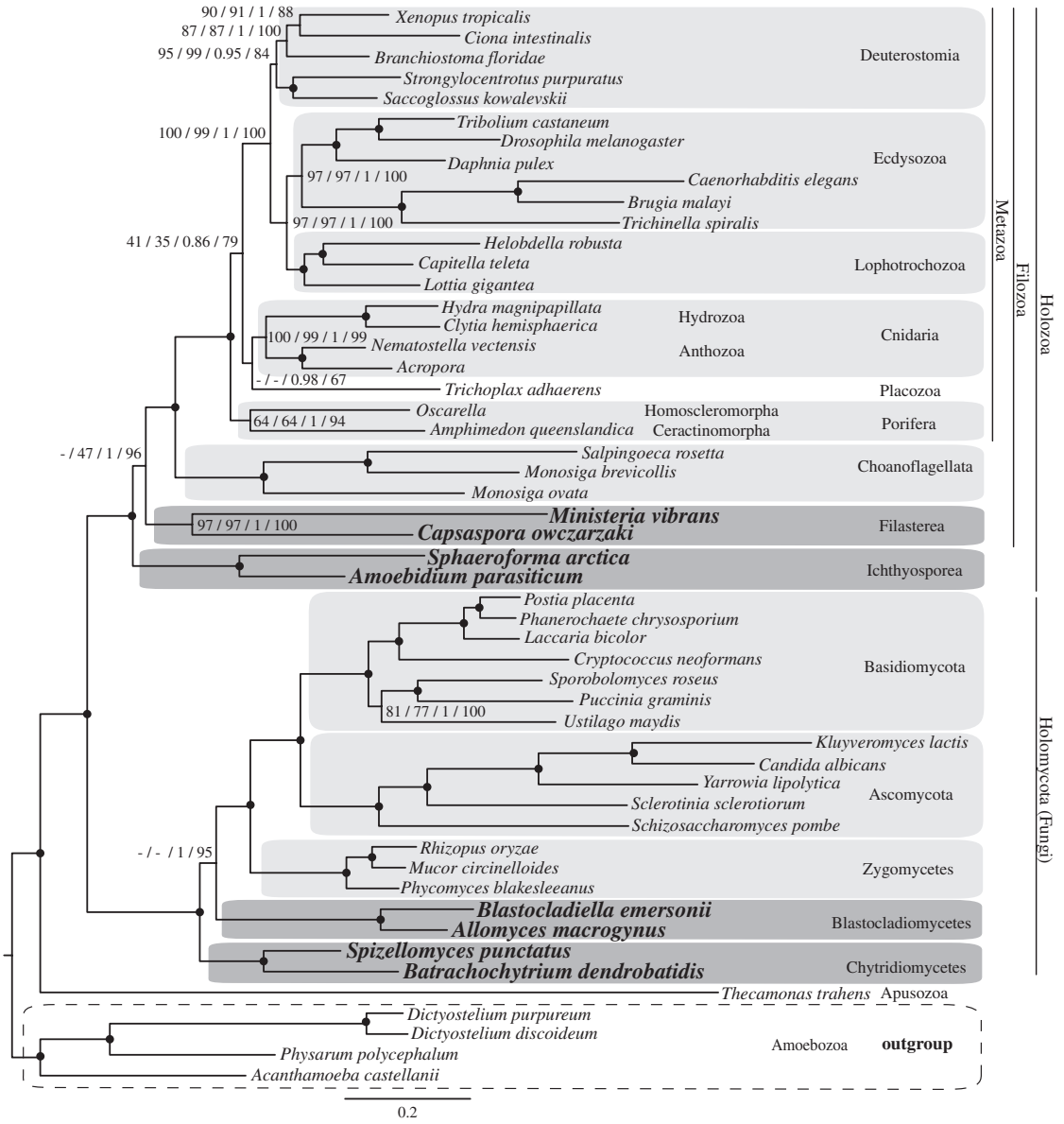


FIG. 3. Phylogenetic tree from BI for data set 2. Amoebozoa and Apusozoa are the outgroup. The numbers on branches indicate, from left to right, the following statistical support values: nonparametric bootstrap using WAG model (ML analysis with 500 replicates), nonparametric bootstrap using LG model (ML analysis with 500 replicates), posterior probability using CAT model (BI analysis), and delete 40% jackknife using CAT model (BI analysis). Nodes with maximum support values for all four analyzes (100 nonparametric bootstraps, 1 posterior probability, and 100 jackknife runs) are depicted with black circles. Clades without support are marked with "-."

differed from the BI analyses employing the CAT model (fig. 2, supplementary figs. S1 and S3, Supplementary Material online). Specifically, the ML analyses grouped Ichthyosporea with Filasterea together, although with low-to-moderate bootstrap values (BV = 65% using WAG and 42% using LG) and place the placozoan *T. adhaerens* within sponges, with no statistical support (WAG BV = 38% and LG BV = 56%). For the partitioned analyses, the removal of taxa with the most missing data or the fastest-evolving sites (i.e., those belonging in rate category 8) made little difference to the results (see supplementary figs. S2, S5, S6, and S7 in supplementary file 3, Supplementary Material online). However, when the sites from both rate categories 7 and 8 were removed (supplementary fig. S8, Supplementary Material online), Filasterea appeared as the sister group to Choanoflagellata and Metazoa (BV = 44%) to the exclusion of Ichthyosporea. Interestingly, the analysis with the amino acid positions recoded into functional categories had the same outcome as the removal of the two fastest-evolving site categories, although with low statistical support. The recoding analyses moderately supported the Filozoa hypothesis (BV = 63%) and the Chytridiomycetes as the earliest-branching fungal lineage (BV = 46%, see supplementary fig. S4 and table S11 in supplementary file 3, Supplementary Material online for an overall view of all analyses performed for all data sets).

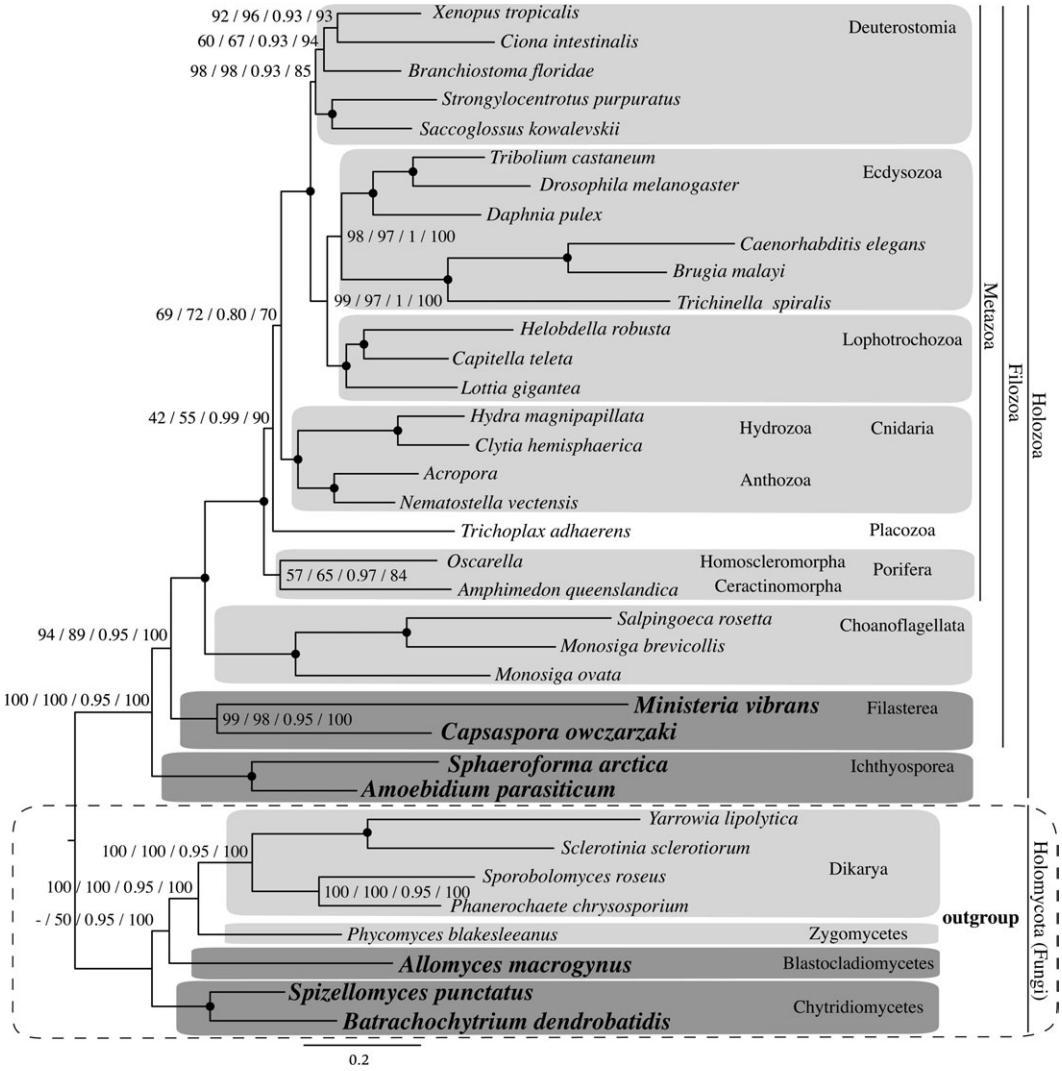


Fig. 4. Phylogenetic tree from BI for data set 3. Fungi is the outgroup. The numbers on branches indicate, from left to right, the following statistical support values: nonparametric bootstrap using WAG model (ML analysis with 500 replicates), nonparametric bootstrap using LG model (ML analysis with 500 replicates), posterior probability using CAT model (BI analysis), and delete 40% jackknife using CAT model (BI analysis). Nodes with maximum support values for all four analyzes (100 nonparametric bootstraps, 1 posterior probability and 100 jackknife runs) are depicted with black circles. Clades without support are marked with "-."

Data Set 2: Amoebozoa and Apusozoa as the Outgroup to Opisthokonta

The BI tree estimated with the CAT model for this data set, in which Amoebozoa and a single Apusozoa species were used as the outgroup to Opisthokonta, had a very similar topology to the data set 1 for the nodes of interest. The following were supported: the Filozoa hypothesis (PP = 1; JV = 96%), the Blastocladiomycota hypothesis (PP = 1; JV = 95%), and the branching of Placozoa as a sister group to Cnidaria (PP = 0.98; JV = 67%; see fig. 3).

The ML tree estimated using LG + GAMMA model also recovered the Filozoa (BV = 47%; see supplementary fig. S11, Supplementary Material online). In contrast, the WAG + GAMMA model analysis recovered the Ichthyosporea and the Filasterea grouping together at the base of Holozoa, with low statistical support (BV = 58%; see supplementary fig. S9, Supplementary Material online).

The ML analyses of data set 2 revealed a different topology for basal Fungi than that recovered by BI, with Chytridiomycetes as the sister group to Blastocladiomycetes (BV = 63% for WAG and 55% for LG). Finally, in contrast to the BI tree, *T. adhaerens* was located between Porifera and Cnidaria as an independent lineage but with no statistical support in the ML analyses (supplementary figs. S9 and S11, Supplementary Material online). Similar to the analyses of data set 1, partitioning the data set, the exclusion of the taxa with the most missing data or removal of the fastest-evolving sites did not have much impact on the results (supplementary table S11 and figs. S10, S13–S15, Supplementary Material online). Only the exclusion of sites with rate categories 7 and 8 (supplementary fig. S16, Supplementary Material online) and the recoding of amino acid positions into functional categories (supplementary fig. S12, Supplementary Material online) affected the

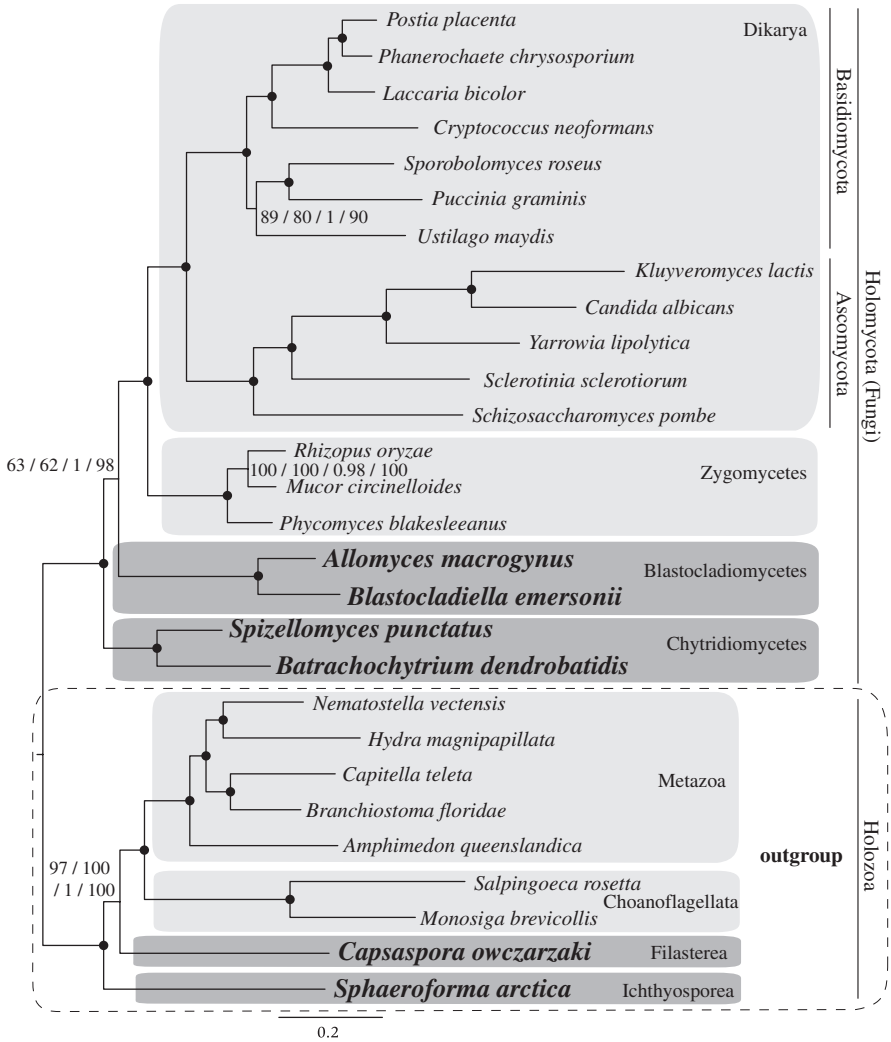


Fig. 5. Phylogenetic tree from BI for data set 4. Holozoa is the outgroup. The numbers on branches indicate, from left to right, the following statistical support values: nonparametric bootstrap using WAG model (ML analysis with 500 replicates), nonparametric bootstrap using LG model (ML analysis with 500 replicates), posterior probability using CAT model (BI analysis), and delete 40% jackknife using CAT model (BI analysis). Nodes with maximum support values for all four analyzes (100 nonparametric bootstraps, 1 posterior probability and 100 jackknife runs) are depicted with black circles. Clades without support are marked with "-."

topology recovered. In this case, Ichthyosporea appeared as the sister group to the rest of Holozoa, although with low statistical support (the Filozoa hypothesis with BV = 76% for recoding and BV = 42% for removing the fastest positions), and Placozoa was recovered as a sister group to Cnidaria (BV = 76% in the recoding analysis).

Data Set 3: Fungi as an Outgroup to Holozoa

This data set was specifically designed to assess the branching order of the first holozoan lineages. Like data set 1, the saturation analysis of data set 3 showed even less substitutional saturation in the amino acid composition between species (see supplementary graphs 9*b*, *d*, and *f* in supplementary file 2, Supplementary Material online). The ML and BI analyzes both showed, with reasonable statistical support (fig. 4), Ichthyosporea as the first lineage to branch off the Holozoa followed by the Filozoa (BV = 94% for WAG; BV = 89% for LG; PP = 0.95; JV = 100%; see supplementary table S11 and figs. S17–S19 in supplementary file 3, Supplementary Material online). This topology is in contrast to most of the ML trees inferred from data sets 1 and 2 that had a more distantly related outgroup. Importantly, analyses of data set 3 also recovered Placozoa as a sister group to Eumetazoa (Cnidaria + Bilateria) in both the BI (PP = 0.99; JV = 90%) and the ML trees (BV = 42% for WAG; BV = 55% for LG; see fig. 4, supplementary figs. S17 and S19, Supplementary Material online). Similar to the other data sets, partitioned analyses had the same result (supplementary fig. S18, Supplementary Material online). However, the position of *T. adhaerens* changes to that of sister group to Cnidaria or to Porifera when the amino acids are recoded into functional categories or when taxa with missing data are excluded, as well as when the fastest-evolving sites are removed (see supplementary table S11 and figs. S20–S23, Supplementary Material online).

Data Set 4: Holozoa as an Outgroup to the Fungi

Data set 4 (fig. 5) was specifically designed to assess the phylogeny of early-branching Fungi. The BI tree shows Chytridiomycetes as the sister group to the remaining Fungi (PP = 1; JV = 98%), according to the BI trees inferred by data sets 1 and 2. The same topology was obtained on the ML trees with WAG and LG models, although with low statistical support (BV = 63% and 62%; supplementary figs. S25 and S27, Supplementary Material online). The statistical support for this topology increased when data were recoded into functional categories (BV = 87%; see supplementary fig. S28, Supplementary Material online) or when the fastest-evolving sites were excluded (both BV = 78%; see supplementary figs. S31 and S32, Supplementary Material online). Interestingly, the analyses of the data set that excluded the taxa with the most missing data (supplementary figs. S29 and S30, Supplementary Material online) did not estimate this topology but did show the grouping of Blastocladiomycetes and Chytridiomycetes (BV = 57% and 58%, respectively).

Comparison of Topologies

To test whether alternative topologies can be statistically rejected, the AU test (Shimodaira 2002) was used for each data set. The test did not significantly exclude most of the tested alternative topologies, except for Placozoa placed as the sister group to Bilateria, which was rejected for all the data sets (with P values < 0.05) and Blastocladiomycetes as the sister group to Chytridiomycetes and other Fungi, which was rejected for data sets 2 and 3 (supplementary table S13 in supplementary file 2, Supplementary Material online).

Discussion

A New Phylogenomic Data Set

Previous phylogenomic analyses of eukaryotes, the Opisthokonta, or the Metazoa, have inferred phylogenetic relationships that have since become widely accepted, such as Opisthokonta monophyly (Cavalier-Smith and Chao 1995; Lang et al. 2002; Medina et al. 2003; Ruiz-Trillo et al. 2004, 2006, 2008; Steenkamp et al. 2006; Shalchian-Tabrizi et al. 2008); the adjacency of Opisthokonta and Amoebozoa; or the sister group relationship between Choanoflagellata and Metazoa (Lang et al. 2002; Philippe et al. 2004; Ruiz-Trillo et al. 2008; Shalchian-Tabrizi et al. 2008). However, some specific relationships remain highly disputed and it was not clear whether this is a methodological problem or a consequence of rapid diversification having occurred in particular parts of the tree (i.e., the origin of Metazoa). Two possible causes of incongruence among published analyses are the use of data sets with limited taxon sampling and/or orthology miss-assignments. Here, we have tried to avoid both problems by using the widest taxon sampling to date for organisms with completed full genome or transcriptome sequences and using conserved single-copy domains as markers. Furthermore, as there is minimal overlap between the data set we assembled and those previously published (Philippe et al. 2004, 2009; Brinkmann et al. 2005; Burki et al. 2007; Ruiz-Trillo et al. 2008; Shalchian-Tabrizi et al. 2008), our analyses serve as an independent test of phylogenetic hypotheses. We used this data set to test hypotheses in three different parts of the opisthokont tree and to evaluate the effect of several methods that are assumed to minimize systematic errors, such as closing the outgroup, excluding the fastest-evolving positions, and recoding the amino acids into functional categories (Philippe 1997; Rodriguez-Ezpeleta et al. 2007; Ruiz-Trillo et al. 2008).

The Early-Branching Fungi

Our data show that the distance of outgroup sequences to the ingroup has a considerable impact on the topology of the Fungi, at least when using ML methods with standard protein models. Analysis of data set 4 (fig. 5), that uses only holozoan lineages as an outgroup, shows Chytridiomycetes as the sister group to the remaining Fungi, supporting previous multigene trees (James et al. 2000, 2006; Liu et al. 2009). In contrast, data sets 1 and 2, which have a distantly related outgroup, show Blastocladiomycetes as the sister group to Chytridiomycetes. Neither the recoding strategy nor excluding the fastest evolving sites showed strongly supported differences in recovered topologies. One possible explanation for this observation is that the phylogenetic signal for branches separating the early-branching Fungi is weak in our data set. Only by using the CAT mixture model and/or an adequate outgroup (less prone to suffer from substitutional saturation and compositional heterogeneity, respectively), a consensus relationship is recovered among most data sets analyzed, although with low statistical support. Unfortunately, the taxon sampling for Fungi in our data set remains very limited compared with other recent studies (James et al. 2006; Liu et al. 2009). In future, our results should be tested with wider fungal taxon sampling, including lineages closest to Fungi, such as the nucleariids and *F. alba*.

The Branching Order within Holozoa

There has been controversy over the branching order within Holozoa and, particularly, over whether Ichthyosporea is the sister group to Filasterea or to Filozoa. On balance, our data supports the Filozoa hypothesis (Ruiz-Trillo et al. 2008; Shalchian-Tabrizi et al. 2008; Liu et al. 2009) rather than a sister group relationship between Filasterea and Ichthyosporea (tentatively named in this manuscript "Filasporea") (Ruiz-Trillo et al. 2008; Liu et al. 2009).

As with Fungi, the influence of outgroup choice is significant, with the closest outgroup-rooted analyses supporting the Filozoa hypothesis most strongly. The ML analyses only recover the Filozoa grouping with data set 3 (supplementary figs. S17 and S19, Supplementary Material online), which is also recovered by the CAT mixture model under BI (fig. 4). Use of the CAT model also has a considerable impact on the estimated topology as it also retrieves the Filozoa topology even when more distantly related outgroups are used, as in data sets 1 and 2 (figs.

2 and 3). A possible explanation is that the CAT model deals better with substitutional saturation (although not with compositional heterogeneity [Nesnidal et al. 2010]) than other models, such as WAG (Lartillot and Philippe 2008). Interestingly, the topology of Holozoa was also affected by recoding the amino acids into functional categories or by removing the fastest-evolving positions (categories 8 and 7). Both measures allowed recovery of Filozoa in analyses of data sets 1 and 2 (supplementary figs. S4, S8, S12, and S16 in supplementary file 3, Supplementary Material online), probably because they reduced error from substitutional saturation (see supplementary graphs 9*a*–*f* in supplementary file 2, Supplementary Material online).

The finding that Ichthyosporea is the sister group to the remaining Holozoa has deeper evolutionary implications. The filasterean *C. owczarzaki* has recently been shown to have several genes that are involved in multicellularity. These genes were previously believed to be metazoan specific (and are not present in Choanoflagellata) and include genes for integrins, T-box proteins (including a Brachyury type), or Runx (Sebé-Pedrós, de Mendoza, et al. 2010; Sebé-Pedrós, Roger, et al. 2010; Sebé-Pedrós and Ruiz-Trillo 2010). Thus, if Ichthyosporea is indeed the earliest branching lineage within the Holozoa, the importance of obtaining genome data from ichthyosporeans increases (Ruiz-Trillo et al. 2007), since the elucidation of the genetic machinery of Ichthyosporea will be crucial to pinpoint the evolutionary origins of these cell-adhesion and intercellular "communication" genes.

The Early Branching Metazoans

Finally, we tested how our new data set behaves with regard to the phylogeny of early-branching metazoans. To avoid problems with missing data, we only included diploblast phyla from which at least one complete genome sequence is available (i.e., Porifera, Placozoa, and Cnidaria). Unfortunately, our data set does not include Ctenophora, which has recently been proposed to be the earliest-branching metazoan phyla (Dunn et al. 2008; Hejnol et al. 2009; but see Philippe et al. 2009, 2011). Although most of the analyzes fail to provide a strongly supported answer to the branching order of diploblastic metazoans, mainly due to the unstable position of *T. adhaerens*, the CAT model and a close outgroup provide a moderately well-supported topology (fig. 4). Although most of the trees (either with ML or BI estimation) show *T. adhaerens* as the sister group to Cnidaria (see fig 3 and supplementary table S11 for summary, Supplementary Material online), the ML and especially the BI analyses for data set 3 show Placozoa in an intermediate position between sponges and cnidarians, that is, as a sister group to the Eumetazoa (Cnidaria + Bilateria). Since other Holozoa interrelationships are well resolved with this data set, we suggest an emergence between sponges and cnidarians is the most likely position of *T. adhaerens*, although we cannot rule out a sister group relationship this organism and Cnidaria. In fact, this position as sister group to Eumetazoa could explain that the *T. adhaerens* genome seems to encode a gene repertoire in between what it is found in sponges and cnidarians for some gene families such as MAGUK, bHLH, or homeobox, although secondary gene loss can not be ruled out (Schierwater et al. 2008; Srivastava et al. 2008; de Mendoza et al. 2010; Sebé-Pedrós, de Mendoza, et al. 2010; Ryan et al. 2010). In any case, further phylogenomic analyses should be carried out with more sponges and with at least one representative of the phylum Ctenophora to test whether this phylogenetic position is stable.

Conclusions

Our data show that concatenated alignments of protein domains rather than genes or complete proteins are a useful alternative strategy for inferring phylogenies and testing previous hypotheses that minimizes potential orthology assignment issues. We found that the selected evolutionary model and the outgroup have a considerable impact on the recovery of deep phylogenetic relationships within the opisthokonts. Other methods that are known to minimize systematic errors, such as recoding the amino acid into functional categories, or excluding the fastest-evolving sites have less impact but still provide important information regarding the kinds and locations of conflicting signals within data sets. Interestingly, our data supports the group Filozoa (Shalchian-Tabrizi et al. 2008) whereby Ichthyosporea are the sister group of the remaining Holozoa and do not branch as the sister group to Filasterea. We also find that the Chytridiomycetes may be the sister group to the rest of Fungi (James et al. 2006; Liu et al. 2009), although our taxon sampling is very limited. Finally, we find some support for the placement of Placozoa as occupying an intermediate position between Porifera and Cnidaria (Srivastava et al. 2008) or as a sister group to Cnidaria (Cavalier-Smith and Chao 2003). Although more analyses are certainly needed, our new approach based on conserved single-copy protein domains has proved to be an invaluable independent data set to infer phylogenies. As new genome or complete transcriptome sequences become available, this new data set can be expanded and further tested.

Supplementary Material

Supplementary files 1–3 are available at *Molecular Biology and Evolution* online (http://www.mbe.oxfordjournals.org/)

Acknowledgments

The genome sequences of *C. owczarzaki*, *A. macrogynus*, *S. punctatus*, *S. rosetta*, and *T. trahens* are being determined by the Broad Institute of MIT/Harvard University under the auspices of the National Human Genome Research Institute (NHGRI) and within the UNICORN initiative. We thank JGI, BI, and BCM for making data publicly available and Dr Nicole King for allowing us to use unpublished data from *S. rosetta*. We also thank Dr Kim Worley and her colleagues in the Human Genome Sequencing Center of Baylor College of Medicine for allowing us to use *A. castellanii* genome sequence. The *A. castellanii* sequencing was

funded by grant U54 HG003273 from NHGRI, National Institutes of Health to Richard Gibbs. We thank Hiroshi Suga, Arnau Sebé-Pedrós, and Alexandre de Mendoza for helpful insights. We also thank Ares Rocañín for her personal support. K.S.-T. thanks University of Oslo and Research Council of Norway for research grants. A.J.R. acknowledges support from the Canada Research Chairs Program. This work was supported by an ICREA contract, an European Research Council Starting Grant (ERC-2007-StG-206883) and a grant (BFU2008-02839/BMC) from Ministerio de Ciencia e Innovación (MICINN) to I.R.-T. G.T.'s work was supported by a pre-graduate FI grant from Generalitat de Catalunya (AGAUR).

References

Adl SM, Simpson AG, Farmer MA, et al. (28 co-authors). 2005. The new higher level classification of eukaryotes with emphasis on the taxonomy of protists. *J Eukaryot Microbiol*. 52:399–451.

Aguinaldo AM, Turbeville JM, Linford LS, Rivera MC, Garey JR, Raff RA, Lake JA. 1997. Evidence for a clade of nematodes, arthropods and other moulting animals. *Nature* 387:489–493.

Ahola V, Aittokallio T, Vihinen M, Uusipaikka E. 2008. Model-based prediction of sequence alignment quality. *Bioinformatics* 24: 2165–2171.

Aleshin VV, Konstantinova AV, Mikhailov KV, Nikitin MA, Petrov NB. 2007. Do we need many genes for phylogenetic inference? *Biochemistry (Mosc)* 72:1313–1323.

Baldauf SL, Palmer JD. 1993. Animals and fungi are each other's closest relatives: congruent evidence from multiple proteins. *Proc Natl Acad Sci U S A*. 90:11558–11562.

Blanquart S, Lartillot N. 2006. A Bayesian compound stochastic process for modeling nonstationary and nonhomogeneous sequence evolution. *Mol Biol Evol*. 23:2058–2071.

Blanquart S, Lartillot N. 2008. A site- and time-heterogeneous model of amino acid replacement. *Mol Biol Evol*. 25:842–858.

Bridge D, Cunningham CW, DeSalle R, Buss LW. 1995. Class-level relationships in the phylum Cnidaria: molecular and morphological evidence. *Mol Biol Evol*. 12:679–689.

Brinkmann H, van der Giezen M, Zhou Y, Poncelin de Raucourt G, Philippe H. 2005. An empirical assessment of long-branch attraction artefacts in deep eukaryotic phylogenomics. *Syst Biol*. 54:743–757.

Brown MW, Spiegel FW, Silberman JD. 2009. Phylogeny of the "forgotten" cellular slime mold, Fonticula alba, reveals a key evolutionary branch within Opisthokonta. *Mol Biol Evol*. 26:2699–2709.

Burge C, Karlin S. 1997. Prediction of complete gene structures in human genomic DNA. *J Mol Biol*. 268:78–94.

Burki F, Shalchian-Tabrizi K, Minge M, Skjaeveland A, Nikolaev SI, Jakobsen KS, Pawlowski J. 2007. Phylogenomics reshuffles the eukaryotic supergroups. *PLoS One*. 2:e790.

Caetano-Anolles G, Caetano-Anolles D. 2003. An evolutionarily structured universe of protein architecture. *Genome Res*. 13:1563–1571.

Castresana J. 2000. Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Mol Biol Evol*. 17:540–552.

Cavalier-Smith T. 1987. The origin of fungi and pseudofungi. In: Rayner ADM, Brasier CM, Moore D, editors. Evolutionary biology of the fungi. Cambridge: Cambridge University Press. p. 339–353.

Cavalier-Smith T. 1998. A revised six-kingdom system of life. *Biol Rev Camb Philos Soc*. 73:203–266.

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, Chao EE. 1995. The opalozoan Apusomonas is related to the common ancestor of animals, fungi, and choanoflagellates. *Proc R Soc Lond B Biol Sci*. 261:1–6.

Cavalier-Smith T, Chao EE. 2003. Phylogeny of choanozoa, apusozoa, and other protozoa and early eukaryote megaevolution. *J Mol Evol*. 56:540–563.

Cavalier-Smith T, Chao EE. 2010. Phylogeny and Evolution of Apusomonadida (Protozoa: Apusozoa): New Genera and Species. Protist. 161:549–576.

Dayhoff MO. 1978. Atlas of protein sequence and structure. Washington (DC): National Biomedical Research Foundation.

de la Torre-Barcena JE, Kolokotronis SO, Lee EK, Stevenson DW, Brenner ED, Katari MS, Coruzzi GM, DeSalle R. 2009. The impact of outgroup choice and missing data on major seed plant phylogenetics using genome-wide EST data. *PLoS One*. 4:e5764.

de Mendoza A, Suga H, Ruiz-Trillo I. 2010. Evolution of the MAGUK protein gene family in premetazoan lineages. *BMC Evol Biol*. 10:93.

de Queiroz A, Gatesy J. 2007. The supermatrix approach to systematics. *Trends Ecol Evol*. 22:34–41.

Dellaporta SL, Xu A, Sagasser S, Jakob W, Moreno MA, Buss LW, Schierwater B. 2006. Mitochondrial genome of Trichoplax adhaerens supports placozoa as the basal lower metazoan phylum. *Proc Natl Acad Sci U S A*. 103:8751–8756.

Delsuc F, Brinkmann H, Philippe H. 2005. Phylogenomics and the reconstruction of the tree of life. *Nat Rev Genet*. 6:361–375.

Doolittle RF. 1995. The multiplicity of domains in proteins. *Annu Rev Biochem*. 64:287–314.

Dunn CW, Hejnol A, Matus DQ, et al. (18 co-authors). 2008. Broad phylogenomic sampling improves resolution of the animal tree of life. *Nature* 452:745–749.

Ender A, Schierwater B. 2003. Placozoa are not derived cnidarians: evidence from molecular morphology. *Mol Biol Evol*. 20:130–134.

Felsenstein J. 1989. PHYLIP—Phylogeny Inference Package (Version 3.2). *Cladistics* 5:164–166.

Gabaldon T. 2008. Large-scale assignment of orthology: back to phylogenetics? *Genome Biol*. 9:235.

Gasteiger E, Gattiker A, Hoogland C, Ivanyi I, Appel RD, Bairoch A. 2003. ExPASy: the proteomics server for in-depth protein knowledge and analysis. *Nucleic Acids Res*. 31:3784–3788.

Gatesy J, DeSalle R, Wahlberg N. 2007. How many genes should a systematist sample? Conflicting insights from a phylogenomic matrix characterized by replicated incongruence. *Syst Biol*. 56:355–363.

Guindon S, Gascuel O. 2003. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst Biol*. 52:696–704.

Hall TA. 1999. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symp Ser*. 41:95–98.

Hejnol A, Obst M, Stamatakis A, et al. (17 co-authors). 2009. Assessing the root of bilaterian animals with scalable phylogenomic methods. *Proc Biol Sci*. 276:4261–4270.

Huang J, Xu Y, Gogarten JP. 2005. The presence of a haloarchaeal type tyrosyl-tRNA synthetase marks the opisthokonts as monophyletic. *Mol Biol Evol*. 22:2142–2146.

Huelsenbeck JP, Bollback JP, Levine AM. 2002. Inferring the root of a phylogenetic tree. *Syst Biol*. 51:32–43.

James TY, Letcher PM, Longcore JE, Mozley-Standridge SE, Porter D, Powell MJ, Griffith GW, Vilgalys R. 2006. A molecular phylogeny of the flagellated fungi (Chytridiomycota) and description of a new phylum (Blastocladiomycota). *Mycologia*. 98:860–871.

James TY, Porter D, Leander CA, Vilgalys R, Longcore JE. 2000. Molecular phylogenetics of the Chytridiomycota supports the utility of ultrastructural data in chytrid systematics. *Can J Bot*. 78:336–350.

Jeffroy O, Brinkmann H, Delsuc F, Philippe H. 2006. Phylogenomics: the beginning of incongruence? *Trends Genet*. 22:225–231.

Jenner RA, Littlewood DT. 2008. Problematica old and new. *Philos Trans R Soc Lond B Biol Sci*. 363:1503–1512.

Katoh K, Misawa K, Kuma K, Miyata T. 2002. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res*. 30:3059–3066.

King N. 2004. The unicellular ancestry of animal development. *Dev Cell*. 7:313–325.

Kumar S, Nei M, Dudley J, Tamura K. 2008. MEGA: a biologist-centric software for evolutionary analysis of DNA and protein sequences. *Brief Bioinform*. 9:299–306.

Lanave C, Preparata G, Saccone C, Serio G. 1984. A new method for calculating evolutionary substitution rates. *J Mol Evol*. 20:86–93.

Lang BF, O'Kelly C, Nerad T, Gray MW, Burger G. 2002. The closest unicellular relatives of animals. *Curr Biol*. 12:1773–1778.

Lartillot N, Philippe H. 2004. A Bayesian mixture model for across-site heterogeneities in the amino-acid replacement process. *Mol Biol Evol*. 21:1095–1109.

Lartillot N, Philippe H. 2008. Improvement of molecular phylogenetic inference and the phylogeny of Bilateria. *Philos Trans R Soc Lond B Biol Sci*. 363:1463–1472.

Le SQ, Gascuel O. 2008. An improved general amino acid replacement matrix. *Mol Biol Evol*. 25:1307–1320.

Liu Y, Steenkamp ET, Brinkmann H, Forget L, Philippe H, Lang BF. 2009. Phylogenomic analyzes predict sistergroup relationship of nucleariids and fungi and paraphyly of zygomycetes with significant support. *BMC Evol Biol*. 9:272.

Medina M, Collins AG, Taylor JW, Valentine JW, Lipps JH, Amaral-Zettler L, Sogin ML. 2003. Phylogeny of Opisthokonta and the evolution of multicellularity and complexity in Fungi and Metazoa. *Int J Astrobiol*. 2:203–211.

Mendoza L, Taylor JW, Ajello L. 2002. The class mesomycetozoea: a heterogeneous group of microorganisms at the animal-fungal boundary. *Annu Rev Microbiol*. 56:315–344.

Minge MA, Silberman JD, Orr RJ, Cavalier-Smith T, Shalchian-Tabrizi K, Burki F, Skjaeveland A, Jakobsen KS. 2009. Evolutionary position of breviate amoebae and the primary eukaryote divergence. *Proc Biol Sci*. 276:597–604.

Nesnidal M, Helmkampf M, Bruchhaus I, Hausdorf B. 2010. Compositional heterogeneity and phylogenomic inference of metazoan relationships. *Mol Biol Evol*. 27:2095–2104.

Page RD. 1996. TreeView: an application to display phylogenetic trees on personal computers. *Comput Appl Biosci*. 12: 357–358.

Paps J, Ruiz-Trillo I. 2010. Animals and their unicellular ancestors. Encyclopedia of life sciences. Chichester (UK): John Wiley & Sons, Ltd p. 1–8.

Patterson DJ. 1999. The diversity of eukaryotes. *Am Nat*. 154:S96–S124.

Philippe H. 1997. Rodent monophyly: pitfalls of molecular phylogenies. *J Mol Evol*. 45:712–715.

Philippe H, Brinkmann H, Lavrov DV, Littlewood DT, Manuel M, Worheide G, Baurain D. 2011. Resolving difficult phylogenetic questions: why more sequences are not enough. PLoS Biol 9: e1000602.

Philippe H, Delsuc F, Brinkmann H, Lartillot N. 2005. Phylogenomics. *Annu Rev Ecol Evol Syst*. 36:541–562.

Philippe H, Derelle R, Lopez P, et al. (29 co-authors). 2009. Phylogenomics revives traditional views on deep animal relationships. *Curr Biol*. 19:706–712.

Philippe H, Snell EA, Bapteste E, Lopez P, Holland PW, Casane D. 2004. Phylogenomics of eukaryotes: impact of missing data on large alignments. *Mol Biol Evol*. 21:1740–1752.

Pick KS, Philippe H, Schreiber F, et al. (11 co-authors). 2010. Improved phylogenomic taxon sampling noticeably affects nonbilaterian relationships. *Mol Biol Evol*. 27:1983–1987.

Ponting CP, Russell RR. 2002. The natural history of protein domains. *Annu Rev Biophys Biomol Struct*. 31:45–71.

Rodriguez-Ezpeleta N, Brinkmann H, Roure B, Lartillot N, Lang BF, Philippe H. 2007. Detecting and overcoming systematic errors in genome-scale phylogenies. *Syst Biol*. 56:389–399.

Roger AJ, Simpson AG. 2009. Evolution: revisiting the root of the eukaryote tree. *Curr Biol*. 19:R165–R167.

Rokas A. 2008. The origins of multicellularity and the early history of the genetic toolkit for animal development. *Annu Rev Genet*. 42:235–251.

Roure B, Rodriguez-Ezpeleta N, Philippe H. 2007. SCaFoS: a tool for selection, concatenation and fusion of sequences for phylogenomics. *BMC Evol Biol*. 7(Suppl 1):S2.

Ruiz-Trillo I, Burger G, Holland PW, King N, Lang BF, Roger AJ, Gray MW. 2007. The origins of multicellularity: a multi-taxon genome initiative. *Trends Genet*. 23:113–118.

Ruiz-Trillo I, Inagaki Y, Davis LA, Sperstad S, Landfald B, Roger AJ. 2004. Capsaspora owczarzaki is an independent opisthokont lineage. *Curr Biol*. 14:R946–R947.

Ruiz-Trillo I, Lane CE, Archibald JM, Roger AJ. 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, Riutort M, Littlewood DT, Herniou EA, Baguna J. 1999. Acoel flatworms: earliest extant bilaterian Metazoans, not members of Platyhelminthes. *Science* 283:1919–1923.

Ruiz-Trillo I, Roger AJ, Burger G, Gray MW, Lang BF. 2008. A phylogenomic investigation into the origin of metazoa. *Mol Biol Evol*. 25:664–672.

Ryan JF, Pang K, Mullikin JC, Martindale MQ, Baxevanis AD. 2010. The homeodomain complement of the ctenophore Mnemiopsis leidyi suggests that Ctenophora and Porifera diverged prior to the ParaHoxozoa. *Evodevo* 1:9.

Sanderson MJ, McMahon MM, Steel M. 2010. Phylogenomics with incomplete taxon coverage: the limits to inference. *BMC Evol Biol*. 10:155.

Schierwater B, Eitel M, Jakob W, Osigus HJ, Hadrys H, Dellaporta SL, Kolokotronis SO, DeSalle R. 2009. Concatenated analysis sheds light on early metazoan evolution and fuels a modern "urmetazoon" hypothesis. *PLoS Biol*. 7:e20.

Schierwater B, Kamm K, Srivastava M, Rokhsar D, Rosengarten RD, Dellaporta SL. 2008. The early ANTP gene repertoire: insights from the placozoan genome. *PLoS One*. 3:e2457.

Schmidt HA, Strimmer K, Vingron M, von Haeseler A. 2002. TREE-PUZZLE: maximum likelihood phylogenetic analysis using quartets and parallel computing. *Bioinformatics* 18:502–504.

Sebé-Pedrós A, de Mendoza A, Lang BF, Degnan BM, Ruiz-Trillo I. 2010. Unexpected repertoire of metazoan transcription factors in the unicellular holozoan Capsaspora owczarzaki. *Mol Biol Evol*. 28:1241–1254.

Sebé-Pedrós A, Roger AJ, Lang FB, King N, Ruiz-Trillo I. 2010. Ancient origin of the integrin-mediated adhesion and signaling machinery. *Proc Natl Acad Sci U S A*. 107:10142–10147.

Sebé-Pedrós A, Ruiz-Trillo I. 2010. Integrin-mediated adhesion complex: cooption of signaling systems at the dawn of Metazoa. *Commun Integr Biol*. 3:475–477.

Shalchian-Tabrizi K, Minge MA, Espelund M, Orr R, Ruden T, Jakobsen KS, Cavalier-Smith T. 2008. Multigene phylogeny of choanozoa and the origin of animals. *PLoS One*. 3:e2098.

Shimodaira H. 2002. An approximately unbiased test of phylogenetic tree selection. *Syst Biol*. 51:492–508.

Shimodaira H, Hasegawa M. 2001. CONSEL: for assessing the confidence of phylogenetic tree selection. *Bioinformatics* 17:1246–1247.

Srivastava M, Begovic E, Chapman J, et al. (21 co-authors). 2008. The Trichoplax genome and the nature of placozoans. *Nature* 454:955–960.

Stamatakis A. 2006. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyzes with thousands of taxa and mixed models. *Bioinformatics* 22:2688–2690.

Stamatakis A, Auch AF, Meier-Kolthoff J, Göker M. 2007. AxPcoords & parallel AxParafit: statistical co-phylogenetic analyzes on thousands of taxa. *BMC Bioinformatics*. 8:405.

Stanke M, Keller O, Gunduz I, Hayes A, Waack S, Morgenstern B. 2006. AUGUSTUS: ab initio prediction of alternative transcripts. *Nucleic Acids Res*. 34:W435–W439.

Steenkamp ET, Wright J, Baldauf SL. 2006. The protistan origins of animals and fungi. *Mol Biol Evol*. 23:93–106.

Strimmer K, von Haeseler A. 1996. Quartet puzzling: a quartet maximum-likelihood method for reconstructing tree topologies. *Mol Biol Evol*. 13:964–969.

Susko E, Spencer M, Roger AJ. 2005. Biases in phylogenetic estimation can be caused by random sequence segments. *J Mol Evol*. 61:351–359.

Wheeler WC. 1990. Nucleic acid sequence phylogeny and random outgroups. *Cladistics* 6:363–367.

Whelan S, Goldman N. 2001. A general empirical model of protein evolution derived from multiple protein families using a maximum-likelihood approach. *Mol Biol Evol*. 18:691–699.

Woese CR, Achenbach L, Rouviere P, Mandelco L. 1991. Archaeal phylogeny: reexamination of the phylogenetic position of Archaeoglobus fulgidus in light of certain composition-induced artifacts. *Syst Appl Microbiol*. 14:364–371.

Yang S, Bourne PE. 2009. The evolutionary history of protein domains viewed by species phylogeny. *PLoS One*. 4:e8378.

Yang S, Doolittle RF, Bourne PE. 2005. Phylogeny determined by protein domain content. *Proc Natl Acad Sci U S A*. 102: 373–378.

Yeh RF, Lim LP, Burge CB. 2001. Computational inference of homologous gene structures in the human genome. *Genome Res*. 11:803–816.

Results R3

Phylotranscriptomics reveals ancient features in *Corallochytrium* and *Ministeria* (Holozoa, Opisthokonta)

Resum article 3: Filotranscriptòmica revela característiques ancestral a *Corallochytrium* i *Ministeria* (Holozoa, Opisthokonta)

Els opistoconts són el supergrup d'eucariotes que conté animals i fongs, però també diversos llinatges de protists incloent Coanoflagel·lats, Filastera, Opisthosporidia, Nucleariids, Ichthyosporea i l'enigmàtic *Corallochytrium limacisporum*. Establir les relacions filogenètiques entre aquests grups és fonamental per entendre les transicions evolutives que han succeït al grup, com l'origen de la multicel·lularitat. Per resoldre-ho vam realitzar filogenòmica amb dades de transcriptomes de representats de tots els llinatges protists, essent l'estudi amb un mostreig més extensiu fins ara en opistoconts. Hem seqüenciat els transcriptomes de dues soques de *C. limacisporum* aïllades de Hawaii i Índia, cinc Ichthyosporea, una *Nuclearia* sense caracteritzar i l'Ancyromonadida *Nutomonas longa* (possible grup germà dels opistoconts. El nostre estudi resol l'arbre filogenètic dels opistoconts i emplaça a *C. limacisporum* com a llinatge germà dels Ichthyosporea, tots junts representant el primer llinatge dels holozous. Així la transició de la fagotrofia a l'osmotrofia va passar dos cops als opistoconts, un als holozous i l'altre als holomicots. Per investigar més a fons aquesta hipòtesi, vam emprar la genòmica comparada per traçar l'historia evolutiva de dos caràcters clau en l'evolució dels fongs: el flagell i les quitina sintases. Les nostres dades mostren que els holozous *Ministeria vibrans* i *C. limacisporum* expressen un aparell flagel·lar abans inadvertit, amb un patró de reducció observat en fongs i altres eucariotes. També demostrem que l'avantpassat comú de tots els opistoconts tenia un repertori complex de quitina sintases que s'ha simplificat en la majoria de llinatges actuals i secundàriament diversificat en animals, fongs i els holozous osmòtrofs. En conjunt, el nostre estudi aporta un marc filogenètic per estudiar les transicions evolutives dels opistoconts emprant la genòmica comparada.

Phylotranscriptomics reveals ancient and convergent features in Corallochytrium and Ministeria (Holozoa, Opisthokonta)

Guifré Torruella,[1,2,f] Alex de Mendoza,[1,2,f] Xavier Grau-Bové,[1,2] Stuart Donachie,[3] Gregorio Pérez-Cordón,[4] Ariadna Sitjà-Bobadilla,[4] Richard Paley,[5] Cathrine S Manohar,[6] Krista Nichols,[7] Laura Eme,[8] Javier del Campo,[9] and Iñaki Ruiz-Trillo[1,2,10*]

[1]Institut de Biologia Evolutiva (UPF-CSIC), Passeig Maritim de la Barceloneta 37-49, 08003 Barcelona, Catalonia.
[2]Departament de Genètica, Universitat de Barcelona, Av. Diagonal, 645, 08028 Barcelona, Catalonia.
[3]Department of Microbiology, University of Hawai'I, Honolulu, HI, USA.
[4]Institute of Aquaculture Torre de la Sal, IATS-CSIC, 12595 Ribera de Cabanes s/n, Castelló.
[5]Centre for Environment Fisheries and Aquaculture Science, Weymouth Laboratory, Barrack Road, The Nothe, Weymouth, Dorset DT4 8UB, UK.
[6]CSIR-National Institute of Oceanography, Dona Paula, Goa 403 004, India; Ecology Group, University of Kaiserslautern, Kaiserslautern 67663, Germany.
[7]Department of Biological Sciences, Purdue University, West Lafayette, IN, 47907, USA.
[8]Centre for Comparative Genomics and Evolutionary Bioinformatics, Department of Biochemistry and Molecular Biology, Dalhousie University, Halifax B3H 4R2, Canada.
[9]University of British Columbia, Vancouver, BC, Canada.
[10]Institució Catalana per a la Recerca i Estudis Avançats (ICREA), Barcelona, Catalonia.

[f]Co-authorship [*]Correspondence: inaki.ruiz@ibe.upf-csic.es

Summary

Opisthokonta is a well known eukaryotic supergroup as it contains fungi and animals; but it also contains diverse protistan lineages including choanoflagellates, filastereans, nucleariids, opisthosporidians, ichthyosporeans and the enigmatic *Corallochytrium limacisporum* [1]. The phylogenetic relationships between those protistan lineages are fundamental to understand the evolutionary transitions that lead to the current diversity, including the transition to multicellularity in animals [2]. In order to resolve opisthokont phylogeny we took a phylogenomic approach gathering transcriptomic data from representatives of all protistan lineages; being this study the most extensive sampling to date. Here we include transcriptomic data from two strains of *C. limacisporum* isolated from Hawaii and India, five ichthyosporeans; an uncharacterized *Nuclearia* and the ancyromonad *Nutomonas longa*, a putative outgroup to Opisthokonta [3, 4]. Our phylogenetic analyses resolve the backbone of opisthokont phylogeny, and places *C. limacisporum* as sister group to the Ichthyosporea. This grouping indicates that the osmotrophic lifestyle evolved once in the Holozoa, mirroring fungi in Holomycota. To further investigate this hypothesis we undertook a comparative genomics approach to trace the evolution of two key characters in fungal evolution: the cilium and the chitin synthases. Our data shows that *Ministeria vibrans* and *C. limacisporum* have a flagellar apparatus previously unnoticed, and that the pattern of flagellum reduction observed resembles that in fungi and other eukaryotes [5]. We also show that the Last Opisthokont Common Ancestor had a complex toolkit of chitin synthases that has been simplified in extant lineages and that has specifically suffered secondary expansions in animals [6], fungi [7], but also in Ichthyosporea and *C. limacisporum*, blurring classic synapomorphies of fungi [8]. Overall, our study provides a phylogenetic framework to study the evolutionary transitions in Opisthokonta at a genomic scale.

Results and Discussion

Broadest taxonomic sampling of Opisthokonta places *Corallochytrium* sister group to Ichthyosporea

To properly infer the opisthokont phylogeny, we used the broadest taxon sampling to date using RNAseq. Previous attempts to solve opisthokont phylogeny balanced between species-rich datasets with poor deep node resolution based on small ribosomal subunit [9, 10] and multigene supermatrices that included few lineages [11–14]. In the study here presented we updated our previously published phylogenomic dataset [14] including representative species of all described protist lineages in Opisthokonta. Of special interest is *Corallochytrium limacisporum*, a spherical walled free-living saprotroph found in coral reefs [15]. Originally classified as a Thraustochytrid for its morphology, *C. limacisporum* has been unstably placed within opisthokonta in all molecular phylogenies to date because of the scarce molecular data available [4, 16–19]. To obtain full transcriptome data we isolated two *C. limacisporum* strains from coral reefs, one from India and the other from Hawaii (see **Experimental Procedures**), and established them as axenic cultures. We also sampled Ichthyosporean representatives, including *Creolimax fragrantissima* [20], *Pirum gemmata*, *Abeoforma whisleri* [21], two strains of *Ichthyophonus hoferi* [22] (all the former belonging to the Ichthyophonida clade) and *Sphaerothecum destruens*, a member of the Dermocystida, which is the deepest branching ichthyosporean lineage according to small ribosomal subunit phylogenies [23]. We also included the uncharacterized *Nuclearia* ATCC 50694, a filose amoeba sister group to fungi, and *Nutomonas longa* CCAP 1958/5 [3], a putative outgroup of Opisthokonta related to Apusozoa [4, 24]. In the dataset we also increased the data for *Ministeria vibrans* and *Amoebidium parasiticum* using newly sequenced RNAseq, as well as updating data for choanoflagellates, microsporidia and early branching fungi and animal phyla (see **Table S1** and **Supplemental Experimental Procedures**). Therefore, we generated transcriptomic data for 12 protistan species and placed them in an updated phylogenomic framework.

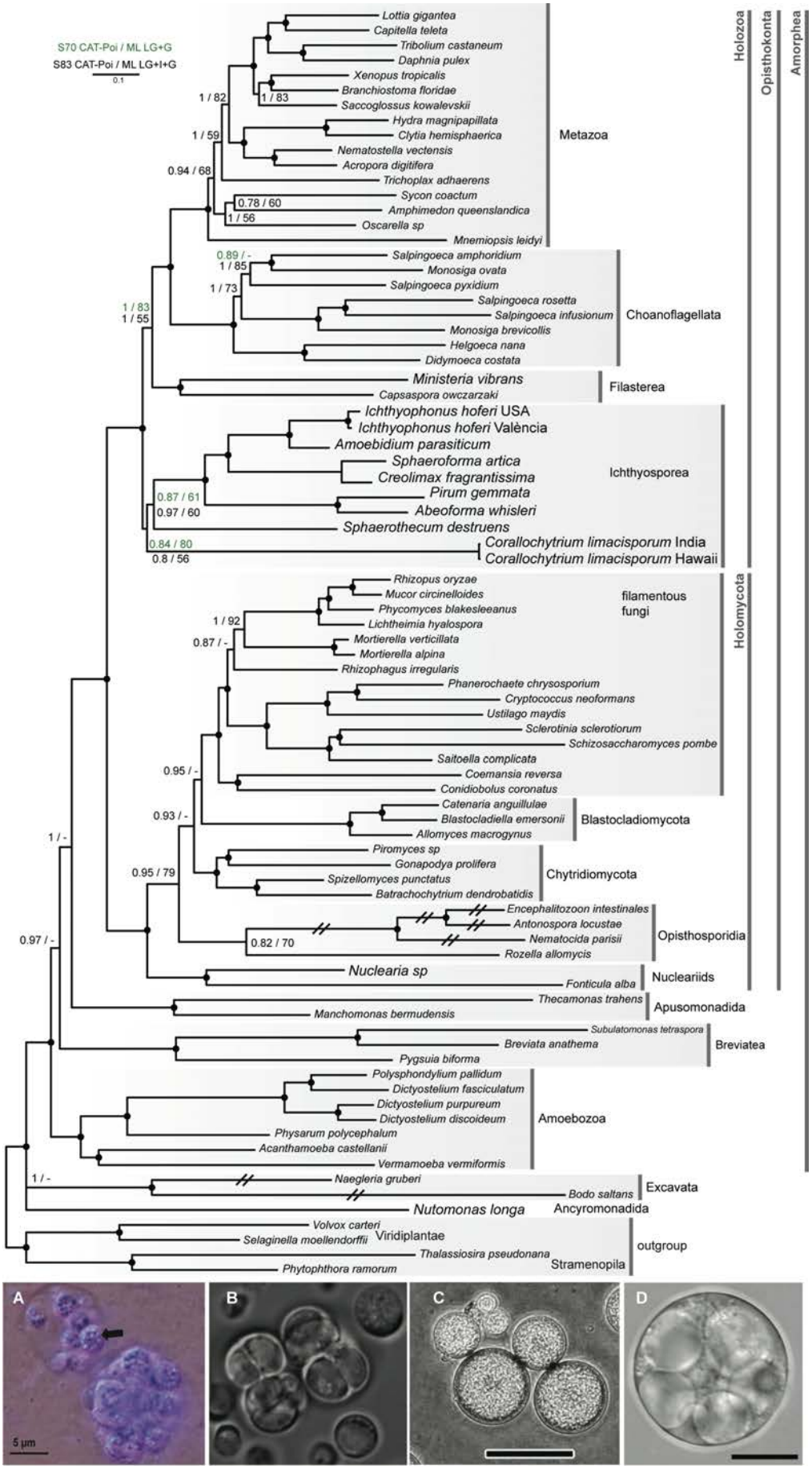


Figure 1. Phylogenetic tree estimated from the 83-taxa matrix (see **Table S1**, **S2** and **Supplemental Information**) inferred by Phylobayes under the CAT-Poisson model. The topology is the consensus of two MCMC chains run for 1,500 generations, saving every 10 trees and after a burnin of 25%. Split supports are posterior probabilities (pp) and nonparametric bootstrap (bs) values using 200 ML replicates using the LG +Γ model implemented in RAxML. Supports from 95 to 100 bs and 0.95 to 1 pp are not shown.

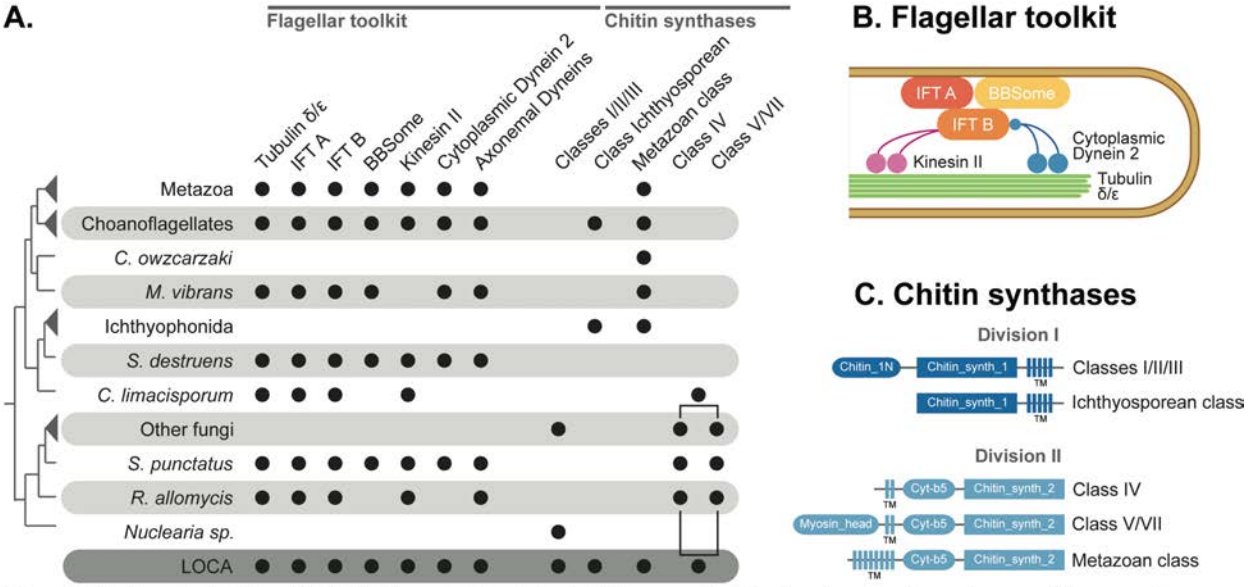


Figure 2. The Flagellar toolkit and Chitin Synthases in Opisthokonta. A. Components of the flagellar apparatus and names of the molecular complexes; adapted from ref. B. Distinct Chitin Synthases (CHS) classes and their domain structure. C. Presence/Absence of the key molecular complexes of the flagellar apparatus and CHS in distinct Opisthokonta lineages.

In order to obtain the phylogenetic relationships, we generated two datasets composed of 93 Protein Domains: one bearing 83 taxa and 18,218 amino acidic positions (S83), and the other designed to minimize topological artifacts/systematic error excluding taxa with long-branches (e.g., Microsporidia, Excavata) and taxa with high percentages of missing data (e.g., EST-based taxa), resulting in a 70 taxa matrix with 22,313 amino acidic positions (S70). Both datasets were consistent in recovering the backbone of eukaryotic phylogeny using both Bayesian inference (BI, CAT-poi) and maximum-likelihood (ML, LG model) (**Figure 1**) approaches (details in **Experimental Procedures**). As sister groups to Opisthokonta, we systematically recovered Apusomonadida and Breviatea as reported in a recent study by Brown et al. [25]. But the topology of the S83 dataset recovered *Nutomonas longa* (Ancyromonadida) branching closer to Excavata than to Apusomonadida, contradicting small ribosomal subunit based phylogenies [3, 4] but confirming recent results based in multigene matrix approach (TCS14). Within Holomycota (which includes fungi and their unicellular relatives), we recovered the clade formed by *Nucleariа* sp. and *Fonticula alba* (Discicristoidea) as the earliest branching lineage [13, 19]; followed by *Rozella allomycis* + Microsporidia [26, 27]; and the paraphyly between Chytridiomycota and Blastocladiomycota [28], although the coenocytic fungi in alternative split positions in ML and BI (see phylogenetic trees in **Supplemental Information**). Finally, within Holozoa we recovered the topologies of previous studies [12, 14], and most importantly, we recovered *C. limacisporum* as sister to Ichthyosporea in both ML and BI methods. The S83 dataset recovers *C. limacisporum* position with weak support (56bs/0.8pp), but excluding the long-branch taxa (S70, see **Table S1**) we recovered higher support for that branching event (80bs/0.84pp) (see **Figure 1**). The position of *Sphaerothecum destruens* is slightly unstable (S83: 60bs/0.97pp; S70: 61bs/0.87pp), but we consistently recovered monophyletic Ichthyosporea (including the two major groups Icthyophonida and Dermocystida) [29] and *C. limacisporum* as the earliest branching lineage within this clade (tentatively named as "Teretosporea" in this manuscript). Thus Teretosporea is the most basal holozoan lineage, and it implies that the transition from phagotrophy to osmotrophy happened only once in the whole holozoan lineage.

C. limacisporum is the only free-living osmotroph within Holozoa, while ichthyosporeans are known to be associated to animal hosts as parasites or commensals [29], despite they are also found in environmental surveys [10]. The life cycle of *C. limacisporum* and Ichthyosporeans [30] is strikingly similar; both start as a single cell that grows as a coenocyte until it reaches maturation, when it suffers schizogony. The offspring (merozoites) have dispersive amoeboid or flagellated forms that once settled, they close the cycle [15, 30, 31]. Within Holomycota, chytrid fungi show a similar developmental mode, with both coenocytic growth and amoeboflagellated stages. Moreover fungi also evolved from phagotrophic ancestors (Nucleariids, *Rozella* and Aphelida [32]) to become saprotrophs and parasites. To investigate the parallelisms between fungi and Teretosporea, we undertook comparative genomics to clarify the evolution of the flagellum and chitin synthases within Opisthokonta.

Unexpected flagellar toolkit in *Corallochytrium limacisporum* and *Ministeria vibrans*

Opisthokonta are characterized by a posterior single motile flagellum [33], nevertheless it has been secondarily lost in many lineages. The flagellum, together with the basal body, is an ancestral eukaryotic organelle with a complex structure encoded by a specific molecular toolkit [34, 35]. Among the molecules that shape the flagellum there are specialized tubulins (Epsilon, Delta) [36], the intraflagellar transport system (which include the IFT-A, IFT-B and BBSome complexes) and the motor molecules, which are specialized subfamilies of Dyneins and Kinesins [34, 37]. Large scale genomic analyses have shown that the presence of those genes in a given genome clearly indicate the presence of a flagellum, even revealing the presence of an unnoticed flagellar stage in some eukaryotes [37].

To clarify the evolution of the flagellum we searched homologs of a set of over 60 flagellum specific proteins [34] in our set of opisthokont protists (see **Experimental Procedures** and **Table S3 Flagellum toolkit**). As expected, non-flagellated lineages such as Dikarya, Discicristoidea, Ichthyophonida and the filasterean *Capsaspora owczarzaki* retrieved no positive hits. But in the transcriptome of the other filasterean species, *M. vibrans*, we found several proteins belonging to all of the

key flagellar molecular complexes. The presence of axonemal dyneins, *epsilon* tubulin and IFT-A/B complexes clearly suggested the presence of a flagellum. *M. vibrans* was originally described as a filose amoeba suspended in the water column thanks to a stalk attached to the substrate, which resembled a modified flagellum based on TEM observations [9]. To test if the stalk is a modified flagellum we performed tubulin immunostaining on the original *M. vibrans* ATCC 50519 strain (see **Experimental Procedures**). Confocal microscopy revealed a tubulin-protrusion branching from the cell body (see **Figure 2?** and/or **Figure SX**). Therefore we find evidence of a flagellar-like structure in *M. vibrans* that conserves most of the ancestral flagellar toolkit.

Even more striking is the finding of multiple flagellum specific genes in the transcriptomes of both *C. limacisporum* strains. Those genes include δ/ε tubulins, IFT-A and IFT-B components and the retrograde motor Kinesin-II (**Figure 2**). Conversely we observe neither cytoplasmic dynein 2 nor any of the axonemal dyneins (heavy, light and intermediate chains, **Table S3**). The absence of those genes may be due the incompleteness of the transcriptome or suggest the presence of a derived non-motile flagellum in *C. limacisporum*, despite it has HEATR2 (ref Diggle 2014 PlosGen). Previous description of *C. limacisporum* [15] and our own observations using light and TEM microscopy did not retrieve any clear flagellar structure at least in culture conditions. Nevertheless molecular data suggest that the life cycle of *C. limacisporum* should present a cryptic flagellated stage, as it has been stated regarding the genome content of *Aureococcus* and *Ostreococcus* (Wickstead and Gull, 2012).

Therefore, within Teretosporea amoeboid and flagellated stages are features shared by *C. limacisporum* and Dermocystida (with no amoeba stage), while flagellated stages were lost in Ichthyophonida (**Figure 3**). This pattern is similar to fungi, were basal lineages retain flagellated stages and later lifestyle specializations lost the flagellar apparatus [38]. The absence of a complete BBSome in basal fungi, Teretosporea and *M. vibrans* reinforces the hypothesis that the loss of this module usually precludes the loss of the complete fagellum in closely related lineages (Van Dam 13). Besides the pattern of flagellar simplification, another feature shared between fungi and Teretosporea is the presence of a cell wall.

At least four Chitin Synthases in the Last Opisthokonta Common Ancestor

To gain insights into the evolution of the cell wall we analysed the evolutionary history of chitin synthases (CHS) within opisthokonta. Chitin is a key biopolymer in fungal cell walls and animal cuticules [39] and is synthesized by CHS, which conform a large and complex multigene family. Several CHS classes have been described within fungi (Classes I/II/III from Division I and classes IV/V/VI/VII from Division II) [7, 40] and three ancestral classes are known in animals [6]. Some of the fungal CHS classes are held as molecular synapomorphies of fungi, as they have been found exclusively in fungal genomes, including *Rozella allomycis* and Microsporidia [8]. Moreover, CHS homologs have been found in other eukaryotes such as the oomycete *Saprolegnia monoica* [41], in diatoms [42] and in unicellular holozoans, including Ichthyosporea [8, 29].

To test which CHS are present in Teretosporea and clarify their phylogenetic relationships compared to those of fungi and animals we gathered CHS homologs from all eukaryotic supergroups and build a tree based on the Chitin synthase domain (see **Supplemental Experimental Procedures**). (**Figure 2** and **Figure SX CHStree**). This revealed three genes belonging to Division II CHS in *C. limacisporum* branching within the clade that comprises fungal classes IV/V/VII and also presenting the canonical functional motifs of fungal sequences (**Table S4 CHSmotifs**). Interestingly two of those genes contain a N-terminal Myosin Head domain, resembling genes from fungal classes V/VII [38]. The myosin head of *C. limacisporum* CHS is sister group to fungal V/VII CHS, forming the Myosin class XVII [43]. Therefore, we propose that the ancestral form of CHS Classes IV/V/VII had a Myosin domain ancestral in Opisthokonta.

We also found that ichthysoporeans have CHS from both Division I and Division II clades. Ichthyosporeans homologs from Division I form a new clade with various eukaryotic sequences, that include diatoms, choanoflagellates and amoebozoans (**Figure 2** and **Figure SX CHStree**), revealing that it is also an ancestral class within eukaryotes. Ichthyosporean Division II CHS homologs belong to the Metazoan class, which is also present in other unicellular holozoans, apusozoans and amoebozoans while it was secondarily lost in fungi. Finally, fungal class I/II/III is found in several bikonts, including oomycetes and chlorophytes, suggesting an ancestral origin and secondary loss in holozoa. In summary, at least 4 ancestral lineages of structurally different CHS were found in the Last Opisthokont Common Ancestor (LOCA), and secondary loss was common in descendant lineages.

The presence of a complex CHS repertoire in the ancestor of all opisthokonta and the retention of rich CHS repertoires in the cell-walled lineages suggests that the presence of chitin in the cell wall was an ancestral feature and not a synapomorphy of fungi [8, 44].

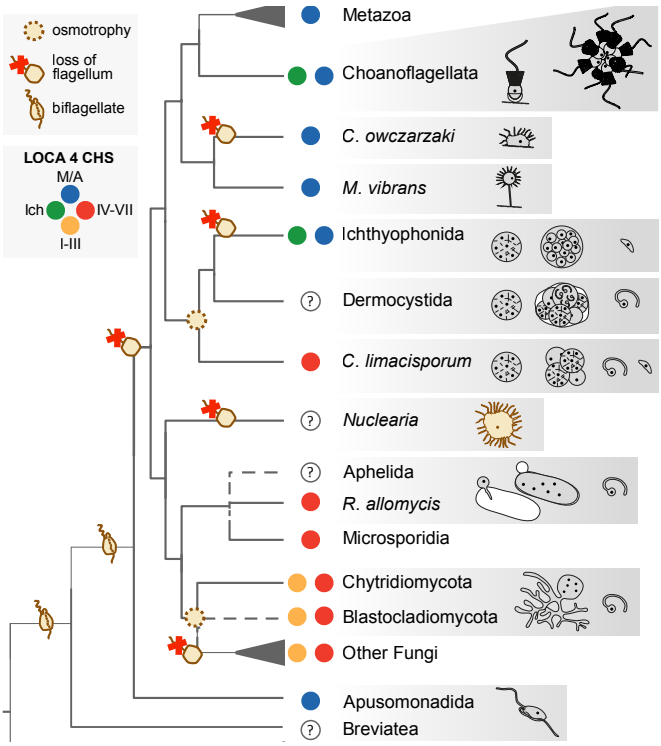


Figure 3. Opisthokonta cladogram displaying lifestyle characteristics such as feeding mode, flagellated stage, CHS repertoire and developmental mode.

The opisthokonta phylogenetic framework with all known protistan lineages

Sequencing the transcriptome of more than 10 opisthokont protists allowed us to reassess their phylogenetic relationships only a couple of years after our last study [14]. Moreover, it opened the doors to large-scale comparative genomic studies. Our transcriptome information from several unicellular opisthokonts already shed light on how basic gene families in fungi and animals evolved. Here we presented case examples for the CHS and the flagellar toolkit, but previous reports revealed complex repertoires of multicellularity related genes in some of those lineages, including Transcription Factors and Tyrosine Kinases [48, 49]. The wealth of these data will provide a less biased representation of genomic information [10, 50], which is not only severely required to understand the basic biology of each species but also to have a broad spectrum of molecular data to investigate the diversification of opisthokonts, the origins of multicellularity in animals or the independent adaptations to parasitism in this eukaryotic supergroup [22, 23, 51].

Overall we propose that the earliest splitting clade within holozoa comprises *C. limacisporum* and Ichthyosporea (Teretosporea). This has important evolutionary implications, as the clade presents shared features, such as an osmotrophic feeding mode, cell wall and a similar life cycle. Interestingly those characters present similarities with fungi, which evolved independently. Fungi and Teretosporea show diversity of overlapping lifestyles, ranging from saprotrophic to parasitic (animal-dwelling) species. Therefore their similar adaptations represent a pattern of convergent evolution, which is not rare in eukaryotes. Similar adaptations are found in stramenopiles such oomycetes and thraustochytrids [52–54]. The data generated in this study will allow new avenues to identify the trends and patterns regarding molecular adaptations to saprotrophy and to parasitism in distantly related eukaryotes.

Experimental Procedures

The Supplemental Information details complete experimental procedures. In summary, total RNA was obtained from 11 species using a Trizol reagents and DNAse treatment. Purified RNA samples were sent to Beijing Genome Institute, where 100 basepair paried-end cDNA libraries for Illumina HiSeq 2000 were sequenced (see RNAseq in **Supplemental Experimental Procedures**). Moreover newly sequenced transcriptomes of *Ministeria vibrans* ATCC 50519 and *Amoebidium parasiticum* JAP-7-2 were obtained from Origins of Multicellularity Database supported by the Broad Institute (http://www.broadinstitute.org/annotation/genome/multicellularity_project).

All transcriptomes were assembled using de Trinity *de novo* assembly pipeline [55], and further translated to 6 frames. The phylogenomic dataset based on conserved Single Copy Protein Domains [14] was updated with new data from assembled transcriptomes and published genome sequences [8, 19, 25, 45, 47] (see Dataset preparation and **Table S1** in **Supplemental Experimental Procedures**). Using in-home scripts based on tBLASTn and BLASTp followed by manual inspection of single domain trees obtained by RaXML we obtained the hits for each species, which were aligned using MAFFT [56] and trimmed with BMGE [57] to obtain a concatenated supermatrix (dataset S83) covering 83 taxa and 18218 amino acids. Appropriate evolutionary model was selected using ProtTest [58], LINKAR PROTEST AMB RAXML, no? which retrieved LG model + Γ distribution. Thus Maximum likelihood trees were performed using Le and Gascuel (LG) model + Γ distribution with four rate categories and invariant sites (LG+I+G) with RAxML [59]. Statistical support was computed with 200 non-parametric bootstrap replicates. Alternative splits for the key phylogenetic questions were performed using AU test [60]. Bayesian analyses were performed with Phylobayes [61] under CAT-Poisson model. We assembled the second dataset excluding the longest branches from each alignment (dataset S70, see **Table S1**) and reassembling the matrix retaining ~4,000 more amino acid positions. The same phylogenetic methods were applied.

Accession Numbers

All new sequence data used in this study include SRA for *Pirum gemmata*, *Abeoforma whisleri*, *Corallochytrium limacisporum* India and Hawaii strain, *Sphaerothecum destruens*, *Ichthyophonus hoferi*, *Nutomonas longa* CCAP 1985/5 and *Nuclearia sp*. ATCC 50694.

Supplemental Information

Supplemental Information includes 2 figures, 4 tables, Supplemental Experimental Procedures can be found with this article online.

Acknowledgments

Dan Richter, Franz Lang, Matt Brown, Joseph Ryan, Ana Riesgo and Scott Nichols for providing data for six choanoflagellates, *Manchomonas bermudensis*, *Pygsuia biforma*, *Mnemiopsis leidyi*, *Sycon coactum*, *Oscarella carmela* and *Oscarella sp*. respectively. Thanks to Romain Derelle, Jordi Paps, Diego Mallo, Martin Kolisko and Andrew Roger for providing scripts and methodological help. Thanks to Cadaques and Floquet IBE cluster, Philippe Lopez and Eric Bapteste for computational help. We also thank Arnau Sebé-Pedrós for insightful comments on the manuscript and methodological help regarding microscopy.

References

1. Adl, S. M., Simpson, A. G. B., Lane, C. E., Lukeš, J., Bass, D., Bowser, S. S., Brown, M. W., Burki, F., Dunthorn, M., Hampl, V., et al. (2012). The revised classification of eukaryotes. J. Eukaryot. Microbiol. *59*, 429–514.
2. Richter, D. J., and King, N. (2013). The genomic and cellular foundations of animal origins. Annu. Rev. Genet. *47*, 509–37.
3. Glücksman, E., Snell, E. a, and Cavalier-Smith, T. (2013). Phylogeny and evolution of Planomonadida (Sulcozoa): Eight new species and new genera Fabomonas and Nutomonas. Eur. J. Protistol.
4. Paps, J., Medina-Chacón, L. a, Marshall, W., Suga, H., and Ruiz-Trillo, I. (2013). Molecular phylogeny of unikonts: new insights into the position of apusomonads and ancyromonads and the internal relationships of opisthokonts. Protist *164*, 2–12.
5. Van Dam, T. J. P., Townsend, M. J., Turk, M., Schlessinger, A., Sali, A., Field, M. C., and Huynen, M. a (2013). Evolution of modular intraflagellar transport from a coatomer-like progenitor. Proc. Natl. Acad. Sci. U. S. A. *110*, 6943–8.
6. Zakrzewski, A.-C., Weigert, A., Helm, C., Adamski, M., Adamska, M., Bleidorn, C., Raible, F., and Hausen, H. (2014). Early divergence, broad distribution, and high diversity of animal chitin synthases. Genome Biol. Evol. *6*, 316–25.
7. Pacheco-Arjona, J. R., and Ramirez-Prado, J. H. (2014). Large-scale phylogenetic classification of fungal chitin synthases and identification of a putative cell-wall metabolism gene cluster in Aspergillus genomes. PLoS One *9*, e104920.
8. James, T. Y., Pelin, A., Bonen, L., Ahrendt, S., Sain, D., Corradi, N., and Stajich, J. E. (2013). Shared Signatures of Parasitism and Phylogenomics Unite Cryptomycota and Microsporidia. Curr. Biol., 1–6.
9. Cavalier-Smith, T., and Chao, E. E.-Y. (2003). Phylogeny of choanozoa, apusozoa, and other protozoa and early eukaryote megaevolution. J. Mol. Evol. *56*, 540–63.

10. Del Campo, J., and Ruiz-Trillo, I. (2013). Environmental Survey Meta-analysis Reveals Hidden Diversity among Unicellular Opisthokonts. Mol. Biol. Evol. *30*, 802–5.

11. 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*, 664–72.

12. 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.

13. Liu, Y., Steenkamp, E. T., Brinkmann, H., Forget, L., Philippe, H., and Lang, B. F. (2009). Phylogenomic analyses predict sistergroup relationship of nucleariids and fungi and paraphyly of zygomycetes with significant support. BMC Evol. Biol. *9*, 272.

14. Torruella, G., Derelle, R., Paps, J., Lang, B. F., Roger, A. J., Shalchian-Tabrizi, K., and Ruiz-Trillo, I. (2012). Phylogenetic relationships within the Opisthokonta based on phylogenomic analyses of conserved single copy protein domains. Mol. Biol. Evol. *29*, 531–544.

15. Raghu-Kumar, S. (1987). Occurrence of the thraustochytrid, Corallochytrium limacisporum gen. et sp. nov. in the coral reef lagoons of the Lakshadweep Islands in the Arabian Sea. Bot. Mar. *30*, 83–89.

16. Sumathi, J. C., Raghukumar, S., Kasbekar, D. P., and Raghukumar, C. (2006). Molecular evidence of fungal signatures in the marine protist Corallochytrium limacisporum and its implications in the evolution of animals and fungi. Protist *157*, 363–76.

17. Steenkamp, E. T., Wright, J., and Baldauf, S. L. (2006). The protistan origins of animals and fungi. Mol. Biol. Evol. *23*, 93–106.

18. Carr, M., Leadbeater, B. S. C., Hassan, R., Nelson, M., and Baldauf, S. L. (2008). Molecular phylogeny of choanoflagellates, the sister group to Metazoa. Proc. Natl. Acad. Sci. U. S. A. *105*, 16641–6.

19. Brown, M. W., Spiegel, F. W., and Silberman, J. D. (2009). Phylogeny of the “forgotten” cellular slime mold, Fonticula alba, reveals a key evolutionary branch within Opisthokonta. Mol. Biol. Evol. *26*, 2699–709.

20. Marshall, W. L., Celio, G., McLaughlin, D. J., and Berbee, M. L. (2008). Multiple isolations of a culturable, motile Ichthyosporean (Mesomycetozoa, Opisthokonta), Creolimax fragrantissima n. gen., n. sp., from marine invertebrate digestive tracts. Protist *159*, 415–33.

21. Marshall, W. L., and Berbee, M. L. (2011). Facing Unknowns : Living Cultures (Pirum gemmata gen . nov ., sp . nov ., and Abeoforma whisleri , gen . nov ., sp . nov .) from Invertebrate Digestive Tracts Represent an Undescribed Clade within the Unicellular Opisthokont Lineage Ichthyosporea (Mesom (Elsevier GmbH.).

22. Kocan, R., LaPatra, S., and Hershberger, P. (2013). Evidence for an amoeba-like infectious stage of Ichthyophonus sp. and description of a circulating blood stage: a probable mechanism for dispersal within the fish host. J. Parasitol. *99*, 235–40.

23. Paley, R. K., Andreou, D., Bateman, K. S., and Feist, S. W. (2012). Isolation and culture of Sphaerothecum destruens from Sunbleak (Leucaspius delineatus) in the UK and pathogenicity experiments in Atlantic salmon (Salmo salar). Parasitology *139*, 904–14.

24. Atkins, M., McArthur, A., and Teske, A. (2000). Ancyromonadida: a new phylogenetic lineage among the protozoa closely related to the common ancestor of metazoans, fungi, and choanoflagellates (Opisthokonta). J. Mol. Evol. *51*, 278–285.

25. Brown, M. W., Sharpe, S. C., Silberman, J. D., Heiss, A. A., Lang, B. F., Simpson, A. G. B., and Roger, A. J. (2013). Phylogenomics demonstrates that breviate flagellates are related to opisthokonts and apusomonads. Proc. R. Soc. B Biol. Sci. *280*, 1–9.

26. Cavalier-Smith, T., Chao, E. E., Snell, E. a, Berney, C., Fiore-Donno, A. M., and Lewis, R. (2014). Multigene eukaryote phylogeny reveals the likely protozoan ancestors of opisthokonts (animals, fungi, choanozoans) and Amoebozoa. Mol. Phylogenet. Evol.

27. James, T. Y., Kauff, F., Schoch, C. L., Matheny, P. B., Hofstetter, V., Cox, C. J., Celio, G., Gueidan, C., Fraker, E., Miądlikowska, J., et al. (2006). Reconstructing the early evolution of Fungi using a six-gene phylogeny. Nature *443*, 818–22.

28. Lara, E., Moreira, D., and López-García, P. (2010). The environmental clade LKM11 and Rozella form the deepest branching clade of fungi. Protist *161*, 116–21.

29. James, T. Y., Letcher, P. M., Longcore, J. E., Mozley-Standridge, S. E., Porter, D., Powell, M. J., Griffith, G. W., and Vilgalys, R. (2006). A molecular phylogeny of the flagellated fungi (Chytridiomycota) and description of a new phylum (Blastocladiomycota). Mycologia *98*, 860–71.

30. Glockling, S. L., Marshall, W. L., and Gleason, F. H. (2013). Phylogenetic interpretations and ecological potentials of the Mesomycetozoea (Ichthyosporea). Fungal Ecol. *6*, 237–247.

31. Suga, H., and Ruiz-Trillo, I. (2013). Development of ichthyosporeans sheds light on the origin of metazoan multicellularity. Dev. Biol. *377*, 284–292.

32. Kocan, R. M. (2013). Proposed changes to the nomenclature of Ichthyophonus sp. life stages and structures. J. Parasitol. *99*, 906–9.

33. Karpov, S. a, Mamkaeva, M. a., Aleoshin, V. V., Nassonova, E., Lilje, O., and Gleason, F. H. (2014). Morphology, phylogeny, and ecology of the aphelids (Aphelidea, Opisthokonta) and proposal for the new superphylum Opisthosporidia. Front. Microbiol. *5*, 1–11.

34. Wainright, P., Hinkle, G., Sogin, M., and Stickel, S. (1993). Monophyletic origins of the metazoa: an evolutionary link with fungi. Science (80-.). *260*, 340–342.

35. Carvalho-Santos, Z., Azimzadeh, J., Pereira-Leal, J. B., and Bettencourt-Dias, M. (2011). Tracing the origins of centrioles, cilia, and flagella. J. Cell Biol. *194*, 165–175.

36. Hodges, M. E., Scheumann, N., Wickstead, B., Langdale, J. a, and Gull, K. (2010). Reconstructing the evolutionary history of the centriole from protein components. J. Cell Sci. *123*, 1407–13.

37. Findeisen, P., Muhlhausen, S., Dempewolf, S., Hertzog, J., Zietlow, A., Carlomagno, T., and Kollmar, M. (2014). Six subgroups and extensive recent duplications characterize the evolution of the eukaryotic tubulin protein family. Genome Biol. Evol. *6*, 2274–2288.

38. Wickstead, B., and Gull, K. (2012). The evolutionary biology of dyneins. In Dyneins: Structure, biology and disease, Elsevier Inc., ed. (London, UK), pp. 89–121.

39. James, T. Y., and Berbee, M. L. (2012). No jacket required--new fungal lineage defies dress code: recently described zoosporic fungi lack a cell wall during trophic phase. Bioessays *34*, 94–102.

40. Merzendorfer, H. (2011). The cellular basis of chitin synthesis in fungi and insects: common principles and differences. Eur. J. Cell Biol. *90*, 759–69.

41. Ruiz-Herrera, J., and Ortiz-Castellanos, L. (2010). Analysis of the phylogenetic relationships and evolution of the cell walls from yeasts and fungi. FEMS Yeast Res. *10*, 225–43.

42. Leal-Morales, C. a, Gay, L., Fèvre, M., and Bartnicki-García, S. (1997). The properties and localization of Saprolegnia monoica chitin synthase differ from those of other fungi. Microbiology *143 (Pt 7*, 2473–83.

43. Durkin, C. a, Mock, T., and Armbrust, E. V. (2009). Chitin in diatoms and its association with the cell wall. Eukaryot. Cell *8*, 1038–50.

44. Sebé-Pedrós, A., Grau-Bové, X., Richards, T. a, and Ruiz-Trillo, I. (2014). Evolution and classification of myosins, a paneukaryotic whole-genome approach. Genome Biol. Evol. *6*, 290–305.

45. Nagy, L. G., Ohm, R. a, Kovács, G. M., Floudas, D., Riley, R., Gácser, A., Sipiczki, M., Davis, J. M., Doty, S. L., de Hoog, G. S., et al. (2014). Latent homology and convergent regulatory evolution underlies the repeated emergence of yeasts. Nat. Commun. *5*, 4471.

46. Grigoriev, I. V, Nikitin, R., Haridas, S., Kuo, A., Ohm, R., Otillar, R., Riley, R., Salamov, A., Zhao, X., Korzeniewski, F., et al. (2014). MycoCosm portal: gearing up for 1000 fungal genomes. Nucleic Acids Res. *42*, D699–704.

47. Cavalier-Smith, T., and Chao, E. E. (2010). Phylogeny and Evolution of Apusomonadida (Protozoa: Apusozoa): New Genera and Species. Protist, 1–28.

48. Grant, J. R., Lahr, D. J. G., Rey, F. E., Burleigh, J. G., Gordon, J. I., Knight, R., Molestina, R. E., and Katz, L. a. (2012). Gene discovery from a pilot study of the transcriptomes from three diverse microbial eukaryotes: Corallomyxa tenera, Chilodonella uncinata, and Subulatomonas tetraspora. Protist Genomics *1*, 3–18.

49. De Mendoza, A., Sebé-Pedrós, A., Sestak, M. S., Matejcic, M., Torruella, G., Domazet-Lošo, T., and Ruiz-Trillo, I. (2013). Transcription factor evolution in eukaryotes and the assembly of the regulatory toolkit in multicellular lineages. Proc. Natl. Acad. Sci. U. S. A. *110*, E4858–66.

50. Suga, H., Torruella, G., Burger, G., Brown, M. W., and Ruiz-Trillo, I. (2014). Earliest holozoan expansion of phosphotyrosine signaling. Mol. Biol. Evol. *31*, 517–28.

51. Keeling, P. J., Burki, F., Wilcox, H. M., Allam, B., Allen, E. E., Amaral-Zettler, L. a, Armbrust, E. V., Archibald, J. M., Bharti, A. K., Bell, C. J., et al. (2014). The Marine Microbial Eukaryote Transcriptome Sequencing Project (MMETSP): Illuminating the Functional Diversity of Eukaryotic Life in the Oceans through Transcriptome Sequencing. PLoS Biol. *12*, e1001889.

52. Gozlan, R. E., Marshall, W. L., Lilje, O., Jessop, C. N., Gleason, F. H., and Andreou, D. (2014). Current ecological understanding of fungal-like pathogens of fish: what lies beneath? Front. Microbiol. *5*, 62.

53. Spanu, P., and Kämper, J. (2010). Genomics of biotrophy in fungi and oomycetes--emerging patterns. Curr. Opin. Plant Biol. *13*, 409–14.

54. Beakes, G. W., Glockling, S. L., and Sekimoto, S. (2012). The evolutionary phylogeny of the oomycete “fungi”. Protoplasma *249*, 3–19.

55. Richards, T. a, and Talbot, N. J. (2013). Horizontal gene transfer in osmotrophs: playing with public goods. Nat. Rev. Microbiol. *11*, 720–7.

56. Grabherr, M. G., Haas, B. J., Yassour, M., Levin, J. Z., Thompson, D. a, Amit, I., Adiconis, X., Fan, L., Raychowdhury, R., Zeng, Q., et al. (2011). Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nat. Biotechnol. *29*, 644–652.

57. 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*, 3059–66.

58. Criscuolo, A., and Gribaldo, S. (2010). BMGE (Block Mapping and Gathering with Entropy): a new software for selection of phylogenetic informative regions from multiple sequence alignments. BMC Evol. Biol. *10*, 210.

59. Darriba, D., Taboada, G. L., Doallo, R., and Posada, D. (2011). ProtTest 3: fast selection of best-fit models of protein evolution. Bioinformatics *27*, 1164–5.

60. Stamatakis, A. (2006). RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. Bioinformatics *22*, 2688–90.

61. Shimodaira, H., and Hasegawa, M. (2001). CONSEL: for assessing the confidence of phylogenetic tree selection. Bioinformatics *17*, 1246–7.

62. Blanquart, S., and Lartillot, N. (2008). A site- and time-heterogeneous model of amino acid replacement. Mol. Biol. Evol. *25*, 842–58.

Supplemental Figures and Tables

Figure S1, related to Figure 2. Flagellum-like stalk in *Ministeria vibrans*. Confocal microscopy showing *Ministeria vibrans* ATCC 50519 stained with DAPI and beta-tubulin antibody (Antimouse Alexa 514, Life Technologies A-31555). *M. vibrans* grows feeding on bacteria, which can be seen as the DAPI staining outside the cell body. Scale-bar 5 µm.

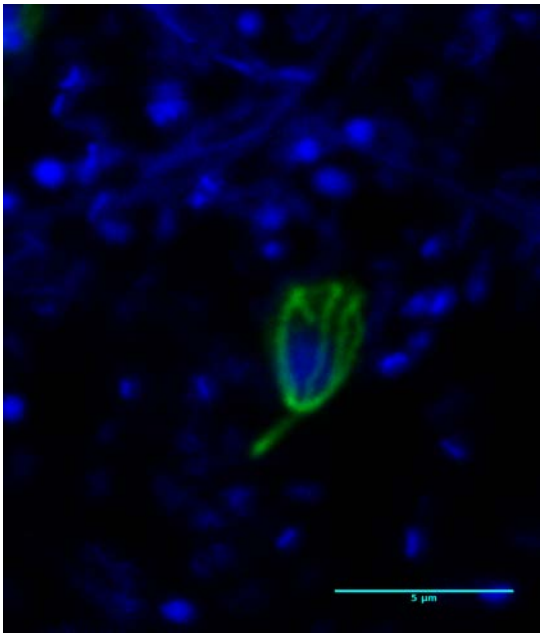
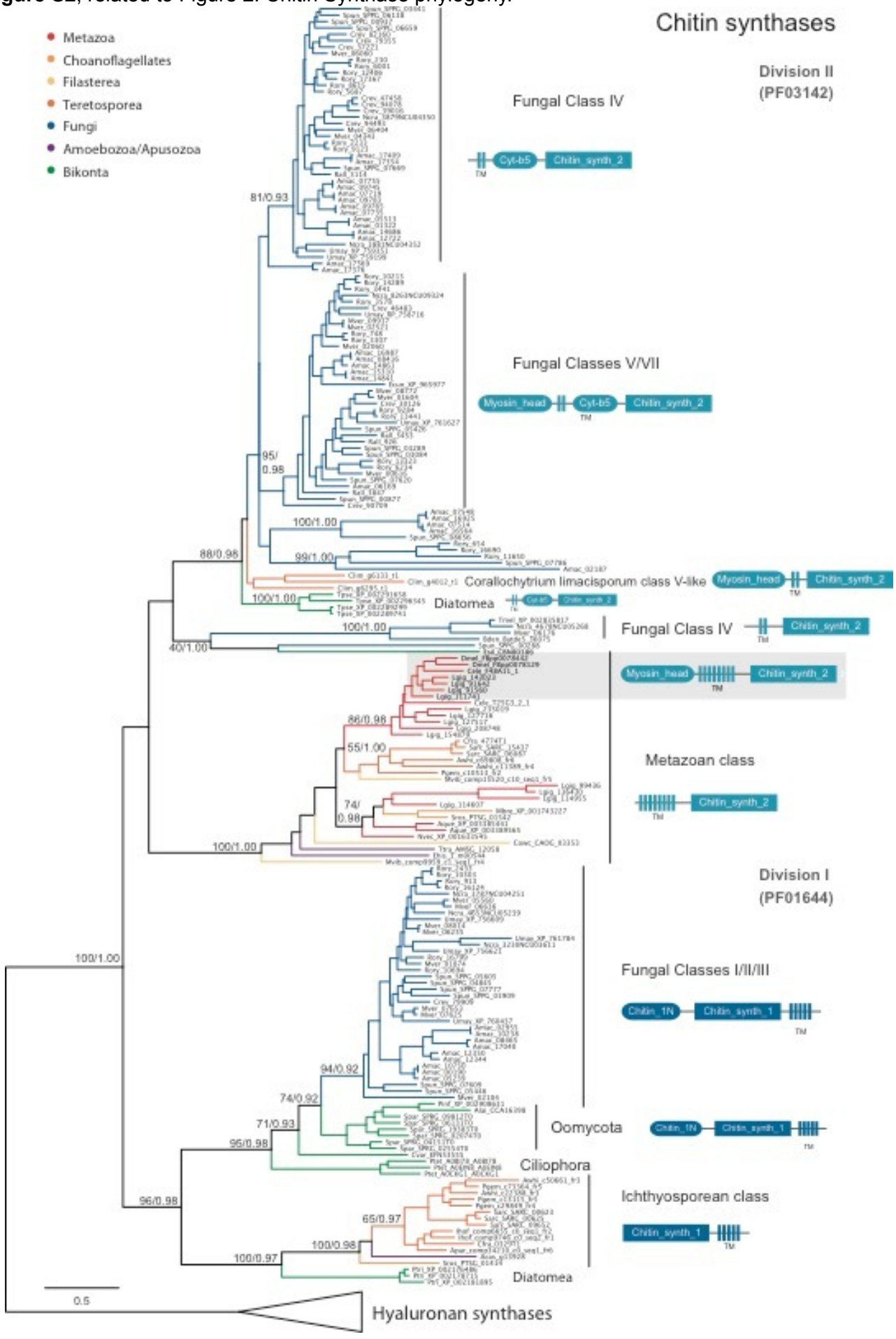


Figure S2, related to Figure 2. Chitin Synthase phylogeny.



Supplemental Experimental Procedures

RNAseq

Cultures for distinct species were kept in specific media and temperature conditions. Typically they were grown in 75 cm^2 culture flasks with 25mL of total volume. Ichthyophonida (*Pirum gemmata*, *Abeoforma whisleri* [1]) and two *Corallochytrium* strains (sampled in Kavaratti island as the original isolation [2], and in Kāneʻohe Bay Hawaii in October 2011) grew for one to two weeks with marine broth medium. *Ichthyophonus hoferi* (USA and València) *Sphaerothecum destruens* (Paley UK) from fathead minnow in UK *Planomonas carolina* CCAP 1958/5 (a.k.a. *Nutomonas longa* [3]) was grown in mineral water with bacteria provided in the acquired culture with 2 grains of cereal grass to feed bacteria. *Nuclearia sp*. ATCC 50694 was grown using ATCC Medium 802 bacterized with *Klebsiella pneumoniae subsp. pneumoniae* (ATTC 700831).

Total RNA was extracted from dense cultures with standard protocol of TRIzol (Ambion), precipitated with etOH, incubated with DNAse I (Roche) and purified with RNeasy mini Kit (Quiagen). Quantification was performed with Nanodrop (Thermo Scientific) and quality assessment with Bioanalyzer (Agilent Technologies, Inc). Valid samples had ≥ 10µg of total RNA, ≥200ng/µl, OD260/280 = 1.8~2.2, RNA 28S: 18S > 1.0, RIN ≥ 7.0. For the phagotrophic *Nuclearia sp*. and *Planomonas* dense cultures cannot reach the minimum quantity of total RNA, therefore different samples were combined to reach 100µg as bacterial RNA contamination was supposed to highly contribute in the samples. Samples were shipped to Beijin Genome Institute (BGI) for standard *de novo* mRNA sequencing transcriptome protocol. There, 200bp short-insert libraries were constructed followed by Illumina HiSeq 2000 101 paired-end sequencing. Raw reads were processed using BGI's automatic pipeline. Contig assembly was performed with Trinity program [4] and CDS annotation with BLAST against distinct protein databases including EST.

Ministeria vibrans ATCC 50519 and *Amoebidium parasiticum* JAP-7-2 RNAseq raw data from Broad Institute Origins of Multicellularity Database (http://www.broadinstitute.org/annotation/genome/multicellularity_project/MultiHome.html) were assembled in-house with MIRA 3 program [5].

Dataset preparation

Previously assembled 93 Single Copy Protein Domain (SCPD) dataset [6] was updated with the broadest possible taxonomic sampling, see Table S1 for details. A total number of 83 taxa were included: 16 Metazoa (7 Bilateria, 4 Cnidaria, 1 Placozoa, 3 Porifera and 1 Ctenophora), 20 Holozoa protists (8 Choanomonada, 2 Filasterea, 8 Ichthyosporea and 2 *Corallochytrium*), 22 Fungi (6 Dikarya, 1 Glomeromycota, 6 Mucoromycotina, 1 Entomophthoramycotina, 1 Kickxellomycotina, 3 Blastocladiomycota and 4 Chytridiomycota), 6 Holomycota protists (3 Microsporidia, 1 Rozellomycota=Cryptosporidia, and 2 Nucleariids), 2 Apusomonadida, 3 Breviatea, 1 Ancyromonadida, 2 Excavata, 2 Vidiriplantae and 2 Stramenopila. So, a part from the previous taxon sampling [6] and nine newly sequenced taxa, we included genome data of *Mortierella verticillata* and *Fonticula alba* from the Origins of Multicellularity project (http://www.broadinstitute.org/annotation/genome/multicellularity_project); draft genome data of *Creolimax fragrantissima* [7]; transcriptomic data of six choanoflagellates (described elsewhere); genomes of ten holomycotan species from Mycocosm and other projects in Joint Genome Institute (http://genome.jgi.doe.gov/programs/fungi/index.jsf); transcriptomic data of *Acropora digitifera* (http://www.compagen.org/datasets.html); genome data of *Physarum polycephalum* (http://genome.wustl.edu/genomes/detail/physarum-polycephalum/) and *Polysphondylium pallidum* (http://genomes.dictybase.org/pallidum/current), *Bodo saltans* from ncbi and *Naegleria gruberi* from JGI; as well as EST data of *Breviata anathema* [8], *Subulatomonas tetraspora* [9] and *Manchomonas bermudensis*.

We used all 78 SCPDs from *Amphimedon*, *Batrachochytrium*, *Capsaspora*, *Nematostella* and *Xenopus* (comprising all markers with the highest percentage of data possible) to mine new genomes and transcriptomes in order to comprise well sampled Opisthokonta lineages and its outgroup. Taxon mining was performed with BLASTp or tBLASTn depending on the data format for each species to include (predicted proteome or genome/transcriptome respectively). Thanks to different bioperl modules and scripts we added the 4 best hits (e-value 10^{-5}, minimum fraction length of 0.2) fusing different hits when overlapping as we were dealing with SCPDs. The output fasta files were automatically aligned with

MAFFT v7 [10] L-INS-I with 1000 iterations and trimmed from ambiguously aligned regions with BMGE (Blossum95) [11].
Phylogenomic studies require semi-automatic pipelines to deal with the big amounts of data and keep reproducibility. Here we used automatic steps for main processes like sequence retrieval, multiple alignment, trimming or tree inference. But in between, manual curation was performed. In its basic sense, it follows the same protocol of a recently published pipeline, which automatize the fundamental steps in phylogenomic studies with fine details [12]. Particularly for the paralogy or HGT, we inferred single marker trees with RAxML and 100 rapid bootstrap [13] for each alignment (visualized with Seaview v4 [14]) and checked suspicious clusterings (for further insights see figures 1 and 2 and main text in TreSpEx paper [12]). Regarding contamination, as most data is sequenced with next-generation sequencing, there is a always a chance for cross-contamination between transcriptomes whose mRNA has been amplified in the same lane. If two sequences had identical sites, the proper one was established checking the tree. Also, species that grow with bacteria (*Nuclearia* and *Planomonas*) have also a percentage of bacterial sequences. Manual Blast Reverse was performed for specific taxa with long branches such Microsporidia and Excavata (but due to miss-alignment suspicion, some sequences were discarded); also extremely divergent sequences can lead to model violations. In case of uncertainty, sequences were discarded. Once alignments were confidently clean, we automatically concatenated them using Alvert.py script from the package Barrel-o-Monkeys, as it provides percentage of present data for each species (see **Table S1**). This resulted in the main dataset S83 with 18218 amino acids.
Dataset S70 was prepared from individual alignments with S83 taxon sampling once we excluded Long Branch taxa such Microsporidia, Excavata, *Nutomonas longa*, Breviatea, *Manchomonas bermudensis* or *Mnemiopsis leidyi* (see **Table S1**). Then alignments were realigned as in S83 but trimmed with BMGE (Blossum 70, more relaxed because alignments were more reliable without Long Branch taxa).

Table S1. Taxonomic sampling. Presenting the percentage of data for S83 matrix.

Species name	#Genes	#Missing	%Missing	%Gaps	
Breviata anathema	4	74	94,87	96,69	ESTs
Manchomonas bermudensis	11	67	85,9	91,74	ESTs
Subulatomonas tetraspora	14	64	82,05	88,73	ESTs
Vermamoeba vermiformis	27	51	65,38	78,92	ESTs
Blastocladiella emersonii	46	32	41,03	64,39	Torruella et al. 2012 ESTs
Ichthyophonus_hoferi USA	69	9	11,54	59,68	RNAseq
Antonospora locustae	43	35	44,87	53,43	JGI
Nematocida parisii	41	37	47,44	52,9	JGI
Encephalitozoon intestinales	43	35	44,87	51,22	JGI
Clytia hemisphaerica	50	28	35,9	49,97	Torruella et al. 2012 ESTs
Piromyces sp	53	25	32,05	34,98	JGI
Bodo saltans	62	16	20,51	31,04	ncbi
Physarum polycephalum	62	16	20,51	29,92	http://genome.wustl.edu/
Fonticula alba	61	17	21,79	23,1	Broad Institute
Sycon coactum	64	14	17,95	22,77	Ana Riesgo
Gonapodya prolifera	62	16	20,51	21,3	JGI
Rhizophagus irregularis	71	7	8,97	21,18	JGI
Salpingoeca rosetta	70	8	10,26	19,42	Dan Richter
Nutomonas longa	63	15	19,23	19,1	RNAseq
Pygsuia biforma	67	11	14,1	16,67	Matt Brown
Saccoglossus kowalevskii	74	4	5,13	15,5	Torruella et al. 2012
Oscarella sp	72	6	7,69	14,72	Scott Nichols
Rozella allomycis	71	7	8,97	12,98	JGI
Naegleria gruberi	67	11	14,1	12,91	JGI
Sclerotinia sclerotiorum	70	8	10,26	12,86	Torruella et al. 2012
Thecamonas trahens	65	13	16,67	12,13	Torruella et al. 2012
Sphaerothecum destruens	72	6	7,69	11,6	RNAseq
Ichthyophonus_hoferi valencia	73	5	6,41	11,18	RNAseq
Mortierella alpina	76	2	2,56	11,02	ncbi
Phytophthora ramorum	69	9	11,54	10,92	Torruella et al. 2012
Conidiobolus coronatus	70	8	10,26	10,39	JGI
Corallochytrium limacisporum india	72	6	7,69	10,15	RNAseq
Sphaeroforma artica	73	5	6,41	9,17	Broad Institute
Didymoeca costata	71	7	8,97	9,03	Dan Richter
Schizosaccharomyces pombe	71	7	8,97	8,82	Torruella et al. 2012
Thalassiosira pseudonana	70	8	10,26	8,71	Torruella et al. 2012
Monosiga ovata	72	6	7,69	8,07	Dan Richter
Corallochytrium limacisporum hawaii	69	9	11,54	7,96	RNAseq
Helgoeca nana	73	5	6,41	7,85	Dan Richter
Salpingoeca pyxidium	73	5	6,41	7,83	Dan Richter
Dictyostelium fasciculatum	77	1	1,28	6,87	dictybase
Creolimax fragrantissima	74	4	5,13	6,71	RNAseq/genome
Salpingoeca amphoridium	74	4	5,13	6,64	Dan Richter
Lottia gigantea	75	3	3,85	6,44	Torruella et al. 2012
Polysphondylium pallidum	77	1	1,28	6,31	dictybase
Catenaria anguillulae	75	3	3,85	6,16	JGI
Acanthamoeba castellanii	74	4	5,13	6,11	Torruella et al. 2012
Daphnia pulex	73	5	6,41	5,54	Torruella et al. 2012
Mnemiopsis leidyi	75	3	3,85	5,23	Joseph Ryan
Nuclearia sp	77	1	1,28	5,22	RNAseq
Salpingoeca infusionum	72	6	7,69	5,21	Dan Richter
Coemansia reversa	75	3	3,85	4,89	JGI
Ustilago maydis	74	4	5,13	4,79	Torruella et al. 2012
Acropora digitifera	76	2	2,56	4,63	Compagen.org
Trichoplax adhaerens	76	2	2,56	4,01	Torruella et al. 2012
Allomyces macrogynus	74	4	5,13	3,99	Torruella et al. 2012
Hydra magnipapillata	77	1	1,28	3,95	Torruella et al. 2012
Amphimedon queenslandica	75	3	3,85	3,83	Torruella et al. 2012
Monosiga brevicollis	76	2	2,56	3,69	Torruella et al. 2012
Ministeria vibrans	76	2	2,56	3,64	RNAseq Broad Institute
Pirum gemmata	76	2	2,56	3,63	RNAseq
Saitoella complicata	75	3	3,85	3,55	JGI
Lichtheimia hyalospora	77	1	1,28	3,53	JGI
Abeoforma whisleri	76	2	2,56	3,53	RNAseq
Branchiostoma floridae	76	2	2,56	3,44	Torruella et al. 2012
Spizellomyces punctatus	75	3	3,85	3,39	Torruella et al. 2012
Rhizopus oryzae	76	2	2,56	3,15	Torruella et al. 2012
Volvox carteri	77	1	1,28	3,11	Torruella et al. 2012
Xenopus tropicalis	75	3	3,85	2,83	Torruella et al. 2012
Capsaspora owczarzaki	77	1	1,28	2,71	Torruella et al. 2012
Capitella teleta	78	0	0	2,67	Torruella et al. 2012
Dictyostelium purpureum	77	1	1,28	2,55	Torruella et al. 2012
Selaginella moellendorffii	75	3	3,85	2,24	Torruella et al. 2012
Phycomyces blakesleeanus	77	1	1,28	2,23	Torruella et al. 2012
Dictyostelium discoideum	77	1	1,28	2,19	Torruella et al. 2012
Amoebidium parasiticum	77	1	1,28	1,32	Torruella et al. 2012
Nematostella vectensis	78	0	0	1,21	Torruella et al. 2012
Cryptococcus neoformans	77	1	1,28	1,12	Torruella et al. 2012
Tribolium castaneum	77	1	1,28	0,81	Torruella et al. 2012
Mortierella verticillata	77	1	1,28	0,71	Broad Institute
Mucor circinelloides	78	0	0	0,61	Torruella et al. 2012
Phanerochaete chrysosporium	78	0	0	0,38	Torruella et al. 2012
Batrachochytrium dendrobatidis	78	0	0	0,24	Torruella et al. 2012

Phylogenetic analyses

Best fitting protein evolutionary model for ML was chosen using Prottest 3 [15], which computes the likelihoods with PhyML [16] for ML.
Maximum likelood phylogenies were inferred using RAxML [13] under the LG model + Γ distribution with 4 discrete categories and invariant sites. Statistical support for the topology was estimated with 200 non-parametric bootstrap replicates. Bayesian phylogenies were inferred using Phylobayes [17] under the CAT-Poisson evolutionary model. Due to computational burdens, 2 independent MCMC chains for each dataset were run for ca. 1500 generations, saving every 10 trees and a burn-in of ~25%. In fact, runs were stop once convergence thresholds were achieved (basically maximum discrepancy < 0.1 and minimum effective size > 100 using bpcomp; but see Phylobayes manual for details).

Alternative topology test was performed using Mesquite to provide fixed constrained splits to RAxML (-g option), for which a best tree was computed. Then all trees likelihood (obtained with -f g option in RAxML) were input in CONSEL [18] to perform the AU test.

Table S2. Topology test using 9 alternative splits of *C. limacisporum* within Holozoa. In green significantly acceptable topologies, in red discarded hypotheses (p<0.05).

S83 PROTGAMMAILG	**rank**	**item**	**obs**	**au**	**(se)**	**np**	**(se)**
Corallochytrium,Ichthyosporea	1	7	-6,7	0,641	-0,009	0,39	-0,002
Corallochytrium,Filasterea	2	3	6,7	0,496	-0,01	0,331	-0,002
Corallochytrium,Sphaerothecum	3	9	9,7	0,42	-0,011	0,164	-0,001
Corallochytrium,Ichthyophonida	4	6	13,9	0,322	-0,011	0,112	-0,001
Corallochytrium,Holozoa	5	5	32,3	0,045	-0,007	0,007	0
(Corallochytrium,Filozoa)	6	4	32,3	0,045	-0,007	0,007	0
Corallochytrium,(Choanoflagellates,Metazoa)	7	2	95,4	0,002	-0,001	0,001	0
Corallochytrium,Metazoa	8	8	244	6,00E-50	0	9,00E-17	0
Corallochytrium,Choanoflagellates	9	1	235,4	2,00E-79	0	7,00E-22	0

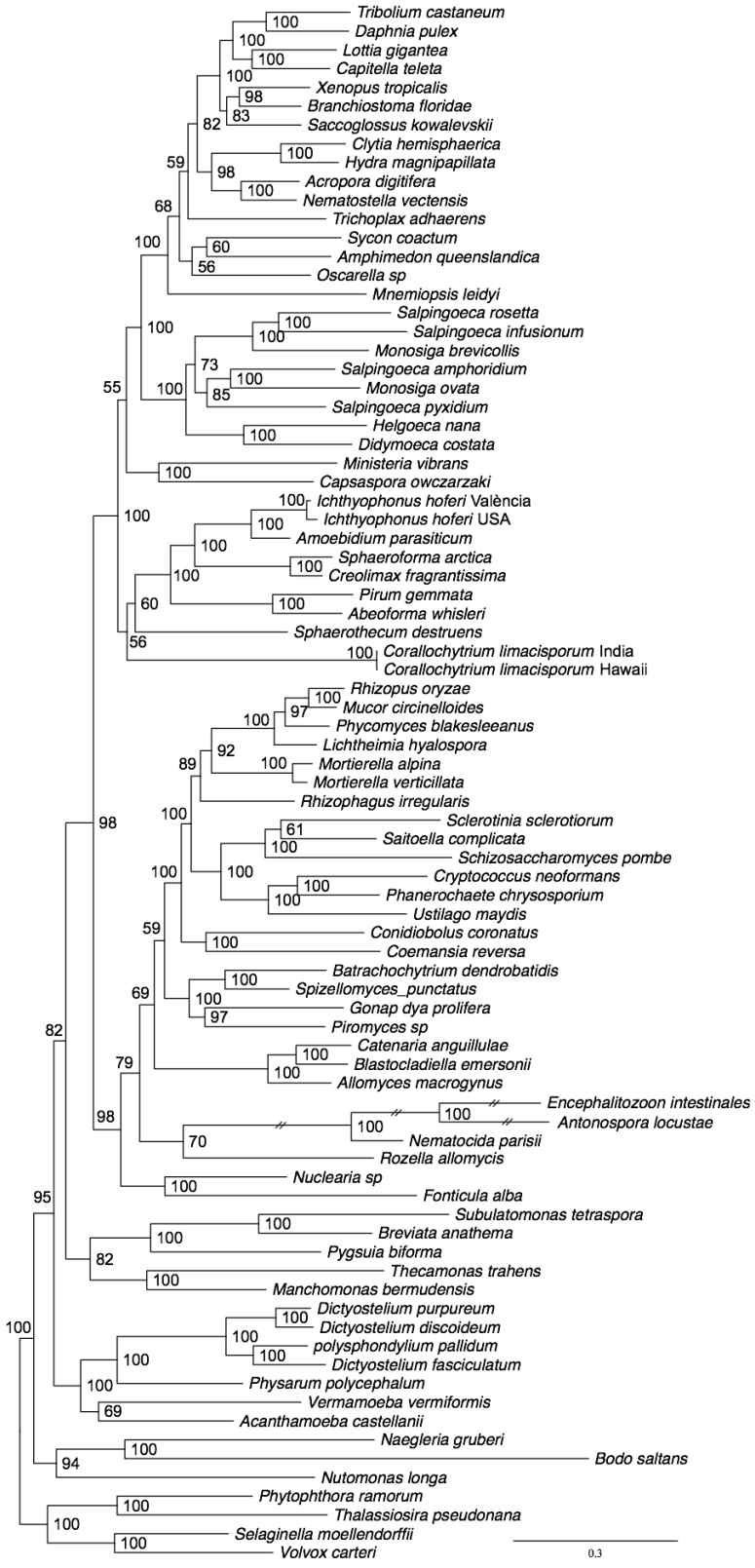
S70 PROTGAMMALG	**rank**	**item**	**obs**	**au**	**(se)**	**np**	**(se)**
Corallochytrium,Ichthyosporea	1	7	-6.3	0,52847	(0.008)	0,27361	(0.002)
Corallochytrium,Sphaerothecum	2	9	6.3	0,32569	(0.011)	0,17292	(0.001)
Corallochytrium,Ichthyophonida	3	6	9.1	0,28472	(0.011)	0,13611	(0.001)
Corallochytrium,Filasterea	4	3	32.3	0,16389	(0.009)	0,10694	(0.001)
Corallochytrium,Holozoa	5	5	40.9	0.049	(0.007)	0.010	(0.000)
(Corallochytrium,Filozoa)	6	4	43.2	0.034	(0.009)	0.002	(0.000)
Corallochytrium,Choanoflagellates	7	1	11,54	0.003	(0.016)	3,00E-06	(0.000)
Corallochytrium,Metazoa	8	8	89.5	4,00E-05	(0.000)	2,00E-05	(0.000)
Corallochytrium,(Choanoflagellates,Metazoa)	9	2	89.5	4,00E-05	(0.000)	2,00E-05	(0.000)

Phylogenetic Trees Pertaining to Experimental Procedures in newick format

Tree A

Unrooted maximum likelihood phylogenetic tree estimated from the 83-taxa matrix. Best tree was estimated using RAxML with LG+I+GAMMA model, and statistical support from 200 nonparametric bootstrap replicates.

Tree A =



Tree B

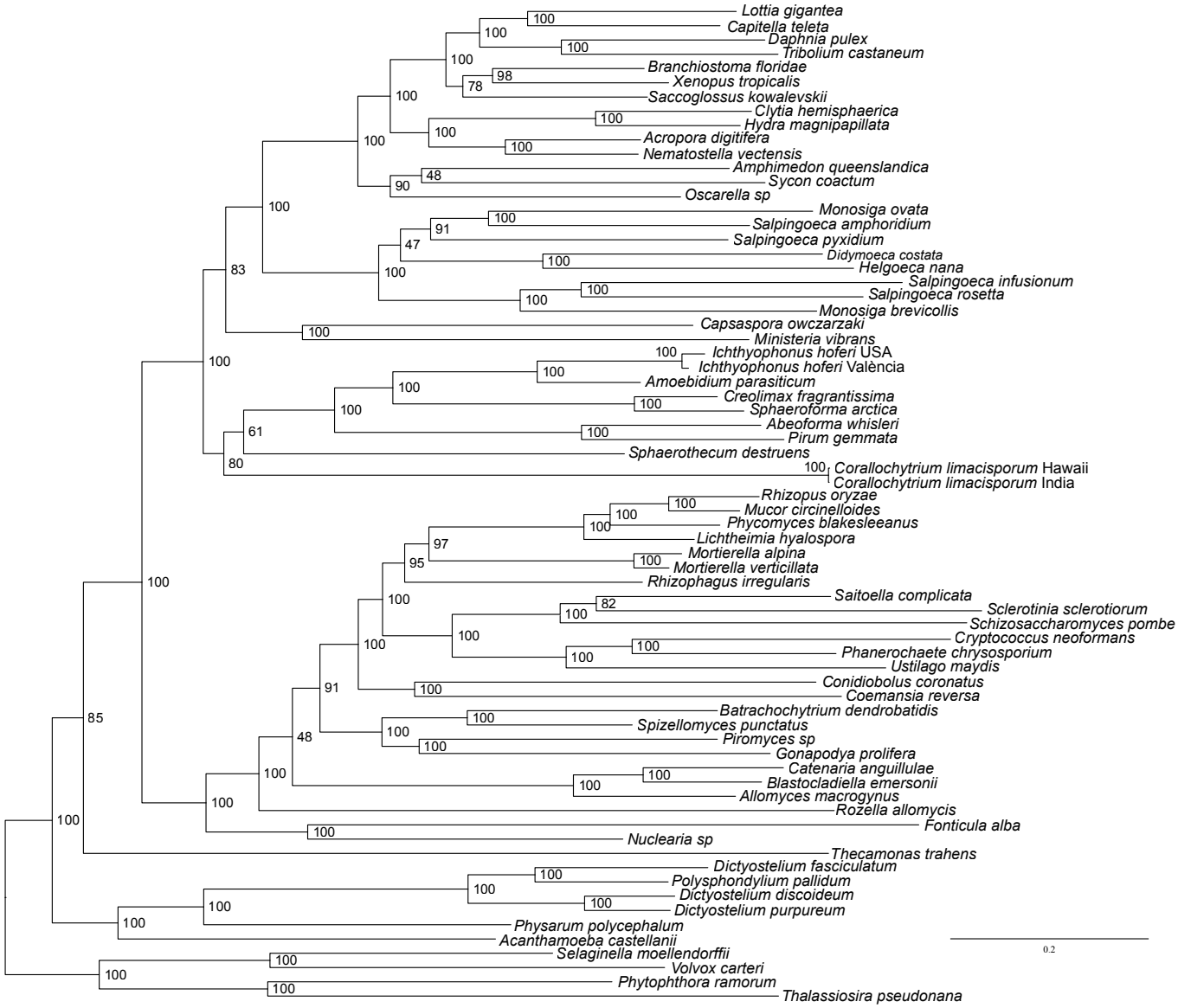
Unrooted Bayesian phylogenetic tree estimated from the 83-taxa matrix. Consensus tree was estimated from two MCMC chains with CAT Poisson model run for 1,500 generations, saving every 10 trees and after a burnin of 25%.

Tree B = The same as in figure 1

Tree C

Unrooted maximum likelihood phylogenetic tree estimated from the 70-taxa matrix. Best tree was estimated using RAxML with LG+GAMMA model, and statistical support from 200 nonparametric bootstrap replicates.

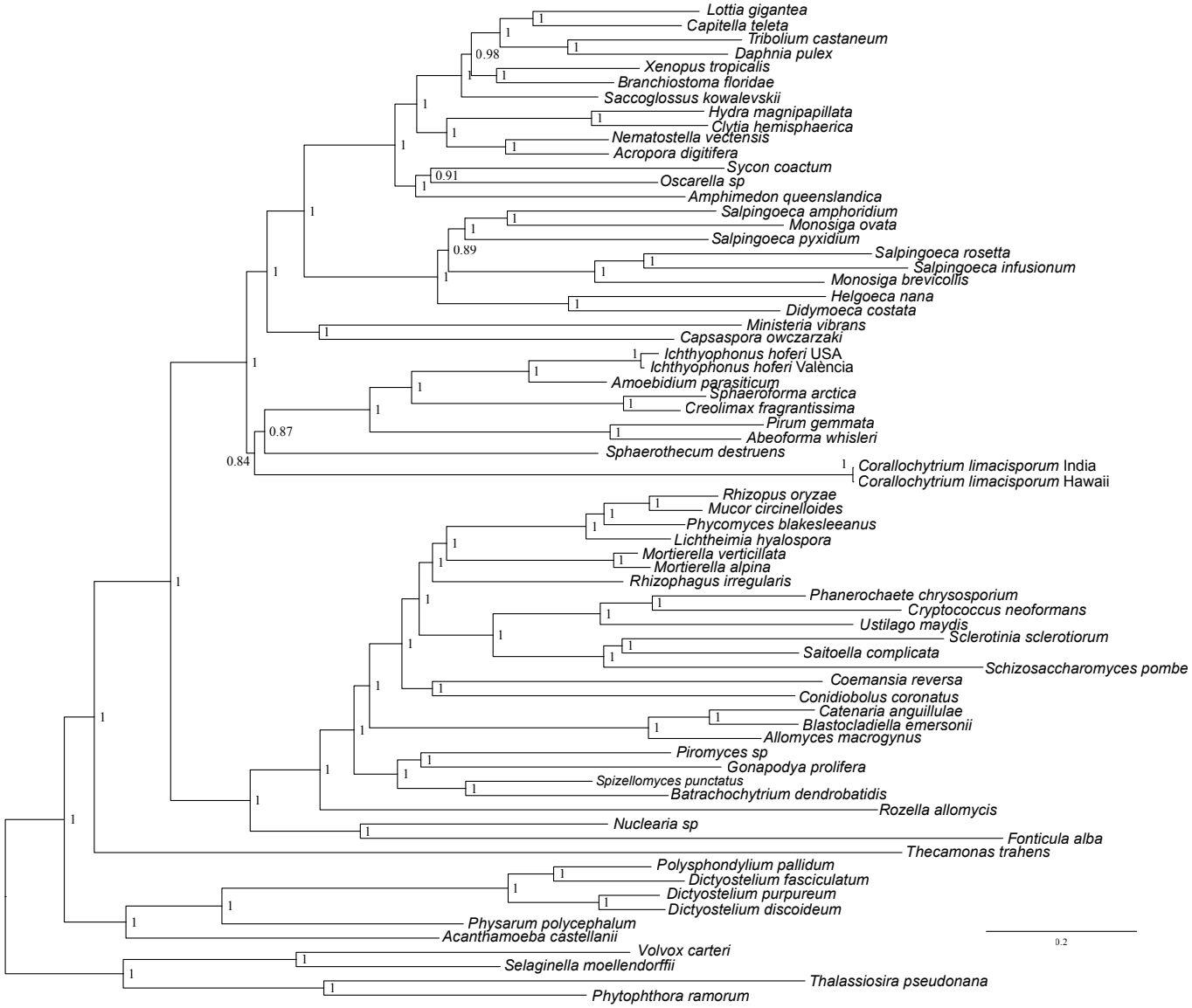
Tree C =



Tree D

Unrooted Bayesian phylogenetic tree estimated from the 70-taxa matrix. Consensus tree was estimated from two MCMC chains with CAT Poisson model run for 1500 generations, saving every 10 trees and after a burnin of 25%.

Tree D =



Gene family analyses

Identification of the flagellar toolkit genes was obtained using Best Reverse Blast Hit using *Homo sapiens* aminoacid sequences as queries. Best Reverse Blast retrieved a primary list of flagellar genes for each species, which were further filtered through manually inspecting multi-species alignments with MAFFT L-ins-I algorithm and performing RAxML trees (**Table S3 Flagellum toolkit**). For multigene families we performed HMMER 3.0 [19] searches against proteomes and 6 frame translations of the query transcriptomes using PFAM seed alignments. Multigene families analysed in this study include Kinesins (PF00225), Dynein Heavy Chains, Dynein Intermediate Chains, Dynein Light Chains (PF01221), Dynein Light Intermediate Chains (PF05783), Tubulins (PF00091) and Chitin synthases I (PF01644) and Chitin synthases II (PF03142). To gather putative fragmentary hits due to transcriptome miss-assembly we performed an additional search step using tBlastn. The hits obtained through the previous methods were aligned with MAFFT L-ins-I algorithm and trimmed manually. Phylogenetic inference was performed with RAxML using LG model and gamma distribution, while statistical support was computed with 100 bootstraps. Phylogenetic trees were validated against previously published topologies [20, 21]. Chitin synthases alignment restricted to the CHS domain and phylogenetic inference was also obtained with PhyloBayes, using two chains and sampling every 100 generations, stopping the process when the two chains converged (maxdiff < 0.3). Bayesian posterior probabilities (BPPs) were

used for assessing the statistical support of each bipartition. All the original alignments and trees are available upon request.

Table S4. Chitin synthase conserved motifs. Refered to [22]

		Substrate binding	Substrate binding	Catalytic base	Processivity	Processivity	Extra domain
	Chitin division I fungal consensus*	TMYNE	DXGT	LAEDRIL	QRRRW	(S/T)WG	Chitin_synth_1N (PF08407)
Class I/II/III	**Oomycetes**	**TMYNE**	**DVGT**	**LAEDRIL**	**QR(K/R)W**	**STW**	**Chitin_synth_1N (PF08407)** and MIT (PF0421
Class I/II/III	*Chlorella variabilis*	TLYNE	**DAGT**	LAEDRVL	**QRRRW**	**STW**	**Chitin_synth_1N (PF08407)**
Class I/II/III	*Paramecium tetraurelia*	TMYSE	DCGA	LAEDRVL	**QRRRW**	**STW**	-
Ichthyosporean class	*Salpingoeca rossetta*	PCYNE	**DCGT**	LAEDRFL	**QRRRW**	**STW**	-
Ichthyosporean class	*Phaeodactylum tricornutum*	PCYNE	DCGI	LAEDRFP	**QRRRW**	**STW**	-
Ichthyosporean class	*Acanthamoeba castellanii*	PHQDT	**DVGT**	LAEDRVL	**QRRRW**	**STW**	-
Ichthyosporean class	**Ichthyosporeans**	PF(Y/N)NE	**DCGT**	**LAEDR(I/V)L**	**QRRRW**	**STW**	SH2 (PF00017) only *C.fragrantissima*
	Chitin division II fungal consensus*	(T/P)(A/C) Y(S/T)E	DADT	LGEDR(YFE)L	Q(R/G)RRW	(S/T)WG	Myosin head (PF00063) and Cyt-b5 (PF00173)
Class IV/V/VII	***Corallochytrium limacisporum***	**TCY(S/T)E**	**DADT**	**LGEDR(Y/F)L**	**QRRRW**	**SWG**	**Myosin head (PF00063)**
Class VI	*Neurospora crassa* 4678NCU05268	*PCYKE*	DSDC	LGEDRWL	**QRRRW**	**TWG**	-
Class VI	*Spizellomyces punctatus* SPPG_00288	**PCYTE**	**DADT**	LGEDRRF	**QRRRW**	GWG	-
Class VI	*Ectocarpus siliculosus*	P(V/T)Y(N/K)E	(D/E)(G/S)DT	QGTDRRY	**QRRRW**	(A/G)W(G/V)	-
Metazoan class	*Entamoeba histolytica*	TMYHE	DGDT	FGEDRWL	**QRRRW**	**SWG**	-
Metazoan class	*Techamonas trahens*	TLWHE	DGDT	MGEDRWL	QRKRW	**SWG**	-
Metazoan class	**Ichthyosporeans**	T(L/M)WHE	DGD(V/T)	QGEDRWL	**QRRRW**	**SWG**	-
Metazoan class	*Salpingoeca rossetta*	TMYNE	DADI	MGEDRWL	**QRRRW**	**SWG**	SAM2
Metazoan class	**Filastereans**	TMWRE	DGD(V/I)	(M/Q)GEDRWL	**QRRRW**	(S/R)WG	-
Metazoan class	**Metazoans**	TMWHE	DGDV	QGEDR(S/R)(F/L	**QRRRW**	**SWG**	**Myosin head (PF00063)** only *Lottia gigantea*

Immunostaining

For both S. rosetta and C. owczarzaki, cells were fixed for 5min with 6% acetone and for 15minwith 4% formaldehyde. The coverslips were washed gently four times with 100mM Pipes at pH 6.9, 1mM EGTA, and 0.1mM MgSO4 (PEM), incubated for 30min in blocking solution (PEM+:1%BSA, 0.3% Triton X-100), 1h in primary antibodies solution (in PEM+), and after further washes (PEM+), 1h in the dark with fluorescent secondary antibodies (1:100 in PEM+,Alexa Fluor 488 goat anti-mouse, and Alexa Fluor 568 goat anti- rabbit; Invitrogen) and washed again four times (PEM). To visualize F-actin coverslips were incubated for 15min in the dark with rhodamine phallidin (6 U/ml in PEM; Molecular Probes). After 3 washes (PEM), coverslips were mounted onto slides with Fluorescent Mounting Media (4mL; Prolong Gold Antifade, Invitrogen). The following primary antibodies have been used: mouse monoclonal antibody against b-tubulin (E7, 1:400; Developmental Studies Hybridoma Bank); mouse monoclonal antibody against Fascin (ab78487, 1:100; Abcam). Images were taken with a 100? oil immersion objective on an inverted Leica microscope

Table S3, related to figure 2. Flagellar toolkit.

	Monosiga brevicollis	*Salpingoeca rossetta*	*Capsaspora owczarzaki*	***Ministeria vibrans***	*Creolimax fragrantissima*	*Sphaeroforma arctica*	*Abeoforma whisleri*	*Pirum gemmata*	*Amoebidium parasiticum*	*Ichthyophonus hoferi*	***Sphaerotecum destruens***	***Corallochytrium limacisporum***	*Rhizophagus irregularis*	*Coemansia reversa*	*Conidiobolus coronatus*	*Allomyces macrogynus*	*Piromyces sp.*	*Spizellomyces punctatus*	*Gonapodia prolifera*	*Encephalitozoon cuniculi*	*Rhozella allomycis*	*Nuclearia sp.*	*Fonticula alba*	*Thecamonas trahens*	*Dictyostelium discoideum*	*Chlamydomonas reindhardtii*	
Epsilon Tubulin	1	1	0	1	0	0	0	0	0	0	1	1	0	0	0	1	0	1	1	0	1	0	0	1	0	1	
Delta Tubulin	1	1	0	0	0	0	0	0	0	0	0	1	0	0	0	1	1	0	1	0	1	0	0	0	0	1	
SAK/PLK4	0	0	0	0	0	0	0	0	1	0	1	1	0	1	0	0	1	1	1	0	1	0	0	0	0	0	centriole kinase
SAS6	1	1	0	1	0	0	1	1	1	0	1	1	0	0	0	1	1	1	1	1	0	0	0	0	0	1	Centriolar structure and percentriolar material
SAS4/CPAP	1	1	0	1	0	0	0	0	1	0	1	1	0	0	0	1	1	1	1	1	1	0	0	1	0	1	
D10/CEP135	0	1	0	1	0	0	0	0	1	1	1	1	1	1	0	1	1	1	1	0	1	0	0	1	0	1	
POC1	1	1	0	0	0	0	0	0	0	1	0	1	0	0	0	1	0	1	1	0	1	0	0	0	0	1	
Centrin-2	1	1	0	1	1	1	0	1	1	0	1	1	1	0	0	1	0	1	1	0	1	0	0	1	0	1	
CEP110	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
CEP164	1	1	0	0	0	0	0	0	0	0	1	1	0	0	0	1	1	1	1	0	1	0	0	1	0	1	
PD2/CEP192	1	1	0	1	0	0	0	0	0	0	0	1	1	0	0	0	0	1	0	0	0	0	0	1	1	0	
rless/CEP152	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
BBS1	1	1	0	0	0	0	0	0	0	0	1	0	0	0	0	1	1	0	0	0	0	0	0	1	0	1	BBSome and MKS
BBS4	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	1	0	1	
BBS5	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	
BBS8	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	1	0	1	
BBS9/PTHB1	1	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	1	0	1	
BBS2	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	1	0	1	
BBS3/ARL6	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	1	0	1	0	0	0	0	0	0	0	1	
BBS7	1	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	1	0	1	
MKS1	1	1	0	0	0	0	0	0	0	0	1	0	0	0	0	0	1	1	1	0	0	0	0	1	0	1	
T144/WDR19	1	1	0	0	0	0	0	0	0	0	0	1	0	0	0	1	1	1	1	0	1	0	0	1	0	1	IFT-A (retrograde) with dynein
IFT43	1	1	0	0	0	0	0	0	0	0	1	1	0	0	0	0	1	1	0	0	0	0	0	1	0	1	
T139/TTC21B	1	1	0	1	0	0	0	0	0	0	1	1	0	0	0	0	1	1	0	0	0	0	0	1	0	1	
IFT122	1	1	0	1	0	0	0	0	0	0	0	1	0	0	0	1	1	1	1	0	1	0	0	1	0	1	
T121/WDR35	1	1	0	1	0	0	0	0	0	0	1	1	0	0	0	1	1	1	1	0	1	0	0	1	0	1	
IFT140	1	1	0	1	0	0	0	0	1	0	1	1	0	0	0	1	1	1	1	0	1	0	0	1	0	1	
IFT70	1	1	0	0	0	0	0	0	0	0	0	1	0	0	0	1	1	1	1	0	1	0	0	1	0	1	IFT-B (anterograde) with kinesin-2
54/TRAF3IP1	1	1	0	0	0	0	0	0	0	0	0	1	0	0	0	1	1	1	1	0	1	0	0	1	0	1	
IFT46	1	1	0	0	0	0	0	0	0	0	1	0	0	0	0	1	1	1	1	0	1	0	0	1	0	1	
IFT27	1	1	0	1	0	0	0	0	0	0	0	0	0	0	0	1	1	1	0	0	0	0	0	0	0	1	
IFT25	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	0	0	0	0	0	1	0	1	
IFT20	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	1	0	1	
IFT52	1	1	0	0	0	0	0	0	0	0	1	1	0	0	0	1	1	1	1	0	1	0	0	1	0	1	
IFT57	1	1	0	0	0	0	0	0	0	0	1	1	0	0	0	1	1	1	1	0	1	0	0	1	0	1	
IFT80	1	1	0	1	0	0	0	0	0	0	1	0	0	0	0	1	1	1	1	0	0	0	0	1	0	1	
IFT88	1	1	0	0	0	0	0	0	0	0	1	1	0	0	0	1	1	1	1	0	1	0	0	1	0	1	
IFT172	1	1	0	0	0	0	0	0	0	0	1	0	0	0	0	1	1	1	1	0	1	0	0	1	0	1	
IFT72/74	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	1	0	1	0	0	1	0	1	
IFT81	1	1	0	0	0	0	0	0	1	0	1	1	0	0	0	1	1	1	1	0	1	0	0	1	0	1	
CLUAP1	1	1	0	1	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	0	1	0	0	1	0	1	
TTC26	1	1	0	1	0	0	0	0	0	0	1	1	0	0	0	1	0	1	1	0	1	0	0	1	0	1	
Hydin	1	1	0	0	0	0	0	0	0	0	1	0	0	0	0	1	1	1	0	0	1	0	0	1	0	1	central mt
DYNC1H1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	0	Cytoplasmic Dynein-1
ytoplasmatic	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	
ytoplasmatic	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	0	1	1	1	1	1	0	
DYNC2H1	1	1	0	1	0	0	0	0	0	0	1	0	0	0	0	0	1	1	0	1	0	0	0	1	0	1	Cytoplasmic Dynein-2
ediate_Chain	1	1	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	
C_axonemal	1	1	0	0	0	0	0	0	0	0	1	0	0	0	0	1	0	1	1	0	1	0	0	1	0	1	
C_axonemal	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	0	0	0	0	0	0	0	0	
*alpha*_DYH5	1	1	0	1	0	0	0	0	0	0	1	0	0	0	0	1	1	1	1	0	1	0	0	1	0	1	Outer Dynein
A*beta*_DYH9	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	1	0	0	1	0	0	
IDA-1 *alpha*	1	1	0	1	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	0	1	0	0	1	0	1	double-headed
IDA-1 *beta*	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	0	1	0	0	0	0	1	
IAD3_DHY7	1	1	0	1	0	0	0	0	0	0	1	0	0	0	0	1	1	1	1	0	1	0	0	0	0	1	single-headed inner arms
IAD4_DHY1	1	1	0	1	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	0	1	0	0	1	0	1	
IAD5	1	1	0	0	0	0	0	0	0	0	1	0	0	0	0	1	1	1	1	0	1	0	0	0	0	1	
HEATR2	1	1	0	1	0	0	0	0	0	0	1	1	0	0	0	1	1	1	1	0	0	0	0	1	0	1	
Kinesin II	1	1	0	0	0	0	0	0	0	0	1	1	0	0	0	1	1	1	1	0	1	0	0	1	0	1	Kinesin Motors
Kinesin 13	1	1	0	0	0	0	1	1	1	0	1	1	1	1	0	1	1	1	1	1	1	0	0	1	1	1	
Kinesin 9	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	0	1	0	0	1	0	1	
Protein/KAP	1	1	0	0	0	0	0	0	0	0	0	1	0	0	0	1	1	1	1	0	1	0	0	1	0	1	

Supplemental References

1. Marshall, W. L., and Berbee, M. L. (2011). Facing Unknowns : Living Cultures (Pirum gemmata gen. nov., sp. nov., and Abeoforma whisleri , gen. nov., sp. nov.) from Invertebrate Digestive Tracts Represent an Undescribed Clade within the Unicellular Opisthokont Lineage Ichthyosporea (Mesomycetozoea. Protist *162*, 33–57.
2. Raghu-Kumar, S., Chandramohan, D., and Ramaiah, N. (1987). Contribution of the Thraustochytrid Corallochytrium limacisporum Raghu-kumar to Microbial Biomass in Coral Reef Lagoons. Indian J. Mar. Sci. New Delhi *16*, 122–125.
3. Glücksman, E., Snell, E. a, and Cavalier-Smith, T. (2013). Phylogeny and evolution of Planomonadida (Sulcozoa): Eight new species and new genera Fabomonas and Nutomonas. Eur. J. Protistol. *49*, 179–200.
4. Grabherr, M. G., Haas, B. J., Yassour, M., Levin, J. Z., Thompson, D. a, Amit, I., Adiconis, X., Fan, L., Raychowdhury, R., Zeng, Q., et al. (2011). Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nat. Biotechnol. *29*, 644–652.
5. Chevreux, B., Wetter, T., and Suhai, S. (1999). Genome sequence assembly using trace signals and additional sequence information. Comput. Sci. Biol. Proc. Ger. Conf. Bioinforma., 45–56.
6. Torruella, G., Derelle, R., Paps, J., Lang, B. F., Roger, A. J., Shalchian-Tabrizi, K., and Ruiz-Trillo, I. (2012). Phylogenetic relationships within the Opisthokonta based on phylogenomic analyses of conserved single copy protein domains. Mol. Biol. Evol. *29*, 531–544.
7. Suga, H., and Ruiz-Trillo, I. (2013). Development of ichthyosporeans sheds light on the origin of metazoan multicellularity. Dev. Biol. *377*, 284–292.
8. Minge, M. a, Silberman, J. D., Orr, R. J. S., 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. R. Soc. B Biol. Sci. *276*, 597–604.
9. Grant, J. R., Lahr, D. J. G., Rey, F. E., Burleigh, J. G., Gordon, J. I., Knight, R., Molestina, R. E., and Katz, L. a. (2012). Gene discovery from a pilot study of the transcriptomes from three diverse microbial eukaryotes: Corallomyxa tenera, Chilodonella uncinata, and Subulatomonas tetraspora. Protist Genomics *1*, 3–18.
10. 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*, 3059–66.
11. Criscuolo, A., and Gribaldo, S. (2010). BMGE (Block Mapping and Gathering with Entropy): a new software for selection of phylogenetic informative regions from multiple sequence alignments. BMC Evol. Biol. *10*, 1–21.
12. Struck, T. H. (2014). TreSpEx—Detection of Misleading Signal in Phylogenetic Reconstructions Based on Tree Information. Evol. Bioinforma. *10*, 51–67.
13. Stamatakis, A. (2006). RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. Bioinformatics *22*, 2688–90.
14. Gouy, M., Guindon, S., and Gascuel, O. (2010). SeaView version 4: A multiplatform graphical user interface for sequence alignment and phylogenetic tree building. Mol. Biol. Evol. *27*, 221–4.
15. Darriba, D., Taboada, G. L., Doallo, R., and Posada, D. (2011). ProtTest 3: fast selection of best-fit models of protein evolution. Bioinformatics *27*, 1164–5.
16. Guindon, S., and Gascuel, O. (2003). A Simple, Fast, and Accurate Algorithm to Estimate Large Phylogenies by Maximum Likelihood. Syst. Biol. *52*, 696–704.
17. Lartillot, N., and Philippe, H. (2004). A Bayesian mixture model for across-site heterogeneities in the amino-acid replacement process. Mol. Biol. Evol. *21*, 1095–109.
18. Shimodaira, H., and Hasegawa, M. (2001). CONSEL: for assessing the confidence of phylogenetic tree selection. Bioinformatics *17*, 1246–7.
19. Eddy, S. R. (2011). Accelerated Profile HMM Searches. PLoS Comput. Biol. *7*, e1002195.
20. Wickstead, B., and Gull, K. (2012). The evolutionary biology of dyneins. In Dyneins: Structure, biology and disease, Elsevier Inc., ed. (London, UK), pp. 89–121.
21. Zakrzewski, A.-C., Weigert, A., Helm, C., Adamski, M., Adamska, M., Bleidorn, C., Raible, F., and Hausen, H. (2014). Early divergence, broad distribution, and high diversity of animal chitin synthases. Genome Biol. Evol. *6*, 316–25.
22. James, T. Y., and Berbee, M. L. (2012). No jacket required--new fungal lineage defies dress code: recently described zoosporic fungi lack a cell wall during trophic phase. Bioessays *34*, 94–102.

4. DISCUSSION

Kaneda: *what the hell is this Akira?*
Kei: *All I know is what Roy said. He said Akira has achieved ultimate energy.*
Kaneda: *ultimate energy?*
Kei: *A human being, you know, achieves a whole lot of things in a lifetime, right? Like discovering and making things like houses, motorbikes, bridges, towns, rockets… Where does that tremendous knowledge and energy come from?*
Kaneda: *mm-um. (Puts his hands in the air)*
Kei: *After all, humans descend from monkeys, right? And before that, like reptiles and fish. And even before that, plankton and amoebas. Creatures like those have incredible amounts of energy inside.*
Kaneda: *Well. I, uh, that's evolution. It's in the genes right?*
[…]
Kei: *And what if everyone shares those ancient memories, what if there were some mistakes in the progression. Then something goes wrong like an amoeba is suddenly given the higher powers than a human has.*
Kaneda: *What? Is that what Akira is?*
Kei: *Amoeba's don't make motorcycles and atomic bombs! They only eat up anything that happens into their way.*

Akira 1988

4.1 Evaluation of phylogenomic methods to infer ancient speciation events

Classical morphological datasets, single gene trees, and, probably, rare genomic changes are not suitable to analyze deep eukaryotic phylogenetic nodes (**section 1.4**). A clear example of the latest is stated with the reanalysis of the presence of the *AAR* gene across the Opisthokonta, a proposed synapomorphy between fungi and the enigmatic protist *Corallochytrium limacisporum* (**section 1.2.4**) presented in **section 3.1 R1**. The taxonomic sampling of the original study (Sumathi et al. 2006) missed members from several holozoan lineages. Thus, when representatives from the Filasterea, Ichthyosporea and Choanoflagellata were included, the story changed completely (Torruella et al. 2009). All sampled holozoan lineages possess the gene; therefore the gene was at least present in the Last Opisthokonta Common Ancestor (LOCA), and lost in animals. Moreover, the wider study on distinct genes of the pathway showed patchy distributed gene trees that could be examples of hidden paralogy or HGT.

In order to properly establish a robust phylogenetic backbone of Opisthokonta, phylogenomics methods were used with the new genome data generated through the UNICORN project (Ruiz-Trillo et al. 2007). In particular, we developed a new dataset based on Single Copy Protein Domains, rather than full proteins (**section 3.2 R2**). (Torruella et al. 2012) The aim was to minimize orthology assignment problems related to deep speciation events, while maximizing the taxon occupancy and symmetry. Most interestingly, the dataset was composed of independent markers compared to the majority of previous studies addressing similar issues. This novel dataset was thoroughly tested to overcome systematic errors. Among the different methodologies, we found that using CAT evolutionary model (Lartillot and Philippe 2004) and removing distant outgroups were indeed very useful, because it diminished the saturation and probably other sources of systematic error such as compositional heterogeneity. Similarly, excluding the most variable sites as classified by the gamma category (a proxy to exclude the fastest evolving sites) or recoding amino acids into functional categories should also minimize all these problems (Rodríguez-Ezpeleta, Brinkmann, Roure, et al. 2007) but in our case they had no clear impact. Other strategies to minimize systematic error in Maximum Likelihood, such as using distinct evolutionary models, partitioning the data or excluding taxa with substantial percentages of missing data, had no influence at all on the resulting topologies.

As taxonomic sampling seemed to clearly influence Opisthokonta phylogeny, our next step was to dedicate our efforts to dramatically increase the genomic information available for the opisthokonts. Thus, we obtained RNAseq from 10 novel opisthokont taxa and included as well new taxa obtained by other authors, while substituting most EST-based taxa for more complete transcriptomic and genomic data. Our latest phylogenomic analysis (**section 3.3 R3**) provides a more robust phylogenetic framework for Opisthokonta (**Figure 16**) that allows us to infer evolutionary hypotheses. This analysis confirms that Choanomonada is sister to Metazoa and corroborates its internal relationships (Nitsche et al. 2011). The Ichthyosporea and Filasterea are clearly independent lineages, and therefore the Filozoa hypothesis is confirmed (Shalchian-Tabrizi et al. 2008). Finally, and although not strongly supported, the enigmatic *C. limacisporum* appears as sister group to the Ichthyosporea (together as the earliest branching holozoan lineage). Within Ichthyophonida, the paraphyly between APCT and spherical groups proposed by SSU rDNA is confirmed (Glockling et al. 2013) (see also **Figure 18**).

Within the Holomycota, Nucleariids are shown to be the earliest branching lineage followed by Opisthosporida (only represented by *Rozella* and Microsporidia) and “fungi”. The Opisthosporidia still needs phylogenomic validation by including Aphelida representatives, as some ribosomal trees show paraphyly between Aphelida and Cryptomycota (Karpov, Mamkaeva, Benzerara, et al. 2014). As mentioned in **section 1.2.6**, both Aphelida and Cryptomycota/Rozellida groups are very similar in terms of morphology and life cycle, although further studies are needed in ultrastructure and molecular comparison to understand if these similarities are true. Within traditional “fungi”, Chytridiomycota and Blastocladiomycota are mostly recovered as paraphyletic although the branching order is still not clear.

Finally, regarding the outgroup, our analysis confirm Apusomonadida and Breviatea as paraphyletic (Brown et al. 2013), with clear evolutionary implications (see **section 4.2.1**). More interestingly, Ancyromonadida does not branch with Apusomonadida, where they have been traditionally placed (Glücksman et al. 2013; Paps et al. 2013).

Classifications based on morphology, molecular synapomorphies or SSU rDNA:

- *C. owczarzaki* as *Nuclearia* (Owczarzak et al. 1980)

- *C. limacisporum* as Thraustochytrid (Raghu-Kumar et al. 1987)

- Amoebidiales & Eccrinales as Trichomycetes (Lichtwardt et al. 2001)

- Aphelida as Ichthyosporea (Adl et al. 2005)

- *Nuclearia* as Filosea (Patterson 1983, 1999)

- Apusomonadida and Ancyromonadida as Apusozoa (Cavalier-Smith et al. 2008)

- *M. vibrans* & *C. limacisporum* as *incertae sedis* Holozoa (Cavalier-smith and Chao 2003)

Classification based on phylogenomics:

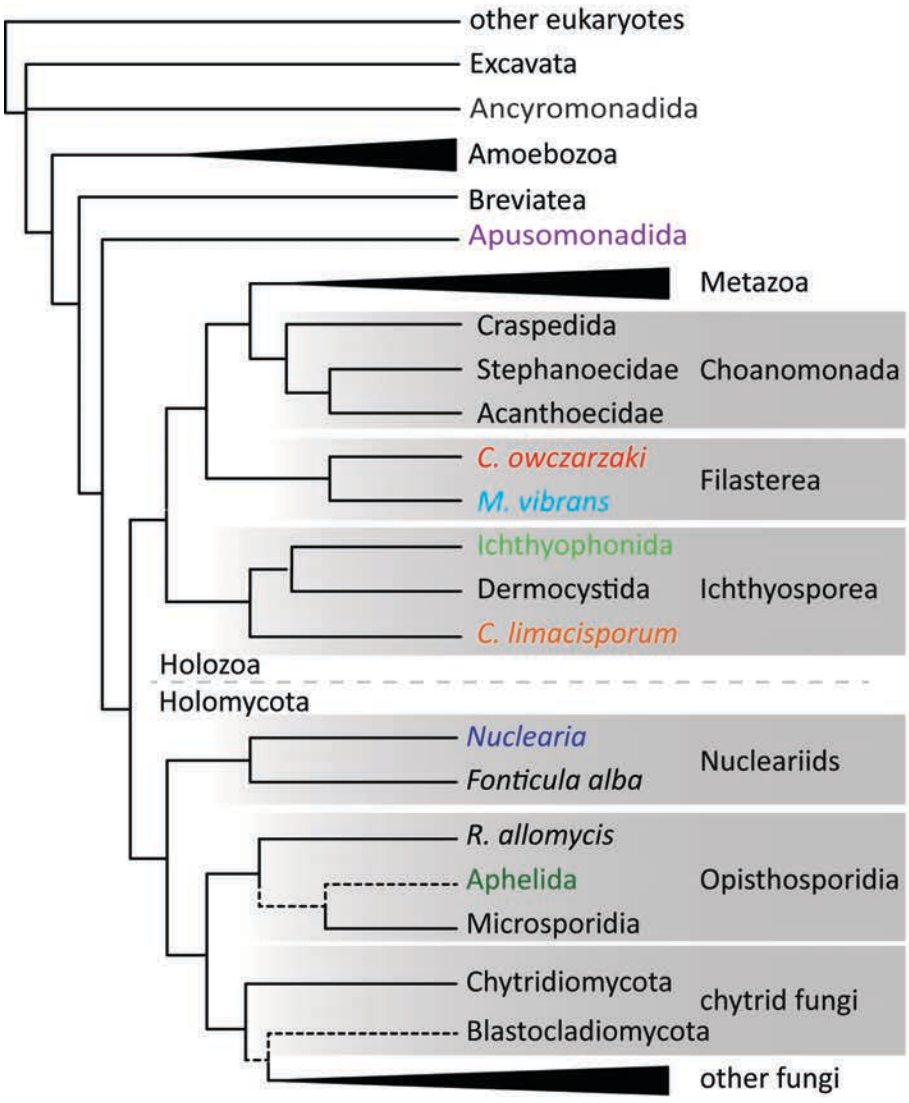


Figure 16. Some examples of original classifications for distinct opistocont protists based on morphology, molecular synapomorphies or SSU rDNA (top) the state-of-the-art classification based on phylogenomics (bottom).

4.2 Evolutionary Hypotheses on the Opisthokonta

The phylogenetic framework generated in this PhD thesis, together with the amount of biological knowledge collected during the past few years **(Figure 16)**, allows us to infer some evolutionary hypotheses. An important observation derived from the branching order is how morphologically different are some of the sister lineages within the Opisthokonta. In contrast, such difference is not observed among the more ancestral biflagellated outgroups (Beviatea and Apusomonadida), probably due to lifestyle constrains (Cavalier-Smith 2013). Another interesting fact is the similarity observed between some holozoan and holomycotan lineages. Thus, the holozoan filastereans resemble much more in terms of morphology and lifestyle to the holomycotan nucleariids than to the rest of holozoans (**section 4.2.3.1**). Moreover, the holozoan ichthyosporeans and *Corallochytrium* share many similarities with the holomycotan chytrid fungi (or even other osmotrophic eukaryotes such as the oomycetes) (**section 4.2.3.1**).

4.2.1 Reconstructing the ancestral character states

Reconstruction of the ancestral character states for each lineage may provide additional information regarding the evolution of morphological characters and lifestyles within the Opisthokonta. The thing is that if we want to infer how the Last Opisthokonta Common Ancestor (LOCA) was, we can not just compare its closest outgroup (i.e., Apusomonadida) and the earliest branching Holozoa or Holomycota lineage (Ichthyosporea or Nucleariids, respectively), but we need to infer ancestral states for all distinct Opisthokonta lineages, as well as in the distinct outgroups (**section 1.3**): Apusomonadida, Breviatea, even Amoebozoa and further back if possible (Cavalier-Smith 2013), tracing back the very root of eukaryotes (although the latter is out of the scope of this work).

Regarding the outgroups, both Apusomonadida (Cavalier-Smith and Chao 2010) and Breviatea (Brown et al. 2013) are small amoeboid phagotrophic lineages with two flagella (one for gliding motility and the other for feeding), and a more or less complex microtubular cytoskeleton (Walker et al. 2006; Heiss et al. 2011; Brown et al. 2013; Heiss et al. 2013) associated with a specific ventral feeding mode on benthonic areas or surfaces (**section 1.3**). These amoeboid cells are complex and versatile, in the sense that the same individual cell expresses flagellar structures, or filopodia; distinct cell shapes that can be remodelled several times according to environmental needs until the cell divides. However, the knowledge on the complete life cycles of these organisms is limited. Although breviates are specialized in microaerophylic environments it seems both

paraphyletic lineages may come from marine benthic bacteriophagous ancestors as SSU rDNA phylogenies suggest (see **Figure 20**).

The distinct **choanoflagellate** lineages so far described (**section 1.2.1**) are quite similar in their basic bauplan and lifestyle. In particular, the Acanthoecidae share the siliceous lorica with the Stephanoecidae, but a life cycle comprising a sedentary vegetative stage and a swimmer dispersal stage with the Craspedida. Thanks to a solid phylogenetic backbone (Nitsche et al. 2011; **section R3**) and comparative studies (Leadbeater 2008), it seems clear that choanoflagellates were ancestrally small marine sedentary suspension filter feeders similar to current craspedids (although migration to fresh-water and reversions may have happened multiple times) (Carr et al. 2008). The ancestral choanoflagellate probably had no siliceous lorica, which is likely a secondary acquisition of the Acanthoecida through HGT from diatom silicon transporters (Marron et al. 2013). The lorica allowed successful diversification in pelagic marine environments where Stephanoecidae have few competitors (Leadbeater 2008). It is more parsimonious as well to infer that the organic theca was ancestral to the group, with multiple secondary losses (Nitsche et al. 2011). Similarly, the ancestral choanoflagellate probably had already a complex and dynamic life cycle (Dayel et al. 2011) with swimmer, feeder, gamete, clonal colonial cell types (Fairclough et al. 2013). Thus, the non-loricate bauplan of the Craspedida with an organic theca and a complex life cycle is most likely the ancestral choanoflagellate form (**Figure 17**).

The **Filasterea** (**section 1.2.2**) are especially difficult to discuss in terms of ancestry, as only two or three species (two genera, in any case) have so far been described. Both are small naked filose amoeba. *Ministeria vibrans* is a marine phagotrophic benthic bacterivore, while *Capsaspora owczarzaki* is a fresh-water phagotroph with no clear ecological nature. We have also shown that *M. vibrans* has a stalk-like flagellum (Figure 2 in (Cavalier-Smith and Chao 2003)), which probably is ancestral given that the flagellar machinery is conserved across all eukaryotes (Carvalho-Santos et al. 2011) (**section 3.3 R3**). If *M. vibrans* is solely static, it may represent a similar form of lifestyle as in choanoflagellates: suspension filter feeders (Cavalier-Smith 2013), but using distinct molecular structures as stalks (choanoflagellate stalks are not flagellar-structures) (**Figure 17**). Indeed, both filastereans have filopodia-based crawling motility, similar to the choanoflagellate *Salpingoeca rosetta,* which uses filopodia to attach to substratum before creating the theca and stalk (Dayel et al. 2011). The aggregative behaviour and dormant cystic stages of *C. owczarzaki* has not been described in *M. vibrans*, so we cannot distinguish whether those characters are ancestral or not, or if they are homologous to other holozoans. Thus, the ancestral filasterean was probably a filopodiated amoeba with some sort of derived flagellum.

The earliest branching lineage of Holozoa is composed of ***Corallochytrium* plus Ichthyosporea** (**sections 1.2.3** and **1.2.4**). Within Ichthyophonida (the ichthyosporean amoeboid family), the paraphyly of the APCT and the "spherical" groups (**Figure 18**) leads to treat their shared traits as ancestral. *Abeoforma*, *Pirum*, *Creolimax* and *Sphaeroforma* are all marine spherical schizonts with no clear parasitic nature (Glockling et al. 2013). In contrast, *Psorospermium*, *Ichthyophonus*, Amoebidiales and Eccrinales have more complex shapes (i.e., polarized growth), probably as a result of their host-dependent lifestyle. Within Dermocystida (the ichthyosporean flagellated family), *Sphaerothecum destruens* seems to be the earliest divergent lineage (Vilela and Mendoza 2012). It is indeed the phylotype with large host-range, clear flagellated stage and the smaller schizonts of the whole group. It has a rosette-shaped colonial form that reminds to the ones in *Corallochytrium limacisporum*. Positioned as sister to the Ichthyosporea, *C. limacisporum* looks like the missing link between both families, since it is a small marine free-living walled schizont with both flagellated and amoeboid forms (**section 3.3 R3**).

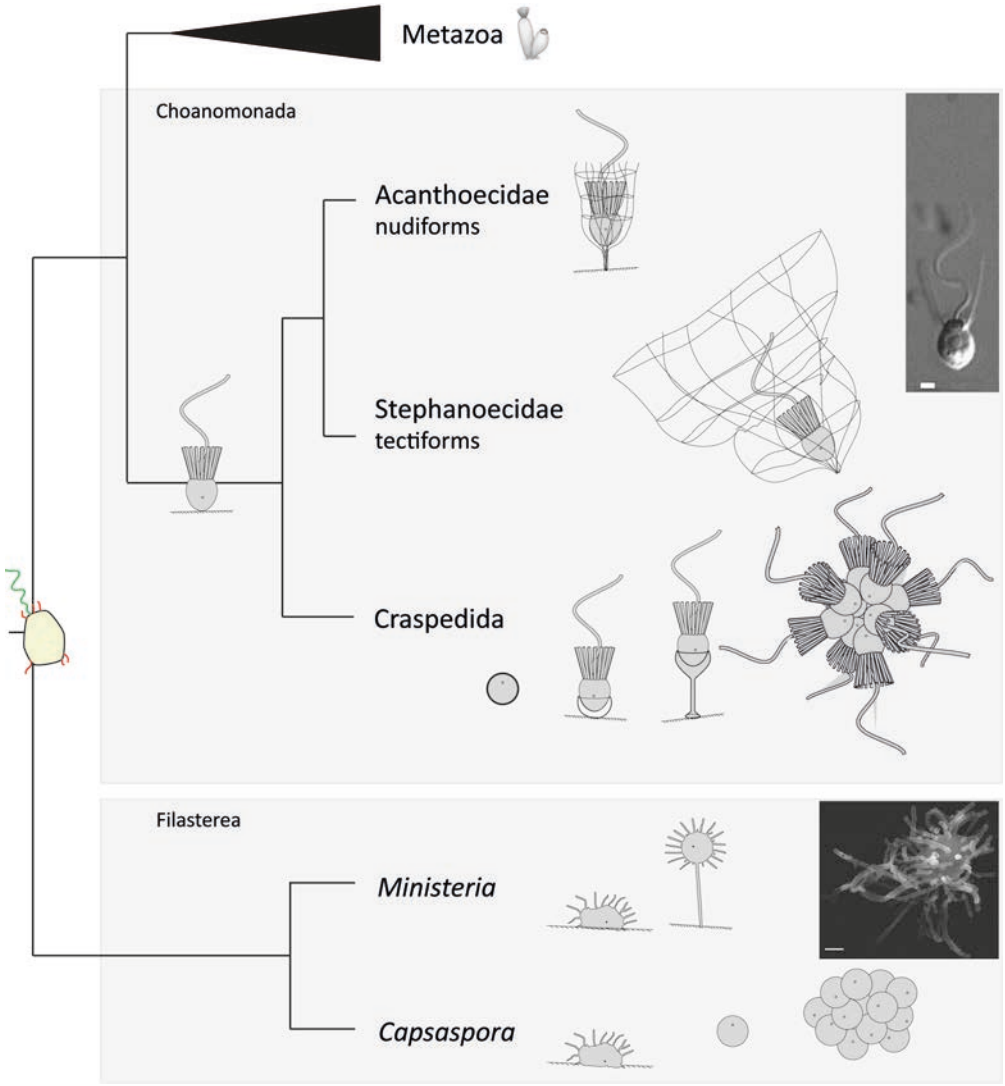


Figure 17.Filozoa ancestral character states. Their Last Common Ancestor probably had a single flagellum (green) and filopodia (red).

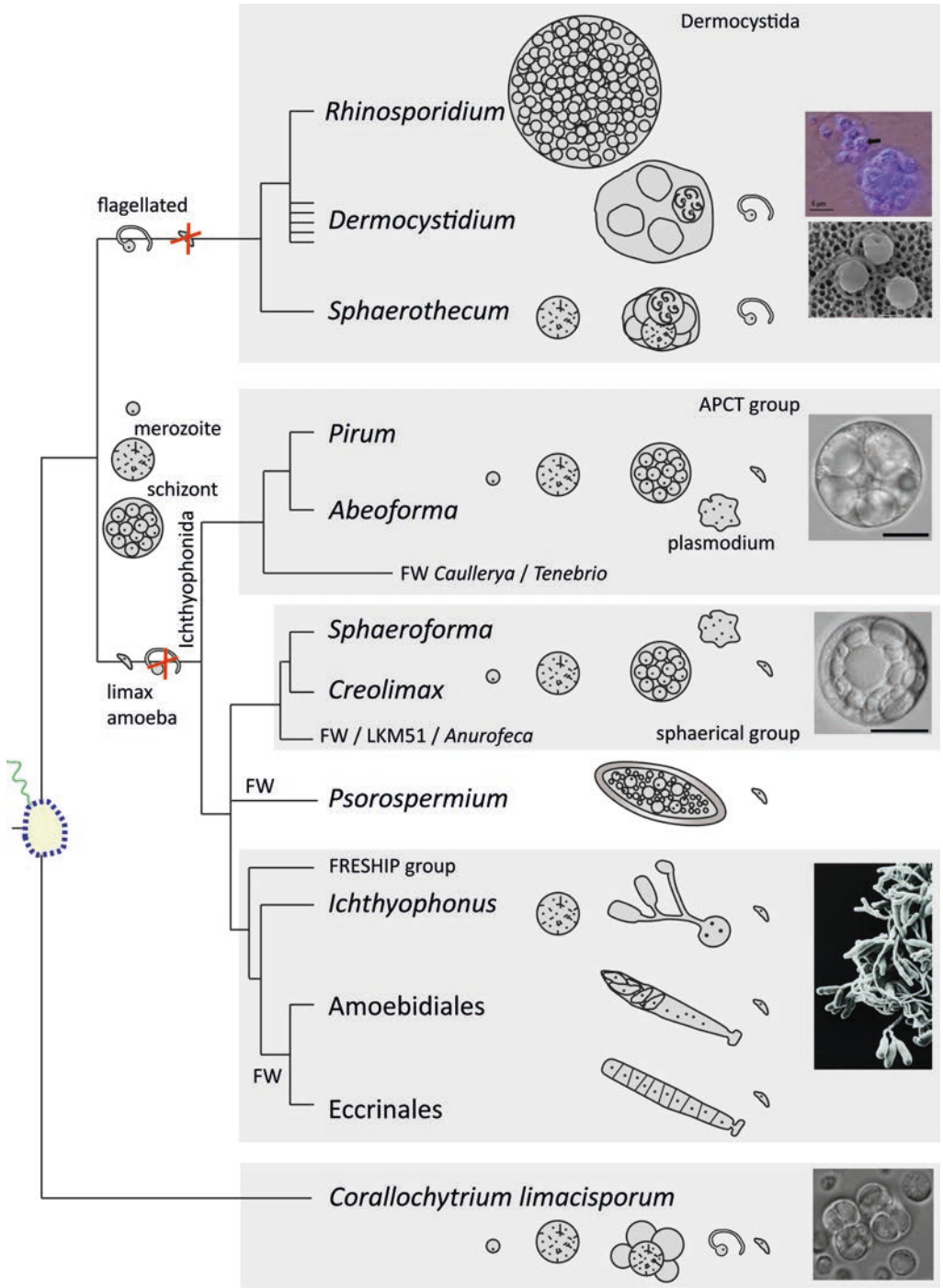


Figure 18. Osmotroph holozoans ancestral character states. Their Last Common Ancestor probably had a single flagellum but no filopodia. It was already osmotroph surrounded by a cell wall (blue) and had coenocytic growth.

Within the Holomycota, the data suggest the ancestral **nucleariid** (**section 1.2.5**) was a fresh-water non-flagellated, naked filose amoeba, probably of smaller size than current *Nuclearia* species, which are much bigger than *Fonticula alba* and other Opisthokonta protists. In fact, a recent finding of an uncharacterized fresh-water bacterivorous *Nuclearia sp*. (~5 *μ*m) that is positioned between *F. alba* lineage and canonical *Nuclearia* clade supports this hypothesis on nucleariid ancestral morphology (**Figure 19**). The aggregative multicellular fruiting body of *F. alba* probably arose independently to any other aggregative behaviour in Opisthokonta. Regarding the **Opisthosporidia** clade (Karpov, Mamkaeva, Aleoshin, et al. 2014) (**section 1.2.6**), Microsporidia are clearly highly derived parasites. It is the naked phagotrophic vegetative stage and life cycle of the parasitic aphelids and *Rozella* that could be considered the ancestral state to this lineage. Indeed, there are flagellates, amoeboflagellates and even podiated amoebae aphelids. This means that among the aphelids, most of the key ancestral characteristics of Opisthokonta are still present (Corsaro et al. 2014; Karpov, Mamkaeva, Aleoshin, et al. 2014).

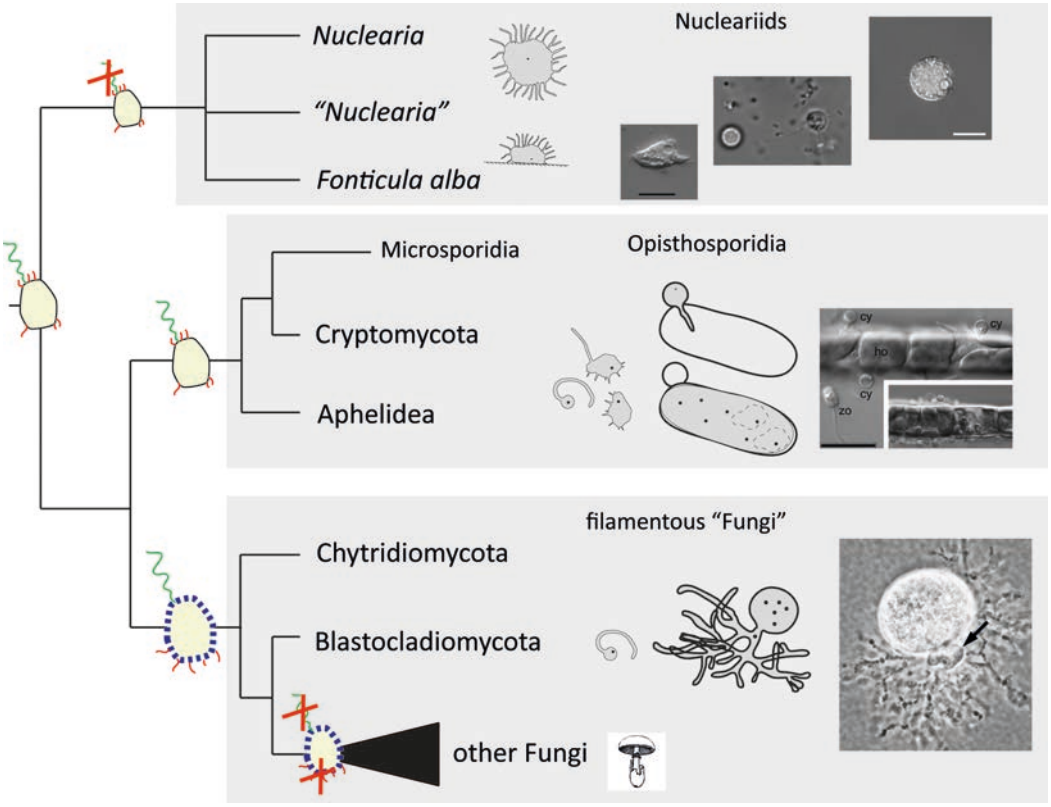


Figure 19. Holomycota ancestral character states. Their Last Common Ancestor probably was a phagotrophic amoeboflagellate. Flagellum (green), filopodia (red) and osmotrophy (dashed blue).

4.2.2 Reconstructing the Last Opisthokont Common Ancestor (LOCA)

If one looks to the reconstructed ancestors of Apusomonadida and Breviatea (**Figure 20**), and the reconstructed ancestor of each Opisthokonta lineage (**Figures 17**, **18** and **19**), no clear link between them points to how Opisthokonta originated and diversified. However, some facts are worth considering. LOCA was probably a marine benthonic phagotroph since the ancestor of both Holozoa and Holomycota was probably marine. Most likely it was an amoeboflagellate with already **one flagellum** as no Opisthokonta lineage presents more than one cilium. Whether the ancestral flagellar structure was similar to any of the extant lineages is unclear. Ultrastructural similarities between flagellated stages in chytrids and Dermocystida (i.e., rumpusome organization) (Pekkarinen et al. 2003; Letcher and Powell 2014) must be further explored. Also, **comparative studies of the microtubular cytoskeletons** in Opisthosporidia, *Ministeria vibrans* and Choanoflagellates may provide transitional forms or links from biflagellated ancestors. This is important, since Apusomonadida and Breviatea have more "rigid" or organized microtubular cytoskeletons (Yubuki and Leander 2013; Azimzadeh 2014), and it is proposed to be highly adaptive for specific lifestyles. Indeed, Cavalier-Smith sustains that the lost of ventral feeding mode, allowed the Opisthokonta to feed differently and therefore diversify morphologically (Cavalier-Smith 2013; Cavalier-Smith et al. 2014). Each Opisthokonta lineage is particularly different from one another, but most of them show rather **complex life cycles, with multiple cell types** (at least five in choanoflagellates (Dayel et al. 2011), three in *Capsaspora owczarzaki* (Sebé-Pedrós et al. 2013), five in Aphelida (Karpov, Mamkaeva, Benzerara, et al. 2014), etc.). Each of this cell types has specific functions already present in the LOCA; such feeding, dispersion, resistance or reproduction (already studied in the blastocladiomycetes *Allomyces macrogynus* (Pommerville 1982) and the choanoflagellate *Salpingoeca rosetta* (Levin and King 2013)). These cell types are usually connected in the cell cycle by cell division or directional transformation, in contrast to biflagellated ancestors, in which it seems that a single individual cell performs most functions modifying its behaviour. Such complex life cycles must be compared once the outgroup lineages are better understood. However, under our current understanding, it seems plausible that LOCA already had multiple cell types and had a more complex life cycle than its biflagellated ancestors. Another putative innovation is that all Opisthokonta lineages contain at least one species with pluricellularity or **colonialism** (but see (Walker et al. 2006)). *Fonticula alba* (Nucleariids) and *Capsaspora owczarzaki* (Filasterea) perform aggregative stages during their life cycle, and *Corallochytrium* plus Ichthyosporea, together with Opisthosporidia and chytrid fungi have coenocytic growth (multiple nuclei for the same cytoplasm). Finally, some choanoflagellates form clonal colonies by avoiding complete cell division (**section 1.2**).

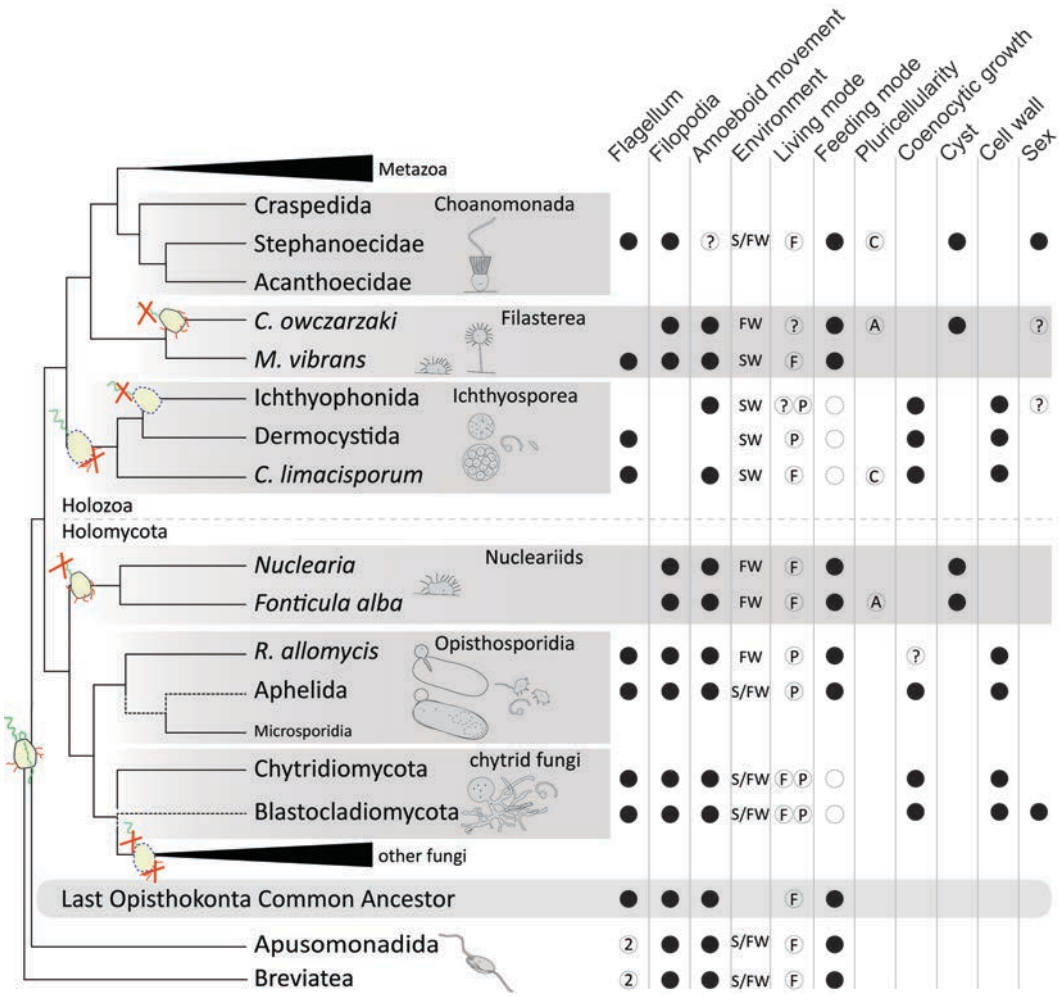


Figure 20. Summary of opisthokonta phylogeny and evolutionary important characters. SW and FW means sea and fresh water; F: free living; P: parasitic; C: clonal; A: aggregative. ?: there is data but not conclusive. All dots mean presence of the character, except in feeding mode: dashed, empty circles mean osmotrophy/saprotrophy; filled dots mean phagotrophy.

To fully reconstruct these evolutionary transitions it is also important to understand the environment or ecosystems where the distinct opisthokonts inhabit. Not only the current habitats are important, but any paleobiology data may be relevant (Knoll 2014). Unfortunately, paleogeological and paleoclimate data is limited and too generalist for the time range of Opisthokonta diversification. No fossils have been found for amorphean protists, besides a >700 million year old testate Amoebozoa (Porter et al. 2003). Regarding molecular clocks, a recent study (Eme et al. 2014) considered Opisthokonta the youngest eukaryotic supergroup (between 904 and 1579 Ma, in the late Mesoproterozoic or during Neoproterozoic eras). The lack of proper calibration points and precision for existing models lead the authors to advise that any data about this may be taken with caution. Indeed, as no hard body parts exist for Opisthokonta protists, researchers will have to look into molecular fossils or biomarkers (biochemical traces in fossilized soils), for example sterols which are found only in current sponges, but not in *M. brevicollis* (choanoflagellate) (Kodner et al. 2008). This means that Opisthokonta protist lipids or any other exclusive molecules must be analysed for each lineage to improve the

quality of molecular clock studies. Additionally, some authors propose that molecular phylogenies may provide hints to infer ancient macroevolutionary events (Pagel 1999; Fritz et al. 2013). Although sampling biases are assumed (e.g., no fossil record, undersampled extant diversity) the tree-shape (tree balance, diversification rates or branch-lengths) may indicate potential paleobiological events such as mass extinctions, fast diversifications or periods of stasis (Mooers and Heard 1997).

4.2.3 Convergent evolution within Opisthokonta

Finally, as mentioned at the beginning of this discussion, certain similarities observed between non-related holozoan and holomycotan lineages exist, and they are worth studying in the near future to fully understand early opisthokont evolution and diversification.

Osmotrophic opisthokonts

Both chytrids and *Corallochytrium* plus Ichthyosporea have a very specific feeding mode: they are saprotrophic/osmotrophic. Instead of ingesting through phagocytosis, they externally digest food – in the form of simple amino acids, fatty acids and disaccharides – to absorb it through specific transporters, similarly to other eukaryotes such as fungi or oomycetes do (Richards and Talbot 2013). Then, it is no surprising that *Corallochytrium* and many Ichthyosporea (e.g., Amoebidiales and Eccrinales) were originally classified as filamentous fungi (within thraustochytrids and trichomycetes, respectively) (Raghu-Kumar et al. 1987; Cafaro 2005) based solely on their morphology. Indeed, osmotrophic organisms seem to require specific bauplans (Olejarz and Nowak 2014). In our particular case, both chytrid fungi and osmotrophic holozoans have big coenocytic spherical protoplasms surrounded by chitinous cell walls and special stage for aquatic dispersion: small amoeboflagellated, flagellated or amoeboid (Hoffman et al. 2008). Whether the osmotrophic-specific features are homologous or not require thorough comparative studies, but most probably they evolved independently from phagotrophic ancestors, as such characters are adaptive for the same lifestyle. A spherical shape allows better absorption of digested nutrients. Also, cell walls are fundamental to get proper isolation or permeability from the environment (Pacheco-Arjona and Ramirez-Prado 2014). In **section 3.3 R3**, there is a putative example of parallel evolution for cell walls in fungi, *Corallochytrium* and Ichthyophonids. Chitin synthases, the key enzymes for chitin production in fungi, are present and diversified in both Holozoan lineages. Another important feature for any osmotroph is the transporter machinery and metabolism, which in the case of both groups probably evolved independently, maybe even through HGT. Thus, further comparative genomics analyses will be key to elucidate the origin of both lineages and the key aspects of the transition from phagotrophy to osmotrophy as well.

The naked filose amoebae

Morphological studies also artificially grouped some amoeboid opisthokonts. The filasterean *C. owczarzaki* (**section 1.2.2**) was originally classified as a *Nuclearia* due to morphological similarities (Stibbs and Owczarzak 1979), an even *Nuclearia* (**section 1.2.5**) were classified with other filose amoeboid eukaryotes (Patterson 1984). Some authors argued for distinct specific filopodia structures (e.g., branched or not) between *Nuclearia* and Filasterea (Shalchian-Tabrizi et al. 2008), but with our observations we conclude that both are branched and of undistinguishable size (see **Figures 5** and **9**). The main differences are that *Nuclearia* are dozens of microns big and feed from cyanobacteria, whereas Filasterea are about 3-5 *µ*m bacteriophagous. It is worth mentioning that in the lab, we recently managed to grow *C. owczarzaki* from *Klebsiella pneumoniae* which suggest that this amoeba, originally isolated from the haemolymph of a snail, might be free-living or have a free-living stage, in agreement with a study in which multiple protists were found within the same snail (Hertel et al. 2004). Interestingly, the recent finding of a smaller uncharacterized *Nuclearia sp*. (**Figure 19**) with no clear physical differences between Filasterea indicates a putative convergent evolution between these two lineages, or at least parallel evolution, since their last common ancestor may have already been a naked filose phagotrophic amoeba. The presence of a flagellum-like structure in *M. vibrans* supports this idea (**section 3.3 R3**), since LOCA was likely flagellated. For now, clear differences in transcription factor and tyrosine kinase content between Nucleariids and Filasterea (de Mendoza et al. 2013; Suga et al. 2014) clearly points to this convergent evolutionary landscape. Thus, again, we have very similar morphologies and even lifestyles that appear in two different parts of the opisthokont tree. The question here is whether they use the same molecular toolkits or gene regulation, meaning an ancestral trait, or simply the benthic lifestyle analogously shaped these amoebae.

5. CONCLUSIONS

The taxonomist nightmare is the evolutionist delight.

Unknown

The fragments of the natural method are to be diligently sought out. This is the first and last desideratum in botanical study. Nature does not make leaps. All plants show affinities on either side, like territories in a geographical map.

Linné, aphorism 77 of his Philosophia Botanica, 1751.

The main conclusions of the present work are the following:

1. Morphological data, molecular synapomorphies and rare genomic changes (e.g., gene fusions, miRNA, metabolic and biochemical pathways, etc.) are not proper phylogenetic markers to resolve deep phylogenetic questions, as they must ultimately be confirmed with phylogenomics.

2. Phylogenomic studies using concatenated data sets are not free of biases. Proper studies must involve parallel analyses of most complex evolutionary models to minimize systematic error (e.g., using proper outgroups or excluding data violating the models). In particular, Single Copy Protein Domain markers provide an independent source of phylogenetic data to classical data sets.

3. *de novo* RNAseq is a valid approach to increase taxonomic sampling to infer phylogenetic relationships, in this regard we also conclude that:

- Apusomonadida, and not Amoebozoa, is the sister group to the Opisthokonta. Also, Ancyromonadida is not closely related to Apusomonadida or Opisthokonta.

- Filasterea and Ichthyosporea are not monophyletic, and *Corallochytrium limacisporum* is more likely the sister group to the Ichthyosporea and together the earliest branching lineage among the Holozoa.

4. RNAseq data also provides valuable information to explore comparative genomics, such:

- the presence of cryptic flagellum-like structures in *Corallochytrium limacisporum* and *Ministeria vibrans*, putative examples of progressive lose of the cilium.

- the repertoire of Chytin Synthases in opisthokonts and the diversification in each osmotrophic lineage.

5. Strong similarities in morphology and lifestyle in independent opisthokont lineages: Filasterea and Nucleariids; or chytrid fungi and *Corallochytrium* plus Ichthyosporea are examples of parallel evolution or convergence.

6. REFERENCES

A

Adl SM, Simpson AGB, Farmer MA, et al. 2005. The new higher level classification of eukaryotes with emphasis on the taxonomy of protists. *J. Eukaryot. Microbiol*. 52:5 399–451.

Adl SM, Simpson AGB, Lane CE, et al. 2012. The revised classification of eukaryotes. *J. Eukaryot. Microbiol*. 59:5 429–514.

Alegado RA, Brown LW, Cao S, Dermenjian RK, Zuzow R, Fairclough SR, Clardy J, King N. 2012. A bacterial sulfonolipid triggers multicellular development in the closest living relatives of animals. *Elife* 1 e00013.

Andreou D, Arkush KD, Guégan JF, Gozlan RE. 2012. Introduced pathogens and native freshwater biodiversity: A case study of Sphaerothecum destruens. *PLoS One* 7:5 e36998.

Arkush KD, Mendoza L, Adkinson MA, Hedrick RP. 2003. Mesomycetozoean Fish Pathogen Formally Referred to as the Rosette Agent. *Nat. Hist*. 50:6 430–438.

Azimzadeh J. 2014. Exploring the evolutionary history of centrosomes. *Philos. Trans. R. Soc. Lond. B. Biol. Sci*. 369 1–12.

B

Baker GC, Beebee TC, Ragan MA. 1999. Prototheca richardsi, a pathogen of anuran larvae, is related to a clade of protistan parasites near the animal-fungal divergence. *Microbiology* 145 1777–1784.

Baldauf SL, Palmer JD. 1993. Animals and fungi are each other's closest relatives : Congruent evidence from multiple proteins. *Evolution* 90:24 11558–62.

Baldauf SL, Roger AJ, Wenk-Siefert I, Doolittle WF. 2000. A Kingdom-Level Phylogeny of Eukaryotes Based on Combined Protein Data. *Science* 290:5493 972–977.

Bangyeekhun E. 2001. Sequence analysis of the ribosomal internal transcribed spacer DNA of the crayfish parasite Psorospermium haeckeli. *Dis. Aquat. Organ*. 46 217–222.

Bapteste E, Brinkmann H, Lee JA, 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*. 99:3 1414–1419.

Bass D, Chao EE, Nikolaev S, Yabuki A, Ishida KI, Berney C, Pakzad U, Wylezich C, Cavalier-Smith T. 2009. Phylogeny of novel naked Filose and Reticulose Cercozoa: Granofilosea cl. n. and Proteomyxidea revised. *Protist* 160:1 75–109.

Bergsten J. 2005. Cladistics A review of long-branch attraction. *Cladistics* 21:2 163–193. Bininda-Emonds ORP. 2004. The evolution of supertrees. *Trends Ecol. Evol*. 19:6 315–322.

Blanc-Brude R, Skreb Y, Dragesco J. 1955. Sur la biologie de Nuclearia delicatula.

Cienkowski. *Bull. Micr. Appl*. 2 113–117.Boore JL. 2006. The use of genome-level characters for phylogenetic reconstruction.

Trends Ecol. Evol. 21:8 439–446.

Brown MW, Spiegel FW, Silberman JD. 2009. Phylogeny of the “forgotten” cellular slime mold, Fonticula alba, reveals a key evolutionary branch within Opisthokonta. *Mol. Biol. Evol*. 26:12 2699–2709.

Brown MW, Sharpe SC, Silberman JD, Heiss AA, Lang BF, Simpson AGB, Roger AJ. 2013. Phylogenomics demonstrates that breviate flagellates are related to opisthokonts and apusomonads. *Proc. R. Soc. B Biol. Sci*. 280:20131755 1–9.

Burki F, Shalchian-Tabrizi K, Minge M, Skjaeveland A, Nikolaev SI, Jakobsen KS, Pawlowski J. 2007. Phylogenomics reshuffles the eukaryotic supergroups. *PLoS One* 2:8 e790.

Burki F. 2014. The Eukaryotic Tree of Life from a Global Phylogenomic Perspective. *Cold Spring Harb. Perspect. Biol.* 6:5

Cafaro MJ. 2005. Eccrinales (Trichomycetes) are not fungi, but a clade of protists at the early divergence of animals and fungi. *Mol. Phylogenet. Evol*. 35:1 21–34.Cann JP. 1986. The feeding behavior and structure of Nuclearia delicula. *J. Protozool*. 33:3 392–396.

Capella-Gutiérrez S, Silla-Martínez JM, Gabaldón T. 2009. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* 25:15 1972–1973.

Capella-Gutiérrez S, Marcet-Houben M, Gabaldón T. 2012. Phylogenomics supports microsporidia as the earliest diverging clade of sequenced fungi. *BMC Biol*. 10:47 1-13.

Carr M, Leadbeater BSC, Hassan R, Nelson M, Baldauf SL. 2008. Molecular phylogeny of choanoflagellates, the sister group to Metazoa. *Proc. Natl. Acad. Sci. U. S. A*. 105:43 16641–16646.

Carvalho-Santos Z, Azimzadeh J, Pereira-Leal JB, Bettencourt-Dias M. 2011. Tracing the origins of centrioles, cilia, and flagella. *J. Cell Biol*. 194:2 165–175.

Cavalier-Smith T. 1993. Kingdom protozoa and its 18 phyla. *Microbial Rev*. 57:4 953-994.

Cavalier-Smith T, Paula Allsopp MTE. 1996. Corallochytrium, an enigmatic non-flagellate protozoan related to choanoflagellates. Eur. J. Protistol. 32.3 306–310.

Cavalier-Smith T. 1998. A revised six-kingdom system of life. *Biol. Rev. Camb. Philos. Soc*. 73:3 203–266.

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, Chao EE, Oates B. 2004. Molecular phylogeny of Amoebozoa and the evolutionary significance of the unikont Phalansterium. *Eur. J. Protistol*. 40:1 21–48.

Cavalier-Smith T, Chao EE, Stechmann A, Oates B, Nikolaev S. 2008. Planomonadida ord. nov. (Apusozoa): ultrastructural affinity with Micronuclearia podoventralis and deep divergences within Planomonas gen. nov. *Protist* 159:4 535–562.

Cavalier-Smith T, Chao EE. 2010. Phylogeny and Evolution of Apusomonadida (Protozoa: Apusozoa): New Genera and Species. *Protist* 161:4 549–576.

Cavalier-Smith T, Chao EE. 2012. Oxnerella micra sp. n. (Oxnerellidae fam. n.), a tiny naked centrohelid, and the diversity and evolution of heliozoa. *Protist* 163:4 574–601.

Cavalier-Smith T. 2013. Early evolution of eukaryote feeding modes, cell structural diversity, and classification of the protozoan phyla Loukozoa, Sulcozoa, and Choanozoa. *Eur. J. Protistol*. 49:2 115–178.

Cavalier-Smith T, Chao EE, Snell E A, Berney C, Fiore-Donno AM, Lewis R. 2014. Multigene eukaryote phylogeny reveals the likely protozoan ancestors of opisthokonts (animals, fungi, choanozoans) and Amoebozoa. *Mol. Phylogenet. Evol*. In Press.

Chan CX, Ragan MA. 2013. Next-generation phylogenomics. *Biol. Direct* 8:3 1-6.

Chen F, Mackey AJ, Stoeckert CJ, Roos DS. 2006. OrthoMCL-DB: querying a comprehensive multi-species collection of ortholog groups. *Nucleic Acids Res*. 34 D363–8.

Corradi N, Keeling PJ. 2009. Microsporidia: a journey through radical taxonomical revisions. *Fungal Biol. Rev*. 23:1-2 1–8.

Corsaro D, Walochnik J, Venditti D, Steinmann J, Müller K-D, Michel R. 2014. Microsporidia-like parasites of amoebae belong to the early fungal lineage Rozellomycota. *Parasitol. Res*. 113:5 1909-18.

Couradeau E, Benzerara K, Moreira D, Gérard E, Kaźmierczak J, Tavera R, López-García P. 2011. Prokaryotic and eukaryotic community structure in field and cultured microbialites from the alkaline Lake Alchichica (Mexico). *PLoS One* 6:12 e28767.

Criscuolo A, Gribaldo S. 2010. BMGE (Block Mapping and Gathering with Entropy): a new software for selection of phylogenetic informative regions from multiple sequence alignments. *BMC Evol. Biol*. 10:210.

D

Dayel MJ, Alegado R a, Fairclough SR, Levin TC, Nichols S a, McDonald K, King N. 2011. Cell differentiation and morphogenesis in the colony-forming choanoflagellate Salpingoeca rosetta. *Dev. Biol*. 357:1 73–82.

Dayel MJ, King N. 2014. Prey Capture and Phagocytosis in the Choanoflagellate Salpingoeca rosetta. *PLoS One* 9:5 e95577.

Dayhoff MO, Schwartz RM, Orcutt BC. 1978. A Model of Evolutionary Change in Proteins. In: Atlas of protein sequence and structure. National Biomedical Research Foundation, Washington, D.C. p. 345–352.

Deasey MC, Olive LS. 1981. Role of Golgi Apparatus in Sorogenesis by the Cellular Slime Mold Fonticula alba. *Science*. 213:4507 561–563.

Del Campo J, Massana R. 2011. Emerging diversity within chrysophytes, choanoflagellates and bicosoecids based on molecular surveys. *Protist* 162:3 435–448.

Del Campo J, Ruiz-Trillo I. 2013. Environmental Survey Meta-analysis Reveals Hidden Diversity among Unicellular Opisthokonts. *Mol. Biol. Evol*. 30:4 802–805.

De Mendoza A, Sebé-Pedrós A, Sestak MS, Matejcic M, Torruella G, Domazet-Lošo T, Ruiz-Trillo I. 2013. Transcription factor evolution in eukaryotes and the assembly of the regulatory toolkit in multicellular lineages. *Proc. Natl. Acad. Sci. U. S. A*. 110:50 E4858–66.

De la Torre-Bárcena JE, Kolokotronis SO, Lee EK, Stevenson DW, Brenner ED, Katari MS, Coruzzi GM, DeSalle R. 2009. The impact of outgroup choice and missing data on major seed plant phylogenetics using genome-wide EST data. *PLoS One* 4:6 e5764.

De Queiroz A, Gatesy J. 2007. The supermatrix approach to systematics. *Trends Ecol. Evol*. 22:1 34–41.

Delsuc F, Brinkmann H, Philippe H. 2005. Phylogenomics and the reconstruction of the tree of life. *Nat. Rev. Genet*. 6:5 361–375.

Derelle R, Lang BF. 2012. Rooting the Eukaryotic Tree with Mitochondrial and Bacterial Proteins. *Mol. Biol. Evol.* 29:4 1277–1289.

Dyková I, Veverková M, Fiala I, Macháčková B, Pecková H. 2003. Nuclearia pattersoni sp. n. (Filosea), a new species of amphizoic amoeba isolated from gills of roach (Rutilus rutilus), and its rickettsial endosymbiont. *Folia Parasitol*. 50:3 161–170.

E

Ebersberger I, de Matos Simoes R, Kupczok A, Gube M, Kothe E, Voigt K, von Haeseler A. 2012. A consistent phylogenetic backbone for the fungi. *Mol. Biol. Evol*. 29:5 1319–1334.

Ekelund F, Patterson DJ. 1997. Some Heterotrophic Flagellates from a Cultivated Garden Soil in Australia. *Arch. für Protistenkd*. 148:4 461–478.

Elias M. 2008. The guanine nucleotide exchange factors Sec2 and PRONE: candidate synapomorphies for the Opisthokonta and the Archaeplastida. *Mol. Biol. Evol*. 25:8 1526–1529.

Eme L, Sharpe SC, Brown MW, Roger AJ. 2014. On the age of eukaryotes: evaluating evidence from fossils and molecular clocks. *Cold Spring Harb Perspect Biol*. 6:8.

FFairclough SR, Dayel MJ, King N. 2010. Multicellular development in a choanoflagellate. *Curr. Biol*. 20:20 875–876.

Fairclough SR, Chen Z, Kramer E, et al. 2013. Premetazoan genome evolution and the regulation of cell differentiation in the choanoflagellate Salpingoeca rosetta. *Genome Biol*. 14:2 R15.

Feist SW, Longshaw M, Hurrell RH, Mander B. 2004. Observations of Dermocystidium sp. infections in bullheads, Cottus gobio L., from a river in southern England. *J. Fish Dis.* 27:4 225–231.

Felsenstein J. 1985. Confidence limits on phylogenies: An approach using the bootstrap. *Evolution*. 39:4 783–791.

Felsenstein J. 2004. Inferring phylogenies. Sinauer Associates, Inc.

Figueras A, Lorenzo G, Orda MC, Gouy M, Novoa B. 2000. Sequence of the Small Subunit Ribosomal RNA Gene of Perkinsus atlanticus-like Isolated from Carpet Shell Clam. *Mar. Biotechnol*. 2:5 419–428.

Foster PG. 2004. Modeling Compositional Heterogeneity. *Syst. Biol*. 53:3 485–495.Fritz SA, Schnitzler J, Eronen JT, Hof C, Böhning-Gaese K, Graham CH. 2013. Diversity in time and space: wanted dead and alive. *Trends Ecol. Evol.* 28:9 509–516.

Gabaldón T. 2008. Large-scale assignment of orthology: back to phylogenetics? *Genome Biol*. 9:10 235.

Glockling SL, Marshall WL, Gleason FH. 2013. Phylogenetic interpretations and ecological potentials of the Mesomycetozoea (Ichthyosporea). *Fungal Ecol.* 6:4 237–247.

Glöckner G, Noegel AA. 2012. Comparative genomics in the Amoebozoa clade. *Biol. Rev. Camb. Philos. Soc*. 88:1 215-25.

Glücksman E, Snell E a, Berney C, Chao EE, Bass D, Cavalier-Smith T. 2010. The Novel Marine Gliding Zooflagellate Genus Mantamonas (Mantamonadida ord. n.: Apusozoa). *Protist* 162:2 207–221.

Glücksman E, Snell EA, Cavalier-Smith T. 2013. Phylogeny and evolution of Planomonadida (Sulcozoa): Eight new species and new genera Fabomonas and Nutomonas. *Eur. J. Protistol*. 49:2 179-200.

Grant JR, Katz LA. 2014. Building a Phylogenomic Pipeline for the Eukaryotic Tree of Life – Addressing Deep Phylogenies with Genome-Scale Data. *PloS Curr.* Apr 2 1–19.

Guindon S, Gascuel O. 2003. A Simple, Fast, and Accurate Algorithm to Estimate Large Phylogenies by Maximum Likelihood. *Syst. Biol.* 52:5 696–704.

H

Hannen EJ Van, Mooij W, Agterveld MP Van, Gons HJ, Laanbroek HJ. 1999. Detritus-Dependent Development of the Microbial Community in an Experimental System: Qualitative Analysis by Denaturing Gradient Gel Electrophoresis. *Appl. Environ. Microbiol*. 65:6 2478–2484.

Hartmann S, Vision TJ. 2008. Using ESTs for phylogenomics: can one accurately infer a phylogenetic tree from a gappy alignment? *BMC Evol. Biol*. 8:95 1-13.

He D, Fiz-Palacios O, Fu CJ, Fehling J, Tsai CC, Baldauf SL. 2014. An alternative root for the eukaryote tree of life. *Curr. Biol*. 24:4 465–470.

Heiss AA, Walker G, Simpson AGB. 2010. Clarifying the taxonomic identity of a phylogenetically important group of eukaryotes: Planomonas is a junior synonym of Ancyromonas. *J. Eukaryot. Microbiol*. 57:3 285–293.

Heiss AA, Walker G, Simpson AGB. 2011. The ultrastructure of Ancyromonas, a eukaryote without supergroup affinities. *Protist* 162:3 373–393.

Heiss AA, Walker G, Simpson AGB. 2013. The Microtubular Cytoskeleton of the Apusomonad Thecamonas, a Sister Lineage to the Opisthokonts. *Protist* 164:5 598–621.

Henttonen P, Huner JV, Rata P, Lindqvist OV. 1997. A comparison of the known life forms of Psorospermium spp . in freshwater crayfish (Arthropoda, decapoda) with emphisis on Astacus astacus L . (Astacidae) and Procambarus clarkii (Girard) (Cambaridae). *Aquaculture* 149:11 15–30.

Herr RA, Ajello L, Taylor JW, Arseculeratne SN, Mendoza L. 1999. Phylogenetic analysis of Rhinosporidium seeberi's 18S small-subunit ribosomal DNA groups this pathogen among members of the protoctistan Mesomycetozoa clade. *J. Clin. Microbiol.* 37:9 2750–2754.

Herr RA, Ajello L, Taylor JW, Arseculeratne SN. 1999. Phylogenetic Analysis of Rhinosporidium seeberi's 18S Small- Subunit Ribosomal DNA Groups This Pathogen among Members of the Protoctistan Mesomycetozoa Clade. *J. Clin. Microbiol*. 37:9 2750–2754.

Hertel LA, Bayne CJ, Loker ES. 2002. The symbiont Capsaspora owczarzaki, nov. gen. nov. sp., isolated from three strains of the pulmonate snail Biomphalaria glabrata is related to members of the Mesomycetozoea. *Int. J. Parasitol*. 32:9 1183–1191.

Hertel LA, Barbosa CS, Santos RAAL, Loker ES. 2004. Molecular identification of symbionts from the pulmonate snail Biomphalaria glabrata in Brazil. *J. Parasitol*. 90:4 759–763.

Hess S, Sausen N, Melkonian M. 2012. Shedding light on vampires: the phylogeny of vampyrellid amoebae revisited. *PLoS One* 7:2 e31165.

Hibbett DS, Binder M, Bischoff JF, et al. 2007. A higher-level phylogenetic classification of the Fungi. *Mycol. Res*. 111:5 509–547.

Hinchliff CE, Roalson EH. 2013. Using supermatrices for phylogenetic inquiry: an example using the sedges. *Syst. Biol*. 62:2 205–219.

Hoffman Y, Aflalo C, Zarka A, Gutman J, James TY, Boussiba S. 2008. Isolation and characterization of a novel chytrid species (phylum Blastocladiomycota), parasitic on the green alga Haematococcus. *Mycol. Res*. 112 70–81.

Huang J, Xu Y, Gogarten JP. 2005. The presence of a haloarchaeal type tyrosyl-tRNA synthetase marks the opisthokonts as monophyletic. *Mol. Biol. Evol*. 22:11 2142–2146.

Huson DH, Rupp R, Scornavacca C. 2010. Phylogenetic Networks: Concepts, Algorithms and Applications. Cambridge University Press

James TY, Kauff F, Schoch CL, et al. 2006. Reconstructing the early evolution of Fungi using a six-gene phylogeny. *Nature* 443:7113 818–822.

James TY, Letcher PM, Longcore JE, Mozley-Standridge SE, Porter D, Powell MJ, Griffith GW, Vilgalys R. 2006. A molecular phylogeny of the flagellated fungi (Chytridiomycota) and description of a new phylum (Blastocladiomycota). *Mycologia* 98:6 860–871.

James TY, Berbee ML. 2012. No jacket required--new fungal lineage defies dress code: recently described zoosporic fungi lack a cell wall during trophic phase. *Bioessays* 34:2 94–102.

James TY, Pelin A, Bonen L, Ahrendt S, Sain D, Corradi N, Stajich JE. 2013. Shared Signatures of Parasitism and Phylogenomics Unite Cryptomycota and Microsporidia. *Curr. Biol*. 23:16 R679-R681.

Jeffroy O, Brinkmann H, Delsuc F, Philippe H. 2006. Phylogenomics: the beginning of incongruence? *Trends Genet*. 22:4 225–231.

Jones MDM, Forn I, Gadelha C, Egan MJ, Bass D, Massana R, Richards TA. 2011. Discovery of novel intermediate forms redefines the fungal tree of life. *Nature* 474:7350 200–203.

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

Jøstensen J, Sperstad S, Johansen S, Landfald B. 2002. Molecular-phylogenetic, structural and biochemical features of a cold-adapted, marine ichthyosporean near the animal-fungal divergence, described from in vitro cultures. *Europ J Protistol*. 38:2 93–104.

K

Kamikawa R, Brown MW, Nishimura Y, et al. 2013. Parallel re-modeling of EF-1α function: divergent EF-1α genes co-occur with EFL genes in diverse distantly related eukaryotes. *BMC Evol. Biol*. 13:131 1-12.

Karpov SA, Leadbeater BSC. 1997. Cell and nuclear division in a freshwater choanoflagellate, Monosiga ovata Kent. *Eur. J. Protistol*. 33:3 323–334.

Karpov SA, Leadbeater BSC. 1998. Cytoskeleton Structure and Composition in Choanoflagellates. *J. Eukaryot. Microbiol.* 45:3 361–367.

Karpov SA, Mikhailov K V, Mirzaeva GS, Mirabdullaev IM, Mamkaeva K A, Titova NN, Aleoshin V V. 2013. Obligately phagotrophic aphelids turned out to branch with the earliest-diverging fungi. *Protist* 164:2 195–205.

Karpov SA, Mamkaeva MA., Aleoshin VV., Nassonova E, Lilje O, Gleason FH. 2014. Morphology, phylogeny, and ecology of the aphelids (Aphelidea, Opisthokonta) and proposal for the new superphylum Opisthosporidia. *Front. Microbiol*. 5 1–11.

Karpov SA, Mamkaeva MA., Benzerara K, Moreira D, López-García P. 2014. Molecular Phylogeny and Ultrastructure of Aphelidium aff. melosirae (Aphelida, Opisthosporidia). *Protist* 165:4 512–526.

Katoh K, Misawa K, Kuma K, Miyata T. 2002. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res*. 30:14 3059–3066.

Katz LA, Grant J, Parfrey LW, Gant A, O'Kelly CJ, Anderson OR, Molestina RE, Nerad T. 2011. Subulatomonas tetraspora nov. gen. nov. sp. is a member of a previously unrecognized major clade of eukaryotes. *Protist* 162:5 762–773.

Keeling PJ, Inagaki Y. 2004. A class of eukaryotic GTPase with a punctate distribution suggesting multiple functional replacements of translation elongation factor 1α. *Proc. Natl. Acad. Sci. U. S. A.* 101:43 15380–15385.

Kim E, Simpson AGB, Graham LE. 2006. Evolutionary relationships of apusomonads inferred from taxon-rich analyses of 6 nuclear encoded genes. *Mol. Biol. Evol*. 23:12 2455–2466.

King N, Westbrook MJ, Young SL, et al. 2008. The genome of the choanoflagellate Monosiga brevicollis and the origin of metazoans. *Nature* 451:7180 783–788.

Knoll AH. 2014. Paleobiological perspectives on early eukaryotic evolution. *Cold Spring Harb. Perspect. Biol*. 6:1 1–14.

Kocan R, LaPatra S, Hershberger P. 2013. Evidence for an amoeba-like infectious stage of Ichthyophonus sp. and description of a circulating blood stage: a probable mechanism for dispersal within the fish host. *J. Parasitol*. 99:5 235–240.

Kocan RM. 2013. Proposed changes to the nomenclature of Ichthyophonus sp. life stages and structures. *J. Parasitol*. 99:2 906–909.

Kodner RB, Summons RE, Pearson A, King N, H. Knoll A. 2008. Sterols in a unicellular relative of the metazoans. *Proc. Natl. Acad. Sci. U. S. A*. 105:29 9897–9902.

Lang BF, O'Kelly C, Nerad T, Gray MW, Burger G. 2002. The closest unicellular relatives of animals. *Curr. Biol*. 12:20 1773–1778.

Lara E, Moreira D, López-García P. 2010. The environmental clade LKM11 and Rozella form the deepest branching clade of fungi. *Protist* 161:1 116–121.

Lartillot N, Philippe H. 2004. A Bayesian mixture model for across-site heterogeneities in the amino-acid replacement process. *Mol. Biol. Evol*. 21:6 1095–1109.Leadbeater BSC, Hassan R, Nelson M, Carr M, Baldauf SL. 2008. A new genus,

Helgoeca gen. nov., for a nudiform choanoflagellate. *Eur. J. Protistol*. 44:3 227–237. Leadbeater BSC. 2008. Choanoflagellate evolution: the morphological perspective. *Protistology* 5:4 256–267.

Leadbeater BSC, Yu Q, Kent J, Stekel DJ. 2009. Three-dimensional images of choanoflagellate loricae. *Proc. R. Soc. B Biol. Sci*. 276:1654 3–11.

Lemmon AR, Brown JM, Stanger-Hall K, Lemmon EM. 2009. The effect of ambiguous data on phylogenetic estimates obtained by maximum likelihood and Bayesian inference. *Syst. Biol*. 58:1 130–145.

Leonard G, Richards T A. 2012. Genome-scale comparative analysis of gene fusions, gene fissions, and the fungal tree of life. *Proc. Natl. Acad. Sci. U. S. A.* 109:52 21402–21407.

Letcher PM, Lopez S, Schmieder R, Lee P A, Behnke C, Powell MJ, McBride RC. 2013. Characterization of Amoeboaphelidium protococcarum, an Algal Parasite New to the Cryptomycota Isolated from an Outdoor Algal Pond Used for the Production of Biofuel. *PLoS One* 8:2 e56232.

Letcher P, Longcore J, Powell M. 2014. Dendrochytridium crassum gen. et sp. nov., a taxon in Chytridiales with unique zoospore ultrastructure. *Mycologia* 106:1 145–153.

Letcher PM, Powell MJ. 2014. Hypothesized evolutionary trends in zoospore ultrastructural characters in Chytridiales (Chytridiomycota). *Mycologia* 106:3 379–396.

Levin TC, King N. 2013. Evidence for sex and recombination in the choanoflagellate Salpingoeca rosetta. *Curr. Biol*. 23:21 2176–2180.

Lichtwardt RW, Cafaro MJ, White MM. 2001. The Trichomycetes: Fungal Associates of Arthropods. Trichomycetes fungal Assoc. arthropods, by Robert W. Lichtwardt 1986 Available from: http://www.nhm.ku.edu/~fungi/monograph/text/contents.htm

Liu L, Yu L, Kubatko L, Pearl DK, Edwards S V. 2009. Coalescent methods for estimating phylogenetic trees. Mol. Phylogenet. Evol. 53.1 320–328.

Liu YJ, Hodson MC, Hall BD. 2006. Loss of the flagellum happened only once in the fungal lineage: phylogenetic structure of kingdom Fungi inferred from RNA polymerase II subunit genes. *BMC Evol. Biol*. 6:74 1-13.

Liu Y, Steenkamp ET, Brinkmann H, Forget L, Philippe H, Lang BF. 2009. Phylogenomic analyses predict sistergroup relationship of nucleariids and fungi and paraphyly of zygomycetes with significant support. *BMC Evol. Biol*. 9:272 1-11.

Lohr JN, Laforsch C, Koerner H, Wolinska J. 2010. A Daphnia parasite (Caullerya mesnili) constitutes a new member of the Ichthyosporea, a group of protists near the animal-fungi divergence. *J. Eukaryot. Microbiol*. 57:4 328–336.

Longcore JE, Simmons DR. 2012. The Polychytriales ord. nov. contains chitinophilic members of the rhizophlyctoid alliance. *Mycologia* 104:1 276–294.

Lopez P, Casane D, Philippe H. 2002. Heterotachy, an important process of protein evolution. *Mol. Biol. Evol.* 19:1 1–7.

Lord JC, Hartzer KL, Kambhampati S. 2012. A Nuptially Transmitted Ichthyosporean Symbiont of Tenebrio molitor (Coleoptera: Tenebrionidae). *J. Eukaryot. Microbiol*. 59:3 1-5.

M

Mah JL, Leys SP. 2014. Choanoflagellate and choanocyte collar-flagellar systems and the assumption of homology. Evol. Dev. 16.1 25–37.

Manktelow M. 2010. History of Taxonomy.Manohar CS, Raghukumar C. 2013. Fungal diversity from various marine habitats deduced through culture-independent studies. *FEMS Microbiol. Lett*. 341:2 69–78.

Margulis L. 1990. Handbook of protoctista : the structure, cultivation, habitats, and life histories of the eukaryotic microorganisms and their descendants exclusive of animals, plants, and fungi : a guide to the algae, ciliates, foraminifera, sporozoa, water molds, slime. Boston: Jones and Bartlett Publishers.

Marks J. 2008. Great chain of being. In: Moore JH, editor. The Encyclopedia of Race and Racism. Vol. 2. Detroid, MI: Macmillan Reference USA. 68–73.

Marron AO, Alston MJ, Heavens D, Akam M, Caccamo M, Holland PWH, Walker G. 2013. A family of diatom-like silicon transporters in the siliceous loricate choanoflagellates. *Proc. R. Soc. B Biol. Sci*. 280:1756 20122543.

Marshall WL, Celio G, McLaughlin DJ, Berbee ML. 2008. Multiple isolations of a culturable, motile Ichthyosporean (Mesomycetozoa, Opisthokonta), Creolimax fragrantissima n. gen., n. sp., from marine invertebrate digestive tracts. *Protist* 159:3 415–433.

Marshall WL, Berbee ML. 2010. Population-level analyses indirectly reveal cryptic sex and life history traits of Pseudoperkinsus tapetis (Ichthyosporea, Opisthokonta): a unicellular relative of the animals. *Mol. Biol. Evol*. 27:9 2014–2026.

Marshall WL, Berbee ML. 2011. Facing Unknowns : Living Cultures (Pirum gemmata gen. nov., sp. nov., and Abeoforma whisleri, gen. nov., sp. nov.) from Invertebrate Digestive Tracts Represent an Undescribed Clade within the Unicellular Opisthokont Lineage Ichthyosporea (Mesomycetozoea). *Protist* 162:1 33-57

Marshall WL, Berbee ML. 2013. Comparative morphology and genealogical delimitation of cryptic species of sympatric isolates of sphaeroforma (ichthyosporea, opisthokonta). *Protist* 164:2 287–311.

Mayr E. 1982. The Growth of Biological Thought Diversity, Evolution, and Inheritance. eleventh 2. Massachusetts London, England: The Belknap Press of Harvard University Press Cambridge

McLaughlin DJ, Hibbett DS, Lutzoni F, Spatafora JW, Vilgalys R. 2009. The search for the fungal tree of life. *Trends Microbiol*. 17:11 488–497.

Medina M, Collins AG, Taylor JW, Valentine JW, Lipps JH, Amaral-Zettler L, Sogin ML. 2003. Phylogeny of Opisthokonta and the evolution of multicellularity and complexity in Fungi and Metazoa. *Int. J. Astrobiol.* 2.3 203–211.

Mendoza L, Taylor JW, Ajello L. 2002. The class mesomycetozoea: a heterogeneous group of microorganisms at the animal-fungal boundary. *Annu. Rev. Microbiol*. 56 315–344.

Mignot JP, Savoie A. 1979. Observations ultrastructurales sur Nuclearia simplex Cienkowski (Protozoa, Rhizopodea, Filosia). *Protistologica* 15:1 23–32.

Mikrjukov KA, Mylnikov AP. 2001. A study of the fine structure and the mitosis of a lamellicristate amoeba, Micronuclearia podoventralis gen. et sp. nov. (Nucleariidae, Rotosphaerida). *Eur. J. Protistol*. 37:1 15–24.

Minge MA, Silberman JD, Orr RJS, Cavalier-Smith T, Shalchian-Tabrizi K, Burki F, Skjaeveland A, Jakobsen KS. 2009. Evolutionary position of breviate amoebae and the primary eukaryote divergence. *Proc. R. Soc. B Biol. Sci.* 276:1657 597–604.

Mooers AO, Heard SB. 1997. Inferring evolutionary process from phylogenetic tree shape. Q. Rev. Biol. 72:1 31–54.

Moreira D, von der Heyden S, Bass D, López-García P, Chao E, 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.* 44:1 255–266.

N

Nesnidal MP, Helmkampf M, Bruchhaus I, Hausdorf B. 2010. Compositional heterogeneity and phylogenomic inference of metazoan relationships. *Mol. Biol. Evol*. 27:9 2095–2104.

Nichols SA, Roberts BW, Richter DJ, Fairclough SR, King N. 2012. Origin of metazoan cadherin diversity and the antiquity of the classical cadherin/β-catenin complex. *Proc. Natl. Acad. Sci. U. S. A*. 109:32 13046–13051.

Nielsen C. 2008. Six major steps in animal evolution: are we derived sponge larvae? *Evol. Dev*. 10:2 241–257.

Nitsche F, Weitere M, Scheckenbach F, Hausmann K, Wylezich C, Arndt H. 2007. Deep Sea Records of Choanoflagellates with a Description of Two New Species. *Acta Protozool*. 46 99–106.

Nitsche F, Carr M, Arndt H, Leadbeater BSC. 2011. Higher Level Taxonomy and Molecular Phylogenetics of the Choanoflagellatea. *J. Eukaryot. Microbiol*. 58:5 452-462.

O

Olejarz JW, Nowak M A. 2014. Evolution of staying together in the context of diffusible public goods. J. Theor. Biol. 360C 1–12.

Owczarzak A, Stibbs HH, Bayne CJ. 1980. The destruction of Schistosoma mansoni mother sporocysts in vitro by amoebae isolated from Biomphalaria glabrata: an ultrastructural study. J. Invertebr. Pathol. 35.1 26–33.

P

Pacheco-Arjona JR, Ramirez-Prado JH. 2014. Large-scale phylogenetic classification of fungal chitin synthases and identification of a putative cell-wall metabolism gene cluster in Aspergillus genomes. PLoS One 9.8 e104920.

Pagel M. 1999. Inferring the historical patterns of biological evolution. Nature 401.6756 877–884.

Paley RK, Andreou D, Bateman KS, Feist SW. 2012. Isolation and culture of Sphaerothecum destruens from Sunbleak (Leucaspius delineatus) in the UK and pathogenicity experiments in Atlantic salmon (Salmo salar). Parasitology 139.7 904–914.

Paps J, Ruiz-Trillo I. 2010. Animals and their unicellular ancestors. In: Encyclopedia of Life Sciences. 1–8.

Paps J, Medina-Chacón LA, Marshall W, Suga H, Ruiz-Trillo I. 2013. Molecular phylogeny of unikonts: new insights into the position of apusomonads and ancyromonads and the internal relationships of opisthokonts. *Protist* 164:1 2–12.

Parfrey LW, Barbero E, Lasser E, Dunthorn M, Bhattacharya D, Patterson DJ, Katz L a. 2006. Evaluating support for the current classification of eukaryotic diversity. *PLoS Genet*. 2:12 e220.

Pascolini R, Daszak P, Cunningham AA, Tei S, Vagnetti D, Bucci S, Fagotti A, Di Rosa I. 2003. Parasitism by Dermocystidium ranae in a population of Rana esculenta complex in Central Italy and description of Amphibiocystidium n. gen. *Dis. Aquat. Organ*. 56:1 65–74.

Patterson DJ. 1983. On the Organization of the Naked Filose Amoeba, Nuclearia moebiusi Frenzel, 1897 (Sarcodina, Filosea) and Its Implications1. *J. Protozool*. 30:2 301–307.

Patterson DJ. 1984. The genus Nuclearia (Sarcodina, Filosea): Species Composition and Characteristics of the Taxa. *Arch. Mikrobiol*. 127–139.

Patterson DJ. 1985. On the Organization and Affinities of the Amoeba, Pompholyxophrys punicea Archer, Based on Ultrastructural Examination of Individual Cells from Wild Material. *J. Eukaryot. Microbiol.* 32:2 241–246.

Patterson D, Nygaard K, Steinberg G, Turley C. 1993. Heterotrophic flagellates and other protists associated with oceanic detritus throughout the water column in the mid North Atlantic. J. Mar. Biol. Ass. U.K. 73 67–95.

Patterson DJ. 1999. The Diversity of Eukaryotes. *Am. Nat*. 154.S4 S96–S124.

Paul M. 2012. Acanthocorbis mongolica nov. spec.: description of the first freshwater loricate choanoflagellate (Acanthoecida) from a Mongolian lake. *Eur. J. Protistol*. 48:1 1–8.

Pekkarinen M, Lom J, Murphy C. 2003. Phylogenetic Position and Ultrastructure of Two Dermocystidium Species (Ichthyosporea) from the Common Perch (Perca fluviatilis). *Acta Protozool*. 42 287–307.

Pereira CN, Rosa I Di, Fagotti A, Simoncelli F, Pascolini R. 2005. The Pathogen of Frogs Amphibiocystidium ranae Is a Member of the Order Dermocystida in the Class Mesomycetozoea. *J clin microbiol* 43:1 192–198.

Pettitt ME, Orme BAA, Blake JR, Leadbeater BSC. 2002. The hydrodynamics of filter feeding in choanoflagellates. *Eur. J. Protistol*. 38 313–332.

Phillips MJ, Delsuc F, Penny D. 2004. Genome-scale phylogeny and the detection of systematic biases. *Mol. Biol. Evol.* 21:7 1455–1458.

Philippe H, Snell EA, Bapteste E, Lopez P, Holland PWH, Casane D. 2004. Phylogenomics of eukaryotes: impact of missing data on large alignments. *Mol. Biol. Evol*. 21:9 1740–1752.

Philippe H, Delsuc F, Brinkmann H, Lartillot N. 2005. Phylogenomics. *Annu. Rev. Ecol. Evol. Syst*. 36:1 541–562.

Philippe H, Brinkmann H, Copley RR, Moroz LL, Nakano H, Poustka AJ, Wallberg A, Peterson KJ, Telford MJ. 2011. Acoelomorph flatworms are deuterostomes related to Xenoturbella. *Nature* 470:7333 255–258.

Philippe H, Brinkmann H, Lavrov DV, Littlewood DTJ, Manuel M, Wörheide G, Baurain D. 2011. Resolving Difficult Phylogenetic Questions: Why More Sequences Are Not Enough. *PLoS Biol*. 9:3 e1000602.

Pinevich A, Gromov B, Mamkaeva K, Nasonova E. 1997. Study of Molecular Karyotypes in Amoeboaphelidium protococcarum,the Endotrophic Parasite of Chlorophycean alga Scenedesmus. *Curr. Microbiol*. 34:2 122–126.

Pommerville J. 1982. Morphology and physiology of gamete mating and gamete fusion in the fungus allomyces. *J. Cell Sci*. 53:1 193–209.

Porter S, Meisterfeld R, Knoll A. 2003. Vase-Shaped Microfossils From The Neoproterozoic Chuar Group, Grand Canyon: A Classification Guided By Modern Testate Amoebae. *J. Paleont*. 77:3 409–429.

Porter TM, Martin W, James TY, Longcore JE, Gleason FH, Adler PH, Letcher PM, Vilgalys R. 2011. Molecular phylogeny of the Blastocladiomycota (Fungi) based on nuclear ribosomal DNA. *Fungal Biol.* 115:4-5 381–392.

R

Raffel TR, Bommarito T, Barry DS, Witiak SM, Shackelton L A. 2008. Widespread infection of the Eastern red-spotted newt (Notophthalmus viridescens) by a new species of Amphibiocystidium, a genus of fungus-like mesomycetozoan parasites not previously reported in North America. *Parasitology* 135:2 203–215.

Ragan MA, Goggin CL, Cawthorn RJ, Cerenius L, Jamieson A V, Plourde SM, Rand TG, Söderhäll K, Gutell RR. 1996. A novel clade of protistan parasites near the animal-fungal divergence. *Proc. Natl. Acad. Sci. U. S. A*. 93:21 11907–11912.

Ragan MA. 2009. Trees and networks before and after Darwin. *Biol. Direct* 4:43 1-38

Raghu-Kumar S, Chandramohan D, Ramaiah N. 1987. Contribution of the Thraustochytrid Corallochytrium limacisporum Raghu-kumar to Microbial Biomass in Coral Reef Lagoons. Indian *J. Mar. Sci. New Delhi* 16 122–125.

Raghu-Kumar S. 1987. Occurrence of the thraustochytrid, Corallochytrium limacisporum gen. et sp. nov. in the coral reef lagoons of the Lakshadweep Islands in the Arabian *Sea. Bot. Mar*. 30:1 83–89.

Rannala B, Yang Z. 1996. Probability distribution of molecular evolutionary trees: a new method of phylogenetic inference. *J. Mol. Evol*. 43:3 304–311.

Richards TA, Cavalier-Smith T. 2005. Myosin domain evolution and the primary divergence of eukaryotes. *Nature* 436:7054 1113–1118.

Richards TA, Jones MDM, Leonard G, Bass D. 2012. Marine Fungi: Their Ecology and Molecular Diversity. *Ann. Rev. Mar. Sci.* 4:1 495–522.

Richards TA, Talbot NJ. 2013. Horizontal gene transfer in osmotrophs: playing with public goods. *Nat. Rev. Microbiol*. 11:10 720–727.

Rodríguez-Ezpeleta N, Brinkmann H, Burger G, Roger AJ, Gray MW, Philippe H, Lang BF. 2007. Toward resolving the eukaryotic tree: the phylogenetic positions of jakobids and cercozoans. *Curr. Biol*. 17:16 1420–1425.

Rodríguez-Ezpeleta N, Brinkmann H, Roure B, Lartillot N, Lang BF, Philippe H. 2007. Detecting and overcoming systematic errors in genome-scale phylogenies. *Syst. Biol*. 56:3 389–399.

Roger AJ, Simpson AGB. 2009. Evolution: revisiting the root of the eukaryote tree. *Curr. Biol*. 19:4 R165–7.

Roger AJ, Kolisko M, Simpson AGB. 2012. Phylogenomic Analysis. In: Evolution of Virulence in Eukaryotic Microbes, First Edition. Edited by L. David Sibley, Barbara J. Howlett, and Joseph Heitman. John Wiley & Sons, Inc. 44–69.

Rogozin IB, Basu MK, Csürös M, Koonin E V. 2009. Analysis of rare genomic changes does not support the unikont-bikont phylogeny and suggests cyanobacterial symbiosis as the point of primary radiation of eukaryotes. *Genome Biol. Evol*. 1 99–113.

Ronquist F, Huelsenbeck JP. 2003. MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* 19:12 1572–1574.

Roper M, Dayel MJ, Pepper RE, Koehl M a. R. 2013. Cooperatively Generated Stresslet Flows Supply Fresh Fluid to Multicellular Choanoflagellate Colonies. *Phys. Rev. Lett.* 110:22 228104.

Roure B, Rodríguez-Ezpeleta N, Philippe H. 2007. SCaFoS: a tool for selection, concatenation and fusion of sequences for phylogenomics. *BMC Evol. Biol*. 7 Suppl 1:S2.

Roure B, Baurain D, Philippe H. 2013. Impact of missing data on phylogenies inferred from empirical phylogenomic data sets. *Mol. Biol. Evol*. 30:1 197–214.

Ruiz-Trillo I. 1999. Acoel Flatworms: Earliest Extant Bilaterian Metazoans, Not Members of Platyhelminthes. *Science*. 283:5409 1919–1923.

Ruiz-Trillo I, Inagaki Y, Davis LA, Landfald B, Roger AJ. 2004. Capsaspora owczarzaki is an independent opisthokont lineage. *Curr. Biol*. 14:22 946–947.

Ruiz-Trillo I, Lane CE, Archibald JM, Roger AJ. 2006. Insights into the evolutionary origin and genome architecture of the unicellular opisthokonts Capsaspora owczarzaki and Sphaeroforma arctica. *J. Eukaryot. Microbiol*. 53:5 379–384.

Ruiz-Trillo I, Burger G, Holland PWH, King N, Lang BF, Roger AJ, Gray MW. 2007. The origins of multicellularity: a multi-taxon genome initiative. *Trends Genet*. 23:3 113–118.

Ruiz-Trillo I, Roger AJ, Burger G, Gray MW, Lang BF. 2008. A phylogenomic investigation into the origin of metazoa. *Mol. Biol. Evol*. 25:4 664–672.

Ryberg M, Matheny PB. 2011. Dealing With Incomplete Taxon Sampling and Diversification of a Large Clade of Mushroom-Forming Fungi. *Evolution* 65:7 1862–1878.

S

Salichos L, Rokas A. 2013. Inferring ancient divergences requires genes with strong phylogenetic signals. *Nature* 497:7449 327–331.

Sanderson MJ, McMahon MM, Steel M. 2010. Phylogenomics with incomplete taxon coverage: the limits to inference. *BMC Evol. Biol*. 10:155 1-13.

Sebé-Pedrós A, Roger A, Lang FB, King N, Ruiz-Trillo I. 2010. Ancient origin of the integrin-mediated adhesion and signaling machinery. *Proc. Natl. Acad. Sci. U. S. A.* 107:22 10142–10147.

Sebé-Pedrós A, Irimia M, del Campo J, Parra-Acero H, Russ C, Nusbaum C, Blencowe BJ, Ruiz-Trillo I. 2013. Regulated aggregative multicellularity in a close unicellular relative of metazoa. *Elife* 2 e01287–e01287.

Sekimoto S, Rochon D, Long JE, Dee JM, Berbee ML. 2011. A multigene phylogeny of Olpidium and its implications for early fungal evolution. *BMC Evol. Biol*. 11:331 1–10.

Shadwick JDL, Ruiz-Trillo I. 2012. A genomic survey shows that the haloarchaeal type tyrosyl tRNA synthetase is not a synapomorphy of opisthokonts. *Eur. J. Protistol*. 48:1 89–93.

Shalchian-Tabrizi K, Minge MA, Espelund M, Orr R, Ruden T, Jakobsen KS, Cavalier-Smith T. 2008. Multigene phylogeny of choanozoa and the origin of animals. *PLoS One* 3:5 e2098.

Simpson AGB, Roger AJ. 2004. The real “kingdoms” of eukaryotes. *Curr. Biol*. 14:17 R693–6.

Simpson AGB, Inagaki Y, Roger AJ. 2006. Comprehensive multigene phylogenies of excavate protists reveal the evolutionary positions of “primitive” eukaryotes. *Mol. Biol. Evol*. 23:3 615–625.

Sjölander K, Datta RS, Shen Y, Shoffner GM, Sjolander K. 2011. Ortholog identification in the presence of domain architecture rearrangement. *Brief. Bioinform*. 12:5 413–422.

Smirnov AV, Chao E, Nassonova ES, Cavalier-Smith T. 2011. A revised classification of naked lobose amoebae (Amoebozoa: lobosa). *Protist* 162:4 545–570.

Smith JE. 2009. The ecology and evolution of microsporidian parasites. *Parasitology* 136:14 1901–1914.

Spanggaard B. 1995. Morphology of Ichthyophonus hoferi assessed by light and scanning electron microscopy. *J. Fish Dis*. 18:6 567–577.

Stairs CW, Eme L, Brown MW, Mutsaers C, Susko E, Dellaire G, Soanes DM, van der Giezen M, Roger AJ. 2014. A SUF Fe-S Cluster Biogenesis System in the Mitochondrion-Related Organelles of the Anaerobic Protist Pygsuia. *Curr. Biol*. 24:11 1176–86.

Stamatakis A. 2006. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. Bioinformatics 22.21 2688–2690.

Stechmann A, Cavalier-Smith T. 2002. Rooting the eukaryote tree by using a derived gene fusion. Science. 297:5578 89–91.

Stechmann A, Cavalier-Smith T. 2003. The root of the eukaryote tree pinpointed. *Curr. Biol*. 13:17 R665–6.

Steenkamp ET, Wright J, Baldauf SL. 2006. The protistan origins of animals and fungi. *Mol. Biol. Evol*. 23:1 93–106.

Stibbs H, Owczarzak A. 1979. Schistosome Sporocyst-Killing Amoebae Isolated from Biomphalaria glabrata. *J. Invertebr. Pathol*. 33:2 159–170.

Stoupin D, Kiss AK, Arndt H, Shatilovich A V, Gilichinsky D A, Nitsche F. 2012. Cryptic diversity within the choanoflagellate morphospecies complex Codosiga botrytis - phylogeny and morphology of ancient and modern isolates. *Eur. J. Protistol*. 48:4 263–273.

Struck TH. 2014. TreSpEx—Detection of Misleading Signal in Phylogenetic Reconstructions Based on Tree Information. *Evol. Bioinforma*. 10 51–67.

Suga H, Ruiz-Trillo I. 2013. Development of ichthyosporeans sheds light on the origin of metazoan multicellularity. *Dev. Biol*. 377:1 284–292.

Suga H, Torruella G, Burger G, Brown MW, Ruiz-Trillo I. 2014. Earliest holozoan expansion of phosphotyrosine signaling. *Mol. Biol. Evol*. 31_3 517–528.

Sumathi JC, Raghukumar S, Kasbekar DP, Raghukumar C. 2006. Molecular evidence of fungal signatures in the marine protist Corallochytrium limacisporum and its implications in the evolution of animals and fungi. *Protist* 157:4 363–376.

Susko E, Field C, Blouin C, Roger AJ. 2003. Estimation of Rates-Across-Sites Distributions in Phylogenetic Substitution Models. *Syst. Biol*. 52:5 594–603.

Susko E, Roger AJ. 2007. On reduced amino acid alphabets for phylogenetic inference. *Mol. Biol. Evol*. 24:9 2139–2150.

T

Talavera G, Castresana J. 2007. Improvement of phylogenies after removing divergent and ambiguously aligned blocks from protein sequence alignments. *Syst. Biol*. 56:4 564–577.

Tanabe Y, Watanabe MM, Sugiyama J. 2005. Evolutionary relationships among basal fungi (Chytridiomycota and Zygomycota): Insights from molecular phylogenetics. *J. Gen. Appl. Microbiol*. 51:5 267–276.

Thomson RC, Plachetzki DC, Mahler DL, Moore BR. 2014. A critical appraisal of the use of microRNA data in phylogenetics. *Proc. Natl. Acad. Sci. U. S. A*. 111:35 E3659–E3668

Tong S, Vørs N, Patterson D. 1997. Heterotrophic flagellates , centrohelid heliozoa and filose amoebae from marine and freshwater sites in the Antarctic. *Polar Biol*. 18:2 91–106.

U

Umen J, Heitman J. 2013. Evolution of sex: mating rituals of a pre-metazoan. *Curr. Biol*. 23:22 R1006–8.

V

Vávra J, Lukeš J. 2013. Chapter Four - Microsporidia and 'The Art of Living Together', In: D. Rollinson, Editor(s), Advances in Parasitology, Academic Press, Volume 82, 253-319

Vilela R, Mendoza L. 2012. The taxonomy and phylogenetics of the human and animal pathogen Rhinosporidium seeberi: a critical review. *Rev. Iberoam. Micol*. 29:4 185–199.

Vogt G, Rug M. 1995. Microscopic anatomy and histochemistry of the crayfish parasite Psorospermium haeckeli. *Dis. Aquat. Organ*. 21 79–90.

W

Walker G, Dacks JB, Martin Embley T. 2006. Ultrastructural description of Breviata anathema, n. gen., n. sp., the organism previously studied as "Mastigamoeba invertens". *J. Eukaryot. Microbiol*. 53:2 65–78.

Wang H-C, Susko E, Roger AJ. 2009. PROCOV: maximum likelihood estimation of protein phylogeny under covarion models and site-specific covarion pattern analysis. *BMC Evol. Biol*. 9:225.

Wiens JJ. 2006. Missing data and the design of phylogenetic analyses. *J. Biomed. Inform*. 39:1 34–42.

Woese CR, Fox GE. 1977. Phylogenetic structure of the prokaryotic domain: the primary kingdoms. *Proc. Natl. Acad. Sci. U. S. A*. 74:11 5088–5090.

Worley AC, Raper KB, Hohl M. 1979. Fonticula alba: a new cellular slime mold (acrasiomycetes). *Mycologia* 71:4 746–760.

Wylezich C, Karpov SA, Mylnikov AP, Anderson R, Jürgens K. 2012. Ecologically relevant choanoflagellates collected from hypoxic water masses of the Baltic Sea have untypical mitochondrial cristae. *BMC Microbiol*. 12:271.

Yang Z. 1993. Maximum-likelihood estimation of phylogeny from DNA sequences when substitution rates differ over sites. *Mol. Biol. Evol.* 10:6 1396–1401.

Yoshida M, Nakayama T, Inouye I. 2009. Nuclearia thermophila sp. nov. (Nucleariidae), a new nucleariid species isolated from Yunoko Lake in Nikko (Japan). *Eur. J. Protistol*. 45:2 147–155.

Yubuki N, Leander BS. 2013. Evolution of microtubule organizing centers across the tree of eukaryotes. *Plant J*. 75:2 230–244.

Zettler L, Nerad TA, O'Kelly CJ, Sogin ML. 2001. The nucleariid amoebae: more protists at the animal-fungal boundary. *J. Eukaryot. Microbiol*. 48:3 293–297.Zettler L, Gómez F, Zettler E. 2002. Eukaryotic diversity in Spain's River of Fire. *Nature* 417: 6885 137.

Zhao S, Burki F, Bråte J, Keeling PJ, Klaveness D, Shalchian-Tabrizi K. 2012. Collodictyon--An Ancient Lineage in the Tree of Eukaryotes. *Mol. Biol. Evol*. 29:6 1–12.

Zhao S, Shalchian-Tabrizi K, Klaveness D. 2013. Sulcozoa revealed as a paraphyletic group in mitochondrial phylogenomics. *Mol. Phylogenet. Evol.* 69:3 462–468.

Zhou Y, Brinkmann H, Rodrigue N, Lartillot N, Philippe H. 2010. A dirichlet process covarion mixture model and its assessments using posterior predictive discrepancy tests. *Mol. Biol. Evol*. 27:2 371–384.

7. Resum en català

INTRODUCCIÓ

1.1 Perspectiva històrica de la taxonomia dels protistes

Des de temps antics els humans hem necessitat entendre i estudiar els organismes vius per sobreviure. Intrínsecament necessitem classificar el coneixement per estructurar-lo i comunicar-nos. En el cas dels éssers vius les classificacions han depengut al llarg de la història de la cosmogonia del moment. Civilitzacions antigues com la xinesa o l'egípcia classificaven les plantes medicinals, però grecs i romans va establir classificacions basades en creences religioses com l'escala de la natura aristotèlica. No va ser fins al segle XVIII que Linné va establir una classificació biològica de les espècies, i fins la teoria de l'evolució al XIX no es va començar a classificar amb l'única finalitat de reconstruir l'història evolutiva de les espècies, representades en diagrames d'arbre o xarxes.

Gràcies a la invenció de la microscopia i el descobriment de la cèl·lula dos segles abans, les classificacions evolutives, com la de Haeckel, contemplaven espècies unicel·lulars. De fet al segle XIX és quan es van començar a descriure força grups taxonòmics involucrats en aquest treball. Aleshores les classificacions estaven limitades a comparar pocs trets morfològics. Fins els avenços en microscopia electrònica al segle XX no es va poder descriure les espècies de protistes amb detall, a nivell subcel·lular. A més, l'estudi de l'evolució i la seqüenciació molecular van permetre la comparació de milers de posicions nucleotídiques per agrupar objectivament les espècies per parentiu evolutiu. És l'exemple de la subunitat petita de l'ARN ribosomal (18S), el marcador filogenètic més emprat en la història de la filogènia i que va modernitzar la classificació dels éssers vius unicel·lulars en sis supergrups: Amoebozoa, Excavata, Archaeplastida, Rhizaria, Chromalveolata i Opisthokonta.

1.2 Qui són els opistoconts?

Els opistoconts són el llinatge evolutiu que conté fongs i diversos grups de protistes (branca dels holomicots) i per altra banda els animals i els seus parents unicel·lulars (branca dels holozous). La filogènia com a tal es va començar a resoldre gràcies als arbre de 18S, que agrupaven aquests llinatges que en comú només tenen un flagell posterior. A continuació hi ha una breu descripció dels diferents grups de protistes.

1.2.1 Choanomonada

També coneguts com coanoflagel·lats, són nanoflagel·lats bacterívors de vida lliure que es troben a tot tipus d'ambients aquàtics. Es coneixen unes 250 espècies tot i que s'han detectat llinatges desconeguts a través de tècniques ambientals. La morfologia del grup és força conservada: una cèl·lula oval amb un flagell a la base del qual hi ha un collar de microvilli. Són filtradors de matèria en suspensió atreta per corrents provocats amb el flagell, quan la partícula contacta, aquesta és fagocitada (similar al tipus cel·lular

anomenat coanòcit de les esponges). Per aconseguir aquests corrents, han de compensar la força motriu del flagell i per això es fixen al substrat. Excepte els Stephanoecida, que juntament amb els Achantoecida recobreixen la cèl·lula amb una cota de malla, però a diferència d'aquests l'estructura és prou gran com per viure en suspensió a la columna d'aigua. Amb tot, els Craspedids són el grup més estudiat i del que es coneix el cicle vital que conté diferents estadis tals com nedador, filtrador, colonial o cístic.

1.2.2 Filasterea

És un grup format per només dos gèneres i dues espècies ben descrites: *Ministeria vibrans* i *Capsaspora owczarzaki*. Són amebes nues amb filopodis que repten pel substrat i fagociten bacteris. La segona és la més ben estudiada, tot i que fins que no es van emprar filogènies moleculars, la classificació era errònia amb altres amebes. C. owczarzaki en cultiu també presenta un estadi colonial, agregatiu en aquest cas, i un estadi cístic.

1.2.3 Ichthyosporea

Aquest grup d'unes 40 espècies d'osmòtrofs (no fagociten sinó que capten els nutrients del medi) s'havien classificat amb els fongs per similitud morfològica i de forma de viure fins que no es van emprar tècniques moleculars. Creixen com esquizonts (creixement d'un únic cos cel·lular amb múltiples nuclis) envoltats de paret cel·lular, que un cop madurs produeixen la divisió cel·lular i una nova generació que té dues formes diferents segons el grup taxonòmic intern. Els Dermocystida són paràsits estrictes d'animals (des de peixos fins a humans) que es dispersen en forma de flagel·lats. Aquest grup conté espècies com *Sphaerothecum destruens* o *Rhinosporidium seeberi*. L'altre grup s'anomena Ichthyophonida i es dispersa en forma d'amebes. És el grup més abundant i del que es coneix més la biologia, tot i que tots ells s'han trobat dins d'animals, la seva funció ecològica no és clara, com per exemple *Psorospermium haeckeli*, amb una morfologia força diferent a la resta de grups. *Ichthyophonus hoferi* sí que és un llinatge paràsit de peixos, que juntament amb Amoebidiales i Eccrinales (ectobionts d'artròpodes) formen estructures polaritzades. La resta de grups són simplement esfèrics (com els Dermocystida) i són els més estudiats. Espècies com Abeoforma whisleri o Sphaeroforma tapetis poden adquirir, a més, formes plasmodials.

1.2.4 Corallochytrium

Corallochytrium limacisporum també és un osmòtrof com el grup anterior, però en aquest cas de vida lliure. El cicle vital també és molt semblant: l'esquizont comença a créixer fins que es cel·lularitza, tot i que enlloc d'alliberar la propera generació en forma d'amebes, es manté en un estadi colonial de fins a 32 cèl·lules durant un temps. El problema és que ni el 18S ha posicionat clarament aquesta espècies dins la branca dels holozous.

1.2.5 Nucleariids

Aquest grup representa el llinatge que primer es va bifurcar dels holomicots. Són amebes nues amb filopodis d'aigua dolça i de mida considerable, ja que s'alimenten de cianobacteris. Gràcies al 18S, una altra ameba (Fonticula alba) es va posicionar com a grup germà d'aquestes amebes, i en conjunt separant el gènere Nucleariа d'altres amebes. A diferència de Nucleariа, F. alba és més petita i fagocita bacteris, però a més les amebes es poden agregar per formar un cos fructífer i al cap de munt generar la propera generació i dispersar-la.

1.2.6 Opisthosporidia

Aquest és el grup d'opistoconts més recentment descrit i es composa dels Cryptomycota, els Aphelida i els Microsporidia; aquests últims són paràsits intracel·lulars molt derivats. Els altres dos també són paràsits, però conserven formes més ancestrals. Tot i que la filogènia interna no és clara, tenen un cicle vital molt similar. Comença amb una cèl·lula adherint-se a la paret de l'hoste. L'organisme hi penetra gràcies a la producció de paret cel·lular pròpia (probablement de quitina com els fongs) i un tub germinal per on accedeix al protoplasma de l'hoste que és fagocitat fins que no queda res. Aleshores el paràsit és multinucleat i quan es divideix en diferents cèl·lules, la progènia pot ser en forma d'ameba nua, flagel·lat o ameboflagel·lat, començant així un nou cicle.

1.2.7 Els fongs "basals"

Per entendre l'evolució dels opistoconts és necessari parlar del primer llinatge dels fongs. Anomenats vulgarment quítrids. Són esquizonts osmòtrofs amb paret de quitina que projecten pseudohifes (anomenats rizòids) ja que no són septades; tot és una única cèl·lula i per tant es poden considerar protistes. Al ser aquàtics, la progènie s'esdevé en forma de flagel·lats o ameboflagel·lats fins i tot amb filopodis.

1.3 El grup germà dels opistoconts: Amoebozoa, Apusozoa and Breviatea

Abans de començar aquesta tesi els amebozous eren considerats el grup germà dels opistoconts per filogènia molecular. Majoritàriament uniflagel·lats, ambdós grups eren anomenats uniconts, separats de la resta d'eucariotes considerats biconts. Tot i així els apusozous (Apusomonadida i Ancyromonadida) són biflagel·lats que s'han proposat propers a opistoconts per estudis amb 18S. Aquests dos biflagel·lats són bentònics i bacterívors, per tant tenen una morfologia específica que s'ajusta a la funció. A més, el grups dels Breviatea, són biflagel·lats bentònics d'ambients hipòxics, que s'han mogut entre grup germà d'amebozous o recentment amb els Apusomonadida. En qualsevol cas, sembla que els Opisthokonts provenen d'ancestres biflagel·lats que ja presentaven la majoria de caràcters funcionals com filopodis, cists, flagell, etc.

1.4 Reconstruint filogènies profundes

L'origen dels eucariotes, les relacions entre supergroups d'eucariotes o l'origen i diversificació d'aquests, són exemples de preguntes filogenètiques profundes, com la dels opistoconts, l'objectiu principal d'aquest treball. Els caràcters morfològics (com la forma

de les crestes mitocondrials) s'han demostrat molt plàstics entre grups d'eucariotes i per tant no són marcadors evolutius vàlids. El 18S, així com arbres basats en un o pocs gens no acostumen a tenir prou senyal filogenètic com per reconstruir esdeveniments d'especiació tan antics, sobretot en relació a les posicions nucleotídiques o amino acídiques saturades (on han hagut més canvis dels que es poden detectar). A més, sovint poden tenir històries evolutives diferents a la de les espècies (esdeveniments de duplicació i pèrdua, transferència horitzontal, etc.).

1.4.1 Advertències en inferència d'especiacions antigues

Els anomenats canvis genòmics poc freqüents (sintènia, indels, fusions gèniques, microRNAs o presència/absència de gens) tampoc són marcadors vàlids, i és un exemple la divisió dels eucariotes entre uniconts i biconts pel sol fet de tenir uns gens fusionats i uns no, ja que a l'augmentar el mostreig taxonòmic, aquests marcadors es va diluir i per tant són menys estables del pensat fa uns anys quan es comparaven poques espècies.

1.4.2 Un protocol per filogenòmica

Dels diferent mètodes de filogenòmica, l'emprat en aquest treball es basa en la concatenació de múltiples seqüències. Aquest mètode es basa en augmentar la relació entre senyal filogenètic i soroll per recuperar l'arbre d'espècies correcte. El passos principals són semblants a l'ús de gens individuals però en aquest cas és necessari emprar processos automàtics per tractar amb volums de dades importants i per això cal tenir en consideració factors importants com els següents. 1) Tenir un mostreig taxonòmic representatiu del grup d'interès i de les espècies externes més properes. 2) Assegurar-se d'usar seqüències ortòlogues, que provinguin d'esdeveniments d'especiació. Per això s'han desenvolupat diferents mètodes que agrupen seqüències per similitud (blast), entre d'altres. 3) Amb tot, és important que la matriu per analitzar contingui el major número de posicions per espècie, el que s'anomena simetria de la matriu. Tot i que alguns autors han proposat que poden provocar biaixos, altres diuen que matrius asimètriques, si impliquen especies necessàries per trencar branques llargues o incrementar clades poc poblats, són útils. Finalment, és important considerar els errors sistemàtics ja que els mètodes de suport dels arbres poden ser alts per topologies errònies. Es deu a que els models evolutius no s'adapten a la realitat de l'evolució molecular, bàsicament no modelitzen canvis en la taxa de canvi de les posicions entre espècies, entre regions de les seqüències i al llarg del temps. De tal manera que hi ha molt esforç a desenvolupar models més complexos però també en detectar i excloure posicions de les seqüències que puguin violar models senzills. Entre d'altres, excloure espècies i posicions amb alta taxa evolutiva o recodificar aminoàcids en categories funcionals.

OBJECTIUS

1. Emprar les dades genòmiques obtingudes al projecte Unicorn per resoldre la filogènia dels opistoconts.
2. Incrementar el mostreig taxonòmic per transcriptòmica d'espècies sense afinitat clara.
3. Re-avaluar les relacions filogenètiques dels opistoconts emprant tots els llinatges coneguts.
4. Inferir i reconstruir les transicions evolutives entre llinatges a través de caràcters funcionalment importants.

RESULTATS I DISCUSSIÓ

4.1 Avaluació de la filogenòmica

La classificació dins els fongs de Corallochytrium limacisporum pel sol fet de compartir un enzim del metabolisme de la lisina no és vàlid com s'ha exposat a la introducció. A R1 demostrem que ampliant el mostreig taxonòmic, els llinatges holozous també tenen aquesta ruta metabòlica, només l'han perdut els animals. A més la història evolutiva d'aquestes rutes metabòliques involucra esdeveniments de paralogia amagada o transferència horitzontal de gens.

Les dues filogenòmiques a R2 i R3 posicionen tots els llinatges coneguts amb més o menys congruència un cop comprovat que la matriu de seqüències, preparada especialment per no patir errors d'ortologia, no pateix d'errors sistemàtics.

Es recuperen congruentment els llinatges Apusomonadida i Breviatea com a grup germà d'opistoconts, probablement de manera parafilètica.

Dins opistoconts, els Holomycota s'enbraquen començant pels Nucleariids, seguits d'opisthosporidia (representats per Rozella i Microsporidia) i seguits pels fongs, on els Chytridiomycota i Blastocladyomycota són probablement parafilètics.

Dins Holozoa, animals i Choanomonada són grups germans, seguits dels Filasterea (tots junts agrupats com a Filozoa). Com a primer grup d'holozous recuperem Corallochytrium limacisporum com a grup germà d'Ichthyosporea, amb important implicació evolutiva, tot i que amb suport moderat. Aquesta agrupació implica que l'osmotrofia es va inventar només dos cops dins els opistoconts, a l'ancestre de fongs i a l'acestre de Corallochytrium més Ichthyosporea.

4.2 Hipòtesis actuals de l'origen dels opistoconts

Per inferir transicions evolutives entre llinatges, primer cal reconstruir els caràcters ancestrals dels diferents grups d'opistoconts i els seus ancestres. La parafília d'Ancyromonadida i Apusomonadida, amb la probable de Breviatea convida a interpretar que els caràcters en comú són homòlegs tot i els milions d'anys d'evolució que els separen. La seva morfologia cel·lular es caracteritza per un citoesquelet complex de

tubulina que confereix una forma específica d'alimentació. Tenen un solc ventral per on el flagell posterior passa tot fent un corrent per atraure els bacteris que fagocitaran específicament per aquella regió. Sembla doncs, que aquest modus vivendi és el que ha mantingut una morfologia concreta durant tant de temps.

4.2.1 Reconstruint caràcters ancestrals

Entre els coanoflagel·lats, els craspedids (nus i sèssils) són probablement els que retenen més característiques ancestrals comparats amb la resta, que tenen una cota de malla d'adquisició probablement per transferència horitzontal. Amb només dos Filasterea és difícil consiliar caràcters ancestrals, però la presència de flagell a M. vibrans, un caràcter clarament ancestral, fa pensar que el grup podria haver sigut originàriament marí i amb estadi sèssil. Dins els holozous osmòtrofs, Corallochytrium presenta la morfologia cel·lular més semblant a Ichthyosporea ancestrals, mantenint a més estadi flagel·lat i ameboide per a la dispersió. Per tant, és probablement una espècie que reté caràcters ancestrals.

Dins els holomycots els Nucleariids són ancestralment d'aigua dolça i no flagel·lats, però els Opisthosporidia contenen tots els caràcters presents als avantpassats biflagel·lats, incloent la fagotrofia. A més tenen quitina, cosa que la resta de llinatges fúngics han diversificat tot passant-se a l'osmotrofia.

4.2.2 Reconstruint l'avantpassat comú dels opistoconts

Entre els actuals ancestres i els actuals opistoconts no hi ha una graduació o una clara interpretació d'estadis intermedis. Els caràcters ancestrals en comú entre opistoconts són un únic flagell, la fagotrofia i l'ús de filopodis per predar i adherir-se als substrats (vida bentònica). Cap d'aquests caràcters és una innovació d'opistoconts. La principal diferencia és l'explotació d'aquests per a funcions diverses.

4.2.3 Convergències evolutives

El que sí que tenen en comú certs opistocots, és precisament entre llinatges no directament emparentats.

Els fongs quítrids són clarament similars al Corallochytrium i Ichthyosporea. Tots osmòtrofs, aquàtics, creixement coenocític i estadis dispersius/reproductius flagel·lats i o ameboids. A R3 hem demostrat que els components ancestrals de la paret cel·lular de fongs es troben diversificats de manera independent als osmotrofs holozous, posant un clar exemple d'evolució paral·lela.

Les amebes nues amb filopodis també són massa semblants com per no explorar aquesta possible convergència evolutiva. El flagell de Ministeria demostra que el modus vivendi entre els nucleariids i els Filastera ha de ser convergent i no ancestral.

Així doncs, les dades produïdes i la filogènia establerta serviran per comprar el contingut genòmic per estudiar factors clau per entrendre la transició de fagotrofia a osmotrofia i de flagel·lat a ameba filopodial.

CONCLUSIONS

1. Les “sinapomorfies” o caràcters morfològics i moleculars compartits entre llinatges, no són marcadors filogenètics vàlids. Només poden validar-se gràcies a la filogènia, en especial la filogenòmica.
2. La filogenòmica no està lliure de problemes. Estudis curosos han d’involucrar anàlisis paral·leles fent servir diferents models evolutius complexos tot minimitzant l’error sistemàtic. En particular, el nostre dataset representa una font independent vàlida per testar hipòtesis filogenètiques.
3. La transcriptòmica és una aproximació vàlida per incrementar el mostreig taxonòmic en filogenòmica. En aquest sentit resolem que:
    a. els Apusomonadida i no els Amoebozoa són el grup germà dels opistoconts. A més, els Ancyromonadida són més llunyans dels opistoconts que els anterior.
    b. Filasterea i Ichthyosporea no són monofilètics, Corallochytrium és probablement el llinatge més proper a Ichthyosporea i junts el primer llinatge que ramifica dels holozous.
4. La transcriptòmica també proveeix informació valuosa per explorar la genòmica comparada, com:
    a. la presència d’estructures flagel·lars a Corallochytrium i Ministeria, possibles exemples de pèrdua progressiva del flagell.
    b. el repertori de quitina sintases en opistoconts i la diversificació independent en cada llinatge osmòtrof.
5. Fortes semblances en morfologia i estil de vida en llinatges independents d’opistoconts: Filasterea i Nucleariids, o fongs quítrids i Corallochytrium més Ichthyosporea són exemples d’evolució paral·lela o convergent.

Agraïments

A hores d'ara, després de ser el primer a entrar al protoMCG i ser l'últim de la fornada a sortir-ne, ben lluny queda aquell 2006 en que vaig anar a buscar-me la vida al departament de genètica animat per la Roser. L'Iñaki em va captar ben ràpid amb les seves idees sobre l'origen dels animals i ben aviat vaig començar a jugar amb els seus bitxos, passant tardes sols entre el lab de la Marta Riutort i el seu despatx compartit amb postdocs. Moltes gràcies a tots els que ens vau ajudar en una època tan incerta. Tot i la meva mala memòria, les primeres PCR i alineaments queden marcats. A final d'aquella època va estar marcat per l'estiu al lab amb un jovial Mendoza i un tocat del bolet com l'Isaac. Però aviat l'Iñaki es va muntar un bon xiringuito al PCB, inflat de calers europeus i el lab estava irreconeixible amb tanta activitat. L'Àlex Pérez bombolleta ens va muntar un lab de luxes i coloraines, el Hiroshi ens va ajudar moltíssim a establir la infraestructura computacional i amb el primer article (arigatou gozaimasu) i l'Arnau va venir amb ganes de quedar-se, enamorat per la boleta amb bracets que és la Capsaspora. Amb la idea de fer una tesi modesta, a l'alçada d'un servidor, vaig passar a ser el ximp de la filogènia (gràcies Mendo pels teus mítics afalacs, ben enregistrats per quan sigui de menester). Sota la tutel·la de l'indescriptible Romain Derelle, entre els seus daltabaixos i el seu talent es va establir un projecte collonut, que no hagués pogut dur a terme sense l'ajuda i la paciència del personatge més entranyable del món, l'amic Jordi (Paaaaaaaps). La Núria (Poooooons) va ser el meu contacte més ferm amb la biologia molecular. Un breu tast que va durar fins que em va tocar encarregar-me d'un petit zoo de protists que no ha parat de créixer gràcies en gran part a l'ecòleg microbià per antonomàsia, gràcies Javi. Gràcies també a tots als qui han passat pel lab (Lluís, Lora, John, Majo) i als que l'han hagut de suportar al PCB, a la UB i a l'IBE. La nova fornada del biotechMCG, liderada pel gentilhome Francesc Xavier, i seguit de prop per la fada Parra, el feliç David (i per molts anys sigui) i la treballadora Nurieta donareu molt a parlar. Moltes gràcies i molts ànims.

Moltes gràcies també al Diego (grazas), l'Ana i de nou al Romain, per ajudar-me a explorar la filogenòmica fins a límits insospitats. També als que m'han acollit quan he sortit a explorar món.

A Halifax, gràcies a l'avi Andrew, Javi, Martin, Eleni, Matt, Michelle & Co. per la gran acollida, però en especial a la Laura per ajudar-me a programar i a la Courtney per no deixar-nos estar massa sobris.

Ja en l'última etapa, gràcies al P59: Joan, Javi, Ana i Marina.

Sobretot gràcies al Philippe i l'Eric per acollir-me a París a canvi de res, i en especial a Slim, PA i Raph, el meu primer èxit en exportar els crits i les freqüents pauses (remercier à tous).

Evidentment una menció especial als germans que no per no ser de sang ho són menys. Arnau i Àlex m'heu permès aprendre molt i superar-me, encara que us costi de creure, sense vosaltres al voltant encara hauria acabat aquesta etapa molt més verd del que estic. Gràcies en especial a l'Arnau per ser el públic més agraït que he trobat mai, en aquest sentit també al Mendo per no deixar-me baixar la guàrdia ni un moment. Gràcies per les festes, els happenings, les cançons i en general a la vida que hem viscut tots aquest anys junts (des dels tirs de pedra a la bassa de la facultat o els cants comunistes a l'ovella negra, i fins on arribem).

I ara moltes gràcies a tothom que m'ha suportat i ajudat no només durant aquesta etapa de doctorat, sinó també a arribar-hi. Les i els que tot i tenir-vos abandonats per treballar massa o per estar repartits per mig món, feu que tot valgui la pena.
Gràcies a la gent de bio, sobretot els que hem superat la separació post-llicenciatura: Cris&Mi, Maria princeseta, Sergi, Trícia, Bet, Roser (no hi haurà mai prou per agrair-t'ho tot), Anna i Marc (a vosaltres us he d'agrair que sempre ho doneu tot per la immersió i per la festa).
I això em porta a la moderna Baetulo, a un informàtic diver salsero pare i millor persona. Gràcies Tonikos, per ser tan bona persona i tant motivat. Tot i estar separats de fa uns anys, gràcies Pablers per ser com ets, per acompanyar-nos des de fa tant en bons i mals moments (per la sinceritat) i perquè tenies una Playstation. Gràcies també a la resta de Badafuckers Beeeto, Pep, Eric, Guillem i Mata i sobretot a la saxofonista, gran amiga i millor persona.
Gràcies també a la família, al Tximo, a la Merceditas i al Manetokun, per fer de la vida una excentricitat constant.

En un espai especial, on pertoca, l'Ares, la persona que fa que qualsevol cosa valgui la pena. Des d'esperar-te dos hores a la fnac, fins a emigrar a París.