Deep eukaryotic phylogenomics : the holomycota branch

Luis Javier Galindo González

université PARIS-SACLAY

Deep Eukaryotic Phylogenomics: The Holomycota Branch

Thèse de doctorat de l'université Paris-Saclay

École doctorale n° 577,
Structure et Dynamique des Systèmes Vivants (SDSV)
Spécialité de doctorat : Science de la Vie et de la Santé
Unité de recherche : Université Paris-Saclay, CNRS, AgroParisTech, Ecologie Systématique et Evolution, 91405, Orsay, France.
Référent : Faculté des sciences

Thèse présentée et soutenue à Orsay, le 11 juin 2020, par

Luis Javier GALINDO GONZÁLEZ

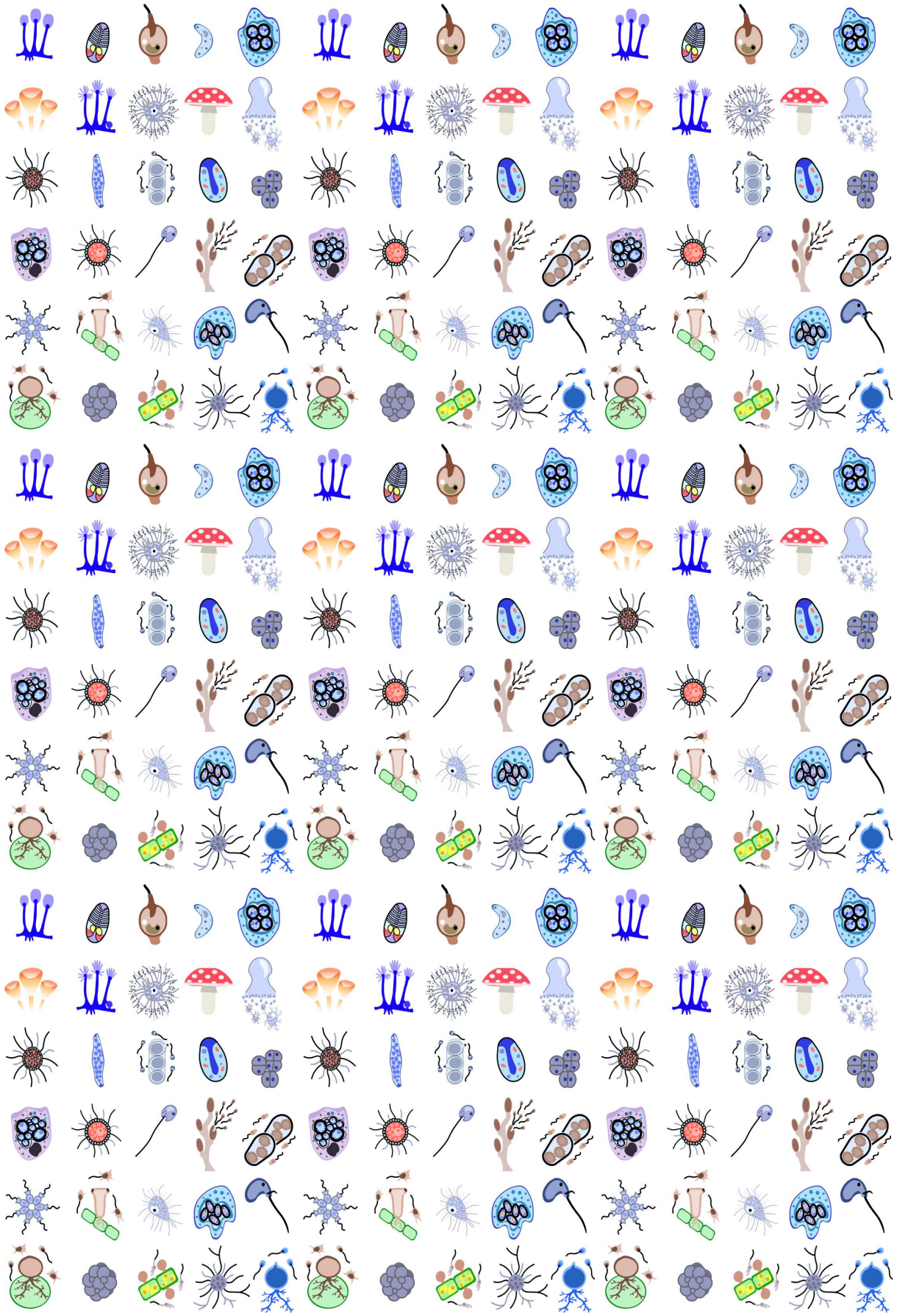
Composition du Jury

Cécile FAIRHEAD Professeur, Université Paris-Saclay	Présidente
Ingrid LAFONTAINE Maître de conférences HDR, UPMC	Rapportrice & Examinatrice
Philippe SILAR Professeur, Université Paris-Diderot	Rapporteur & Examinateur
Ramon MASSANA Professeur, ICM-CSIC (Biologia Marina i Oceanografia)	Examinateur
Tom RICHARDS Professeur, University of Oxford	Examinateur

Purificación LÓPEZ-GARCÍA Directeur de recherche, Université Paris-Saclay	Directrice de thèse
David Moreira Directeur de recherche, Université Paris-Saclay	Co-Directeur de thèse

Thèse de doctorat

NNT : 2020UPASS050



Funded by the Horizon 2020
Framework Programme of the
European Union

Acknowledgements

Developing my PhD in Paris was the best choice I could have ever made both professionally and personally. I would like to thank in the following lines to some of the people that made it possible.

Quiero agradecer a mi familia que siempre ha estado ahí para mí y siempre me han apoyado en mis “rarezas de animalitos y bichos”. Tanto a los que siguen en Venezuela como los que están en España, Republica Dominicana o cualquier rincón del mundo, gracias.

Mamá y **Papá**, pienso que podría escribir otra tesis solo agradeciéndoles todo lo que han hecho por mí y, aun así, no sería suficiente. Ustedes han creído en mí cuando yo no creía, me han guiado cuando yo estaba perdido y me han llevado a donde ustedes sabían que estaban mis sueños. Ustedes tomaron la decisión de iniciar una nueva vida conmigo en España, y lo hicieron con todo en contra, incluyéndome a mí, un adolescente impertinente negado a irse de Venezuela y de su zona de confort. Ustedes se fueron sin trabajo, sin casa, sin red de seguridad y dejaron atrás amigos y familia. Dejaron todo atrás para dármelo todo a mí y nunca en mi vida podré devolverles lo suficiente. Si hoy soy biólogo y, quizás doctor, es gracias a ustedes.

Gracias **tía Margarita**, gracias, nos acogiste a mis padres y a mí cuando fuimos a Tenerife a iniciar una nueva vida y a día de hoy nos sigues acogiendo con una sonrisa, gracias.

A mis hermanos **Claudito**, **Yorimar** y **Maryori**, me duele no poder verles más a ustedes y a todas mis sobrinas. Siempre les llevo conmigo.

Ori, eres una prima fantástica y nunca olvidaré nuestras sesiones terapéuticas amorosas. Que haya más sesiones si hace falta, pero sobre todo que haya más cervezas y risas.

Puri y **David**, gracias por haber confiado en mí para llevar a cabo este proyecto y gracias por toda vuestra ayuda durante la realización del mismo. Especialmente quiero agradecer que no importa cuando, la puerta de vuestro despacho siempre ha estado abierta. Este ha sido un camino muy exploratorio para todos, pero vuestra seguridad en que todo iba a llegar a buen puerto siempre me tranquilizó y ayudó a mantener la vista en el objetivo. Todo lo que he aprendido estos años gracias a ustedes y al increíble DEEM team que habéis creado lo llevare conmigo durante toda mi carrera.

Office 208! I will never forget that number nor the people within that crazy office.

Guillaume, these years would not have been the same without you, you crazy and sexy bastard! Because of you our office was a place of laughter and craziness, the perfect environment for making a PhD. You were there in beers, parties and pain. Additionally, I cannot give you credit enough for all the work-related help. If I am a little bit more of a “puro hacker” is thanks to you. Thank you for the patience every time I came to you with a bioinformatic question.

Gwendoline, I mean, how many favors and help have you given to me during the last years? Without you I would probably still be making the first inscription to the École doctorale! And not only that, you also turn out to be a great friend, companion, and desk neighbor. Thank you for always being there with a huge smile on your face.

Jodie, oh Jodie, oh Jodie playing jokes to each other (ok, maybe I did a bit more jokes) made these years so much fun. I will always remember your epic dramas and situations you always get yourself into. I want you to know that no matter how hard things have been for you lately I will always be there for you.

Rafael, Raphou, respondiéndole a los agradecimientos de tu tesis he de decir que nunca es tarde para pedir una orden de alejamiento! Gracias por ser un pinche puto cabrón de amigo.

To the new guys! **Thomas**, thank you for showing me two things, that everything can be bought in amazon and that you can laugh even after burning half of your body. **Fabian**, Dutch jokes and all! You are a great friend and I really enjoy having you in the office.

Naoji san, I will always remember our trip to Cyprus. Thank you for also for sharing your infinite skills and knowledge about protist and the microscopic world. P.D.: NUTSCAPING!

Ana, miau, miau, miau miau miau. Traducción: ¡Gracias por todo, me he reído muchísimo contigo estos últimos años y que sigan las risas por muchos más!

Miguel, si no te lo han dicho ya te lo digo yo, eres un cacho pan, de verdad gracias por tu amabilidad eterna.

To **Paola**, **Maria**, **Ludwig** and **Philippe** thank you for your technical assistance in the lab, with the computer and in other science related problems, working with you makes everything easier.

Guifré, has sido uno de mis supervisores y consejeros durante esta tesis y por eso siempre te agradeceré. Me acogiste y me has enseñado todo lo que sabías con toda la paciencia del mundo, del “punk science” al “puro hackerismo”. Que este manuscrito vea la luz es en gran parte gracias a tu tutorización y sinceridad en todo momento. Y, por si fuera poco, eres además un gran amigo,

con el que he viajado, compartido y vivido muchos momentos que siempre recordaré. Moltes gràcies!

Mariano, has sido mi padre en la ciencia desde que me adoptaste en tu laboratorio durante la carrera, estoy aquí hoy gracias a que apostaste por mi. Cada navidad vuelvo con ganas de asistir a otra jornada del almendro. Igualmente gracias **Teresa (Titi)** por tu tutela y asesoramiento en el laboratorio durante mi TFG, y por supuesto, por esa risa tan contagiosa.

Thanks to all the people involved in the SINGEK network. Thanks to all the organization and PIs for making this network the best possible scenario to develop a PhD.
Thanks to all my SINGEKers ESRs! You are all not just some of the best scientist I have ever met but also great friends and companions during this journey. **Kons**! μαλάκα! Thank you for your friendship and help during these years. **Max**, you rock in so many ways, I have never laughed so much like those days sharing a flat with you and François in Barcelona. **Xavi**, **Imer**, **Raphael** and **Felipe (my boy)** I will never forget all the fun we have had.
François my boy, my brother, you are the most precious thing I will take home with me after the SINGEK years. I have never had a strongest positive feedback effect over my stupidity as the one that I have whenever I am with you. All good memories from SINGEK have your face in it, see you soon.

A mis amigos de Venezuela: **Héctor (Randal), José, Edu, Ana, CJ, Edmir, Victor, Ker (Chino)** y **Robert,** jamás me olvidaré de ustedes y lo compartido.
A mis amigos de Tenerife. **David**, tío, que nunca dejemos de irnos de bares para crear nuevos recuerdos que atesorar, gracias por estar siempre ahí. **Néstor**, tenerte durante los años de la carrera fue genial, y todas esas noches de estudio y muchas de fiesta te tienen de protagonista. **Samu**, gracias por estar ahí y por todas las paridas. **Cristian**, tu risa y tus “chistacos” siempre resonarán en mi cabeza. **Ale** y **Adri**, gracias por todas las noches de fiesta y los días de amistad. **Miguel**, no solo eres el mejor guitarrista que conozco sino una de las personas más genuinamente buenas que he conocido. By The Way, nunca olvidemos Garachico. **Santi**, Santi, Santi chiquito de payaso er nota. Sabes que estoy aquí porque te encontré durante la carrera, ¿no? Espero que seas consciente de que aquellas noches en las que arreglábamos el mundo, en las que

nos creíamos Darwin y Wallace descifrando el universo científico, me han inspirado más que muchos profesores de la carrera. No solo eres genial intelectualmente sino el mejor amigo que puedo desear. **Tania**, eres de las cosas más bonitas que tengo, no puedo resumirlo de otra forma, mil gracias por estar siempre ahí. La de risas que me he echado contigo y la de risas que vendrán son incalculables. No importa lo que pase, ya sé que te tendré para toda la vida, y recordarlo siempre me saca una sonrisa.

El **Colesp**, aquí también podría estar páginas agradeciendo a toda la increíble gente con la que he convivido durante los últimos años, sois tantos y todos tan especiales que pido disculpas a todos aquellos y aquellas que, aunque no estén reflejados en este manuscrito, sí lo están en mis recuerdos. **Manuel**, sé que no te gustan las ñoñadas pero te vas a aguantar un poquito. Y te vas a aguantar porque no puedo agradecerte lo suficiente todo lo que has hecho por mí. Tu amistad ha sido uno de los pilares principales de mi vida en París. Hay tantos recuerdos que tengo contigo que pensar en mi doctorado será siempre sinónimo de pensar en ti. Gracias por todo, y siempre estaré a tu lado más allá de París. **Ari**, mi urraca, sin tí el Colesp no hubiera sido lo que ha sido y sabes que cuando marchaste nunca ha vuelto a ser igual. Mil gracias por estar a mi lado, por ser mi consejera y mi desahogo. **Marta**, otra urraca catalana más. Gracias por todas las risas, que han sido muchas, y que no me cabe duda de que serán más. **Errepunto**, no te imaginas cuánto me has enseñado sobre esa cosa llamada vivir. Viéndote seguir tus sueños he aprendido una cosa o dos sobre seguir los míos. **Jorge**, mi matemático favorito, si el primer año del Colesp ha sido el mejor es en gran parte gracias a ti. **Ana**, mi barbateña, gracias por haber estado y seguir ahí. **Julia**, mi murciana pelirroja, mi presi, cuánto apoyo y cuánta amistad me has dado, nunca pares, porfa. Gracias a las **Paulas**, entre otras muchas cosas, a **Portugal** por toda la comida que se ha comido en mi cara y a **Jaca** por todos esos postres riquísimos. **Jorge** y **Miguel**, sea el individuo o sea la población ustedes han sido mi selección favorita (os habéis quedado locos). **Andrés**, illo, tu guitarra, tu persona, tus tocs, tu arte, tú y solo tú, mil gracias por ser tú. **Silvia**, gracias por tu amistad incondicional y por tu dulzura imperecedera. **Gabi**, la arepa es venezolana. **Sonia**, aún escucho tu risa retumbando por el Colesp. **Marta Asturias**, gracias por toda tu bondad. **Cesar**, de manifa en manifa. Y gracias a **Sergio**, **María**, **Nico**, **los Manu**, **Aida**, **los Carlos**, **Rubén** y a todo el resto de las personas que han hecho del Colesp una casa.

Por último, gracias a ti, **Alba**, mi compañera. Tenerte a mi lado compartiendo los momentos que hemos vivido lo ha sido todo. Has estado durante los momentos más duros de la escritura de esta tesis manteniéndome cuerdo, recordándome que no me estaba muriendo y que esto pasaría pronto. Gracias por gustarte mis frikadas y mis bichos, gracias por hacer de mis gustos los tuyos, gracias por hacer de mi vida, nuestra vida.

Index

1. Introduction

"For a deeper understanding of our present development and the view of the world that is built upon it, there can be few branches of natural science of such importance as the natural history of the lower forms of life, the so-called Protista"

Ernst Haeckel. *Das Protistenreich* (1878)

1. Introduction

1.2. A historical perspective on protistology

Letters among scientists have been for long time the carriers of knowledge around the world, allowing great minds to share ideas, in a polite, or not so polite manner. On February 5, 1676 Sir Isaac Newton wrote to his long-standing rival Robert Hooke, one of these not so polite letters. He wrote:

"What Descartes did was a good step. You have added much several ways, & especially in taking the colours of thin plates into philosophical consideration. **If I have seen further it is by standing on the shoulders of giants**."

Some have seen in this last phrase nothing more than a sarcastic mock from Newton towards Hooke's not so short stature. However, I consider it is worth spending some time to analyse both the phrase and the target of Newton's (potential) mock.

This PhD work has come into existence only because we all stand on the shoulder of great scientists that have made history; these are our giants. Giants that in many occasions discussed and argued ideas and hypothesis among them. Importantly, these discussions (beyond sarcastic mockery to someone's physical aspect), sometimes led to key findings and on the whole, contributed to their particular field of knowledge, in our case, protistology. In the following lines, I will try to highlight some of the most important scientists, findings, and discussions on microbial eukaryotes or, generally, protists, to which my work intends to contribute.

It is precisely Robert Hooke one of the first relevant names worth mentioning. Not only famous because of his confrontations with Sir Isaac Newton, Robert Hooke is seen as one of the most important experimental scientists in history. He largely contributed to different fields of science, but it is the publication in 1665 of his book *Micrographia* which had the most impact (Hooke, 1665). Using an originally designed microscope, he observed, described and illustrated for the first time the microscopic structure of cork, fleas, fossils, etc (Figure 1). These findings pushed him to first coin the term "cell" to describe the most basic structural and functional unit in biology. Cellular biology was born.

We only had to wait 9 years for another giant to make his contribution, and again, there were letters. On September 7, 1674, a 42 years old Antonie van Leeuwenhoek decided to write a letter

to the Secretary of the Royal Society, Mr. Henry Oldenburg. Beyond the description of a new group of organisms, the content of this letter would directly imply the starting point of a new scientific field, protistology (Rothschild, 1989). After using microscopes of his own making on water samples from a nearby lake, Leeuwenhoek reported that he "found, floating therein… very many little animalcules". In October 9, 1676 Leeuwenhoek wrote the first detailed description of these animalcules to the Royal Society, which included rotifers, bacteria, and both pigmented and non-pigmented eukaryotes (Dobell, 1932). Nowadays, we can distinguish in his drawing's several protists including an array of ciliates species (e.g. *Vorticella* sp.) (Figure 1).

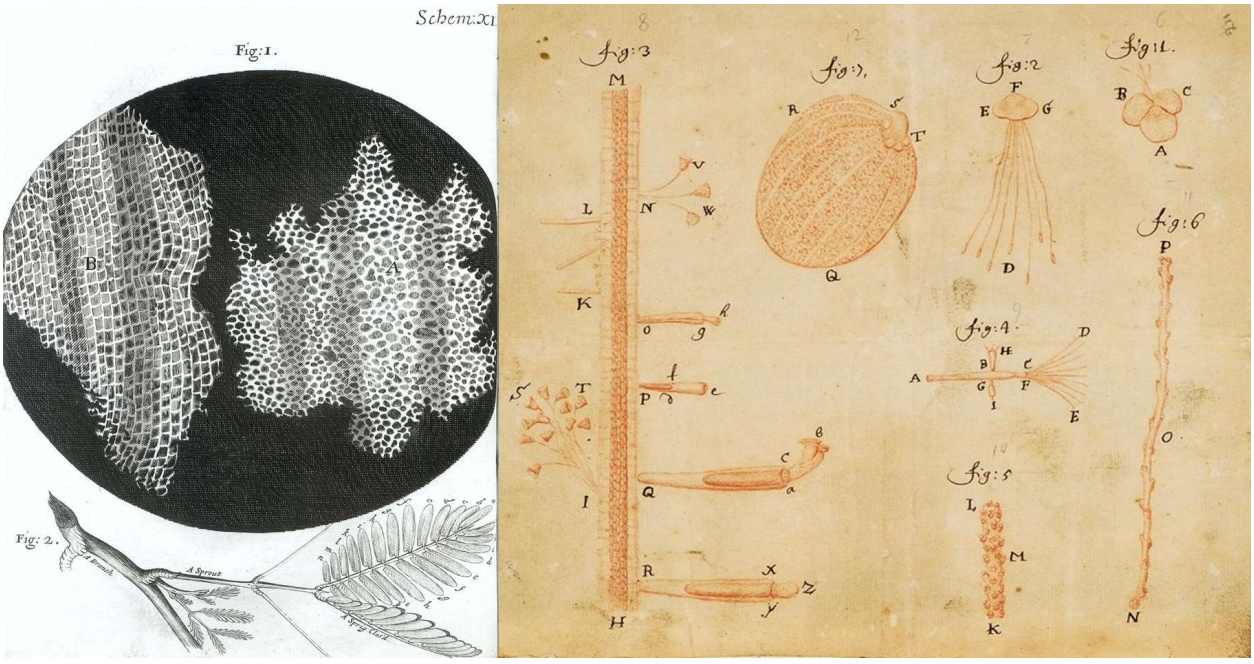


Figure 1. Left panel: Plate XI "Texture of Cork" from *Micrographia* by Robert Hooke, (1665) (Science Museum / Science & Society Picture Library). Right panel: Illustrations of duckweed and freshwater microorganisms ("animalcules") attached to it, included in a letter from Antonie van Leeuwenhoek to the Royal Society, 25 December 1702 (The Royal Society). Images obtained from: National Museums of Scotland (https://www.nms.ac.uk/).

Among the first protist ever classified was *Euglena viridis* (1696), which John Harris and Carl Linnaeus classified as an animal due to its movement. Later on, Linnaeus, in his *Systema Naturae,* classified the majority of known unicellular eukaryotes in the class Vermes, within Zoophyta (Linnaeus, 1758).

During the last half of the 18th century the new term "Infusion animals" or "Infusoria" was coined by Martin Ledermüller to group a large collection of small organisms that included unicellular eukaryotes, bacteria, worms, planarians, etc., all argued to appear in infusions (Cole, 1926). In

1807, Jean Baptiste Lamarck classified the Infusoria as Class 1 of the Animal kingdom and admitted some similarities with plants including the capability of living "entirely by absorption". Not only Lamarck, but also his strong opponent and competitor Georges Cuvier placed the "Infusoires" with animals, in the Zoophyta.

The merit of first coining the name "Protozoa" is usually attributed to Georg A. Goldfuss in his 1817 book *Ueber die Entwicklungsstufen des Thieres*. The term comes from the Greek words "proto" meaning "first" and "zoon" meaning animal". "Protozoa" was created to include unicellular and multicellular organisms from Infusoria, Lithozoa, Phytozoa and Medusar (Dobell, 1932). It was only after the cell theory was enounced by Schleiden and Schwann in the late 1830's that the term protozoa was used to refer to only unicellular organisms, as proposed by Carl Theodor Ernst von Siebold in 1845.

The year 1859 was an outstanding year for biology, naturally famous by the publishing of The Origin of Species by Charles Darwin. However, on that year another milestone in the history of biology occurred (Rothschild, 1989), again related with a discussion between two scientists. During that year, in a conference about palaeontology, Sir Richard Owen set the basis for a text published in 1860, only four months after the publication of The Origin of Species. This text was an attempt to refute Darwin's theory. Although Owen believed in some form of evolutionary theory, he was completely against the idea of evolution by natural selection. Despite Darwin turned out to be right, Owen's rebuttal attempt offered the first delimitation of three divisions of life, including plants, animals and protozoa (separately from animals).

However, protozoa includes the Greek word "zoa", which means animals, preventing these unicellular organisms to form an independent group. John Hogg tried to solve this in 1860 by coining the term "protoctista" meaning "first created beings", implying that the other two kingdoms appeared from these protoctista.

The next remarkable scientist who made considerable contributions to the field of protistology was Ernst Haeckel who, following up Johannes Müller's work on radiolarians and foraminiferans, published his famous monograph *Die Radiolarien* in 1862 (Haeckel Ernst, 1862). Haeckel created some of the most beautiful and intricate artwork on metazoans and protists ever made (Figure 2).

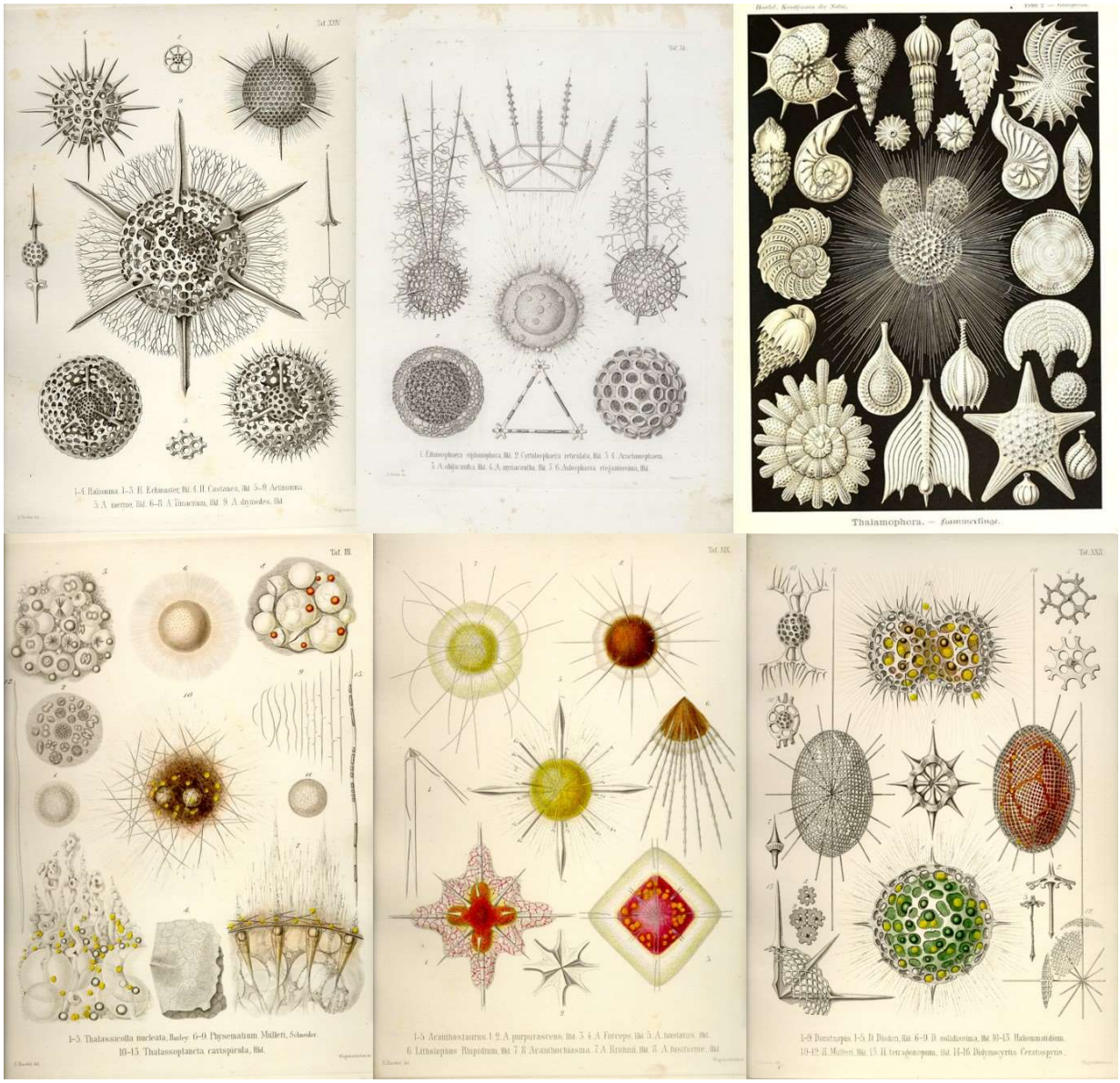


Figure 2. Different panels showing Ernst Haeckel exquisite illustrations on protists (Adapted from Haeckel's compilation books Art Forms in Nature and Art Forms from The Ocean: The Radiolarian Atlas of 1862).

Haeckel was a declared Darwin admirer and decided to spread Darwinism as his scientific life goal. In 1866, he published *Generelle Morphologie der Organismen* (Haeckel, 1866), a manuscript regarded as a foundational work for many biologists and, more specifically, protistologists. In this work, for the first time, Haeckel talked about "phylogenetic" relationships to refer to the natural, evolutionary relationships of beings. Also, not without hesitation (since he considered his knowledge on the topic was limited), he created the kingdom Protista "the ones who came first in time", the term that has been now more widely adopted to refer to unicellular eukaryotes (Figure

3). Haeckel even doubted that the Protista were a monophyletic group. Haeckel also included within the Protista, the Moneres (bacteria without the cyanobacteria), and he created many characteristic protistan groups like Flagellata, Diatomeae, Myxomycetes, Rhizopoda, etc. Later on, in 1880, he resurrected Protozoa to refer to those Protista ancestral to animals, and "Protophyta", to name those protists ancestral to plants.

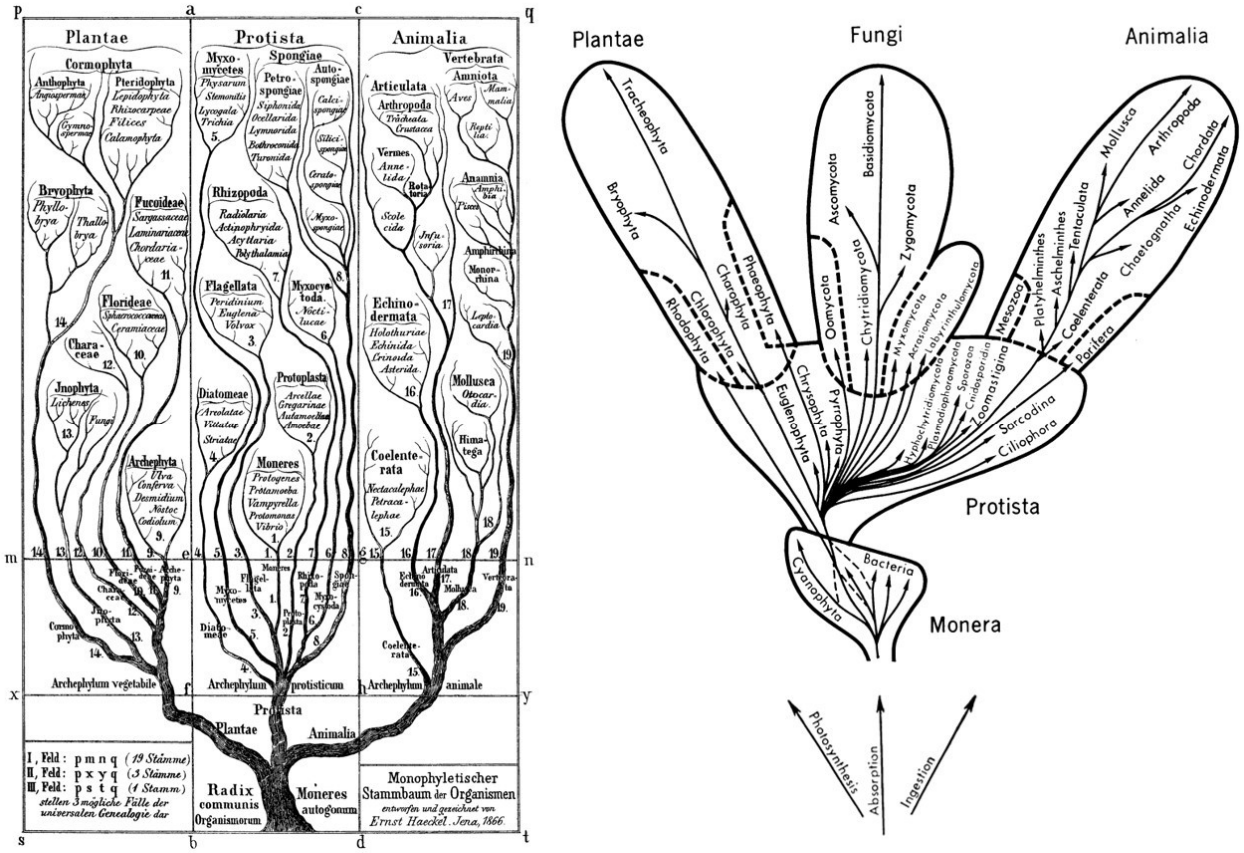


Figure 3. Left: *Monophyletischer Stambaum der Organismen* from *Generelle Morphologie der Organismen* (Haeckel, 1866). Right: Adaptation from Figure 3 of Whittaker's New Concepts of Kingdoms of Organisms (Whittaker, 1969) showing a five-kingdom system based on three levels of organization.

At the beginning of the 20th century several scientists popularized the term Protista. Max Hartmann in 1902 created the first Protista specific journal *Archly für Protistenkunde*, from which important works including "The principles of Protistology" by C. Clifford Dobell was published (Dobell, 1911). In 1938, Herbert F. Copeland proposed a new classification of life including the already established kingdoms Monera (Haeckel), Plantae (Linnaeus), Animalia (Linnaeus) and Protista (Haeckel) in which the Fungi where included.

Several more classifications succeeded until that from Whittaker's work. Robert H. Whittaker's life-time taxonomic work came to a climax with his publication "New Concepts of Kingdoms of Organisms" (Whittaker, 1969). This publication and its proposal of five kingdoms became the standard and part of all textbooks. The main difference of Whittaker's classification with previous ones was that Monera were raised to have their own Kingdom, instead of being a group within Protista (Figure 3). He kept being exceptive with his own classification and pointed problems including the division between unicellular and multicellular eukaryotes. A problem that keeps being of current debate (see chapter 1.8. Multicellularity: Fungi vs Protist). Beyond the debate over the five kingdoms and the modern awareness of their flaws for our current Domain-based (Woese, Kandler, & Wheelis, 1990) classification of life, Whittaker's classification increased the awareness that most protist are not related with neither animal nor plants.

The second half the 20th century was marked by the molecular revolution. After the structure of DNA was determined (Watson & Crick, 1953) thanks to work by Franklin, Crick and Watson, major findings in other fields led to the current modern synthesis view of evolution (Mayr, 1982). A key step was the recognition that evolutionary information can be stored in biopolymers (Zuckerkandl & Pauling, 1965), which paved the way for the development of molecular phylogeny. Molecular markers started to be sequenced, providing a new kind of homologous characters to study the evolution of life on earth (Mayr, 1982; Kimura, 1983)

During this period scientist began to use ribosomal RNAs as markers to reconstruct phylogenetic trees. This probably represented the biggest revolution in microbiology since Leeuwenhoek picked up a microscope. In 1977, Carl Woese attempted to reconstruct the first universal molecular phylogenetic tree analysing rRNAs from a wide range of organisms. In doing so, he discovered fundamental differences within bacteria. He discovered that a group of bacteria was so different from the rest, as eukaryotes from prokaryotes. He called this group Archaebacteria (Woese & Fox, 1977), which was later renamed Archaea when he established his classification of life in three domains: Archaea, Bacteria and Eucarya (Woese *et al.*, 1990)

During this period, the interest in mycology research as we now know it developed in parallel. One of the most important advances on this field was Fleming's discovery of penicillin from the fungi *Penicillium*, which led him to win the Nobel Prize in 1945. Fundamental mycological research led also to the 'one gene, one enzyme' hypothesis and to a second Nobel Prize for fungal research granted to Beadle and Tatum in 1958. During the mid-60's, new research in biochemistry and

genetics in other fungi (especially in *Saccharomyces cerevisiae*) started to develop. These studies somehow led fungal systems to be able to compete with bacterial systems in the molecular arena.

Another outstanding biologist from this era was Lynn Margulis, whose work significantly contributed to shape biology as we know it today. Margulis is most famous for her "Serial Endosymbiotic Theory" to explain the origin of the eukaryotic cell. However, she also contributed to the increase the knowledge of protists, and their importance. In particular, she highlighted protists as the basis of eukaryotic evolution, including the origin of the multicellular fungi, plants and animals (Margulis, Soyer-Gobillard, & Corliss, 1984; Margulis *et al.*, 1990).

During this period, several molecular studies showed that classical "Protista" were paraphyletic and had different degree of relatedness with multicellular lineages (Cavalier-Smith, 1993a; Baldauf *et al.*, 2000). Increasingly refined molecular phylogenetic trees of eukaryotes led to the recognition of several eukaryotic super-groups, one of which was the Opisthokonta.

The 20th century added two key developments for biology and protistology. One of them was the application of transmission electron microscopy (TEM) to the study of unicellular eukaryotes (Patterson, 1999). The other, more general, was the beginning of the Human Genome Project in 1990, which started the countdown for the genomic era. In terms of protistology and mycology, this genomic age started with the sequencing of the first eukaryotic genome, that of the unicellular fungi *Saccharomyces cerevisae,* in 1996 (Goffeau *et al.*, 1996).

During the 21st century, most advances in eukaryotic microbiology have come from the genomic field. The human genome finally got sequenced in 2003 and, with the development of high-throughput sequencing in 2005 (Loman *et al.*, 2012), thousands of new sequenced eukaryotic genomes followed. The obtention of that wealth of new genomic data made it possible to combine multiple conserved markers into new multi-gene phylogenomic studies (e.g. Bapteste *et al.*, 2002; Ruiz-Trillo *et al.*, 2004; Burki *et al.*, 2007; Torruella *et al.*, 2015; Derelle *et al.*, 2016).

All these breakthroughs in molecular techniques and new microscopy data have led to a considerable improvement on protist evolution studies in recent decades. This is largely due to the contribution of single-gene and multi-gene molecular phylogenies, which have consistently improved the tree of eukaryotes, now divided into six or eight supergroups (Simpson, Inagaki, & Roger, 2004; Keeling *et al.*, 2005; Burki *et al.*, 2019). However, the root of the eukaryotes and the relationships among most eukaryotic supergroups remain uncertain (Figure 4).

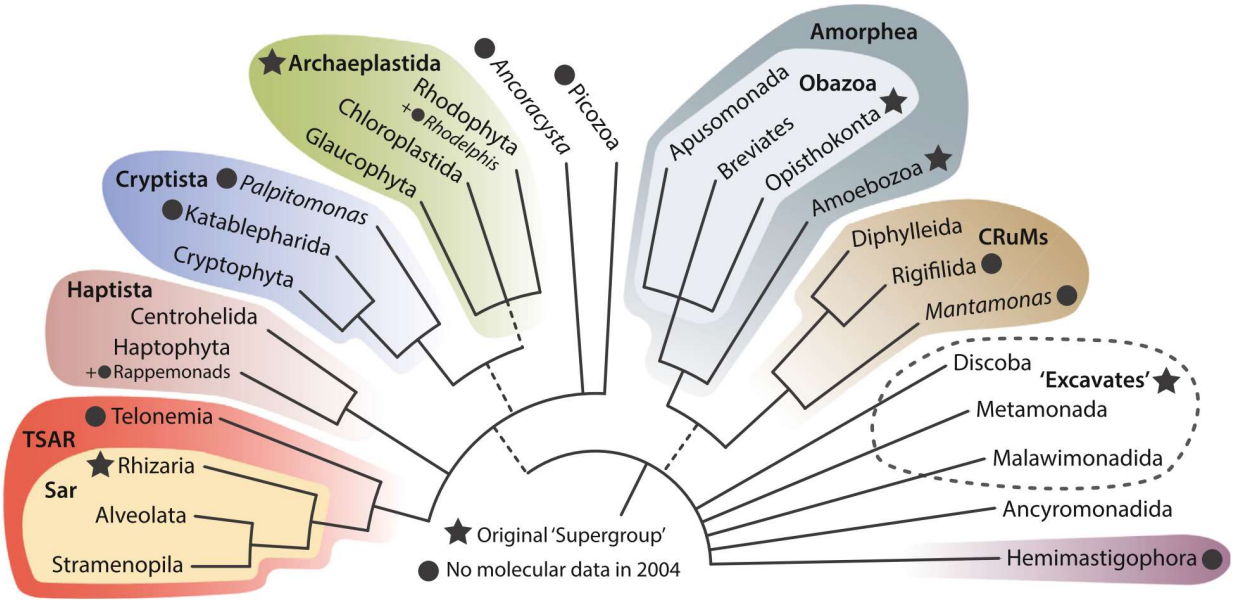


Figure 4. The most recent version of the eukaryotic tree of life. Adapted from Figure 1 from The New Tree of Eukaryotes (Burki *et al.*, 2019).

Recently, the discovery of the Asgard archaea first through metagenomics techniques and more recently by culture has strengthened the idea that eukaryotes derived from archaea-bacteria symbiosis and that the tree of life is formed by two primary domains - the 2 domain (2D) hypothesis (Williams *et al.*, 2013, 2020; Spang *et al.*, 2015). In this hypothesis eukaryotes are nested within archaea (López-García & Moreira, 2015). Thus, eukaryotes are a secondary domain arising from the evolutionary merging of two primary domains (Archaea and Bacteria). A representative species from this group of Asgard archaea was recently cultured for the first time, highlighting the importance of cell-culturing in the middle of the genomic era (Imachi *et al.*, 2020).

1.3. Opisthokonta

Opisthokonta is one of the main eukaryotic supergroups, formally proposed for the first time by Thomas Cavalier-Smith in 1987. The clade was proposed based in two main morphological synapomorphies: a single posterior flagellum (secondarily lost in some members) and flat mitochondrial cristae (Cavalier-Smith, 1987). Opisthokonta encompass exclusively lineages of heterotrophic organisms. Since very early, the monophyly of Opisthokonta proved to be well supported by both, molecular phylogenies of the SSU rRNA gene and multigene phylogenomic

studies (Baldauf *et al.*, 2000; Lang *et al.*, 2002; Medina *et al.*, 2003; Ruiz-Trillo *et al.*, 2004, 2008; Steenkamp, Wright, & Baldauf, 2006; Torruella *et al.*, 2012, 2015, 2018; Paps *et al.*, 2013; Del Campo *et al.*, 2014; del Campo *et al.*, 2015; Arroyo *et al.*, 2018; López-Escardó *et al.*, 2018). Molecular synapomorphies have been also suggested, such as a 9-17 amino acid insertion in the elongation factor 1 alpha (Baldauf & Palmert, 1993; Baldauf & Steenkamp, 2004; Steenkamp *et al.*, 2006).

Opisthokonta comprise the multicellular Metazoa (Haeckel, 1874) and Fungi (Moore, 1980) together with several groups of unicellular relatives. New representatives of these unicellular lineages have been described since the proposal of the original opisthokont clade as well as several new lineages without known representatives inferred from environmental molecular studies (Paps & Ruiz-Trillo, 2010; Nagahama *et al.*, 2011; Del Campo & Ruiz-Trillo, 2013; del Campo *et al.*, 2015; Arroyo *et al.*, 2018). All these analyses suggest that all the opisthokontan diversity falls into two main clades: the Holozoa (Lang *et al.*, 2002), including Metazoa and unicellular relatives, and the Holomycota/Nucletmycea (Brown, Spiegel, & Silberman, 2009; Liu *et al.*, 2009) containing, in a similar way, Fungi and their unicellular relatives (Figure 5).

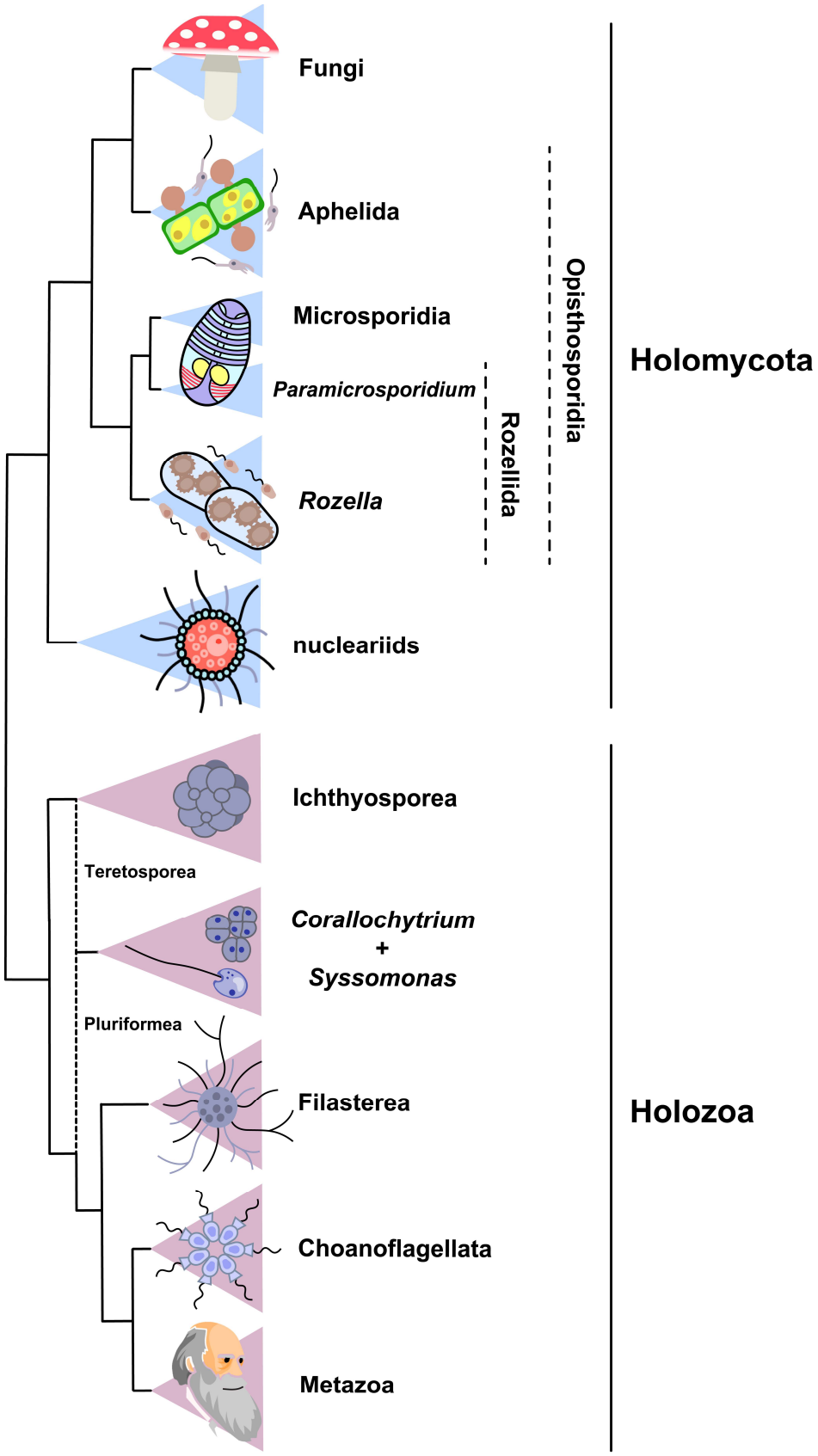

Fungi
Aphelida
Microsporidia
Paramicrosporidium
Rozella
nucleariids
Rozellida
Opisthosporidia
Holomycota
Ichthyosporea
Teretosporea
Corallochytrium
+
Syssomonas
Pluriformea
Filasterea
Choanoflagellata
Metazoa
Holozoa

Figure 5. Overview of the Opisthokonta supergroup phylogeny. Cladogram showing the relationship between the main Opisthokonta lineages, grouped in two main branches Holozoa and Holomycota. Relationships among groups are based on the latest results from phylogenomic studies (cited in text).

1.4. Holozoa

Holozoa was first coined for the clade formed by Ichtyosporea (also named: Mesomycetozoa or DRIPs) (Ragan et al. 1996; Cavalier-Smith 1998; Herr et al. 1999), Choanoflagellata (Clark, 1866) and Metazoa (Lang *et al.*, 2002), which appeared monophyletic in phylogenetic trees of mitochondrial markers. Later on, another group, the Filasterea, was found to be the sister lineage of Choanoflagellata and Metazoa (Ruiz-Trillo *et al.*, 2004; Schalchian-Tabrizi *et al.*, 2008). The position of an enigmatic holozoan, *Corallochytrium limacisporum*, remained controversial within the clade in few-gene phylogenies (Zettler, Nerad, & Sogin, 2001; Steenkamp *et al.*, 2006; Carr *et al.*, 2008; Paps *et al.*, 2013). Later, phylogenomic analyses using more markers derived from transcriptomic data suggested that *Corallochytrium* formed a monophyletic group with Ichtyosporea, both being the sister lineage to all other Holozoans; the name Teretosporea was used to engulf both lineages (Torruella *et al.*, 2015). However, more recently phylogenomic analyses including a new holozoan species, *Syssomonas multiformis*, placed this organism as sister to *Corallochytrium* (Hehenberger *et al.*, 2017). This new holozoan clade has been called Pluriformea and could branch either with filastereans, within the Teretosporea, or between ichtyosporeans and filastereans, forming a new independent clade. Supports are still low to solve their affiliation within Holozoa (Figure 5).

Unicellular holozans present very diverse morphologies (Figure 5) and lifestyles, from free-living forms to symbiotic organisms. Choanoflagellata is a group of free-living bacterivorous flagellates. They are mainly marine but are distributed in other environments including freshwater systems, abyssal plains and anoxic habitats (Nitsche *et al.*, 2007; Wylezich *et al.*, 2012). There are around 250 described species (Carr *et al.*, 2008; Leadbeater, 2014), being the most sampled group of unicellular holozoans, with more cryptic diversity found in environmental studies (e.g. Del Campo & Ruiz-Trillo 2013). They are classified into two monophyletic groups, Craspedida and Acanthoecida, the latter being characterized by the presence of a siliceous cover, the lorica, as synapomorphy (Nitsche *et al.*, 2011).

Filasterea are a considerably less diverse group of unicellular holozoans with only 4 described species. These include two filose amoebae: the snail "symbiont" *Capsaspora owczarzaki* (Owczarzak, Stibbs, & Bayne, 1980; Hertel, Bayne, & Loker, 2002) and the free-living *Ministeria vibrans* (Steinberg, Nygaard, & Steinberg, 1993; Cavalier-Smith & Chao, 2003). Recently, two new species of flagellated filastereans were isolated, *Pigoraptor chileana* and *Pigoraptor vietnamica* both bacterivorous, although they are also the first known unicellular holozoans known to be also capable of feeding on eukaryotic prey.

Ichthyosporea (also named: Mesomycetozoa or DRIPs) are a morphologically and ecologically diverse group of around 40 species, with an osmotrophic/saprotrophic fungal-like lifestyle (Mendoza, Taylor, & Ajello, 2002; Glockling, Marshall, & Gleason, 2013). They are coenocytic parasites or commensals of freshwater, marine and terrestrial Metazoans. They are divided into two main groups differentiated by the presence of a flagellated dispersal stage on the zoosporic Dermocystida, and the ancestral absence of any flagellated form in the amoeba-like Ichthyophonida (Torruella *et al.*, 2015; Donachie, Suga, & Grau-bove, 2017).

The newly proposed clade Pluriformea contains two different species: the free-living coenocytic protist *Corallochytrium limacisporum*, which seems to be osmotrophic and present an exclusively amoeboid dispersal stage, and the phagotrophic predator *Syssomonas multiformis,* displaying a variety of cellular morphologies from amoeboid to flagellated (Hehenberger *et al.*, 2017).

The large morphology and lifestyle diversity observed in Holozoa makes the reconstruction of last universal holozoan ancestor traits difficult; the same is true for Holomycota (see chapter 1.7.5. Fungal synapomorphies). Thus, we need to assess their molecular characters looking for synapomorphies using comparative genomics (see chapter 2 of Martín-durán & Vellutini 2019). Some of the few molecular synapomorphies of holozoans include the presence of *de novo* transcription factors as Nf-kB, p53 or RUNX (Sebé-Pedrós *et al.*, 2011). However, before further reading, we do need to highlight that the molecular synapomorphies that differentiate unicellular holozoans/holomycotans from their multicellular relatives are rarer than previously thought, especially in Holomycota (see chapter 1.7.5. Fungal synapomorphies). Comparative genomic studies keep finding evidence that many traits thought to be synapomorphies for multicellular opisthokonts were already present in the unicellular relatives of, respectively, animals and true multicellular Fungi (e.g. King et al. 2008; Richter & King 2013; Donachie et al. 2017; Richards et al. 2017; Kiss et al. 2019).

1.5. History of holomycotan (Fungi) taxonomy

Humans know fungi probably since the early history of humankind and its relationship with nature. One of the first texts in which fungi are mentioned is attributed to the Greek poet Euripides. On the 5th century BC (450-456 BC), he was on a visit to Icarus when a woman and all her sons died from mushroom poisoning after they had collected them in the fields. Euripides wrote an epigram translated by Hougthon (original cite in Ainsworth, 1976) as follows:

"O Sun, that cleavest the undying vault of heaven, hast thou ever seen such a calamity as this? – a mother and maiden daughter and two sons destroyed by pitiless fate in one day?"

From this first written acknowledgment of the impact of fungi on human life, our knowledge of fungi and its classification has come a long way.

The first fungal classification was established by the Dutch Christiaan Hendrik Persoon (1761-1836), who is now considered the founder of modern mycology. He created the first taxonomy of mushrooms based on the binomial nomenclature system proposed by Carl Linnaeus in his Species plantarum (1753). His work *Synopsis Methodica Fungorum* was considered the golden book of mycology (Persoon, 1801).

Decades later, Whittaker's classification of the natural world in Kingdoms, attributed Fungi its own kingdom (Whittaker, 1969). This classification, mainly based on the osmotrophic capabilities of Fungi, led to the incorporation within Fungi of several organisms that turned out to lack close phylogenetic association. Subsequent molecular phylogenetic analyses placed them in other branches of the tree of life (Barr, 1992; Van De Peer & De Wachter, 1997; Dick, 1999; Riethmüller, Weiß, & Oberwinkler, 1999; Hausner, Belkhiri, & Klassen, 2000), including: the phyla Myxomycota (Amoebozoa), Acrasiomycota (Excavata), Labyrinthulomycota (Stramenopiles) and Oomycota (Stramenopiles), all into the subkingdom Gymnomycota. The community did not have to wait for the molecular revolution to obtain a reclassification of these lineages, since the first correction came from Whittaker himself. On May of 1969, five months after the publishing of its classification, Lindsay S. Olive together with Whittaker decided to move Gymnomycota outside of Fungi (Olive & Whittaker, 1969). However, four phyla sharing morphological and reproductive characteristics withstood and were popularized among the

community, classified as Eumycota: Chytridiomycota, Zygomycota, Ascomycota and Basidiomycota.

For decades that Whittaker's classification of Fungi comprising only four phyla was the norm (Barr, 1992; Cannon & Hawksworth, 1995; Alexopoulos, Mims, & Blackwell, 1996). But the molecular era and the first molecular phylogenies of the SSU rRNA gene, changed everything. These studies showed that the phyla Chytridiomycota and Zygomycota were not monophyletic (Nagahama *et al.*, 1995; Tanabe, Watanabe, & Sugiyama, 2005; James *et al.*, 2006a, 2006b; White *et al.*, 2006). This led to the proposal of new phyla: Blastocladiomycota (previously the class Blastocladiales within chytrids) (James *et al.*, 2006a) and Zoopagomycota and Mucoromycota (previously grouped together as the Zygomycota) (Spatafora *et al.*, 2016b).

In parallel, the addition of unicellular relatives to the fungal taxonomic landscape began with Microsporidia. Microsporidia are a lineage that had an uncertain affinity to any eukaryotic group in early molecular phylogenies due to their fast-evolving genomes (Philippe *et al.*, 2000), although they were suggested to be related with Fungi (Hirt *et al.*, 1999; Keeling, Luker, & Palmer, 1999; Fischer & Palmer, 2005). *Rozella,* a genus of flagellated endobiotic parasites, also displayed affinity with Fungi, and later was shown to form a monophyletic clade with Microsporidia (James *et al.*, 2013a). Environmental studies showed that the group represented by *Rozella* was widely diverse and received several synonymic designations including Rozellida (Lara, Moreira, & López-García, 2010), Cryptomycota (Jones *et al.*, 2011b) and Rozellomycota (Corsaro *et al.*, 2014b). The also flagellated endobiotic parasitoids known as Aphelida were included together with Microsporidia and rozellids in the clade known as Opisthosporidia (Karpov *et al.*, 2014), sister to all other Fungi. However, that affiliation was based in very few gene markers and recent phylogenomic analyses with many more markers suggest that Opisthosporidia are paraphyletic, aphelids branching as the sister lineage to (classical) Fungi to the exclusion of rozellids and Microsporidia (Tedersoo *et al.*, 2018; Torruella *et al.*, 2018). Lastly, the filopodiated nucleariid amoeba englobing the genus *Nuclearia* (Zettler *et al.*, 2001; Medina *et al.*, 2003), the aggregative amoeba *Fonticula* (Brown *et al.*, 2009), and the small filopodiated amoeba *Parvularia atlantis* (López-Escardó *et al.*, 2018) were shown to be part of a monophyletic lineage sister to all other holomycotan clades (Opisthosporidia + Fungi) (Figure 5).

As seen, the evolutionary, phylogenetic and taxonomic history of Holomycota is tightly linked to those of Fungi. Thus, both terms mix frequently in general bibliography. It is also worth having in

mind that phylogenomic analyses are constantly reshaping holomycotan classification. Thus, the holomycotan classification is still undergoing changes, which we will review in depth in the following chapters.

Nevertheless, we will attempt to follow a universal terminology, for the present manuscript. In the following, we refer to "Fungi" as including osmotrophic organisms forming a monophyletic group that comprises: the unicellular coenocytic Chytridiomycota, Blastocladiomycota and Zoopagomycota, and the multicellular Mucoromycota and Dikarya. All other unicellular lineages related to Fungi, including aphelids, rozellids, Microsporidia and nucleariids will be referred to as unicellular holomycotans (see chapter 1.8. Multicellularity: Fungi vs Protist).

1.6. Pioneer women in mycology

Women have historically struggled to have access to education and, consequently, their contribution in science is limited compared to that of men (Maroske & May, 2018). Universities were restricted only to male students in the majority of Western countries until the late 19th century. For instance, women could obtain degrees and develop careers in neither Oxford nor Cambridge until the twentieth century (Heilbron, 2003). This possibly explains why I found so few women scientists before the 1900s during my bibliographic research for the historical aspects of this manuscript. Despite these restrictions, several women did actually contribute to the field in those early times. This chapter tries to bring awareness about the role of women during the early years of fungal studies, through the biography of 4 of them (Figure 6).

My first example can be traced back to the 18th century. Contemporary of Linnaeus, Catharina Helena Dörrien (1717-1795) came from a family of scholars and learned botany in an autodidactic manner by observing her father (Viereck, 2000). She was an outstanding taxonomist who described dozens of fungi (e.g. within *Agaricus*), and she became the first woman to name a new fungal taxon (*Lichen centrifugus*).

Half a century later, Marie-Anne Libert (1782-1865) equaled Dörrien and became the second woman to formally name a fungal taxon (*Lejeunia* Lib.; latter renamed *Lejeunea*). She was also self-taught and had two mentors. She was a prolific author on the description of fungal taxa and is an established figure in the history of mycology.

The third example corresponds to Mariette Rousseau (1850-1926) born within an educated middle-class family, his father being a zoology professor at Brussels University. Rousseau was a curator of the mycological collection at the Brussels Botanic Garden. She also organized public exhibitions of fungi in which she encouraged younger workers to follow mycological studies. She became a Knight of the Order of Leopold, Belgium's highest order (Maroske & May, 2018).

Lastly, I will mention Annie Lorrain Smith (1854-1937). In 1878, she took botany classes at the Royal College of Science, London (Imperial College) as an 'occasional student' (a consequence of the restrictive university rules on women). She then dedicated herself to mycology, becoming responsible for identifying most of the incoming fungi for the collections of the Natural History Museum of London, again, as an "unofficial worker". She was also a founding member of the British Mycological Society (of which she was president twice). In 1904, she was in the first group of female fellows of the Linnean Society and served as part of its council.

Figure 6. From left to right and up and down: Catharina Helena Dörrien (Portrait by Friedrich Hauck. Museum Wiesbaden). Marie-Anne Libert (Downloaded from Biodiversity Heritage Library). Mariette Rousseau (Detail from photograph '*Mariette Rousseau and James Ensor in the Rousseau family garden in Brussels*). Annie Lorrain Smith. (The Trustees of the Natural History Museum, London) (Maroske & May, 2018).

1.7. Holomycota

As already mentioned, Holomycota is the second of the two main supergroups of the opisthokont branch. The term Holomycota was first proposed by Liu et al. (2009) to refer to nucleariids plus Fungi. Its synonym Nucletmycea was proposed by Brown et al. (2009), its etymology of Nucletmycea referred to the nucleariid amoeba (Nucl) and (et) Fungi (myc). The name Nucletmycea was proposed slightly earlier but both terms were in principle equivalent and used indistinctively. However, Nucletmycea was created to group Fungi only with *Nuclearia* and *Fonticula*, and from this point of view, Holomycota is a more inclusive term since it was conceived to include all nucleariid amoeba. Holomycota is a term bearing some parallel to Holozoa; both use the prefix "Holo", suggesting an inclusive aspect but also a similar taxonomic status within Opisthokonta. For these reasons, we will use the term Holomycota to refer to this supergroup in rest of the manuscript (Figure 5).

Since 2009, Holomycota has been extensively used in taxonomy, diversity and phylogenetic studies to refer to the defined clade grouping Fungi and their unicellular relatives within Opisthokonta (Del Campo & Ruiz-Trillo, 2013; Torruella *et al.*, 2015, 2018; Arroyo *et al.*, 2018; López-Escardó *et al.*, 2018; Tedersoo *et al.*, 2018; Adl *et al.*, 2019; Keeling & Burki, 2019). Holomycota is a clade that was born in the molecular era from phylogenetic analyses that grouped organisms that otherwise, by morphological characters, were never considered to be related. Due to the large diversity in morphology and lifestyles of its members, the reconstruction of these traits for the last universal holomycotan ancestor is a difficult task. Nonetheless, it is possible to assess its molecular characters looking for synapomorphies via comparative genomics. However, as with the Fungi (see chapter 1.7.5. Fungal synapomorphies), Holomycota synapomorphies remain elusive, and neither molecular nor morphological synapomorphies have been found yet.

The overall classification of Holomycota that we are going to discuss in detail in the next chapters includes 5 major groups as follows: nucleariids, Rozellida, Microsporidia, Aphelida, and Fungi.

Fungi in turn encompass Chytridiomycota, Blastocladiomycota, Zoopagomycota, Mucoromycota, and Dikarya (Figure 5).

1.7.1. Nucleariids

Nucleariid amoeba are an understudied group of non-flagellated, free-living filose amoeba known since 1865 (Cienkowski, 1865). Their overall morphological characteristics include cells with a variable morphology from spherical to flattened, and one of its main characteristics is a prominent nucleus with central nucleolus, that can be syncytial or multinucleated. Nucleariid cells are covered by a mucous envelope or glycocalyx, and present multiple food vacuoles which contract constantly. Another defining trait is the presence of hyaline thin radiating filopodia (branching or not), with knobs of cytoplasm along the filopod that help to elongate and retract. The nucleariid filopodia are never stiff and do not present extrusomes, anastomoses or reticulation. They have mitochondria with either discoid or flat cristae, abundant dictyosomes and lack detectable cytoskeletal elements. Their cell sizes cover a wide range reaching from nanoplantonic to small microplantonic sizes. They seem to be more abundant in freshwater environments but there are also nucleariids in marine and brackish waters (Schulze, 1874; Patterson, 1984).

These morphological features are widely found among different eukaryotic lineages, leading to their historical classification as part of different amoeboid taxa (Cavalier-Smith, 1993a; Patterson, Simpson, & Rogerson, 2000). Cienkowski described the first nucleariid of the genus *Nuclearia* (Cienkowski, 1865), due to the already mentioned morphological features and its ecological relevance in freshwater environments. Since then, it has been one of the most observed nucleariids (Patterson, 1984; Dirren & Posch, 2016; Dirren *et al.*, 2017), and until the late 20th century, many other naked filose amoeba were associated with *Nuclearia* in conflictive taxonomies (Cann & Page, 1979; Patterson *et al.*, 2000; Adl *et al.*, 2019).

Nucleariids have passed through several reclassifications round during history. Rainer (Rainer, 1968) first consider them as heliozoans in the suborder Rotosphaeridia. Then, Page (Page, 1987) proposed the taxon Cristidiscoidida including the families Nucleariidae and Pompholyxophryidae; the latter including the genera *Pompholyxophrys* and *Vampyrellidium* (Patterson, 1983a, 1985; Patterson, Surek, & Melkonian, 1987), and later extended to include *Pinaciophora* (Thomsen, 1978), *Rabdiaster* (Nicholls, 2012a) and *Rabdiophrys* (Roijackers & Siemensma, 1988). Mikjukov

somehow merged both classifications and promoted the use of the name Rotosphaerid(i)a, treating Cristidiscoidida as a junior synonym (Mikrjukov, 1999a) and including in it both *Elaeorhanis* (Greeff, 1873) and *Lithocolla* (Schulze, 1874). Since then, studies have classified scale-bearing rotosphaerids (e.g. *Pompholyxophrys* and *Lithocolla*) under the Rotosphaerida nomenclature (Esteban, Gooday, & Clarke, 2007; Leonov, 2012; Nicholls, 2012b, 2012a; Wujek, 2015). However, molecular phylogenetic studies tend to promote the use of the term Cristidiscoidea, as proposed by Cavalier-Smith (Cavalier-Smith, 1993b). We leave the taxonomical discussion of the clade and the issue of whether some terms may have historical priority over others to the specialists. We will simply refer to the whole clade as nucleariids.

The first molecular phylogenies of the small subunit of the rRNA (18S rRNA gene) placed *Nuclearia* as a deep lineage sister to Fungi (Steenkamp et al. 2006; Brown et al. 2009), a placement later confirmed by multi-gene phylogenies (Shalchian-tabrizi *et al.*, 2008; Liu *et al.*, 2009). The aggregative amoeba *Fonticula alba* (Brown *et al.*, 2009) and the small nucleariid *Parvularia atlantis* (López-Escardó *et al.*, 2018) also proved to be related with nucleariids by phylogenomic studies (Torruella *et al.*, 2015). Therefore, nucleariids appeared as the sister group to all other holomycotan lineages (Figure 5). Environmental metabarcoding studies showed a wide diversity of undescribed nucleariids related with *Nuclearia*, *Fonticula* and *Parvularia* from different freshwater and marine environments (Zettler, Gómez, & Zettler, 2002; Lara *et al.*, 2010; Simon *et al.*, 2015; Arroyo *et al.*, 2018; Heger *et al.*, 2018; López-Escardó *et al.*, 2018; Rodríguez-Martínez *et al.*, 2020). Some of the most characteristic genera of nucleariids are:

- *Nuclearia* is a naked filose freshwater amoeba with large sizes from 10 to 60 µm, feeding from a wide range of eukaryotic organisms, including algae, and cyanobacteria (Dirren *et al.*, 2017) (Figure 7A-B). It is the widest known nucleariid genus with around a dozen of described species, although not all by ultrastructure or molecular phylogeny (Patterson, 1983a, 1984; Yoshida, Nakayama, & Inouye, 2009; Dirren & Posch, 2016). Patterson, based on electron microscopy observations, was capable of separating and unifying nucleariids from other filose amoeba genera including *Nuclearella*, *Nuclearina* or *Nucleosphaerium* (Patterson, 1983b, 1984), into the genus *Nuclearia*. The morphological traits of *Nuclearia* include a highly polymorphic cell shape, cyst formation, multiple nuclei and, at least in the case of *N. moebiusi,* microtubules (Patterson, 1983a). Besides ribosomal rRNA gene data, the only additional molecular data available for the genus comes from two ESTs (Expressed Sequence Tags) sequencing runs of two putative *Nuclearia*

simplex strains (CCAP 1552/4 and CCAP 1552/2), the cultures of which are no longer available. They were analysed with limited taxon sampling and they were claimed to be renamed *N. moebiusi* (CCAP 1552/4) and *N. pattersoni* (CCAP 1552/2).

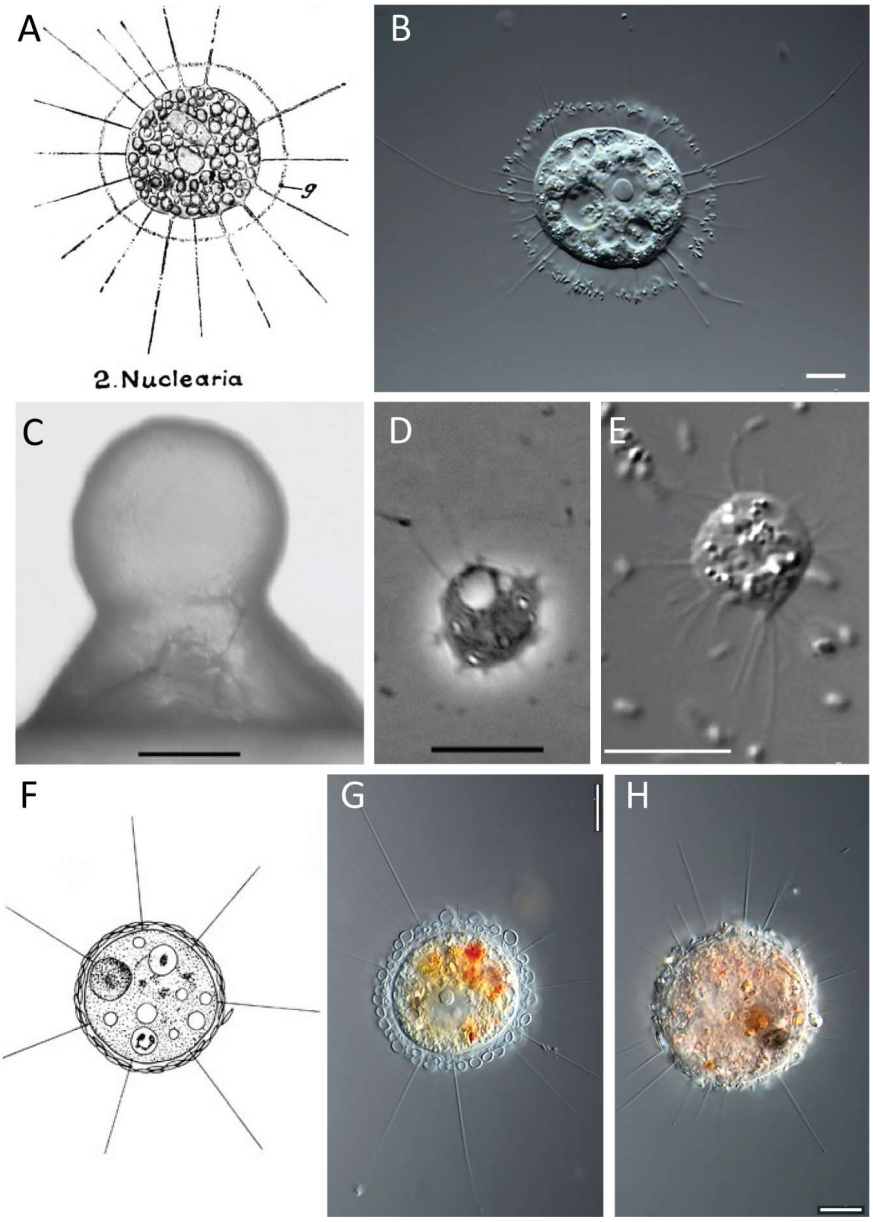


Figure 7. Adapted illustrations and light microscopy images of nucleariids A) Illustration of *Nuclearia* sp. (Parker & Haswell, 1900). B) *Nuclearia* sp. light microscopy image. C-D) *Fonticula alba* sorocarp and individual cell (Brown *et al.*, 2009). E) *Parvularia atlantis* (López-Escardó *et al.*, 2018). F) *Pompholyxophrys stammeri* illustration (Page & Siemensma, 1991) G) *Pompholyxophrys* sp. (http://protistology.com/Tidbits/). H) *Lithocolla* sp. (*Nuclearia*, *Lithocolla* and *Pompholyxophyrs* light microscopic images from: http://www.voelcker.com/). Scale bars: B = 10 µm, C = 100 µm, D = 10 µm, E = 5 µm, G-H = 10 µm.

- *Fonticula alba* (Worley, Raper, & Hohl, 1979) is a small bacterivorous filose amoeba with sizes from 6 to 12 μm, that was isolated from a dog excrement (Figure 7C-D). It has also a variable shape from rounded to elongated and, due to its life cycle including an aggregative stage with the formation of a volcanic-shaped "sorocarp", it was originally described as a myxamoeba or slime mould. This multicellular development has been described only in this species (Prohaska, 1981). *F. alba* has been both molecularly and morphologically described (Worley *et al.*, 1979; Brown *et al.*, 2009). Molecular data for another *Fonticula*-like species (SCN 57-25) is available from a metagenomic study (Kantor *et al.*, 2015). Both fonticulids are related, according to phylogenomic analyses from the same study.

- *Parvularia* is a freshwater small filopodiated bacterivorous amoeba with cell sizes from 3 to 5 μm and represents the most recently described nucleariid genus (Figure 7E). It was obtained from a culture collection under the name *Nuclearia* sp. ATCC 50694 originally isolated from a freshwater sample. Cells can present a single reflecting vacuole occupying the cytoplasm, they have mucous coat with one or, occasionally, two nuclei. They also present radiating branching actin-based filopodia with variable lengths (López-Escardó *et al.*, 2018). Both 18S rRNA gene and transcriptomic data were obtained and used in a single gene and multi-gene phylogenetic frameworks to prove the relatedness of *Parvularia* with *Nuclearia* and *Fonticula* (Torruella *et al.*, 2015; López-Escardó *et al.*, 2018).

Previous studies of nucleariids have failed to resolved the relationships within the clade (e.g. López-Escardó et al. 2018). This is mainly due to the lack of phylogenetic signal and to the limited taxon sampling for nucleariid species. To better resolve their relationships, a wider taxon sampling of nucleariids is needed. In particular, scaled (covered) nucleariids, which morphologically should belong within the Rotosphaerida, were missing in those phylogenetic trees. Unfortunately, no genomic data are available for any of these nucleariid species. Two of the main representative species of the clade are *Pompholyxophrys* and *Lithocolla*.

- *Pompholyxophrys* (Archer, 1869), is a planktonic freshwater filose amoeba that feeds on algae and detritus (Figure 7F-G). Seven species have been described, with sizes ranging from 15 to 66 μm (Roijackers & Siemensma, 1988). Cells are always nearly spherical, tightly surrounded by spherical, ovoid, discoid or bone-shaped perforated silica pearls. These hollow scales are formed endogenously as observed in *P. punicea* (Patterson, 1985) and embedded in a mucilaginous coat. Regarding ultrastructure, it resembles *Nuclearia*, having mitochondria with flat cristae, nucleus

with prominent nucleolus and perinuclear dictyosomes; no microtubule-organizing centre or extrusomes have ever been observed.

- *Lithocolla* (Schulze, 1874), is a genus of filose amoeba ranging from 10 to 50 μm in size found in both marine and freshwater environments (Figure 7H). Similar to *Pompholyxophrys*, *Lithocolla* feeds on algae, diatoms and detritus and possess a silica-based cover which, in this case is formed by exogenous material embedded in its mucilaginous glycocalyx. The exogenous material can be small quartz grains, diatom frustules, or even chalk particles depending on the medium conditions. Cells are mostly spherical, although can be flattened or a bit elongated, and present multiple radiating and variable filopodia, sometimes narrowing at the base and branching.

Other *incertae sedis* filopodiated amoeba, both with naked and covered cells, that have been morphologically related with nucleariids, and will need further molecular assessment are: *Vampyrellidium* (Zopf, 1885b), *Elaeorhanis* (Greeff, 1873), *Pinaciophora* (Greeff, 1873), *Rabdiophrys* (Rainer, 1968), *Rabdiaster* (Mikrjukov, 1999a) and *Thomseniophora* (Nicholls, 2012b).

1.7.2. Opisthosporidia

Opishosporidia (Karpov *et al.*, 2014) is a clade proposed to group three phyla that appear monophyletic in 18S rRNA and RNA polymerase genes: Microsporidia, Rozellida (=Cryptomycota) and Aphelida. The name comes from the combination of Opisthokont and spore. All members are intracellular parasites or parasitoids with an amoeboid vegetative stage. They exhibit spores or cysts with chitin cell wall and possess a specialized apparatus for penetration into the host cell. In some cases, they produce zoospores with filopodia and/or one posterior flagellum (reduced in some cases). Organisms in this clade can be phagotrophic (rozellids and aphelids) or osmotrophic (Microsporidia).

However, new multi-gene molecular analyses suggest the paraphyly of Opisthosporidia. Two studies recover the position of aphelids as the sister lineage to Fungi, in one case in a phylogenomic framework using several proteins from the transcriptome of the aphelid *Paraphelidium tribonemae* (Torruella *et al.*, 2018) and in other case using single-protein or 18S + 28S rRNA gene trees (Tedersoo *et al.*, 2018). Thus, the Opisthosporidia might not be a monophyletic group. We decided to approach Microsporidia, Rozellida and Aphelida as individual lineages.

1.7.2.1.Rozellida

Rozellida, Cryptomycota, or Rozellomycota (Lara *et al.*, 2010; Jones *et al.*, 2011b; Corsaro *et al.*, 2014b) are all terms to refer to the same group of flagellated intracellular parasites of zoosporic Fungi (Chytridiomycota and Blastocladiomycota), Oomycetes and green algae (Gleason *et al.*, 2012). Rozellids possess flagellated zoospores that attach to the host cell initiating an infective process, then the amoeboid protoplasm invades the host through an infection tube, feeds from it and then sporulates, closing the cycle (Powell *et al.*, 2017; Powell & Letcher, 2019). Some species also form resting spores with the presence of chitin. Chitin is also present when invading the hosts cytoplasm's in which is linked to the physical penetration of the infection tube (James & Berbee, 2012; James *et al.*, 2013b).

For some time *Rozella* remained as the only known characterized representative of the lineage. Thus, most discussions were based in conjectures when it came to study rozellids as a whole, since most of them have not been cultured and most of the data came from only one genus (*Rozella*). Recently, a few other representatives of Rozellida, including *Paramicrosporidium* and *Nucleophaga,* have been described. These are microsporidia-like intranuclear parasites without flagellated spore formation (Corsaro *et al.*, 2014b, 2014a, 2016; Quandt *et al.*, 2017). Rozellids are thought to feed through phagotrophy of the host cytoplasm. If true, this is a trait that clearly supports the separation of rozellids from Fungi *sensu stricto*, which are osmotrophic. Evidence of phagotrophic feeding include observations of pseudopod-like digitiform protrusions in both *Rozella* and *Nucleophaga* (Powell, 1984; Corsaro *et al.*, 2014a, 2016). In addition, molecular evidence for phagotrophy in *Rozella* has been put forward (Torruella *et al.*, 2018).

Environmental studies yield a highly diverse array of 18S rRNA gene sequences related with Rozellida from all kind of aquatic environments. This diversity seems so outstanding that might be perhaps compared with the known diversity of Fungi (Lara *et al.*, 2010; Jones *et al.*, 2011a, 2011b; Mohamed & Martiny, 2011; Lazarus & James, 2015; Corsaro *et al.*, 2016; Grossart *et al.*, 2016). However, rather than being similar to *Rozella,* some rozellids might represent intermediate reduction states to extreme parasitism, such as that observed in the microsporidia-like *Paramicrosporidium* and *Nucleophaga*. Accordingly, it would be possible that high evolutionary rates could inflate these taxonomic diversity estimates. Also a diversity overestimation could come from the fact that some of these studies considered as members of Rozellida all organisms

branching between Fungi and nucleariids in phylogenetic trees, due to a clear lack of borders between lineages (Karpov *et al.*, 2014; Naranjo-Ortiz & Gabaldón, 2019a).

Rozella has been historically classified as a clade related to Fungi (e.g. Held, 1981). A position that was later confirmed when *Rozella* was recovered in molecular phylogenies as an unicellular lineage within Holomycota (James *et al.*, 2006a, 2006b). After its genome was sequenced, phylogenomic studies started to place systematically *Rozella* as the sister lineage of Microsporidia, the two forming a clade sister to Fungi (James *et al.*, 2013b; Mikhailov, Simdyanov, & Aleoshin, 2017; Torruella *et al.*, 2018). Later, the intranuclear parasites *Paramicrosporidium* and *Nucleophaga* were shown to branch within the radiation of rozellid environmental sequences (Corsaro *et al.*, 2014b, 2014a, 2016; Bass *et al.*, 2018; Richardson *et al.*, 2019). The genome of *Paramicrosporidium saccamoebae* allowed to carry out phylogenomic analyses in which this rozellid is sister to all Microsporidia, and *Rozella* is sister to *P. saccamoebae* + Microsporidia (Quandt *et al.*, 2017) (Figure 8). These results suggested that rozellids are paraphyletic, a result that has been confirmed by posterior phylogenomic analyses (Torruella *et al.*, 2018). Thus, these microsporidia-like intranuclear parasites of amoebae seam to represent a group of highly modified Rozellomycota, representing intermediate reduction steps towards the Microsporidia (Michel *et al.*, 2000). For more about the relationship between Rozellida and Microsporidia see chapter 1.2.1.1. Microsporidia. Some of the most characteristic genera of Rozellida are:

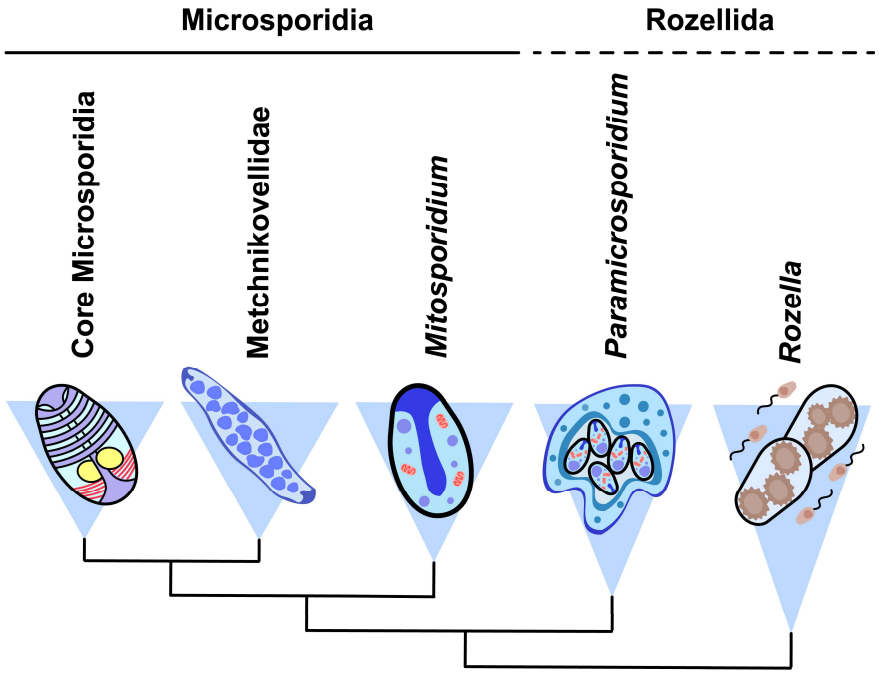


Figure 8. Cladogram showing the relationships between the main lineages of the Rozellida + Microsporidia clade. The relationship shown are based on the latest results from phylogenomic studies (cited in text).

- *Rozella* (Cornu, 1872) is a genus originally erected to unify four described species which had in common the presence of a plasmodial thallus (protoplasm), posteriorly uniflagellate zoospores, and the formation of thick walled spherical resting spores presenting spines in some cases (Figure 9A-C). Two morphologically distinct forms where described within *Rozella*. One species formed multiple sporangia separated from each other by cross walls (polysporangiate), and the other three formed a single sporangium (monosporangiate). However, *Rozella allomycis* which probably is the most widely recognized species, was first described in 1937 (Foust, 1937) and represented the second described polysporangiate species.

Currently the genus *Rozella* is composed of 27 species, 24 monosporangiate and 3 polysporangiate (Letcher & Powell, 2018). After first molecular analyses the relationship of both monosporangiate and polysporangiate was confirmed (James *et al.*, 2006b). However, the importance of the presence of one or multiple sporangia as a taxonomical differentiating trait among *Rozella* species, awaits validation, since environmental studies show a large radiation of *Rozella*-like species with unknown sporangia arrangement (e.g. Lara *et al.*, 2010; Karpov *et al.*, 2014; Grossart *et al.*, 2016; Tedersoo *et al.*, 2017). More molecular data from different *Rozella* species is still needed. The genome of *R. allomycis* was the first rozellid genome sequenced in 2013, and remains the only one obtained of a *Rozella* species (James *et al.*, 2013b). Its genome showed a high level of reduction almost comparable to that of Microsporidia, albeit still possessing a reduced but functional mitochondrion. *Rozella* also possess a Microsporidia specific ATP/ADP transporter originated from a horizontal gene transfer HGT even from bacteria (Heinz *et al.*, 2014)

- *Paramicrosporidium* (Corsaro *et al.*, 2014b) is a genus of non-flagellated intranuclear parasites of amoebae with infective spores having a chitinous cell wall and an anchoring disc with an inactive polar filament (Figure 9D). The organism multiplies as unwalled cells by merogony within the nucleus of the amoebae hosts. *Paramicrosporidium* resembles morphologically to most Microsporidia in traits shared including the presence of a chitin/cellulose cell wall and the absence of a flagellated stage.

However, 18S rRNA gene phylogenies show that some Rozellida members are possibly related to Microsporidia (Corsaro *et al.*, 2014b), such as *Paraphelidium saccamoeba* and Nucleophaga,

named based on their respective amoebal hosts. The genome of *P. saccamoebae* was the second to be sequenced for a rozellid (Quandt *et al.*, 2017). Multi-gene phylogenies confirmed its relationship with Microsporidia and the paraphyly of Rozellida. Unlike Microsporidia, these species retain fully functioning mitochondrion with all elements of the electron transport chain present; including the respiratory Complex 1 that has been lost in *Rozella*. *Paramicrosporidium* do not possess the Microsporidia specific ATP/ADP transporter previously mentioned.

- *Nucleophaga* was described as a non-flagellated intranuclear parasite of amoeba together with another parasite named *Sphaerita* in 1895 (Dangeard, 1895) (Figure 9E). It remained elusive since then, only appearing in scarce reports (Karling, 1972; Anderson, Stewart, & Allen, 1995). A strain of *N. amoebae* (KTq-2) was recovered from a *Thecamoeba* (Corsaro *et al.*, 2014a) followed by a strain of *N. terricolae* (KTt1) (Corsaro *et al.*, 2016). The infectious stage corresponds to a non-flagellated walled spore, which is engulfed by amoebal phagocytosis. Then the parasite invades the nucleus of the host amoeba, feeding from it and developing finger-like extensions at the surface (suggesting phagocytosis), an endogenous unicellular sporangium is then formed to develop a new generation of spores. Both strains present an anchoring disc and atypical polar filament. 18S rRNA gene phylogenies have shown that *Nucleophaga*, branches at an intermediate position between *Rozella* and *Paramicrosporidium*, representing a unique lineage within the Rozellomycota. There is no genomic data available for *Nucleophaga* species.

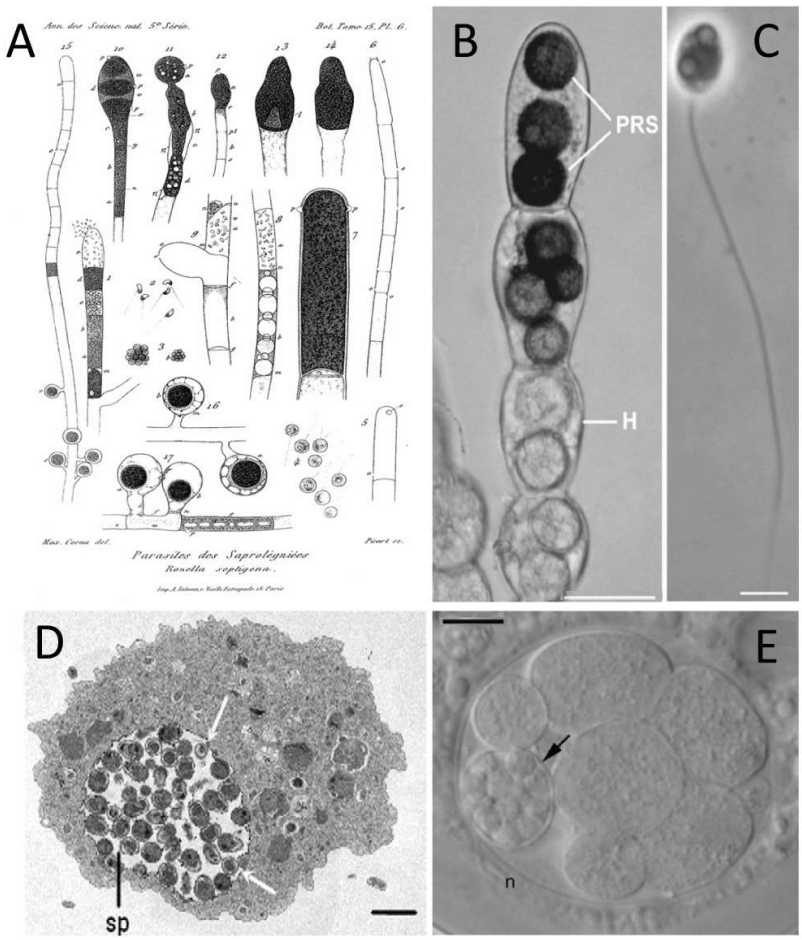


Figure 9. Illustrations and images of Rozellida. A) Illustrations of *Rozella septigena* (Cornu, 1872). B) *R. allomycetis* parasite resting spores (PRS) and its host (H) *Allomyces macrogynus* (Letcher & Powell, 2018). C) Motile zoospore of *R. rhizoclosmatii* (Letcher & Powell, 2018). D) TEM image of *Saccamoeba* with the nucleus filled of *Paramicrosporidium saccamoebae* spores (sp) (Corsaro *et al.*, 2014b). E) *Thecamoeba quadrilineata* nucleus (n) filled with plasmodia of *Nucleophaga amoebae* (Gordetskaya *et al.*, 2019). Scale bars: B = 15 µm, C = 1 µm, D = 2 µm, E = 5 µm.

1.7.2.2. Microsporidia

Microsporidia are a highly diverse and specialized group of intracellular obligate parasites of Metazoan and in some cases, protists, with more than 1300 described species grouped in 187 genera (Vávra & Lukeš, 2013) (Figure 10).

They share a relatively uniform life cycle in which the infective stage is the dormant spore, whose size ranges from 1 to 12 µm (Figure 10A-B). Spores possess the most important synapomorphy of the clade, the polar filament (Figure 10C) (also known as polar tube, injection tube or invasion

tube) (Kudo, 1918; Weidner, 1972; Xu & Weiss, 2005). The polar filament is a coiled organelle which evaginates during infection. It is divided in two regions: an anterior straight end surrounded by a polaroplast (an infection vacuole) and an attachment structure known as the anchoring disc; the posterior end is a coiled region that, depending on the species, may form from four to around 30 coils within the cytoplasm of the spore (Wittner & Weiss, 1999). During spore germination, the polar filament physically pierces the host cell plasma membrane. This is followed by an expansion of the polaroplast vacuole, which forces the content of the spore through the tube and into the host cell forming the first intracellular stage known as meront (Vávra & Larsson, 2014). The meront then divides into more meront cells, which progressively fill the cytoplasm of the host cell. The chitin cell wall material starts to deposit on the spores plasma membrane, helping to form the sporont. The sporont in some species may continue dividing, producing daughter sporonts, but finally, cells develop during sporogony into mature infective spores (Bohne *et al.*, 2000; Hayman *et al.*, 2001; Brosson *et al.*, 2005).

Microsporidia have undergone a process of extreme cellular and genomic reduction. Microsporidians lack all flagellum components and, consequently, do not form zoosporic stages. Since the flagellum is an ancestral character in Holomycota and the flagellum is still present in *Rozella,* this implies that Microsporidia have secondarily loss their flagellum (James *et al.*, 2006a). They also have a reduced Golgi apparatus, and they do not present dictyosomes or peroxisomes (Beznoussenko *et al.*, 2007). However, one of the most widely known simplified eukaryotic features in Microsporidia is the mitochondria (Embley & Martin, 2006). Microsporidia lack canonical mitochondria, they do however, present mitochondrial-derived organelles called mitosomes. Mitosomes were first detected by immunolocalization of an mitochondrial Hsp70 protein derived from an endosymbiont (Williams *et al.*, 2002). The complete genome sequence of Encephalitozoon cuniculi (Katinka *et al.*, 2001) revealed that the mitosome had lost its genome; its main function seems to be the assembly of iron-sulphur clusters (Stairs, Leger, & Roger, 2015). About a dozen sequenced genomes of microsporidian are available. The analysis of the first genome from *E. cuniculi* started to give insights into the particular characteristics of these reduce genomes. The microsporidian cell simplification is also reflected in reduction and compaction at the genomic level. Microsporidia genomes are among the smallest found in any eukaryote, *E. intestinalis* genome has a size of only 2.3 Mbp (Corradi *et al.*, 2010). Genome reduction is accompanied by gene loss of metabolic genes, (due to dependency on the host's metabolic

products), gene length reduction and genome compaction (Vivarès & Méténier, 2000; Corradi & Slamovits, 2011; Heinz *et al.*, 2012).

At the same time, Microsporidia genomes have acquired genes essential to import metabolites from the host cell. These genes include ATP/ADP translocase transporters, nucleoside transporters and other genes involved in metabolism. These genes were acquired by ancestral HGT from prokaryotes (from *Rickettsia* and *Chlamydia* lineages) or in few cases from an animal hosts (Tsaousis *et al.*, 2008; Lee, Weiss, & Heitman, 2009; Cuomo *et al.*, 2012; Pombert *et al.*, 2012).

Microsporidia, have been known and studied for over 150 years; the first of them to be described in 1857 by Nägeli caused a silkworm disease (Nägeli, 1857). They were originally included within the Schizomycete fungi, now known to be a collection of different unrelated eukaryotic and prokaryotic species. They were transferred to the class Sporozoa within Protozoa (Balbiani, 1882), and then to Cnidosporidia within Mixozoa for over 70 years (Döflein, 1901). At one point, they even constituted an independent phylum (Sprague, 1977). In 1983, Cavalier-Smith included Microsporidia in the subkingdom Archezoa (Cavalier-Smith, 1983) that grouped a collection of eukaryotes having in common an apparent absence of mitochondria. Archezoa were supposed to be ancestrally amitochondriate, grouping lineages that predated the mitochondrial acquisition by the ancestor of the rest of eukaryotes. The hypothesis of Archezoa and the existence of truly amitochondriate eukaryotes was later refuted (Embley & Hirt, 1998; Hirt *et al.*, 1999; Philippe *et al.*, 2000).

Microsporidia have always been a group difficult to place in the tree of life. Being structurally simple and different from other protists, molecular phylogenies seemed to be a useful tool to resolve their relationships. However, due to their extremely accelerated rate of sequence evolution (Thomarat, Vivarès, & Gouy, 2004), microsporidian phylogenies suffered from long branch attraction artefacts (Philippe *et al.*, 2000). This was typically the case for the earliest 18S rRNA gene phylogenetic trees where microsporidia were the deepest branching eukaryotes (Leipe *et al.*, 1993; Kamaishi *et al.*, 1996). Subsequent phylogenetic trees using RPB1, α and β -tubulin, and other genes, suggested a fungal affinity for the microsporidia (Hirt *et al.*, 1999; Keeling, 2003) and, more specifically a possible relationship with *Rozella* (James *et al.*, 2006a), later confirmed by more complete phylogenomic studies (Capella-Gutiérrez, Marcet-Houben, & Gabaldón, 2012; James *et al.*, 2013a).

Recently, organisms branching early within Microsporidia have been described and their genomes sequenced. These include *Mitosporidium daphniae* and *Amphiamblys* sp. (Metchnikovellidae) (Figure 10 D-G), two lineages branching at the base of the "core Microsporidia" (canonical long branch Microsporidia) (Haag *et al.*, 2014; Mikhailov *et al.*, 2017) (Figure 8). *M. daphniae* possesses a functional mitochondrion and the metchnikovellids display a short and straight polar filament, lacking also a polaroplast. These phylogenetic and structural particularities prove that these early branching Microsporidia stand in a key position to understand the progressive reductive evolutionary transition between rozellid-like ancestors and Microsporidia. Some of the most characteristic representatives outside core Microsporidia are:

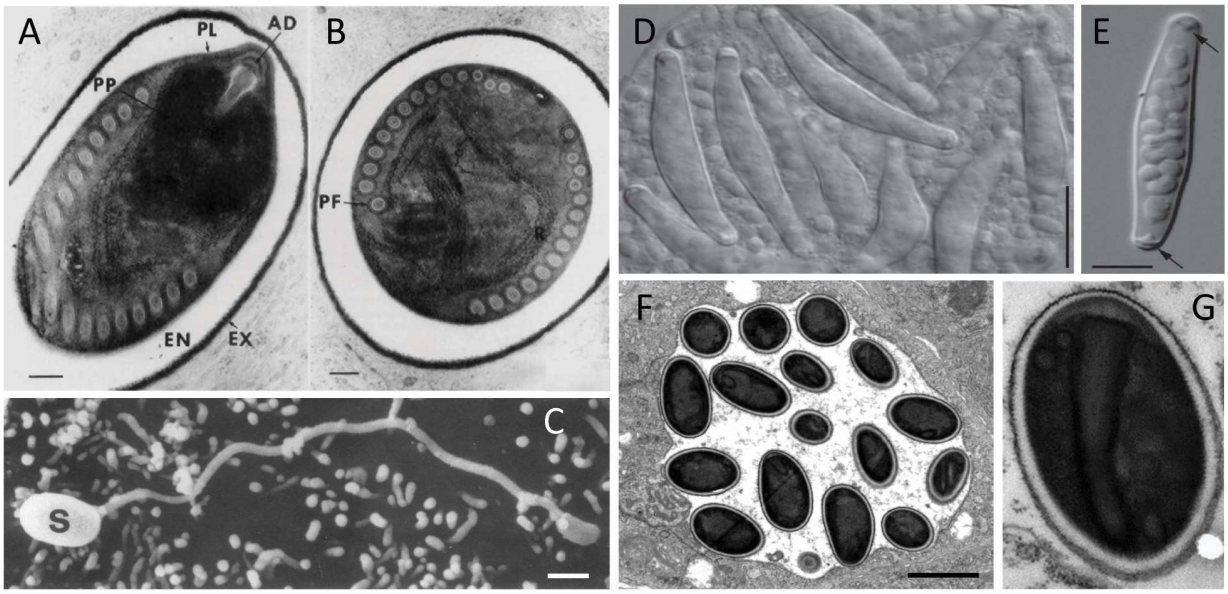


Figure 10. Electron transmission, scanning and light microscopy images of Microsporidia. A-B) Electron micrographs of spores of *Nosema scripta*, a core Microsporidia (Bauer & Pankratz, 1993). C) Scanning electron micrograph of an *Encephalitozoon hellem* spore (s) with an extended polar tube (Visvesvara *et al.*, 1991; Weber *et al.*, 1994). D) *Metchnikovella incurvata* spore sacs in the cytoplasm of the gregarine *Polyrhabdina* sp. under high magnification and E) detail of a spore sac (Sokolova *et al.*, 2013). F) Spores of *Mitosporidium daphniae* and G) in detail (Haag *et al.*, 2014). AD (anchoring disk), PL (plasmalemma), PP (polaroplast), PF (polar filament), EN (endospore), EX (exospore). Scale bars: A-B = 0.2 µm, C = 1 µm, D = 10 µm, E = 5 µm, F close to 2 µm, G close to 1 µm.

- *Mitosporidium daphniae* (Haag *et al.*, 2014) is an intracellular non-flagellated parasite found in the hind gut of *Daphnia magna* crustaceans (Figure 10 F-G). Morphologically, they resemble gut Microsporidia and they were used in other studies, without a formal classification, to negatively affect *Daphnia*'s fitness (Refardt & Ebert, 2012). However, a finer ultrastructural examination

revealed morphological similarities to Microsporidia, including a polar tube. *M. daphniae* genome differs from those of any other known Microsporidia. It encompasses a complete mitochondrial genome with genes coding for ATP production both from glucose, via the citrate cycle, and by oxidative phosphorylation. By contrast, the also mitochondriate *Paramicrosporidium* and *Mitosporidium* do not possess the canonical microsporidian ATP/ADP transporters acquired by HGT. In phylogenomic trees, *Mitosporidium* branches between *Rozella* and all other microsporidia with a considerably shorter branch (Figure 8). In subsequent phylogenomic analyses, *Mitosporidium* branches between *Paramicrosporidium* and the rest of Microsporidia (Quandt *et al.*, 2017; Torruella *et al.*, 2018), supporting the paraphyletic character of Rozellida (or highlighting the need to revise the taxonomy of this group). However, the topology of 18S rRNA phylogenies with larger taxon sampling (although less phylogenetic signal) is not congruent with that of phylogenomic trees for *Paramicrosporidium*, *Mitosporidium* and Microsporidia (Corsaro *et al.*, 2016; Bass *et al.*, 2018; Richardson *et al.*, 2019). Further taxon sampling may help better resolving the relative branching order along the *Rozella*-Microsporidia continuum.

- Metchnikovellidae is a particular group of microsporidian hyperparasites of gregarines that inhabit the intestinal tract of marine annelids (Vivier, 1975) (Figure 10 D-E). The clade only accounts for a few described species (~25 spp), grouped in three genera *Amphiamblys*, *Amphiacantha* and *Metchnikovella*. Metchnikovellids possess a long description history, since they were first described in the 19th century (Caullery & Mesnil, 1897) and all three genera were already proposed by the beginning of the 20th century (Caullery & Mesnil, 1914, 1919).

Metchinikovellidae remained *incertae sedis* for a long time and their phylogenetic affiliation was debated over time. Their morphological and ultrastructural characteristics suggested that metchnikovellids were related to Microsporidia (Sprague, 1977) and, like most Microsporidia, they lack a canonical mitochondria. However, their spores also lack some key microsporidian features, including a coiled polar filament (they possess a short and straight polar filament), the polaroplast and a merogonial proliferation stage (Larsson, 2000; Larsson & Køie, 2006; Sokolova *et al.*, 2013). Accordingly, metchnikovellids have been commonly treated as an early diverging group within the Microsporidia, and their spore ultrastructural features considered as “primitive”. Nonetheless, metchnikovellid spores also hold ultrastructural similarities with those of intranuclear microsporidian-like parasites of amoeba like *Nucleophaga* and the rozellid *Paramicrosporidium* (Corsaro *et al.*, 2014a, 2014b; Rotari, Paskerova, & Sokolova, 2015).

Recently, the genome analysis of the *Amphiamblys* sp. confirmed the microsporidian affinity of metchnikovellids, since *Amphiamblys* branched midway between *M. daphniae* and core Microsporidia in phylogenomic trees (Figure 8). Its metabolic profile resembles more to that of derived core Microsporidia than to rozellids and it does not possess a functional mitochondrion buy a mitosome. Interestingly, similar to *P. saccamoebae* and *M. daphniae*, *Amphiamblys* lacks the specific HGT acquired ATP/ADP transporters from Microsporidia, raising the question of how metchnikovellids obtain ATP from their host. Recent phylogenomic analyses confirmed their position branching after the *P. saccamoebae* and *M. daphniae* radiation (Torruella *et al.*, 2018). Thus, *Amphiamblys* sp. seems to be closer to core Microsporidia in the rozellid-Microsporidia continuum. Nevertheless, the *Amphiamblys* genome is partial, such that sequencing more metchnikovellid genomes is essential to validate the metchnikovellid phylogenetic position and the presence or absence of proteins (e.g. ADP/ATP transporters).

- Chytridiopsida is a microsporidian group that mainly infects terrestrial arthropods (e.g. Purrini & Weiser, 1985; Larsson, Steiner, & Bjørnson, 1997; Radek *et al.*, 2015). Chytridiopsida exhibit several morphological traits similar to metchnikovellids, including a usually short polar filament, the absence of a posterior vacuole and a highly reduced or lacking polaroplast. Like in metchnikovellids, the chytridiopsid life cycle generally lacks a merogony stage and the spores form through endogenous sporogony (Corsaro *et al.*, 2019).

Recently, phylogenies based on SSU and SSU + 5.8S + LSU rRNA genes suggested that *Chytridiopsis typographi* branches earlier than metchnikovellids in the microsporidian branch, forming one of the most basal lineages within the Microsporidia (Corsaro *et al.*, 2019). To further support these results, Corsaro et al. compared the structure of *C. typographi* SSU rRNA and 5.8S/LSU rRNA gene region, which are typically fused in core Microsporidia, with other related members (including rozellids). A progressive reduction of the intergenic transcribed spacer (ITS) is observed along the microsporidial branch (shorter in metchnikovellids than in chytridiopsids), as expected. However, the support values for the chytridopsids as the sister lineage to metchnikovellids + core Microsporidia remains low, and only one species represents this group in molecular phylogenies. Genomic data from several representatives of the clade will be needed to confirm their branching order and if they form a coherent group.

The overall evolutionary relationship between *R. allomycis*, *P. saccamoebae*, *M. daphniae*, metchnikovellids and chytridiopsids remain somehow uncertain an under debate (Corsaro *et al.*,

2016). What it seems clear is that they fall into a continuum from a more *Rozella*-like ancestor towards a more canonical core Microsporidia-like organisms (Haag *et al.*, 2014; Mikhailov *et al.*, 2017; Quandt *et al.*, 2017).

It now seems clear that the array of similarities between rozellids and microsporidians goes beyond morphological, or phylogenetic placement aspects. The presence of specific transporters acquired by HGT, the loss of amino acid biosynthesis routes or the mitochondrial respiratory chain reduction are patchy in both Microsporidia and Rozellida (Heinz *et al.*, 2014; Quandt *et al.*, 2017). This has led some authors to propose a new reclassification in which the term Microsporidia is extended to include all these parasitic lineages (Bass *et al.*, 2018). However, this classification is also based on many 18S rRNA gene environmental sequences. These represent a high diversity, but the phylogenetic signal from a single marker is low such that, to solve these deep level relationships, the description and genomic data of new organisms branching along this continuum (e.g. *Nucleophaga*, chytridiopsids or metchnikovellids) will be required.

1.7.2.3. Aphelida

Aphelida (=Aphelidea) (Gromov, 2000; Karpov *et al.*, 2014) is a diverse group of zoosporic parasites of freshwater and marine algae, characterized by an intracellular amoeboid vegetative stage (Figure 11). There are four described genera differing mostly in the morphology of their zoospores: flagellated (*Aphelidium*, *Pseudoaphelidium*), mostly ameboid (*Amoebaphelidium*) or both (*Paraphelidium*) (Karpov *et al.*, 2017b) (Figure C-D). Their life cycle is very similar in all four genera, and similar to that of chytrids. The only difference is the occurrence of an endobiotic development and a phagotrophic feeding mode (Gromov, 2000; Karpov, Mikhailov, & Mirzaeva, 2013). Their life cycle starts when zoospores attach to the surface of the algal host cell and encyst. This cyst can now act as a resistance form or start the process of infection that involves penetration in the algal cell. In this process, the posterior vacuole develops, pushing the cyst content through an infection tube into the hosts cell (Figure 11B, 11E). The parasite (in reality a parasitoid) becomes a phagotrophic amoeba that develops pseudopods and starts engulfing the host cytoplasm, developing into an endobiotic plasmodium (Figure 11E-F). A multinucleated plasmodium then develops using the host's cell wall as its own sporangium wall before start dividing into several mononucleated cells. These cells develop into zoospores that are released from the empty cell

through the orifice left by the infection tube (Karpov *et al.*, 2014). The aphelid life cycle resembles that of *Rozella* in the presence of an infection tube, phagotrophic feeding, and an unwalled endobiotic sporangium.

The scientific history of aphelids goes back to the 19th century when the genus *Aphelidium* was first described in 1885 by Zopf (Zopf, 1885a). *Amoeboaphelidium* was described in 1925 (Scherffel, 1925) and included in the Monadinea group. During the 1950s and 60s, aphelids were included in the order Proteomyxida within Rhizopoda (Hall, 1953; Honigberg *et al.*, 1964). After a few publications on aphelids in the following decades (e.g. Schnepf *et al.*, 1971) Gromov in 2000 created the class Aphelidia within Rhizopoda (Gromov, 2000). However, two years earlier Cavalier-Smith suggested that the genus *Aphelidium* belonged to Opisthokonta, due to their uniflagellated zoospores and flat mitochondrial cristae (Cavalier-Smith, 1998).

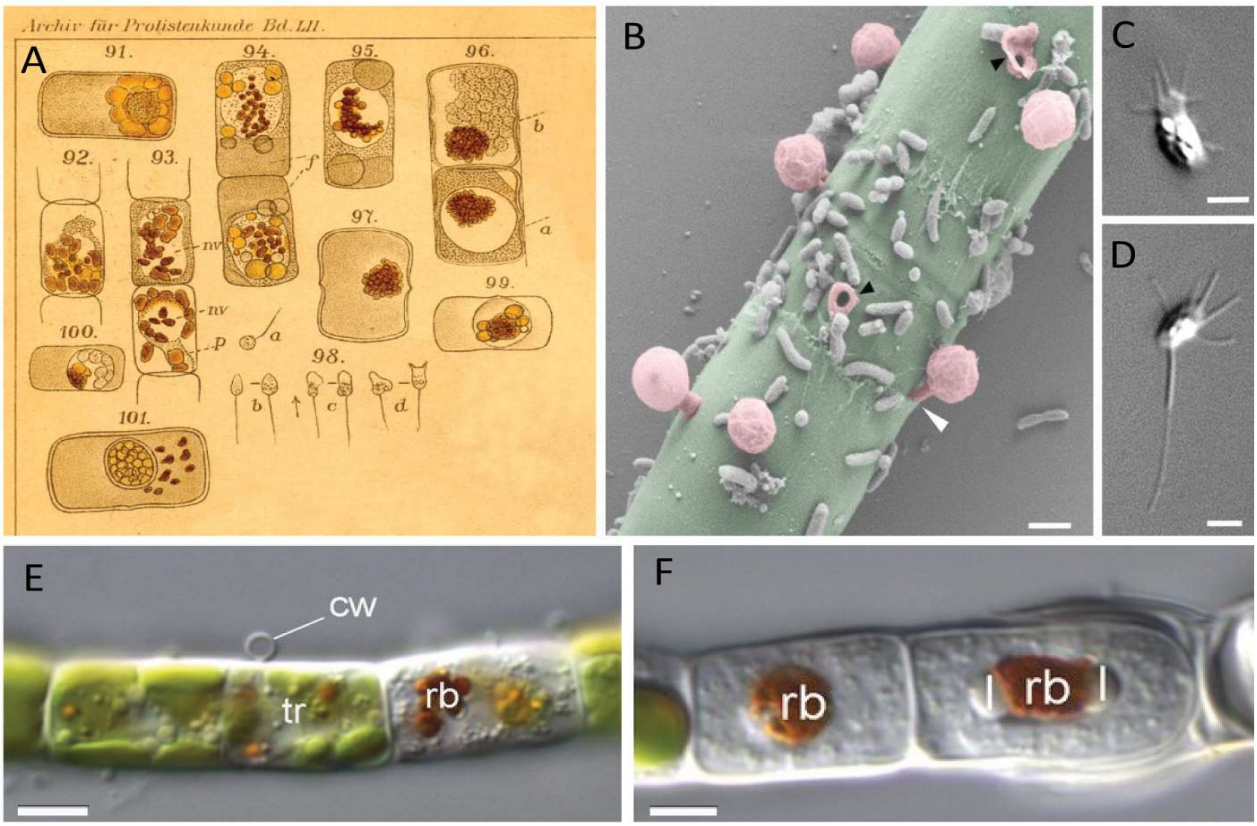


Figure 11. Illustrations and electron scanning and light microscopy images of Aphelida. A) Illustration of *Aphelidium melosirae* in diatoms (Scherffel, 1925). B) False-colored scanning-electron microscopy image of a filament infected by several *Paraphelidium tribonemae* cysts in C-D) detail of ameboflagellated spores of *P. tribonemae* (Torruella *et al.*, 2018). E-F) *Tribonema* filament after recent injection of cyst contents of *P. letcheri* into the host and developing plasmodia (Karpov *et al.*, 2017c). cw (cyst wall), tr (trophont), rb (residual bodies), l (lipid globules). Scale bars: B-D = 1 µm, E-F = 5 µm.

The first 18S rRNA gene phylogenies of *Aphelidium* confirmed the inclusion of aphelids within opisthokonts, but associated them with different lineages within Holozoa, as members of Choanozoa and Ichtyosporea (Pinevich *et al.*, 1997; Adl *et al.*, 2005; Shalchian-tabrizi *et al.*, 2008). Due to the previously mentioned morphological similarities and with *Rozella* and the previously obtained results with 18S rRNA genes, aphelids were grouped with them in the class Rozellidea, within Choanozoa (Cavalier-smith, 2013).

Aphelid classification within Holozoa was shown to be an artefact due to low phylogenetic signal. Posterior studies used a combination of RNA polymerase and rRNA gene trees for *Amoeboaphelidium* (Karpov *et al.*, 2013; Letcher *et al.*, 2013) and, later, *Paraphelidium* (Karpov *et al.*, 2016, 2017c). These studies suggested that the phagotrophic aphelids and rozellids relate to the osmotrophic Microsporidia, and were proposed to form a monophyletic clade of endobiotic organisms sister to fungi called ARM and later named Opisthosporidia (Karpov *et al.*, 2013, 2014). These sequences together with diverse environmental sequences seemed to support this classification (James *et al.*, 2013b; Karpov *et al.*, 2013, 2014; Letcher *et al.*, 2013).

However, these genes provide limited phylogenetic signal. This together with the inclusion of fast-evolving sequences from microsporidians led to incongruent tree topologies. These studies recovered both rozellids (Karpov *et al.*, 2013; Letcher *et al.*, 2013) or aphelids (Mikhailov *et al.*, 2017; Bass *et al.*, 2018; Richardson *et al.*, 2019) as the sister lineage to the rest of the Opisthosporidia, and Opisthosporidia was always recovered with moderate support.

Recently, the sequencing of the transcriptome of *Paraphelidium tribonemae* turned the table again. Phylogenomic analysis of this aphelid within the Holomycota suggested the paraphyly of Opisthosporidia, and pinpointed aphelids as the sister lineage to all Fungi (Torruella *et al.*, 2018) (Figure 5). Furthermore, by studying the phagotrophic molecular profile (similar to *Rozella*) of *P. tribonemae*, this study showed that the aphelid-like ancestors of Fungi were phagotrophic. This position has also been recovered recently by a study using 18S + 28S rRNA gene phylogenies (Tedersoo *et al.*, 2018). Nevertheless, more genomic data from other aphelid genera needs to be obtained to further confirm the branching order of aphelids as sister to Fungi.

1.7.3. Fungi

Fungi (Moore, 1980) is a highly diverse eukaryotic clade playing key roles on Earth's nutrient (e.g. nitrogen, phosphorus) and carbon cycling. Fungi can be found in practically all terrestrial environments, from the stratosphere (Wainwright *et al.*, 2003), to glaciers (Freeman *et al.*, 2009), deserts (Gonçalves *et al.*, 2016) and the bottom of the sea (Nagahama *et al.*, 2011). In some of these systems they are present simply as dispersal forms (resistance spores), but fungi are also one of the eukaryotic groups that copes the best with extreme conditions, including desiccation, high salt or metal content, low pH or relatively high temperature (Kubicek & Druzhinina, 2007). Fungi play a variety of roles in ecosystems, they are mostly saprotrophs, degrading organic matter in soils or sediments but can also be, pathogens and symbionts of a wide variety of organisms (Sanders, 2002; Stajich *et al.*, 2009). Their symbioses with plants are of special importance. It has been shown that the terrestrialization and radiation of both fungi and green algae occurred at least twice, possibly through an endomycorrhizal and possible endophytic symbioses (Lutzoni *et al.*, 2018). Thus, plant:fungus ratios have been used to estimate fungal diversity. All estimates of fungal species number tend include rozellids and microsporidia and would comprise about 1.5 million species (Hawksworth, 1991). However, since 1700's the number of described fungal species has not stop growing. New estimates based on environmental molecular studies and plant:fungus ratios, suggest that there are 2.2 to 3.8 million fungal species (Hawksworth & Lücking, 2017). From this, only 120,000 fungal species are officially described, implying only a 3-8% of the estimated fungal diversity. This relatively small numbers are partially due to the fact that fungi have been traditionally studied through their morphology and cytology (lichens and mushrooms), and using culture based studies (Lawrey & Diederich, 2003; Stajich *et al.*, 2009; de Mattos-Shipley *et al.*, 2016). To unveil the fraction of organism that cannot be cultured using traditional methods (known as "microbial dark matter"), molecular methods, including single-cell omics, need to be used (see chapter "Single cell genomics applied to the fungal dark mater").

The current diversity of Fungi can have been grouped in 6 major lineages: The zoosporic unicellular Chytridiomycota and Blastocladiomycota, the filamentous Zoopagomycota and Mucoromycota (previously within the paraphyletic Zygomycota) and the complex multicellular Dikarya fungi Ascomycota and Basidiomycota (Figure 12) (Spatafora *et al.*, 2017; Naranjo-Ortiz & Gabaldón, 2019a). The phylogenetic position of many fungal clades remains under discussion,

with several competing hypothesis for several fungal nodes (Ebersberger *et al.*, 2012). One of these nodes is the one that defines branching order between the zoosporic Fungi Chytridiomycota and Blastocladiomycota. This node remains unresolved largely due to low phylogenetic signal (see chapter 1.7.3.3.Phylogenetic relationship of Chytridiomycota and Blastocladiomycota) making it essential the generation of genomic data from new representatives of these two clades (Chang *et al.*, 2015). The first clade to diverge after them are the non-flagellated Zoopagomycota, marking a large flagellum loss event that paralleled fungi evolution in terrestrial environments (Lutzoni *et al.*, 2018). The branching order continues with the (mostly) endomycorrhizal fungi from the Mucoromycota, which putatively include Glomeromycota. Lastly, Dikarya comprise the best-studied groups of Fungi: Ascomycota and Basidiomycota (Figure 12).

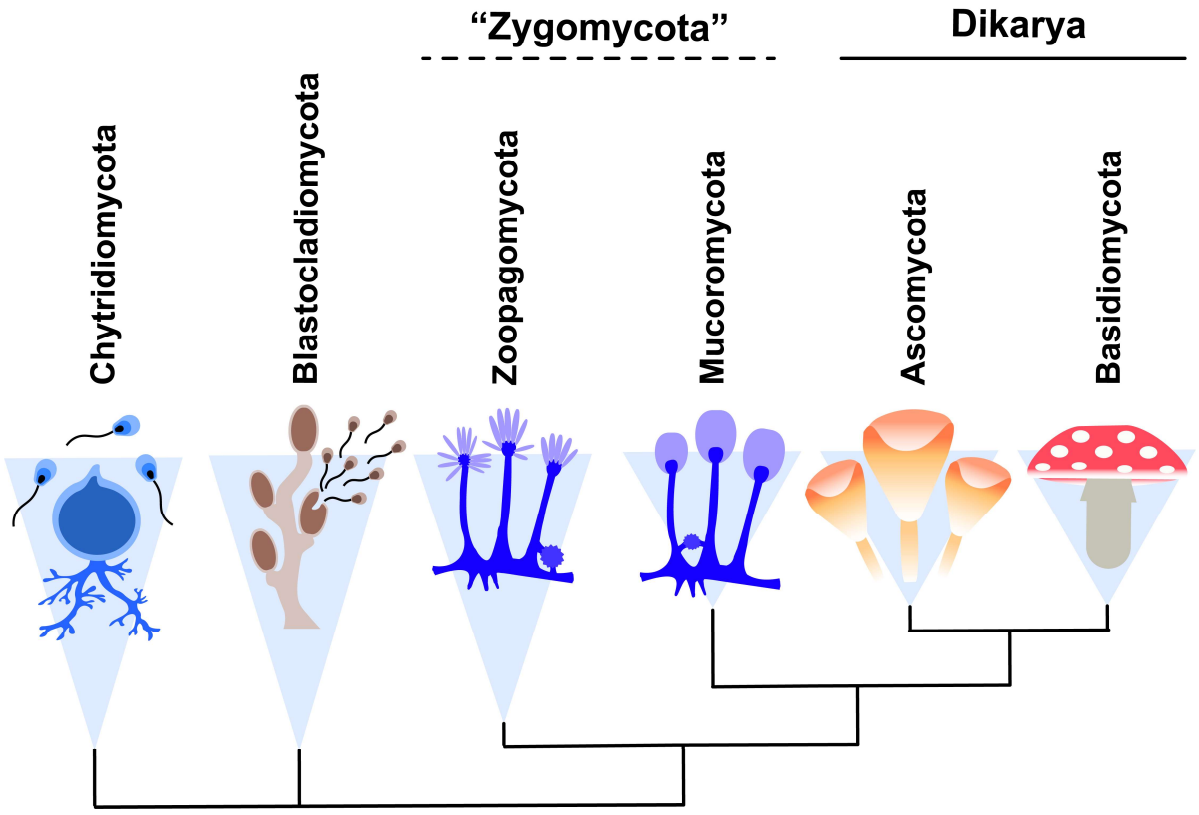


Figure 12. Cladogram showing the relationships between the main lineages of Fungi. The relationship shown are based on the latest results from phylogenomic studies (cited in text).

1.7.3.1. Chytridiomycota

Chytridiomycota (Doweld, 2001) is a clade of coenocytic zoosporic fungi, pathogens of plants, animals, and several groups of algae, as well as an important group of saprotrophs of organic matter including pollen and cellulose (Naranjo-Ortiz & Gabaldón, 2019a) (Figure 13). Chytrids seem to play a key role in nutrient recycling in aquatic environments; the so called 'mycoloop' (Kagami, Miki, & Takimoto, 2014). Chytrids are ubiquitous, occurring in almost all habitats from the tropics to the arctic regions (Powell, 1993; Freeman *et al.*, 2009). Over 1,000 species of chytrids have been described based on classical taxonomy and morphology, but recent environmental molecular and ultrastructural studies indicate the richness of the group is even greater than previously thought (Letcher *et al.*, 2005; Freeman *et al.*, 2009; Seto, Kagami, & Degawa, 2017).

The most characteristic trait of chytrids (and Blastocladiomycota) within Fungi, is their zoospores, which carry a single posterior whiplash flagellum. This flagellum moves using energy from lipid and glycogen reservoirs (Figure 13F). The overall chytrid life cycle starts when the zoospores find a suitable substrate by chemotaxis or phototaxis. Then the zoospores retract their flagella, produce a wall and develop into coenocytic thalli that will eventually generate a sporangia. The sporangia are highly variable among chytrid clades. They can form a monocentric (one sporangium) sporangium or in polycentric (many sporangia) hyphae-like structure bearing numerous sporangia (e.g. Monoblepharidomycetes). The sporangium develops root-like structures called rhizoids involved in the osmotrophic uptake of nutrients from the food source (Figure 13A-E). Lastly, the coenocytic sporangia produce numerous zoospores which are then released through an operculum (Powell, 2017a). Sexual reproduction has been described in a few representatives including *Chytriomyces* and *Zygorhizidium* (Miller & Dylewski, 1981; Doggett & Porter, 1996).

Chytrids were first related to Fungi with the description of *Chytridium olla* by Braun in 1851 (Braun, 1851), and followed by the description of *Monoblepharis* by Cornu (Maxime Cornu, 1871). For decades chytrids were regarded as Phycomycetes (Fitzpatrick, 1930), a junk drawer of heterotrophic fungal and fungal-like organisms with both coenocytic thalli and spores formed within sporangia. Sparrow (Sparrow, 1960) provided an early systematic history of chytrids. Later on, Bartnicki-Garcia (1970) based on biochemical characteristics classified the chytrids as 'true'

Fungi (Bartnicki-Garcia, 1970). Others also started considering chytrids as an intermediate lineage between protists and Fungi due to the production of motile zoospores (Barr, 1990).

The first 18S rRNA gene phylogenies confirmed the position of Chytridiomycota as the earliest divergent branch within Fungi (Forster *et al.*, 1990; Bruns *et al.*, 1992). The discovery of *Batrachochytrium dendrobatidis* strongly affected chytrid studies.

B. dendrobatidis (Longcore, Pessier, & Nichols, 1999) infects over 700 species across the three amphibian orders, causing species extinctions and mass mortality events with large population declines (Vredenburg *et al.*, 2010; Olson *et al.*, 2013). All the array of studies on this chytrid focused on finding a solution to the amphibian crisis and led to the sequencing of its genome in 2009. This represented the first chytrid genome ever sequenced and, since then, around 26 chytrid genomes have been made available (https://genomeportal.jgi.doe.gov/portal/). This genomic data has allowed a better understanding of the relationship of chytrids within the clade and its relationship with Fungi in a multigene-phylogenetic framework (Chang *et al.*, 2015; McCarthy & Fitzpatrick, 2017; Ahrendt *et al.*, 2018; Torruella *et al.*, 2018) (Figure 12).

Chytridiomycota includes three large groups with uncertain phylogenetic relationship between them (Sekimoto *et al.*, 2011; Chang *et al.*, 2015): Chytridiomycetes, Monoblepharidomycetes and Neocallimastigomycota. Now I will review these three main groups within Chytridiomycota:

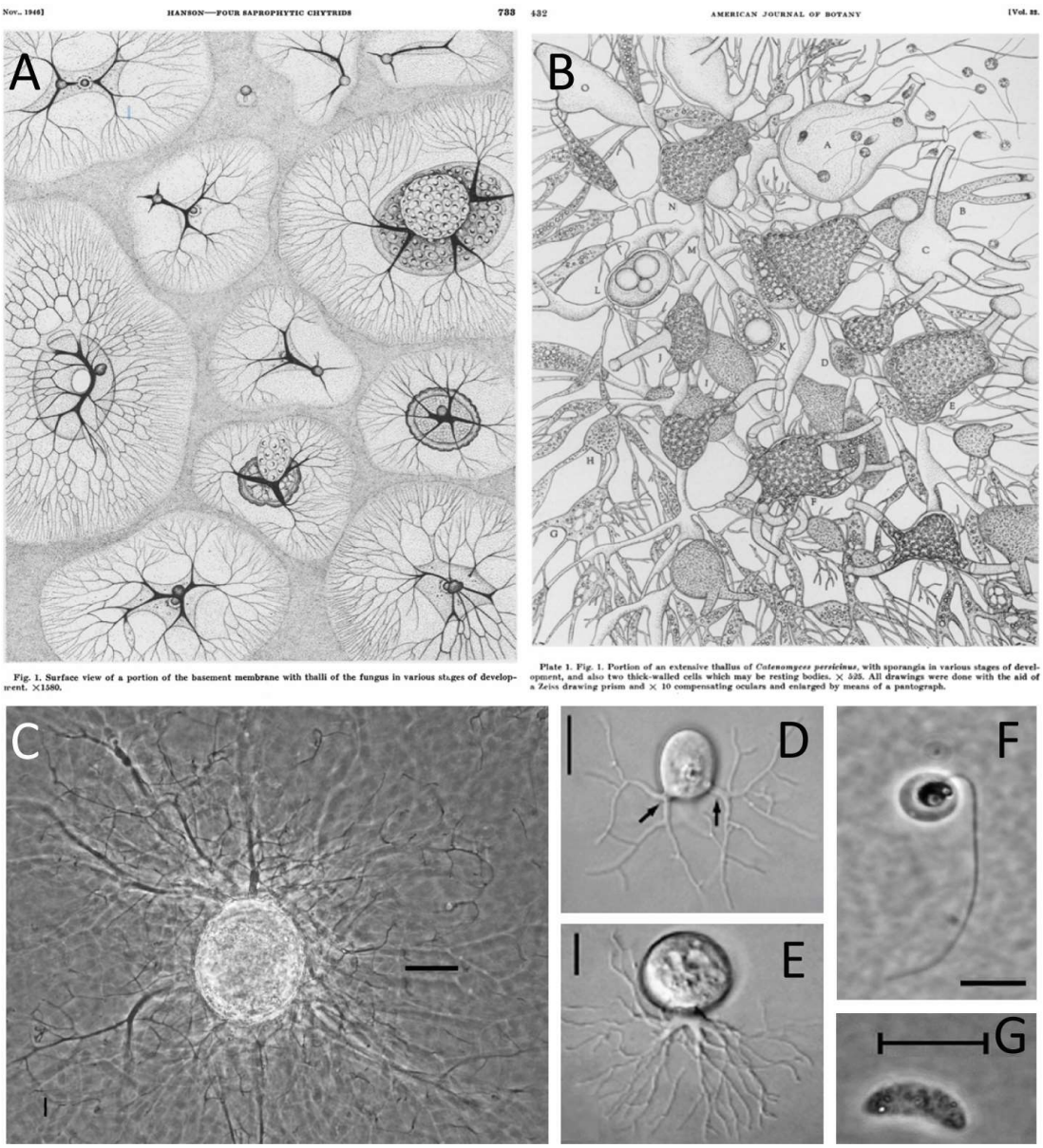


Figure 13. Illustrations and light microscopy images of Chytridiomycota. A-B) Illustrations showing saprophytic chytrids made by Anne Marie Hanson (Hanson, 1946). C) Developmental morphology of *Neokarlingia chitinophila* (Longcore & Simmons, 2012). D) Developmental thallus with two rhizoidal axes of *Aquamyces* sp. E) Developmental thallus of *Rhizophydium* sp. (Letcher *et al.*, 2008). F) Spherical zoospore of *Karlingiomyces asterocystis* with small lipid body. G) *Hyaloraphidium curvatum* (https://jgi.doe.gov/). Scale bars: C = 20 µm, D-F = 10 µm, G = 20 µm.

- Chytridiomycetes form the largest clade of chytrids comprising the majority of the chytrid 1000 described species (Naranjo-Ortiz & Gabaldón, 2019a). Chytridiomycetes are widespread in aquatic and soil environments developing parasitic (plants, animals and protozoa) or saprobiotic lifestyles (Spatafora *et al.*, 2017). Chytridiomycetes are classified in around 90 genera (James *et*

al., 2006a; Hibbett *et al.*, 2007) and 10 described orders: Chytridiales, Spizellomycetales, Cladochytriales, Rhizophydiales, Polychytriales Rhizophlyctidales, Lobulomycetales, Synchytriales, Gromochytriales and Mesochytriales. The relationships of the different orders within Chytridiomycetes remains unresolved (Misra, Tewari, & Deshmukh, 2012; Powell & Letcher, 2014). This lineage includes the already mentioned *B. dendrobatidis* from which the clade gained attention.

- Monoblepharidomycetes is a group of freshwater saprotrophic chytrids with a developed coenocytic mycelial thallus. It groups about 30 species under 6 genera, including *Gonapodya* from which there is genomic data available (Chang *et al.*, 2015; Spatafora *et al.*, 2017). Their thalli form hyphae-like structures with centrioles and no Spitzenkörper, suggesting an independent origin from the hyphae observed in non-flagellated fungi (Sekimoto *et al.*, 2011; Dee *et al.*, 2015). Another key trait of Monoblepharidomycetes is the presence of a sexual reproduction method unique in fungi involving flagellated male gametangia and a non-flagellated female gametangium, called oogonic sexual cycle. One member of the clade worth highlighting is *Hyaloraphidium curvatum* (Figure 13G), previously classified as a colourless algae (Ustinova, Krienitz, & Huss, 2000; Forget *et al.*, 2002). Phylogenetic studies showed that *H. curvatum* grouped with the Monoblepharidomycetes (Sekimoto *et al.*, 2011; Tedersoo *et al.*, 2018). Its particularity is that has an holocarpic thallus (all the thallus is the sporangium) and its sporangia produce non-flagellated spores suggesting an independent loss of the flagellum in this lineage (James *et al.*, 2006a).

- Neocallimastigomycota (Neocallimastigomycetes) is a clade of anaerobic fungi from the rumen and hindgut of mammal and reptile herbivores, where they produce cellulases and xylanases helping to digest plant cell walls (Ho & Barr, 1995; Tsai, Qiu, & Liu, 2003; Huang, Huang, & Hseu, 2005; Gruninger *et al.*, 2014). There are around 20 described species grouped in 6 genera. Neocallimastigomycota develop a monocentric or polycentric thalli with extensive rhizoids penetrating the fibrous plant material (Ho & Barr, 1995). They present single or multiple flagellated, unwalled zoospores which vary in size, even within the same strain (Powell, 2017a). Genomes from this group are large and characterized by low GC content (Youssef *et al.*, 2013). The phylogenetic position of Neocallimastigomycota remains unresolved including its position as an independent phylum or a class within Chytridiomycota (James *et al.*, 2006a, 2006b; Sekimoto *et al.*, 2011; Ebersberger *et al.*, 2012).

1.7.3.2. Blastocladiomycota

Blastocladiomycota (James *et al.*, 2006b) is a group of coenocytic zoosporic fungi with unique morphological and life history traits (Figure 14). They can be found both in aquatic and soil environments and include genera traditionally classed as water molds. Blastocladiomycota members can be saprotrophs on decaying organic matter or obligated parasites of invertebrates, plants, and algae (James, Porter, & Martin, 2014). Blastocladiomycota present a wide range of growth morphologies, from monocentric with limited development to polycentric with highly developed coenocytic hyphae-like structures (Spatafora *et al.*, 2017).

Some of the characteristics that differentiate Blastocladiomycota from chytrids is the presence of a resting sporangium, which is usually thick-walled and with pits and spines ornamentation. This sporangium germinates by cracking and cleavage of the external wall, releasing the zoospores. Blastocladiomycota also differ chytrids in their meiosis which is typically sporic, while in chytrids is zygotic (Lange & Olson, 1980; Olson, 1984). Another morphological characteristic of Blastocladiomycota is the presence of a prominent nuclear cap in zoospores (Couch & Whiffen, 1942) (Figure 14E), which may lead to bipolar germination in species developing hyphae-like structures (e.g. *Allomyces*) (Figure 14A-B).

Blastocladiomycota zoospores also present prominent lipid bodies called side bodies or microbody lipid globule complexes (MLCs). This structure consists of an ordered arrangement of different microbodies, lipid globules and one or more mitochondria, which locate in one side along the axially arranged nucleus with its prominent nuclear cap; all is englobed by an extra membrane (Lovett, 1975; Powell, 1978; Lange, 1979). Distinct types are distinguished depending on the number, localization a size of the mitochondria associated with them.

Zoospore locomotion in Blastocladiomycota is driven by both chemotaxis and phototaxis. Chemotaxis has been observed in zoospores of *Allomyces* attracted towards cellulose, chitin and certain amino acids (Machlis, 1969; Stumm *et al.*, 1976; Mitchell & Deacon, 1986). On the other hand positive phototaxis, thus, attraction towards light, has been proven in *Blastocladiella* (Avelar *et al.*, 2014), *Allomyces* (Robertson, 1972; Olson, 1984) and *Coelomomyces* (Martin, 1969). This type of phototaxis may provide a mechanism for zoospores to move from dark sediments towards more illuminated areas, richer in food sources.

Recently, *B. emersonii* has been found to possess a unique novel gene derived from the fusion of a bacterial-derived rhodopsin domain, and a guanylyl cyclase domain (BeGC1). The protein controls zoospore phototaxis in response to levels of cGMP. The response depends on a BeCNG1 cyclic nucleotide-gated channel. The BeGC1 fusion protein was localized by immunofluorescence to the external membrane of the side bodies and functions as an eyespot, at the base of the flagellum and controlling its beating (Avelar *et al.*, 2014, 2015; Richards & Gomes, 2015). The BeGC1 fusion and the channel BeCNG1 proteins were also found in other Blastocladiomycota fungi such as *Allomyces macrogynus* and *Catenaria anguillulae*. These findings seem to confirm that this light sensing pathway is a synapomorphy of Blastocladiomycota.

Another unique trait of Blastocladiomycota is related with their life cycle, since they are the only known fungal group to present alternation of haploid (gametophyte) and diploid (sporophyte) generations.

There is not a unifying life cycle for Blastocladiomycota, since variations are as common as members of the clade (James *et al.*, 2014). However, the life cycle of some species of *Allomyces* (Hatch, 1935; Emerson, 1941) is often taken as canonical. In *Allomyces* a haploid thallus produces male and female gametangia, which usually differ in color and size. Both gametangia produce motile planogametes that fuse to create a zygote that germinates into a diploid thallus. Diploid zoospores produced by the diploid thallus form other diploid thalli. Resting resistant sporangia eventually form. These sporangia are the meiosis site, leading to the formation of haploid zoospores, which can lead to haploid thalli formation and start the cycle again.

Blastocladiomycota were first described within the “chytrid” genus *Physoderma* in 1833 (Wallroth, 1833). The first classification including Blastocladiales as a separate order was made by Petersen in 1909 (Petersen, 1909). Blastocladiomycota belonged within the Chytridiomycota for a long time since they have posteriorly uniflagellated zoospores. However, the first molecular studies including members of this group already showed a great phylogenetic distance between Blastocladiomycota and chytrids (Bruns *et al.*, 1992; Nagahama *et al.*, 1995; James *et al.*, 2000). Eventually, molecular phylogenies proved chytrid paraphyly and Blastocladiomycota were promoted as an independent phylum (James *et al.*, 2006b). The phylum Blastocladiomycota contains only the order Blastocladiales (Petersen, 1909), which is divided in five families (Barr, 2001; Porter *et al.*, 2011). The position of Blastocladiomycota as an independent clade from

chytrids was later confirmed by phylogenomic studies (Chang *et al.*, 2015; McCarthy & Fitzpatrick, 2017; Ahrendt *et al.*, 2018; Torruella *et al.*, 2018) (Figure 12).

Environmental metabarcoding studies have recovered sequences for Blastocladiomycota from freshwater environments, and showed some affinity for oxygen-depleted environments (Lefèvre *et al.*, 2007; James *et al.*, 2014; Tedersoo *et al.*, 2017). No species of Blastocladiomycota from marine environments are known (James *et al.*, 2014; Berbee, James, & Strullu-Derrien, 2017; Powell, 2017b). Some of the most characteristic genera of Blastocladiomycota are:

- *Allomyces*, exhibits polycentric development (multiple sporangia). It represents a model of study for the clade because it grows easily in synthetic media. *Allomyces* is the only known genus that displays 'true' hyphae, which form a mycelium with incomplete septa having central and lateral perforations (Spatafora *et al.*, 2017). It can also form structures similar to the Spitzenkörper typical non-flagellated fungi (Vargas, Aronson, & Roberson, 1993). However, since *Allomyces* is phylogenetically distant and the variability of these structures in fungi, they are most likely to have evolved independently (Richards, Leonard, & Wideman, 2017a).
- *Blastocladiella* is a genus similar to *Allomyces* with the difference of a monocentric growth, developing only one sporangium (James *et al.*, 2014).
- *Catenaria* is a polyphyletic clade (Porter *et al.*, 2011) of parasitic Blastocladiomycota known mainly as pathogens of nematodes and small flies (Martin, 1987), although they can also parasitize other Blastocladiomycota (Couch, 1945). They are characterized by a rhizomycelium with or sporangia separated by narrow links.
- *Coelomomyces* is another genus of obligate parasites, requiring two aquatic hosts during different stages of their lifecycle, including mosquito larvae and copepods (Kerwin & Petersen, 1997). *Coelomomyces* is unique among all fungi because they lack cell walls in hyphal bodies and gametangial stages (Whisler, Zebold, & Shemanchuk, 1975).
- *Paraphysoderma* is the only known parasite of the algae with industrial interest *Haematococcus pluvialis* (Boussiba, 2000) (Figure 14C-D). It clusters with *Physoderma* as sister group to all other Blastocladiomycota (Porter *et al.*, 2011). For a long time, it was distinctively characterized for producing non-flagellated aplanospores instead of canonical zoospores (Letcher *et al.*, 2016). However, recent examinations detected different types of zoospores including a fast-swimming, uniflagellated zoospores, which rapidly transform into infectious amoeboid aplanospores (Strittmatter *et al.*, 2016; Asatryan, Boussiba, & Zarka, 2019).

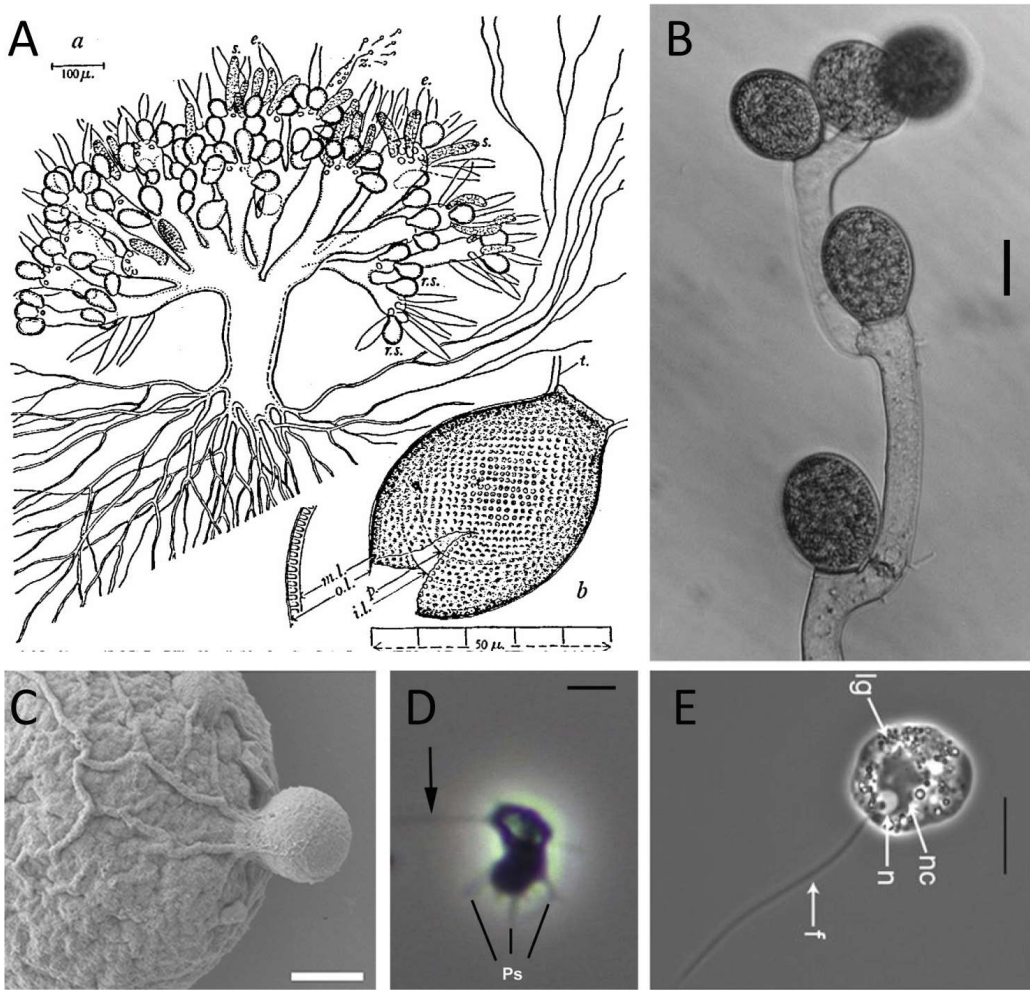


Figure 14. Illustrations and scanning electron and light microscopy images of Blastocladiomycota. A) Illustration of *Blastocladiella pringsheimii* by Elizabeth Blackwell (Blackwell, 1940). B) *Allomyces anomalus* thallus (Wijayawardene *et al.*, 2018). C) SEM image of *Paraphysoderma sedebokerense* propagule encysted on *H. pluvialis* cells (Asatryan *et al.*, 2019). D) An amoeboid swarmer zoospore of *P. sedebokerense* with possible pseudocilium (arrow) (Letcher *et al.*, 2016). E) Uniflagellate zoospore of *Allomyces javanicus* (James *et al.*, 2014). Ps (pseudopodium), f (flagellum), n (nucleus), nc (nuclear cap), lg (lipid globules). Scale bars: B around 30 µm, C = 2 µm, D = 1.25 µm, E = 10 µm.

1.7.3.3. Phylogenetic relationship of Chytridiomycota and Blastocladiomycota

Both Blastocladiomycota (Chang *et al.*, 2015; McCarthy & Fitzpatrick, 2017; Ahrendt *et al.*, 2018; Torruella *et al.*, 2018) and Chytridiomycota (Sekimoto *et al.*, 2011; Letcher *et al.*, 2013; Torruella *et al.*, 2015; Spatafora *et al.*, 2016b; Mikhailov *et al.*, 2017; Tedersoo *et al.*, 2018) have been recovered as the sister lineage to all other fungi in phylogenetic and phylogenomic studies. Thus,

the overall position of Blastocladiomycota and chytrids within Fungi remains uncertain (Figure 12).

These contradictory results are most likely the consequence of insufficient phylogenetic signal to resolve these fungal nodes (Chang *et al.*, 2015). The age of these splits, between 1 billion and 400 million years old (Taylor, Hass, & Remy, 1992; Strullu-Derrien *et al.*, 2018b; Loron *et al.*, 2019), and the accompanying erosion of phylogenetic signal with time, might partially explain the low observed phylogenetic signal. However, this might also result from the radiation underwent during early fungi evolution, which would have prevented the accumulation of the required amount of substitutions to have phylogenetic resolution at these levels (Chang *et al.*, 2015).

To resolve phylogenetic relationships among zoosporic fungi, two approaches will be needed: 1) improving taxon sampling for early diverging taxa around the Blastocladiomycota-Chytridiomycota split (e.g. from genomes of sanchytrids, *Olpidium*, Nephridiophagidae, chytrid-like-clade-1) and 2) improving methods of phylogenetic reconstruction associated with genome and gene content and composition (Spatafora *et al.*, 2017).

1.7.3.4. Zygomycota

Zygomycota is a paraphyletic phylum of fungi with the common trait of presenting a sexual structure called the zygospore. Molecular phylogenies have rejected their monophyly (Spatafora *et al.*, 2016b) dividing them in two independent phyla, Zoopagomycota and Mucoromycota (maybe three including Glomeromycota; Figure 12; Figure 15A-H). Thus, the zygospore was present in the common ancestor of Zoopagomycota, Mucoromycota and Dikarya. They also shared the loss of the flagellum, which was one of the steps leading to the evolution of filamentous terrestrial forms that we observed in the apical part of the fungal tree (Naranjo-Ortiz & Gabaldón, 2019b).

1.7.3.5. Zoopagomycota

Zoopagomycota is a group of filamentous fungi composed of parasites and commensals of opisthokonts (Figure 15A-D). This clade is the sister group to all other non-flagellated fungi, and

its position its key to understand the terrestrialization of Fungi (Spatafora *et al.*, 2017; Naranjo-Ortiz & Gabaldón, 2019b).

Fungal flagellum loss and terrestrialization seems to have preceded the evolution of embryophytes and was likely associated with two independent origins of terrestrial green algae. These clades seem to have co-evolved, facilitating the colonization of land by plants thorough the development of mycorrhizae (Lutzoni *et al.*, 2018).

The flagellar loss also led to one of the main characteristics of filamentous fungi, the lack of centrioles and its substitution by the spindle pole body. The spindle pole body is a functional equivalent of centrioles and the place of attachment during chromosome segregation. There is evidence that Zoopagomycota retain a reduced 9+2 microtubular system in pole bodies. This suggests a possible origin of the poles bodies from centrioles (McLaughlin *et al.*, 2015).

After the paraphyly of Zygomycota was confirmed (Spatafora *et al.*, 2016b; Davis *et al.*, 2019a), Zoopagomycota was established as a phylum containing 3 classes: Zoopagomycotina and Kickxellomycotina, which are sister groups (White *et al.*, 2006; Hibbett *et al.*, 2007), and Entomophthoromycotina (Humber, 2012).

Zoopagomycotina is a class containing predators and parasites of nematodes and their eggs, amoebae, micro-invertebrates and other fungi, divided in five families and 20 genera (Hibbett *et al.*, 2007; Degawa, 2014; Benny *et al.*, 2016b). Some genera produce septa. The evolutionary relationships within the clade are not resolved due to poor sampling (many are obligated symbionts and obtaining cultures is not an easy task) (Davis *et al.*, 2019b).

Kickxellomycotina is a class created to unify several Zoopagomycota representatives which possess regularly compartmented hyphae separated by bifurcated septa that are blocked by a unique lenticular plug (Tanabe *et al.*, 2004; Hibbett *et al.*, 2007; Benny, Humber, & Voigt, 2014)(Figure 15D). They encompass four orders and are associated with the digestive tract of arthropods; some act as parasites of other fungi, including Mucoromycota (Benjamin, 1965; Valle & Cafaro, 2008).

Entomophthoromycotina represent a group of metazoan-associated fungi, either as pathogens of insects or as commensals that have been isolated from animal dung. They comprise three classes and three orders. It is worth highlighting the Basidiobolomycetes, with its class Basidiobolales (Hibbett *et al.*, 2007; Humber, 2012; Benny *et al.*, 2014) (Figure 15C). The placement of Basidiobolales remains problematic due to its long branches (Gryganskyi *et al.*, 2012). Members

from this clade, including *Basidiobolus* and *Conidiobolus,* are unique among other Zoopagomycota fungi due to the presence of a true Spitzenkörper (Roberson *et al.*, 2011; Fisher *et al.*, 2018)

1.7.3.6. Mucoromycota

Mucoromycota is the largest fungal clade of the old "Zygomycota" and is composed mostly of plant symbionts and saprotrophs (e.g. mycorrhizae, root endophytes, etc.) (Figure 15E-H). A few are opportunistic animal and fungal parasites (Hoffmann *et al.*, 2013), including some causing infections in human (Kwon-Chung, 2012; Serris, Danion, & Lanternier, 2019). Mucoromycota are the sister lineage to Dikarya. The two are plant-associated groups indicating that their common ancestor had likely established an intimate relationship with plants (Spatafora *et al.*, 2017; Strullu-Derrien *et al.*, 2018a). It is also in this group where the first complex multicellular sporocarps first developed (Bidartondo *et al.*, 2011; Smith *et al.*, 2013) This makes of Mucoromycota a key group to understand the transition from unicellular coenocytic structures to true complex multicellular thalli (Nagy, Kovács, & Krizsán, 2018).

Mucoromycota is composed by three subgroups Mucoromycotina, Morteriellomycotina and Glomeromycotina (Glomeromycota) (Spatafora *et al.*, 2016b). We have included Glomeromycota within Mucoromycota given recent taxonomic evidence suggesting their association (Chang *et al.*, 2015; Spatafora *et al.*, 2016b; Tedersoo *et al.*, 2018). However, their elevation as an individual phylum as suggested by early phylogenies (Schüßler, Schwarzott, & Walker, 2001) is still under discussion due to their phenotypic peculiarities and their historical use (Naranjo-Ortiz & Gabaldón, 2019a).

- Mucoromycotina is a group of mainly saprotrophs, and occasional parasites of animals and fungi, including some obligate parasites of fungi (Benny *et al.*, 2014). Fungal sexual reproduction was first demonstrated in this group. They also exhibit a phototactic response. Consequently, they are important model organisms (Blakeslee, 1904; Tisch & Schmoll, 2010). Ectomycorrhizal associations with plants and liverworts have been described (Orchard *et al.*, 2017).
- Mortierellomycotina has been created exclusively in the base of phylogenetic analyses (White *et al.*, 2006; Spatafora *et al.*, 2016b).The paraphyletic clade *Mortierella* contains numerous species

(Petkovits *et al.*, 2011). These organisms are commonly isolated from soil. Many are plant root endophytes with an undetermined effect on the host fitness (Hoff *et al.*, 2004; Summerbell, 2005).
- Glomeromycota is exclusively composed by obligate plant symbionts through the formation of arbuscular mycorrhizae (AM). The only known exception is *Geosiphon,* which establish symbioses with cyanobacteria (Redecker & Schüßler, 2014) (Figure 15G-H). AM symbiosis is the most common and widespread mycorrhizal symbiosis on the planet. AM is present in phylogenetically distinct vascular plant lineages and early-diverging land plants, including liverworts (Marchantiophyta) and hornworts (Anthocerotophyta) (Smith & Read, 2008; Pressel *et al.*, 2016). Hyphae from these species grow into the apoplastic space between plant cells, penetrating within cells forming arbuscules. Arbuscules are the key branched structures involved in nutrient exchange between the plant and the fungus (Bonfante & Genre, 2010).

1.7.3.7. Dikarya

Dikarya are most species-rich and by far the best studied group of Fungi (Figure 15I-L). It is a widely recognized and phylogenetically stable group including mainly saprotrophic fungi associated to the decomposition of organic matter (Chang *et al.*, 2015; Hibbett *et al.*, 2018; Tedersoo *et al.*, 2018) (Figure 12). Since the focus of this manuscript is far from this clade, we will only provide a few introductory notes to the group.

The name Dikarya derives from the key trait of the clade: the presence of two (sometimes genetically distinct) nuclei in some stages of their life cycle (Spatafora *et al.*, 2017). Dikarya species exhibit regularly septate hyphae with two nuclei; if their hyphal cells possess the same genotype they are referred to as homokaryotic and if the genotype is different, they are referred as heterokaryotic.

Based on the nature of these nuclei in each stage, two main groups are distinguished within Dikarya, Basidiomycota (Figure 15I-J) and Ascomycota (Figure 15K-L). During their lifecycle, homokaryon hyphae fuse during plasmogamy to produce the heterokaryon stage, which is the main vegetative stage in Basidiomycota. This heterokaryon suffers karyogamy followed by meiosis to form basidiospores. Thus, time and space separates plasmogamy from karyogamy and meiosis in Basidiomycota.

In Ascomycota, the heterokaryon state is only found in sexual cells, the vegetative mycelium is always homokaryotic. Female gametangia (ascogonia) and male gametangia (antheridia) fuse,

forming the heterokaryotic stage. This is followed shortly after by karyogamy and meiosis, generating ascospores. Thus, all three process co-exist in time and space.

The Basidiomycota main trait is the formation of the basidium, a specialized structure from which, with few exceptions, four sexual spores form in outgrows of the basidia (Hibbett *et al.*, 2007; Adl *et al.*, 2012). In Ascomycota, meiosis leads to the formation of a sac-like structure containing (normally) eight spores.

Ascomycota is the largest clade of all fungi, encompassing up to two thirds of all described species (Lutzoni *et al.*, 2004; Schoch *et al.*, 2009).They notably include model species that have helped make breakthrough discoveries in molecular biology field (e.g. *Saccharomyces cerevisae*, *Neurospora crassa*, etc).

Dikarya include organisms capable of forming complex multicellular reproductive structures, and secondarily unicellular organisms referred to as yeasts (Nagy *et al.*, 2018) (Figure 15L). The secondary transition to unicellular lifestyles occurred independently in several fungal lineages (Nagy *et al.*, 2014; O'Malley, Wideman, & Ruiz-Trillo, 2016).

Many species of Ascomycota, Basidiomycota form ectomycorrhiza (ECM) with plants, mostly shrubs and trees from temperate, boreal and Mediterranean regions (Yamamoto *et al.*, 2017). In addition, two plant families (Orchidaceae an Ericaceae) display endomycorrhizas involving Basidiomycota and Ascomycota (Dearnaley, Martos, & Selosse, 2012; Lallemand *et al.*, 2016).

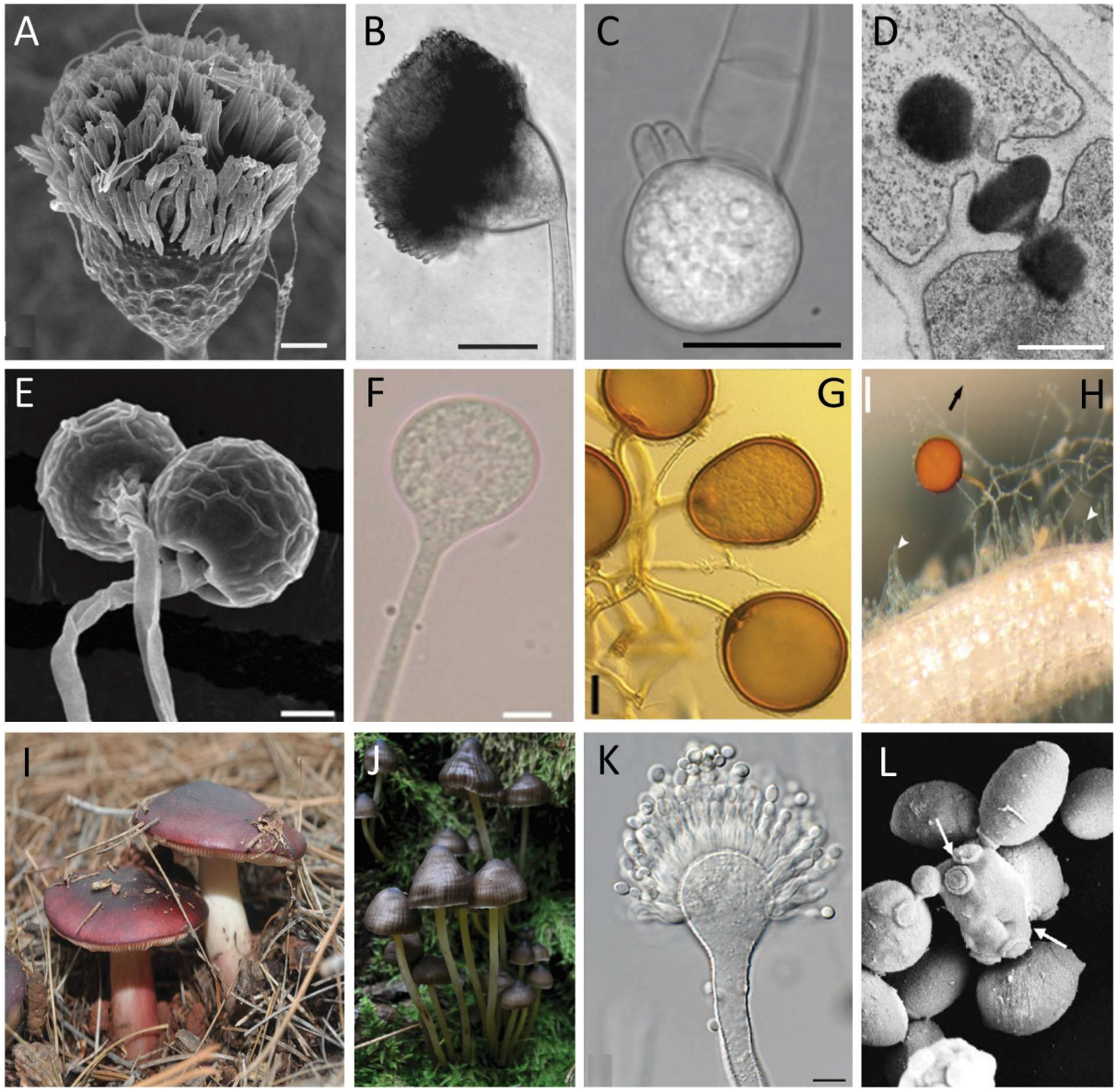


Figure 15. Photographs and transmission/scanning electron and light microscopy images of Zoopagomycota (A-D), Mucoromycota (E-H) and Dikarya (I-L). (A) SEM and light microscopy (B) of *Syncephalis pseudoplumigaleata* sporangiophore (Benny *et al.*, 2016a). C) Immature zygospore of *Basidiobolus ranarum* (Benny *et al.*, 2014). D) TEM longitudinal section of a hyphae of *Dimargaris cristalligena*, with focus on the lenticular plug connection (Jeffries & Young, 1979). (E) SEM and light microscopy (F) of *Absidia glauca* and *Absidia pseudocylindrospora* respective sporangia (Nguyen *et al.*, 2016). G) *Glomus tetrastratosum* spores (Błaszkowski *et al.*, 2015). H) A *Diversispora epigaea* spore attached to a root extraradical hyphae (black arrows) and root hairs (white arrows) (Sun *et al.*, 2018). I) *Russula queletii* fruiting body. J) *Mycena* sp. fruiting body. K) *Aspergillus aflatoxiformans* conidiophores and conidia (Frisvad *et al.*, 2019). L) SEM of *Saccharomyces cerevisiae* cells (Agizzio *et al.*, 2006). Scale bars: A = 10 µm, B = 20 µm, C = 70 µm, D = 0.5 µm, E = 30 µm, F = 20 µm, G-H = 50 µm, K = 10 µm.

1.7.4. *Incertae sedis* lineages

1.7.4.1. *Olpidium*

Olpidium (including *O. brassicae* and *O. bornovanus*) (Uebelmesser, 1956; Barr & Hadland-Hartmann, 1977) is a genus of morphologically reduced zoosporic fungi that infect nematodes, rotifers (Barron & Szijarto, 1986; Meirinho *et al.*, 2013) and Brassicaceae plant roots (Powell & Letcher, 2014) (Figure 16A-D). It was originally assigned to the Spizellomycetales (Barr, 1980). However, early ultrastructural analyses already showed unique features of *O. brassicae* zoospores, including cone-shaped striated rhizoplasts, gamma-like particles and rough endoplasmic reticulum (Lange & Olson, 1976). This, together with ultrastructural features including the holocarpic and endobiotic nature of its sporangia suggested a potential relationship between *Olpidium* and *Rozella* (Held, 1975). However, unlike *Rozella*, *Olpidium*'s sporangia develop a cell wall inside the host cytoplasm (Held, 1981). The surprise came when the first molecular phylogenies for this group placed *Olpidium* within Zoopagomycota, in most of the cases in close association with Basidiobolaceae (James *et al.*, 2006a; Sekimoto *et al.*, 2011; Tedersoo *et al.*, 2018). *Basidiobolus* possesses a spindle pole body that contains 11–12 singlet microtubules similar to a centriole, suggesting a recent transition from a zoosporic state (McKerracher & Heath, 1985). If *Olpidium* indeed branched within the terrestrial non-flagellated Fungi, this would imply an independent flagellar loss event (James *et al.*, 2006a).

Few ecological and morphological features unite *Olpidium* and Zoopagomycota. *Olpidium* are mostly cucurbit root pathogens, whereas most Zoopagomycota are associated with animals (see Zoopagomycota chapter). At the same time, *Olpidium* is fast evolving, and it might be speculated that a change in lifestyle led to that acceleration of evolutionary rate and, potentially, to phylogenetic reconstruction artifacts. Based on phylogenetic analyses and estimates of divergence time, it was recently suggested that *Olpidium* could form an independent phylum, Olpidiomycota (Tedersoo *et al.*, 2018). Nevertheless, further genomic data will be needed to robustly determine the phylogenetic position of *Olpidium*. Multi-gene phylogenomic studies should in particular help to solve the relationship of this clade with non-flagellated terrestrial fungi.

Beyond *Olpidium,* it is worth mentioning other chytrid species, such as *Caulochytrium protostelioides* since their zoospores display similar ultrastructure and aerially-produced sporangia

to *Olpidium*. Thus, it has been suggested that *C. protostelioides* could branch together with *Olpidium* (James *et al.*, 2006b).

1.7.4.2. Sanchytrids

Sanchytriaceae (sanchytrids) is a recently described clade of zoosporic chytrid-like parasitic fungi of the yellow-green algae *Tribonema gayanum*. They are characterized by monocentric and epibiotic thallus, which penetrates the hosts cell wall with rhizoids (Karpov *et al.*, 2017a) (Figure 16E-H).

There are two described species, *Amoeboradix gromovi* (Figure 16E, G) and *Sanchytrium tribonematis* (Figure 16F,H), both have amoeboid zoospores with a posterior pseudocilium; no swimming zoospores have ever been observed. They have also a reduced flagellum (= pseudocillium) with an axoneme having only 9 singlets, without the central pair, or as few as only 4 microtubules. The sanchytrid kinetosome, which in eukaryotes would generally present a more conserved 9x3 plus the central pair, possesses only 9 singlet microtubules in *S. tribonematis* and 9 singles or doublets (depending on the strain) in *A. gromovi*. However, despite this flagellar ultrastructural reduction, sanchytrids zoospores have a remarkable long kinetosome. *A. gromovi* possesses one of the longest kinetosomes known for a eukaryotic cell, reaching a length of 2.2 µm (Karpov *et al.*, 2018, 2019). Sanchytrids unique flagellar composition might help us to get insights into the evolution of flagella in Holomycota. But for that, it is essential to resolve their phylogenetic relationships with other fungal lineages.

Sanchytrium was first described and initially classified as a Monoblepharidomycetes by SSU and LSU rRNA gene phylogenies (Karpov *et al.*, 2017a). *Amoeboradix* was described a bit later, and both were confirmed to be part of the same clade by 18S + 28S rRNA gene phylogenies. Surprisingly, they branched at the base of Glomeromycota + Dikarya without good phylogenetic support, such that their position within Fungi remains undetermined (Karpov *et al.*, 2018). As in previous examples, this uncertainty was probably a by-product of low phylogenetic signal and a long phylogenetic branch, making the clade susceptible to long branch attraction artefacts. Similar to what it is observed for *Olpidium*, genomic sequence data will be necessary to better resolve the phylogenetic position of *Amoeboradix* and *Sanchytrium* within the fungal tree.

1.7.4.3. Other *incertae sedis* lineages

Other Holomycota lineages of uncertain phylogenetic affinity for which genomic/transcriptomic data will be needed include:

- Nephridiophagidae. It is a unique clade of non-flagellated unicellular eukaryotes, of uncertain phylogenetic position (Figure 16I-J). Members of this group parasitize the Malpighian tubules of insects and myriapods, and have been extensively characterized by ultrastructure (Fabel, Radek, & Storch, 2000; Radek *et al.*, 2017). They are extracellular, multinucleated and produce an amoeboid stage that attaches to the host's Malpighian tubule lumen microvilli. 18S rRNA gene trees have only confirmed that they form a distinct clade within Fungi, probably branching near the root of Fungi (Radek *et al.*, 2017).
- Several groups that are known only by environmental samples: The Basal Clone Group 1 (BCG1) (Nagahama *et al.*, 2011; Bass *et al.*, 2018; Tedersoo *et al.*, 2018; Chambouvet *et al.*, 2019), which is a marine clade with an apparent relationship to Rozellida. This clade has also been called novel chytrid-like-clade-1 (NCLC1) (Richards *et al.*, 2015). Recently, using rRNA-targeted fluorescent in situ hybridization (FISH) microscopy, NCLC1 was demonstrate to form intracellular infections in key diatom species (Chambouvet *et al.*, 2019) (Figure 16K). They seem to have an "Opisthosporidia"-like lifestyle.

The Namako-37, LKM47, LMK11 and LKM15 are environmental clades branching within the rozellids radiation (Takishita *et al.*, 2007; Lara *et al.*, 2010; Nagahama *et al.*, 2011; Bass *et al.*, 2018; Corsaro *et al.*, 2019). Lastly, the Basal Clone Group 2 (BCG2) is a freshwater lineage that seems to branch at the base of the Aphelida + Fungi clade (Monchy *et al.*, 2011; Tedersoo *et al.*, 2017, 2018).

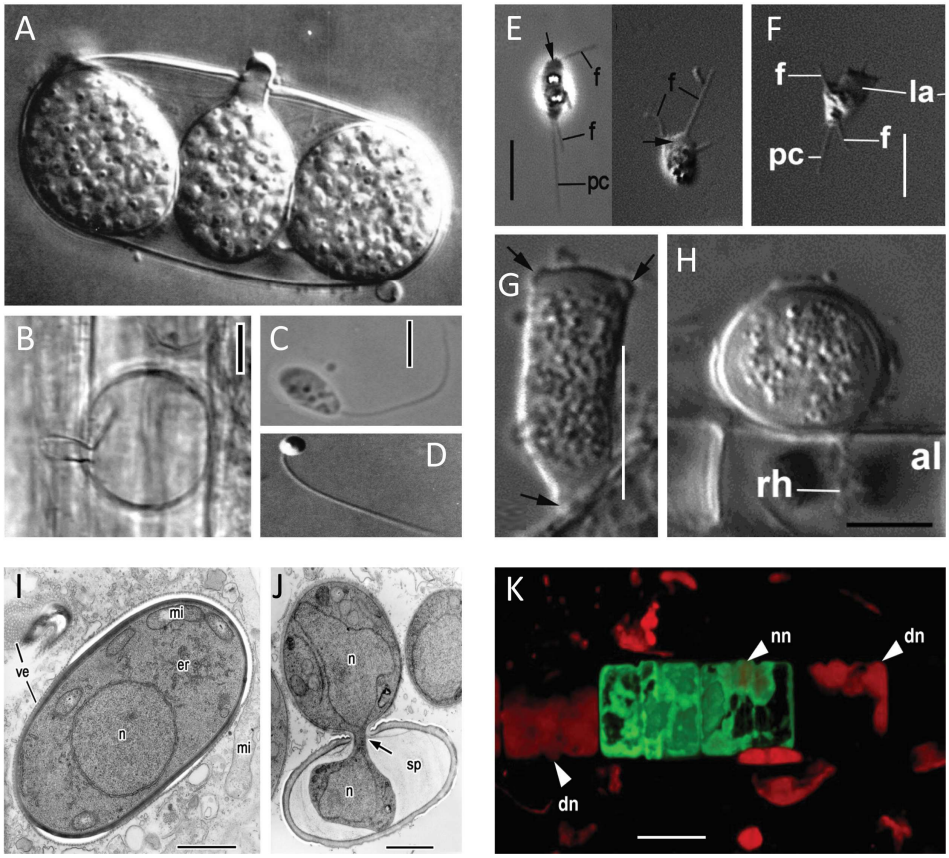


Figure 16. Illustrations and transmission electron and light microscopy images of *Incertae sedis* lineages. A) *Olpidium vermicola* sporangia in a nematode egg just before zoospore release (x 1200) (Barron & Szijarto, 1986). B) *O. bornovanus* empty sporangium, after zoospore release (Sekimoto *et al.*, 2011). C) *O. bornovanus* zoospore (Sekimoto *et al.*, 2011). D) *O. vermicola* zoospore (x 3000). E, G) *Amoeboradix gromovi* zoospores and sporangium respectively (Karpov *et al.*, 2018). F,H) *Sanchytrium tribonematis* zoospores and sporangium respectively (Karpov *et al.*, 2019). I-J) *Nephridiophaga blattellae* spore before and after hatching (Radek *et al.*, 2017). K) 3D confocal micrograph of a *Chaetoceros*-like diatom intracellularly infected by NCLC1 (BCG1) members as detected by FISH (Chambouvet *et al.*, 2019). pc (pseudocilium), f (filopodium), la (lamellipodium), al (alga), rh (rhizoid), sp (spores), mi (mitochondria), n (nucleus), er (endoplasmic reticulum), dn (diatom nucleus) and nn (NCLC1 nucleus). Scale bars: B = 10 µm, C = 5 µm, E-F = 5 µm, G = 10 µm, H = 5 µm, I-J = 1 µm, K = 10 µm.

1.7.5. Fungal synapomorphies

Several morphological and molecular characters have been stated as synapomorphies for the clade. However, if we define synapomorphies as characters that must be ancestrally present in the clade (and virtually in most of the taxa within a clade), and at the same time absent from the rest of the tree of life, no fungal character fits the requisites.

Among the supposedly unique and defining traits of fungi is the synthesis of the amino acid lysin via the α-aminoadipate pathway. However, it has been shown to be present in other eukaryotes including Euglena, or even in prokaryotes (Vogel, 1964; Kosuge & Hoshino, 1999; Torruella *et al.*, 2009). Some other characters have failed to be considered synapomorphies not only because they are found in other parts of the tree of life but also because they have a patchy distribution within Fungi. Examples of this are the presence of ergosterol, missing in several fungal taxa (Weete, Abril, & Blackwell, 2010) but present in numerous protists (e.g. Stern et al. 1960; Scotia & Columbia 2013), or the presence of a chitin cell wall. However, elements of the chitin synthase pathway are missing in some fungi (Bruns *et al.*, 1992; Ma *et al.*, 2016), and chitin cell walls occur in other eukaryotes (Lin & Aronson, 1970; Das *et al.*, 2006; Mélida *et al.*, 2013).

At any rate, two major features seem to define Fungi: osmotrophic feeding and polarized cell/hyphal growth. The coupling of both features is most likely one of the causes of the ecological success of fungi (Richards *et al.*, 2017a).

Osmotrophy, is the absorption of dissolved nutrients by diffusion across a plasma membrane. It is usually opposed to phagotrophy, an ancestral eukaryotic feature which consists of engulfing food particles by phagocytosis prior to digestion. The early divergent fungi Chytridiomycota and Blastocladiomycota were probably the first to evolve this trait since their rhizoids, similarly to hyphae, secrete enzymes prior to the import of dissolved organics. Osmotrophy is present in all other Fungi including Zygomycota and Dikarya. However, other members within Holomycota are most likely phagotrophs. There is compelling evidence for phagotrophy in trophic stages of nucleariids (Brown *et al.*, 2009), rozellids and aphelids (Powell, 1984; Karpov *et al.*, 2013; Torruella *et al.*, 2018).However, phagotrophy was secondarily lost in Microsporidia (Bass *et al.*, 2018). This implies that the ancestor of Holomycota was most likely phagotrophic (Torruella *et al.*, 2018). Phagotrophy is indeed widespread in non-fungal opisthokonts. A similar example of convergent phagotrophy loss in the holozoan branch is that of the osmotroph *Corallochytrium limacisporum* (Torruella *et al.*, 2015).

Polarized growth in fungi is involves continuous vesicle flow from the hyphal cell body to the growing hyphal tip, a process essential for cell growth. Several “thought to be” unique organelles (e.g. Spitzenkörper) and multiprotein complexes (e.g. exocyst and polarisome) are involved (Kiss *et al.*, 2019). This process was thought to be limited to Dikarya and the paraphyletic Zygomycota. However, some exceptions have been found like the early divergent blastoclad *Allomyces*, that

presents similar structures to the *Spitzenkörper* (Vargas *et al.*, 1993). This might suggest that polar growth evolved earlier than thought in the fungal tree.

Some components of the exocyst (e.g. Sec4), another characteristic complex involved in polarized growth, have orthologs in other eukaryotes. The exocyst seems to play a global role in eukaryotic vesicles, cytoskeleton and membranes (Koumandou *et al.*, 2007; Elias *et al.*, 2012). In the case of another typical fungal complex, the polarisome, available data suggest it might be a true synapomorphy, although more extensive comparative genomic analyses of the unexplored protist diversity will be need for confirmation.

Thus, it has been claimed that truly "fungal-specific" characters might not exist either because of patchy distribution within fungi, or by its presence elsewhere in the tree of life (Figure 17).

Nonetheless, the gene families *MedA* and *APSES* have more recently been suggested as candidate molecular fungal synapomorphies. These gene families contain fungal-specific protein domains that evolved in early fungi and potentially represent fungal-specific innovations for hyphal growth (Kiss *et al.*, 2019). Additional taxon sampling might however unveil more complex evolutionary histories.

The difficulties to define exclusive synapomorphies for the Fungi creates uncertainty on where to draw the limit of Fungi in phylogenetic trees. This uncertainty comes from the fact that the limit of fungi in the eukaryotic tree of life keeps changing and being redrawn constantly, every time a new lineage is discovered. This is a direct consequence of the fact that the fungal tree remains incomplete both in genomic data and in diversity sampled. Thus, it has been argued that for the moment the only certainty about Fungi is that there are no certainties about Fungi. Once we understand this, it is easy to get around the idea that for now the only way to name a given "X" taxon within Fungi is as the sister group to the "Y" taxon. This clearly changes how we approach their biology. As stated by Richards et al. (2017), the focus should be less on constructing categories and more on seeking explanations for major transitions that occurred during the evolution and diversification of the clade (e.g. flagellum loss).

1.8. Multicellularity: Fungi vs Protist

The lack of exclusive fungal synapomorphies opened another debate, related with whether some organisms in this branch should be classified as Fungi, or as protists (meaning unicellular

eukaryotes) (Figure 17). At a first glance, it should be straightforward to distinguish between them. If a given organism from Opisthokonta is multicellular, then is a metazoan or a fungus; if it is unicellular, then is a protist. Nevertheless, as it usually happens in nature, it is not that simple.

In Holozoa and Holomycota, the absence of evident synapomorphies occurs in both multicellular and unicellular lineages (Donachie *et al.*, 2017; Richards *et al.*, 2017b). Nevertheless, when the scientific community draws the line between metazoans and their unicellular relatives there seems to exist a clear consensus (e.g. Lang et al. 2002; Torruella et al. 2015; Hehenberger et al. 2017). But when it comes to Holomycota, there is a trend to keep englobing as Fungi organism that are unicellular (rozellids, Microsporidia, aphelids) or coenocytic (Chytridiomycota, Blastocladiomycota and most Zoopagomycota) (e.g. Chang et al. 2015; Spatafora et al. 2016; Tedersoo et al. 2018). To try to understand why this occurs, it is essential to first define fungal multicellularity.

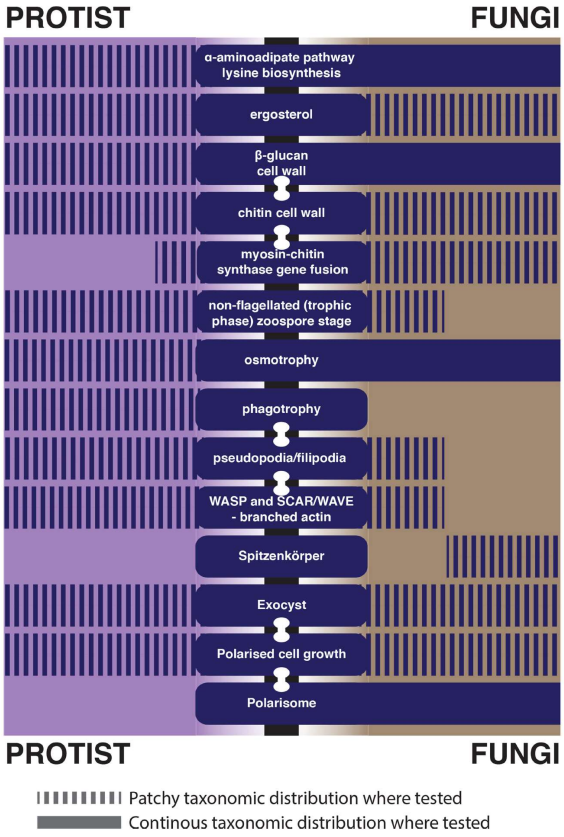


Figure 17. Figure from Richards *et al.*, (2017) illustrating a summary of how features previously thought to define the protist-fungal boundary have been observed to have a mosaic distribution. This distribution has been observed both within and outside the Fungi (Holomycota). White connecting nodes illustrate linked characters.

One definition of fungal multicellularity would be that of a thallus made of hyphae. Hyphae are basically tubular structures that grow apically to form a mycelium that explores and extends in its substrate (Kiss *et al.*, 2019). Hyphae are hypothesized to have evolved from the gradual elongation of substrate anchoring rhizoids of unicellular ancestors resembling modern Chytridiomycota and Blastocladiomycota (Harris, 2011). Both Blastocladiomycota and chytrids form unicellular coenocytic thalli with one or multiple nuclei, although some species do form structures that resemble true hyphae (e.g. *Allomyces*) or narrow exit tubes on sporangia (e.g. *Catenaria* spp.) (Sanders, 2002; Archibald *et al.*, 2017; Berbee *et al.*, 2017). Nevertheless, since most taxa of these clades are unicellular, we cannot discard that these hyphal-like forms result from convergent evolution.

Primitive "hyphae" would be unicellular structures which lack compartmentation, forming coenocytic multinucleated structures without regulation of the cytoplasm content (Kiss *et al.*, 2019). This organization resembles that of some Monoblepharidomycetes chytrids and most Zoopagomycota (James *et al.*, 2006a; Sekimoto *et al.*, 2011; Spatafora *et al.*, 2017). Taking of this into consideration, all chytrids, Blastocladiomycota and most Zoopagomycota could be considered protists. Then consequently, like in the case of Holozoa, these would have to be treated as 'unicellular relatives of Fungi' (Torruella *et al.*, 2015). This is an open debate, and if instead of drawing the line on unicellularity we draw the line on osmotrophy (for example), zoosporic fungi and Zoopagomycota would be considered Fungi, maybe even "fungal protists" if we use both traits as criteria.

Fungal multicellularity is unique and different from other eukaryotic multicellularities. Its origin does not depend on the expansion of kinases, receptors or adhesive proteins, like observed in animals (e.g. Sebé-pedrós et al. 2017). Its origin seems to be more related with the co-option and exaptation of ancient eukaryotic genes than with gene duplications (Kiss *et al.*, 2019). Instead of clade-specific innovations as in Holozoa (Donachie *et al.*, 2017); in Holomycota limited innovations occurred, which is also another potential explanation to the limited amount of available synapomorphies for the Fungi.

The origin of multicellularity in Fungi seems to have occurred in the split of Blastocladiomycota, Chytridiomycota and Zoopagomycota (Kiss *et al.*, 2019). Since as in Holozoa (Sebé-pedrós *et al.*, 2017), unicellular representatives already had evolved and developed many of the genes needed for the multicellular transition (even when the mechanisms were different). This would also

explain some of the already mentioned intermediate forms observed in the earlier branches (e.g. *Allomyces*, Monoblepharidomycetes).

Another layer of complexity has been added by the distinction between simple and complex multicellularity (Nagy *et al.*, 2018). Simple multicellularity (SM) is observed in Mucoromycota, which evolved from organisms with primitive coenocytic "hyphae". SM is involved in the hyphal growth of osmotrophic fungi in the substrate, i.e. in the feeding process. Complex multicellularity (CM) is that of structures showing a three-dimensional differentiated organization resulting from a spatial and temporal developmental program that develops until a genetically determined shape and size is reached. Organisms with CM develop sexual fruiting bodies, i.e. CM is involved in reproductive roles. Both SM and CM occur in different life stages of CM organisms.

In summary, according to these criteria, only Dikarya and Mucoromycota possess true hyphae and can be stated as true multicellular lineages within Fungi (Figure 18. from Nagy *et al.*, 2018). However, as mentioned, a significant diversity of coenocytic forms exists in the Blastocladiomycota, Chytridiomycota and to a smaller extent in the Zoopagomycota, which can be stated as unicellular fungi and are therefore comparable to all other unicellular members of Holomycota. Future discoveries, discussions and conventions might help to better define the boundaries between unicellular and multicellular fungi.

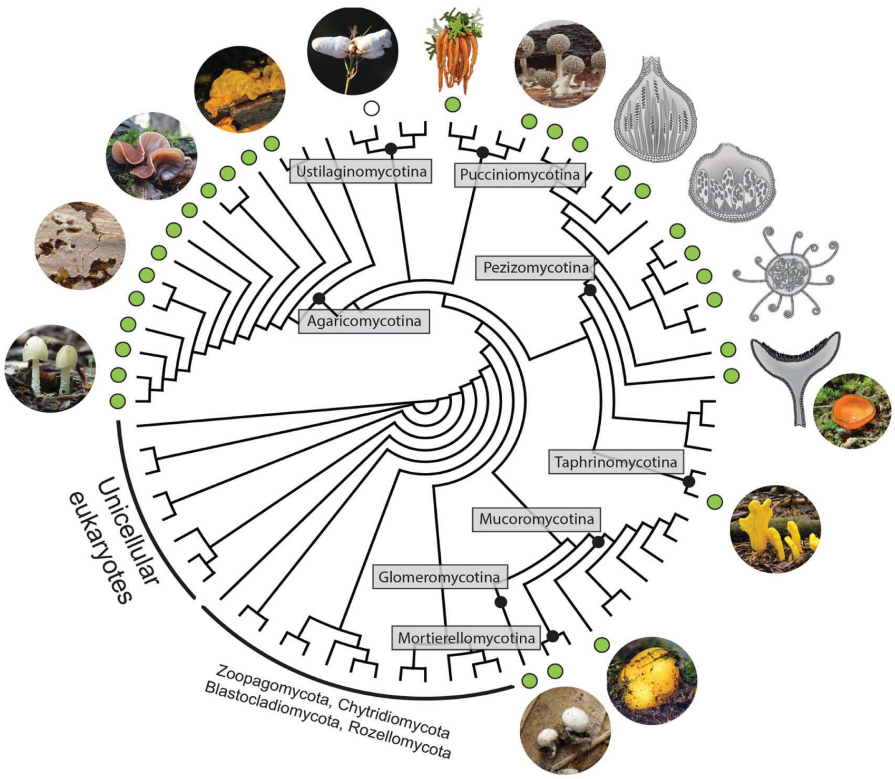


Figure 18. Figure from Nagy *et al.*, (2018) showing the current known phylogenetic distribution of complex multicellularity in fungi. Images show examples of typical complex multicellular morphologies of sexual fruiting bodies for each major clade within Fungi. Green dots indicate lineages with known complex multicellular representatives, and an empty circle indicates an uncertain status.

1.9. Early fungal evolution and the holomycotan flagellum

Molecular clock analyses suggest that the opisthokont most recent common ancestor (MRCA) dates back to 1-1.5 billion years ago. Holomycota might have evolved around 1 billion years ago with the rise of the nucleariid lineage (Zettler *et al.*, 2001). The diversification of "Opisthosporidian" holomycotans may have occurred soon after. Lastly, the ancestor of Chytridiomycota and Blastocladiomycota might have evolved around 750 Mya (Douzery *et al.*, 2004; Parfrey *et al.*, 2011; Eme *et al.*, 2014; Berbee *et al.*, 2017).

At any rate, the earliest non-ambiguous fungal fossils in the geological record correspond to zoosporic lineages resembling Chytridiomycota and Blastocladiomycota. They include 407 My-old chytrid- and blastoclad-like organisms from the Rhynie Chert (Krings *et al.*, 2010; Strullu-Derrien *et al.*, 2016, 2018b). More recently, a potential 1 Ga-old fungus microfossil from the Grassy Bay Formation in Canada has been described (Loron *et al.*, 2019). If this is confirmed, estimates of opisthokontan emergence would be older by half a million years.

The respective MRCA of Opisthokonta, Holozoa and Holomycota likely had a posterior flagellum (Cavalier-Smith, 1987; Torruella *et al.*, 2015, 2018).

Most Holozoa have retained that flagellated state at some stage during their life cycle. Although some independent losses took place in some groups, traces of its former presence can be found. For instance, the filasterean *Ministeria* has been observed to possess a rudimentary flagellum, and molecular components of the flagellar toolkit have been found in both *Ministeria* and *Corallochytrium* (Torruella *et al.*, 2015), confirming a secondary flagellar loss from the flagellated holozoan ancestor. In Holomycota, ancestral reconstruction analyses point out to the same trend, with independent flagellar losses from a flagellated holomycotan (opisthokont) flagellum (Torruella *et al.*, 2018)

Within Fungi, and as we have seen (see previous chapters), phylogenetic evidence showed that flagellated fungi (Chytridiomycota and Blastocladiomycota) were sister to the remaining clade grouping non-flagellated fungi (Zygomycota, Glomeromycota and Dikarya). This suggested a

single flagellum loss around the time when fungi are supposed to have colonized land (Liu, Hodson, & Hall, 2006a). The panorama recently changed, after more recent phylogenetic studies started questioning the monophyly of Chytridiomycota and Zygomycota (Tanabe *et al.*, 2004; James *et al.*, 2006b; Spatafora *et al.*, 2016a), and suggested the relationship of nucleariids (Brown *et al.*, 2009; Liu *et al.*, 2009) and Microsporidia (Hirt *et al.*, 1999; Keeling, 2003) with Fungi.
Current estimates infer 4 to 6 independent flagellum losses during the evolution of Holomycota (James *et al.*, 2006a). Two losses occurred independently in early holomycotan branches, one for Microsporidia, and another for the nucleariid clade. Another well-supported loss took place along the branch leading to the chytrid *Hyaloraphidium curvatum* (see chapter 1.7.3.1. Chytridiomycota). However, of the number of total losses (2–4 losses) within the Zoopagomycota + Mucoromycota + Dikarya clade remains uncertain. This is due to the problematic placement of *Olpidium* (see *Olpidium* chapter). Solving the phylogenetic placement of *Olpidium* and other *incertae sedis* organisms like sanchytrids (which present a reduced flagellar structure), will be essential to understand the number of independent losses in Holomycota.

1.10. Single cell genomics applied to the study of the holomycotan 'dark matter'

The genomic era in Fungi and Holomycota started with the pioneer sequencing of yeast *Saccharomyces cerevisae* genome (Goffeau *et al.*, 1996). Today, more than 800 holomycotan genomes have been sequenced (Spatafora *et al.*, 2017). These genomes (and transcriptomes) have helped reconstructing the holomycotan tree of life with unprecedented resolution. However, if many relationships between important lineages have been solved, many new questions arise (Chang *et al.*, 2015; Mikhailov *et al.*, 2017; Torruella *et al.*, 2018).
The molecular era also brought environmental sequencing studies including 18S/ITS rRNA (meta)barcoding approaches. These studies have consistently shown a wide diversity of holomycotan clades, some of which are novel and remain unexplored (Lilleskov *et al.*, 2002; Cox *et al.*, 2010; Tedersoo *et al.*, 2014; Yahr, Schoch, & Dentinger, 2016; Bass *et al.*, 2018). However, 18S rRNA gene analyses are insufficient to solve deep relationships among many holomycotan clades.
To solve this problem and reconstruct robust phylogenetic trees, it is essential to obtain genomic and transcriptomic data from these new clades. However, especially for the unicellular uncultured

fraction of Holomycota, isolating the organisms and then obtaining enough material to sequence can be a difficult task.

Recently, single-cell omics have risen as one of the best approaches to unveil this unicellular unculturable fraction, also known as microbial 'dark matter'. Single-cell 'omic' techniques represent a powerful approach to generate high phylogenetic signal from genomic/transcriptomic data from an otherwise inaccessible diversity as revealed from environmental studies. An overall similar pipeline to generate genomes/trasncriptomes from single cells can be followed.

The pipeline starts with the collection of samples from e.g. aquatic or soil environments. Single cells from the samples can be isolated using three main methodologies: micromanipulation, microdroplet isolation by microfluidics, and fluorescence-activated cell sorting (FACS). Each technique has advantages and disadvantages. Microfluidics and FACS are high-throughput techniques and can be easily applied to small free cells. However, manual micromanipulation can more easily target particular cells from several size ranges (Kolisko *et al.*, 2014; Liu *et al.*, 2017) or attached to larger structures. FACS protocols are currently more adapted to protist single-cell fractionation than most microfluidic techniques, still requiring improvement, and have already led to the generation of high quality genomes and transcriptomes (Rinke *et al.*, 2014; Ahrendt *et al.*, 2018).

Once individual cells are isolated, their genomes/transcriptomes have to be amplified to have enough material for sequencing. The most commonly used technique is Multiple Displacement Amplification (MDA), with other alternative techniques like Multiple Annealing and Looping–Based Amplification Cycles (MALBAC). MDA techniques are useful but sometimes lead to a low genome recovery rate. This is due to the fact that random hexamers primers, used for amplification, might preferentially hybridize to specific areas of the genomes. This technique may also create of chimeric reads, and false positive related problems (Lasken & Stockwell, 2007; Nurk *et al.*, 2013). Before genome sequencing, genomes (and transcriptomes) can be screened for their SSU rRNA genes to confirm the identification of the single-amplified genome and limit potential contamination. Once genomes and transcriptomes are sequenced, the treatment of the data is similar to other studies with phases including assembly, annotation and comparative study genomic/transcriptomic data.

Single-cell genomics (SCG) has already been applied successfully in several studies. Some studies worth highlighting include the recent generation of 206 eukaryotic single-cell genomes from

heterotrophic flagellates (Wideman *et al.*, 2020), and eight uncultured species genomes across the fungal tree of life (Ahrendt *et al.*, 2018).

Single-cell genomics it is still under development and needs to overcome problems related with cell lysis and amplification methods. However, it has the potential to unveil a great proportion of the unculturable fungal fraction from the eukaryotic 'dark matter' (Grossart *et al.*, 2016).

The SINGEK (SINgle cell Genomics to explore the ecology and evolution of hidden microeuKaryotes) network emerged in this context (http://www.singek.eu/). SINGEK was an EU H2020 Marie Skłodowska-Curie Innovative Training Network created to provide a unique and structured PhD training program to a new generation of scientists with the highest expertise in Single Cell Genomics. The network was composed of a multidisciplinary team of researchers from nine institutions with high expertise in eukaryotic SCG. I was lucky enough to be selected to carry out my PhD within this network and apply SCG approaches to generate genomic data to better resolve the holomycotan tree of life and gain insights about their evolution. Based on this main premise, the specific objectives or my PhD are as follows.

2. Objectives

"Así, las células más simples que nos es posible estudiar, no tienen nada de "primitivo". Son el producto de una selección que ha podido, a través de medio billón a un billón de generaciones, acumular un aparejo teleonómico tan poderoso que los vestigios de las estructuras verdaderamente primitivas son indiscernibles."

Jacques Monod. Le hasard et la nécessité. Essai sur la philosophie naturelle de la biologie moderne (1970)

2. Objectives

1. Resolve the inner phylogenetic relationships of nucleariids using single-cell genomic and culture-based techniques.

Nucleariid amoeba appear as the sister lineage to all other holomycotans. They have been known since the 19th century and their relationship with Fungi was recovered only recently by several studies. However, there is almost no genomic nor transcriptomic data from members of the clade, and environmental data is insufficient so resolve the inner relationships of the clade. The first objective of my PhD implied combining single-cell and traditional cultured-based approaches to obtain genomic and transcriptomic nucleariid data in order to solve inner phylogenetic relationships within the clade. To do so we will generate and analyze data from the nucleariid *Nuclearia* and two putative cover-bearing nucleariid species without any molecular data available, *Pompholyxophrys* and *Lithocolla*. By confirming the position of these putative nucleariid species, we hope to reconstruct inner relationships in the clade and the characters present in the nucleariid ancestor. Additionally, single-cell data might give us some insight into the ecology of the clade, and help us compare the effectiveness of single-cell vs culture-based approaches.

2. Ascertain the phylogenetic position of metchnikovellids within Microsporidia and study synapomorphies for the Microsporidia + Rozellida clade.

Metchnikovellidae is a family of atypical Microsporidia. They have been known from over a century and their unique morphological features suggested their basal position within Microsporidia. Recently the partial genome of the metchnikovellid *Amphiamblys* sp. suggested a basal position within Microsporidia but it remains the only available genome from a non-taxonomically validated metchnikovellid representative. To confirm the phylogenetic placement of metchnikovellids we will sequence the single-cell genome of *Metchnikovella incurvata,* from a taxonomically verified metchnikovellid species. We will see if the two metchnikovellids cluster together and eventually confirm their position as the sister clade to core Microsporidia. Gene content analyses should provide insights into genome evolution along the Microsporidia branch and eventually point to the gain and loss of particular functions.

3. Resolve the phylogenetic placement of sanchytrids within Fungi and study their specific and shared life history traits.

The sanchytrids *Amoeboradix gromovi* and *Sanchytrium tribonematis* are two recently described enigmatic zoosporic fungi with an unresolved taxonomic placement. Phylogenies of the 18S + 28S rRNA gene confirmed their association in the same clade, but they failed to prove their affinity to any other fungal lineage. They also have atypical traits including a reduced flagellum with an extremely long kinetosome. Our third and last objective is to resolve their phylogenetic position of this enigmatic fungal clade and potentially improve the global fungal phylogeny generating single-cell genomic data from this new clade. In particular, we will compare their genomes to those of other zoosporic fungi to determine: (i) the number of independent flagellum losses in Holomycota by analyzing its flagellum toolkit, (ii) study their primary metabolic composition, and (iii) understand the molecular determinants of the reduced but peculiar sanchytrid flagellum.

3. Materials and Methods

“The method of science, as stodgy and grumpy as it may seem, is far more important than the findings of science.”

Carl Sagan. The Demon-Haunted World: Science as a Candle in the Dark (1996)

3. Materials and Methods

We used phylogenomic approaches to infer the phylogenetic position of several organisms belonging to the Holomycota and comparative genomic approaches to get insights into the evolution of several genomic characteristics of these organisms. Even if our different studies had some specificities, we followed a similar overall methodological approach that is summarized in this section. For a detailed description of specific Materials and Methods used, see the corresponding chapter.

3.1. Samples, cell isolation, single-cell genome and transcriptome amplification and sequencing

Most of our samples and organisms were obtained from freshwater and marine environments by our collaborators (*M. incurvata*, sanchytrids and most nucleariids) or from culture collections (*N. thermophila* and *N. delicatula*). There were two main types of cells used for isolation and DNA/RNA extraction: Those that were present in stable mixed or pure cultures (sanchytrids, *Lithocolla*, *N. thermophila* and *N. delicatula*) and those that we directly collected from fresh samples (*M. incurvata*, *Pompholyxophrys* and *N. pattersoni*).

Single-cell isolation was done in all cases using a micromanipulator (Eppendorf PatchMan NP2) or by manual pipette-picking of the cells, which were always washed several times to reduce contamination. Single-cell isolation was applied to all cells coming from fresh samples but also to most cells growing in culture (except *N. thermophila* and *N. delicatula*), since the last ones were present in mixed cultures and we wanted to avoid contamination from other sources (e.g. bacteria or algal food). Nevertheless, to facilitate subsequent *in silico* sequence decontamination we also extracted RNA from cultures of the food organisms. For these cultures and *Lithocolla*, *N. delicatula* and *N. thermophila,* we extracted whole RNA using the RNeasy Micro (Qiagen, Venlo, The Netherlands). For single cells, DNA and RNA were extracted using commercial kits (e.g. PicoPure from Applied biosystems). However, in some cases micromanipulated single cells were lysed during the first steps of the single-cell amplification protocol.

Whole transcriptome amplification (WTA) and whole genome amplification (WGA) of micromanipulated single cells and DNA/RNA extracted from single cells, were carried out using

the REPLI-g WTA/WGA Kits (Qiagen). In all cases WTA/WGA success was confirmed by amplification and sequencing of the 18S rRNA gene of the target species, used to confirm the identity of the amplified genome/transcriptome. This material was paired-end sequenced (2 × 100 bp or 2 x 125 bp, depending on the study) with an Illumina HiSeq 2500 instrument with chemistry v4.

3.2. Sequence quality assessment, trimming, assembly, decontamination and annotation

Quality assessment of the Illumina reads was performed with FastQC (Andrews, 2010) before and after quality and Illumina adapter trimming with Trimmomatic (Bolger, Lohse, & Usadel, 2014) in paired-end mode. Different trimming parameters were used according to the study. Resulting reads were assembled with SPAdes (with -rna mode for transcriptomes) (Bankevich *et al.*, 2012). Decontamination was always carried out in a multi-step process, by which in all cases we performed two rounds of assembly before and after bacterial sequence removal with BlobTools (Laetsch & Blaxter, 2017). Then, open-reading frames were predicted and translated from the assembled contigs using Transdecoder (http:transdecoder.github.io) with default parameters and Cd-hit (Li & Godzik, 2006) with 100% identity to produce protein sequences. After this stage, in some cases (e.g. sanchytrids and *Lithocolla*) we removed contaminating eukaryotic sequences using the predicted protein sequences from the corresponding hosts or food sources (e.g. *Tribonema* and *Navicula*) searched by BLASTp (Camacho *et al.*, 2009). Statistics of the final assemblies were assessed with QUAST (Gurevich *et al.*, 2013). To assess genome and transcriptome completeness, we used BUSCO (Simão *et al.*, 2015) with different reference databases according to the study. Functional annotation of the predicted proteins was made with different programs according to the study (e.g. eggNOG mapper) (Huerta-cepas *et al.*, 2016).

3.3. Phylogenomic and comparative genomic analyses

We used two main protein datasets for all phylogenomic analyses: dataset GBE (264 proteins) modified from Mikhailov et al. (Mikhailov *et al.*, 2016; Torruella *et al.*, 2018) and dataset SCPD (74 single-copy domains) from Torruella et al. (Torruella *et al.*, 2012, 2015, 2018). Both datasets

were updated with data of the new sequenced species retrieved by tBLASTn (Camacho *et al.*, 2009).

All alignments were performed using MAFFT (Katoh & Standley, 2013) with default parameters. Alignments were inspected manually and edited using Geneious (Kearse *et al.*, 2012), and trimmed using trimAl in automated1 mode (Capella-Gutiérrez, Silla-Martínez, & Gabaldón, 2009). In the case of 18S rRNA gene sequences in some studies we trimmed the alignments manually. Single protein trees were reconstructed with FastTree (Price, Dehal, & Arkin, 2009) and were then manually checked to detect possible paralogous and/or contaminating sequences. Finally, alignments were concatenated into a supermatrix with Alvert.py from the package Barrel-o-Monkeys (http://rogerlab.biochemistryandmolecularbiology.dal.ca/Software/Software.htm) or with Geneious.

All phylogenetic/phylogenomic trees (18S rRNA gene trees and multi-protein trees) were reconstructed using maximum likelihood (ML) (Felsenstein, 1981) and Bayesian inference (BI) (Huelsenbeck & Ronquist, 2001). The software used was IQ-TREE (Nguyen *et al.*, 2015) for ML analyses under different evolutionary models (see the different chapters) and PhyloBayes-MPI (Lartillot, Lepage, & Blanquart, 2009) for BI analyses always under the CAT-Poisson model. All trees were visualized using FigTree (Rambaut, 2016).

Comparative genomic analyses were carried out on the proteomes of the new sequenced lineages and their relatives. Annotated proteins were analyzed by comparing different GO terms, orthologs identified by Othofinder (Emms & Kelly, 2015), COG categories and/or KEGG pathways profiles (Kanehisa, 2000) (for more details see each chapter).

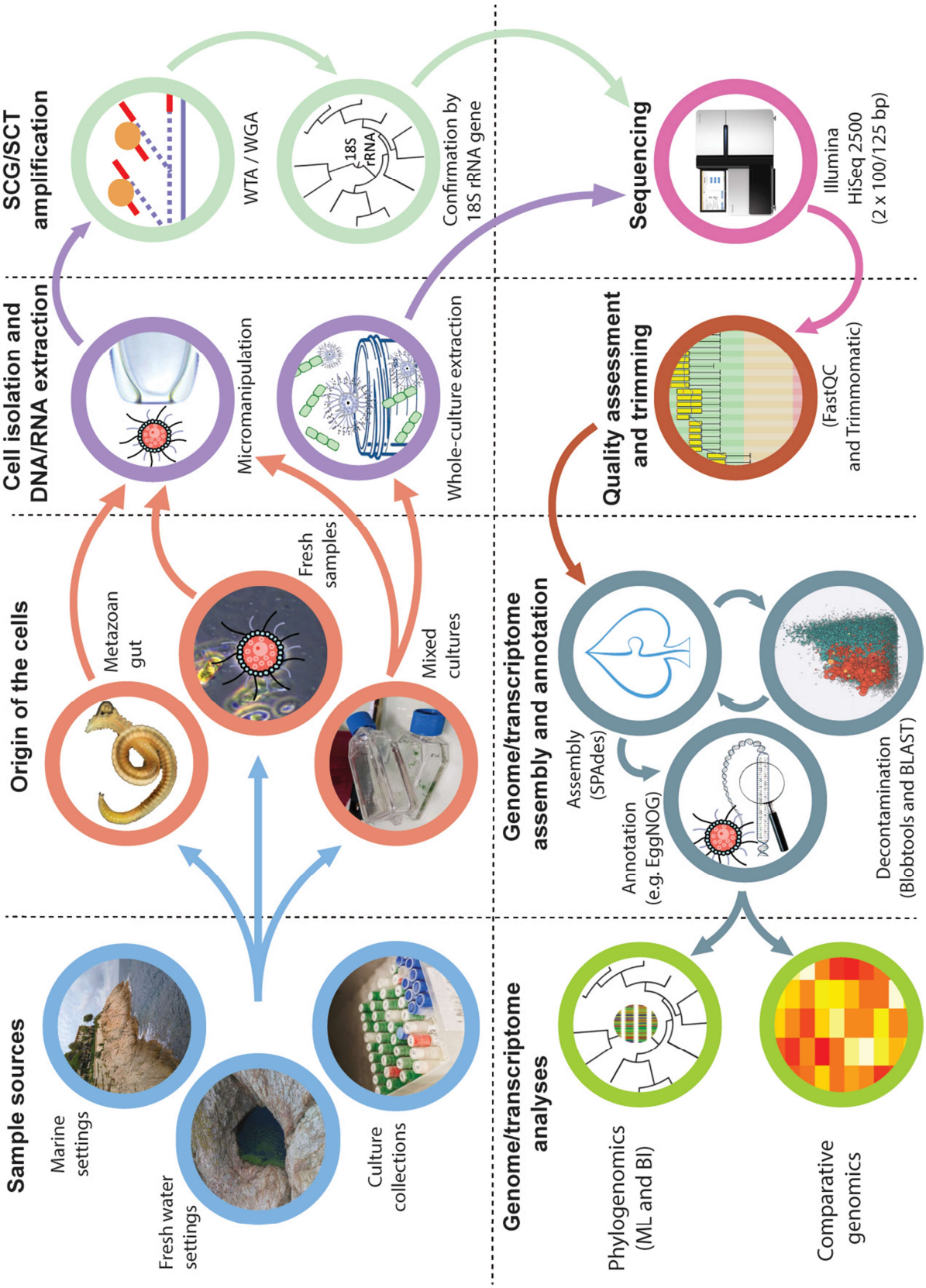


Figure 19. Flow diagram of the overall methodology used during this PhD project.

4. Combined cultivation and single-cell approaches to the phylogenomics of nucleariid amoebae, close relatives of fungi

"Evolution is no linear family tree but change in the single multidimensional being that has grown to cover the entire surface of Earth."

Lynn Margulis and Dorion Sagan. What is Life? (1995)

4. Combined cultivation and single-cell approaches to the phylogenomics of nucleariid amoebae, close relatives of fungi

4.1. Context and objectives

As previously described (see chapter 1.7.1. Nucleariids), nucleariids have been systematically recovered as the first branch to emerge in the Holomycota clade, branching sister to all other members of this lineage. However, even if nucleariids are known to be a widespread and diverse group from environmental metabarcoding analyses, the amount of available genomic/transcriptomic data for the clade was very scarce. There was data available only for members of the fonticulids (the genome sequence of *Fonticula alba* and one fonticulid metagenome), *Nuclearia* (EST data for *N. pattersoni* and *N. moebiusi*) and, recently, the transcriptome sequence of *Parvularia atlantis*. Thus, most of the nucleariid diversity remained unsampled in terms of genome and/or transcriptome sequencing. The phylogeny of nucleariids has been studied mostly using 18S rRNA gene sequences but, as a result of the low phylogenetic signal of this marker, the relationships between members of this lineage remained unresolved. In addition, for some species such as the cover-bearing nucleariids, there were no molecular data available.

The combined efforts of our team with partner laboratories led to the isolation of samples containing the cover-bearing nucleariids *Lithocolla* (a marine culture) and *Pompholyxophrys* (fresh water lake samples). Additionally, for the traditional naked nucleariids we could get several *Nuclearia* species from both culture collections and from micromanipulated tadpole gut sampled. With all the new collected data we pursued the following objectives:

1) Resolve the relationships of the different clades within the nucleariid clade. We did that by sequencing genomes and transcriptomes of new covered and naked nucleariid species and performing multi-gene phylogenomic analyses of the whole clade. Given the mixed origin of our data, we used both single-cell and culture-based approaches.

2) Additionally, given the fact that we used single-cell and culture-based approaches to the sequencing of both transcriptomic and genomic data from nucleariid species, we wanted to assess which technique performs best and if different approaches can give us different insights for different aspects of these organisms (e.g. ecological aspects).

4.2. Results

Our phylogenomic analyses have shown that the cover-bearing organisms *Lithocolla* and *Pompholyxophrys* belong to the nucleariids, forming a monophyletic clade sister to the *Nuclearia* clade, and both creating a clade sister to the lineage of the small filose amoebas *Parvularia* and *Fonticula*. The reconstruction of phylogenetic relationships among nucleariids also allowed us to infer that the MRCA of nucleariids was most likely a freshwater, bacterivorous, non-flagellated filose mucilaginous amoeba. Finally, even if culture-based approaches appeared to perform better than single-cell techniques, single-cell data also allowed us to carry out robust phylogenomic analyses and to get insights into ecological aspects such as the characterization of bacterial endosymbionts in *Pompholyxophrys*.

4.3. Manuscript of article 1

Combined cultivation and single-cell approaches to the phylogenomics of the nucleariid amoebae, close relatives of Fungi

(Phil. Trans. R. Soc. B 374: 20190094)

Luis Javier Galindo[1*], Guifré Torruella[1*], David Moreira[1], Yana Eglit[2], Alastair G.B. Simpson[2], Eckhard Völcker[3], Steffen Clauß[3], Purificación López-García[1]

[1]Unité d'Ecologie, Systématique et Evolution, CNRS, Université Paris-Sud, Université Paris-Saclay, AgroParisTech, 91400 Orsay, France.

[2]Department of Biology, and Centre for Comparative Genomics and Evolutionary Bioinformatics, Dalhousie University, Halifax, Nova Scotia, Canada.

[3]Penard Laboratory, Cape Town, 7708 South Africa.

*authors contributed equally

Authors for correspondence:

Guifré Torruella

e-mail: guifftc@gmail.com

Purificación López-García

e-mail: puri.lopez@u-psud.fr

PHILOSOPHICAL TRANSACTIONS B

royalsocietypublishing.org/journal/rstb

Research

Cite this article: Galindo LJ, Torruella G, Moreira D, Eglit Y, Simpson AGB, Völcker E, Clauß S, López-García P. 2019 Combined cultivation and single-cell approaches to the phylogenomics of nucleariid amoebae, close relatives of fungi. *Phil. Trans. R. Soc. B* **374**: 20190094.
http://dx.doi.org/10.1098/rstb.2019.0094

Accepted: 26 May 2019

One contribution of 18 to a discussion meeting issue 'Single cell ecology'.

Subject Areas:
evolution, ecology, genomics, microbiology, taxonomy and systematics

Keywords:
nucleariid amoebae, opisthokont, single-cell, *Pompholyxophrys*, *Lithocolla*, Rotosphaerida

Authors for correspondence:
Guifré Torruella
e-mail: guifftc@gmail.com
Purificación López-García
e-mail: puri.lopez@u-psud.fr

†These authors contributed equally to the study.

Electronic supplementary material is available online at https://doi.org/10.6084/m9.figshare.c.4646486.

THE ROYAL SOCIETY PUBLISHING

Combined cultivation and single-cell approaches to the phylogenomics of nucleariid amoebae, close relatives of fungi

Luis Javier Galindo[1,†], Guifré Torruella[1,†], David Moreira[1], Yana Eglit[2], Alastair G. B. Simpson[2], Eckhard Völcker[3], Steffen Clauß[3] and Purificación López-García[1]

[1]Unité d'Ecologie, Systématique et Evolution, CNRS, Université Paris-Sud, Université Paris-Saclay, AgroParisTech, 91400 Orsay, France
[2]Department of Biology, and Centre for Comparative Genomics and Evolutionary Bioinformatics, Dalhousie University, Halifax, Nova Scotia, Canada
[3]Penard Laboratory, Cape Town, 7708 South Africa

LJG, 0000-0003-3642-3200; GT, 0000-0002-6534-4758; DM, 0000-0002-2064-5354; YE, 0000-0002-2178-3062; AGBS, 0000-0002-4133-1709; EV, 0000-0002-6841-1389; PL-G, 0000-0002-0927-0651

Nucleariid amoebae (Opisthokonta) have been known since the nineteenth century but their diversity and evolutionary history remain poorly understood. To overcome this limitation, we have obtained genomic and transcriptomic data from three *Nuclearia*, two *Pompholyxophrys* and one *Lithocolla* species using traditional culturing and single-cell genome (SCG) and single-cell transcriptome amplification methods. The phylogeny of the complete 18S rRNA sequences of *Pompholyxophrys* and *Lithocolla* confirmed their suggested evolutionary relatedness to nucleariid amoebae, although with moderate support for internal splits. SCG amplification techniques also led to the identification of probable bacterial endosymbionts belonging to Chlamydiales and Rickettsiales in *Pompholyxophrys*. To improve the phylogenetic framework of nucleariids, we carried out phylogenomic analyses based on two datasets of, respectively, 264 conserved proteins and 74 single-copy protein domains. We obtained full support for the monophyly of the nucleariid amoebae, which comprise two major clades: (i) *Parvularia–Fonticula* and (ii) *Nuclearia* with the scaled genera *Pompholyxophrys* and *Lithocolla*. Based on these findings, the evolution of some traits of the earliest-diverging lineage of Holomycota can be inferred. Our results suggest that the last common ancestor of nucleariids was a freshwater, bacterivorous, non-flagellated filose and mucilaginous amoeba. From the ancestor, two groups evolved to reach smaller (*Parvularia–Fonticula*) and larger (*Nuclearia* and related scaled genera) cell sizes, leading to different ecological specialization. The *Lithocolla* + *Pompholyxophrys* clade developed exogenous or endogenous cell coverings from a *Nuclearia*-like ancestor.

This article is part of a discussion meeting issue 'Single cell ecology'.

1. Introduction

Nucleariids are non-flagellated, free-living, phagotrophic filose amoebae [1]. 18S rRNA gene molecular phylogenies placed *Nuclearia* as a deep branch within the opisthokonts [2,3], particularly as sister clade to fungi [4,5], as subsequently corroborated by phylogenomic analyses [6,7]. They are thus part of the Holomycota (Nucletmycea), the opisthokont lineage containing fungi and its relatives [8]. The last opisthokont common ancestor probably was a phagotrophic cell with a single flagellum and polarized cell shape, a feature that is shared with the deepest-branching fungi and their aphelid [9] and rozellid [10] relatives [11].

Therefore, nucleariids underwent substantial evolutionary change from that ancestor which we need to understand to infer the global evolutionary history of Holomycota, including key biological traits such as the fungal multicellularity [12] or the transition to parasitism [13].

So far, only a few studies of nucleariid species are available, including some morphological descriptions [1,14–16] and molecular phylogenetic studies [2–5,17–20]. Nevertheless, many *incertae sedis* species await molecular characterization [21–25]. Historically, owing to the lack of clear external features distinguishable under optical microscopy, nucleariids have been assigned to a variety of amoeboid taxa [26,27]. *Nuclearia* Cienkowski, 1865, is the most commonly observed and characterized genus [1,28,29]. Until the late twentieth century, this genus was associated with other naked filose amoebae in several different and conflicting taxonomies [8,30,31]. Patterson, using transmission electron microscopy data, separated nucleariids from other filose amoebae, united distinct genera (e.g. *Nuclearella* Frenzel, 1897; *Nuclearina* Frenzel, 1897, *Nucleosphaerium* Cann and Page, 1979) into *Nuclearia*, clarified its systematics [1,14], and confirmed its relationship with *Vampyrellidium perforans* [16,32] (not to be confused with the cercozoan *Vampyrella* [33]) and the scale-bearing filose amoeba *Pompholyxophrys* [15]. It was further proposed that other silica-scaled amoebae with a secreted silica-mineral coat composed of silicified particles (i.e. idiosomes), like *Pinaciophora* and *Rabdiophrys* (not to be confused with the centrohelid *Raphidiophrys* [34]) were related to *Nuclearia* [19,20,22,25,33]. In agreement with Patterson, Page grouped *Nuclearia* and *Pompholyxophrys* inside the Cristidiscoidida [35]. Later, Mikrjukov suggested that *Elaeorhanis* [36] and *Lithocolla* [37], two scaled filose amoebae with coats composed of aggregated exogenous material (i.e. xenosomes), were also related to nucleariids [22] and claimed priority of the name Rotosphaerida over Cristidiscoidea to group all nucleariid amoebae [38]. Since then, molecular phylogeny analyses have placed *Fonticula* [5,26,39] and *Parvularia* [20,40] together with *Nuclearia* as a sister clade to the rest of Holomycota, although the 18S rDNA gene marker could not resolve the internal relationships between nucleariid clades.

To solve some of these uncertainties, we sampled putative nucleariid species from freshwater and marine environments, including naked (*Nuclearia* sp.) and scale-bearing (*Pompholyxophrys* sp. and *Lithocolla* sp.) amoebae. We obtained molecular data using traditional culturing and single-cell genomic techniques and inferred a robust phylogenetic framework that leads to an improved understanding of the biodiversity of these organisms and a clarification of the systematics of the whole nucleariid clade.

2. Methods

(a) Biological material

Lithocolla globosa (electronic supplementary material, figure S1) was isolated from a marine sediment sample from Splitnose Point near Ketch Harbour, Nova Scotia, Canada (44.477 N, 63.541 W) and grown in culture with *Navicula pseudotenelloides* NAVIC33 as food source. Single *Lithocolla* cells were micromanipulated with an Eppendorf PatchMan NP2 micromanipulator using a 110 µm VacuTip microcapillary (Eppendorf) in an inverted microscope Leica Dlll3000 B, cells were washed in clean water drops before storing them into individual tubes. *Pompholyxophrys* cells (electronic supplementary material, figure S2) were collected from a freshwater lake near Zwönitz, Germany both by manual micromanipulation and by using the previously described equipment into tubes in sets of 20–30 cells or as single cells (without washing steps when manually collected) [41]. Both *Nuclearia delicatula* and *Nuclearia thermophila* (electronic supplementary material, figure S3) were isolated from the mixed freshwater culture JP100 from Sciento (UK) maintained with *Oscillatoria*-like filamentous cyanobacteria, and with the presence of *Poterioochromonas*-like (stramenopile) and *Echinamoeba*-like (amoebozoan) contaminants (electronic supplementary material, figure S3A–D). *Nuclearia thermophila* was isolated by micromanipulation (using previously cited equipment) from the initial JP100 culture. Individual *Nuclearia pattersoni* XT1 cells were collected after washing steps using the previously described micromanipulator equipment from the intestine of a dissected *Xenopus tropicalis* tadpole grown in the laboratory.

(b) DNA and RNA purification, 18S rRNA gene amplification and sequencing

To assess the identity of our nucleariid amoebae, we first obtained 18S rRNA gene sequences from cultures and single-cell isolates by polymerase chain reaction (PCR) amplification using distinct combinations of primers 82F (5′-GAAACTGCGAATGGCTC-3′), 612F (5′-GCAGTTAAAAAGCTCGTAGT-3′), 1379R (5′-TGTGTACAAAGGGCAGGGAC-3′) and 1498R (5′-CACCTACGGAAACCTTGTTA-3′). Amplicon cloning was performed with the TOPO-TA cloning kit (Invitrogen) following the instructions of the manufacturers. RNA was purified from the cultures of *N. delicatula*, *N. thermophila*, the mixed culture of *L. globosa* and its food *Navicula* sp. using the kit RNeasy Micro (Qiagen, Venlo, The Netherlands) including a DNAse treatment. In addition, whole transcriptome amplification (WTA) and whole genome amplification (WGA) of micromanipulated single cells was carried out using REPLI-g WTA/WGA Kits (Qiagen) for *N. pattersoni*, *L. globosa* and *Pompholyxophrys*. For a batch of 20 *Pompholyxophrys* cells, DNA was first released with the PicoPure DNA extraction kit (Applied Biosystems) and then WGA was performed (table 1). Paired-end sequences were obtained by polyA RNAseq or Nextera library construction and sequencing was performed with an Illumina HiSeq SBS Kit v4 2500 2 × 125 bp by Eurofins Genomics (Ebersberg, Germany) or by the Centre Nacional d'Anàlisi Genòmica (CNAG, Barcelona, Spain) for the Nextera libraries.

(c) Molecular data assembly, decontamination and annotation

Reads were screened with FastQC [42] before and after quality/Illumina adapter trimming with Trimmomatic v0.33 [43] in paired-end mode with the following parameters: ILLUMINACLIP:adapters.fasta:2:30:10 LEADING:20 TRAILING:20 SLIDINGWINDOW:4:28. Resulting reads were assembled with SPAdes 3.9.1 [44]. To predict protein sequences, we co-assembled the *L. globosa* dataset and sequences from the two *Pompholyxophrys* species (*P.* sp. and *P. punicea*), after verifying that they belonged to the same species by 18S rRNA gene phylogenetic analyses. Two co-assembly rounds were performed before and after decontamination by BlobTools v0.9.19 [45]. In the case of *Lithocolla*, the predicted *Navicula* proteome was used to further eliminate sequences from its prey using BLASTp [46]. Decontaminated predicted protein sequences were obtained using Transdecoder v2 (http:transdecoder.github.io) with default parameters and Cd-hit v4.6 [47] with 100% identity. Proteins were annotated with the EggNOG v4.5 [48] database with DIAMOND as mapping mode, and the taxonomic scope to adjust automatically (table 1). We have deposited the new nucleariid 18S rRNA gene sequences in GenBank with accession numbers MK547173–MK547179, and *Pompholyxophrys* bacterial endosymbionts 16S rRNA gene

Table 1. List of protist single-cells/culture samples, sequence statistics and number of phylogenetic markers retrieved from genome/transcriptome datasets. (WTA stands for whole transcriptome amplification and WGA for whole genome amplification.)

						individual species assemblies				
cell/culture identifier	DNA/RNA (culture or few/single-cell)	read-pairs	yield (Gb)	GBE 264 markers (%)	SCPD 74 markers (%)	no. of contigs/ scaffolds	no. of proteins	no. of 'clean' proteins	GBE 264 markers (%)	SCPD 74 markers (%)
L. globosa MK547176										
culture SnPLi with *Navicula*	RNAseq (culture)	41 033 000	23 854	199 (75.37)	60	70 737	72 580	9277	211 (79.92)	65 (87.83)
LG140, LG144, LG145	WGA (few-cells)	77 313 319	15 462	35 (13.25)	27					
LG147	WTA (single-cell)	37 212 410	18 681	81 (30.68)	24					
N. pseudotenelloides										
NAVIC33 culture	RNAseq (culture)	44 463 054	13 428	—	—	36 618	3350	—	—	—
Pompholyxophrys sp. MK547174										
LG126	WGA (single-cell)	73 107 816	14 621	3 (1.13)	1 (1.34)	86 851	39 399	1094	82 (31.06)	19 (25.67)
LG130	WTA (single-cell)	37 135 207	18 642	80 (30.3)	18 (24.32)					
P. punicea MK547175										
LG129	WTA (single-cell)	39 500 923	19 829	125 (47.34)	31 (41.89)	227 098	82 091	3121	144 (54.54)	34 (45.94)
20cellsWGA	WGA (few-cells)	47 517 660	23 854	36 (13.63)	9 (12.16)					
LG127	WGA (single-cell)	68 532 623	13 706	0	0	2356	—	—		
N. pattersoni XT1 MK547179										
XT1	WTA (single-cell)	7 062 454	4237	33 (12.5)	0	453 169	41 060	—	—	—
N. delicatula JP100 MK547177										
culture JP100 contaminated with other eukaryotes	RNAseq (culture)	83 127 257	10 390	234 (88.63)	59 (79.72)	56 177	54 191	—	—	—
N. thermophila JP100 MK547178										
culture JP100 cleaned Sep/Nov	RNAseq (culture)	128 552 236	32 139	251 (95.07)	72 (97.29)	70 205	65 150	—	—	—

sequences with accession numbers MK616425–MK616429. Transcriptome and genome sequence data have been submitted to NCBI SRA under the Bioproject PRJNA517920. Decontaminated predicted proteins, phylogenetic datasets and trees have been deposited in Figshare [49].

(d) 18S and 16S rRNA gene phylogenies

We compiled the 18S rRNA gene sequences included in three previous studies of nucleariids, including environmental sequences [20,50,51], and aligned them with our newly obtained sequences. We generated a dataset of 207 sequences and 1756 bp. For bacterial endosymbionts, we used the 16S rRNA gene sequences of *Nuclearia* sp. endosymbionts identified in the previous study [28] as queries to find homologues by BLASTn [46] in all nucleariid assemblies (*Parvularia*, 2 *Nuclearia* and 2 *Fonticula* species). Selected sequences of potential endosymbionts along with their closest BLAST hits were included in phylogenetic trees to have representatives of closely related bacteria. We worked with three datasets, one complete dataset of 100 sequences and 1503 bp, and two subsets of this first dataset for the Chlamydiae group (18 sequences and 1454 bp) and the Rickettsiales group (26 sequences and 1390 bp). All alignments were made using MAFFT v7 [52]. Trimming of the alignment was performed manually for the 18S rRNA gene sequences and with TrimAl in automated1 mode [53] for the 16S rRNA gene sequences.

Maximum-likelihood (ML) phylogenetic trees were inferred using IQTree v1.6 [54]. For the 18S rRNA gene ML trees, the GTR + R8 + F0 evolutionary model was used to assess branch support with 1000 ultrafast bootstraps (UFBS), single branch tests SH-like approximate likelihood ratio test based on the Shimodaira-Hasegawa (SH) algorithm for tree comparison [55] and approximate Bayes test [56]. In addition, 1000 non-parametric bootstraps [57] were obtained with the TIM3 + F + I + G4 model as the best-fitting one based on the Bayesian information criterion (BIC) from ModelFinder [58]. For the 16S rRNA gene ML trees, the best fit model chosen by BIC [59] was the GTR model (for the complete dataset and for the Rickettsiales dataset) and the TIM3 model (for the Chlamydiae dataset) both with F + I + G4. Bayesian inference (BI) phylogenies were inferred using MrBayes v3.2.6 [60]. For both the 16S and 18S rRNA gene BI trees, the GTR + G + I model was used, with four Markov chain Monte Carlo (MCMC) chains for 1 000 000 generations, sampling every 100 trees and burn-in of the first 2500 saved trees.

(e) Phylogenomic analyses

Two distinct datasets, a dataset modified from Mikhailov *et al*. [9,61] (dataset GBE: 264 protein alignments) and Torruella *et al*. [9] (dataset SCPD: 74 single-copy domains) were updated with data from seven new nucleariid species. For both datasets, orthologues were identified by tBLASTn, aligned with MAFFT v7 and trimmed with TrimAl with the automated1 option. Alignments were visualized and manually edited with Geneious v6.0.6 and single gene trees obtained with FastTree v2.1.7 [62] with default parameters. Single gene trees were then manually checked and corrected for paralogous and/or contaminating sequences. All datasets were assembled into a supermatrix with Alvert.py from the package Barrel-o-Monkeys [63]. Resulting matrices were called SCPD21_23481aa and GBE22_97918aa. No orthologous markers were retrieved for *N. pattersoni* XT1 in the SCPD dataset. For both datasets, BI phylogenetic trees were reconstructed using PhyloBayes-MPI v1.5 [64] under the CAT-Poisson model, two MCMC chains for each dataset were run for greater than 15 000 generations, saving one every 10 trees. Analyses were stopped once convergence thresholds were reached after a burn-in of 25% (i.e. maximum discrepancy less than 0.1 and minimum effective size greater than 100 calculated using bpcomp). ML phylogenetic trees were inferred with IQ-Tree v1.6 under the LG + R5 + C60 model. Statistical support was obtained with 1000 UFBS [65] and 1000 replicates of the SH-like approximate likelihood ratio test [56]. All trees were visualized with FigTree [66].

Fully detailed materials and methods can be found in the electronic supplementary material.

3. Results and discussion

(a) *Pompholyxophrys* and *Lithocolla* are free-living nucleariid amoebae

We obtained 18S rRNA gene sequences from two cultures of *Nuclearia* (*N. delicatula* JP100 and *N. thermophila* JP100), one single cell from another *Nuclearia* species (*N. pattersoni* XT1), two single cells and one few cells (20 cells) from *Pompholyxophrys* species and one culture of *L. globosa* (table 1 and the electronic supplementary material). This represents the first molecular data for both *Pompholyxophrys* and *Lithocolla*. We included our new sequences in a large 18S rRNA gene dataset containing all available nucleariid sequences. Phylogenetic analyses of this dataset confirmed the monophyly of *Nuclearia* species and their relationship with the environmental sister clade NUC-1, whereas the environmental clade NUC-2 was sister to the *Parvularia* clade (figure 1 and electronic supplementary material, figure S4A–C) [20]. *Fonticula alba* exhibited a long branch sister to the group containing the *Pompholyxophrys* and *Lithocolla* sequences. This group also contained several environmental sequences originally called marine fonticulids [19] but recent metabarcoding studies [45,46] have found freshwater representatives intermixed with the marine ones. The morphology and behaviour of *Lithocolla* cells in culture strongly resemble *Nuclearia* (electronic supplementary material, figure S1). Also its exogenous aggregative cell covering suggests a higher similarity to naked *Nuclearia* than to *Pompholyxophrys* [22]. However, our results support a closer phylogenetic relationship of *Pompholyxophrys* and *Lithocolla* as compared to *Nuclearia* (figure 1). Nevertheless, the internal topology of this large Pompholyxophrys–Lithocolla group, which additionally encompasses two large clades of environmental sequences (with currently not known representative species), remains unclear. This is probably owing to the limited signal of the 18S rRNA marker at this level of resolution.

Although some *Nuclearia* have been found in brackish water [1], all published environmental sequences clustering with *Nuclearia* come from soil or freshwater systems (as deduced from sequence metadata deposited in GenBank) and *Parvularia*, as *Nuclearia*, seems to be exclusively freshwater. *Pompholyxophrys* has also been found only in freshwater systems [15,22] but it is sister to a clade of marine environmental sequences (figure 1 and electronic supplementary material, figure S4A–C). Although our *Lithocolla* sequence clustered within an exclusively marine clade, this genus has been observed also in freshwater environments [37].

Nuclearia species are capable of growing in eutrophic and/or contaminated environments. For example, they can ingest toxic filamentous cyanobacteria that can thrive in perturbed environments as their sole food source [29,41]. This capability appears to be related to their association with symbiotic bacteria that degrade toxic metabolites, as microcystin, contained in the cyanobacteria ingested by *Nuclearia* [28,29,67]. Our *N. pattersoni* single cell was recovered by micromanipulation from the gut content of a dissected *X. tropicalis* tadpole grown in the laboratory. When collected, this cell was alive and moving, suggesting that it was a commensal in the amphibian gut. In agreement

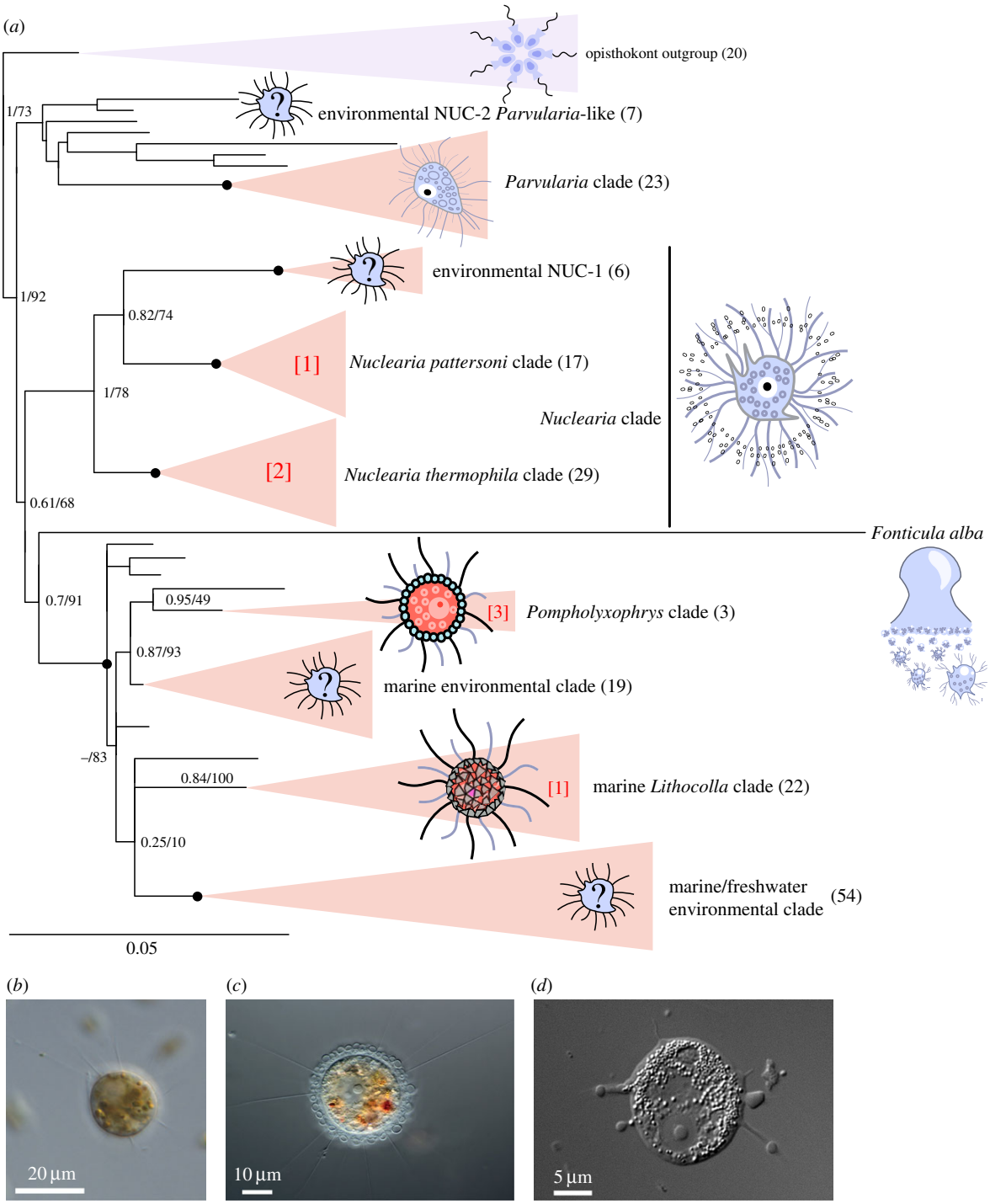


Figure 1. (*a*) ML phylogenetic tree of nucleariid 18S rRNA gene sequences. The tree was reconstructed from an alignment of 1756 nucleotide positions of 207 sequences, including the three *Nucleariа*, three *Pompholyxophrys* and one *L. globosa* sequences obtained in this study as well as all nucleariid sequences available in GenBank with the GTR + R8 model. Major groups were collapsed (the complete tree is shown in the electronic supplementary material, figure S4A). Statistical supports are Bayesian posterior probabilities (PP) obtained under the GTR + G + I model on the left and ML ultrafast bootstrap (UFBS) on the right. Branches with support values higher or equal to 0.99 PP and 95% UFBS are indicated by black dots. Clades without known representatives are indicated with a question mark. The number of sequences is shown in parenthesis and the number of sequences obtained in this study is shown in red brackets. (*b–d*) From left to right optical microscopy images of *L. globosa*, *P. punicea* and *N. thermophila* JP100. Scale bars are 20, 10 and 5 µm, respectively. (Online version in colour.)

with this idea, *N. pattersoni* was originally described from fish gills [17]. Whether *Nucleariа* maintains preferential ecological interactions with metazoans or not remains to be determined. By contrast, multiple observations suggest that *Pompholyxophrys* species, as many other silica-based scale-bearing amoebae, are free-living and develop in clear freshwater bodies, wet *Sphagnum* moss, and peat bogs [68,69].

(b) Endosymbiotic bacteria in nucleariids

Single-cell approaches allowed us to examine an important ecological aspect of these amoebae, namely their relationships with intracellular bacteria. Bacterial endosymbionts have been previously observed in nucleariids [31], with the first molecular data coming from a *Rickettsia* endosymbiont in *N. pattersoni* [17] and the gammaproteobacterium *Candidatus*

Endonucleobacter rarus in *N. thermophila* [67]. Dirren and Posch [28] characterized several bacterial endosymbionts in different species and strains of *N. thermophila* and *N. delicatula*. They observed that the specificity of the symbiosis might vary depending on the host *Nuclearia* species. In some cases, the same endosymbiont species was found in the same host (*N. thermophila*) from different places, but in other cases, the same host (*N. delicatula*) may harbour different endosymbionts.

We generated four single/few-cell transcriptomes (SCT) and four single/few-cells genomes (SCG) for *Lithocolla*, *Pompholyxophrys* and *Nuclearia* (table 1), and using as a reference the bacterial endosymbiont 16S rRNA gene dataset from Dirren and Posch [28], we searched for endosymbiotic candidate species. However, we not only searched in our SCTs/SCGs but also in our RNAseq data and in all other nucleariid data available in public databases (*Parvularia*, two *Fonticula* species and two *Nuclearia* species).

We retrieved 13 bacterial 16S rRNA gene sequences, five of which branched together with well-known bacterial intracellular lineages (figure 2; electronic supplementary material, figure S5). These sequences were only found in the *Pompholyxophrys* assemblies, including two SCTs (*Pompholyxophrys* LG126 and LG127) and one SCG from *P. punicea* (20-cells WGA).

One of these bacterial sequences (*Pompholyxophrys* sp. LG126 (2)) branched within the Chlamydiae (figure 2*a*), along with sequences of known bacterial endosymbionts of the amoebae *Acanthamoeba* sp. and *Hartmannella vermiformis*. The other four sequences branched within the Rickettsiales (figure 2*b*). *Pompholyxophrys punicea* LG127 seemed to harbour two different *Rickettsia*-like endosymbionts. One of them, LG127 (1), branched within a clade of *Rickettsia* species endosymbionts of different hosts, including metazoans and, interestingly, *N. pattersoni* [70]. The second sequence LG127 (2) and a second sequence from *Pompholyxophrys* sp. LG126 (1) were identical. The last endosymbiont candidate sequence came from the *P. punicea* 20-cells WGA assembly and, although clearly branching within the Rickettsiales, had no close relatives. Thus, the same endosymbiont can be found in different cells from the same natural sample, as in the case of *Pompholyxophrys* sp. LG126 (1) and LG127 (2). Conversely, different endosymbionts can coexist in the same cell as well, as seen in *P. punicea* LG127 (1 and 2), in this case belonging to the same bacterial clade of Rickettsiales. A single cell can also harbour endosymbionts from phylogenetically distant groups as seen in *Pompholyxophrys* sp. LG126 (1 and 2), containing representatives of Chlamydiae and Rickettsiales (figure 2).

Our results are consistent with the findings of Dirren and Posch [28], showing that symbiont acquisition in nucleariids seems to be rather promiscuous. It is also worth noting that we only found endosymbiont sequences in the *Pompholyxophrys* assemblies. We could not recover any bacterial sequence from our *Nuclearia* assemblies, even though we have worked with the same *Nuclearia* species studied by Dirren and Posch [28]. However, because we only analyzed with *Nuclearia* transcriptome sequences, we cannot completely discard the presence of endosymbionts.

(c) Phylogenomic analyses

To establish a solid phylogenetic framework for nucleariids, and because the 18S rRNA gene has limited resolution power, we generated genome and transcriptome data for several nucleariids (table 1). Although the percentage of orthologue gene markers recovered for the two datasets was low (especially for *Pompholyxophrys* assemblies) (table 1), we could retrieve a sufficient number of gene marker sequences from our new assemblies for three *Nuclearia* species, two *Pompholyxophrys* species and *Lithocolla* (table 1). We also used publicly available data from two *Nuclearia* species [7], two *Fonticula* species [5,71] and *Parvularia atlantis* [20], adding representative members of other opisthokont lineages as outgroup. With these sequence datasets, we updated two datasets of conserved phylogenetic markers previously used to study the phylogeny of holomycotan clades [9,61]: the GBE dataset (264 proteins) and the SCPD dataset (74 single-copy protein domains—without *N. pattersoni* XT1 as no gene markers were retrieved for this species) (electronic supplementary material, figures S6A–D). As in the 18S rRNA gene phylogeny, all previously recognized nucleariids (*Nuclearia*, *Fonticula* and *Parvularia*) clustered together with *Lithocolla* and the two *Pompholyxophrys* species with maximum support in ML and BI analyses for both datasets, forming a sister clade to other Holomycota (figure 3). However, the relationships between the different genera were not the same as in the 18S rRNA gene tree, in particular regarding the placement of *Fonticula*. *Fonticula* appeared as a long branch sister clade to *Lithocolla* and *Pompholyxophrys* (with low statistical support) in the 18S rRNA gene tree (figure 1). However, in the phylogenomic analyses, the two *Fonticula* species clustered with *Parvularia* with high statistical support (figure 3). All the five *Nuclearia* species (with the same internal topology as in the 18S rRNA gene tree) clustered with *Lithocolla* and the two *Pompholyxophrys*. Thus, two separated clades formed, one containing all *Nuclearia* species and one containing the scale-bearing *Pompholyxophrys* and *Lithocolla*, both with maximum support values.

(d) Evolutionary implications

Our robust phylogenomic tree of nucleariids allows us to discuss the evolutionary history of several nucleariid characters, although molecular data are still missing for genera putatively related to nucleariids, such as *Vampyrellidium*, *Pinaciophora*, *Elaeorhanis* or *Rabdiophrys* (see the electronic supplementary material for detailed taxonomical discussion).

The last common ancestor of opisthokonts was probably phagotrophic with amoeboid polarized cell shape and a single flagellum, features that can be found in extant examples such as choanoflagellates [72], pigoraptors [73] or aphelids [9]. All known nucleariids lack flagella, suggesting that the last common ancestor of all nucleariids had already lost the ancestral flagellum. It is also worth mentioning that the nucleariid ancestor probably originated in freshwater environments, as suggested by the 18S rRNA gene tree analysis in which all the basal branches (including environmental clades) are occupied by freshwater lineages. The non-polarized and plastic cell shape surrounded by hyaline pseudopodia (branching filopodia) of nucleariids seems concomitant with the loss of flagella. Although there are few studies on nucleariid biology, cell movement by 'walking' on the benthos [29] and planktonic stages with equally radiating filopodia (electronic supplementary material, figures S1–S3) arise as major common features of nucleariids, together with a mucilaginous coat involved in different functions (from encystation to encapsulation of ectosymbionts or scales [1,14,29,67]. Although the current knowledge about this group is limited, we can already speculate about evolutionary patterns regarding cell size, food

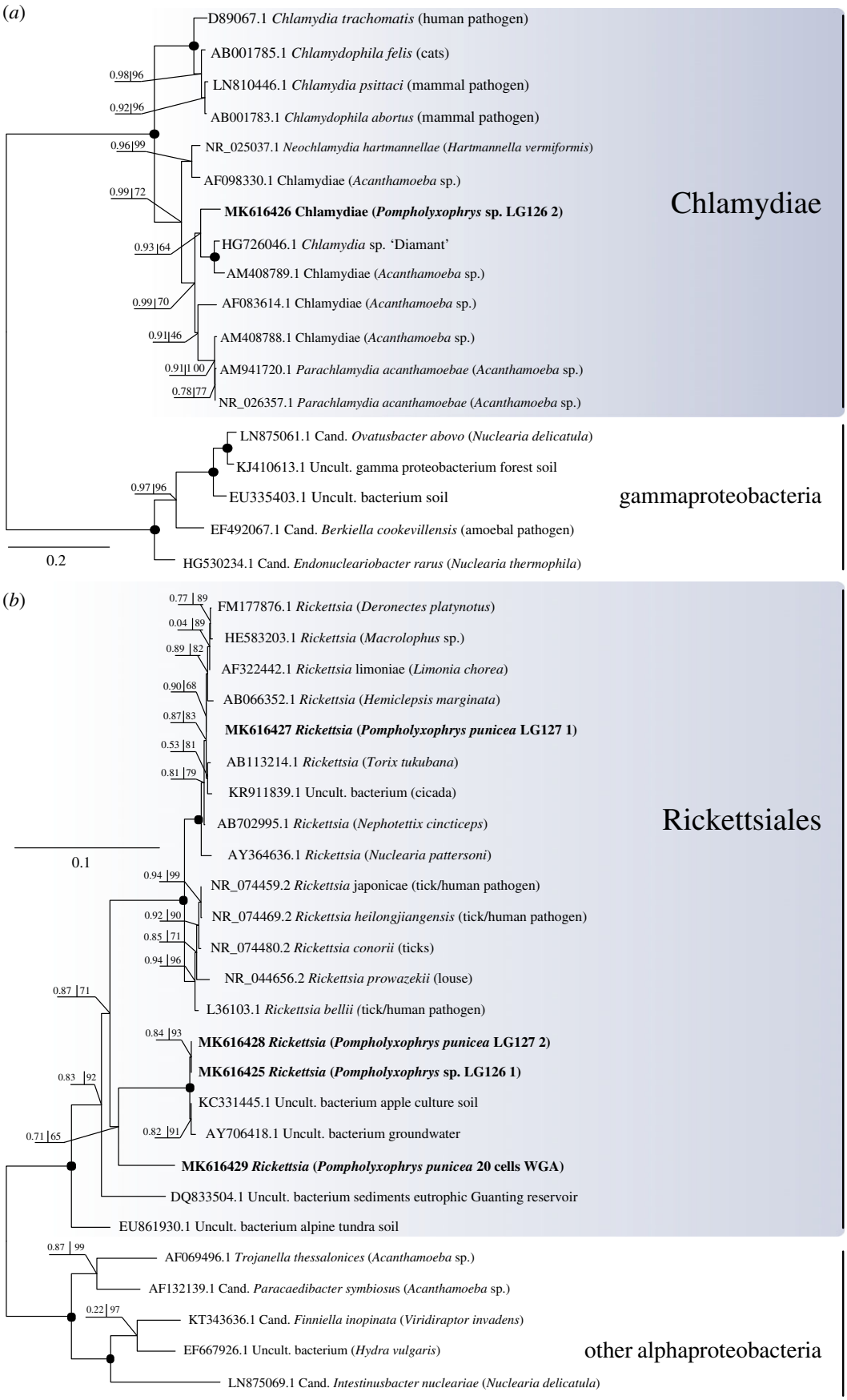


Figure 2. ML phylogenetic tree of 16S rRNA genes showing likely nucleariid bacterial endosymbionts (in bold). (*a*) Chlamydiae tree including one sequence from *Pompholyxophrys* sp. LG126 (2) and inferred under the TIM3 + F + I + G4 model using 1454 conserved nucleotide positions. (*b*) Rickettsiales tree including four sequences obtained in this study and inferred under the GTR + F + I + G4 model using 1390 conserved nucleotide positions. Statistical supports shown are Bayesian PP obtained under GTR + G + I on the left and ML UFBS on the right. Endosymbiont hosts are indicated in parenthesis.

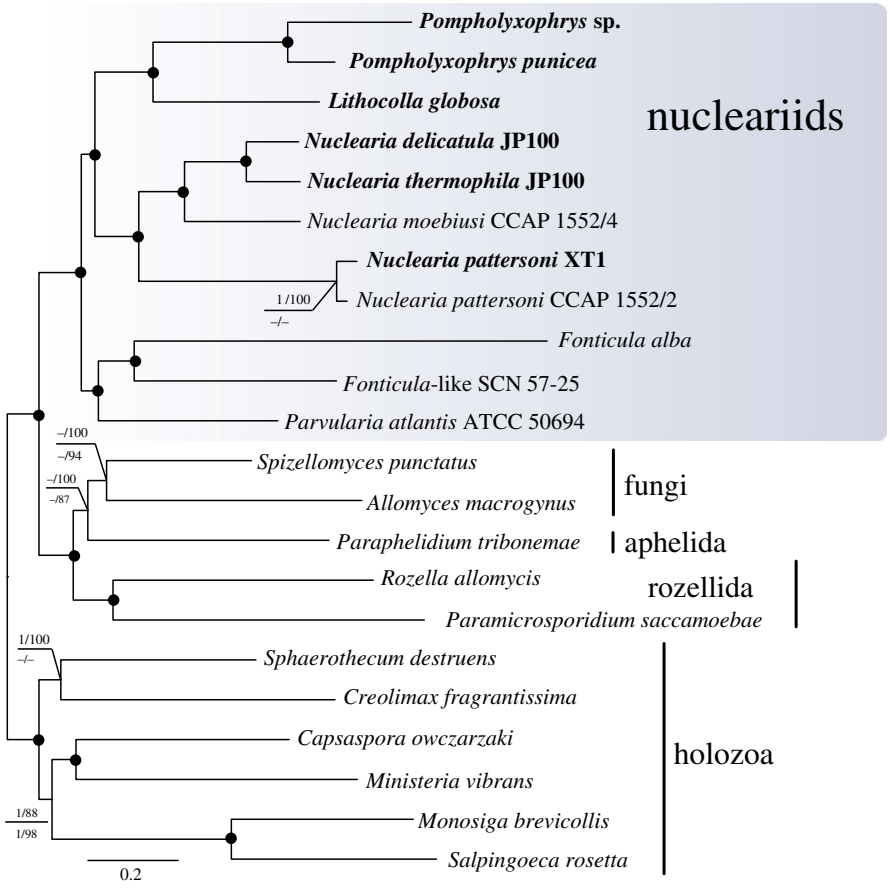


Figure 3. ML phylogenomic tree based on the GBE protein dataset. The tree was reconstructed using 264 conserved proteins, 22 species and 96 276 conserved amino acid positions with the LG + R5 + C60 model. Upper values correspond to supports obtained from the GBE dataset and lower values to those obtained from the single-copy protein domain (SCPD21; without *N. pattersoni* XT1) dataset. Bayesian PP under the CAT-Poisson model are shown on the left and ML UFBS supports are shown on the right. Branches with support values higher or equal to 0.99 PP and 95% UFBS are indicated by black dots. Species names in bold correspond to those for which we have obtained transcriptome and/or genome sequences in this study. (Online version in colour.)

source, ecological niche and cell-coverings (figure 4). From the last common nucleariid ancestor, two clades evolved, one characterized by smaller cells (*Parvularia–Fonticula*) and other with larger cells (*Nucleria* and scaled nucleariids). These differential cell sizes correlate with different ecological specializations in terms of prey and lifestyle. *Parvularia* and *Fonticula* are both exclusively bacterivorous and part of nanoplankton, the first never reaches more than 6 µm [20] and the latter no more than 12 µm [5,39] in size. *Fonticula alba*, which seems to evolve faster than other nucleariids (see branch lengths in figures 1 and 2), grows better in agar plates than in liquid medium (D. López-Escardó 2017, personal communication), and uses its mucilaginous coat to aggregate cells and form fruiting bodies [74]. Hence, *F. alba* looks more adapted to soil environments than to the water column preferred by other nucleariids. Although *Parvularia* and *Nucleria* share many common features (justifying the original identification of *Parvularia* as a nucleariid [20]) *Nucleria* cells are much bigger (from approximately 10 up to 60 µm, depending on the life stage and culture conditions [28]; electronic supplementary material, figure S3). *Lithocolla* (electronic supplementary material, figure S1) and *Pompholyxophrys* (electronic supplementary material, figure S2) range from 20 to 45 µm [15,23,37]. This microplanktonic (greater than 20 µm) size allows them to feed on filamentous cyanobacteria, algae or even other eukaryotes. Finally, *Fonticula*, *Parvularia* and *Nucleria* seem very plastic in terms of cell shape, being round, amorphous or extremely elongated. However, cells became less polymorphic in the genera that acquired the capacity to cover themselves either with xenosomes (probably as a by-product of phagocytosis), as in *Lithocolla* and maybe *Elaeorhanis* [27] (electronic supplementary material, figure S1, [41]), or with idiosomes, as in *Pompholyxophrys* (electronic supplementary material, figure S2) and maybe *Pinaciophora* [25].

Despite these evolutionary implications, deciphering the evolutionary history of nucleariids will require additional data. Indeed, although nucleariids are a pivotal group at the onset of the Holomycota divergence, they remain an under-sampled group, as suggested by environmental data and the many described and likely related species that still lack molecular data. As most nucleariids lack cultured representatives in the laboratory, single-cell techniques will be an invaluable tool to expand the known diversity of uncultured nucleariids, helping to reconcile genomic information with morphology and ecological features, including the presence and role of ecto- and endo-symbiotic bacteria.

(e) Culturing versus single-cell genomes/transcriptomes

In this study, we have used a combination of single-cell techniques (including steps of whole genome/transcriptome amplification) and whole RNA extraction from cultured

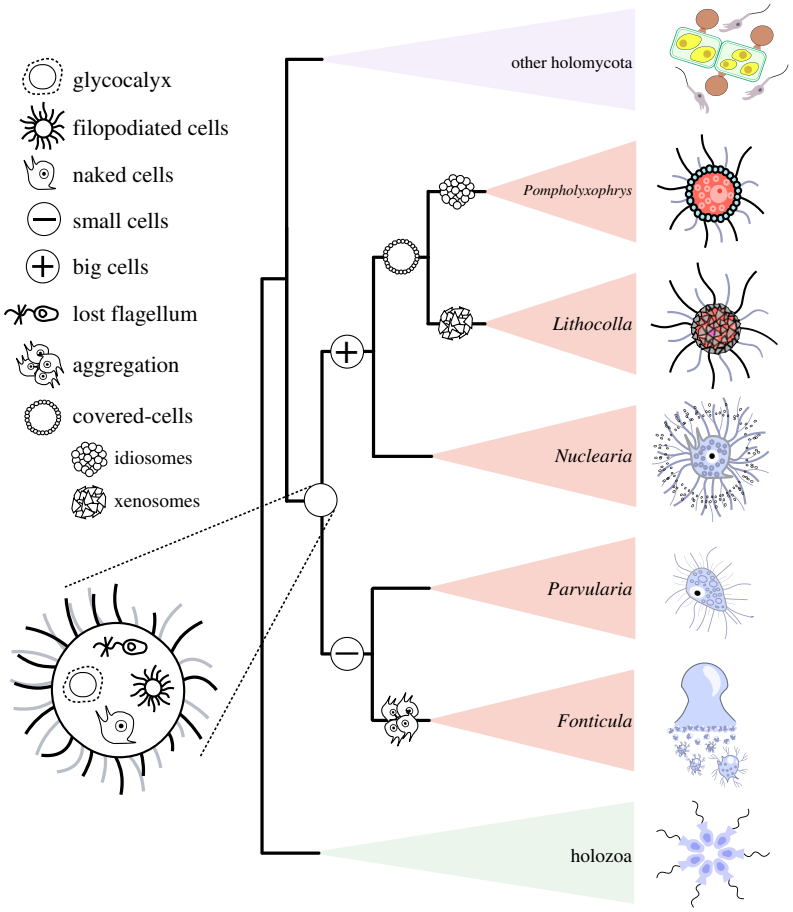


Figure 4. Schematic opisthokont phylogeny displaying cellular characteristics of nucleariids (cell size, presence/absence of cell-cover and its nature, lack of flagellated stages, filopodia, and the presence of a glycocalyx and the capacity to aggregate) and their probable ancestral states in some nodes. (Online version in colour.)

material (without amplification steps) to sequence genomic material from several nucleariid species. Our SCGs/SCTs obtained after WGA/WTA steps produced different results when blasted against our biggest and most complete multigene dataset (GBE). In the case of *Lithocolla*, we obtained two single-cell assemblies, one from a few-cells genome amplification (SCG; LG140,144,145) and one from an SCT (SCT; LG147), recovering 30.68% and 13.25% of the GBE dataset proteins, respectively. The SCTs outperformed the SCGs in *Lithocolla*. In comparison, we recovered 75.37% of the proteins when performing traditional whole RNA extraction and sequencing from a culture.

In the case of the two *Pompholyxophrys* species, we only could obtain single/few-cell genomes/transcriptomes, because no cultures were available. Our *Pompholyxophrys* assemblies displayed different protein recovery percentages ranging from 30 to 47% for the SCTs (LG130 and LG129) and 0 to 13.63% for the SCGs and few-cells genome (LG127, LG126 and 20cellsWGA). Again, the SCTs seemed to outperform the SCGs in terms of protein recovery in this particular case.

Both SCGs/SCTs proved to be useful to obtain enough data to place *Lithocolla* and *Pompholyxophrys* in our multigene phylogeny with strong support. It also allowed us to unveil the hidden diversity in the group, because what initially we thought to be a single *Pompholyxophrys* species were actually two different species (*Pompholyxophrys* sp. and *P. punicea*) as revealed by both 18S rRNA gene and multigene trees.

Nevertheless, not surprisingly, the best results were obtained after RNA extraction of cultures, e.g. *Lithocolla* (75.37%), a result that we confirmed for *N. delicatula* JP100 and *N. thermophila* JP100, for which we recover 88.63% and 95.07% of the dataset proteins, respectively. Culturing approaches, if achievable, remain the best strategies to produce a high amount of high-quality data. However, most protist species are not easily amenable to culture. Therefore, single-cell 'omics', although still far from allowing high or even levels of completeness often allow, as in this particular study, retrieving enough conserved markers to run robust phylogenomic analyses. Further progress in single-cell approaches leading to the retrieval of higher and more homogeneous coverages will hopefully allow more in-depth comparative genomics and population genomics of protists directly sampled from natural communities.

Data accessibility. 18S and rRNA gene sequences have been deposited in GenBank with accession nos. MK547173–MK547179 and MK616425–MK616429, respectively. Transcriptome and genome sequence data have been submitted to NCBI SRA under the Bioproject PRJNA517920.

Authors' contributions. L.J.G., G.T., D.M. and P.L.-G. conceived, coordinated the study and wrote the manuscript. G.T. micromanipulated and obtained *Nuclearia*, *Lithocolla* and *Navicula* RNA. L.J.G. micromanipulated and obtained *Lithocolla* and *Pompholyxophrys* DNA and RNA. D.M. micromanipulated and amplified *N. pattersoni* RNA. S.C. collected freshwater samples and micromanipulated *Pompholyxophrys* cells. Y.E. isolated, cultured and characterized *Lithocolla*. E.V. identified and obtained images from *Pompholyxophrys*. G.T., L.J.G.

and D.M. reconstructed 18S and 16S rRNA gene phylogenies. G.T. and L.J.G. assembled genome and transcriptome sequences, cleaned the assemblies, performed phylogenomic analyses and contributed equally to this work. All authors gave final approval for publication.

Competing interests. We have no competing interests to declare.

Funding. This work was funded by the European Research Council Advanced Grant 'ProtistWorld' (grant no. 322669) and the Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie ITN project SINGEK (http://www.singek.eu/; grant agreement no. H2020-MSCA-ITN-2015-675752). G.T. was financed by the European Marie Sklodowska-Curie Action (704566 AlgDates).

Acknowledgements. We thank David López-Escardó for help with *Nuclearia* molecular identification and advice on duplex PCR, John O'Brien for sample collection for *Lithocolla*, Hélène Timpano and the UNICELL platform for help with single-cell methods, Giselle Walker for help with an unpublished taxonomy of eukaryotes, and the website repositories Microworld, Protist information server, The World of Protozoa and Biodiversity Heritage Library for access to valuable information.

References

1. Patterson DJ. 1984 The genus *Nuclearia* (Sarcodina, Filosea): species composition and characteristics of the taxa. *Arch. Mikrobiol.* **128**, 127–139. (doi:10.1016/s0003-9365(84)80034-2)
2. Zettler L, Nerad TA, O'Kelly CJ, Sogin ML. 2001 The nucleariid amoebae: more protists at the animal-fungal boundary. *J. Eukaryot. Microbiol.* **48**, 293–297. (doi:10.1111/j.1550-7408.2001.tb00317.x)
3. 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. (doi:10.1017/S1473550403001551)
4. Steenkamp ET, Wright J, Baldauf SL. 2006 The protistan origins of animals and fungi. *Mol. Biol. Evol.* **23**, 93–106. (doi:10.1093/molbev/msj011)
5. 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. (doi:10.1093/molbev/msp185)
6. 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. (doi:10.1371/journal.pone.0002098)
7. 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**, 1–11. (doi:10.1186/1471-2148-9-272)
8. Adl SM *et al.* 2018 Revisions to the classification, nomenclature, and diversity of eukaryotes. *J. Eukaryot. Microbiol.* **66**, 4–119. (doi:10.1111/jeu.12691)
9. Torruella G, Grau-Bové X, Moreira D, Karpov SA, Burns JA, Sebé-Pedrós A, Völcker E, López-García P. 2018 Global transcriptome analysis of the aphelid *Paraphelidium tribonemae* supports the phagotrophic origin of fungi. *Commun. Biol.* **1**, 231. (doi:10.1038/s42003-018-0235-z)
10. 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**, 1548–1553. (doi:10.1016/j.cub.2013.06.057)
11. 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. (doi:10.3389/fmicb.2014.00112)
12. Nagy LG, Kovács GM, Krizsán K. 2018 Complex multicellularity in fungi: evolutionary convergence, single origin, or both? *Biol. Rev.* **93**, 1778–1794. (doi:10.1111/brv.12418)
13. Magallón S *et al.* 2018 Contemporaneous radiations of fungi and plants linked to symbiosis. *Nat. Commun.* **9**, 1–11. (doi:10.1038/s41467-018-07849-9)
14. Patterson DJ. 1983 On the organization of the naked filose amoeba, *Nuclearia moebiusi* Frenzel, 1897 (Sarcodina, Filosea) and its implications 1. *J. Protozool.* **30**, 301–307. (doi:10.1111/j.1550-7408.1983.tb02920.x)
15. 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**, 241–246. (doi:10.1111/j.1550-7408.1985.tb03044.x)
16. Patterson DJ, Surek B, Melkonian M. 1987 The ultrastructure of *Vampyrellidium perforans* Surek & Melkonian and its taxonomic position among the naked filose Amoebae1. *J. Protozool.* **34**, 63–67. (doi:10.1111/j.1550-7408.1987.tb03133.x)
17. 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.* (*Praha*). **50**, 161–170. (doi:10.14411/fp.2003.030)
18. 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**, 147–155. (doi:10.1016/j.ejop.2008.09.004)
19. del Campo J, Mallo D, Massana R, de Vargas C, Richards TA, Ruiz-Trillo I. 2015 Diversity and distribution of unicellular opisthokonts along the European coast analysed using high-throughput sequencing. *Environ. Microbiol.* **17**, 3195–3207. (doi:10.1111/1462-2920.12759)
20. López-Escardó D, López-García P, Moreira D, Ruiz-Trillo I, Torruella G. 2017 *Parvularia atlantis* Gen. et Sp. Nov., a nucleariid filose amoeba (Holomycota, Opisthokonta). *J. Eukaryot. Microbiol.* **38**, 42–49. (doi:10.1111/jeu.12450)
21. Thomsen HA. 1978 On the identity of the Heliozoan *Pinaciophora fluviatilis* and *Potamodiscus kalbei*; with the description of eight new *Pinaciophora* species. *Protistologica* **14**, 359–373. (doi:10.1016/s0003-9365(78)80027-x)
22. Mikrjukov KA. 1999 Taxonomic revision of scale-bearing Heliozoon-like amoebae (Pompholyxophryidae, Rotosphaerida). *Acta Potozool.* **38**, 119–131.
23. Nicholls KH, Durrschmidt M. 1985 Scale structure and taxonomy of some species of *Raphidocystis*, *Raphidiophrys*, and *Pompholyxophrys* (Heliozoea) including descriptions of six new taxa. *Can. J. Zool.* **63**, 1944–1961. (doi:10.1139/z85-288)
24. Roijackers RMM, Siemensma FJ. 1988 A study of cristidiscoidid amoebae (Rhizopoda, Filosea), with descriptions of new species and keys to genera and species. *Arch. fur Protistenkd.* **135**, 237–253. (doi:10.1016/S0003-9365(88)80072-1)
25. Nicholls KH. 2012 New and little-known marine species of *Pinaciophora*, *Rabdiaster* and *Thomseniophora* gen. nov. (Rotosphaerida: Pompholyxophryidae). *J. Mar. Biol. Assoc. United Kingdom* **93**, 1211–1229. (doi:10.1017/S002531541200135X)
26. Cavalier-Smith T. 1993 Kingdom protozoa and its 18 phyla. *Microbiol. Rev.* **57**, 953–994.
27. Patterson DJ, Simpson AGB, Rogerson A. 2000 Amoebae of uncertain affinities. In *An illustrated guide to the protozoa, 2nd edn*, vol. II (eds JJ Lee, GF Leedale, P Bradbury), pp. 804–827. Lawrence, KS: Society of Protozoologists.
28. Dirren S, Posch T. 2016 Promiscuous and specific bacterial symbiont acquisition in the amoeboid genus *Nuclearia* (Opisthokonta). *FEMS Microbiol. Ecol.* **92**, 1–16. (doi:10.1093/femsec/fiw105)
29. Dirren S, Pitsch G, Silva MOD, Posch T. 2017 Grazing of *Nuclearia thermophila* and *Nuclearia delicatula* (Nucleariidae, Opisthokonta) on the toxic cyanobacterium *Planktothrix rubescens*. *Eur. J. Protistol.* **60**, 87–101. (doi:10.1016/j.ejop.2017.05.009)
30. Cienkowski L. 1865 Beiträge zur Kenntnis der Monaden. *Arch. für mikroskopische Anat.* **1**, 203–232. (doi:10.1007/BF02961414)
31. Cann JP, Page FC. 1979 *Nucleosphaerium tuckery* nov. ge. nov. sp—a new freshwater filose amoeba without motile form in a new family Nucleariidae (Filosea: Aconchulinida) feeding by ingestion only. *Arch. für Protistenkd.* **122**, 226–240. (doi:10.1016/S0003-9365(79)80034-2)

32. Surek B, Melkonian M. 1980 The filose amoeba *Vampyrellidium perforans* nov. sp. (Vampyrellidae, Aconchulinida): axenic culture, feeding behaviour and host range specifity. *Arch. für Protistenkd.* **123**, 166–191. (doi:10.1016/S0003-9365(80)80003-0)
33. Hess S, Sausen N, Melkonian M. 2012 Shedding light on vampires: the phylogeny of vampyrellid amoebae revisited. *PLoS ONE* **7**, e31165. (doi:10.1371/journal.pone.0031165)
34. Burki F, Kaplan M, Tikhonenkov DV, Zlatogursky V, Minh BQ, Radaykina LV, Smirnov A, Mylnikov AP, Keeling PJ. 2016 Untangling the early diversification of eukaryotes: a phylogenomic study of the evolutionary origins of Centrohelida, Haptophyta and Cryptista. *Proc. R. Soc. B Biol. Sci.* **283**, 20152802. (doi:10.1098/rspb.2015.2802)
35. Page FC. 1987 The classification of 'Naked' Amoebae (Phylum Rhizopoda). *Arch. für Protistenkd.* **133**, 199–217. (doi:10.1016/S0003-9365(87)80053-2)
36. Schulze FE. 1874 Rhizopodenstudien II. *Arch. für mikroskopische Anat.* **10**, 377–400. (doi:10.1007/BF02960329)
37. Penard E. 1904 *Les héliozoaires d'eau douce*. Genève: H. Kündig. See https://www.biodiversitylibrary.org/item/16282.
38. Rainer H. 1968 Urtiere, Protozoa—Wurzelfuessler, Rhizopoda—Sonnentierchen, Heliozoa–Systematik und Taxonomie, Biologie, Verbreitung und Ökologie der Arten der Erde. In *Die Tierwelt Deutschlands und der angrenzenden Meeresteile nach ihren Merkmalen und nach ihrer Lebensweise* (eds M Dahl, F Peus), p. 56. Jena, Germany: Gustav Fischer Verlag.
39. Worley AC, Raper KB, Hohl M. 1979 *Fonticula alba*: a new cellular slime mold (Acrasiomycetes). *Mycologia* **71**, 746–760. (doi:10.1080/00275514.1979.12021068)
40. Torruella G *et al*. 2015 Phylogenomics reveals convergent evolution of lifestyles in close relatives of animals and fungi. *Curr. Biol.* **25**, 2404–2410. (doi:10.1016/j.cub.2015.07.053)
41. Cann JP. 1986 The feeding behavior and structure of *Nuclearia delicatula*. *J. Protozool.* **33**, 392–396. (doi:10.1111/j.1550-7408.1986.tb05629.x)
42. Andrews S. 2010 FastQC: a quality control tool for high throughput sequence data. See https://www.bioinformatics.babraham.ac.uk/projects/fastqc/.
43. Bolger AM, Lohse M, Usadel B. 2014 Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120. (doi:10.1093/bioinformatics/btu170)
44. Nurk S *et al*. 2013 Assembling single-cell genomes and mini-metagenomes from chimeric MDA products. *J. Comput. Biol.* **20**, 714–737. (doi:10.1089/cmb.2013.0084)
45. Laetsch DR, Blaxter ML, Leggett RM. 2017 BlobTools: interrogation of genome assemblies. *F1000Research* **6**, 1–16. (doi:10.12688/f1000research.12232.1)
46. Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL. 2009 BLAST plus: architecture and applications. *BMC Bioinf.* **10**, 1–9. (doi:10.1186/1471-2105-10-421)
47. Li W, Godzik A. 2006 Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics* **22**, 1658–1659. (doi:10.1093/bioinformatics/btl158)
48. Huerta-Cepas J *et al*. 2016 EGGNOG 4.5: a hierarchical orthology framework with improved functional annotations for eukaryotic, prokaryotic and viral sequences. *Nucleic Acids Res.* **44**, D286–D293. (doi:10.1093/nar/gkv1248)
49. Galindo LG, Torruella G, Eglit Y, Simpson AGB, Völcker E, Clauß S, Moreira D, López-García P. 2019 Data from combined cultivation and single-cell approaches to the phylogenomics of nucleariid amoebae, close relatives of Fungi. *Figshare* (doi:10.6084/m9.figshare.7814411.v1)
50. Heger TJ, Giesbrecht IJW, Gustavsen J, del Campo J, Kellogg CTE, Hoffman KM, Lertzman K, Mohn WW, Keeling PJ. 2018 High-throughput environmental sequencing reveals high diversity of litter and moss associated protist communities along a gradient of drainage and tree productivity. *Environ. Microbiol.* **20**, 1185–1203. (doi:10.1111/1462-2920.14061)
51. Arroyo AS, López-Escardó D, Kim E, Ruiz-Trillo I, Najle SR. 2018 Novel diversity of deeply branching holomycota and unicellular holozoans revealed by metabarcoding in Middle Paraná River, Argentina. *Front. Ecol. Evol.* **6**, 1–17. (doi:10.3389/fevo.2018.00099)
52. 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. (doi:10.1093/nar/gkf436)
53. Capella-Gutiérrez S, Silla-Martínez JM, Gabaldón T. 2009 TrimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* **25**, 1972–1973. (doi:10.1093/bioinformatics/btp348)
54. Nguyen LT, Schmidt HA, Von Haeseler A, Minh BQ. 2015 IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol. Biol. Evol.* **32**, 268–274. (doi:10.1093/molbev/msu300)
55. Guindon S, Gascuel O. 2003 A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst. Biol.* **52**, 696–704. (doi:10.1080/10635150390235520)
56. Anisimova M, Gil M, Dufayard J-F, Dessimoz C, Gascuel O, Nisimova MAA, Il MAG, Ranc JEAN. 2011 Survey of branch support methods demonstrates accuracy, power, and robustness of fast likelihood-based approximation schemes. *Syst. Biol.* **60**, 685–699. (doi:10.1093/sysbio/syr041)
57. Felsenstein J. 1985 Confidence limits on phylogenies: an approach using the bootstrap. *Evolution (N. Y).* **39**, 783–791. (doi:10.2307/2408678)
58. Kalyaanamoorthy S, Minh BQ, Wong TKF, von Haeseler A, Jermiin LS. 2017 ModelFinder: fast model selection for accurate phylogenetic estimates. *Nat. Methods* **14**, 587–589. (doi:10.1038/nmeth.4285)
59. Posada D. 2008 jModelTest: phylogenetic model averaging. *Mol. Biol. Evol.* **25**, 1253–1256. (doi:10.1093/molbev/msn083)
60. Ronquist F, Huelsenbeck JP. 2003 MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* **19**, 1572–1574. (doi:10.1093/bioinformatics/btg180)
61. Mikhailov KV, Simdyanov TG, Aleoshin VV. 2016 Genomic survey of a hyperparasitic microsporidian *Amphiamblys* sp. (Metchnikovellidae). *Genome Biol. Evol.* **9**, 454–467. (doi:10.1093/gbe/evw235)
62. Price MN, Dehal PS, Arkin AP. 2009 Fasttree: computing large minimum evolution trees with profiles instead of a distance matrix. *Mol. Biol. Evol.* **26**, 1641–1650. (doi:10.1093/molbev/msp077)
63. Leigh JWJW, Roger AJ. 2008 Barrel-o-Monkeys. See http://rogerlab.biochemistryandmolecularbiology.dal.ca/Software/Software.htm.
64. Lartillot N, Lepage T, Blanquart S. 2009 PhyloBayes 3: a Bayesian software package for phylogenetic reconstruction and molecular dating. *Bioinformatics* **25**, 2286–2288. (doi:10.1093/bioinformatics/btp368)
65. Minh BQ, Nguyen MAT, Von Haeseler A. 2013 Ultrafast approximation for phylogenetic bootstrap. *Mol. Biol. Evol.* **30**, 1188–1195. (doi:10.1093/molbev/mst024)
66. Rambaut A. 2009 FigTree. See http://tree.bio.ed.ac.uk/software/figtree/.
67. Dirren S, Salcher MM, Blom JF, Schweikert M, Posch T. 2014 Ménage-à-trois: the Amoeba *Nuclearia* sp. from Lake Zurich with its ecto- and endosymbiotic bacteria. *Protist* **165**, 745–758. (doi:10.1016/j.protis.2014.08.004)
68. Leonov MM. 2012 Heliozoans (Heliozoa, Sarcodina, Protista) of fresh and marine waters of the European part of Russia: species composition, morphology, and distribution. *Inl. Water Biol.* **3**, 344–355. (doi:10.1134/s1995082910040073)
69. Wujek DE. 2015 Freshwater silica-scaled heterotrophic protista: heliozoa, thaumatomonad flagellates, amoebae, and bicosoecids, from the Lake Itasca Region, Minnesota. *J. Minnesota Acad. Sci.* **78**, 1–14. (doi:10.1007/s10773-016-2998-7)
70. Feehan CJ, Johnson-Mackinnon J, Scheibling RE, Lauzon-Guay JS, Simpson AGB. 2013 Validating the identity of *Paramoeba invadens*, the causative agent of recurrent mass mortality of sea urchins in Nova Scotia, Canada. *Dis. Aquat. Organ.* **103**, 209–227. (doi:10.3354/dao02577)
71. Kantor RS, van Zyl AW, van Hille RP, Thomas BC, Harrison STL, Banfield JF. 2015 Bioreactor microbial ecosystems for thiocyanate and cyanide degradation unravelled with genome-resolved metagenomics. *Environ. Microbiol.* **17**, 4929–4941. (doi:10.1111/1462-2920.12936)
72. Hoffmeyer TT, Burkhardt P. 2016 Choanoflagellate models—*Monosiga brevicollis* and *Salpingoeca rosetta*. *Curr. Opin. Genet. Dev.* **39**, 42–47. (doi:10.1016/j.gde.2016.05.016)
73. Hehenberger E, Tikhonenkov DV, Kolisko M, del Campo J, Esaulov AS, Mylnikov AP, Keeling PJ. 2017 Novel predators reshape holozoan phylogeny and reveal the presence of a two-component signaling system in the ancestor of animals. *Curr. Biol.* **27**, 2043–2050. (doi:10.1016/j.cub.2017.06.006)
74. Deasey MC, Olive LS. 1981 Role of golgi apparatus in sorogenesis by the cellular slime mold *Fonticula alba*. *Science* **213**, 561–563. (doi:10.1126/science.213.4507.561)

5. Evolutionary genomics of Metchnikovella incurvata (Metchnikovellidae): An early branching microsporidium

“Only in the higher organic world is it possible to distinguish at opposite poles animal and plant life, and that these contrasting features do not exist for many lower forms of life"

Friedrich Kützing. *Ueber die Verwandlung der Infusorien in niedere Algenformen* (1844)

5. Evolutionary genomics of *Metchnikovella incurvata* (Metchnikovellidae): An early branching microsporidium

5.1. Context and objectives

The next branch to diverge after nucleariids on the Holomycotan tree is the one composed by Microsporidia and Rozellids. Many relationships of the different lineages within this clade remain unresolved, including the possibility that Rozellida forms a paraphyletic group basal to Microsporidia (see chapters 1.7.2.2. Microsporidia and 1.7.2.3 Rozellida). One poorly-known lineage with unresolved relationships are the metchnikovellids (Metchnikovellida). Metchikovellids are a group of microsporidians with traits that have been considered historically as "primitive", including a short polar tube and the lack of polaroplast (see chapter 1.7.2.2. Microsporidia). Recently, the genome of the first metchnikovellid, *Amphiamblys* sp. was sequenced (Mikhailov *et al.*, 2016) and allowed to show that metchnikovellids are the sister group to all other "core" Microsporidia. However, that study only sequenced the genome of this single species, which was morphologically uncharacterized. Thus, it was important to sequence more metchnikovellid genomes (if possible, from morphologically characterized species) to confirm that metchnikovellids form a coherent group sister to core Microsporidia. Additionally, this would allow to study if the gene content of metchnikovellid resembles more to canonical Microsporidia or to rozellids.

Thanks to the collaboration with other groups, we could get single-cell samples of the well-described metchnikovellid species *Metchnikovella incurvata*, which had been morphologically characterized.

We established the following objectives:

1) Confirm the branching order of metchnikovellids as sister clade to core Microsporidia. By sequencing the single-cell genome of *M. incurvata* we could run phylogenomic analyses to test that hypothesis.

2) Determine if the gene content and main metabolic pathways of *M. incurvata* and *Amphiamblys* sp. resembles more to core Microspordiaia or to microsporidia-like rozellids or rozellids.

5.2. Results

Our multi-gene phylogenomic studies with the genomes of both *Metchnikovella incurvata* and *Amphiamblys* sp. confirmed that *M. incurvata* belongs to the Metchnikovellidae, which formed a congruent clade sister to all other, long-branching, core Microsporidia. After comparing main metabolic GO term categories, we also confirmed that the metchnikovellid genome content resembled more that of core Microsporidia than that of other member of the Microsporidia + Rozellida clade. Additionally, gain and loss analysis of protein orthologous groups indicated that genome reduction and the appearance of new genes have co-occurred during the adaptation of Microsporidia to their diverse hosts.

5.3. Manuscript of article 2

Evolutionary Genomics of Metchnikovella incurvata (Metchnikovellidae): An Early Branching Microsporidium

(Genome Biol. Evol. 10(10):2736–2748)

Luis Javier Galindo[1]*, Guifré Torruella[1], David Moreira[1], Hélène Timpano[1], Gita Paskerova[2], Alexey Smirnov[2], Elena Nassonova[2,3], and Purificación López-García[1]*

[1]Ecologie Systématique Evolution, CNRS, Université Paris-Sud, AgroParisTech, Université Paris-Saclay, Orsay, France.

[2]Department of Invertebrate Zoology, Faculty of Biology, St Petersburg State University, Russia

[3]Laboratory of Cytology of Unicellular Organisms, Institute of Cytology Russian Academy of Sciences, St. Petersburg, Russia

* Corresponding authors

E-mails: puri.lopez@u-psud.fr; luis.galindo@u-psud.fr.

Evolutionary Genomics of *Metchnikovella incurvata* (Metchnikovellidae): An Early Branching Microsporidium

Luis Javier Galindo[1,*], Guifré Torruella[1], David Moreira[1], Hélène Timpano[1], Gita Paskerova[2], Alexey Smirnov[2], Elena Nassonova[2,3], and Purificación López-García[1,*]

[1]Ecologie Systématique Evolution, CNRS, Université Paris-Sud, AgroParisTech, Université Paris-Saclay, Orsay, France

[2]Department of Invertebrate Zoology, Faculty of Biology, St Petersburg State University, Russia

[3]Laboratory of Cytology of Unicellular Organisms, Institute of Cytology Russian Academy of Sciences, St. Petersburg, Russia

*Corresponding authors: E-mails: puri.lopez@u-psud.fr; luis.galindo@u-psud.fr.

Accepted: September 12, 2018

Data deposition: This project has been deposited at DDBJ/ENA/GenBank under the BioProject number PRJNA477760, the SRA accession number SRP151413, and the Whole Genome Shotgun project QXFS00000000.

Abstract

Metchnikovellids are highly specialized hyperparasites, which infect and reproduce inside gregarines (Apicomplexa) inhabiting marine invertebrates. Their phylogenetic affiliation was under constant discussion until recently, when analysis of the first near-complete metchnikovellid genome, that of *Amphiamblys* sp., placed it in a basal position with respect to most Microsporidia. Microsporidia are a highly diversified lineage of extremely reduced parasites related to Rozellida (Rozellosporidia = Rozellomycota = Cryptomycota) within the Holomycota clade of Opisthokonta. By sequencing DNA from a single-isolated infected gregarine cell we obtained an almost complete genome of a second metchnikovellid species, and the first one of a taxonomically described and well-documented species, *Metchnikovella incurvata*. Our phylogenomic analyses show that, despite being considerably divergent from each other, *M. incurvata* forms a monophyletic group with *Amphiamplys* sp., and confirm that metchnikovellids are one of the deep branches of Microsporidia. Comparative genomic analysis demonstrates that, like most Microsporidia, metchnikovellids lack mitochondrial genes involved in energy transduction and are thus incapable of synthesizing their own ATP via mitochondrial oxidative phosphorylation. They also lack the horizontally acquired ATP transporters widespread in most Microsporidia. We hypothesize that a family of mitochondrial carrier proteins evolved to transport ATP from the host into the metchnikovellid cell. We observe the progressive reduction of genes involved in DNA repair pathways along the evolutionary path of Microsporidia, which might explain, at least partly, the extremely high evolutionary rate of the most derived species. Our data also suggest that genome reduction and acquisition of novel genes co-occurred during the adaptation of Microsporidia to their hosts.

Key words: Microsporidia, Metchnikovellidae, Holomycota, phylogenomics, phylogeny, comparative genomics.

Introduction

Microsporidia (Opisthokonta) are a highly specialized group of intracellular parasites. This phylum currently includes between 1,300 and 1,500 described species, which parasitize diverse animal groups and, less frequently, protists (Vávra and Lukeš 2013). Among their animal hosts, many have economic importance, such as silkworms, honey bees, and fish. They can also be opportunistic parasites of humans, being particularly harmful in immunosuppressed patients (Didier et al. 2004; Didier and Weiss 2006, 2011). Microsporidia harbor some of the most reduced genomes among eukaryotes (Corradi et al. 2010). During the course of evolution, members of this lineage have lost or drastically simplified several typical eukaryotic features, including canonical mitochondria (Embley and Martin 2006), the flagellum (James et al. 2006) and a conventional Golgi apparatus (Beznoussenko et al. 2007). Although their evolutionary history has been essentially reductive, Microsporidia have also developed key evolutionary innovations, such as a unique infection apparatus, the polar tube, which serves to penetrate the host (Wittner and Weiss 1999).

For a long time Microsporidia were erroneously thought to be the deepest branching eukaryotes (Leipe et al. 1993; Kamaishi et al. 1996) until the discovery that this result was likely due to a long branch attraction artefact produced by the fast evolutionary rate of these organisms (Philippe et al. 2000). Recently, the position of Microsporidia as close fungal relatives has been recurrently substantiated (Capella-Gutiérrez et al. 2012; James et al. 2013). In particular, Microsporidia appeared to be related to a diverse clade of parasitic organisms known as rozellids—synonymic taxonomic designations: Rozellida (Lara et al. 2010), Cryptomycota (Jones et al. 2011), Rozellomycota (Corsaro et al. 2014), Rozellosporidia (Karpov et al. 2017). However, the precise reconstruction of their phylogenetic relationships and evolutionary traits remains problematic due to their high evolutionary rates, reduction or loss of cellular organelles, and loss of core metabolic routes (Williams et al. 2002; Thomarat et al. 2004; Corradi et al. 2010). Until recently, the only rozellid species with a sequenced genome was *Rozella allomyces*. In recent years, more sequence data for more or less distant relatives of *Rozella* and early-branching Microsporidia have been made available, opening up the possibility to carry out comparative genomic analyses and gain insights in the genome reduction process that seemingly occurred along the Microsporidia branch. These include the genomic surveys on the early-branching *Mitosporidium daphniae* (Haag et al. 2014), which is the only microsporidium with functional DNA-containing mitochondria described to date, and the rozellid *Paramicrosporidium saccamoebae*, branching at a somewhat intermediate position between *Rozella* and *Mitosporidium* (Quandt et al. 2017). Another key deeply branching lineage along the Microsporidia branch is that of metchnikovellids.

The metchnikovellids (taxonomically designated as the family Metchnikovellidae; Caullery and Mesnil, 1914) unites hyperparasites of gregarines (Apicomplexa) that inhabit the intestinal tract of marine annelids (Vivier 1975). Only a few genera have been described to date, including *Amphiamblys*, *Amphiacantha*, and *Metchnikovella*. Some members of the clade, such as *Metchnikovella incurvata*, have been known for >100 years (Caullery and Mesnil 1914). The phylogenetic affiliation of this long-standing *incertae sedis* group has been debated over time. Because of their morphological and ultrastructural characteristics, metchnikovellids were often thought to be related to Microsporidia (Sprague 1977). Indeed, like most Microsporidia, they lack canonical mitochondria. However, their spores do not exhibit some key microsporidian features, such as the coiled polar filament, the polaroplast and a merogonial proliferation in the life cycle (Sokolova et al. 2013). Phylogenomic analysis of the first available genome of a metchnikovellid, that of *Amphiamblys* sp. (Mikhailov et al. 2017) placed this lineage as the sister group of all derived Microsporidia with the exception of *M. daphniae*, which was placed close to the root of Microsporidia, thereby confirming the long-held suspicion that the metchnikovellids are early diverged Microsporidia. The analysis of the *Amphiamblys* sp. genome revealed some remarkable features, such as the absence of the ATP/ADP translocase family, which is ubiquitous in all derived Microsporidia (Tsaousis et al. 2008), and raised the question of how metchnikovellids obtain ATP without this transporter. However, although seemingly quite complete, the amplified *Amphiamblys* sp. genome is nonetheless partial and these peculiar features need to be verified in other members of the group. Obtaining novel metchnikovellid genome sequences might thus be very useful to determine synapomorphies for the clade and refine the evolutionary path to extreme genome reduction observed along the Microsporidia branch.

In this study, we have analyzed the genome of a second metchnikovellid species, *M. incurvata*. This is the first genomic and phylogenetic study of a taxonomically described and well-documented metchnikovellid species (Sokolova et al. 2013; Rotari et al. 2015). Our results confirm the monophyly of *Amphiamblys* and *Metchnikovella* and strongly support the notion that metchnikovellids branch deeply in the Microsporidia lineage, providing insights into the evolution of the Microsporidia proteome along the diversification of this lineage.

Materials and Methods

Biological Samples

Individual cells of the gregarine *Polyrhabdina* sp. infected with the metchnikovellid *M. incurvata* were isolated from the intestinal tract of the polychaete *Pygospio elegans*. Polychaetes were collected from the Levin Reach silt littoral zone in the Chupa Inlet of the Kandalaksha Gulf, located in the White Sea (66°17′52.68″N, 33°27′46.44″E). Polychaetes were dissected and infected gregarine cells were individually isolated, washed by successive passage into filtered seawater droplets, and sorted into separate tubes for further analyses. At light microscopic level the infected gregarines were easily recognized by the presence of rounded and oval inclusions (fig. 1*A* and *B*) or by elongated and slightly curved cysts inside the cytoplasm (fig. 1*C*). On the basis of the individual characters used for the species identification within the genus *Metchnikovella*, the super-host range, the host range, the size and shape of the observed cysts (Sokolova et al. 2013; Rotari et al. 2015) the microsporidium was identified as *M. incurvata*.

Single-Cell Genome Amplification and Sequencing

Total DNA extraction was performed on one single-isolated infected gregarine cell (containing proliferating cells of *M. incurvata*; see fig. 1*A*) using the PicoPure kit (Thermo Fisher Scientific) according to the manufacturer's protocol. Whole genome amplification (WGA) of the DNA purified from this single infected gregarine was carried out using two methodologies, either multiple displacement amplification (MDA)

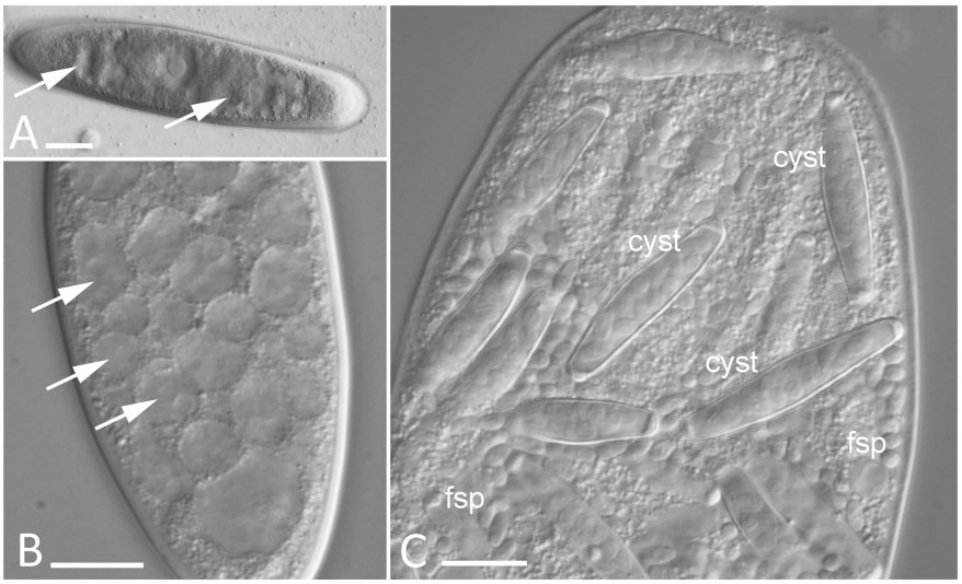


Fig. 1.—*Metchnikovella incurvata*, a hyperparasite of gregarines *Polyrhabdina* sp. from the polychaete *Pygospio elegans*. (*A*, *B*) Infected gregarine cells filled with rounded and oval inclusions (arrowed) corresponding to the early stages of hyperparasite proliferation. Panel *A* shows the cell from which DNA extraction and the subsequent whole genome amplification by MDA were done and which was further used for single-cell genome sequencing. (*C*) Infected gregarine cell filled with the cysts (cyst) and free spores (fsp). Scale bar: 5 μm.

with the REPLI-g kit (QIAGEN) or amplification with the MALBAC single cell WGA kit (Yikon Genomics), following the manufacturer's protocols. We then proceeded to exclude single amplified genomes (SAGs) which did not yield the expected metchnikovellid 18S rDNA amplicons when tested by PCR amplification. Three independent WGA reactions were done on DNA extracted from the individually isolated gregarine cell documented in figure 1*A*: Two MDA reactions (LNA5-MDA1 and LNA5-MDA2) and one MALBAC reaction (LNA5-MALBAC). DNA amplification was confirmed by assessing the DNA quantity using Qubit fluorometric quantification (Life Technologies), together with PCR amplification and Sanger sequencing of the 18S rRNA gene. From the two MDA reactions performed for the same sample, we obtained a DNA concentration of 60 ng/µl (LNA5-MDA1) and 129.2 ng/µl (LNA5-MDA2), and for the MALBAC reaction, we obtained 23.4 ng/µl (LNA5-MALBAC). Since the MALBAC sample did not yield a high DNA amount and since we failed to amplify the 18S rRNA gene, we proceeded to sequence one of the MDA samples (LNA5-MDA1). We prepared two WGA TruSeq Paired-end libraries from this sample and sequenced them on a HiSeq 2500 Illumina instrument (2 × 125 bp) chemistry v4. For each library, we obtained 102,700,329 reads for a total length of 25,675 Mbp and 99,095,593 reads for a total length of 24,774 Mbp, respectively.

Genome Decontamination, Assembly, and Annotation

The quality of the paired-end reads was assessed with FastQC (Andrews 2010) before and after quality trimming. We then trimmed the Illumina adapters with Trimmomatic v0.32 in Paired End mode (Bolger et al. 2014), with a minimal length of 100 bp, removing the first 15 bp, a minimum quality allowed of 20 at the beginning and end of the read. Trimmed pair-end reads were assembled using SPAdes 3.9.1 in single-cell mode (Bankevich et al. 2012), with four k-mer values (25, 77, 99, 117); two assembly rounds were performed, one before and one after decontamination. The first round resulted in an assembly of 6.45 Mb formed by 1,667 contigs of prokaryotic and eukaryotic origin. To decontaminate the assembly, we used BlobTools (Kumar et al. 2013; Laetsch and Blaxter 2017), generating taxon-annotated GC plots. We then eliminated abundant contaminant prokaryotic reads identified in the generated plots. We also removed a few reads that were suspected to belong to the apicomplexan or to the polychaetan hosts by the BlobTools taxonomic identity; these reads were manually inspected with BlastN (Altschul et al. 1990) prior to removal. Surprisingly, there were almost no contaminating reads coming from the apicomplexan host, likely due to the advanced stage of *M. incurvata* infection. After decontamination, the remaining reads underwent a second round of assembly, and were again analyzed with BlobTools to confirm the success of the decontamination procedure (supplementary fig. S1, Supplementary Material online). The final *M. incurvata* assembly had 5.39 Mb and 1,257 contigs. The statistics of the final assembled genome were assessed with QUAST 4.5 (Gurevich et al. 2013) and Qualimap v2.2.1 (Okonechnikov et al. 2015) for coverage estimation. De novo functional gene annotation for the *M. incurvata* genome was performed using two gene

prediction programs: Augustus 3.0.3 (Stanke and Morgenstern 2005) and GeneMarkS v3.26 (Besemer et al. 2001). A few potential introns were predicted by Augustus, but further exploration using BlastX (BLAST 2.6.0; Altschul et al. 1997) against a wide diversity of eukaryotic proteomes included in our local database, rejected them. We used GeneMarkS, an intronless gene prediction algorithm, for gene prediction in the *M. incurvata* genome. Repetitive elements in the genome of *M. incurvata* were searched using RepeatModeler 1.0.10 (Smit et al. 1996). We generated a custom library of repetitive families with RepeatModeler. The resulting sequences were subsequently aligned to the *M. incurvata* assembly and masked by RepeatMasker 4.0 (Smit et al. 2013) to provide a table of the distribution of repetitive element families for the group. This annotation of families of repetitive elements was done including several control genomes to confirm the consistency and quality of the annotation. To assess the completeness of the *M. incurvata* genome, we used BUSCO v2.0.1 (Simão et al. 2015) on the annotated genes with the Fungi and Microsporidia data sets of near-universal single-copy orthologs.

Phylogenetic Analyses

We reconstructed molecular phylogenetic trees of the 18S rRNA genes and phylogenomic analyses of a multigene data set using maximum likelihood (ML) methods (Felsenstein 1981) and Bayesian inference (BI; Huelsenbeck and Ronquist 2001). All alignments were performed using MAFFT v7.388 (Katoh and Standley 2013) with default parameters. Alignments were inspected manually using Geneious v6.0.6 (Kearse et al. 2012), and trimmed from ambiguously aligned regions and gaps using trimAl v1.2 in automated1 mode (Capella-Gutiérrez et al. 2009). For the 18S rRNA, ML inferences were done using IQ-TREE v1.6.2 (Nguyen et al. 2015) applying the TIM3 model with four gamma categories and empirical base frequencies (F+G4), which was the best fit model chosen by BIC (Posada, 2008). 18S rRNA BI analyses were performed using Phylobayes v1.5a (Lartillot and Philippe 2004, 2006; Lartillot et al. 2007), under the CAT-Poisson evolutionary model. Two independent MCMC chains for each data set were run for 10,000 cycles and summarized with a 25% burn-in. For the multigene phylogenomic analyses we used a previously assembled 56 Single-Copy Protein Domains (SCPD) data set (Torruella et al. 2012) containing 32 representatives of the Holomycota clade and 5 other Amorphea species as outgroup (2 Holozoa, 1 Apusomonadida, and 2 Amoebozoa). Proteome data were obtained from GenBank (http://www.ncbi.nlm.nih.gov/genbank, last accessed February 25, 2018), except for the proteomes of *Antonospora locustae*, *Mortierella alpina*, *Rhizopus oryzae*, and *Lichtheimia corymbifera*, which were obtained from the JGI Genome Portal (http://genome.jgi.doe.gov/, last accessed January 22, 2018). ML analyses were performed using IQ-TREE v1.6.2 (Nguyen et al. 2015) applying the LG evolutionary model with four gamma categories, empirical amino acid frequencies and a proportion of invariable sites (LG + F + I + G4), which was the best fit model according to BIC. The BI analyses of the SCPD data set were performed with Phylobayes v1.5a (Lartillot and Philippe 2004, 2006; Lartillot et al. 2007), under the CAT-Poisson evolutionary model. Two independent MCMC chains for each data set were run for 10,000 cycles and summarized with 25% burn-in. All trees were visualized using FigTree v1.4.3 (Rambaut 2016).

Functional Annotation of the Predicted Proteome

Assignment of Gene Ontology (GO) terms to the *M. incurvata* proteome was done using the eggNOG-mapper from the EggNOG v4.5 (Huerta-Cepas et al. 2017) database, using DIAMOND as mapping mode, and the taxonomic scope to adjust automatically. We did this for 12 Holomycota proteomes (1 Fungi, 2 Cryptomycota, and 9 Microsporidia). We used the native R heatmap function (R Development Core Team 2014) to plot a proteome comparison between 73 different core GO terms. We also compared GO terms of gene/pathways involved in DNA repair for 27 holomycotan proteomes, specifically GOs involved in homologous recombination, nonhomologous end joining, mismatch repair, nucleotide-excision repair, and base-excision repair. In addition, we used HMMER 3.1b2 software (Finn et al. 2011) to search for specific genes of interest in the *M. incurvata* proteome, including 15 genes reported to have been acquired by Microsporidia by horizontal gene transfer (HGT; Tsaousis et al. 2008; Marcet-Houben and Gabaldón 2010; Xiang et al. 2010; Heinz et al. 2012; Pombert et al. 2012; Nakjang et al. 2013; Alexander et al. 2016).

Gene Gain and Loss Analysis

We selected the 37 opisthokont proteomes used for the phylogenomic analysis, including that of *M. incurvata*, and carried out orthologue clustering with OrthoFinder v1.1.20 (Emms and Kelly 2015) with default parameters. We identified 12,448 orthogroups, of which 792 were genome-specific. To infer gene gain and loss in protein families along the Microsporidia line, we then applied the Dollo parsimony method implemented in the Count software (Csurös 2010) on the phylogenomic tree topology obtained by BI. Finally, specific metchnikovellid orthologs were later validated with HMMER search (Finn et al. 2011) on EggNOG v4.5 (Huerta-Cepas et al. 2017).

Data Availability

Data generated for this study has been deposited at DDBJ/ENA/GenBank under the BioProject number PRJNA477760,

Table 1
Key Statistics for Genome Assembly and Annotation

	R. allomycis	*P. saccamoebae*	*M. daphniae*	*M. incurvata*	*Amphiamblys* sp.	*N. parisii* (ERTm1)	*T. hominis*	*E. cuniculi*
Assembly size (Mb)	11.86	7.28	5.64	5.4	5.6	4.15	8.5	2.5
GC%	34.5	46.9	43	32.62	50.2	34.5	34.1	34.5
Number of contigs	1,150	216	612	1,257	1,843	65	1,632	11
N50	48,693	69,936	32,031	14,622	10,678	649,559	9,528	220,294
Protein-coding genes	6,350	3,750	3,331	2,803	3,647	2,726	3,212	1,996
Repetitive elements %	3.23%	5.91%	3.69%	17.53%	32.6%	10.59 %	7.43%	10.78%

the SRA accession number SRP151413, and the Whole Genome Shotgun project QXFS00000000.

Results and Discussion

Genome Organization and Repetitive Elements

After the amplification and Illumina sequencing of the *M. incurvata* genome from one single gregarine cell (fig. 1*A*), we generated an initial draft genome of 6.45 Mb, with an N50 of 14,687 bp. After decontamination of the identified nonmetchnikovellid sequences (supplementary fig. S1, Supplementary Material online), we obtained a genome of 5.4 Mb with a N50 of 14,622 bp and a GC content of 32.62%, encoding a total of 2,803 proteins (table 1). The sequence coverage distribution along the genome followed a normal distribution with a notably high mean of ~7000× (supplementary fig. S2, Supplementary Material online), a result that might be expected after the MDA amplification of a relatively small genome. To assess the completeness of this genome, we conducted a BUSCO analysis using as reference the Fungi and Microsporidia data sets of near-universal single-copy orthologs (supplementary table S1, Supplementary Material online). The fungal data set provided a better coverage of the early branching metchnikovellid genome, as it allowed us to identify 196 complete single-copy, 6 duplicated, 31 fragmented, and 57 missing core orthologs, from a total of 290. This allowed inferring ~80% genome completeness for the *M. incurvata* genome sequence, which is comparable to that for the *Amphiamblys* sp. genome (90%; Mikhailov et al. 2017).

Around 43% of the *M. incurvata* genome corresponded to coding regions (similar to the 46% seen in *Amphiamblys* sp.). Repetitive elements accounted for 17.53% of the whole genome, most of which were interspersed repeats (16.2% of the genome; supplementary table S2, Supplementary Material online). To compare our results to those of previous reports we calculated the percentage of repetitive elements in other members of the group (table 1). For *Amphiamblys* sp. we inferred that 32.6% of the genome corresponded to repetitive sequences; similar to the 30% calculated by Mikhailov et al. (2017). However, the percentage of repetitive elements that we found in *P. saccamoebae*, was 5.92%, ten times the percentage (0.53%) reported by Quandt et al. (2017). Interestingly, the percentage of repetitive elements in organisms with smaller genomes was higher than that in those with larger genomes. This might seem at odds with the tendency of Microsporidia towards genome reduction and compaction. Nonetheless, these repetitive elements could be involved in the genome reduction process, as their presence often accompanies pseudogenization and gene loss (Lynch and Conery 2000; Jurka 2004; Dewannieux and Heidmann 2005). At the same time, the presence and relative abundance of repetitive elements in Microsporidia could also play an important role in the adaptation and evolution of this lineage (Parisot et al. 2014). Repetitive elements contribute to genome plasticity in several organisms (Biémont 2010), including closely related fungi. Accordingly, they might contribute to the adaptation to different hosts (Dean 2005; Amyotte et al. 2012; Raffaele and Kamoun 2012). In the case of the microsporidium *Anncaliia algerae*, these elements might be speculated to help eluding the host immune system by acting as a lure (Panek et al. 2014).

The loss of introns and spliceosome activity is thought to have occurred independently in several Microsporidia lineages. Similarly to *Amphiamblys* sp. (Mikhailov et al. 2017), *M. incurvata* seems to lack introns and, consequently, the spliceosome machinery is practically inexistent. For example, neither *M. incurvata* nor *Amphiamblys* sp. seem to possess genes coding for Sf3b1 and Prp8 (two proteins that form the spliceosome catalytic core); these genes have been found in Microsporidia with active splicing (Desjardins et al. 2015). However, we found a RtcB-like ligase involved in tRNA splicing and repair (supplementary fig. S3, Supplementary Material online). This gene is ubiquitous in all eukaryote groups, except fungi, but is present in Microsporidia including *M. daphniae* (Haag et al. 2014). We have also found this gene in *P. saccamoebae*. Therefore, it was likely present in the last common holomycotan ancestor and retained during the evolution and diversification of Microsporidia as intracellular parasites. It has been hypothesized that, in bacteria, RctB may be involved in repair from stress-induced RNA damage; their homologs might catalyze tRNA repair or splicing reactions in archaea and eukaryotes (Tanaka et al. 2011; Tanaka and Shuman 2011). Consequently, RctB in Microsporidia possibly evolved to cope with RNA damage (Thomarat et al. 2004).

Microsporidia are classically thought to lack canonical mitochondria and to possess instead mitosomes, which are highly reduced mitochondrial-derived organelles without respiratory function, involved in the biosynthesis of iron–sulfur clusters essential for many proteins (Hirt et al. 1997; Tsaousis et al. 2008; Waller et al. 2009; Boniecki et al. 2017; Freibert et al. 2017). However, recent studies have shown that the early-branching microsporidium *M. daphniae* and the closely related *P. saccamoebae* still possess mitochondria. Since the phylogenetic position of Metchnikovellidae lies between Microsporidia with and without mitochondria, they might have retained intermediate mitochondria-related organelles important to understand the transition towards mitosomes. We found genes coding for both mitochondrial Hsp70 and the essential sulfur donor Nfs1 (which is of mitochondrial origin) (Emelyanov 2003) in the *M. incurvata* genome (supplementary fig. S4, Supplementary Material online). These genes were also present in *Amphiamblys* sp. and in all Microsporidia with a mitosome (Tsaousis et al. 2008). They seem to play a key role in mitosomes, being required for the maturation of diverse functional proteins (Kispal et al. 2005). The Hsp70 gene phylogeny was congruent with the known phylogeny of Microsporidia (supplementary fig. S4*A*, Supplementary Material online). We did not find any genes related to the main mitochondrial metabolic routes, such as a functional oxidative phosphorylation or the tricarboxylic acid cycle (supplementary table S3, Supplementary Material online). Therefore, the metchnikovellid mitochondrion-derived organelle seems to resemble more a microsporidian mitosome than the *M. daphniae* or *P. saccamoebae* mitochondria.

Phylogenomics of *M. incurvata*

We identified the 18S rRNA gene sequence in the *M. incurvata* assembled genome and reconstructed the corresponding phylogenetic tree (supplementary fig. S5, Supplementary Material online). As expected, *M. incurvata* formed a clade with the rest of metchnikovellids, and the Metchnikovellidae family formed a strongly supported group (with maximum support values for bootstrap and posterior probabilities) branching basally to the clade of canonical, long-branching Microsporidia (which we name here Core Microsporidia). These results support the monophyly of metchnikovellids and their early divergence in the lineage, even though many other nodes in the 18S rRNA gene tree remain unresolved.

To reconstruct a more robust phylogeny for metchnikovellids, we carried out a multigene phylogenomic analysis for several members of the Holomycota. Our ML and BI trees further confirmed that *M. incurvata* forms a solidly supported monophyletic lineage with *Amphiamblys*. Despite this, *Metchnikovella* is only distantly related to *Amphiamblys* (fig. 2), suggesting that the family Metchnikovellidae might encompass a wide diversity of fast-evolving parasites specialized in various hosts. Interestingly, the Metchnikovellidae branched at the base of the fast-evolving, more derived, Core Microsporidia, right after the basal-branching *M. daphniae* and *P. saccamoebae*. Their intermediate position between classical Microsporidia with mitosomes, and the basal mitochondriate members *M. daphniae* and *P. saccamoebae* and *R. allomycis*, makes this group interesting for studying the evolutionary path to the extreme genome reduction and specialization undergone by the long-branching Microsporidia.

Molecular analyses based on 18S rRNA gene amplification and sequencing from various environments have uncovered a wide diversity of eukaryotes along the lineage leading to Core Microsporidia, from the very basal rozellids to the metchnikovellid divergence (Lara et al. 2010; Bass et al. 2018). In addition to this uncharacterized environmental diversity, a few described genera occupy deep-branching positions in this broad lineage based on 18S rRNA gene phylogenies and their further study should shed some light in the evolutionary history of Microsporidia. An example is *Nucleophaga*, which branches between *Paramicrosporidum* and metchnikovellids plus the Core Microsporidia (Corsaro et al. 2016).

Genome Evolution and Gene Gain and Loss along the Microsporidia Line

We annotated the *M. incurvata* genome and 11 additional members of the Holomycota clade for 73 different GO terms, and then clustered the annotated genomes according to their gene content similarity (fig. 3). In agreement with their phylogeny, *Metchnikovella* and *Amphiamblys* clustered together also according to their gene content. However, the gene content of metchnikovellids has undergone the loss of many core function genes, resembling more that of the highly reduced derived Microsporidia (indicated as Core Microsporidia in fig. 3). This suggests that metchnikovellids rely on the assimilation of metabolites from their hosts, as long-branching Microsporidia do. Remarkably, the metabolic core of *M. daphniae* and *P. saccamoebae* clustered them with *R. allomycis* (fig. 3). Thus, although *Paramicrosporidium* and *Mitosporidium* are considered to be basal microsporidia by some authors (Bass et al. 2018), their functional gene content is more similar to that of rozellids. This, together with the fact that *P. saccamoebae* and *M. daphniae* possess functional mitochondria may question their classification as microsporidia and claims for a taxonomic revision of Microsporidia and Rozellida, the boundaries of which are blurry.

Despite their reduced genomes, *Amphiamblys* and *Metchnikovella* still conserve some basic core metabolic pathways, such as the pentose phosphate and glycolysis pathways and trehalose biosynthesis (supplementary table S3, Supplementary Material online). Nevertheless, like most Microsporidia, metchnikovellids cannot synthesize their own nucleotides, amino acids and lack practically all enzymes involved in fatty acid metabolism (supplementary tables S3, S4, Supplementary Material online).

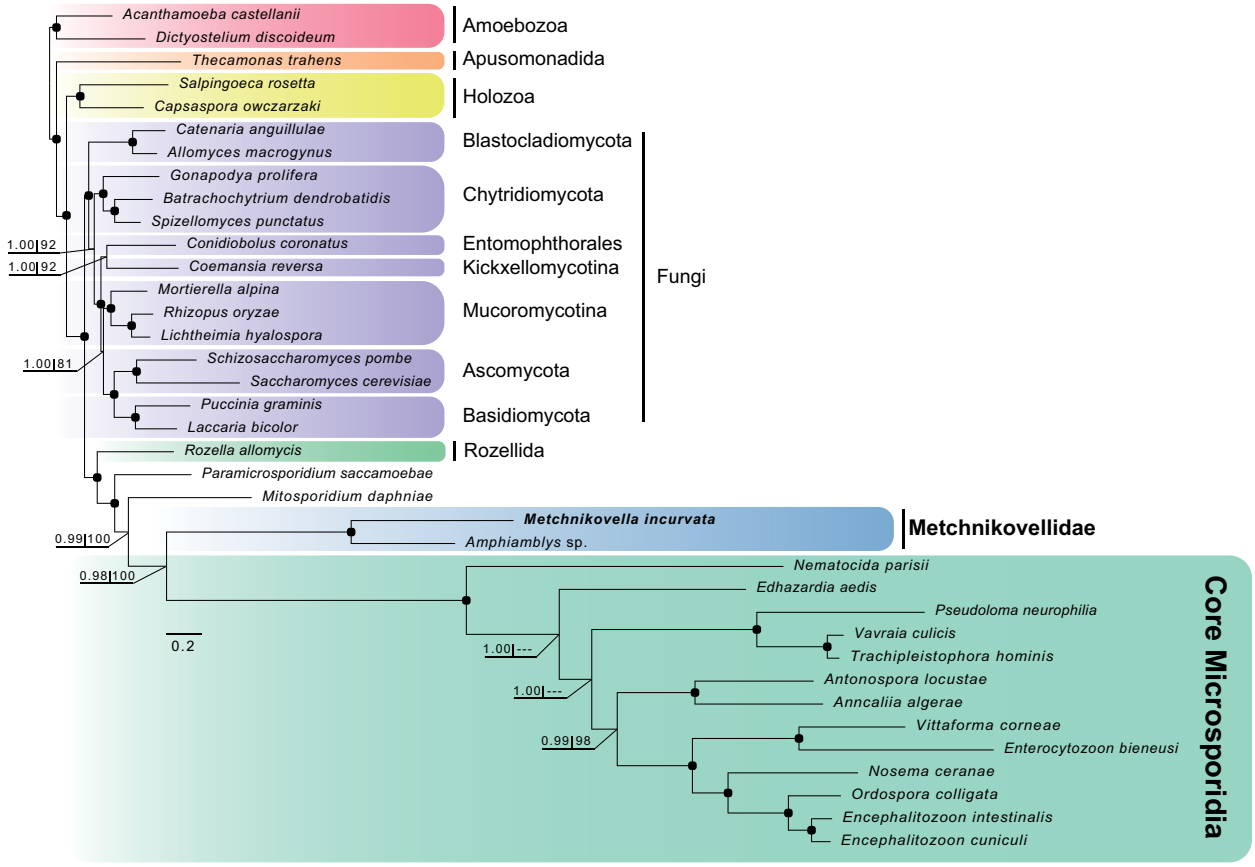


Fig. 2.—Bayesian phylogenomic tree showing the position of metchnikovellids. The tree was reconstructed using a concatenated alignment of 56 single-copy protein domain data set for 32 representatives of the Holomycota clade and 5 other Amorphea species as an outgroup (2 Holozoa, 1 Apusomonadida, and 2 Amoebozoa). Split supports are posterior probabilities (pp) (values on the left) and maximum likelihood (ML) bootstrap (bs) values (on the right). Sequences obtained in this study are highlighted in black. Support values >0.99 pp and >95% bs are indicated by a black bullet.

Interestingly, unlike both divergent Microsporidia and early branching members of this lineage, the two metchnikovellid species lacked an alternative oxidase and the mitochondrial glycerol-3-phosphate dehydrogenase, one of the two enzymes of the glycerol-3-phosphate shuttle. This may appear surprising, since both are thought to be essential for the viability of microsporidian energy metabolism (Dolgikh et al. 2009; Williams et al. 2010). However, metchnikovellids retain a cytosolic glycerol-3-phosphate dehydrogenase (supplementary fig. S6, Supplementary Material online). This enzyme seems to be important in their metabolism as it may allow the synthesis of glycerol 3-phosphate, the starting material for de novo synthesis of glycerolipids and NAD^+, necessary to maintain the adequate NAD^+ cellular levels.

The tricarboxylic acid cycle does not seem to be functional in metchnikovellids, although genes for some enzymes of the pathway are still present, namely the citrate synthase (only detected in *Amphiamblys* sp.) and the malate dehydrogenase (supplementary table S3, Supplementary Material online). The mitochondrial malate dehydrogenase (mMDH) found in *M. incurvata* and *Amphiamblys* sp. is absent in all derived Microsporidia. It has been shown that in some cases the function of this enzyme may shift to a lactate dehydrogenase (LDH; Wilks et al. 1988), which potentially constitutes the final step of anaerobic energy metabolism in metchnikovellids, by balancing the reducing potential of glycolysis. In *Amphiamblys* sp., this change has been reported to occur when an arginine in a key active site changed to a tyrosine (Mikhailov et al. 2017). In *M. incurvata* the change is from arginine to tryptophan, another hydrophobic amino acid (supplementary fig. S7, Supplementary Material online). However, it is likely that the enzyme works as a LDH, since the same substitution (Arg102 to a Trp107) functionally turned the MDH into LDH in Apicomplexa, another group of parasitic protists (Boucher et al. 2014).

One interesting example of differential gene retention in metchnikovellids as compared with core Microsporidia relates to clathrin-coated vesicle formation. Mikhailov et al. (2017) already noted that the *Amphiamblys* sp. genome encoded several proteins required for the formation of clathrin vesicles.

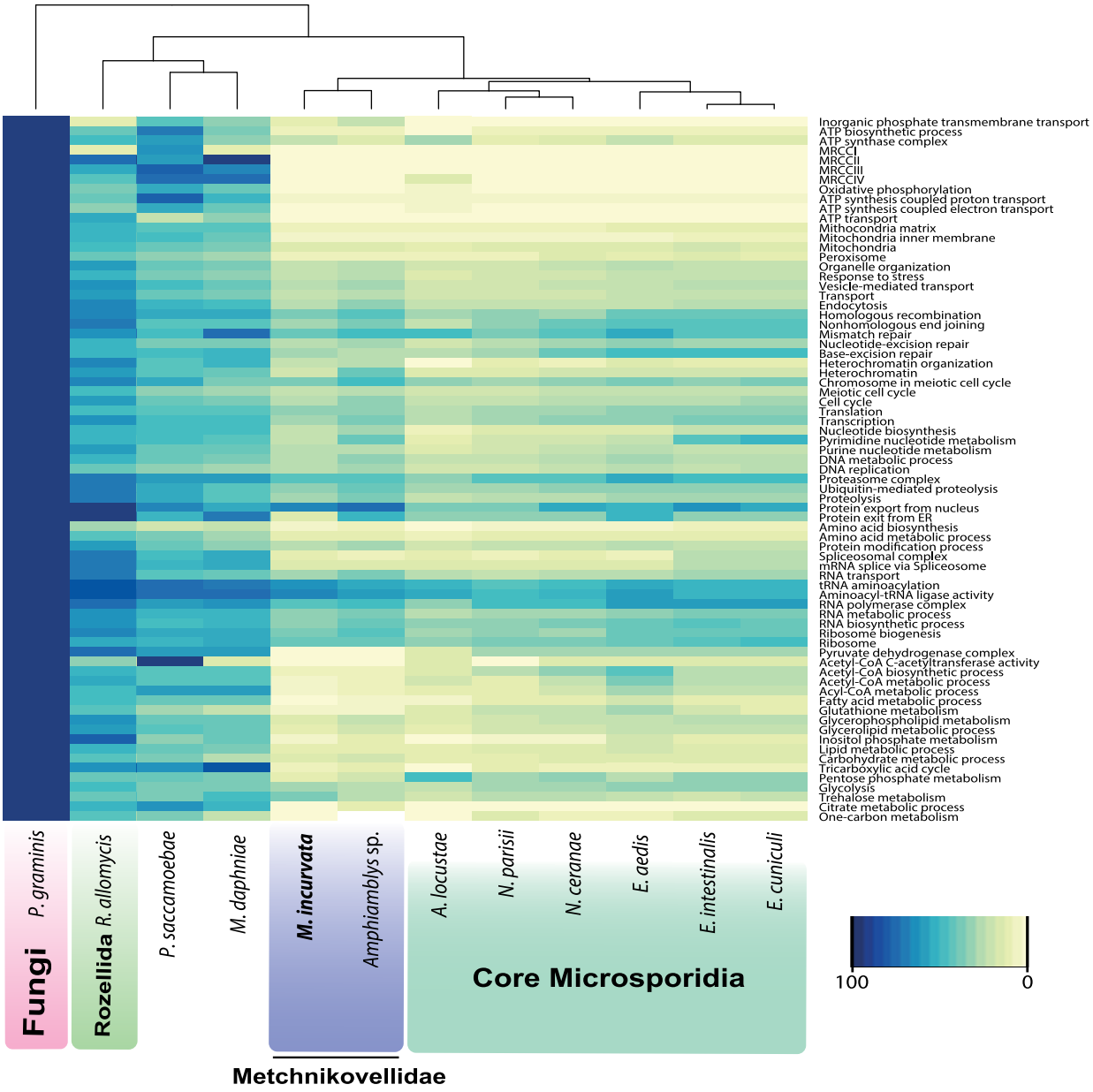


Fig. 3.—Heatmap illustrating the conservation of pathways and structures in metchnikovellids and neighbor lineages. It is based on 73 GO terms, identified using EggNOG (Huerta-Cepas et al. 2017) for 12 Holomycota representative proteomes. Sequences obtained in this study are highlighted in black. Colors indicate the percentage of annotated genes with a GO term.

These are lacking in more derived Microsporidia, which have a highly reduced endocytic machinery. We identified in *M. incurvata* 10 out of the 12 genes involved in clathrin-coated vesicle formation present in *Amphiamblys* sp. (supplementary table S5, Supplementary Material online). We failed to identify an actin-binding WH2 domain (PF02205) and an ARP2/3 complex subunit ARPC3 (PF04062). Since our genome is not complete, the presence of these two genes cannot be discarded. In addition, we do find fragments of other members of the two respective families: WH1 (PF00568) and 2 subunits of the complex ARP2/3 (PF04045 and PF05856). Therefore, with the data available for the two metchnikovellid genomes, the endocytic components of the clathrin vesicle-mediated transport seem to be conserved in metchnikovellids, making it one of the main distinctions of this group from more derived Microsporidia.

Microsporidian genomes are among the fastest-evolving eukaryotic genomes (Thomarat et al. 2004). This high evolutionary rates, which make this group prone to long-branch attraction artifacts, made it difficult for a long time to place

Microsporidia in the tree of life (Capella-Gutiérrez et al. 2012). Since Microsporidia have highly reduced genomes, we asked whether the loss of genes involved in DNA repair might have been responsible for an augmentation of mutation rates. We thus searched for genes in the five main GO terms involved in DNA repair (homologous recombination, nonhomologous end joining, mismatch repair, nucleotide-excision repair, and base-excision repair) in 27 opisthokont genomes. We found that Microsporidia, including metchnikovellids, do indeed have a lower number of genes involved in DNA repair (supplementary table S6; fig. S8, Supplementary Material online). Even in the case of nucleotide-excision repair, where gene loss might appear less important that in other DNA repair systems (supplementary fig. S8, Supplementary Material online), the two metchnikovellid genomes have a gene reduction of circa 50% as compared with more basal opisthokonts (*Amphiamblys*, 41 genes; *M. incurvata*, 46 genes; vs. e.g., *S. rosetta*, 98 genes; *T. trahens*, 78 genes; *R. allomycis*, 80 genes; supplementary table S6, Supplementary Material online). Therefore, this loss of genes involved in DNA repair is likely one of the causes leading to the increased evolutionary rates seen in Microsporidia. A high mutation and recombination rate may be at the origin of pseudogenization and gene loss but, at the same time, offers a powerful mean to successfully cope with the arms race established with the microsporidian hosts.

Microsporidia have acquired several genes through HGT (Alsmark et al. 2013; Alexander et al. 2016). In highly reduced genomes, these genes may play important adaptive roles. A paradigmatic example is the acquisition of a bacterial ATP/ADP translocase of probable chlamydial origin, which is responsible for the import of ATP from the host (Tsaousis et al. 2008). However, we did not find this ATP/ADP translocase family or any other paralogue in the *M. incurvata* genome. The family is also missing in *Amphiamblys* sp., *M. daphniae*, and *P. saccamoebae*. Although the two metchnikovellid genomes available are not fully complete, it is unlikely that this gene was missed in the both genomes. This observation might suggest that the ATP/ADP translocase family was acquired by HGT after the divergence of Metchnikovellidae. However, since this gene family is also present in the rozellid *R. allomycis* (James et al. 2013), an alternative explanation is that the gene was acquired by a common ancestor of *R. allomycis* and Microsporidia (Dean et al. 2018), and later lost in the Metchnikovellidae.

It has been proposed that, in *Amphiamblys* sp., a gene related to the mitochondrial carrier protein family (MCF) might have evolved for nucleotide transport, playing a role in mitosomal metabolism (Mikhailov et al. 2017). This MCF gene is found in fungi (e.g., *Saccharomyces cerevisiae*) and other holomycota, and originally transported inorganic phosphate into the mitochondrion. We found the same gene in *M. incurvata*, and in fact, it is the only MCF member still retained in the metchnikovellid genomes (supplementary fig. S9, Supplementary Material online). Although we observe that the gene is also found in early-branching members that produce ATP with active mitochondria (*M. daphniae*, *P. saccamoebae*, *R. allomycis*, and the aphelid *Paraphelidium tribonemae*), it is unrelated to MCF genes found in the late-branching microsporidium *A. locustae*.

However, the MCF gene previously identified in *A. locustae* (Williams et al. 2008) derived from an EST and is not actually encoded in the available genome sequence for *A. locustae*, such that it represents a potential contaminant gene. If this is indeed the case, the MCF gene would have been retained only in metchnikovellids and basal members of the lineage and later lost in Core Microsporidia. The presence of MCF gene homologues in both *M. incurvata* and *Amphiamblys* sp. supports the idea that MCF is retained in metchnikovellids. Moreover, the *M. incurvata* contig in which the MCF gene is located has a total length of 19,629 bp and encodes 10 eukaryotic proteins. When blasted, all proteins except one (which is fragmented) have homologs in *Amphiamblys* sp. This further confirms that this MCP gene belongs to the *M. incurvata* genome and is not a contaminant. Metchnikovellids are the only Microsporidia with no functional aerobic mitochondria still conserving this gene. Although experimental and cellular localization evidence are still lacking, we might hypothesize that this MCF has become an ATP transporter in metchnikovellids, evolving its function to pump ATP from their hosts.

In addition to the canonical ATP/ADP translocase family, we have studied other 15 published cases of HGT that are widespread in different Microsporidia. We searched for these markers in 19 genomes of derived Microsporidia, metchnikovellids and basal members of the group (supplementary table S7, Supplementary Material online). In the *M. incurvata* genome we found only one out of the 15 transferred genes, the manganese superoxide dismutase (MnSOD; Xiang et al. 2010). The ML tree of this protein including members of bacteria, Microsporidia, and various other eukaryotes supports several independent acquisitions of the MnSOD gene by members of Microsporidia and Rozellida from various bacterial donors (supplementary fig. S10, Supplementary Material online). These transfers affect the Metchnikovellidae (both *M. incurvata* and *Amphiamblys* sp.), the microsporidium *A. locustae*, and the remaining derived Microsporidia. Independent acquisitions of the MnSOD gene occur also in aphelids (*P. tribonemae*; Torruella et al. submitted) and the two parasitic anaerobic gut fungi *Piromyces finnis* and *Anaeromyces robustus* (supplementary fig. S10, Supplementary Material online). These multiple acquisitions likely reflect an important adaptive function. In fact, the MnSOD gene seems to play a key role in protecting anaerobic life from the well-known deleterious effects of oxygen (Holley et al. 2011). Several examples, mainly in bacteria and yeast, have shown that cells expressing MnSOD as a result of stimulation with high oxygen levels were more resistant to hyperbaric oxygen concentrations (Gregory

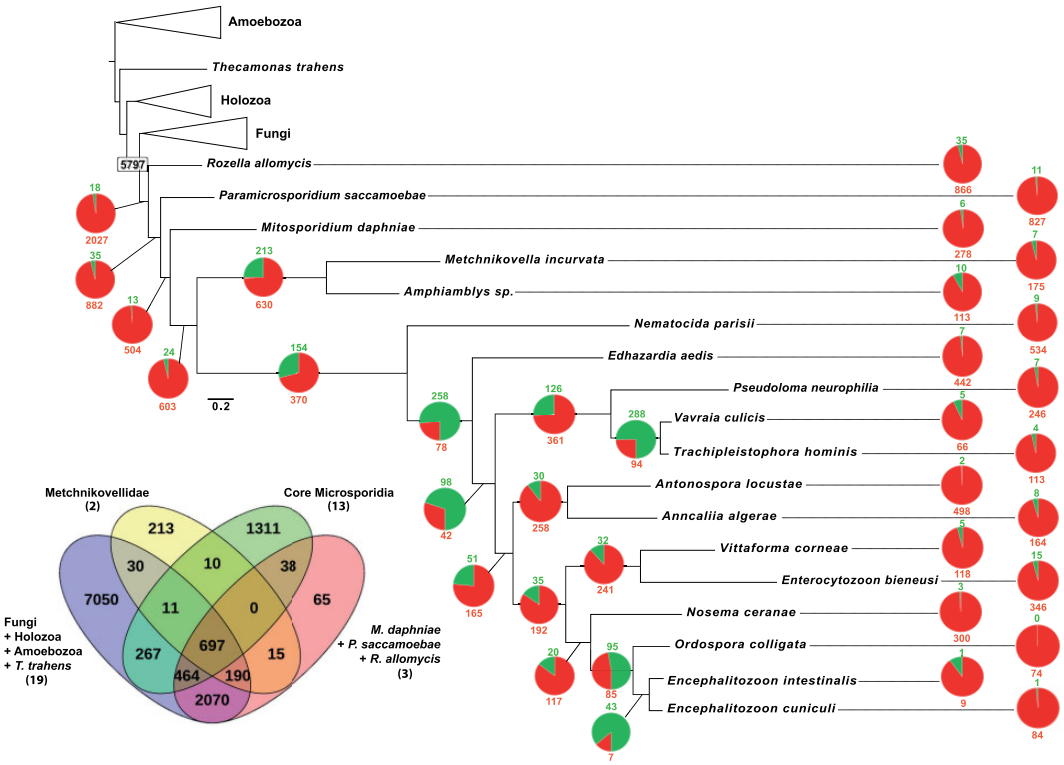


Fig. 4.—Gain and loss of protein orthogroups along the evolutionary lineage of Microsporidia, based on a BI phylogeny from the multigene data set. Pie charts at nodes represent the total gains (green) and losses (red) shown respectively in numbers above and below the pie charts. The deepest node indicates the estimated number of ancestral protein orthogroups (5, 797). The Venn diagram represents the total number of protein orthogroups shared between 37 opisthokont proteomes grouped into metchnikovellids (2 proteomes), Microsporidia (13 proteomes), Fungi + Holozoa + Amoebozoa + *T. trahens* (19 proteomes) and *M. daphniae* + *P. saccamoebae* + *R. allomycis* (3 proteomes).

and Fridovich 1973; Gregory et al. 1974). The recurrent acquisition of the MnSOD gene in Microsporidia might be therefore linked to cell protection against reactive oxygen species.

To have a global overview on the genome evolution in metchnikovellids and in other Microsporidia, we carried out a gain and loss analysis using the reconstructed phylogeny of Microsporidia as a backbone and applying Dollo parsimony on protein orthogroups (fig. 4; supplementary tables S8–S10, Supplementary Material online). As shown in figure 4, an important proportion of orthogroups were lost at the beginning of the diversification of a clade comprising *R. allomycis*, *P. saccamoebae* and the mitochondriate microsporidium *M. daphniae*. This early genome reduction undergone by Microsporidia might relate with the transition towards an obligatory intracellular parasite lifestyle (Heinz et al. 2012; Nakjang et al. 2013). Metchnikovellids occupy an intermediate position along the microsporidian branch. We found 213 gains of specific orthogroups in these organisms, including 561 proteins, of which 251 correspond to *Amphiamblys* sp. and 310 to *M. incurvata*. HMMR searches using EggNOG v4.5 for the 561 metchnikovellid proteins retrieved homologues only for 11 *Amphiamblys* sp. and 15 *M. incurvata* proteins in this database. These proteins included 17 orthogroups comprising 37 proteins (11 of them did not return positive hits; supplementary table S11, Supplementary Material online), which likely correspond to genes retained from the common Microsporidia ancestor. However, we found 524 proteins present in both *M. incurvata* and *Amphiamblys* sp. grouped in 196 orthogroups without relatives in other lineages. These genes were probably gained by metchnikovellids (supplementary table S120, Supplementary Material online), implying that soon after the important gene loss experienced by the microsporidian ancestors, gene gain started to overcome gene loss, coinciding with the diversification of long-branch Microsporidia. Indeed, there is also a remarkable process of gene gain at the node leading to one of the most derived lineages of Microsporidia, comprising *Ordospora colligata* and the *Encephalitozoon* clade. This clade comprises Microsporidia with some of the most simplified genomes (Corradi et al. 2010) suggesting that genome reduction and evolution of new markers have co-occurred during the adaptation of Microsporidia to their hosts. However, we cannot completely exclude the possibility that gene losses and gains actually correspond to genes that have evolved beyond recognition (even if some are likely true new genes and clearly adaptive). Collectively, these observations of gene gain reflect

specialization to different hosts, likely achieved by the acquisition of some adaptive genes. Thus, as derived Microsporidia diversified, gene losses started to overcome gains and they became progressively more adapted to their parasitic lifestyle.

Conclusions

Phylogenomic analyses using data from the *M. incurvata* genome have confirmed that the Metchnikovellidae are a deep-branching group inside the Microsporidia, and the deepest of those without functional mitochondria. Comparative analysis of the two available metchnikovellid genomes confirmed that gene complement resembles more to those of typical derived Microsporidia than to those of the less derived mitochondriate *M. daphniae*, *P. saccamoebae*, and *R. allomycis*. We observed a reduction of DNA repair pathways, which seems to correlate with the high evolutionary rates seen in the clade. Interestingly, the typical microsporidian ATP/ADP translocase family does not seem to be present in any of the two metchnikovellid genomes. We hypothesize that this gene was replaced by a MCF gene that became an ATP transporter. Lastly, our gain and loss analysis suggests that reductive evolution is not the only ongoing process in Microsporidia and that the evolution of new genes has also taken place during the adaptation of Microsporidia to their hosts.

Both phylogenomic and comparative genomic analyses rise the need for taxonomic revision of Microsporidia and Rozellida, since the boundaries between them are blurry. For this, is essential to continue with the surveys and sequencing efforts for new members of the group, which will help to fill the gap in knowledge still present in the evolution of the clade.

Supplementary Material

Supplementary data are available at *Genome Biology and Evolution* online.

Acknowledgments

This work was funded by the European Research Council Advanced Grant "ProtistWorld" (No. 322669) and the Horizon 2020 research and innovation program under the Marie Skłodowska-Curie ITN project SINGEK (http://www.sin-gek.eu/; grant agreement No. H2020-MSCA-ITN-2015-675752). It also received support from the RFBR grant 18-04-01359. G.T. was financed by the European Marie Sklodowska-Curie Action (704566 AlgDates). We thank the Core Facility Centers "Development of molecular and cell technologies," "Culturing of microorganisms," and the Educational Research Station "Belomorskaia" of Saint-Petersburg State University for access to their equipment facilities.

Literature Cited

Alexander WG, Wisecaver JH, Rokas A, Hittinger CT. 2016. Horizontally acquired genes in early-diverging pathogenic fungi enable the use of host nucleosides and nucleotides. Proc Natl Acad Sci USA. 113(15):4116–4121.

Alsmark C, et al. 2013. Patterns of prokaryotic lateral gene transfers affecting parasitic microbial eukaryotes. Genome Biol. 14(2):R19.

Altschul SF, et al. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. 25(17):3389–3402.

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. J Mol Biol. 215(3):403–410.

Amyotte SG, et al. 2012. Transposable elements in phytopathogenic *Verticillium* spp.: insights into genome evolution and inter- and intra-specific diversification. BMC Genomics. 13(1):314.

Andrews S. 2010. FastQC: a quality control tool for high throughput sequence data. http://www.bioinformatics.babraham.ac.uk/projects/.

Bankevich A, et al. 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J Comput Biol. 19(5):455–477.

Bass D, et al. 2018. Clarifying the relationships between microsporidia and cryptomycota. J Eukaryot Microbiol. doi: 10.1111/jeu.12519. [Epub ahead of print]

Besemer J, Lomsadze A, Borodovsky M. 2001. GeneMarkS: a self-training method for prediction of gene starts in microbial genomes. Implications for finding sequence motifs in regulatory regions. Nucleic Acids Res. 29:2607–2618.

Beznoussenko GV, et al. 2007. Analogs of the Golgi complex in microsporidia: structure and avesicular mechanisms of function. J Cell Sci. 120:1288–1298.

Biémont C. 2010. A brief history of the status of transposable elements: from junk DNA to major players in evolution. Genetics 186(4):1085–1093.

Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30(15):2114–2120.

Boniecki MT, Freibert SA, Mühlenhoff U, Lill R, Cygler M. 2017. Structure and functional dynamics of the mitochondrial Fe/S cluster synthesis complex. Nat Commun. 8(1):1287.

Boucher JI, Jacobowitz JR, Beckett BC, Classen S, Theobald DL. 2014. An atomic-resolution view of neofunctionalization in the evolution of apicomplexan lactate dehydrogenases. Elife 3:02304.

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

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.

Caullery M, Mesnil F. 1914. Metchnikovellidae et autres Protistes parasites des Grégarines d'Annélides. CR Soc Biol. 77:527–532.

Corradi N, Pombert JF, Farinelli L, Didier ES, Keeling PJ. 2010. The complete sequence of the smallest known nuclear genome from the microsporidian *Encephalitozoon intestinalis*. Nat Commun. 1(6):1.

Corsaro D, et al. 2014. Microsporidia-like parasites of amoebae belong to the early fungal lineage Rozellomycota. Parasitol Res. 113(5):1909–1918.

Corsaro D, et al. 2016. Molecular identification of *Nucleophaga terricolae* sp. nov. (Rozellomycota), and new insights on the origin of the Microsporidia. Parasitol Res. 115(8):3003–3011.

Csurös M. 2010. Count: evolutionary analysis of phylogenetic profiles with parsimony and likelihood. Bioinformatics 26(15): 1910–1912.

Dean RA, et al. 2005. The genome sequence of the rice blast fungus *Magnaporthe grisea*. Nature 434(7036):980–986.

Dean P, et al. 2018. Transporter gene acquisition and innovation in the evolution of Microsporidia intracellular parasites. Nat Commun. 9(1):1709.
Desjardins CA, et al. 2015. Contrasting host-pathogen interactions and genome evolution in two generalist and specialist microsporidian pathogens of mosquitoes. Nat Commun. 6:7121.
Dewannieux M, Heidmann T. 2005. LINEs, SINEs and processed pseudogenes: parasitic strategies for genome modeling. Cytogenet Genome Res. 110(1–4):35–48.
Didier ES, et al. 2004. Epidemiology of microsporidiosis: sources and modes of transmission. Vet Parasitol. 126(1–2):145–166.
Didier ES, Weiss LM. 2006. Microsporidiosis: current status. Curr Opin Infect Dis. 19(5):485–492.
Didier ES, Weiss LM. 2011. Microsporidiosis: not just in AIDS patients. Curr Opin Infect Dis. 24(5):490–495.
Dolgikh VV, et al. 2009. Heterologous expression of pyruvate dehydrogenase E1 subunits of the microsporidium *Paranosema* (Antonospora) *locustae* and immunolocalization of the mitochondrial protein in amitochondrial cells. FEMS Microbiol Lett. 293(2):285–291.
Embley TM, Martin W. 2006. Eukaryotic evolution, changes and challenges. Nature 440(7084):623–630.
Emelyanov VV. 2003. Phylogenetic affinity of a *Giardia lamblia* cysteine desulfurase conforms to canonical pattern of mitochondrial ancestry. FEMS Microbiol Lett. 226(2):257–266.
Emms DM, Kelly S. 2015. OrthoFinder: solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. Genome Biol. 16:157.
Felsenstein J. 1981. Evolutionary trees from DNA sequences: a maximum likelihood approach. J Mol Evol. 17(6):368–376.
Finn RD, Clements J, Eddy SR. 2011. HMMER web server: interactive sequence similarity searching. Nucleic Acids Res. 39(Suppl.):W29.
Freibert SA, et al. 2017. Evolutionary conservation and in vitro reconstitution of microsporidian iron-sulfur cluster biosynthesis. Nat Commun. 8:13932.
Gregory EM, Fridovich I. 1973. Oxygen toxicity and the superoxide dismutase. J Bacteriol. 114(3):1193–1197.
Gregory EM, Goscin SA, Fridovich I. 1974. Superoxide dismutase and oxygen toxicity in a eukaryote. J Bacteriol. 117(2):456–460.
Gurevich A, Saveliev V, Vyahhi N, Tesler G. 2013. QUAST: quality assessment tool for genome assemblies. Bioinformatics 29(8):1072–1075.
Haag KL, et al. 2014. Evolution of a morphological novelty occurred before genome compaction in a lineage of extreme parasites. Proc Natl Acad Sci USA. 111(43):15480–15485.
Heinz E, et al. 2012. The genome of the obligate intracellular parasite *Trachipleistophora hominis:* new insights into microsporidian genome dynamics and reductive evolution. PLoS Pathog. 8:e1002979.
Hirt RP, Healy B, Vossbrinck CR, Canning EU, Embley TM. 1997. A mitochondrial Hsp70 orthologue in Vairimorpha necatrix: molecular evidence that microsporidia once contained mitochondria. Curr Biol. 7(12):995–998.
Holley AK, Bakthavatchalu V, Velez-Roman JM, St. Clair DK. 2011. Manganese superoxide dismutase: guardian of the powerhouse. Int J Mol Sci. 12(10):7114–7162.
Huelsenbeck JP, Ronquist F. 2001. MR BAYES: bayesian inference of phylogenetic trees. Bioinformatics 17(8):754–755.
Huerta-Cepas J, et al. 2017. Fast genome-wide functional annotation through orthology assignment by eggNOG-mapper. Mol Biol Evol. 34(8):2115–2122.
James TY, et al. 2006. Reconstructing the early evolution of Fungi using a six-gene phylogeny. Nature 443(7113):818–822.
James TY, et al. 2013. Shared signatures of parasitism and phylogenomics unite cryptomycota and microsporidia. Curr Biol. 23(16):1548–1553.
Jones MDM, Richards TA, Hawksworth DL, Bass D. 2011. Validation and justification of the phylum name Cryptomycota phyl. nov. IMA Fungus. 2(2):173–175.
Jurka J. 2004. Evolutionary impact of human Alu repetitive elements. Curr Opin Genet Dev. 14(6):603–608.
Kamaishi T, et al. 1996. Complete nucleotide sequences of the genes encoding translation elongation factors 1 and 2 from a microsporidian parasite, *Glugea plecoglossi*: implications for the deepest branching of eukaryotes. J Biochem. 120(6):1095–1103.
Karpov SA, Torruella G, Moreira D, Mamkaeva MA, López-García P. 2017. Molecular phylogeny of *Paraphelidium letcheri* sp. nov. (Aphelida, Opisthosporidia). J Eukaryot Microbiol. 64(5):573–578.
Katoh K, Standley DM. 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol Biol Evol. 30(4):772–780.
Kearse M, et al. 2012. Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics 28(12):1647–1649.
Kispal G, et al. 2005. Biogenesis of cytosolic ribosomes requires the essential iron–sulphur protein Rli1p and mitochondria. EMBO J. 24(3):589–598.
Kumar S, Jones M, Koutsovoulos G, Clarke M, Blaxter M. 2013. Blobology: exploring raw genome data for contaminants, symbionts, and parasites using taxon-annotated GC-coverage plots. Front Genet. 4:1–12.
Laetsch DR, Blaxter ML. 2017. BlobTools: interrogation of genome assemblies. F1000Research 6:1287.
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, Brinkmann H, Philippe H. 2007. Suppression of long-branch attraction artefacts in the animal phylogeny using a site-heterogeneous model. BMC Evol Biol. 7(Suppl. 1):S4.
Lartillot N, Philippe H. 2004. A Bayesian mixture model for across-site heterogeneities in the amino-acid replacement process. Mol Biol Evol. 21(6):1095–1109.
Lartillot N, Philippe H. 2006. Computing Bayes factors using thermodynamic integration. Syst Biol. 55(2):195–207.
Leipe DD, Gunderson JH, Nerad TA, Sogin ML. 1993. Small subunit ribosomal RNA+ of Hexamita inflata and the quest for the first branch in the eukaryotic tree. Mol Biochem Parasitol. 59(1):41–48.
Lynch M, Conery JS. 2000. The evolutionary fate and consequences of duplicate genes. Science (80-) 290(5494):1151–1155.
Marcet-Houben M, Gabaldón T. 2010. Acquisition of prokaryotic genes by fungal genomes. Trends Genet. 26(1):5–8.
Mikhailov KV, Simdyanov TG, Aleoshin VV. 2017. Genomic survey of a hyperparasitic microsporidian *Amphiamblys* sp. (Metchnikovellidae). Genome Biol Evol. 9:454–467.
Nakjang S, et al. 2013. Reduction and expansion in microsporidian genome evolution: new insights from comparative genomics. Genome Biol Evol. 5(12):2285–2303.
Nguyen LT, Schmidt HA, Von Haeseler A, Minh BQ. 2015. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. Mol Biol Evol. 32(1):268–274.
Okonechnikov K, Conesa A, García-Alcalde F. 2015. Qualimap 2: advanced multi-sample quality control for high-throughput sequencing data. Bioinformatics 32(2):292–294.
Panek J, et al. 2014. Hijacking of host cellular functions by an intracellular parasite, the microsporidian *Anncaliia algerae*. PLoS ONE. 9(6):e100791.
Parisot N, et al. 2014. Microsporidian genomes harbor a diverse array of transposable elements that demonstrate an ancestry of horizontal exchange with metazoans. Genome Biol Evol. 6(9):2289–2300.

Philippe H, et al. 2000. Early-branching or fast-evolving eukaryotes? An answer based on slowly evolving positions. Proc R Soc B Biol Sci. 267(1449):1213–1221.
Pombert J-F, et al. 2012. Gain and loss of multiple functionally related, horizontally transferred genes in the reduced genomes of two microsporidian parasites. Proc Natl Acad Sci USA. 109(31):12638–12643.
Posada D. 2008. jModelTest: phylogenetic model averaging. Mol Biol Evol. 25(7):1253–1256.
Quandt CA, et al. 2017. The genome of an intranuclear parasite, *Paramicrosporidium saccamoebae*, reveals alternative adaptations to obligate intracellular parasitism. Elife 6:pii: e29594.
R Core Team (2014). R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing. URL http://www.R-project.org/.
Raffaele S, Kamoun S. 2012. Genome evolution in filamentous plant pathogens: why bigger can be better. Nat Rev Microbiol. 10(6):417–430.
Rambaut A. 2016. FigTree v1.4.3. Mol. Evol. phylogenet ics Epidemiol.
Rotari YM, Paskerova GG, Sokolova YY. 2015. Diversity of metchnikovellids (Metchnikovellidae, Rudimicrosporea), hyperparasites of bristle worms (Annelida, Polychaeta) from the White Sea. Protistology 9:50–59.
Simão FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM. 2015. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformatics 31(19):3210–3212.
Smit A, Hubley R, Green P. 1996. RepeatMasker Open-3.0. RepeatMasker Open-3.0. www.repeatmasker.org.
Smit A, Hubley R, Green P. 2013. RepeatMasker Open-4.0. 2013–2015. http://repeatmasker.org.
Sokolova YY, Paskerova GG, Rotari YM, Nassonova ES, Smirnov AV. 2013. Fine structure of *Metchnikovella incurvata* Caullery and Mesnil 1914 (microsporidia), a hyperparasite of gregarines Polyrhabdina sp. from the polychaete *Pygospio elegans*. Parasitology 140(7): 855–867.
Sprague V. 1977. In: Bulla LA, Cheng TC, editors. Annotated list of species of microsporidia BT—comparative pathobiology: volume 2 systematics of the Microsporidia. Springer US: Boston, MA. p. 31–334.
Stanke M, Morgenstern B. 2005. AUGUSTUS: a web server for gene prediction in eukaryotes that allows user-defined constraints. Nucleic Acids Res. 33(Web Server):W465.
Tanaka N, Meineke B, Shuman S. 2011. RtcB, a novel RNA ligase, can catalyze tRNA splicing and HAC1 mRNA splicing in vivo. J Biol Chem. 286(35):30253–30257.
Tanaka N, Shuman S. 2011. RtcB is the RNA ligase component of an *Escherichia coli* RNA repair operon. J Biol Chem. 286(10):7727–7731.
Thomarat F, Vivarès CP, Gouy M. 2004. Phylogenetic analysis of the complete genome sequence of Encephalitozoon cuniculi supports the fungal origin of microsporidia and reveals a high frequency of fast-evolving genes. J Mol Evol. 59(6):780–791.
Torruella G, et al. 2012. Phylogenetic relationships within the Opisthokonta based on phylogenomic analyses of conserved single-copy protein domains. Mol Biol Evol. 29(2):531–544.
Tsaousis AD, et al. 2008. A novel route for ATP acquisition by the remnant mitochondria of *Encephalitozoon cuniculi*. Nature 453(7194):553–556.
Vávra J, Lukeš J. 2013. Microsporidia and 'the art of living together'. Adv Parasitol. 82:253–319.
Vivier E. 1975. The microsporidia of the protozoa. Parazitologica 11:345–361.
Waller RF, et al. 2009. Evidence of a reduced and modified mitochondrial protein import apparatus in microsporidian mitosomes. Eukaryot Cell. 8(1):19–26.
Wilks HM, et al. 1988. A specific, highly active malate dehydrogenase by redesign of a lactate dehydrogenase framework. Science 242(4885):1541–1544.
Williams BAP, et al. 2010. A broad distribution of the alternative oxidase in microsporidian parasites. PLoS Pathog. 6(2):e1000761.
Williams BAP, Haferkamp I, Keeling PJ. 2008. An ADP/ATP-specific mitochondrial carrier protein in the microsporidian *Antonospora locustae*. J Mol Biol. 375(5):1249–1257.
Williams BAP, Hirt RP, Lucocq JM, Embley TM. 2002. A mitochondrial remnant in the microsporidian *Trachipleistophora hominis*. Nature 418(6900):865–869.
Wittner M, Weiss LM. eds. 1999. The microsporidia and microsporidiosis. Washington, D.C: American Society of Microbiology.
Xiang H, et al. 2010. A tandem duplication of manganese superoxide dismutase in *Nosema bombycis* and its evolutionary origins. J Mol Evol. 71(5–6):401–414.

Associate editor: Martin Embley

6. A new fungal clade helps reconstructing the tree of Fungi and the evolution of the flagellum in Holomycota

“The more I thought over it the more I became convinced that I had at length found the long-sought-for law of nature that solved the problem of the origin of species. For the next hour I thought over the deficiencies in the theories of Lamarck and of the author of the "Vestiges," and I saw that my new theory supplemented these views and obviated every important difficulty. I waited anxiously for the termination of my fit so that I might at once make notes for a paper on the subject. The same evening I did this pretty fully, and on the two succeeding evenings wrote it out carefully in order to send it to Darwin by the next post, which would leave in a day or two."

Alfred Russel Wallace. My Life (1905)

6. A new fungal clade helps reconstructing the tree of Fungi and the evolution of the flagellum in Holomycota

6.1. Context and objectives

The last large group on the Holomycota branch corresponds to the well-known Fungi. The deepest branches of the fungal tree are composed of the two zoosporic lineages Blastocladiomycota and Chytridiomycota. Not only the relationship between these groups remains uncertain, but also the relationships of several *incertae sedis* groups, for which only 18S rRNA gene data is often available (see chapter 1.7.4. *Incertae sedis* lineages). One of these groups is the Sanchytriaceae, a clade composed of two known species, *Amoeboradix gromovi* and *Sanchytrium tribonematis*. Sanchytrids are highly atypical zoosporic fungi, having a highly reduced flagellum ultrastructure, which seems not functional, and at the same time one of the longest kinetosomes known in eukaryotes. The two sanchytrid species were known to branch together but their affinity with any other fungal lineage remained uncertain. *Olpidium* is another zoosporic genus with an unresolved phylogeny. 18S rRNA gene trees including *Olpidium* seem to indicate that it branches within the non-flagellated Zoopagomycota, although with poor support.

Recently we managed to establish cultures of the two sanchytrids species, allowing us to obtain genome sequence data for the clade. We then set the following objectives:

1) Reconstruct the phylogenetic relationships of sanchytrids and other zoosporic clades within Fungi. By obtaining genomic data from these two sanchytrid species we could reconstruct a phylogenomic tree of Holomycota to try to resolve the position of sanchytrids. Additionally, by including the available genomic data of *Olpidium bornovanus*, we aimed at resolving its branching position. The addition of sanchytrids, *Olpidium* and a broad fungal taxon sampling was used to clarify the branching order between chytrids and Blastocladiomycota.

2) Compare the genomes of sanchytrids, other zoosporic fungi and other holomycotan representatives. This allowed estimating the number of independent flagellar losses in Holomycota and comparing the primary metabolism of these organisms. Finally, we wanted to understand the molecular determinants of the atypical reduced flagellum with a long kinetosome that sanchytrids possess.

6.2. Results

Our phylogenomic analyses placed sanchytrids as the sister lineage of Blastocladiomycota in a well-supported clade. Intriguingly, sanchytrids exhibited a long branch, indicating a fast-evolving genome. Phylogenomic analyses and further tests of the phylogenetic position of chytrids also indicated that chytrids could be the sister lineage to all other fungi. Additionally, for the first time we confirmed in a phylogenomic framework that the zoosporic fungus *Olpidium* is part of a new lineage sister of the major clade of non-flagellated fungi and that it forms its own phylum Olpidiomycota.

Our assessment using COG categories of the primary metabolic capacities of Holomycota showed a highly atypical metabolism in sanchytrids compared with “canonical” fungi. Comparative analyses of the flagellar toolkit indicated that the sanchytrid genomes lacked complete sets of proteins involved in flagellar function and maintenance. These analyses indicated under our current taxon sampling four independent flagellum losses in Fungi.

Finally, sanchytrids possessed the *BeGC1* fusion gene, the *BeCNG1* gene and a lipid intracellular organelle that could indicate the presence of a light sensing pathway shared with Blastocladiomycota. If confirmed, this could explain the presence of the long sanchytrid kinetosome as a support structure for a lipid eye-spot.

6.3. Manuscript of article 3

A new fungal clade helps reconstructing the tree of Fungi and the evolution of the flagellum in Holomycota

(Manuscript in preparation)

Luis Javier Galindo[1], Guifré Torruella[1], Purificación López-García[1], Sergey Karpov[2] and David Moreira[1]

[1]Ecologie Systématique Evolution, CNRS, Université Paris-Sud, AgroParisTech, Université Paris-Saclay, Orsay, France.

[2]Department of Invertebrate Zoology, Faculty of Biology, St Petersburg State University, Russia

* Corresponding authors

E-mails: david.moreira@u-psud.fr; luis.galindo@u-psud.fr.

A new fungal clade helps reconstructing the tree of Fungi and the evolution of the flagellum in Holomycota

Luis Javier Galindo[1], Guifré Torruella[1], Purificación López-García[1], Sergey Karpov[2] and David Moreira[1]

[1]Ecologie Systématique Evolution, CNRS, Université Paris-Sud, AgroParisTech, Université Paris-Saclay, Orsay, France.

[2]Department of Invertebrate Zoology, Faculty of Biology, St Petersburg State University, Russia

* Corresponding authors

E-mails: david.moreira@u-psud.fr; luis.galindo@u-psud.fr.

Abstract

While the majority of known fungi are multicellular (or secondarily unicellular), unicellular species are far from rare. Most unicellular clades have zoosporic free-living flagellated stages and the phylogenetic relationship among the main groups remain unresolved. Among the zoosporic fungi, *Amoeboradix gromovi* and *Sanchytrium tribonematis* (Sanchytriaceae) are two parasitic species with unique features. Although their flagella are structurally reduced, their amoeboid zoospores have some of the longest kinetosomes known in eukaryotes, an extremely unusual feature in eukaryotic cell biology. Molecular phylogenetic analyses of 18S+28S rRNA genes revealed that both species are closely related. However, they did not show any affinity with any other described fungal clade. To assess their phylogenetic position and look into their unique features, we obtained single-cell genomic data for both species. Using a dataset of 264 protein alignments for 84 and 69 species, we show that sanchytrids form a well-supported fast-evolving clade sister to the Blastocladiomycota. Our results also support a more stable fungal global phylogeny in which Chytridiomycota branch as the sister lineage to the rest of fungi and the zoosporic fungi *Olpidium bornovanus* as the sister lineage to non-flagellated fungi. Comparative genomic analyses of several holomycotan genomes proved that the overall metabolic repertoire of sanchytrids is atypically reduced given their placement in the tree. The study of more than 60 flagellum-specific proteins in flagellated and non-flagellated species allowed us to retrace the evolution of the flagellum in fungi. We infer 4 independent flagellum losses in Holomycota. In particular, sanchytrids have a highly reduced flagellar toolkit and the maintenance of this reduced flagellum is most likely related with the possible presence of a light-sensing lipid eyespot supported by the long kinetosome.

Key words: Fungi, Holomycota, phylogenomics, sanchytrids, flagella

Introduction

The Opisthokonta, is one of the main eukaryotic supergroups, containing the well-known multicellular animals and fungi which, together with their unicellular relatives, form the two major branches Holozoa and Holomycota, respectively (Cavalier-Smith, 1998; Brown, Spiegel, & Silberman, 2009; Liu *et al.*, 2009). According to molecular dating, the first opisthokonts evolved between 1 and 1.5 billion years ago (Douzery *et al.*, 2004; Parfrey *et al.*, 2011; Eme *et al.*, 2014) and most likely already possessed a single posterior flagellum to propel them in aquatic environments. This character was also present in the first holomycotan species, which probably evolved around the same time (Loron *et al.*, 2019), and has been retained in most modern fungal lineages at least at some stages of their life cycle (Berbee, James, & Strullu-Derrien, 2017; Spatafora *et al.*, 2017). The free-living non-flagellated nucleariid amoebae were the first lineage to diverge within the Holomycota, followed by several lineages of endoparasitic taxa, collectively known as Opisthosporidia (Karpov *et al.*, 2014): the flagellated Rozellida or Cryptomycota (Lara, Moreira, & López-García, 2010; Jones *et al.*, 2011; James *et al.*, 2013) and Aphelida (Karpov *et al.*, 2014) and the highly reduced non-flagellated Microsporidia (Haag *et al.*, 2014; Bass *et al.*, 2018). Opisthosporidia is now suggested to be paraphyletic (Torruella *et al.*, 2018) with aphelids branching as the sister lineage to all fungi. Among fungi, with the exception of the secondary loss of the flagellum in the chytrid *Hyaloraphydium curvatum* (Ustinova, Krienitz, & Huss, 2000), all known early divergent fungal taxa are zoosporic, namely with at least one flagellated stage during their life cycle.

Zoosporic fungi have been found to be ubiquitous in all types of habitats from tropical to Artic regions (Powell, 1993; Freeman *et al.*, 2009). They thrive both in marine and freshwater systems, where they are specially diverse as parasites and take part in nutrient recycling through the "mycoloop" (Kagami, Miki, & Takimoto, 2014; Frenken *et al.*, 2016, 2017), as well as in soils, where they are saprotrophs and obligate parasites (Tedersoo *et al.*, 2017). Originally part of the same linage, zoosporic fungi have been recently classified into the two major groups Blastocladiomycota and Chytridiomycota based on few-genes molecular phylogenies (James *et al.*, 2006b) later confirmed by multi-gene studies (Chang *et al.*, 2015; Torruella *et al.*, 2018). Blastocladiomycota and Chytridiomycota are sister lineages to the three main lineages of non-flagellated fungi (Zoopagomycota, Mucoromycota and Dikarya), for which a single ancestral loss

of the flagellum has been proposed (Liu, Hodson, & Hall, 2006). The characterization of zoosporic fungi and the loss of the flagellum are key to understand the evolutionary changes that accompanied land colonization by fungi and their adaptation to the currently plant-dominated terrestrial ecosystems (Bidartondo *et al.*, 2011; Lutzoni *et al.*, 2018). A robust phylogeny of Fungi and, most specifically, resolving the relationships between zoosporic lineages, are necessary for this task. However, the later has proven to be a difficult task since both Blastocladiomycota and Chytridiomycota have been alternatively recovered as sister group of the non-flagellated fungi in phylogenomic studies (James *et al.*, 2006b; Sekimoto *et al.*, 2011; Spatafora *et al.*, 2016). This lack of resolution may derive from the old age of these splits, estimated between 0.5 and 1 billion years (Taylor & Berbee, 2006; Parfrey *et al.*, 2011; Loron *et al.*, 2019), and the existence of several rapid evolutionary radiation events in the history of fungi, including during their transition to land with plants (Ebersberger *et al.*, 2012; Lutzoni *et al.*, 2018). This has led to a low phylogenetic signal on early nodes, in particular those concerning the diversification of zoosporic fungi (Chang *et al.*, 2015).

Because of this phylogenetic uncertainty, the total number of flagellum losses in fungi remains under debate, with estimates ranging between four and six for the whole Holomycota clade (James *et al.*, 2006a). Sampling new key divergent zoosporic fungi can help to increase the phylogenetic signal on these nodes. One of such organisms is *Olpidium*, a morphologically reduced parasite of plant roots, nematodes and rotifers (Barr, 1980; Powell & Letcher, 2014). Although it has been associated with the non-flagellated Zoopagomycota by few-genes phylogenies (James *et al.*, 2006a; Sekimoto *et al.*, 2011; Tedersoo *et al.*, 2018), *Olpidium* has an unclear phylogenetic placement that has a strong impact on the inferred number of independent flagellum losses in Holomycota. Sanchytriaceae, a group of chytrid-like algal parasites represented by the genera *Amoeboradix* and *Sanchytrium*, are a second of zoosporic fungi with unclear phylogenetic placement (Karpov *et al.*, 2018). Sanchytrids exhibit some atypical morphological traits, in particular a highly reduced flagellum with an extremely long kinetosome (Karpov *et al.*, 2018, 2019), which might help to shed some light into the process of flagellum reduction and loss in Holomycota.

To both increase the phylogenetic signal for early fungal nodes and to clarify the number of independent flagellum losses, we generated genomic data for the sanchytrids *Amoeboradix gromovi* and *Sanchytrium tribonematis*. We analysed these genomes together with the available

genomic and transcriptomic data for key zoosporic groups, including *Olpidium bornovanus* and several members of Chytridiomycota and Blastocladiomycota, in a multi-gene phylogenomic framework. We showed that sanchytrids belong to a new fast-evolving lineage sister to all Blastocladiomycota and that *Olpidium* forms part of a new and independent lineage sister to all other non-flagellated fungi. We also recovered evidence for the placement of the root of the fungal tree between chytrids and all other fungi. Comparison of protein sets involved in metabolism and in the flagellar toolkit showed reduced metabolic capabilities in sanchytrids and confirmed that they have a non-motile flagellum. The maintenance of this reduced flagellum in sanchytrids is most likely related with the presence of a light-sensing lipid eyespot supported by the long kinetosome. Our new phylogenomic framework of Fungi supports a conservative model of four flagellum losses during the evolution of the Holomycota.

Methods

Biological material. *Sanchytrium tribonematis* strain X-128 and *Amoeboradix gromovi* strain X-113, isolated from freshwater sampling locations in Russia (Karpov *et al.*, 2018, 2019), were maintained in culture with the freshwater yellow-green alga *Tribonema gayanum* Pasch. strain 20 CALU as host as described in Karpov et al. (2017). The algal host was grown in mineral freshwater medium at room temperature under white light. After inoculation with *Sanchytrium* or *Amoeboradix*, cultures were incubated for 1–2 weeks to reach a maximum infection level. We then collected both individual zoospores and sporangia full of moving zoospores by micromanipulation with an Eppendorf PatchMan NP2 micromanipulator using 19 μm VacuTip microcapillaries (Eppendorf) on an inverted Leica Dlll3000 B microscope. Zoospores and sporangia were then washed 2 times in clean sterile water drops before storing them into individual tubes for further analyses.

Genome amplification and sequencing. DNA extraction from single-isolated zoospores and sporangia was done with the PicoPure kit (Thermo Fisher Scientific) according to the manufacturer's protocol. Whole genome amplification (WGA) was carried on the extracted DNA by multiple displacement amplification (MDA) with the single-cell REPLI-g kit (QIAGEN). Sanchytrid DNA amplification was assessed by DNA quantification using Qubit fluorometric

quantification (Life Technologies) and PCR amplification and Sanger sequencing of the 18S rDNA gene. We only kept amplified genomes that yielded high DNA concentration and the expected sanchytrid 18S rDNA amplicons. As expected, WGA from sporangia performed better (due to high zoospore concentration) than individual zoospores and were selected for sequencing. We selected for sequencing the MDA samples K1-9_WGA (*A. gromovi*) and SC-2_WGA (*S. tribonematis*), both with similar DNA concentrations of 152 and 160 ng/µl, respectively. Two TruSeq paired-end single-cell libraries were prepared from these samples and sequenced on a HiSeq 2500 Illumina instrument (2 x 100 bp) chemistry v4. We obtained 121,233,342 reads for a total of 26,245 Mbp for *A. gromovi* and 106,922,235 reads for a total of 21,384 Mbp for *S. tribonematis*.

Genome sequence assembly, decontamination and annotation. Paired-end read quality was assessed with FastQC (Andrews, 2010) before and after quality trimming. We then trimmed the Illumina adapters with Trimmomatic v0.32 in Paired End mode (Bolger, Lohse, & Usadel, 2014), with the following parameters: ILLUMINACLIP:adapters.fasta:2:30:10 LEADING:28 TRAILING:28 SLIDINGWINDOW:4:30. Trimmed paired-end reads were assembled using SPAdes 3.9.1 in single-cell mode (Bankevich *et al.*, 2012). This produced assemblies of 48.7 and 37.1 Mb with 8,420 and 8,015 contigs of prokaryotic and eukaryotic origin for *A. gromovi* and *S. tribonematis*, respectively. The decontamination of the two genomes was carried out by a three-step process. First, the genome sequences were subjected to two rounds of assembly, before and after bacterial sequence removal with BlobTools v0.9.19 (Laetsch & Blaxter, 2017). Second, open-reading frames were predicted and translated from the assembled contigs using Transdecoder v2 (http:transdecoder.github.io) with default parameters and Cd-hit v4.6 (Li & Godzik, 2006) with 100% identity to produce protein sequences for *A. gromovi* and *S. tribonematis*. Finally, to remove possible eukaryotic host (*Tribonema*) contamination, the predicted protein sequences were searched by BLASTp (Camacho *et al.*, 2009) against two predicted yellow-green algae proteomes: the proteome inferred from the transcriptome of the host *Tribonema gayanum* obtained in a previous study (Torruella *et al.*, 2018) and the proteome inferred from the genome of *Heterococcus* sp. DN1 (PRJNA210954) (member of the same Tribonematales order as the host). After blasting *A. gromovi* and *S. tribonematis* against the database of *Tribonema gayanum* + *Heterococcus* sp. DN1 proteins, we excluded those sanchytrid hits that were 100% or >95% identical to them,

respectively. Statistics of the final assembled genomes were assessed with QUAST 4.5 (Gurevich *et al.*, 2013) and Qualimap v2.2.1 (Okonechnikov, Conesa, & García-Alcalde, 2015) for coverage estimation. At the end, we obtained 7,220 and 9,368 protein sequences for *A. gromovi* and *S. tribonematis*, respectively (Table 1). These proteins were functionally annotated with eggNOG mapper (Huerta-Cepas *et al.*, 2017) using DIAMOND as the mapping mode and the eukaryotic taxonomic scope. This resulted in 3,757 (*A. gromovi*) and 4,670 (*S. tribonematis*) annotated peptides for the predicted proteomes (Supplementary Table 1). The mitochondrial genomes of both sanchytrids were identified in single contigs using Blast (Altschul *et al.*, 1990) and annotated with MITOS (Bernt *et al.*, 2013). A posterior Blast search was made to confirm missing proteins. To assess genome completeness, we used BUSCO v2.0.1 (Simão *et al.*, 2015) on the decontaminated predicted proteomes with the fungi_odb9 dataset of 290 near-universal single-copy orthologs.

Phylogenomic analyses and single-gene phylogenies. An updated version of the dataset of conserved phylogenetic markers from (Mikhailov *et al.*, 2016) with 264 protein alignments was used to reconstruct our phylogenomic trees (Torruella *et al.*, 2018; Galindo *et al.*, 2019). This dataset, named 'GBE', was updated with sequences from the two sanchytrid genomes, the non-flagellated chytrid *Hyaloraphidium curvatum* SAG235-1 (SRX4387575), the enigmatic flagellated fungus *Olpidium bornovanus* UBC F19785 (SRX125102, SRX123912, SRX123911), and all publicly available Blastocladiomycota sequences, including the recently sequenced *Paraphysoderma sedebokerense* JEL821 (SRX3538887) and *Coelomomyces lativittatus* ClRM-AVA-1 (SRX2781572). Data was obtained from GenBank (http://www.ncbi.nlm.nih.gov/genbank, last accessed November, 2019), and the Joint Genome Institute (http://www.jgi.doe.gov/; last accessed May 2017). The updated taxon sampling of our study comprises a total of 81 Opisthokonta (2 Holozoa and 79 Holomycota), 2 Amoebozoa and 1 Apusomonadida. Two datasets with two different taxon samplings were prepared, one with all 84 species (GBE84) and one without the long-branch core Microsporidia and metchnikovellids for a total of 69 species (GBE69).

Orthologs of the 264 proteins were searched by tBLASTn (Camacho *et al.*, 2009), incorporated into the individual protein datasets, aligned with MAFFT v7 (Katoh & Standley, 2013) and trimmed with TrimAl with the automated1 option (Capella-Gutiérrez, Silla-Martínez, & Gabaldón, 2009). Alignments were visualized, manually edited and concatenated with Geneious

v6.0.6 (Kearse *et al.*, 2012) and single gene trees obtained with FastTree v2.1.7 (Price, Dehal, & Arkin, 2009) with default parameters. Single gene trees were manually checked to identify and remove paralogous and/or contaminating sequences. The concatenation of the clean trimmed 264 proteins resulted in alignments containing 91,768 (GBE69) and 83,321 (GBE84) amino acid positions. Bayesian inference (BI) phylogenetic trees were reconstructed using PhyloBayes-MPI v1.5 (Lartillot, Lepage, & Blanquart, 2009) under CAT-Poisson model, two MCMC chains for each dataset were run for more than 15,000 generations, saving one every 10 trees. Analyses were stopped once convergence thresholds were reached (i.e. maximum discrepancy less than 0.1 and minimum effective size greater than 100 calculated using bpcomp) and consensus trees constructed after a burn-in of 25%. Maximum likelihood (ML) phylogenetic trees were inferred with IQ-TREE v1.6 (Nguyen *et al.*, 2015) under the LG + R9 + PMSF model for GBE69 and LG + F+ R10 + PMSF for GBE84, selected with the IQ-TREE TESTNEW algorithm as per the Bayesian information criterion (BIC). Statistical support was generated with 1000 ultrafast bootstraps (Minh, Nguyen, & Von Haeseler, 2013) and 1,000 replicates of the SH-like approximate likelihood ratio test (Anisimova *et al.*, 2011). All trees were visualized with FigTree (Rambaut, 2016).

We tested if alternative constrained tree topologies could be rejected. For that, we used Mesquite (Mesquite Project Team, 2014) to constrain the following topologies: 1) chytrids as sister lineage of all other fungi (Blastocladiomycota + Sanchytriaceae + *Olpidium* + Zygomycota + Dikarya), 2) Blastocladiomycota + Sanchytriaceae as sister lineage of all other fungi (Chytridiomycota + *Olpidium* + Zygomycota + Dikarya), and 3) *Olpidium* as an independent lineage sister of all other non-flagellated fungi (Zygomycota + Dikarya) or belonging within Zoopagomycota. The constrained topologies without branch lengths were reanalysed with the -g option of IQ-TREE and the best-fitting model. AU tests were carried out on the resulting trees for each taxon sampling with the -z and -au options of IQ-TREE. Additionally, to minimize possible systematic bias due to the inclusion of fast-evolving sites in our protein alignments, we progressively removed the fastest evolving sites, 5% of sites at a time. For that, among-site evolutionary rates were inferred using IQ-TREE under the -wsr option and the best-fitting model for both taxon samplings for a total of 19 new subsets of each (Supplementary Table 2). We then reconstructed phylogenetic trees for all these subsets using IQ-TREE with the same best-fitting model as for the whole dataset. To assess the support of the alternative topologies in the

bootstrapped trees, we used CONSENSE from the PHYLIP package (Felsenstein, 1993) and interrogated the .UFBOOT file using a Python script (M. Kolisko, pers. comm).

Finally, to remove possible additional compositional heterogeneity within our data, we used the Dayhoff recoding scheme (from 20 to 4 categories) using a Python script. We then reconstructed phylogenetic trees for all the recoded subsets using IQ-TREE with the GTR+F+I+G4 model for ML trees and PhyloBayes-MPI v1.5 with the CAT-Poisson model for BI trees.

Comparative proteomic analysis of primary metabolism. To get insights into the metabolic capabilities of sanchytrids in comparison with other member of Holomycota we carried out statistical multivariate analyses. Protein sets used in this study were obtained between May 2017 and November 2019 from the NCBI protein, genome and SRA databases (https://www.ncbi.nlm.nih.gov/), except the following: *Spizellomyces punctatus*, *Gonapodya prolifera*, *Batrachochytrium dendrobatidis*, *Allomyces macrogynus*, *Catenaria anguillulae* and *Blastocladiella britannica*, retrieved from the MycoCosm portal of the Joint Genome Institute (these sequence data were produced by the US Department of Energy Joint Genome Institute http://www.jgi.doe.gov/ in collaboration with the user community); *Parvularia atlantis* (previously *Nuclearia* sp. ATCC50694, from https://doi.org/10.6084/m9.figshare.3898485.v4); and *Paramicrosporidium saccamoebae*, from NCBI in January 2018. We searched in both sanchytrids for the presence of 1206 eggNOG orthologous groups (Huerta-Cepas *et al.*, 2017) corresponding to 8 primary metabolism categories (Gene Ontology, GO). The correspondence between GO terms and primary metabolism COGs used are the following: [C] Energy production and conversion (227 orthologs); [G] Carbohydrate transport and metabolism (205 orthologs); [E] Amino acid transport and metabolism (200 orthologs); [F] Nucleotide transport and metabolism (87 orthologs); [H] Coenzyme transport and metabolism (94 orthologs); [I] Lipid transport and metabolism (201 orthologs); [P] Inorganic ion transport and metabolism (153 orthologs); and [Q] Secondary metabolites biosynthesis, transport and catabolism (70 orthologs). From these categories, we identified 1158 orthologs non-redundant among categories in the sanchytrid genomes which were shared among a set of 45 species, including 8 opisthosporidians, 30 fungi, 2 holozoans, 2 amoebozoans, and 1 apusomonad (for the complete list, see Supplementary Table 3). We annotated the protein sets of these 45 species using eggNOG-mapper (Huerta-Cepas *et al.*, 2017) with DIAMOND as mapping mode and the eukaryotic taxonomic scope. All ortholog counts

were transformed into a presence/absence matrix (encoded as 0/1) and analysed with the R script (R Development Core Team, 2011) detailed in Torruella et al. (2018) in which similarity values between binary COG profiles of all species were calculated to create a complementary species-distance matrix. We then analysed this distance matrix using a Principal Coordinate Analysis (PCoA) and also plotted binary COG profiles in a presence/absence heatmap. Clustering of the species and orthologs was done by Ward hierarchical clustering (on Euclidean distances for orthologs) of the interspecific Pearson correlation coefficients. The raw species clustering was also represented in a separate pairwise correlation heatmap colour-coded to display positive Pearson correlation values (0–1). Finally, COGs categories of each primary metabolism were also analysed separately (categories C, E, F, G, H, I, P, and Q) using the same workflow (for more details see Torruella et al. 2018 and https://github.com/xgrau/paraphelidium2018).

We compared in more detail the inferred metabolism of a subset of species (the two sanchytrids, *Rozella allomycis* and *Allomyces macrogynus*). The annotation of the proteomes was done using BlastKOALA (Kanehisa, Sato, & Morishima, 2016), with eukaryotes as taxonomy group and the genus_eukaryotes KEGG GENES database, and the annotations uploaded in the KEGG Mapper Reconstruct Pathway platform (Kanehisa, 2000) by pairs. First, we compared the two sanchytrid proteomes to confirm their similarity and, second, we compared them with the proteomes of *R. allomycis* and *A. macrogynus* to study their metabolic reduction.

Homology searches and phylogenetic analysis of specific proteins. To assess the evolution of the flagellum in holomycotan lineages we used the dataset of over 60 flagellum-specific proteins from Torruella et al. (2015) based on previous studies (Carvalho-Santos *et al.*, 2011; Wickstead & Gull, 2012; Van Dam *et al.*, 2013) to examine a total of 47 flagellated and non-flagellated species within and outside the Holomycota. The flagellar toolkit proteins were identified using *Homo sapiens* protein sequences as Blast queries. Candidate proteins were then blasted against the nr GenBank database to confirm their identification and submitted to phylogenetic analysis by multiple sequence alignment with MAFFT, trimming with TrimAl with the automated1 option, and tree reconstruction with FastTree. After inspection of trees, we removed paralogs and other non-orthologous protein sequences. We excluded the proteins with no identifiable presence in any of the 47 species used in the analysis and encoded the presence/absence of the remaining ones in a 1/0 matrix. The native R heatmap function (R Development Core Team, 2011) was used to plot

the flagellar proteome comparison between all species according to their presence/absence similarity profiles.

To study the presence or absence of the fusion of the BeGC1 and BeCNG1 proteins, we blasted them against the proteomes of *S. tribonematis*, *A. gromovi*, *P. sedebokerense* and *C. lativittatus* using the *Blastocladiella emersonii* sequences (BeGC1: AIC07007.1; BeCNG1: AIC07008.1) as queries. We then used MAFFT to include the new sequences in a multiple sequence alignment based on Avelar et al. (2014) for the BeCNG1 protein channel and separately for both the guanylyl-cyclase GC1 domain and the rhodopsin domain of the BeGC1 fusion protein. After trimming with TrimAl we reconstructed phylogenetic trees for the 3 datasets using IQ-TREE with the LG+F+I+G4 model for the rhodopsin domain and BeCNG1 alignments and LG+G4 model for the GC1 alignment. The resulting trees were visualized with FigTree.

Results and discussion

Sanchytrid genomes. We isolated sporangia of the two sanchytrid species *A. gromovi* and *S. tribonematis* by micromanipulation and sequenced their genomes after whole genome amplification. After a three-step decontamination (see Methods), we assembled two high coverage genome sequences (123.9X and 45.9X, respectively) of 10.5 and 11.2 Mb, encoding 7,220 and 9,638 proteins, respectively (Table 1). Comparison with a fungal dataset of 290 near-universal single-copy orthologs (Simão *et al.*, 2015) indicated very high completeness levels for the two genomes (92.41% for *A. gromovi* and 91.72% for *S. tribonematis*). Whereas the two sanchytrid genomes had similar genome statistics (Table 1), they showed important differences when compared with genomes from other well-known zoosporic fungi. Sanchytrid genomes were five time smaller than those of Blastocladiomycota (ranging from 40 to 50 Mb) and average chytrids (~20 to 101 Mb). The number of protein coding genes also followed this trend. The sanchytrid genome GC content (~35%) was also atypical, much lower than in Blastocladiomycota and most chytrids (40-57%, though some chytrids like *Anaeromyces robustus* can have values down to 16.3%, ;-Billon-Grand *et al.*, 1991; Youssef *et al.*, 2013).

A correlation between parasitism and low GC content has already been noticed in eukaryotes (Videvall, 2018). In Holomycota, this pattern can be observed both in anaerobic species with remnant mitochondria, like Microsporidia and Neocallimastigomycota (Chen *et al.*, 2013;

Youssef *et al.*, 2013), and in aerobic ones such as the rozellid *Rozella allomycis* (James *et al.*, 2013). Sanchytrids are aerobic and parasitic and their life cycles (Karpov *et al.*, 2017a, 2018, 2019) do not seem different from those of Blastocladiomycota or chytrids, so it remains unclear why their genome sizes and GC content are considerable lower than in these other fungal lineages. Although we have not identified the precise reasons of the pronounced genome size reduction and GC content bias, they seem to be correlated with a global acceleration of evolutionary rate in sanchytrids (see below), as observed in *R. allomycis* (Thomarat et al. 2004; James et al. 2013; Mikhailov et al. 2016; Galindo et al. 2018).

Table 1. Comparative statistics of sanchytrid genomes before and after decontamination with related zoosporic lineages.

	Amoeboradix gromovi		*Sanchytrium tribonematis*		*Allomyces macrogynus*	*Catenaria anguillulae*	*R. globosum*	*S. punctatus*	*A. robustus*
	Before	After	Before	After					
Genome size (Mb)	48.7	10.5	37.1	11.2	52.62	41.34	57.02	24.1	71.69
GC%	51.06	36.27	42.9	34.64	61.6	56	44.9	47.6	16.3
Number of contigs	8,420	1,167	8,015	1,960	8,973	509	437	329	1,035
N50	27,236	13,376	17,350	11,874	35,497	217,825	292,246	155,888	141,798
Predicted proteins	87,868	7,220	50,780	9,368	19,447	12,763	16,987	9,422	12,083

A new phylogenomic framework for Fungi. To resolve the unstable phylogenetic position of sanchytrids (Karpov *et al.*, 2018) and, more globally, the relationships among the different groups of zoosporic fungi, we carried out phylogenetic analyses on a large dataset of 264 conserved proteins and 91,768 amino acid positions (Mikhailov *et al.*, 2017; Torruella *et al.*, 2018; Galindo *et al.*, 2019) using Bayesian inference (BI) under the CAT-Poisson model (Lartillot & Philippe, 2004) and maximum likelihood (ML) under the PMSF model (Wang *et al.*, 2018). Both mixture models have been proven to be robust against homoplasic positions and long-branch attraction (LBA) artefacts (Lartillot & Philippe, 2004; Chang *et al.*, 2015). We selected a taxon sampling with 69 species, including holomycotan species plus two amoebae and one apusomonad as outgroup (dataset GBE69). In addition to the new sanchytrid data, we incorporated several zoosporic fungi never included before in large phylogenomic analyses: the two Blastocladiomycota *Paraphysoderma sedebokense* (Hoffman *et al.*, 2008; Gutman, Zarka, &

Boussiba, 2009; James *et al.*, 2011) and *Coelomomyces lativittatus* (Couch, 1962), the enigmatic flagellated fungus *Olpidium bornovanus* (Uebelmesser, 1956), and the non-flagellated chytrid *Hyaloraphidium curvatum*, which seems to have completely lost its flagellum (Ustinova *et al.*, 2000). BI and ML phylogenomic analyses yielded the same tree topology for major groups with only minor changes in the position of terminal branches (Figure 1A). We recovered maximum support for both the monophyly of sanchytrids (*A. gromovi* + *S. tribonematis*) and their position as sister group of Blastocladiomycota. Thus, sanchytrids form a new deep-branching group among zoosporic fungi. In contrast with its previous unstable position in few-genes phylogenies (Sekimoto *et al.*, 2011; Tedersoo *et al.*, 2018), we retrieved full support with all methods for *Olpidium bornovanus* as an independent fungal lineage sister to the major non-flagellated fungal clade. We also systematically recovered a deep divergence of chytrids as the sister group of all other fungi with full Bayesian posterior probability but moderate ML bootstrap support (79%). Despite the use of a large dataset, some branches remained unresolved, such as the position of Glomeromycota sister either to Mucoromycota or Dikarya.

In agreement with previous phylogenies based on rRNA coding genes (Karpov *et al.*, 2018), sanchytrids exhibited a very long branch in all our phylogenetic trees (Figure 1A), suggesting that sanchytrids have a fast evolutionary rate. This is a well-known phenomenon in other fungi-related organisms such as the Microsporidia (also with low GC content and small genomes), for which their long branches due to fast evolving genomes (Thomarat *et al.*, 2004) caused LBA artefacts when trying to infer their phylogenetic position (Leipe *et al.*, 1993; Kamaishi *et al.*, 1996; Philippe *et al.*, 2000). To test if LBA affected the position of the long-branching sanchytrids and, eventually, other branches in our tree, we introduced long-branching metchnikovellids and core Microsporidia in our dataset, for a total of 84 species and 83,321 conserved amino acid positions (dataset GBE84). Despite the inclusion of this very long branch, which could attract other long branches by LBA, we globally recovered the same topology as with the previous taxon sampling (GBE69) with just minor changes in the position of lineages inside the large clades (Supplementary Figure 1). The new trees confirmed the maximum support for the sister relationship of sanchytrids and Blastocladiomycota, and for the monophyly of *O. bornovanus* and non-zoosporic fungi. Likewise, the position of chytrids as the sister lineage of all fungi was recovered with full Bayesian posterior probabilities and slightly higher ML bootstrap (82%).

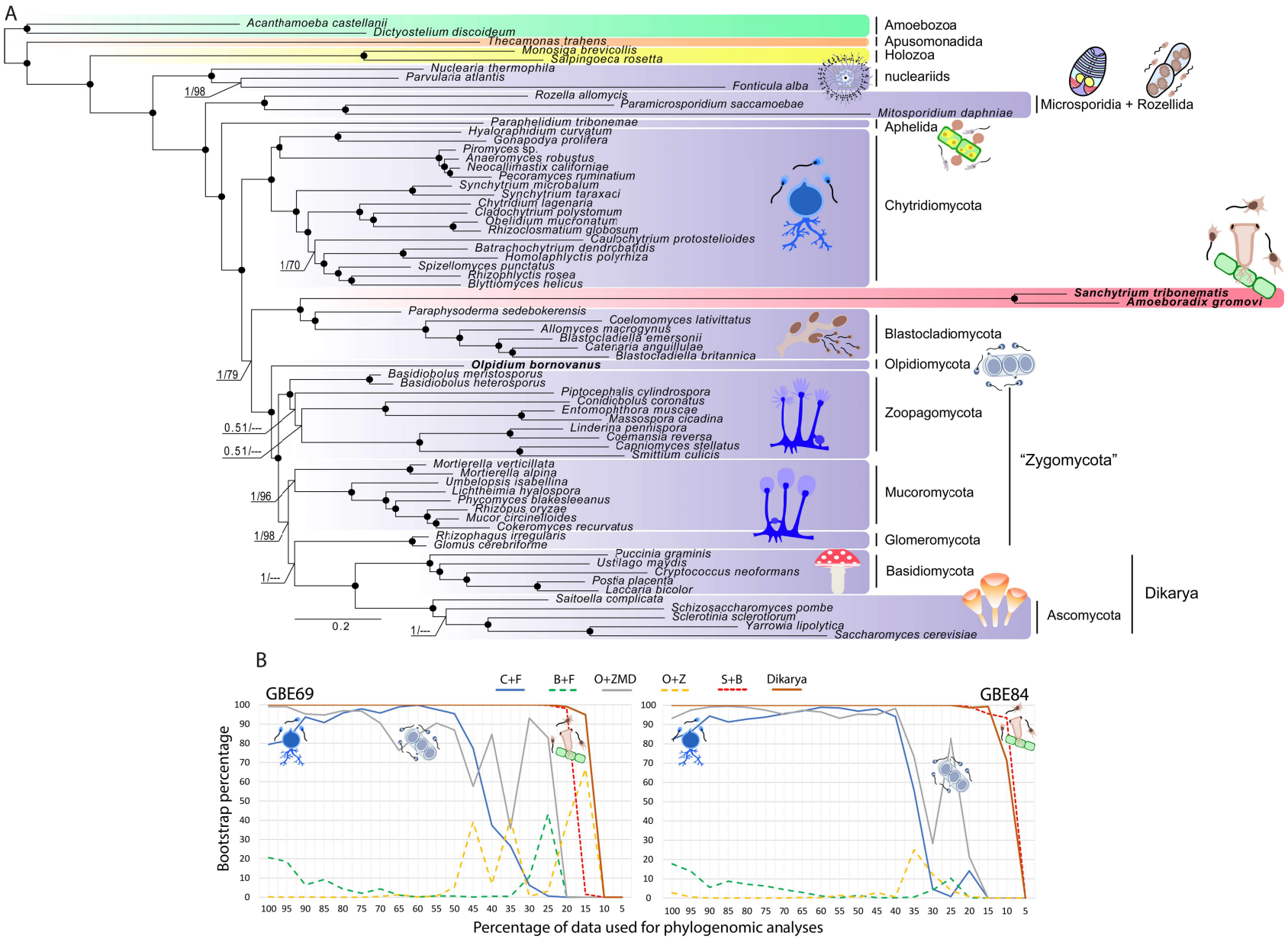

A
Acanthamoeba castellanii
Dictyostelium discoideum
Thecamonas trahens
Monosiga brevicollis
Salpingoeca rosetta
Nuclearia thermophila
Parvularia atlantis
Fonticula alba
Rozella allomycis
Paramicrosporidium saccamoebae
Mitosporidium daphniae
Paraphelidium tribonemae
Hyaloraphidium curvatum
Gonapodya prolifera
Piromyces sp.
Anaeromyces robustus
Neocallimastix californiae
Pecoramyces ruminatium
Synchytrium microbalum
Synchytrium taraxaci
Chytridium lagenaria
Cladochytrium polystomum
Obelidium mucronatum
Rhizoclosmatium globosum
Caulochytrium protostelioides
Batrachochytrium dendrobatidis
Homolaphlyctis polyrhiza
Spizellomyces punctatus
Rhizophlyctis rosea
Blyttiomyces helicus
Sanchytrium tribonematis
Amoeboradix gromovi
Paraphysoderma sedebokerensis
Coelomomyces lativittatus
Allomyces macrogynus
Blastocladiella emersonii
Catenaria anguillulae
Blastocladiella britannica
Olpidium bornovanus
Basidiobolus meristosporus
Basidiobolus heterosporus
Piptocephalis cylindrospora
Conidiobolus coronatus
Entomophthora muscae
Massospora cicadina
Linderina pennispora
Coemansia reversa
Capniomyces stellatus
Smittium culicis
Mortierella verticillata
Mortierella alpina
Umbelopsis isabellina
Lichtheimia hyalospora
Phycomyces blakesleeanus
Rhizopus oryzae
Mucor circinelloides
Cokeromyces recurvatus
Rhizophagus irregularis
Glomus cerebriforme
Puccinia graminis
Ustilago maydis
Cryptococcus neoformans
Postia placenta
Laccaria bicolor
Saitoella complicata
Schizosaccharomyces pombe
Sclerotinia sclerotiorum
Yarrowia lipolytica
Saccharomyces cerevisiae
1/98
1/70
1/79
0.51/---
0.51/---
1/96
1/98
1/---
1/---
0.2
Amoebozoa
Apusomonadida
Holozoa
nucleariids
Microsporidia + Rozellida
Aphelida
Chytridiomycota
Blastocladiomycota
Olpidiomycota
Zoopagomycota
"Zygomycota"
Mucoromycota
Glomeromycota
Basidiomycota
Ascomycota
Dikarya
B
GBE69
GBE84
C+F
B+F
O+ZMD
O+Z
S+B
Dikarya
Bootstrap percentage
Percentage of data used for phylogenomic analyses

Figure 1. (A) Bayesian inference (BI) phylogenomic tree based on 264 conserved proteins. The tree was reconstructed using 69 species and 91,768 amino acid positions with the CAT-Poisson model and the LG+R9+PMSF model for maximum likelihood (ML). Branches with support values higher or equal to 0.99 BI posterior probability and 99% ML bootstrap are indicated by black dots. (B) Evolution of IQ-TREE ML bootstrap support for Chytridiomycota sister of all other fungi (C+F), Blastocladiomycota + Sanchytriaceae sister of all other fungi (B+F), *Olpidium* independent lineage sister to non-flagellated fungi (O+ZMD), *Olpidium* within Zoopagomycota (O+Z), Sanchytriaceae within Blastocladiomycota (S+B), and the monophyly of Dikarya (Dikarya) as a function of the proportion of fast-evolving sites removed from both datasets (GBE69 without long-branching Microsporidia, and GBE84 with long-branching Microsporidia). All phylogenomic trees can be seen in Supplementary Figures 1A-E.

We further tested the robustness of the position of chytrids and *Olpidium* using alternative topology AU tests for the two datasets of 69 and 84 species. The test did not reject alternative positions for the divergence of Chytridiomycota and Blastocladiomycota + Sanchytriaceae (AU test *p*-values >0.05; Supplementary table 2). This most likely reflected a weak phylogenetic signal in the dataset concerning these deep branches (Chang *et al.*, 2015). However, we did observe a clear trend with both taxon samplings by which the *p*-values of trees showing chytrids as sister lineage of all other fungi were higher than those of trees with the Blastocladiomycota + sanchytrids as the first fungal branch to diverge (Supplementary table 2). In the case of *Olpidium*, the tests significantly rejected its previously found position within Zoopagomycota (James *et al.*, 2006a; Sekimoto *et al.*, 2011; Tedersoo *et al.*, 2018) with both datasets (GBE64 AU *p*-value 0.0001; GBE84 AU *p*-value 0.001) and supported its position as an independent lineage sister to the large non-flagellated fungal clade in both datasets (GBE69 AU *p*-value 0.539; GBE84 AU *p*-value 0.554).

We also tested the possible influence of noisy, fast-evolving positions by applying a slow-fast approach (Brinkmann & Philippe, 1999) by progressively removing 5% of the fastest-evolving sites for both the 69 and 84 species datasets. The monophyly of sanchytrids and Blastocladiomycota obtained maximum support (>99% bootstrap) in all steps until only 25/20% (GBE69/GBE84) of the sites remained, when the phylogenetic signal was too low to resolve any deep-level relationship (Figure 1B). The sanchytrids + Blastocladiomycota monophyly was as strongly supported as the Dikarya monophyly. The root of the tree of fungi between the sanchytrids + Blastocladiomycota and the rest of fungi always received poor support (<50% bootstrap). On the contrary, the root between chytrids and the rest of fungi obtained high support (>90% bootstrap) from all datasets with 10 to 50% (GBE69) and 10 to 45% (GBE84) of the fast-evolving

sites removed. The position of *Olpidium* as sister of the non-flagellated fungi was also recovered with high support (>95% bootstrap) until 25/60% (GBE69/GBE84) of the fast-evolving sites were removed. Its possible relationship with Zoopagomycota was always poorly supported, indicating that this position previously found (James *et al.*, 2006b; Sekimoto *et al.*, 2011) was probably an artefact due to low phylogenetic signal.

Finally, we applied a recoding approach (Susko & Roger, 2007) to alleviate possible compositional biases within our data. After recoding, we recovered the same tree topology except for minor changes in deep splits for both BI and ML analyses (Supplementary figs. 2A-D). We recover in all trees chytrids as the sister lineage of all fungi, except for the ML tree for GBE84, where we suspect a LBA artefact attracting the long-branching sanchytrids towards the Microsporidia. In all the recoded trees, we recovered with full support *Olpidium* as a new independent lineage sister to the non-flagellated fungi. Although recoding has been shown to remove informative substitutions (Vera-ruiz *et al.*, 2014), these results were in good general agreement with those from the non-recoded datasets, suggesting that they were not due to compositional biases.

Our phylogenetic analyses confirmed that sanchytrids form a new fast-evolving lineage sister to Blastocladiomycota and supported the position of chytrids as sister group of all other fungi. They also confirmed the position of *Olpidium* as an independent sister lineage to canonical non-flagellated fungi. Based on ultrastructural characteristics of its zoospores, *Olpidium* has been suggested to be related with *Caulochytrium protostelioides* (Olive, 1980; James *et al.*, 2006b). However, in our multi-gene phylogenies *C. protostelioides* branched within Chytridiomycota with maximum support, confirming that any similarity with *Olpidium* most likely reflects convergent evolution. The main hypothesis of *Olpidium* placement within the Zoopagomycota, in most cases in close association with Basidiobolaceae (James *et al.*, 2006a; Sekimoto *et al.*, 2011; Tedersoo *et al.*, 2018), has never received strong phylogenetic support and, on the contrary, our results robustly indicated it constitutes a new independent fungal linage (I.e., a new phylum Olpidiomycota) sister to the non-flagellated fungi.

The relative position of Chytridiomycota and Blastocladiomycota close to the root of the tree of Fungi has remained a major unresolved question and both lineages have alternatively been recovered as sister lineage of all other fungi in phylogenomic studies (James *et al.*, 2006b; Sekimoto *et al.*, 2011; Spatafora *et al.*, 2016). There is evidence that the earliest fungal split may

have occurred as far as 1 billion years ago (Loron *et al.*, 2019) so that phylogenetic signal to infer it may have been largely eroded over time. It has also been suggested that fast radiation events occurred during early fungi evolution (Chang *et al.*, 2015), making difficult the accumulation of enough substitutions to have strong phylogenetic resolution. This phenomenon might also explain the low support and discrepancies observed for the split between Glomeromycota and Mucoromycota, probably related to their symbiotic adaptation to land plants (Bidartondo *et al.*, 2011; Field *et al.*, 2015; Feijen *et al.*, 2018). Nevertheless, our various phylogenetic analyses, based on much larger sequence datasets than previous studies, converged to locate the root of the tree of Fungi between chytrids and all other fungi. This root position agrees with the distribution of characters that are considered to be ancestral (sporic meiosis and relatively small number of genes for carbohydrate metabolism). And fits with the fact that some species within the Blastocladiomycota, present derived traits, including apically growing structures similar to true hyphae (*Allomyces*) and narrow exit tubes on sporangia (e.g. *Catenaria* spp.) (Vargas, Aronson, & Roberson, 1993; Stajich *et al.*, 2009; Archibald *et al.*, 2017; Berbee *et al.*, 2017) .
Nevertheless, to further confirm the relationships among the main clades of zoosporic fungi, there will be needed a combination of both taxon sampling to obtain genome sequences of more early diverging taxa, together with improvements in phylogenetic analyses associated with genome and gene content and composition (Spatafora *et al.*, 2017).

Comparative genomics of primary metabolism. To assess if the metabolic capabilities of the sanchytrids are as reduced as suggested by their very small genome sizes, we compared their metabolic potential with other members inside Fungi, but also other opisthokonts and amoebozoans as outgroups (for a total of 45 species). We looked for the presence of gene orthologous groups in these genomes using eggNOG (Huerta-Cepas *et al.*, 2017). Only half of the proteins predicted for both sanchytrids species got a functional annotation by EggNOG. Nevertheless, this is not atypical in fast-evolving holomycotan lineages. For example, in opisthosporidians a large proportion of their proteins remain without a known function, probably due to the fact that many of their genes have evolved so fast that they cannot longer be recognized by annotation programs (Cuomo *et al.*, 2012; Nakjang *et al.*, 2013). For example, for the Microsporidia *Nosema parisii* and *Encephalitozoon cuniculi*, only 20% and 52% of their genes, respectively, can be assigned to Pfam domains and GO terms (Vivarès *et al.*, 2002; Cuomo *et al.*,

2012). In the case of short-branching opisthosporidians we recover the following percentages of annotated proteins using the same EggNOG parameters: 45.6% in *Amphiamblys* sp. (1,665/3,647), 64.9% in *R. allomycis* (4,123/6,350), and 66,7% in *P. saccamoebae* (2,502/3,750), which are values comparable to those of sanchytrids. Similarly, despite not being part of a fast-evolving lineage in the closely related Blastocladiomycota, only about half of the proteome could be assigned a functional annotation in *Paraphysoderma sedebokerense* (66.8%: 7256/10859), *Catenaria anguillulae* (47.5%: 6,062/12,763), and *Blastocladiella britannica* (44.1%: 11,573/26,214). This indicated that the percentage of functional annotation in sanchytrid proteomes was not unusual given their phylogenetic position and the rapid evolution of their genomes.

To carry out an overall comparison of the metabolism, we focused on the genes involved in eight primary metabolism categories. We identified a total of 1158 orthologous groups in the sanchytrids and the other 43 eukaryotic species, with which we built a presence/absence matrix (Figure 2A). Sanchytrid profiles clustered with *Rozella allomycis*, *Mitosporidium daphniae* and *Paramicrosporidium saccamoebae*, all of them organisms belonging to the Opisthosporidia, a lineage with clear evidence of genome reduction (James *et al.*, 2013; Corsaro *et al.*, 2014; Haag *et al.*, 2014). These profiles also showed that sanchytrids did not cluster with canonical fungi (all fungi excluding the anaerobic Neocallimastigomycota chytrids). This was confirmed by a principal coordinate analysis of the same matrix (Figure 2B). Under the first axis, sanchytrids positioned between canonical fungi and other eukaryotes, including *R. allomycis*, *M. daphniae* and *P. saccamoebae*. However, the second axis clarified that even if sanchytrids cluster among the zoosporic fungi (including also the aphelid *P. tribonemae*), they share the lack of certain metabolic traits with other reduced holomycotan parasites, such as the already mentioned *R. allomycis*.

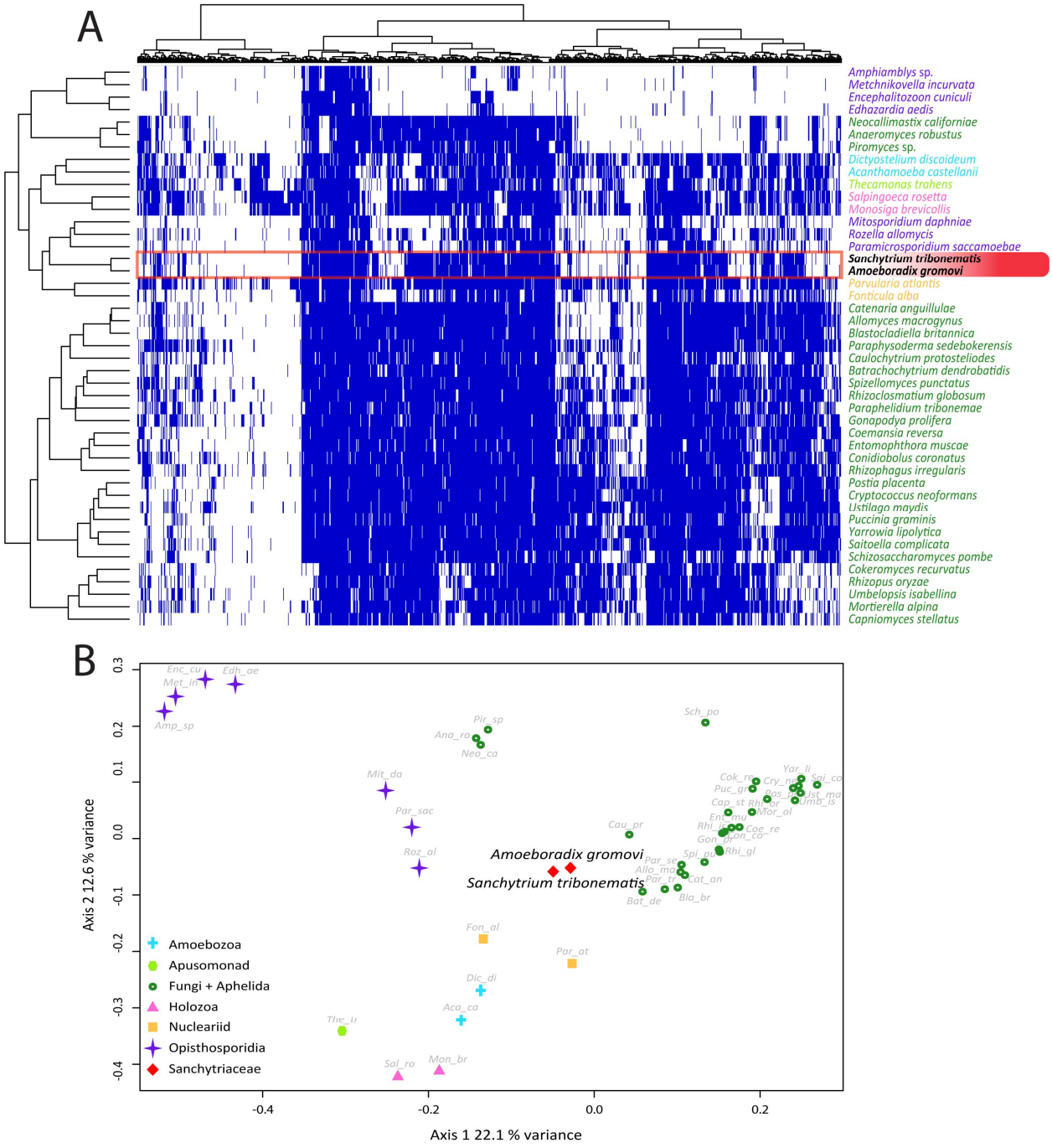


Figure 2. Reduction of *Amoeboradix gromovi* and *Sanchytrium tribonematis* primary metabolism. (A) Binary heat-map and (B) principal coordinate analysis (PCoA) species clustering based on the presence/absence of 1158 orthologous genes belonging to 8 primary metabolism Gene Ontology categories across 45 eukaryotic genomes and transcriptomes. Species are color- and shaped-coded according to their taxonomic affiliation. COG presence is depicted in blue and absence is depicted in white.

To examine in more detail those traits, we looked into each of the eight primary metabolism categories and compared the sanchytrids profiles against those of canonical fungi and opisthosporidians. For most categories, the only truly different profiles were those of Microsporidia and Neocallimastigomycota. All other species, including sanchytrids, chytrids and Blastocladiomycota, clustered together, meaning that the metabolic composition of sanchytrids is much alike canonical fungi, with two exceptions. In fact, for the categories of carbohydrate and lipid transport and metabolism, sanchytrids clustered with opisthosporidians (Supplementary Figure 3). To better understand these similarities, we compared by pairs the metabolic gene composition of sanchytrids against *R. allomycis* (representing opithosporidians) and the Blastocladiomycota *Allomyces macrogynus* (representing canonical fungi) using KEGG annotations. We observed that the sanchytrid metabolic map contained 394/593 more orthologs than *R. allomycis* and 3638/3442 less orthologs than *A. macrogynus* (for a total of 1222/1418 orthologous groups in *A. gromovi*/*S. tribonematis*, 845 in *R. allomycis*, and 4860 in *A. macrogynus*) (Supplementary Figures 4A-C). As previously observed, we found more similarities between canonical fungi and sanchytrids, including the maintenance of amino acid and nucleotide metabolism and energy production with a complete electron transport chain, than with *Rozella*, which lacks these characters (James *et al.*, 2013; Torruella *et al.*, 2018). Most carbohydrate metabolic pathways were retained in sanchytrids and canonical fungi except for the galactose and inositol phosphate pathways, which were lost in sanchytrids and *Rozella*. However, the most important difference between sanchytrids and canonical fungi was found in lipid metabolism. In parallel with *Rozella* (James *et al.*, 2013), sanchytrids lack several lipid metabolic pathways, including steroids and fatty acids metabolism (Supplementary figure 4D-I). This absence of several lipid and carbohydrate metabolic pathways explains the similar primary metabolic profiles of sanchytrids and opisthosporidians. Sanchytrids represent an independent lineage of fast-evolving organisms within the canonical fungi undergoing metabolic reduction.

As mentioned above, sanchytrids share some other opisthosporidian-like traits: small genomes, low GC content, and long branches in phylogenies (due to fast-evolving genomes). However, despite their very reduced genomes and the lack of some metabolic capabilities, our data suggest that the level of host dependence of sanchytrids is most likely much less extreme as in opisthosporidians.

The mitochondrial genomes of sanchytrids. After sequence assembly and decontamination, we recovered the mitochondrial genome of both sanchytrids in single contigs of 24,749 bp and 27,055 bp for *S. tribonematis* and *A. gromovi*, respectively (Supplementary Figure 5), substantially shorter in comparison with Blastocladiomycota species like *B. emersonii* (37,503 bp) or *A. macrogynus* (57,473 bp). Their GC content was also lower (25.86% and 30.69%, respectively) than that of *B. emersonii* (35.09%) and *A. macrogynus* (39.5%). Even when their mitochondrial genomes seemed to follow the same reductive trend as the nuclear genome, most of the core mitochondrial genes were present in both sanchytrids, suggesting that they have a functional mitochondrion including all main elements of the electron transport chain. This concours with the other results that indicate that sanchytrids are less dependent on their hosts than the more metabolically reduced opisthosporidian parasites.

One noticeable loss was ATP8, a subunit of the F-type ATP synthase. ATP8 has been observed to be absent or highly modified in several metazoan groups, including chaetognaths, rotifers, most bivalve molluscs, and flatworms (Gissi, Iannelli, & Pesole, 2008; Egger, Bachmann, & Fromm, 2017). *S. tribonematis* also lacked several mt-tRNA genes and the NAD4L subunit of the NADH dehydrogenase. However, these losses probably do not impact the capacity to produce ATP as shown by the opisthosporidian *R. allomycis*, which possesses a more reduced mitochondrial genome that lacks not only ATP8 but also the complete NADH dehydrogenase complex and most mt-tRNAs (only 4 kept), but still seems to be able to synthesize ATP (James *et al.*, 2013). The mitochondrial gene order was highly variable between the two sanchytrids, a common phenomenon in Fungi, both among and within the major phyla (Aguileta *et al.*, 2014),

Flagellum evolution and reduction in sanchytrids and Holomycota. As in all opisthokonts, the flagellum is one of the most defining traits of Holomycota (Cavalier-Smith & Chao, 2003; Torruella *et al.*, 2015). Its loss is considered to be one key adaptation in the transition of Fungi to land environments (Naranjo-Ortiz & Gabaldón, 2019) so that knowing how many times it occurred and in which lineages remains one of important question in fungal ecology and evolution. The flagellum is completely absent in nucleariids (Bass *et al.*, 2018; Galindo *et al.*, 2019) but can be found in representatives of all other major holomycotan clades including opisthosporidians, such as *Rozella* (Letcher & Powell, 2018) and aphelids (Karpov *et al.*, 2014), and various Fungi, including chytrids (Powell, 2017a), Blastocladiomycota (Hibbett *et al.*, 2007), the enigmatic

Olpidium (James *et al.*, 2006b; Sekimoto *et al.*, 2011), and sanchytrids (Karpov *et al.*, 2017a, 2018, 2019). Among these organisms, sanchytrids represent an atypical case. Their amoeboid zoospores have never been observed swimming, instead they glide on solid surfaces by producing thin filopodia and a hyaline pseudopodium at their anterior end, while their posterior flagellum (pseudocilium) drags behind the cell and does not seem to be used for locomotion (Karpov *et al.*, 2018, 2019). Examination of the ultrastructure of the axoneme and the kinetosome confirmed that sanchytrids do not have a typical functional flagellum. Although the eukaryotic axonemes can be remarkably variable from the typical 9 peripheral microtubules doublets plus a central pair (Mitchell, 2004), it is noteworthy that in sanchytrids the axoneme is formed by just 9 singlets without the central pair or as few as only 4 microtubules (Karpov *et al.*, 2018, 2019). Moreover, the eukaryotic kinetosomes or basal bodies have a much more conserved structure, which normally contains 9 triplets of microtubules, but that is highly reduced in sanchytrids and is formed of 9 singlet microtubules in *S. tribonematis* and 9 singlets or doublets (depending on the strain) in *A. gromovi*. Surprisingly, despite this substantial structural simplification, the sanchytrid zoospores have a remarkably long kinetosome: *A. gromovi* possesses one of the longest known kinetosomes reaching 2.2 µm (Karpov *et al.*, 2018, 2019).

Amoeboid movement has been described in zoospores of several species within the Blastocladiomycota, where amoeboid forms appear to be the infective agents in the vegetative cycle and the flagellated forms most likely gametes (Sparrow, 1960; Letcher *et al.*, 2016; Powell, 2017b). One of these species with both amoeboid and swimming zoospores is *Paraphysoderma sedebokerense* (Strittmatter *et al.*, 2016), the first branch to diverge in Blastocladiomycota in our multi-gene phylogeny (Figure 1A). However, in contrast with sanchytrids *P. sedebokerense* amoeboid zoospores do not present long kinetosomes (Letcher *et al.*, 2016). In fact, such long and, at the same time, reduced kinetosomes have not been reported in any other zoosporic fungi.

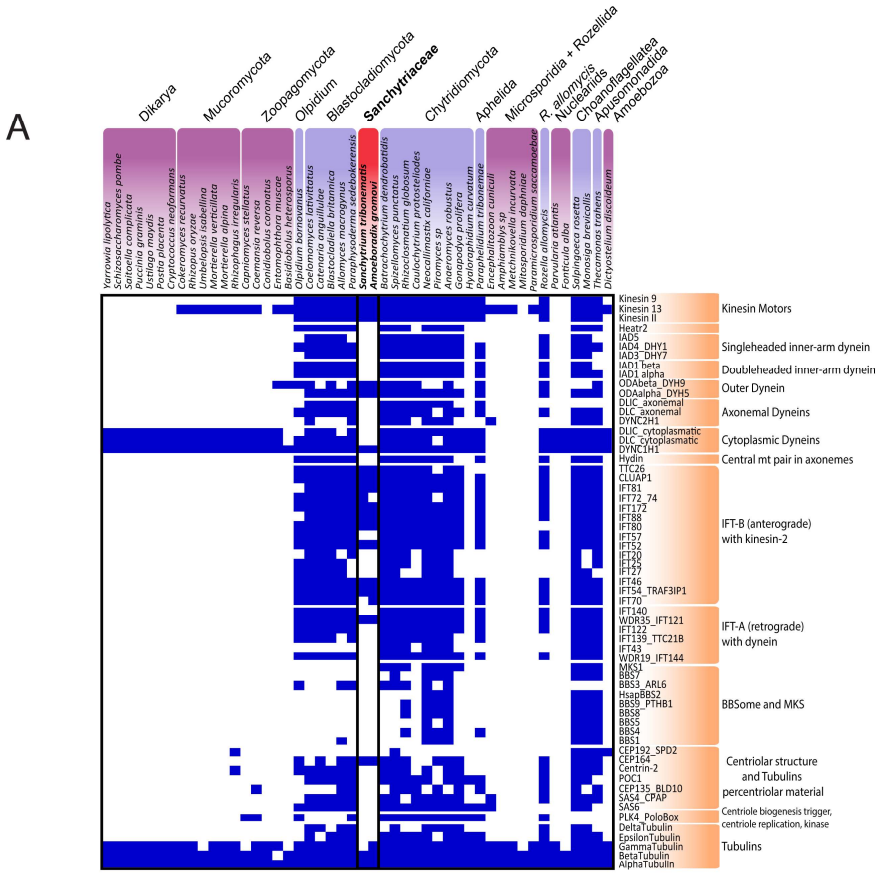

A
Dikarya
Mucoromycota
Zoopagomycota
Olpidium
Blastocladiomycota
Sanchytriaceae
Chytridiomycota
Aphelida
Microsporidia + Rozellida
R. allomycis
Nucleariids
Choanoflagellatea
Apusomonadida
Amoebozoa
Yarrowia lipolytica
Schizosaccharomyces pombe
Saitoella complicata
Puccinia graminis
Ustilago maydis
Postia placenta
Cryptococcus neoformans
Cokeromyces recurvatus
Rhizopus oryzae
Umbelopsis isabellina
Mortierella verticillata
Mortierella alpina
Rhizophagus irregularis
Capniomyces stellatus
Coemansia reversa
Conidiobolus coronatus
Entomophthora muscae
Basidiobolus heterosporus
Mitosporidium daphniae
Paramicrosporidium saccamoebae
Metchnikovella incurvata
Hyaloraphidium curvatum
Encephalitozoon cuniculi
Amphiamblys sp.
Sanchytrium tribonematis
Amoeboradix gromovi
Nuclearia simplex
Spizellomyces punctatus
Rhizoclosmatium globosum
Caulochytrium protostelioides
Paramicrosporidium saccamoebae
Piromyces sp.
Anaeromyces robustus
Gonapodya prolifera
Parapheidium tribonemae
Encephalitozoon cuniculi
Amphiamblys sp.
Metchnikovella incurvata
Mitosporidium daphniae
Paramicrosporidium saccamoebae
Rozella allomycis
Parvularia atlantis
Fonticula alba
Salpingoeca rosetta
Monosiga brevicollis
Thecamonas trahens
Dictyostelium discoideum
Kinesin 9
Kinesin 13
Kinesin II
Kinesin Motors
Heatr2
IAD5
IAD4_DHY1
IAD3_DHY7
Singleheaded inner-arm dynein
IAD1 beta
IAD1 alpha
Doubleheaded inner-arm dynein
ODAbeta_DYH9
ODAalpha_DYH5
Outer Dynein
DLIC_axonemal
DLC_axonemal
DYNC2H1
Axonemal Dyneins
DLIC_cytoplasmatic
DLC_cytoplasmatic
DYNC1H1
Cytoplasmic Dyneins
Hydin
Central mt pair in axonemes
TTC26
CLUAP1
IFT81
IFT72_74
IFT172
IFT88
IFT80
IFT57
IFT52
IFT20
IFT43
IFT27
IFT46
IFT54_TRAF3IP1
IFT70
IFT-B (anterograde)
with kinesin-2
IFT140
WDR35_IFT121
IFT122
IFT139_TTC21B
IFT43
WDR19_IFT144
IFT-A (retrograde)
with dynein
MKS1
BBS7
BBS3_ARL6
HsapBBS2
BBS9_PTHB1
BBS8
BBS5
BBS4
BBS1
BBSome and MKS
CEP192_SPD2
CEP164
Centrin-2
POC1
CEP135_BLD10
SAS4_CPAP
SAS6
Centriolar structure
and Tubulins
percentriolar material
PLK4_PoloBox
Centriole biogenesis trigger,
centriole replication, kinase
DeltaTubulin
EpsilonTubulin
GammaTubulin
BetaTubulin
AlphaTubulin
Tubulins
B
Non-Flagellated lineages
Flagellated lineages
Sanchytrium tribonematis
Amoeboradix gromovi
Hyaloraphidium curvatum
Encephalitozoon cuniculi
Basidiobolus heterosporus
Mitosporidium daphniae
Amphiamblys sp.
Paramicrosporidium saccamoebae
Metchnikovella incurvata
Coemansia reversa
Enterocytozoon bieneusi
Dictyostelium discoideum
Rhizophagus irregularis
Cokeromyces recurvatus
Rhizopus oryzae
Mortierella verticillata
Mortierella alpina
Umbelopsis isabellina
Capniomyces stellatus
Fonticula alba
Yarrowia lipolytica
Schizosaccharomyces pombe
Saitoella complicata
Puccinia graminis
Ustilago maydis
Postia placenta
Cryptococcus neoformans
Parvularia atlantis
Conidiobolus coronatus
Piromyces sp.
Thecamonas trahens
Monosiga brevicollis
Salpingoeca rosetta
Neocallimastix californiae
Anaeromyces robustus
Olpidium bornovanus
Rhizoclosmatium globosum
Paramicrosporidium saccamoebae
Batrachochytrium dendrobatidis
Spizellomyces punctatus
Parapheidium tribonemae
Caulochytrium protostelioides
Rozella allomycis
Gonapodya prolifera
Allomyces macrogynus
Catenaria anguillulae
Blastocladiella britannica
Coelomomyces lativittatus
Kinesin13
DLC_cytoplasmatic
DLC_cytoplasmatic
DYNC1H1
GammaTubulin
BetaTubulin
AlphaTubulin
CEP192_SPD2
DYNC2H1
IFT43
MKS1
BBS4
BBS8
BBS9_PTHB1
BBS7
BBS5
BBS2
BBS1
ODAbeta
BBS3_ARL6
IFT27
IFT25
DLIC_axonemal
IFT20
IFT139_TTC21B
CEP135_BLD10
PLK4_PoloBox
CEP164
TTC26
KinesinII
Kinesin9
ODAalpha
IFT88
IFT81
IFT80
CLUAP1
IFT172
IFT52
IFT46
WDR35_IFT121
IFT54_TRAF3IP1
DeltaTubulin
SAS4_CPAP
Heatr2
Centrin-2
IAD5
IAD3_DHY7
IAD1beta
SAS6
EpsilonTubulin
POC1
DLC_axonemal
IFT72_74
Hydin
IFT70
IAD4_DHY1
IAD1alpha
IFT57
IFT140
WDR19_IFT144
IFT122

Figure 3. Flagellum toolkit reduction. Presence/absence heatmaps of 61 flagellum-specific proteins in 48 eukaryotic flagellated (purple) and non-flagellated (pink) lineages. (A) The right column lists key microtubular genes and flagellum-specific genes. (B) Heatmap clustered by similarity showing sanchytrids in an intermediate position between flagellated and non-flagellated lineages. Gene presence is depicted in blue and absence is depicted in white.

To better understand this reduction pattern, we have studied the evolution of the flagellar gene complement at the whole holomycotan scale. We worked with a set of 61 flagellum-specific proteins on a taxon sampling of 48 flagellated and non-flagellated species. Sanchytrids yielded no significant hits for several key functional and maintenance flagellar components (Figure 3A). In particular, their axonemes had no trace of axonemal dyneins, single-headed and double-headed inner arm dyneins and all intraflagellar transport proteins (IFT) of the group IFT-A, as well as several of the group IFT-B. The sanchytrid kinetosomes have also lost several components of the centriolar structure and tubulins, including the protein Centrin2 that plays a key role in correct basal body anchoring (Aubusson-Fleury *et al.*, 2017), and tubulins Delta and Epsilon, which are essential for the assembly and anchorage of the centriolar microtubules (Dupuis-Williams *et al.*, 2002). Therefore, sanchytrids do not have a functional flagellum from a motility perspective and the observed structural reduction, at both the axonemal and kinetosome levels, has a direct explanation in their reduced flagellar toolkit composition. To better characterize this reduction, we analysed the presence/absence matrix of the flagellum toolkit components by similarity clustering (Figure 3B). We observed that sanchytrids clustered in an intermediate position between flagellated and non-flagellated lineages, indicating that they are a lineage engaged in an ongoing, but not finished, process of complete loss of the flagellum. These organisms are, therefore, interesting models to study the intermediate phases of this process.

In addition to this flagellum degeneration in sanchytrids, complete flagellum loss has been estimated to have happened between 4 and 6 times in Holomycota (James *et al.*, 2006a). If we consider our new phylogenetic framework, including the position of sanchytrids as sister group of Blastocladiomycota, chytrids as sister group of all other fungi, *Olpidium* as an independent new lineage, and the Monoblepharidomycete *Hyaloraphidium curvatum* nested within chytrids, we only infer 4 independent events of complete flagellum loss, plus the ongoing one occurring in sanchytrids (Figure 4). Three are high-rank losses that occurred at the base of the nucleariid clade, the Microsporidia clade, and the Zoopagomycota + Mucoromycota + Dikarya clade. The fourth one occurred in *H. curvatum*, an atypical fungus originally classified as a colourless green alga

(Ustinova *et al.*, 2000) and later reclassified within the Monoblepharidomycetes (Forget *et al.*, 2002; Sekimoto *et al.*, 2011). A putative fifth independent flagellum loss within Holomycota could be the one of nephridiophagids (Nephridiophagida), a clade of non-flagellated eukaryotes which parasitize the Malpighian tubules of insects and myriapods (Fabel, Radek, & Storch, 2000; Radek *et al.*, 2017). Recently, they have been placed within the Fungi using 18S rRNA gene phylogenies (Radek *et al.*, 2017), but they lack affinity with any known fungal lineage. Genomic/transcriptomic data from this lineage will be necessary to clarify their phylogenetic position.

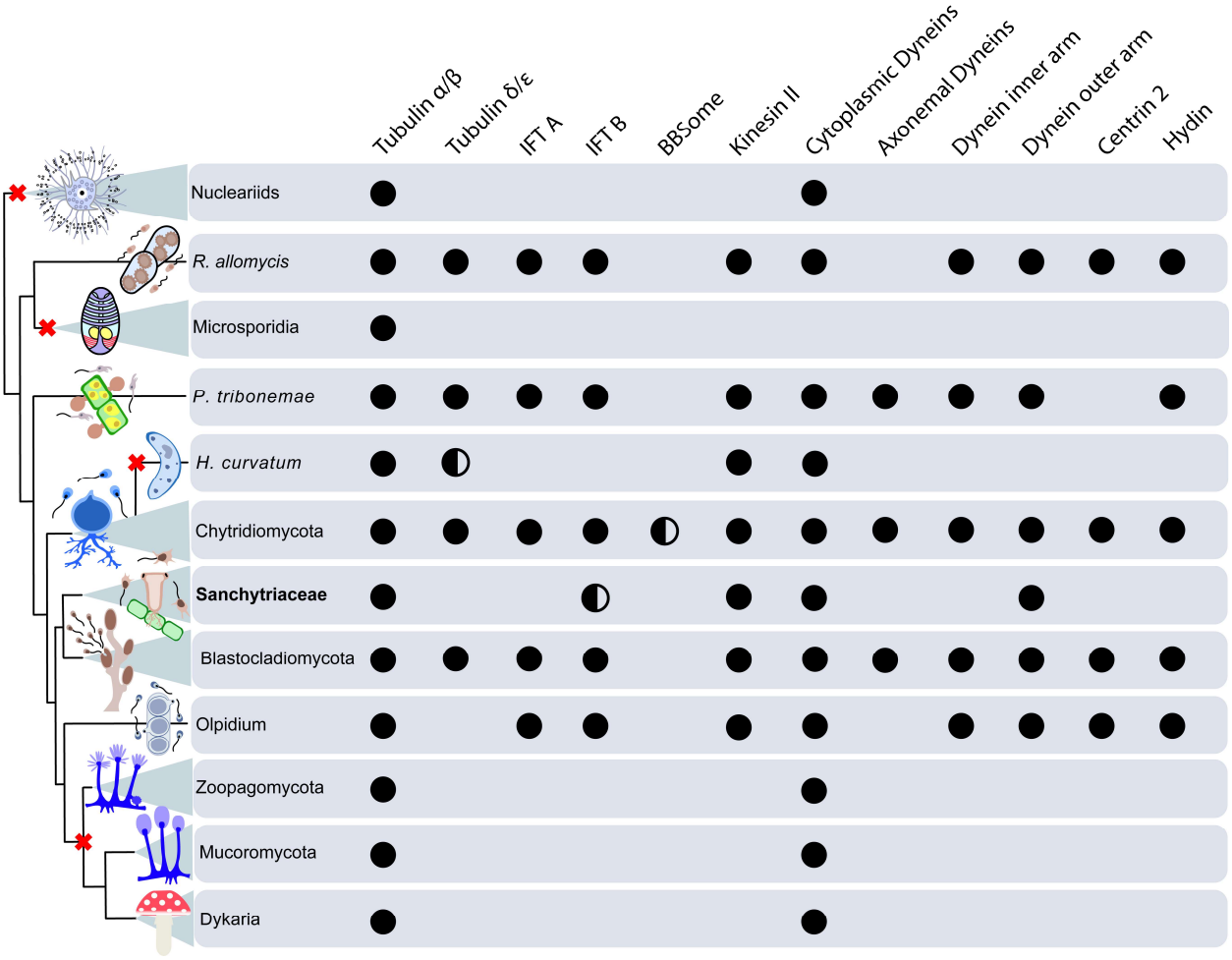


Figure 4. Flagellum loss and reduction in Holomycota. Evolutionary relationships of holomycotan lineages and their patterns of presence/absence of key molecular components of the flagellar apparatus. Black dots on branches indicate complete independent flagellum loss in a lineage.

So far, attempts to place *Olpidium* within a global phylogeny of Fungi have only led to a poorly-supported placement within the Zoopagomycota, usually in association with Basidiobolomycetes (James *et al.*, 2006a; Sekimoto *et al.*, 2011; Tedersoo *et al.*, 2018). Its correct

location in the tree of Fungi remained critical to estimate the number and timing of flagellar losses. Our phylogenomic analyses strongly supported that *Olpidium* forms an independent lineage sister to the main clade of non-flagellated fungi (Figure 1B). We have also shown that, even if *Olpidium* has a functional flagellum, its flagellar toolkit is reduced in comparison with other zoosporic fungi, lacking elements like the Delta and Epsilon tubulins and several dyneins (Figure 4). This adds support to the fact that Olpidiomycota represent an intermediate lineage between zoosporic fungi with fully functional flagella and the species with complete flagellum loss (Zygomycota and Dikarya). This model also agrees with Zoopagomycota being the first major lineage diverging after Olpidiomycota, as it has been observed that these fungi possess a degenerated 9+2 microtubular system that may be a relict of an ancestral flagellum (McKerracher & Heath, 1985; Roberson *et al.*, 2011; McLaughlin *et al.*, 2015). Thus, we propose a global scenario in which Fungi would have undergone a major progressive reduction of the flagellar toolkit from *Olpidium* to Dikarya, with a few additional punctual independent losses or reductions in some distant branches of the tree (*H. curvatum* and sanchytrids).

Light sensing in sanchytrids. Although we have clarified some aspects of the extreme simplification of the atypical sanchytrid flagellar organization, the question of why these organisms have at the same time a reduced flagellum and an unusually long and developed kinetosome remains unanswered. What is true is that selection seems to be maintaining sanchytrids reduced flagellum, rather than completely losing it. One possibility is that, even if the primary flagellar function has been lost in favour of an amoeboid movement, the remaining flagellar structure may have been kept because it was exapted in favour of a new function for the zoospore. This phenomenon has been observed in bacteria, in which the flagellum acquired new roles such as mechanosensitivity (Belas, Simon, & Silverman, 1986; Belas, 2014) and wetness sensing (Wang *et al.*, 2005).

Observation of living cultures has shown that the sanchytrid flagellum is rather labile and can be totally retracted within the cell cytoplasm, with its long kinetosome most likely being involved in this retraction capability (Karpov *et al.*, 2018, 2019). An atypical ultrastructural feature of this kinetosome was the presence of large lipid globules fused in a huge curved rosary chain surrounding the kinetosome structure. These globules were the most conspicuous intracellular structures seen in *A. gromovi* zoospores, often closely associated with mitochondria (Karpov *et*

al., 2018, 2019). In the closely related Blastocladiomycota *B. emersonii*, such structures are also present and have been called "side-body complexes", also closely associated with mitochondria (Lovett, 1975). Recently, it has been found in this species a unique gene fusion of a rhodopsin domain and a guanylyl cyclase domain (BeGC1) of 626 amino acids derived from a bacterial type 1 rhodopsin which, together with a cyclic nucleotide-gated channel (BeCNG1), control phototaxis in zoospores in response to levels of cGMP after exposure to green light (Avelar *et al.*, 2014). The BeGC1 fusion protein was localized by immunofluorescence on the external membrane of the lipid droplets, which function as an eyespot at the base of the flagellum and controls its beating (Avelar *et al.*, 2014, 2015; Richards & Gomes, 2015). The BeGC1 fusion and the channel BeCNG1 proteins have also been found in other blastocladiomycotan species such as *A. macrogynus* and *Catenaria anguillulae*.

We searched for these genes in our sanchytrid genomes and found that both *A. gromovi* and *S. tribonematis* possessed the fused BeGC1 (532 and 535 amino acids, respectively) and the gated channel BeCNG1 (Supplementary figures 6A-C). We also found the BeGC1 fusion in the newly available Blastocladiomycota genomes of *P. sedebokerense* and *Coelomomyces lativittatus*. Therefore, this fusion constitutes a synapomorphy for the whole Blastocladiomycota + Sanchytriaceae clade. On the contrary, the BeCNG1 channel was found in *C. lativittatus* but not in *P. sedebokerense*, suggesting that this species probably lost the capacity of light sensing or that it uses a different protein non-homologous to BeCNG1 to finish the cGMP cascade reaction (Supplementary figure 6C). No phototaxis, functional or cellular localization studies have yet been made for these proteins in sanchytrid zoospores, so we cannot assure that sanchytrids also have an eyespot with a cGMP cascade like in *B. emersonii*. However, the presence of the lipid organelle closely associated with mitochondria and the kinetosome points towards the presence of a similar light-sensing organelle in sanchytrids. If this is confirmed, the maintenance of the reduced flagellum would respond to a selective pressure towards light perception. It is interesting to highlight that sanchytrids showed branch lengths in phylogenies of the rhodopsin domain and the guanylyl cyclase domain that were considerably shorter than those shown in the multi-gene phylogenies, indicating that, in fact, these proteins are subjected to a strong selective pressure and play a key role in these organisms.

Acknowledgments

This work was funded by the European Research Council Advanced Grants "ProtistWorld" (No. 322669) and Plast-Evol (No. 787904) and the Horizon 2020 research and innovation program under the Marie Skłodowska-Curie ITN project SINGEK (http://www.singek.eu/; grant agreement No. H2020-MSCA-ITN-2015-675752).

References

AGUILETA, G., DE VIENNE, D.M., ROSS, O.N., HOOD, M.E., GIRAUD, T., PETIT, E. & GABALDÓN, T. (2014) High variability of mitochondrial gene order among fungi. *Genome Biology and Evolution*.

ALTSCHUL, S.F., GISH, W., MILLER, W., MYERS, E.W. & LIPMAN, D.J. (1990) Basic local alignment search tool. *Journal of Molecular Biology* **215**, 403–410.

ANDREWS, S. (2010) FastQC: A quality control tool for high throughput sequence data. *Http://Www.Bioinformatics.Babraham.Ac.Uk/Projects/Fastqc/*. .

ANISIMOVA, M., GIL, M., DUFAYARD, J.F., DESSIMOZ, C. & GASCUEL, O. (2011) Survey of branch support methods demonstrates accuracy, power, and robustness of fast likelihood-based approximation schemes. *Systematic Biology*.

ARCHIBALD, J.M., SIMPSON, A.G.B., SLAMOVITS, C.H., MARGULIS, L., MELKONIAN, M., CHAPMAN, D.J. & CORLISS, J.O. (eds) (2017) *Handbook of the Protists*. In *Handbook of the Protists* p. Springer International Publishing, Cham.

AUBUSSON-FLEURY, A., BALAVOINE, G., LEMULLOIS, M., BOUHOUCHE, K., BEISSON, J. & KOLL, F. (2017) Centrin diversity and basal body patterning across evolution: New insights from Paramecium. *Biology Open* **6**, 765–776.

AVELAR, G.M., GLASER, T., LEONARD, G., RICHARDS, T.A., ULRICH, H. & GOMES, S.L. (2015) A cyclic GMP-dependent K+ channel in the blastocladiomycete fungus Blastocladiella emersonii. *Eukaryotic Cell* **14**, 958–963.

AVELAR, G.M., SCHUMACHER, R.I., ZAINI, P.A., LEONARD, G., RICHARDS, T.A. & GOMES, S.L. (2014) A Rhodopsin-Guanylyl cyclase gene fusion functions in visual perception in a fungus. *Current Biology* **24**, 1234–1240. Elsevier.

BANKEVICH, A., NURK, S., ANTIPOV, D., GUREVICH, A.A., DVORKIN, M., KULIKOV, A.S., LESIN, V.M., NIKOLENKO, S.I., PHAM, S., PRJIBELSKI, A.D., PYSHKIN, A. V., SIROTKIN, A. V., VYAHHI, N., TESLER, G., ALEKSEYEV, M.A., ET AL. (2012) SPAdes: A New Genome Assembly Algorithm and Its Applications to Single-Cell Sequencing. *Journal of Computational Biology* **19**, 455–477.

BARR, D.J.S. (1980) An outline for the reclassification of the Chytridiales, and for a new order, the Spizellomycetales. *Canadian Journal of Botany* **58**, 2380–2394.

BASS, D., CZECH, L., WILLIAMS, B.A.P., BERNEY, C., DUNTHORN, M., MAHÉ, F., TORRUELLA, G., STENTIFORD,

G.D. & WILLIAMS, T.A. (2018) Clarifying the Relationships between Microsporidia and Cryptomycota. *Journal of Eukaryotic Microbiology* **65**, 773–782.

BELAS, R. (2014) Biofilms, flagella, and mechanosensing of surfaces by bacteria. *Trends in Microbiology* **22**, 517–527. Elsevier Ltd.

BELAS, R., SIMON, M. & SILVERMAN, M. (1986) Regulation of lateral flagella gene transcription in Vibrio parahaemolyticus. *Journal of Bacteriology* **167**, 210–218.

BERBEE, M.L., JAMES, T.Y. & STRULLU-DERRIEN, C. (2017) Early Diverging Fungi: Diversity and Impact at the Dawn of Terrestrial Life. *Annual Review of Microbiology* **71**, 41–60.

BERNT, M., DONATH, A., JÜHLING, F., EXTERNBRINK, F., FLORENTZ, C., FRITZSCH, G., PÜTZ, J., MIDDENDORF, M. & STADLER, P.F. (2013) MITOS: Improved de novo metazoan mitochondrial genome annotation. *Molecular Phylogenetics and Evolution*.

BIDARTONDO, M.I., READ, D.J., TRAPPE, J.M., MERCKX, V., LIGRONE, R. & DUCKETT, J.G. (2011) The dawn of symbiosis between plants and fungi. *Biology Letters*.

BILLON-GRAND, G., FIOL, J.B., BRETON, A., BRUYÈRE, A. & OULHAJ, Z. (1991) DNA of some anaerobic rumen fungi: G + C content determination. *FEMS Microbiology Letters*.

BOLGER, A.M., LOHSE, M. & USADEL, B. (2014) Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120.

BRINKMANN, H. & PHILIPPE, H. (1999) Archaea sister group of Bacteria? Indications from tree reconstruction artifacts in ancient phylogenies. *Molecular Biology and Evolution*.

BROWN, M.W., SPIEGEL, F.W. & SILBERMAN, J.D. (2009) Phylogeny of the 'forgotten' cellular slime mold, Fonticula alba, reveals a key evolutionary branch within Opisthokonta. *Molecular Biology and Evolution*.

CAMACHO, C., COULOURIS, G., AVAGYAN, V., MA, N., PAPADOPOULOS, J., BEALER, K. & MADDEN, T.L. (2009) BLAST+: Architecture and applications. *BMC Bioinformatics*.

CAPELLA-GUTIÉRREZ, S., SILLA-MARTÍNEZ, J.M. & GABALDÓN, T. (2009) trimAl: A tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* **25**, 1972–1973.

CARVALHO-SANTOS, Z., AZIMZADEH, J., PEREIRA-LEAL, J.B. & BETTENCOURT-DIAS, M. (2011) Tracing the origins of centrioles, cilia, and flagella. *Journal of Cell Biology*. .

CAVALIER-SMITH, T. (1998) A revised six-kingdom system of life. *Biological Reviews of the Cambridge Philosophical Society* **73**, 203–266.

CAVALIER-SMITH, T. & CHAO, E.E.Y. (2003) Phylogeny of choanozoa, apusozoa, and other protozoa and early eukaryote megaevolution. *Journal of Molecular Evolution*.

CHANG, Y., WANG, S., SEKIMOTO, S., AERTS, A.L., CHOI, C., CLUM, A., LABUTTI, K.M., LINDQUIST, E.A., NGAN, C.Y., OHM, R.A., SALAMOV, A.A., GRIGORIEV, I. V., SPATAFORA, J.W. & BERBEE, M.L. (2015) Phylogenomic analyses indicate that early fungi evolved digesting cell walls of algal ancestors of land plants. *Genome Biology and Evolution* **7**, 1590–1601.

CHEN, Y. PING, PETTIS, J.S., ZHAO, Y., LIU, X., TALLON, L.J., SADZEWICZ, L.D., LI, R., ZHENG, H., HUANG, S., ZHANG, X., HAMILTON, M.C., PERNAL, S.F., MELATHOPOULOS, A.P., YAN, X. & EVANS, J.D. (2013) Genome

sequencing and comparative genomics of honey bee microsporidia, Nosema apis reveal novel insights into host-parasite interactions. *BMC Genomics*.

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. *Parasitology Research* **113**, 1909–1918.

COUCH, J.N. (1962) VALIDATION OF THE FAMILY COELOMOMYCETACEAE AND CERTAIN SPECIES AND VARIETIES OF COELOMOMYCES. *Journal of the Elisha Mitchell Scientific Society* **78**, 135–138. Temporary Publisher.

CUOMO, C.A., DESJARDINS, C.A., BAKOWSKI, M.A., GOLDBERG, J., MA, A.T., BECNEL, J.J., DIDIER, E.S., FAN, L., HEIMAN, D.I., LEVIN, J.Z., YOUNG, S., ZENG, Q. & TROEMEL, E.R. (2012) Microsporidian genome analysis reveals evolutionary strategies for obligate intracellular growth. *Genome Research*.

VAN DAM, T.J.P., TOWNSEND, M.J., TURK, M., SCHLESSINGER, A., SALI, A., FIELD, M.C. & HUYNEN, M.A. (2013) Evolution of modular intraflagellar transport from a coatomer-like progenitor. *Proceedings of the National Academy of Sciences of the United States of America* **110**, 6943–6948.

DOUZERY, E.J.P., SNELL, E.A., BAPTESTE, E., DELSUC, F. & PHILIPPE, H. (2004) The timing of eukaryotic evolution: Does a relaxed molecular clock reconcile proteins and fossils? *Proceedings of the National Academy of Sciences of the United States of America*.

DUPUIS-WILLIAMS, P., FLEURY-AUBUSSON, A., DE LOUBRESSE, N.G., GEOFFROY, H., VAYSSIÉ, L., GALVANI, A., ESPIGAT, A. & ROSSIER, J. (2002) Functional role of ε-tubulin in the assembly of the centriolar microtubule scaffold. *Journal of Cell Biology* **158**, 1183–1193.

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. *Molecular Biology and Evolution* **29**, 1319–1334.

EGGER, B., BACHMANN, L. & FROMM, B. (2017) Atp8 is in the ground pattern of flatworm mitochondrial genomes. *BMC Genomics*.

EME, L., SHARPE, S.C., BROWN, M.W. & ROGER, A.J. (2014) On the Age of Eukaryotes: Evaluating Evidence from Fossils and Molecular Clocks. *Cold Spring Harbor Perspectives in Biology* **6**, 1–16.

FABEL, P., RADEK, R. & STORCH, V. (2000) A new spore-forming protist, Nephridiophaga blaberi sp. nov., in the Death's head cockroach Blaberus craniifer. *European Journal of Protistology*.

FEIJEN, F.A.A., VOS, R.A., NUYTINCK, J. & MERCKX, V.S.F.T. (2018) Evolutionary dynamics of mycorrhizal symbiosis in land plant diversification. *Scientific Reports*.

FELSENSTEIN, J. (1993) PHYLIP (Phylogenetic Inference Package, version 3.5c). *Dpartement of Genetics, SK, University of Washington, Seattle, WA*. .

FIELD, K.J., RIMINGTON, W.R., BIDARTONDO, M.I., ALLINSON, K.E., BEERLING, D.J., CAMERON, D.D., DUCKETT, J.G., LEAKE, J.R. & PRESSEL, S. (2015) First evidence of mutualism between ancient plant lineages (Haplomitriopsida liverworts) and Mucoromycotina fungi and its response to simulated Palaeozoic changes in atmospheric CO2. *New Phytologist*.

FORGET, L., USTINOVA, J., WANG, Z., HUSS, V.A.R. & LANG, B.F. (2002) Hyaloraphidium curvatum: A linear

mitochondrial genome, tRNA editing, and an evolutionary link to lower fungi. *Molecular Biology and Evolution*.

FREEMAN, K.R., MARTIN, A.P., KARKI, D., LYNCH, R.C., MITTER, M.S., MEYERA, A.F., LONGCORE, J.E., SIMMONS, D.R. & SCHMIDT, S.K. (2009) Evidence that chytrids dominate fungal communities in high-elevation soils. *Proceedings of the National Academy of Sciences of the United States of America*.

FRENKEN, T., ALACID, E., BERGER, S.A., BOURNE, E.C., GERPHAGNON, M., GROSSART, H.P., GSELL, A.S., IBELINGS, B.W., KAGAMI, M., KÜPPER, F.C., LETCHER, P.M., LOYAU, A., MIKI, T., NEJSTGAARD, J.C., RASCONI, S., ET AL. (2017) Integrating chytrid fungal parasites into plankton ecology: research gaps and needs. *Environmental Microbiology*. .

FRENKEN, T., VELTHUIS, M., DE SENERPONT DOMIS, L.N., STEPHAN, S., ABEN, R., KOSTEN, S., VAN DONK, E. & VAN DE WAAL, D.B. (2016) Warming accelerates termination of a phytoplankton spring bloom by fungal parasites. *Global Change Biology*.

GALINDO, L.J., TORRUELLA, G., MOREIRA, D., EGLIT, Y., SIMPSON, A.G.B., VÖLCKER, E., CLAUß, S. & LÓPEZ-GARCÍA, P. (2019) Combined cultivation and single-cell approaches to the phylogenomics of nucleariid amoebae, close relatives of fungi. *Philosophical Transactions of the Royal Society B: Biological Sciences* **374**, 20190094.

GALINDO, L.J., TORRUELLA, G., MOREIRA, D., TIMPANO, H., PASKEROVA, G., SMIRNOV, A., NASSONOVA, E. & LÓPEZ-GARCÍA, P. (2018) Evolutionary genomics of metchnikovella incurvata (metchnikovellidae): An early branching microsporidium. *Genome Biology and Evolution*.

GISSI, C., IANNELLI, F. & PESOLE, G. (2008) Evolution of the mitochondrial genome of Metazoa as exemplified by comparison of congeneric species. *Heredity*. .

GUREVICH, A., SAVELIEV, V., VYAHHI, N. & TESLER, G. (2013) QUAST: Quality assessment tool for genome assemblies. *Bioinformatics* **29**, 1072–1075.

GUTMAN, J., ZARKA, A. & BOUSSIBA, S. (2009) The host-range of Paraphysoderma sedebokerensis, a chytrid that infects Haematococcus pluvialis. *European Journal of Phycology*.

HAAG, K.L., JAMES, T.Y., POMBERT, J.-F., LARSSON, R., SCHAER, T.M.M., REFARDT, D. & EBERT, D. (2014) Evolution of a morphological novelty occurred before genome compaction in a lineage of extreme parasites. *Proceedings of the National Academy of Sciences* **111**, 15480–15485.

HIBBETT, D.S., BINDER, M., BISCHOFF, J.F., BLACKWELL, M., CANNON, P.F., ERIKSSON, O.E., HUHNDORF, S., JAMES, T., KIRK, P.M., LÜCKING, R., THORSTEN LUMBSCH, H., LUTZONI, F., MATHENY, P.B., MCLAUGHLIN, D.J., POWELL, M.J., ET AL. (2007) A higher-level phylogenetic classification of the Fungi. *Mycological Research*.

HOFFMAN, Y., AFLALO, C., ZARKA, A., GUTMAN, J., JAMES, T.Y. & BOUSSIBA, S. (2008) Isolation and characterization of a novel chytrid species (phylum Blastocladiomycota), parasitic on the green alga Haematococcus. *Mycological Research*.

HUERTA-CEPAS, J., FORSLUND, K., COELHO, L.P., SZKLARCZYK, D., JENSEN, L.J., VON MERING, C. & BORK, P. (2017) Fast genome-wide functional annotation through orthology assignment by eggNOG-mapper. *Molecular Biology and Evolution* **34**, 2115–2122.

JAMES, T.Y., HOFFMAN, Y., ZARKA, A. & BOUSSIBA, S. (2011) Paraphysoderma sedebokerense, gen. et sp. Nov., an aplanosporic relative of Physoderma (Blastocladiomycota). *Mycotaxon*.

JAMES, T.Y., KAUFF, F., SCHOCH, C.L., MATHENY, P.B., HOFSTETTER, V., COX, C.J., CELIO, G., GUEIDAN, C., FRAKER, E., MIADLIKOWSKA, J., LUMBSCH, H.T., RAUHUT, A., REEB, V., ARNOLD, A.E., AMTOFT, A., ET AL. (2006a) Reconstructing the early evolution of Fungi using a six-gene phylogeny. *Nature* **443**, 818–822.

JAMES, T.Y., LETCHER, P.M., LONGCORE, J.E., MOZLEY-STANDRIDGE, S.E., PORTER, D., POWELL, M.J., GRIFFITH, G.W. & VILGALYS, R. (2006b) A molecular phylogeny of the flagellated fungi (Chytridiomycota) and description of a new phylum (Blastocladiomycota). *Mycologia* **98**, 860–871.

JAMES, T.Y., PELIN, A., BONEN, L., AHRENDT, S., SAIN, D., CORRADI, N. & STAJICH, J.E. (2013) Shared signatures of parasitism and phylogenomics unite cryptomycota and microsporidia. *Current Biology* **23**, 1548–1553. Elsevier Ltd.

JONES, M.D.M., FORN, I., GADELHA, C., EGAN, M.J., BASS, D., MASSANA, R. & RICHARDS, T.A. (2011) Discovery of novel intermediate forms redefines the fungal tree of life. *Nature*.

KAGAMI, M., MIKI, T. & TAKIMOTO, G. (2014) Mycoloop: Chytrids in aquatic food webs. *Frontiers in Microbiology*.

KAMAISHI, T., HASHIMOTO, T., NAKAMURA, Y., MASUDA, Y., NAKAMURA, F., OKAMOTO, K. -I., SHIMIZU, M. & HASEGAWA, M. (1996) Complete Nucleotide Sequences of the Genes Encoding Translation Elongation Factors 1 and 2 from a microsporidian parasite, Glugea plecoglossi: Implications for the Deepest Branching of Eukaryotes. *Journal of Biochemistry* **120**, 1095–1103.

KANEHISA, M. (2000) KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Research*.

KANEHISA, M., SATO, Y. & MORISHIMA, K. (2016) BlastKOALA and GhostKOALA: KEGG Tools for Functional Characterization of Genome and Metagenome Sequences. *Journal of Molecular Biology*.

KARPOV, S.A., LÓPEZ-GARCÍA, P., MAMKAEVA, M.A., KLIMOV, V.I., VISHNYAKOV, A.E., TCVETKOVA, V.S. & MOREIRA, D. (2018) The Chytrid-like Parasites of Algae Amoeboradix gromovi gen. et sp. nov. and Sanchytrium tribonematis Belong to a New Fungal Lineage. *Protist* **169**, 122–140. Elsevier GmbH.

KARPOV, S.A., MAMANAZAROVA, K.S., POPOVA, O. V., ALEOSHIN, V. V., JAMES, T.Y., MAMKAEVA, M.A., TCVETKOVA, V.S., VISHNYAKOV, A.E. & LONGCORE, J.E. (2017a) Monoblepharidomycetes diversity includes new parasitic and saprotrophic species with highly intronized rDNA. *Fungal Biology* **121**, 729–741. Elsevier Ltd.

KARPOV, S.A., MAMKAEVA, M.A., ALEOSHIN, V. V., NASSONOVA, E., LILJE, O. & GLEASON, F.H. (2014) Morphology, phylogeny, and ecology of the aphelids (Aphelidea, Opisthokonta) and proposal for the new superphylum Opisthosporidia. *Frontiers in Microbiology*. .

KARPOV, S.A., TCVETKOVA, V.S., MAMKAEVA, M.A., TORRUELLA, G., TIMPANO, H., MOREIRA, D., MAMANAZAROVA, K.S. & LÓPEZ-GARCÍA, P. (2017b) Morphological and Genetic Diversity of Opisthosporidia: New Aphelid Paraphelidium tribonemae gen. et sp. nov. *Journal of Eukaryotic Microbiology*.

KARPOV, S.A., VISHNYAKOV, A.E., MOREIRA, D. & LÓPEZ-GARCÍA, P. (2019) The Ultrastructure of Sanchytrium tribonematis (Sanchytriaceae, Fungi incertae sedis) Confirms its Close Relationship to Amoeboradix. *Journal*

of Eukaryotic Microbiology, 0–2.

KATOH, K. & STANDLEY, D.M. (2013) MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. *Molecular Biology and Evolution* **30**, 772–780.

KEARSE, M., MOIR, R., WILSON, A., STONES-HAVAS, S., CHEUNG, M., STURROCK, S., BUXTON, S., COOPER, A., MARKOWITZ, S., DURAN, C., THIERER, T., ASHTON, B., MEINTJES, P. & DRUMMOND, A. (2012) Geneious Basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* **28**, 1647–1649.

LAETSCH, D.R. & BLAXTER, M.L. (2017) BlobTools: Interrogation of genome assemblies. *F1000Research* **6**, 1287.

LARA, E., MOREIRA, D. & LÓPEZ-GARCÍA, P. (2010) The Environmental Clade LKM11 and Rozella Form the Deepest Branching Clade of Fungi. *Protist* **161**, 116–121.

LARTILLOT, N., LEPAGE, T. & BLANQUART, S. (2009) PhyloBayes 3: A Bayesian software package for phylogenetic reconstruction and molecular dating. *Bioinformatics*.

LARTILLOT, N. & PHILIPPE, H. (2004) A Bayesian mixture model for across-site heterogeneities in the amino-acid replacement process. *Molecular Biology and Evolution* **21**, 1095–1109.

LEIPE, D.D., GUNDERSON, J.H., NERAD, T.A. & SOGIN, M.L. (1993) SMALL SUBUNIT RIBOSOMAL RNA+ OF HEXAMITA-INFLATA AND THE QUEST FOR THE 1ST BRANCH IN THE EUKARYOTIC TREE. *Molecular And Biochemical Parasitology*. .

LETCHER, P.M., LEE, P.A., LOPEZ, S., BURNETT, M., MCBRIDE, R.C. & POWELL, M.J. (2016) An ultrastructural study of Paraphysoderma sedebokerense (Blastocladiomycota), an epibiotic parasite of microalgae. *Fungal Biology* **120**, 324–337. Elsevier Ltd.

LETCHER, P.M. & POWELL, M.J. (2018) A taxonomic summary and revision of Rozella (Cryptomycota). *IMA Fungus* **9**, 383–399.

LI, W. & GODZIK, A. (2006) Cd-hit: A fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics*.

LIU, Y., STEENKAMP, E.T., BRINKMANN, H., FORGET, L., PHILIPPE, H. & LANG, B.F. (2009) Phylogenomic analyses predict sistergroup relationship of nucleariids and Fungi and paraphyly of zygomycetes with significant support. *BMC Evolutionary Biology* **9**, 1–11.

LIU, Y.J., HODSON, M.C. & HALL, B.D. (2006) Loss of the flagellum happened only once in the fungal lineage: Phylogenetic structure of Kingdom Fungi inferred from RNA polymerase II subunit genes. *BMC Evolutionary Biology* **6**, 1–13.

LORON, C.C., FRANÇOIS, C., RAINBIRD, R.H., TURNER, E.C., BORENSZTAJN, S. & JAVAUX, E.J. (2019) Early fungi from the Proterozoic era in Arctic Canada. *Nature*. .

LOVETT, J.S. (1975) Growth and differentiation of the water mold Blastocladiella emersonii: cytodifferentiation and the role of ribonucleic acid and protein synthesis. *Bacteriological Reviews* **39**, 345–404.

LUTZONI, F., NOWAK, M.D., ALFARO, M.E., REEB, V., MIADLIKOWSKA, J., KRUG, M., ARNOLD, A.E., LEWIS, L.A., SWOFFORD, D.L., HIBBETT, D., HILU, K., JAMES, T.Y., QUANDT, D. & MAGALLÓN, S. (2018) Contemporaneous radiations of fungi and plants linked to symbiosis. *Nature Communications* **9**, 1–11.

Springer US.

McKerracher, L.J. & Heath, I.B. (1985) The Structure and Cycle of the Nucleus-Associated Organelle in Two Species of Basidiobolus. *Mycologia*.

McLaughlin, D.J., Healy, R.A., Celio, G.J., Roberson, R.W. & Kumar, T.K.A. (2015) Evolution of zygomycetous spindle pole bodies: Evidence from Coemansia reversa mitosis. *American Journal of Botany*.

Mesquite Project Team (2014) Mesquite: A modular system for evolutionary analysis. *Available from: http:// mesquiteproject.wikispaces.com/home*. .

Mikhailov, K. V., Simdyanov, T.G. & Aleoshin, V. V. (2017) Genomic survey of a hyperparasitic microsporidian Amphiamblys sp. (Metchnikovellidae). *Genome Biology and Evolution*.

Mikhailov, K. V, Simdyanov, T.G., Aleoshin, V. V & Belozersky, A.N. (2016) Genomic survey of a hyperparasitic microsporidian Amphiamblys sp. (Metchnikovellidae) Genome Biology and Evolution Advance Access. *at University of Library and Information Science on*.

Minh, B.Q., Nguyen, M.A.T. & Von Haeseler, A. (2013) Ultrafast approximation for phylogenetic bootstrap. *Molecular Biology and Evolution*.

Mitchell, D.R. (2004) Speculations on the evolution of 9+2 organelles and the role of central pair microtubules. *Biology of the Cell*. .

Nakjang, S., Williams, T.A., Heinz, E., Watson, A.K., Foster, P.G., Sendra, K.M., Heaps, S.E., Hirt, R.P. & Embley, T.M. (2013) Reduction and expansion inmicrosporidian genome evolution: New insights from comparative genomics. *Genome Biology and Evolution*.

Naranjo-Ortiz, M.A. & Gabaldón, T. (2019) Fungal evolution: major ecological adaptations and evolutionary transitions. *Biological Reviews* **94**, 1443–1476.

Nguyen, L.T., Schmidt, H.A., Von Haeseler, A. & Minh, B.Q. (2015) IQ-TREE: A fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Molecular Biology and Evolution* **32**, 268–274.

Okonechnikov, K., Conesa, A. & García-Alcalde, F. (2015) Qualimap 2: Advanced multi-sample quality control for high-throughput sequencing data. *Bioinformatics* **32**, 292–294.

Olive, L.S. (1980) Caulochytrium protostelioides sp. nov., a New Chytrid with Aerial Sporangia. *American Journal of Botany*.

Parfrey, L.W., Lahr, D.J.G., Knoll, A.H. & Katz, L.A. (2011) Estimating the timing of early eukaryotic diversification with multigene molecular clocks. *Proceedings of the National Academy of Sciences of the United States of America*.

Philippe, H., Lopez, P., Brinkmann, H., Budin, K., Germot, A., Laurent, J., Moreira, D., Müller, M. & Le Guyader, H. (2000) Early-branching or fast-evolving eukaryotes? An answer based on slowly evolving positions. *Proceedings of the Royal Society B: Biological Sciences* **267**, 1213–1221.

Powell, M.J. (1993) Looking at mycology with a Janus face: a glimpse at Chytridiomycetes active in the environment. *Mycologia*.

Powell, M.J. (2017a) Chytridiomycota. In *Handbook of the Protists: Second Edition* p.

Powell, M.J. (2017b) Blastocladiomycota. In *Handbook of the Protists: Second Edition* p.

POWELL, M.J. & LETCHER, P.M. (2014) Chytridiomycota, Monoblepharidomycota, and Neocallimastigomycota. In *Systematics and Evolution: Part A: Second Edition* p.

PRICE, M.N., DEHAL, P.S. & ARKIN, A.P. (2009) Fasttree: Computing large minimum evolution trees with profiles instead of a distance matrix. *Molecular Biology and Evolution*.

R DEVELOPMENT CORE TEAM, R. (2011) *R: A Language and Environment for Statistical Computing*. In *R Foundation for Statistical Computing* p.

RADEK, R., WURZBACHER, C., GISDER, S., NILSSON, R.H., OWERFELDT, A., GENERSCH, E., KIRK, P.M. & VOIGT, K. (2017) Morphologic and molecular data help adopting the insect-pathogenic nephridiophagids (Nephridiophagidae) among the early diverging fungal lineages, close to the Chytridiomycota. *MycoKeys*.

RAMBAUT, A. (2016) FigTree v1.4.3. *Molecular evolution, phylogenet ics and epidemiology*. .

RICHARDS, T.A. & GOMES, S.L. (2015) How to build a microbial eye. *Nature* **523**, 166–167.

ROBERSON, R.W., SAUCEDO, E., MACLEAN, D., PROPSTER, J., UNGER, B., ONEIL, T.A., PARVANEHGOHAR, K., CAVANAUGH, C. & LOWRY, D. (2011) The hyphal tip structure of Basidiobolus sp.: A zygomycete fungus of uncertain phylogeny. *Fungal Biology*.

SEKIMOTO, S., ROCHON, D., LONG, J.E., DEE, J.M. & BERBEE, M.L. (2011) A multigene phylogeny of Olpidium and its implications for early fungal evolution. *BMC Evolutionary Biology* **11**, 331. BioMed Central Ltd.

SIMÃO, F.A., WATERHOUSE, R.M., IOANNIDIS, P., KRIVENTSEVA, E. V. & ZDOBNOV, E.M. (2015) BUSCO: Assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* **31**, 3210–3212.

SPARROW, F.K. (1960) Aquatic Phycomycetes. *University of Michigan Press* **132**.

SPATAFORA, J.W., AIME, M.C., GRIGORIEV, I. V., MARTIN, F., STAJICH, J.E. & BLACKWELL, M. (2017) The Fungal Tree of Life: from Molecular Systematics to Genome-Scale Phylogenies. *Microbiology Spectrum* **5**, 1–32.

SPATAFORA, J.W., CHANG, Y., BENNY, G.L., LAZARUS, K., SMITH, M.E., BERBEE, M.L., BONITO, G., CORRADI, N., GRIGORIEV, I., GRYGANSKYI, A., JAMES, T.Y., O'DONNELL, K., ROBERSON, R.W., TAYLOR, T.N., UEHLING, J., ET AL. (2016) A phylum-level phylogenetic classification of zygomycete fungi based on genome-scale data. *Mycologia* **108**, 1028–1046.

STAJICH, J.E., BERBEE, M.L., BLACKWELL, M., HIBBETT, D.S., JAMES, T.Y., SPATAFORA, J.W. & TAYLOR, J.W. (2009) The Fungi. *Current Biology*. .

STRITTMATTER, M., GUERRA, T., SILVA, J. & GACHON, C.M.M. (2016) A new flagellated dispersion stage in Paraphysoderma sedebokerense, a pathogen of Haematococcus pluvialis. *Journal of Applied Phycology* **28**, 1553–1558. Journal of Applied Phycology.

SUSKO, E. & ROGER, A.J. (2007) On reduced amino acid alphabets for phylogenetic inference. *Molecular Biology and Evolution*.

TAYLOR, J.W. & BERBEE, M.L. (2006) Dating divergences in the Fungal Tree of Life: Review and new analyses. *Mycologia*. .

TEDERSOO, L., BAHRAM, M., PUUSEPP, R., NILSSON, R.H. & JAMES, T.Y. (2017) Novel soil-inhabiting clades fill gaps in the fungal tree of life. *Microbiome*.

TEDERSOO, L., SÁNCHEZ-RAMÍREZ, S., KÕLJALG, U., BAHRAM, M., DÖRING, M., SCHIGEL, D., MAY, T., RYBERG, M. & ABARENKOV, K. (2018) High-level classification of the Fungi and a tool for evolutionary ecological analyses. *Fungal Diversity* **90**, 135–159.

THOMARAT, F., VIVARÈS, C.P. & GOUY, M. (2004) Phylogenetic analysis of the complete genome sequence of Encephalitozoon cuniculi supports the fungal origin of microsporidia and reveals a high frequency of fast-evolving genes. *Journal of Molecular Evolution* **59**, 780–791.

TORRUELLA, G., GRAU-BOVÉ, X., MOREIRA, D., KARPOV, S.A., BURNS, J.A., SEBÉ-PEDRÓS, A., VÖLCKER, E. & LÓPEZ-GARCÍA, P. (2018) Global transcriptome analysis of the aphelid Paraphelidium tribonemae supports the phagotrophic origin of fungi. *Communications Biology* **1**, 1–11.

TORRUELLA, G., DE MENDOZA, A., GRAU-BOVÉ, X., ANTÓ, M., CHAPLIN, M.A., DEL CAMPO, J., EME, L., PÉREZ-CORDÓN, G., WHIPPS, C.M., NICHOLS, K.M., PALEY, R., ROGER, A.J., SITJÀ-BOBADILLA, A., DONACHIE, S. & RUIZ-TRILLO, I. (2015) Phylogenomics Reveals Convergent Evolution of Lifestyles in Close Relatives of Animals and Fungi. *Current Biology* **25**, 2404–2410.

UEBELMESSER, E. RUTH (1956) Über einige neue Chytridineen aus Erdboden - Olpidium, Rhizophidium, Phlyctochytrium und Rhizophlyctis. *Archiv für Mikrobiologie*.

USTINOVA, I., KRIENITZ, L. & HUSS, V.A.R. (2000) Hyaloraphidium curvatum is not a green alga but a lower fungus; Amoebidium parasiticum is not a fungus, but a member of the DRIPs. *Protist*.

VARGAS, M.M., ARONSON, J.M. & ROBERSON, R.W. (1993) The cytoplasmic organization of hyphal tip cells in the fungus Allomyces macrogynus. *Protoplasma*.

VERA-RUIZ, V.A., LAU, K.W., ROBINSON, J. & JERMIIN, L.S. (2014) Statistical tests to identify appropriate types of nucleotide sequence recoding in molecular phylogenetics **15**, 1–11.

VIDEVALL, E. (2018) Plasmodium parasites of birds have the most AT-rich genes of eukaryotes. *Microbial genomics* **4**.

VIVARÈS, C.P., GOUY, M., THOMARAT, F. & MÉTÉNIER, G. (2002) Functional and evolutionary analysis of a eukaryotic parasitic genome. *Current Opinion in Microbiology*. .

WANG, H.C., MINH, B.Q., SUSKO, E. & ROGER, A.J. (2018) Modeling Site Heterogeneity with Posterior Mean Site Frequency Profiles Accelerates Accurate Phylogenomic Estimation. *Systematic Biology*.

WANG, Q., SUZUKI, A., MARICONDA, S., PORWOLLIK, S. & HARSHEY, R.M. (2005) Sensing wetness: A new role for the bacterial flagellum. *EMBO Journal* **24**, 2034–2042.

WICKSTEAD, B. & GULL, K. (2012) Evolutionary biology of dyneins. In *Dyneins* p.

YOUSSEF, N.H., COUGER, M.B., STRUCHTEMEYER, C.G., LIGGENSTOFFER, A.S., PRADE, R.A., NAJAR, F.Z., ATIYEH, H.K., WILKINS, M.R. & ELSHAHED, M.S. (2013) The genome of the anaerobic fungus orpinomyces sp. strain c1a reveals the unique evolutionary history of a remarkable plant biomass degrader. *Applied and Environmental Microbiology*.

7. Discussion

“Examining this water...I found floating therein divers earthy particles, and some green streaks, spirally wound serpent-wise...and I judge that some of these little creatures were above a thousand times smaller than the smallest ones I have ever yet seen, upon the rind of cheese, in wheaten flour, mould, and the like.” (The first recorded observation of protozoa)

Antonie van Leeuwenhoek. Letter to the Royal Society, London (Sept. 7, 1674).

7. Discussion

7.1. Nucleariids, more than just *Nuclearia* and fonticulids

Nucleariids are a lineage of organisms that have a long taxonomical history suffering from several re-classifications (see chapter 1.7.1. Nucleariids). When the first molecular analyses were done including *Nuclearia* spp. and *Fonticula alba* sequences, the placement of nucleariids as holomycotans became clear (Brown *et al.*, 2009; Liu *et al.*, 2009). Since then, subsequent studies have found a wide diversity of sequences from several environments branching within the nucleariid radiation (Zettler *et al.*, 2002; Lara *et al.*, 2010; Simon *et al.*, 2015; Arroyo *et al.*, 2018; Heger *et al.*, 2018; López-Escardó *et al.*, 2018; Rodríguez-Martínez *et al.*, 2020). However, since molecular data for reference nucleariid species remained limited to *Nuclearia* and *Fonticula* (and, more recently, *Parvularia*), all new sequences have been provisionally classified within those unique two clades (Del Campo & Ruiz-Trillo, 2013; del Campo *et al.*, 2015).

Our results (see chapter 4) have clearly shown that this diversity has indeed a more complex phylogenetic structure than previously thought. Thanks to new reference sequences from identified species, our updated 18S rRNA gene tree (Figure 1 from chapter 4) allowed to distinguish several clades far from *Nuclearia*, *Parvularia* and *Fonticula* and close to *Pompholyxophrys* and *Lithocolla*. What is also worth highlighting is the large number of environmental clades that remain with no known representatives and that represent still undescribed diversity.

These findings imply that many of the previously mentioned studies grouped organisms which are highly diverse both morphologically and in life traits (e.g. food source, cell cover, aggregation, size). Nevertheless, the weak phylogenetic signal of the 18S rRNA gene did not allow us to reconstruct phylogenetic relationships within the clade. To increase the signal, we reconstructed a phylogenomic tree that integrated our new genomic data from three *Nuclearia* species, *Lithocolla globosa* and two *Pompholyxophrys* species and the already available information for *Parvularia atlantis*, two fonticulids, and two *Nuclearia* species (Figure 3 from chapter 4). Our new tree supported the existence of two distinguishable clades, one including *Fonticula* and another including the *Nuclearia* species. However, the clade including the aggregative fonticulids also encompassed the small *Parvularia*, without any known aggregative behaviour. At the same time the clade including the naked *Nuclearia* included the cover-bearing *Lithocolla* and

Pompholyxophrys. Thus, in these two cases key traits of each group (aggregation and naked cells) were discarded as diagnostic characters after we explored part of the unknown diversity of nucleariids.

Nucleariids are highly derived and atypical given their phylogenetic placement within Holomycota. Being free-living and without flagellated stages, they lack many "holomycotan defining traits". Our updated phylogeny of nucleariids corroborated the fact that nested within the "fonticulida" and the "nucleariida" there are diverse groups of organisms that have adapted very differently to their environment, showing the need for future studies of each nucleariid group. This effort will be necessary also to reconstruct the traits of the nucleariid ancestor and to understand how a so unique clade evolved.

7.2. The Microsporidia + Rozellida boundary

Microsporidia and Rozellida are two clades of endobiotic parasitic organisms that recently gave rise to an interesting debate on the taxonomic boundary between the two lineages, a debate that I consider worth getting into (e.g. Richards, Leonard, & Wideman, 2017; Bass *et al.*, 2018). Recent phylogenetic analyses have suggested that Rozellida is actually a paraphyletic group containing the core Microsporidia clade (Haag *et al.*, 2014; Corsaro *et al.*, 2016; Mikhailov *et al.*, 2016; Quandt *et al.*, 2017). The improvement of taxon sampling for these organisms within a phylogenomic framework showed that there was a continuum between *Rozella*-like organisms and the core Microsporidia-like organism (see chapter 1.7.2.2. Microsporidia).

In our phylogenomic analyses we also recovered the same branching pattern, with Rozellida as a paraphyletic group, suggesting that within this continuum different organisms with intermediate traits are classified within the same group. These "intermediate" organisms have provisionally been associated taxonomically with Rozellida (*Paramicrosporidium, Nucleophaga*) or Microsporidia (*Mitosporidium*, *Chytridiopsida,* Metchnikovellidae). This taxonomical uncertainty has led to some authors to propose an expanded definition for the Microsporidia, which would englobe all Microsporidia and Rozellida lineages (Bass *et al.*, 2018). Nevertheless, before getting into a discussion about the taxonomic boundaries of the clade we first need to discuss the main defining traits of these lineages in a comparative framework (Figure 20).

The core Microsporidia contains all the species that traditionally have been classified as Microsporidia and which share the defining characters for the group, including a coiled polar tube, a polaroplast, and the absence of a mitochondrial genome (see chapter 1.7.2.2. Microsporidia). Core Microsporidia also present as a defining characteristic shared with *Rozella* the presence of an ATP transporter (nucleoside NTT transporters) probably acquired from bacteria. Nevertheless, these horizontally acquired NTT transporters of *R. allomycis* and core Microsporidia, have not been found in other related organisms, including *P. saccamoebae*, *Mitosporidium* and metchnikovellids (Figure 20). *Mitosporidium* and *Paramicrosporidium* bear functional mitochondria (Figure 20), thus they are able to produce their own ATP, but metchnikovellids do not and there is no known metabolic pathway for metchnikovellids to obtain ATP. Both the study of the *Amphiamplys* sp. genome and our own study on *Metchnikovella incurvata* tried to find a candidate gene to act as a possible transporter of ATP into the metchnikovellid cell (see chapter 5) (Mikhailov *et al.*, 2016). A possible candidate was found in a gene related with the mitochondrial carrier protein family (MCF), originally a transporter of inorganic phosphate that might have evolved to serve as an ATP pump. The MCF gene was found in all representatives of the linage except in core Microsporidia. However, this hypothesis of MCF genes needed to be tested with functional studies. Recently, a study carried by Major and collaborators (Major *et al.*, 2019) provided new insight into this question by characterizing a family of microsporidian Major Facilitator Superfamily (MFS) transport proteins, which was identified in all known core Microsporidia with available genomes sequenced in previous studies (e.g. Cuomo *et al.*, 2012; Heinz *et al.*, 2012, 2014). Representatives from this family were found also in all rozellids and Microsporidia-like species, including *Amphiamblys* sp. and *Metchnikovella incurvata*. The authors then characterized these proteins by RNA-seq transcript abundance and immunofluorescence analyses in the core Microsporidia species *Trachipleistora hominis*. They showed that the MFS proteins are ATP and GTP transporters which locate on the surface of the microsporidian cells during their growth and replication. Thus, the MFS protein family is most probably in charge of ATP absorption from the host also in metchnikovellid cells.

If we compare metchnikovellids with rozellids and core Microsporidia, the genomes of *M. incurvata* and *Amphiamblys* sp. (Mikhailov *et al.*, 2016) show that metchnikovellids gene content and metabolic repertoire resemble more to core Microsporidia. These analyses based in GO terms clearly trace a line between Metchnikovellida + core Microsporidia, and the rest of related lineages

(see chapter 5). Thus, in this discussion we will refer to metchnikovellids as an early-diverging lineage but branching within Microsporidia. On the other extreme, we have the *Rozella* species. The defining characteristics of *Rozella* include a zoosporic stage in the life cycle, the presence of a functional mitochondrion with an incomplete electron transport chain and phagotrophy (see chapter 1.7.2.1. Rozellida) (Figure 20). *Rozella* also presents an infection tube which morphologically differs from the canonical coiled polar tube of Microsporidia but fulfils the same function (Letcher & Powell, 2018).

Paramicrosporidium is another lineage classified within the Rozellida and marks the beginning of a series of organisms with intermediate traits between rozellids and Microsporidia (Figure 20; Figure 21). Some of the traits of *Paramicrosporidium* that are similar to Microsporidia include a polar tube, which is however inactive (Corsaro *et al.*, 2014b). *Paramicrosporidium* also presents an overall morphological cell shape similar to metchnikovellid Microsporidia with the presence of a horseshoe-shaped nucleus, manubrial cisternae and manubrium-like structures (Michel, 2019). At the same time, *P. saccamoebae* also resembles rozellids in the presence of mitochondria. However, *P. saccamoebae* possesses a more complete mitochondrial genome than *Rozella*, retaining the electron transport complex I. Additionally, the genome of *Rozella* is also more reduced and its sequence is extremely AT-rich (James *et al.*, 2013b). One possible explanation for the maintenance of a fully functional mitochondrion with a complete electron transport chain in *Paramicrosporidium* may be the subcellular localization of its spores within the host (Quandt *et al.*, 2017). It has been shown that *R. allomycis* and some Microsporidia species are surrounded by host mitochondria, probably to facilitate ATP absorption from the host (Hacker *et al.*, 2014; Powell *et al.*, 2017). Thus, it could be that the intranuclear localization of *P. saccamoebae* within the host cell prevents the parasite to associate with the host mitochondria to take up ATP. However, there are core Microsporidia that are also intranuclear parasites but have no mitochondrial ATP production (e.g. *Nucleospora*, *Enterospora*) (Stentiford *et al.*, 2007; Freeman, Kasper, & Kristmundsson, 2013).

The comparison can be extended to *Mitosporidium daphniae. Mitosporidium* has been traditionally classified within Microsporidia due to the presence of a polaroplast, a developed polar tube, and an overall core microsporidian-like morphology and life cycle (see chapter 1.7.2.2. Microsporidia). However, based on the phylogenies published in the current and previous studies (Haag *et al.*, 2014; Torruella *et al.*, 2018), *M. daphniae* does not reside on the long-branching group of core

Microsporidia. In agreement with that, there are several cellular (e.g. presence of a functional mitochondrion) and metabolic (e.g. ability to produce some amino acids) characteristics that suggest that this species is biologically more similar to rozellids than to the core Microsporidia. Also, similarly to *Rozella*, *Mitosporidium* mitochondrial genome is incomplete, lacking the respiratory chain complex 1. The investigation of this wide taxon sampling indicates that many major synapomorphies of the lineage suffer from a patchy distribution in several representatives (e.g. complex 1 of the electron transport chain, NTT transporters, polar tube, polaroplast) leading to the situation in which a given intermediate clade may present a similar number of traits that can be considered either rozellid- or Microsporidia-like.

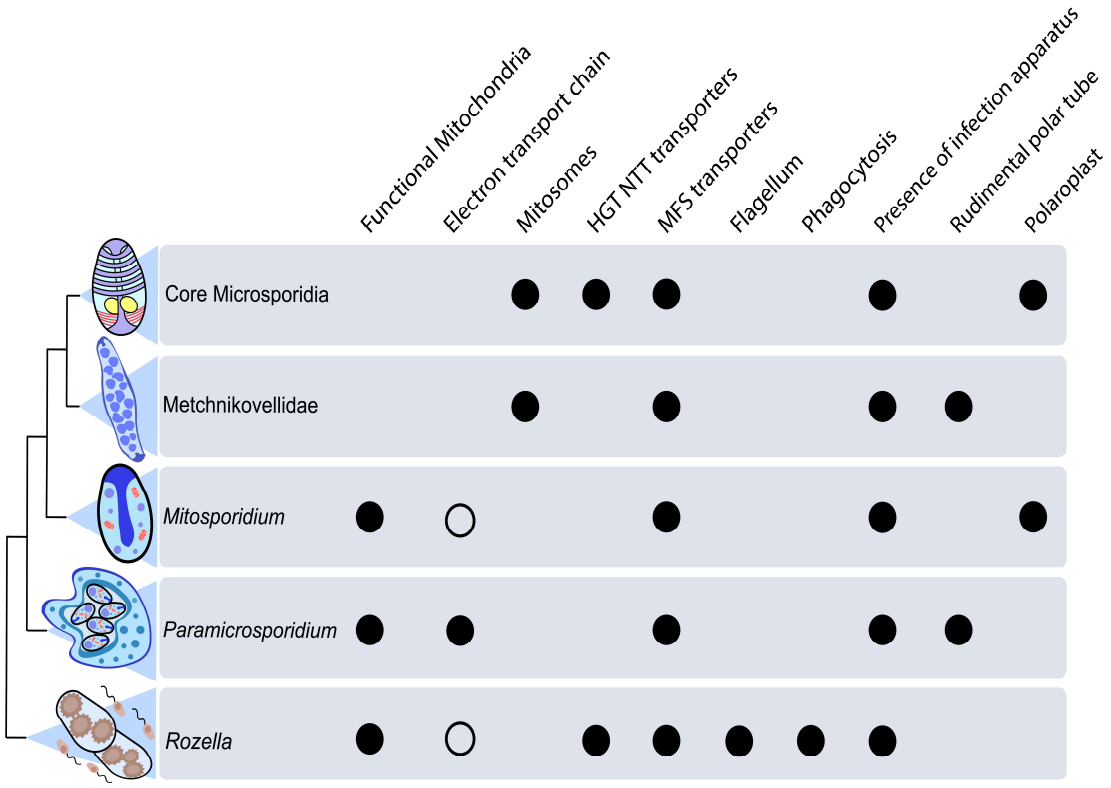


Figure 20. Summary of Rozellida + Microsporidia phylogeny based in phylogenomic data from this study and distribution of evolutionary important characters. Black dots indicate presence of the trait and empty dots indicate presence of a partial trait.

So far in this discussion, we have only considered and evaluated two "intermediate" lineages (*Mitosporidium* and *Paramicrosporidium*), for which genome sequences are available, making possible comparative genomics analyses. However, there are other organisms forming part of this interface for which the only available molecular data comes from rRNA genes (Figure 21).

Consequently, the phylogenetic placement of these lineages has been done using only these genes, which usually have been enough to confirm their association to the Microsporidia + Rozellida clade. However, their precise relationships with other lineages within that clade remain uncertain. For some of these lineages the ultrastructure and life cycle have been characterized, providing more pieces of evidence to evaluate. Two relevant lineages to discuss in this context are *Nucleophaga* and *Chytridiopsis.*

Nucleophaga, like *Paramicrosporidium*, is a microsporidia-like rozellid with non-flagellated spores (see chapter 1.7.2.1. Rozellida). *Nucleophaga* is generally recovered in 18S rRNA gene phylogenies as the sister lineage to Microsporidia (Grossart *et al.*, 2016; Stentiford *et al.*, 2017; Bass *et al.*, 2018; Corsaro *et al.*, 2019). Several traits of this putative rozellid reflect its phylogenetic position within the Microsporidia-Rozellida continuum, since it possesses both a polar filament and an anchoring disc, having a microsporidia-like spore (Corsaro *et al.*, 2016) (Figure 21). Thus, it would be possible to argue that *Nucleophaga* belongs to the Microsporidia but, again, without further genomic characterization this hypothesis lacks the necessary support.

The Chytridiopsida (see chapter 1.7.2.1. Rozellida) branch between *Nucleophaga* and the clade of metchnikovellids + core Microsporidia (Figure 21). Chytridiopsids present similar characteristics to metchnikovellids, including an aberrant polar filament and the loss of the polaroplast. These characteristics are also present in microsporidia-like rozellids, making chytridiopsids morphologically more similar to these rozellids than to metchnikovellids (Corsaro *et al.*, 2019). Thus, even if the phylogenetic signal is weak, this branching order of *Chytridiopsis* possibly reflects well the rozellid-Microsporidia transition. Other evidence supporting and earlier branching position of chytridiopsids than metchnikovellids includes the decreasing size of the ITS region separating the 5.8S and LSU rRNA genes as we move from *Rozella* towards core Microsporidia (Figure 21).

Some other lineages within Rozellida are only known from environmental surveys using the 18S rRNA gene (see chapter 1.7.2.1. Rozellida). However, some of these environmental clades have a characterized zoosporic stage, identified by the TSA-FISH techniques together with antibodies against α-tubulin (Jones *et al.*, 2011a; Corsaro *et al.*, 2019). The existence of these flagellated groups adds evidence of the maintenance of the flagellum beyond the *Rozella* clade and supports that the current distribution of traits in the clade is patchy.

This last sentence can resume the discussion about this group, since it seems that the current patchy distribution of synapomorphies on the whole clade does not allow to address many of the relationships within the lineage. Thus, any interpretation about the relationships between Microsporidia, Microsporidia-like rozellids, and rozellids can be precipitated in the absence of genome sequence data. This includes the previously mentioned inclusion of all members of the clade within an expanded definition of Microsporidia (Bass *et al.*, 2018). What seems sure is that the members of this clade are more diverse and complex than expected, which remarks the importance of efforts to isolate and characterize their genomic and morphological diversity. This would allow a better resolution of the phylogenetic relationships in the clade to get insight into the evolution of the many patchy traits to understand if they descend from a common ancestor at the root of the clade or if they have been acquired independently in several branches (e.g. polar filament, loss of mitochondrial genome, flagellum loss).

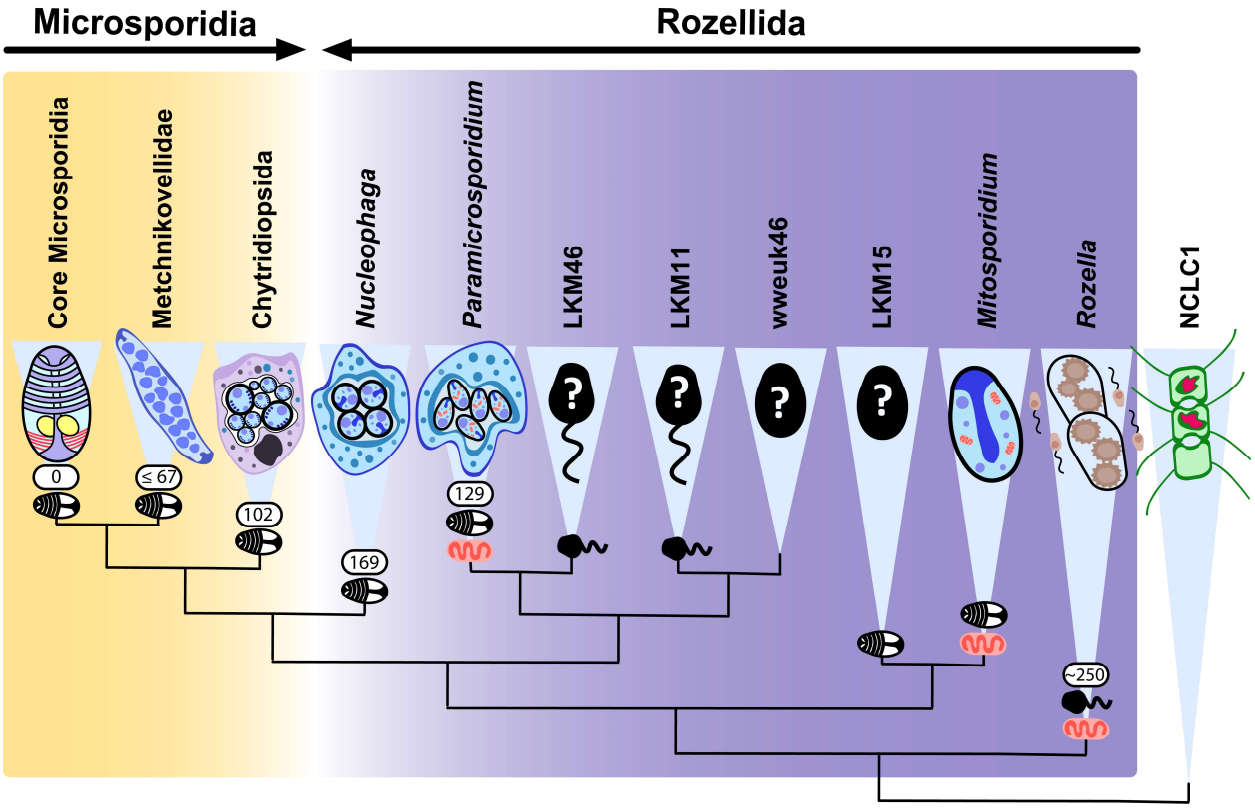


Figure 21. Cladogram based on 18S rRNA gene phylogenies of the Microsporidia + Rozellida clade, adapted mainly from Corsaro et al. but also from Chambouvet et al. (Chambouvet *et al.*, 2019; Corsaro *et al.*, 2019). Black microsporidian cells indicate lineages with microsporidia like spore. Black flagellated cells indicate lineages with zoospores. Small orange mitochondrion indicates lineages that have retained a mitochondrial genome. White circles indicate the number of nucleotides found in the ITS region between the 5.8S and 28S rRNA genes.

7.3. The new fungal tree of life and its implications for early fungi evolution

Our phylogenomic study on the fungal tree of life (see chapter 6) suggests a new panorama for fungal evolution. It has been shaped by the addition of several early-branching zoosporic fungal taxa added for the first time in a multi-gene study (e.g. *Paraphysoderma*, *Coelomomyces*) and, especially, by the addition of two clades, sanchytrids and *Olpidium*, since their phylogenetic position changed the known fungal tree by adding two new branches (Figure 22). The sanchytrids represent a new lineage sister to Blastocladiomycota. Despite the strong support for this relationship (see Figure 1; chapter 6), we argue that sanchytrids may possess enough differences to be erected as a new phylum (Sanchytriomycota), rather than forming part of the Blastocladiomycota. These differences include having totally amoeboid zoospores with a highly reduced non-motile flagellum and an extremely long kinetosome as well as the pronounced simplification of their fast-evolving genomes compared with those of Blastocladiomycota. Yet, it did not escape to our notice that both clades share some key traits that prove their relatedness beyond their phylogenetic position. One interesting example is the possession of the BeGC1 rhodopsin-guanylyl gene fusion, and the BeCNG1 cyclic nucleotide-gated channel. These proteins have evolved (as proved in *Blastocladiella emersonii*) in a cGMP-mediated light sensing cascade, colocalized in a lipid organelle at the base of the flagellum (Avelar *et al.*, 2014), which allows *B. emersonii* zoospores to sense light through an "eyespot". No assay has yet been made to assess a similar light response in sanchytrid zoospores. However, they possess every element needed to have it, including the already mentioned genes and a rosary-chain lipid organelle (potential eyespot). This suggests that the ancestor of both sanchytrids and Blastocladiomycota already presented the machinery necessary for light sensing. The deepest-branching lineage of Blastocladiomycota, *Paraphysoderma*, also shares traits with sanchytrids, including being parasites of algae, having monocentric sporangia and producing spores with no functional flagellum (*Paraphysoderma* produces both flagellated and non-flagellated zoospores) (see chapter 1.7.3.2. Blastocladiomycota).

The second branch that our results support as a new one in the fungal tree of life is that of *Olpidium*. *Olpidium* is a genus of zoosporic fungi with a controversial placement since the first molecular phylogenies placed them within the major clade of non-flagellated fungi (see chapter 1.7.4.1.

Olpidium). In our phylogenomic tree and subsequent analyses (Figure 1 from chapter 6), we recovered *Olpidium bornovanus* as a new independent branch of Fungi, sister to the non-flagellated clades Zoopagomycota + Mucoromycota + Dikarya. Previously, despite its supposed placement and due to its unique characteristics, *Olpidium* had been already considered by some authors as an independent phylum, Olpidiomycota (Tedersoo *et al.*, 2018). Nevertheless, our study is the first that provides phylogenomic support for this hypothesis. Again, the sequencing of genomes of new species of Olpidiomycota will be important to confirm our results.

Chytrids have also played a central role in our study, since resolving their branching order with the Sanchytriomycota + Blastocladiomycota clade was essential to obtain a stable backbone for the Fungi. Several authors have addressed their placement before but failed to recover high statistical support neither for Blastocladiomycota nor chytrids as the sister lineage to all other fungi (see chapter 1.7.3.3. Phylogenetic relationship of Chytridiomycota and Blastocladiomycota). To resolve these relationships, different authors have suggested two obvious approaches: the sequencing of genomes of more early-divergent taxa and the improvement of the phylogenetic analyses (Chang *et al.*, 2015; Spatafora *et al.*, 2017). In our study, we followed the first suggestion to obtain a wide taxon sampling of Blastocladiomycota and chytrids. This taxon sampling includes deep-branching Blastocladiomycota, like *Paraphysoderma*, and the sanchytrid clade. By doing this we hoped to add the necessary phylogenetic signal to clarify which clade is the sister lineage to all other known Fungi.

Even if our results will need further assessment, our analyses clearly point towards the position of Chytridiomycota as the sister lineage to all other fungi (Figure 1 and Supplementary Figure 1; chapter 6). Bayesian posterior probabilities using different datasets were maximal, but the ML ultrafast bootstrap supports remained relatively low (around 80%). We decided to address this ambiguity following several approaches (topology tests, removing fast-evolving sites, and character recoding) and all evidence still pointed towards chytrids being the sister clade of all other fungi. We can also look for other evidence supporting this placement of Chytridiomycota emerging earlier than Blastocladiomycota + Sanchytriomycota. For example, it has been shown that most genes underlying the fungal multicellular hyphae seem to have evolved in unicellular early fungal precursors (Kiss *et al.*, 2019). Blastocladiomycota have features found in multicellular fungi but not in chytrids, which is consistent with a closer relationship with multicellular fungi than with chytrids. Some Blastocladiomycota even exhibit a hyphae-like organization with a Spitzenkörper

(e.g. *Allomyces*) which is similar to that of the hyphae of non-flagellated Fungi. Thus, its presence in Blastocladiomycota suggests that they are closer to the transition towards multicellularity than chytrids. By contrast, in chytrids the only clade that possesses hyphae-like structures are the Monoblepharidomycota, although it lacks a Spitzenköper, which indicates an independent origin (see chapters 1.7.3.1 Chytridiomycota and 1.7.3.2. Blastocladiomycota)

Blastocladiomycota zoospores are also unique, presenting a nuclear cap that is the origin of bipolar growth in some species with hyphae-like development (see chapter 1.7.3.2. Blastocladiomycota). No similar structure related with bipolar growth is present in chytrids. Also similarly to many fungi and differently to chytrids, the poles of the nuclear envelope remain closed during mitosis (Olson, 1984). Lastly, like fungi, most Blastocladiomycota (e.g. *Allomyces*, *Blastocladiella*) have been reported to produce equivalents of a Golgi apparatus with an unstacked single cisterna instead of the stacked cisternae in dictyosomes present in chytrids (Bracker, 1967; Barstow & Lovett, 1974; Feeney & Triemer, 1979; Sewall, Roberson, & Pommerville, 1989). In *Olpidium*, the presence of a single dictyosome (similar to Golgi equivalents) has been reported in zoospores (Lene Lange & Olson, 1978)

Altogether this evidence points towards a "chytrid first" origin of Fungi. Nevertheless, we have always to keep in mind the possibility of an independent origin of some of these traits. For example, the Spitzenköper of Blastocladiomycota has structural differences with that of multicellular fungi (Vargas *et al.*, 1993). Neither *Olpidium* nor sanchytrids or *Paraphysoderma* (which is the sister genera to all other Blastocladiomycota) present Spitzenköper, suggesting that it may have an independent origin in the other Blastocladiomycota. The same is true for the stacked Golgi cisternae. The chytrid-like Golgi apparatus with stacked cisternae is again present in the deep-branching blastocladiomycetes *Coelomomyces*, *Physoderma* and *Paraphysoderma* (Lange & Olson, 1980; Lucarotti & Federici, 1984; Powell, 2017b) and in the sanchytrid species (Karpov *et al.*, 2018, 2019). This also supports a possible independent origin in the other Blastocladiomycota. A Golgi equivalent in Olpidiomycota could make sense given their current phylogenetic placement, although up to now there has not been a clear consensus on the nature of *Olpidium*'s Golgi apparatus.

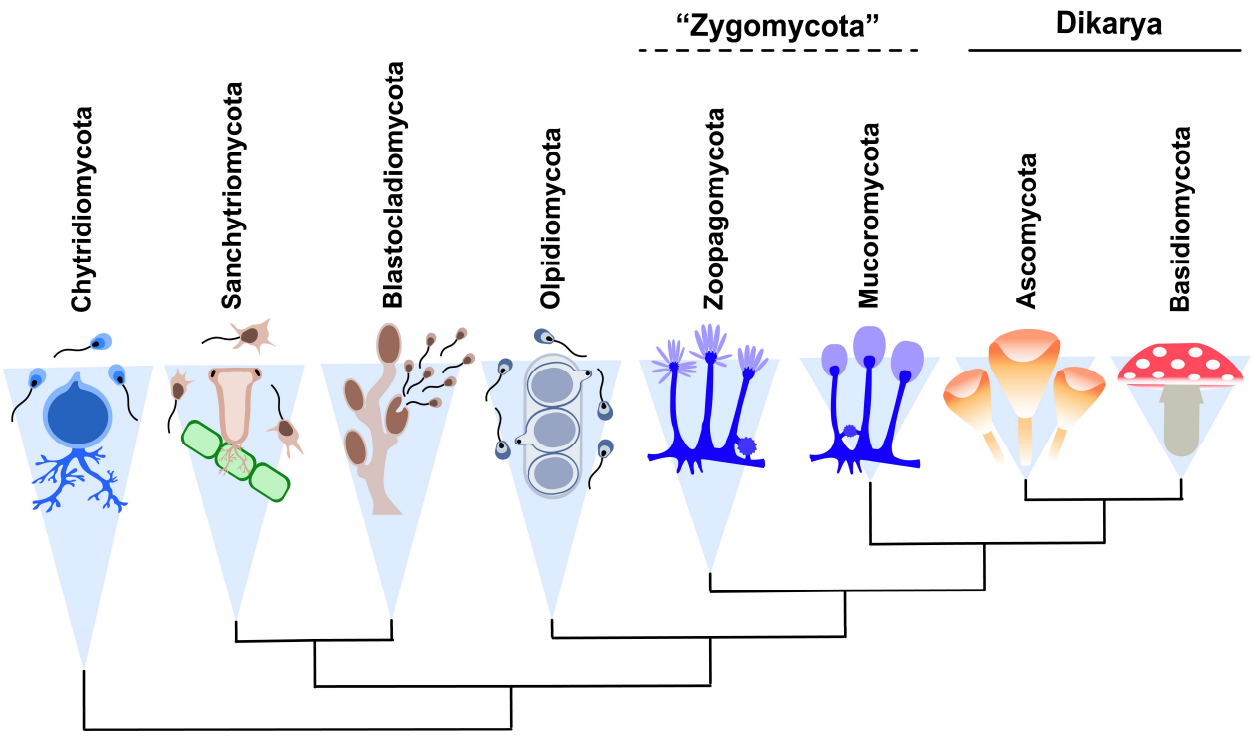


Figure 22. Cladogram of the Fungi showing the relationships between the main lineages based on phylogenomic results from this study (Chapter 6).

Outside of the scope of our study there are other holomycotan lineages with a placement that needs further assessment, or which remain without taxonomic placement at all. A good example are aphelids, a lineage of zoosporic holomycotans previously classified within Opisthosporidia (see chapter 1.7.2.3. Aphelida). Opisthosporidia was a clade composed of endobiotic parasites and initially supported by several phylogenetic studies. Besides aphelids, it also included rozellids and Microsporidia. Recently, Opisthosporidia was recovered as paraphyletic after the sequencing and study in a phylogenetic framework of the near-complete transcriptome of the aphelid *Paraphelidium tribonemae* (Torruella *et al.*, 2018). However, several posterior 18S rRNA-based diversity studies kept recovering aphelids as part of Opisthosporidia (e.g. Bass *et al.*, 2018; Chambouvet *et al.*, 2019). Although these studies possess a larger taxon sampling than phylogenomic studies, they only retrieve a weak support for the monophyly of Opisthosporidia. Our study using a multi-marker approach has recovered again the position of *P. tribonemae* as the sister lineage to all Fungi, with maximum support values in all cases (Figure 1 and Supplementary Figure 1; chapter 6). Thus, we support the hypothesis of Torruella and collaborators that Opisthosporidia are paraphyletic with aphelids as sister clade to Fungi.

Glomeromycota can be considered a complex multicellular clade that also escaped the scope of this study. Nevertheless, its position within Fungi is of phylogenetic interest since it remains unresolved (see chapter 1.2.3.6. Mucoromycota). Some early studies recovered them as an independent phylum (Glomeromycota) sister to Dikarya (Schüßler *et al.*, 2001; Lutzoni *et al.*, 2004; White *et al.*, 2006). However, subsequent studies started proving their potential affinity with Mucoromycota (Nadimi *et al.*, 2012; Lin *et al.*, 2014; Chang *et al.*, 2015; Spatafora *et al.*, 2016b) within the class Glomeromycotina. Tedersoo and collaborators (Tedersoo *et al.*, 2018) made trees of different proteins and recovered both alternatives but suggested that this lineage has enough differential traits from Mucoromycota to form its own independent phylum. In our study, we included data for only two Glomeromycota species and recovered trees supporting both alternative hypotheses. With our dataset of 69 species, we recovered their sister relationship with Dikarya in Bayesian analyses and branching within Mucoromycota in ML reconstruction. In the case of our large dataset of 84 species, we always recovered their association within Mucoromycota but with low support values. Thus, their position remains unresolved. This clade has a particular lifestyle (mainly plant mycorrhizal symbionts) that probably has influenced its genome evolution in ways that can induce phylogenetic artifacts (e.g., compositional bias). Thus, an effort needs to be made to sample additional genomic data from this clade that might add the necessary phylogenetic signal to resolve its position, in addition to keep working on improving the available phylogenetic reconstruction methods.

Overall, our study has provided a new phylogenetic backbone for fungi. This backbone will need further work to clarify the position of several lineages (e.g. aphelids, chytrids, Glomeromycota) but it will serve as a starting point for future research on Holomycota.

7.4. The most recent common ancestor of Holomycota and lineages within

The nature of the most recent common ancestor (MRCA) of Holomycota and Fungi has been addressed in the past (James *et al.*, 2006a; Liu, Hodson, & Hall, 2006b; Torruella *et al.*, 2015, 2018; Naranjo-Ortiz & Gabaldón, 2019b). However, under our new phylogenetic framework, we wanted to revisit this question, not only for the MRCA of Holomycota but also those of opisthosporidians, sanchytrids plus Blastocladiomycota, and the non-flagellated fungi (Figure 23). The MRCA of Holomycota was most likely a phagotrophic free-living amoeboflagellated

unicellular organism. Whether it evolved in marine or freshwater environments remains unknown. Its unicellular nature is most likely given that the deepest branch sister to all opisthokonts is that of Apusomonadida (Torruella *et al.*, 2012, 2015), a unicellular lineage, and that all known deep-branching representatives of Holozoa and Holomycota are unicellular (see chapter 1.3. Opisthokonta). This ancestral organism would probably have also presented filopodia related with its phagotrophic capabilities.

The current lineages related to Holomycota, including Holozoa and Apusomonadida, are mostly composed of organisms that at a given stage of their life cycle possess a flagellum. In the case of Opisthokonta, they possess only one posterior flagellum instead of two. Punctual independent flagellum losses have occurred in several opisthokontan lineages. In Holozoa, at least two losses occurred in Icthyophonida (Ichthyosporea) and *Capsaspora* (Torruella *et al.*, 2015). In Holomycota, the scenario is more complicated, with estimates ranging from 4 to 6 independent flagellum losses (see chapter 1.9. Early fungal evolution and the holomycotan flagellum). These estimates have varied mainly due to the position of *Olpidium* within the non-flagellated Fungi. In our new phylogenetic framework, *Olpidium* branches outside the non-flagellated fungi, as an independent new lineage sister to them. Thus, we infer 4 independent flagellum losses: Olpidiomycota, nucleariids, Microsporidia, and the chytrid *Hyaloraphidium* (see chapter 6). A possible fifth loss event would be the one of nephridiophagids, which have been placed within fungi but with no clear affinity with any known fungal branch (see chapter 1.7.4.3. Other *incertae sedis* lineages).

The MRCA of the paraphyletic Opisthosporidia would also have been a phagotrophic amoeboflagellated unicellular organism, but with the important difference that its life cycle included a parasitic (or intracellular predatory) endobiotic stage (Torruella *et al.*, 2018). All member of Opisthosporidia are by definition endobiotic and parasitic (Karpov *et al.*, 2014), then both traits would have been present in its ancestor. Even if it was secondarily lost in the highly reduced Microsporidia, phagotrophy has been described in rozellids and aphelids (see chapter 1.7.2.1. Rozellida).

The MRCA of Fungi was most likely already an osmotrophic organism since all known fungal diversity is osmotrophic. It would have possessed a unicellular organization (although we do not know if completely unicellular or if unicellular coenocytic) given the fact that all known opisthosporidians and deep-branching zoosporic fungi are unicellular. Its life cycle would have

presented both an amoeboflagellated free-living stage and a parasitic stage, just like all known aphelids and zoosporic fungi (Torruella *et al.*, 2018). This parasitic stage in the MCRA of Fungi would most probably have involved a monocentric thallus (one sporangium). Other characters of the MRCA of Fungi remain uncertain since they are highly variable in contemporary fungi. For example, its environment (freshwater or marine), the nature of its cellular organization (purely unicellular or coenocytic), and its endobiotic or epibiotic nature. These uncertainties are addressed at the end of this chapter.

For similar reasons as for the MRCA of Fungi, the MRCA of Sanchytriomycota and Blastocladiomycota was probably an osmotrophic amoeboflagellated unicellular organism with a coenocytic organization. Its life cycle was similar to that of chytrids and would have included a parasitic epibiotic stage that had most likely a monocentric (and eucarpic) thallus. This MRCA developed its life cycle in freshwater environments given the fact that all current known diversity of sanchytrids and Blastocladiomycota has exclusively been found in freshwater (see chapters 1.7.3.2. Blastocladiomycota). Additionally, all deep-branching lineages in this clade are parasites of algae (the two known sanchytrid species and *Paraphysoderma*). Then it would be logical to consider that their ancestor would likely predate also on algae (see chapters 1.7.3.2. Blastocladiomycota and 1.7.4.2. Sanchytrids). Within Blastocladiomycota, *Paraphysoderma* and *Coelomomyces* have a stacked Golgi organization, whereas *Blastocladiella* and *Allomyces* present Golgi equivalents similar to those present in non-flagellated Fungi. The same is true for the presence of the Spitzenkörper. Thus, the MRCA of sanchytrids and Blastocladiomycota most probably would have lacked Spitzenkörper and presented a stacked Golgi apparatus.

The MRCA of non-flagellated Fungi would have been an osmotrophic non-flagellated unicellular coenocytic organism. Its living mode may have been either parasitic or already saprobiotic. Even if the environmental origin of this ancestor is uncertain, we have hypothesized a possible freshwater origin (see next chapter 7.5). Additionally, this ancestor may have presented polycentric (multiple sporangia) development, and a Spitzenkörper and Golgi equivalents. The deep-branching fungi Zoopagomycota, Olpidiomycota and most zoosporic fungi are coenocytic, and thus the MRCA of non-flagellated fungi would also likely have had this trait. A Spizenkörper has not been observed in *Olpidium* species, but a similar structure to a Golgi equivalent has been observed in its zoospores (see previous chapter 7.3), implying its presence in the ancestor of both Olpidiomycota and non-flagellated Fungi.

The phylogenetic placement of Olpidiomycota implies not only the inclusion of a new branch within the Fungi but also that there are not known flagellated fungal species after this branching point, suggesting that the common ancestor of all non-flagellated fungi already had lost its flagellum. What is more, in our flagellar toolkit analyses, *Olpidium bornovanus* already seems to lack some elements of the canonical toolkit (e.g. gamma and epsilon tubulins, axonemal dyneins) (see chapter 6). Thus, the transition towards the total loss of the flagellum could have already started in the ancestor of both Olpidiomycota and non-flagellated Fungi.

When observing the overall composition of traits present in the different holomycotan lineages (Figure 23) we can detect some patterns. One of them is how the transition from unicellularity, coenocytic unicellularity and towards complex multicellularity follows the branching of the different fungal lineages. This process has three steps. The first one concerns nucleariid and opisthosporidians, all unicellular organisms, the second one includes unicellular coenocytic organisms, which are all zoosporic fungi and Zoopagomycota, and the last step is that including the two complex multicellular lineages, Mucoromycota and Dikarya. These results show the presence of a transition between unicellular and complex multicellularity (CM) in Fungi in the form of an intermediate state, which is the coenocytic development of cells. However, we must take into account a fourth stage of fungal development, simple multicellularity (SM) (see chapter 1.8. Multicellularity: Fungi vs Protist). CM fungi show a patchy distribution in the fungal tree, found in some lineages of Mucoromycota and in most dikarya. Thus, there have been probably several transitions from coenocytic forms towards SM and then to CM. Some estimates suggest 8-11 independent origins of CM (Nagy *et al.*, 2018). Thus, the three-step transition from unicellularity towards multicellularity is most likely an oversimplification of what we really observe in nature.

Another interesting pattern in Holomycota is that being unicellular tends to be correlated with being phagotrophic, with some punctual exceptions derived from morphological simplification (e.g. Microsporidia) or from ancestral osmotrophy (e.g. *Hyaloraphidium* and yeasts). In the same way, osmotrophy appears to be found mostly in unicellular coenocytic and multicellular species. Finally, another pattern can be found in parasitic holomycotans. In this case, unicellular species appear to be endobiotic whereas those that have a unicellular coenocytic or multicellular development are capable of being epibiotic.

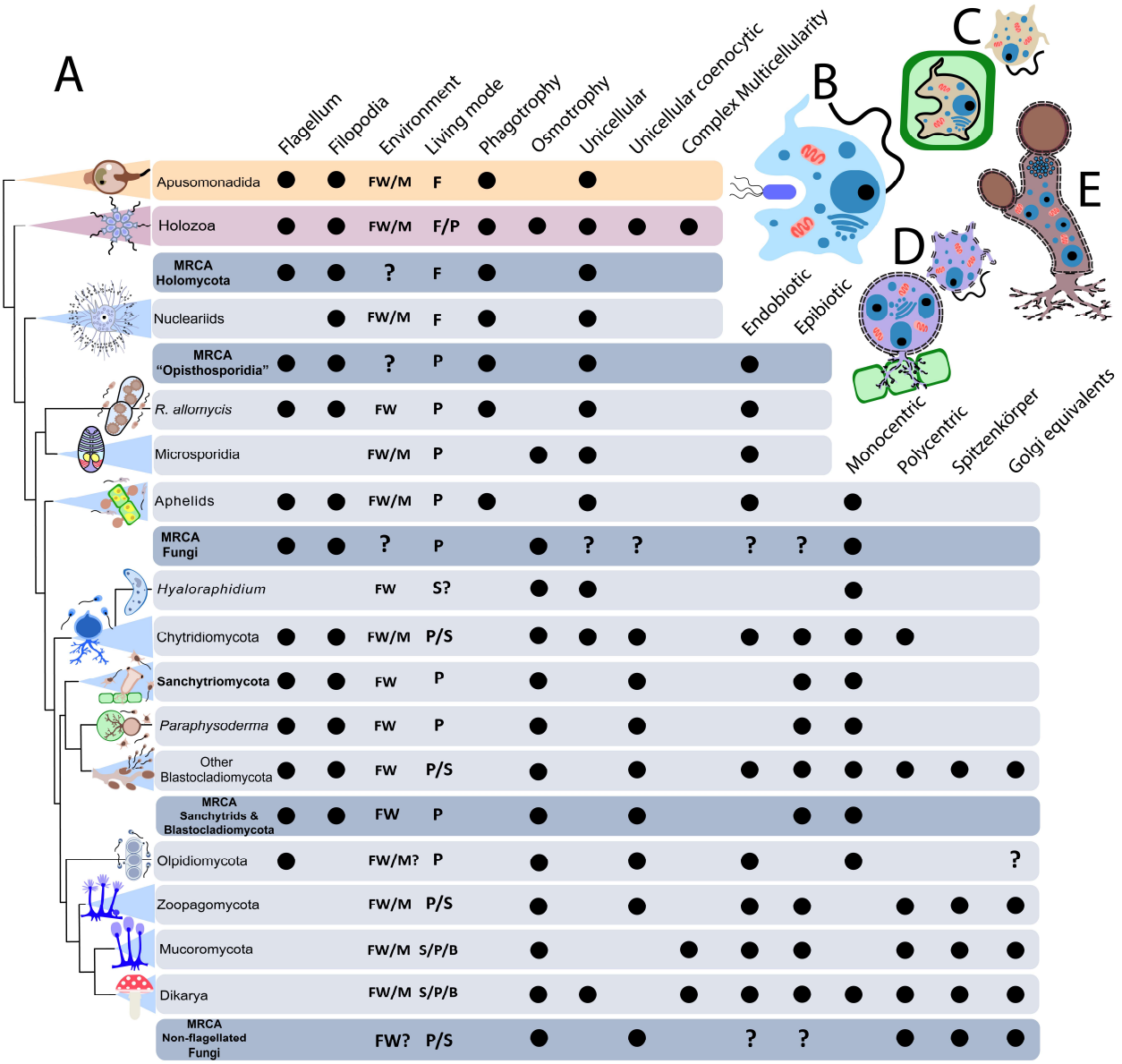


Figure 23. A) Summary of the Holomycota phylogeny based in phylogenomic data from this study and evolutionary important characters. The presence of a black dot indicates presence of the trait in representatives of the clade. An interrogation sign indicates uncertainty over the character. FW and M means fresh water and marine; F: free living; P: parasite/predator; S: saprotroph; B: symbiont. B-F Illustrations depicting in a simplified manner the overall possible appearance of the different MRCA based on their expected characters. B) MRCA of Holomycota. C) MRCA of Opisthosporidia. D) MRCA of both of Fungi and sanchytrids and Blastocladiomycota. E) MRCA of non-flagellated Fungi.

These patterns can help us when trying to infer some unknown traits of the ancestors of several groups within Holomycota. For example, on the debate on whether the MRCA of Fungi was

unicellular or unicellular and coenocytic, given the fact that most coenocytic fungi are osmotrophic, the MRCA of fungi would have been coenocytic. The same its true for its endobiotic or epibiotic nature: since most osmotrophic organisms seem to be primarily epibiotic, the MRCA of all fungi and the MRCA of non-flagellated fungi would have been epibiotic. This implies that unicellular (e.g. *Hyaloraphidium*, yeasts) and endobiotic (e.g. *Olpidium*) traits would have been acquired secondarily. However, I consider that even if these patterns can be helpful, the overall simplification of the actual natural diversity potentially hidden behind these generalizations can lead to errors when trying to infer traits present in these ancient ancestral organisms.

7.5. A possible freshwater origin of sanchytrids and Blastocladiomycota and its implications for the evolution of Fungi

All known species of Blastocladiomycota and sanchytrids are found exclusively in freshwater environments (James *et al.*, 2014; Berbee *et al.*, 2017; Powell, 2017b). This could just reflect a sampling artefact, but no molecular environmental survey has yet found evidence of these lineages in marine environments, so that their restricted association with freshwater habitats appears to be robust. I have developed two hypotheses to explain this distribution. In both cases, the chytrid lineage evolved and ended up conquering both marine and freshwater environments. In the first hypothesis, the MRCA of non-chytrid Fungi diverged from a lineage that found its niche in freshwater settings and diversified there (Figure 24A). In this scenario, the lineage leading towards Olpidiomycota and non-flagellated fungi would have secondarily conquered marine environments. Naturally, in this case we cannot discard non-observed ancient secondary transitions to marine settings within the radiation of non-chytrid Fungi. The second hypothesis is that the lineage evolving towards sanchytrids and Blastocladiomycota diversified both in marine and freshwater environments, just like chytrids (Figure 24B). However, due to an unknown cause (e.g. a competitive pressure coming from chytrids already occupying all available niches), that lineage could only survive in freshwater and marine lineages became extinct. In this scenario, Olpidiomycota and non-flagellated fungi would be part of a lineage that conquered and diversified both in marine and freshwater environments (maybe independently) during their evolutionary history. All these are theoretical possibilities that will need further testing (e.g. time trees, fossil record studies, reconstruction of plant-fungi transition) to corroborate.

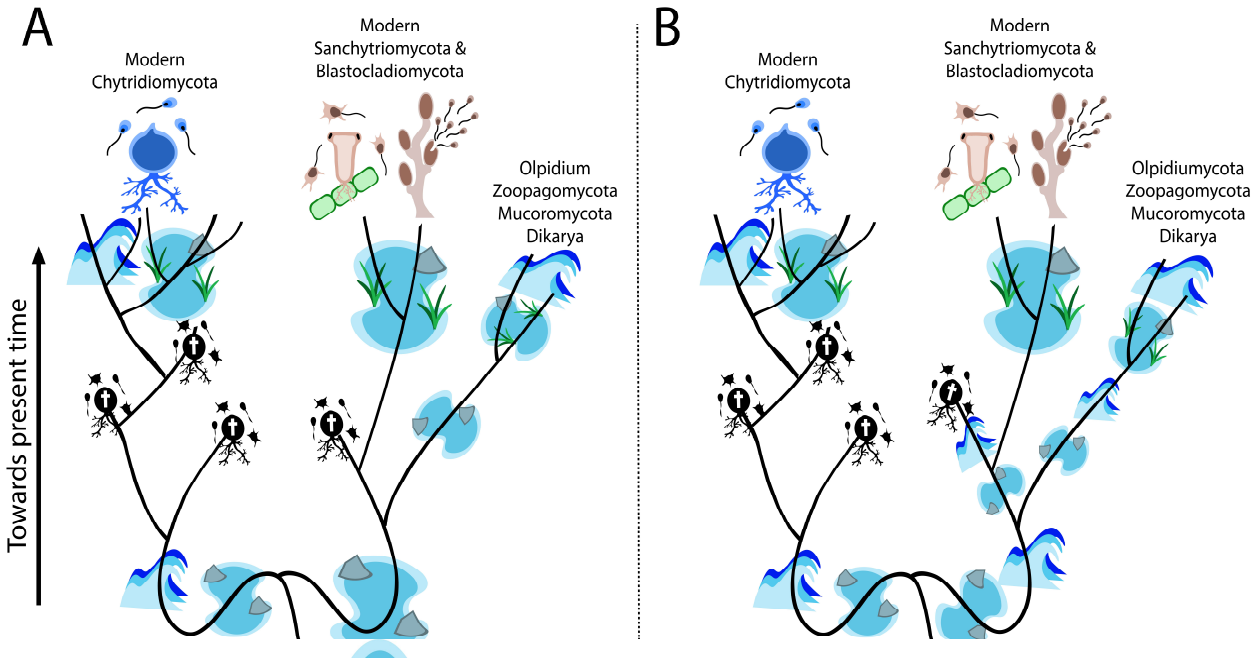


Figure 24. Illustrations of the two possible hypotheses for the transition of modern non-chytrid Fungi to marine and freshwater environments. Black chytrid shadows indicate evolutionary dead-ends; lakes and waves figures indicate freshwater and marine environments, respectively.

Following up on a possible freshwater origin of non-chytrid fungi (Figure 24A), it would be important to clarify the presence of Olpidiomycota in marine habitats. Around 50 species of *Olpidium* have been described (Sekimoto *et al.*, 2011) (see chapter 1.2.4.1 *Olpidium*). However, there are about 12 species of marine parasites of algae that have been ascribed to the genus *Olpidium* in several reports that date back to the 19th and early 20th century and the corresponding type material cannot be verified (Dick, 2001; Jones, 2012). Therefore, up to now this genus has not been confirmed with molecular data to be present in marine environments. If confirmed that Olpidiomycota is composed only of freshwater/soil species, a possible freshwater origin of the lineage heading towards the main group of non-flagellated Fungi would be further supported. A freshwater origin of must fungal diversity is a hypothesis that has been presented in the past since is one of the requisites of the "green" scenario (Naranjo-Ortiz & Gabaldón, 2019b) of fungal terrestrialization (Lücking *et al.*, 2009). In this scenario, Fungi co-evolved with the ancestor of land plants arriving from freshwater bodies as parasites of green algae.

It has been recognized that the earliest plant forms were unicellular freshwater algae and that land plants emerged from a lineage of freshwater charophyte algae (Wickett *et al.*, 2014; Delwiche &

Cooper, 2015). Recent, molecular clock analyses of both plants and Fungi have recovered time estimates that corroborate this co-evolution by showing that fungi probably facilitated plant terrestrially through symbioses, possibly in the form of endomycorrhizae (Lutzoni *et al.*, 2018). Thus, the diversification of non-flagellated fungi may have started in freshwater bodies.

Other hypothesis of fungal terrestrialization is the "brown" scenario in which Fungi colonized land through microbial crust-like communities existing on emerged lands (Taylor & Osborn, 1996). From this point, fungi developing on these proto-soils would have evolved into lineages related with Streptophyta and in parasitic lineages of protists. The "brown" scenario suggests the remains of cyanobacteria, algae, etc. over continental marine shores as the ground for terrestrialization. Nevertheless, as it has been shown for the origin of life, the conditions and composition of freshwater bodies can sometimes be more suitable environments than the oceans for many evolutionary transitions (Mulkidjanian *et al.*, 2012). Even for some major clades, like cyanobacteria, it has been shown that they first evolved in freshwater habitats and then they migrated to marine areas (Blank & Sánchez-Baracaldo, 2010). Additionally, the diversity of environments is larger on land than in marine settings (Hutchinson, 1961) so, theoretically, there would be more niches which would increase the speciation potential (Strother *et al.*, 2011). Thus, imagining a "brown" scenario occurring in freshwater bodies seems likely. Recently, a "white" scenario has been proposed which involves icy environments (Naranjo-Ortiz & Gabaldón, 2019b). In this scenario, zoosporic fungi would have arrived to ice environments as parasites of green algae and the conditions in these environments would have led to adaptations related with water limitation, making possible the colonization of land. The "white" scenario implies the transition of fungi through a freshwater interface that is ice. Putting all these ideas together, an origin and diversification of non-flagellated Fungi in freshwater bodies emerges as a strong hypothesis for fungal terrestrialization.

7.6. Single-cell genomics to resolve the eukaryotic tree of life

My PhD project was embedded in the frame of an international training network (SINGEK ITN) with the goal of training a new generation of scientists to apply single-cell techniques to explore the ecology and evolution of microeukaryotes, including those within the "microbial dark matter" (see chapter 1.10. Single cell genomics applied to the holomycotan dark matter). The microbial

dark matter is defined as the large fraction of microorganisms (~99%) that we are unable to culture in the laboratory (Ishoey *et al.*, 2008; Lasken & McLean, 2014; Wang & Navin, 2015). To study this huge proportion of hidden diversity, single-cell approaches appeared as one of the best solutions. Nevertheless, the application of this approach to environmental samples, and especially for the study of protists, was recent at the time that the ITN project began, in January 1st, 2016. Therefore, many improvements and optimizations were necessary to successfully apply the different single-cell approaches. In the following lines I will resume the application of single cell genomic to this project. I will discuss the advantages and disadvantages of the technique to resolve the questions that we asked about the phylogenetic relationships of organisms within Holomycota and about its capacity to help us gain insights into ecological aspects of some lineages. But first, we must understand how recent were the single-cell approaches at the time the project began.

Single-cell omics were born thanks to the mix of different technologies developed between 1990 and early 2000. They included the whole transcriptome/genome amplification techniques (WTA/WGA) developed in the 90's (Van Gelder *et al.*, 1990; Telenius *et al.*, 1992) and next generation sequencing (NGS) in 2005 (Mardis, 2011; Loman *et al.*, 2012). This technological evolution culminated in the invention of the first single-cell genome/transcriptome methods applied to mammalian cells in 2009 and 2011 (Tang *et al.*, 2009; Navin *et al.*, 2011). However, in the case of microbial cells these approaches had to be modified since in many cases an individual cell would carry only a few femtograms of DNA and RNA, making even harder to amplify them. For single-cell microbial studies, the invention of the Multiple Displacement Amplification technique (MDA) was crucial. MDA made possible to sequence the first genomes of microbial single cells of *Escherichia coli* and *Myxococcus xanthus* isolated by flow cytometry. This study confirmed that it was possible to generate micrograms of DNA from a few femtograms of starting material (Raghunathan *et al.*, 2005). These studies on few microbial cells of species cultivated in the laboratory opened the way for the first large-scale study in 2013 (3 years prior to the beginning of SINGEK). In this study the authors isolated and sequenced 201 uncultivated prokaryotic single cells from environmental samples using fluorescence activated cell sorting (FACS) (Rinke *et al.*, 2013, 2014).

The application of single-cell approaches to microbial eukaryotes was the next milestone to be achieved at that time. A milestone from which the SINGEK network emerged with the intention of optimizing the necessary steps to make protist single-cell genome and transcriptome sequencing

feasible. In the context of my research, we focused on holomycotan lineages that could not or had not been cultured yet. They were mainly lineages with a parasitic lifestyle but also free-living ones. Our research included cells isolated from a wide variety of environments, ranging from the metazoan gut (*Metchnikovella incurvata*) to freshwater and marine samples (nucleariids and sanchytrids). Cell isolation was performed in all cases by micromanipulating individual cells and storing them into individual tubes after several washing steps in clean water. We discarded the application of high throughput methods (e.g. FACS) because, in general, our target cells were in very low abundance and mixed with many other eukaryotes.

From the many challenges that arose from the transition of these methods from prokaryotic to unicellular eukaryotic cells, the main problem that has been identified by many studies concerns the cell lysis step (Krabberød *et al.*, 2017; Onsbring *et al.*, 2019) (http://www.singek.eu/chemical-hammer-needed/). Unicellular eukaryotes in many cases possess cell covers of different chemical composition, including silica (e.g. radiolarians, diatoms), chitin (e.g. unicellular fungi) or cellulose (e.g. unicellular algae). Nevertheless, even if for this study chitin could have been problematic due to the holomycotan nature of our species, lysis was never a major drawback. Standard lysis methods from commercial extraction kits were enough to access the DNA/RNA of our cells. This was probably because chitin is only present during some stages of the life cycle of the organisms that we studied. In the case of zoosporic fungi, they lack chitin during the sporic stages (Powell, 2017b), and sanchytrid DNA was extracted from zoospores or from sporangia fill of zoospores. Metchnikovellid cells were isolated from infected gregarines, thus chitin has not yet accumulated in the spores wall (Vávra & Lukeš, 2013). Nucleariids do not seem to have chitin (James & Berbee, 2012; Torruella *et al.*, 2015) but if they do, chitin would only accumulate when they form cysts (López-Escardó *et al.*, 2018), and all our micromanipulated nucleariid cells were not encysted. We could expect that the silica-based cover of some nucleariid cells (*Pompholyxophrys* and *Lithocolla*) would have make DNA/RNA extraction more difficult, but it was not the case.

Amplification of the extracted nucleic acids was then carried out for both DNA and RNA of single cells. WGA from single cells of *M. incurvata*, nucleariids and sanchytrids was always performed using the MDA REPLI-g WGA Single Cell Kit (QIAGEN). Additionally, for *M. incurvata* an amplification with a MALBAC kit was also performed. However, the MALBAC sample did not yield high DNA amount, and we failed to amplify the 18S rRNA gene of our targeted metchnikovellid using the amplified DNA as template. A lower genome coverage rate in

MALBAC compared with MDA had been observed in previous studies (Gawad, Koh, & Quake, 2016). Since MDA performed better, we used MDA for all our amplification reactions, including for WTA for which we used the REPLI-g WTA Single Cell Kit (QIAGEN).

After sequencing and assembly of our sequences, the completeness estimates of the genomes/transcriptomes, via BUSCO (Simão *et al.*, 2015) or via the percentage of recovered conserved protein in phylogenomic datasets, varied greatly depending on the sample. In nucleariids, a detailed assessment was made (see chapter 4) and the main conclusion was that, despite the relatively low percentage of recovery of conserved proteins in single-cell transcriptomes (SCT) or genomes (SCG), both approaches allowed to recover enough conserved markers to run robust phylogenomic analyses. Starting from this base, we did observed differences and, in general, SCT outperformed SCG by recovering higher percentages. This is a pattern previously observed for SCG of microbial eukaryotes which in many cases present low genome recovery rates (Gawryluk *et al.*, 2016; López-Escardó *et al.*, 2017; Mangot *et al.*, 2017). Additionally, the assembly process can be complicated due to the presence of repetitive intergenic regions (Onsbring *et al.*, 2019). It was not a surprise that culture-based approaches (whole RNA extraction) for *Nuclearia* and *Lithocolla* performed better than SCT/SCG and yielded the highest recovery rates, as it has also been observed in other studies (Kolisko *et al.*, 2014).

Even if the amount of generated sequence data was enough to run phylogenomic analyses, it was insufficient to make further comparative analyses. However, in some cases we got very high BUSCO completeness values, especially in the case of three SCG of *M. incurvata* and our two sanchytrids. These values ranged from 80% for our metchnikovellid to more than 90% for our two sanchytrid species. This allowed to study their genomes in more detail by looking into their metabolic capabilities, structure and similarity with related lineages. What made the difference in these cases was most probably that the starting material for amplification had a higher DNA quantity than a single nucleariid cell. In fact, the SCG of *M. incurvata* was amplified from DNA extracted from a single infected gregarine full of clonal spore cells. Similarly, in sanchytrids the SCGs were amplified from DNA extracted from sporangia full of also genome-wise identical zoospores. Thus, in both cases the amplification reaction had more template DNA to amplify, which implies higher coverage and recovery rates. Therefore, DNA extraction from sporangia is something to consider in future single-cell studies of microbial eukaryotes with sporic stages in their lifecycle.

Finally, concerning ecological aspects, single-cell genomics can also be an approach to consider, for example to assess close ecological interaction between microbes. Single-cell techniques have been used widely to study microbial interactions (Yoon *et al.*, 2011; Benites *et al.*, 2019; Ku & Sebé-Pedrós, 2019; Needham *et al.*, 2019). We identified in this way sequences of several probable bacterial endosymbionts in our *Pompholyxophrys* SCG data, which corresponded to bacteria from the groups Rickettsiales and Chlamydiae, both well-known obligated endosymbionts.

These results show how single-cell techniques are a useful tool to generate ecological and genomic data for organisms that are difficult or impossible to culture. If coupled with FACS or microfluidics, these approaches can be the best high throughput method to assess the microbial diversity of a given environment. Recent studies using these high throughput approaches have successfully led to obtaining gigabases of assembled genomic data from new eukaryotic lineages (e.g. diplonemids, ketablepharids, cercozoans, MAST, fungi) (Gawryluk *et al.*, 2016; Mangot *et al.*, 2017; Ahrendt *et al.*, 2018; Seeleuthner *et al.*, 2018; Sieracki *et al.*, 2019; Wideman *et al.*, 2020). However, these approaches have the drawback that they do not allow to observe the organisms prior to isolating them, whereas simple micromanipulation allows you the observation of the target organism. Observing the target cell gives important information about their morphology, behavior and possible lifestyle that can help to the interpretation of the subsequent genomic data. It is for this last reason that approaches based on the cultivation of microbial diversity need to keep being developed, not only because it is the way to get the highest quality of genomic data, but also because they allow to study these other aspects of the microorganisms (e.g. ecology, symbioses, ultrastructure, function). Taking this into account, it would not be impossible to imagine approaches coupling high throughput techniques with morphological characterization of the microbial eukaryotes. For example, using a microfluidic chip coupled with a microscope and a camera to take pictures or record videos of the organisms within the droplets. Until then, single-cell techniques together with improvements in culture techniques will be one of the best formulas to address the study of the microbial dark matter.

7.7. Perspectives

Every time that an answer is given in science, it is not uncommon that more questions appear. The results of my PhD have derived into new questions and opportunities to be addressed in future studies regarding the Holomycota.

- Sequencing of transcriptomes and genomes from key-branching holomycotan species

Something that became clear during the development of this project was that we are only scratching the surface of the real holomycotan diversity. Therefore, to unveil new major holomycotan lineages we need to keep searching both for new unknown representatives of the clade and already described putative members with no molecular/genomic data (see chapter 1.7.4. *Incertae sedis* lineages). The last option is probably the easiest one since we already know what to look for. The sequences from new key taxa will clarify the branching order of lineages in many groups (e.g. Microsporidia and Rozellida), put a face to environmental groups with not known representatives (e.g. LKM15 and some environmental groups within nucleariids), and help us to reconstruct the evolution of traits within the clades. Thus, in the following lines I will include known holomycotan species that need to be further addressed.

For nucleariids, the presence of a large hidden diversity became obvious after the construction of the 18S rRNA gene trees, in which there were many clades without known representatives. A potentially interesting candidate for the naked nucleariid amoeba clade is *Vampyrellidium* (Zopf, 1885b). *Vampyrellidium* is very similar to *Nuclearia* and, unlike other nucleariid amoebae, it feeds by penetration and uptake of the cell content of prokaryotic or eukaryotic algal cells with a specialized pseudopodium. For the covered nucleariids, it would also be interesting to generate new molecular data to corroborate if the monophyly of covered nucleriids is real or if there have been several independent origins of cell covers in nucleariids. *Elaeorhanis* (Greeff, 1873) is an interesting candidate possibly related with the *Lithocolla* clade since both accumulate silica-based exogenous materials around them. On the other hand, there are several candidates to form part of the clade of nucleariids with covers made of siliceous endogenous scales, from which *Pompholyxohrys* forms part, and include *Pinaciophora* (Greeff, 1873), *Rabdiophrys* (Rainer, 1968), *Rabdiaster* (Mikrjukov, 1999b) and *Thomseniophora* (Nicholls, 2012b).

If we now switch towards the Microsporidia + Rozellida clade, it is clear that the sequencing of genomes from new lineages would be important to add phylogenetic signal in a multi-gene phylogeny to key nodes in the tree of life. Therefore, the main organisms that we need to target are those which seem to branch at the border of the Microsporidia-Rozellida continuum. We thus need to generate genomic data from *Nucleophaga* and Chytridiopsida. Not only this would clarify the respective branching order of metchnikovellids and chytridiopsids and their relative position

within Microsporidia or Rozellida, but also a detailed inspection of their genomes would help clarifying the presence and evolution of traits like the absence of mitochondrial genome, polar tube, etc. Other potential candidates within this clade are of course the environmental lineages only known by their 18S rRNA gene sequences and some characterized by immunofluorescence assays. From these assays we know that some, like LKM11 and LKM46, are flagellated. By obtaining genomic data of LKM11 and LKM46 we could be able to unveil key aspects as the loss of the flagellum in rozellids and Microsporidia. Other environmental clades to consider are LKM15, wwwuk46, the diatom parasites NCLC1 (BCG1), the BCG2 and Namako-37.

Additionally, there are clades only described in ancient literature which might be interesting to target. *Sagittospora*, for example, is a parasite of ciliates that does not possess a known flagellated stage and resembles Microsporidia (Lubinsky, 1955). Some others include parasites of green algae and oomycetes like *Plasmophagus* and *Dictyomorpha*, which are very similar to *Rozella* (Blackwell, Letcher, & Powell, 2016, 2017).

The position of aphelids within the Holomycota tree of life needs to be clarified. So far, all evidence of aphelids as the sister lineage of Fungi is based on the transcriptome of a single species (*Paraphelidium tribonemae*). Thus, it is essential to generate genomic data from other aphelid species from the other known genera *Aphelidiumm*, *Pseudoaphelidium* and *Amoeboaphelidium*

Finally, there is the need of more sampling from zoosporic fungal groups (1.7.4. *Incertae sedis* lineages), including sampling of more sanchytrid species that hopefully would present intermediate traits that could give further insights into their evolution (e.g. flagellum). Also, we need more molecular data from Olpidiomycota species to confirm that they form a coherent group and their branching as the sister lineage to non-flagellated Fungi. That sampling needs to include the Olpidiomycota representatives that parasitize nematodes and rotifers (e.g. *O. vermicola* and *O. nematodae*). We also need genomic data from groups without known affinity among fungi like the enigmatic Nephridiophagidae.

- Light sensing assays in sanchytrids

We have been able to characterize the presence of the *BeGC1* gene fusion and the *BeCNG1* gene in our two sanchytrid genomes and observe in their ultrastructure that they possess a putative "eyespot" organelle in the form of a lipid side body in their zoospores. All these elements have been proven thanks to functional assays performed on zoospores to take part in a light sensing

cascade mediated by cGMP levels in the blastocladiomycete *B. emersonii*. These assays included exposure of zoospores to green light to observe a preferential phototaxis behaviour, selective exposure and inhibition of different elements of the light sensing route followed by measurement of cGMP levels, and immunocolocalization of the BeGC1 protein in the lipid organelle membrane. All these functional and immunocolocalization assays need to be performed also in sanchytrid zoospores to confirm that they do possess phototaxis and that the *BeGC1* and *BeCNG1* genes are not just non-functional remnants of a common ancestor shared with Blastocladiomycota.

- Comparative genomic assessments of other holomycotan relevant functional toolkits.

We have addressed the composition of several toolkits (e.g. flagellum, main metabolic pathways) in our almost complete holomycotan genomes. For example, the study of the flagellar toolkit was essential to understand the number of flagellum losses in Fungi. However, considering our new phylogenetic framework, there are other toolkits involved in fungal traits or lifestyle that remain to be analyzed. The toolkits involved in fungal pathogenesis are one of the main candidates for further assessment. From these toolkit family, the homologs of the cell wall degradation enzymes (e.g. cellulases, pectinases) are highly important in parasitic holomycotans (Kubicek, Starr, & Glass, 2014; Torruella *et al.*, 2018). Other enzymes involved in fungal pathogenesis that need to be studied include the glucosyltransferases, which have been shown to facilitate interactions between plants and fungi by enabling growth on solid matrices (King *et al.*, 2017).

Recently, it has been shown that many of the orthologs involved in hyphal multicellularity evolved within the unicellular fungal ancestors which represent the nodes of Blastocladiomycota, Chytridiomycota and Zoopagomycota (Kiss *et al.*, 2019). Precisely these nodes are the ones that are more taxon-rich in our phylogenomic tree. Thus, a new assessment of the hyphal multicellular toolkit using our new Fungi backbone that includes two new branches (Sanchytriomycota and Olpidiomycota) would be essential to understand the evolution of the genes involved in hyphal multicellularity.

Similarly, it would be important to study other functions, such as the chitin enzymatic toolkit (e.g. chitin synthases, deacetylases, chitinases, 1,3-beta-glucan synthases), phagotrophy/osmotrophy, cytoskeleton, membrane-trafficking, meiosis, etc. More globally, studying the overall gain and loss patterns of orthologous genes along the fungal tree would be useful to detect both missing

elements and key innovations that might give us more insight into the evolution and ecological success of this group.

- Molecular clock estimation of divergence times in the new fungal tree of life.

Several studies have calculated divergence times estimates for the different clades within (and outside) the holomycotan tree of life (Lücking *et al.*, 2009; Chang *et al.*, 2015; Lutzoni *et al.*, 2018; Tedersoo *et al.*, 2018). However, most of these studies have 1) considered Blastocladiomycota to be the sister group of all other Fungi, 2) not included aphelids, 3) not included a large diversity within lineages like nucleariids, microsporidia, rozellids or Blastocladiomycota, 4) not included the two new fungal branches Sanchytriomycota and Olpidiomycota and 5) not included new calibration points as the newly discovered chytrid-like fungi from 1 billion years ago (Loron *et al.*, 2019). A new time calibrated tree could give a new perspective on the divergence time of holomycotan taxa and might change some of the estimates and number of events previously inferred for fungal and plant terrestrialization (Lutzoni *et al.*, 2018).

8. Conclusions

"…la selección natural, que no actúa de manera perfecta, pero tiende exclusivamente a proporcionar a cada una de las especies el mayor éxito posible en sus combates por la vida entablados con otras especies en unas circunstancias maravillosamente complejas y cambiantes"

Charles Darwin. Autobiography (1887)

8. Conclusions

I have carried out a phylogenomic study of the unicellular component of the Holomycota branch of the eukaryotic tree of life through the use of single-cell approaches. Many holomycotan lineages remain understudied, with no genomic data available. In order to robustly reconstruct the phylogenetic relations of several of these groups and to initiate comparative genomic analyses, I produced genomic data for several of these lineages. In particular, I focused on the resolution, via multi-gene analyses, of three main branches of the holomycotan tree: the nucleariids, the Microsporidia + Rozellida clade and the (core) Fungi. The most important outcomes of my work are listed as follows:

1) Both, single-cell and cultured-based approaches can be used to obtain useful holomycotan genomic and transcriptomic data. We have increased the amount of molecular data available for nucleariid, metchinikovellid and sanchytrid species through the sequencing of single-cell transcriptomes and genomes (SCTs and SCGs). We applied and compared both culture-based and single-cell techniques in the case of nucleariids.

2) Single-cell approaches allowed retrieving enough conserved markers for phylogenomic studies. The percentage of conserved phylogenetic protein markers recovered for our phylogenomic studies was higher for SCTs than for SCGs in nucleariid cells. These markers did allow to reconstruct robust phylogenomic trees.

3) *Lithocolla* and *Pompholyxophyrs* belong to the nucleariid clade. Our phylogenomic analyses demonstrate that the cover-bearing organisms *Lithocolla* and *Pompholyxophrys* form a monophyletic group sister to the *Nuclearia* clade. In turn, this monophyletic group is sister to the nucleariids lineage formed by the small filose amoeba *Parvularia* and *Fonticula*.

4) The most recent common ancestor of nucleariids was likely a freshwater, bacterivorous, non-flagellated filose mucilaginous amoeba. We inferred the characteristics of the most recent common ancestor (MRCA) of nucleariids by mapping morphological and ecological traits on their phylogenetic tree. The nucleariid MRCA was most likely a freshwater, bacterivorous, non-flagellated filose and mucilaginous amoeba. From this ancestor, two groups evolved to reach smaller (*Parvularia-Fonticula*) and larger (*Nuclearia* and related cover-bearing genera) cell sizes, leading to different ecological specialization. The *Lithocolla* + *Pompholyxophrys* clade developed exogenous or endogenous cell coverings from a *Nuclearia*-like naked ancestor.

5) *Pompholyxophyrs* bears bacterial endosymbionts. We identified sequences likely corresponding to bacterial endosymbiont in the single-cell genomes of *Pompholyxophrys***.** These potential bacterial endosymbionts belong to the Rikettsiales and Chlamydiae.

6) *Metchnikovella incurvata* forms a monophyletic group with *Amphiamblys* sp. (Metchnikovellidae) that is sister to the fast-evolving core Microsporidia. We have sequenced and assembled a single-cell genome of the microsporidian hyperparasite *Metchnikovella incurvata*, which infects gregarines in turn infecting polychaetes. This is the first genome for a morphologically characterized metchnikovellid species. Phylogenomic analyses allowed us to confirm that *M. incurvata* branches together with *Amphiamblys* sp. and confirm the branching order of this clade as sister to all other core Microsporidia.

7) Metchnikovellid genomes resemble more to fast-evolving, core Microsporidia than to other member of the Microsporidia + Rozellida clade. Gene ontology terms corresponding to main metabolic categories show that the gene complement of *Metchnikovella* and *Amphiamblys* resembles more to those of typical derived core Microsporidia than to early branching Microsporidia (*Mitosporidium*) or Rozellids (*Paramicrosporidium* and *Rozella*).

8) The genome of *Metchnikovella incurvata* acquired genes from bacteria by horizontal gene transfer. Our analyses showed the presence of horizontally acquired genes in the genome of *M. incurvata* including a manganese superoxide dismutase (MnSOD). This enzyme most likely protects the metchnikovellid cell from deleterious effects of oxygen.

9) Genome reduction and the appearance of new genes have co-occurred during the evolutionary adaptation of Microsporidia to their hosts. Gain and loss analysis of orthologous genes along the microsporidian branch, suggests that not only genome reduction but also gene gain occurred during the evolution of Microsporidia, likely as adaptation to their obligate parasitic lifestyle.

10) *Amoeboradix gromovi* and *Sanchytrium tribonematis* form a single and well supported fast-evolving clade sister to the Blastocladiomycota. Phylogenomic analyses of the single-cell genomes (SCGs) of *A. gromovi* and *S. tribonematis* show that they form a monophyletic clade sister to Blastocladiomycota. The presence of specific features (e.g. reduced flagellum, long kinetosome) might justify the inclusion of this fast-evolving group into its own phylum, Sanchytriomycota.

11) Chytrids appear to be the sister group to the rest of Fungi and *Olpidium* might be an independent phylum (Olpidiomycota) forming a clade sister to all other non-flagellated Fungi. Our phylogenetic analyses, together with tests over the position of chytrids and *Olpidium* in the fungal tree, indicate that chytrids could be the sister lineage to the rest of fungi, and that the zoosporic fungi *Olpidium* constitutes an independent lineage sister to the non-flagellated fungi.

12) The primary metabolic profile of Sanchytrids differs from that of canonical Fungi, including that of Blastocladiomycota. Analyses of KOG categories related to primary metabolism in Holomycota showed that sanchytrids have an atypical and reduced lipid and carbohydrate metabolism and transport, different from that of "canonical fungi".

13) There have been at least 4 independent losses of the flagellum in Holomycota and sanchytrids possess a highly reduced flagellar toolkit. Sanchytrid zoospores carry a non-motile structurally reduced flagellum (pseudocilium) which it has been characterized by ultrastructure. We analyzed their genome for the presence of proteins involved in the eukaryotic flagellum toolkit in several holomycotan members. Complete sets of proteins involved in flagellar function and maintenance are missing in the genomes of sanchytrids, explaining their reduced ultrastructure. We infer 4 independent losses of the flagellum in Holomycota.

14) Sanchytrids possess homologs of the *BeGC1* fusion gene and the *BeCNG1* channel gene, involved in phototactic response in *Blastocladiella emersonii*. Our analyses showed that the sanchytrid genomes carry the *BeGC1* fusion gene and the *BeCNG1* channel gene, two genes involved in a light sensing cascade in the closely related Blastocladiomycota *B. emersonii*. *B. emersonii* possesses a lipid organelle where these genes colocalize functioning as an "eyespot". A prominent lipid organelle is also present in sanchytrids. This might suggest that sanchytrids can sense light. If this hypothesis is correct, it could explain the maintenance and selection of a highly reduced flagellum but at the same time long kinetosome. We have hypothesized that the pressure towards light perception could have led to maintain a flagellar support structure (the kinetosome) for the lipid eyespot.

9. Supplementary Material

9.1. Supplementary material of manuscript 1

Combined cultivation and single-cell approaches to the phylogenomics of the nucleariid amoebae, close relatives of Fungi

(Phil. Trans. R. Soc. B 374: 20190094)

Luis Javier Galindo[1*], Guifré Torruella[1*], David Moreira[1], Yana Eglit[2], Alastair G.B. Simpson[2], Eckhard Völcker[3], Steffen Clauß[3], Purificación López-García[1]

[1]Unité d'Ecologie, Systématique et Evolution, CNRS, Université Paris-Sud, Université Paris-Saclay, AgroParisTech, 91400 Orsay, France.
[2]Department of Biology, and Centre for Comparative Genomics and Evolutionary Bioinformatics, Dalhousie University, Halifax, Nova Scotia, Canada.
[3]Penard Laboratory, Cape Town, 7708 South Africa.
*authors contributed equally
Authors for correspondence:
Guifré Torruella
e-mail: guifftc@gmail.com
Purificación López-García
e-mail: puri.lopez@u-psud.fr

Detailed taxonomical discussion
Figures S1-3
Videos S1-4

Detailed methodology
Figures S4-6

Detailed taxonomical discussion

Nucleariid amoebae are abundant in freshwater and known since 1865 when Cienkowski described the characteristic genus *Nuclearia* [1], a filose, polymorphous amoeba displaying an always prominent vesicular nucleus and a central nucleolus. Until the late 20th century, this genus was associated with other naked filose amoebae in several different and conflicting taxonomies; see references in [2–4].

Small subunit rRNA (18S rRNA gene) molecular phylogenies placed *Nuclearia* as a deep branch within the opisthokonts [5], particularly as sister to fungi [6,7] as subsequently corroborated by phylogenomic analyses [8,9]. Up to now, only *Parvularia* [10] and *Fonticula* [7] were molecularly confirmed to be related to *Nuclearia* [11]. Here, we have included molecular information for species of two other genera. *Pompholyxophrys* 1869 Archer and *Lithocolla* Schulze 1874, both described as spherical amoebae with fine filopodia. While *Pompholyxophrys* is covered with a layer of self-secreted siliceous scales (idiosomes), *Lithocolla* uses available xenosomes, preferably small quartz grains or alternatively diatom frustules. Rainer [12] considered these genera as heliozoans and created the suborder Rotosphaeridia. After Patterson had shown that ultrastructural characteristics of *Pompholyxophrys* were similar to *Nuclearia* and *Vampyrellidium* [13–15], Page proposed the order Cristidiscoidida [16] with two families, Nucleariidae Cann & Page 1979 and Pompholyxophryidae Page 1987. Roijackers and Siemensma [17] subsequently supported Page's classification on the basis that Rainer had established an heliozoan suborder and Pompholyxophryidae were actually spherical amoeba with fine, filose pseudopodia and included other amoebae with silica-based scales like *Pinaciophora* [18], *Rabdiaster* [19] and *Rabdiophrys* [17]. Mikrjukov, promoted Rotosphaerid(i)a Rainer 1968 and treated Cristidiscoidida Page 1987 as a as a junior synonym [20], including *Elaeorhanis* [21] and *Lithocolla* [22]. Without molecular phylogenetic data, other authors considered them *incertae sedis* [3]. Since then, on the one hand, morphological observations and diversity studies have been done on scale-bearing rotosphaerids under the Rotosphaerida nomenclature [19,23–26] and, on the other hand, molecular phylogenetic studies have rather used the name Cristidicoidea based on Cavalier-Smith [27] and considering only naked nucleariid sequences [28–33]. Our intention here is to provide a summarized description for each genus, including our own observations and discussing the taxonomy including the possible placement of *incertae sedis* nucleariids. For further details, see the original sources cited below, as well as

the multiple online sources of images and descriptions: Microworld, Penard Lab, Protist Information Server and The world of Protozoa.

Overall nucleariid diagnosis: Free-living filose amoebae lacking flagella, spherical to flattened with a prominent big nucleus with central nucleolus, sometimes multinucleated, or syncytial. Cells are usually covered by a mucous coat, a glycocalyx of unknown composition. Most of them present cystic stages. Cells are full of multiple contractile and food vacuoles. Cells have radiating thin hyaline filopodia, sometimes branching, tapering when attached to substrates, with knobs of cytoplasm along the filopod for elongation or retraction. Filopodia are never stiff nor with extrusomes, never anastomosing nor reticulating. All nucleariids have mitochondria with flat or discoid cristae, multiple dictyosomes and few to undetectable microtubular cytoskeletal elements. Cell body sizes, excluding filopodia, from nanoplanktonic to small microplanktonic. They form a robust clade sister to Holozoa and other holomycotans.

Five nucleariid genera

Genus *Fonticula* Worley, Raper & Hohl, 1979. After Patterson and Simpson 2000.

Fonticula alba is the only isolated species, morphologically [34] and molecularly [7] described. Originally isolated from a dog dung in Kansas, USA, individual cells grow on agar feeding on bacteria (e.g., *Klebsiella*). *F. alba* is a small filose amoeba with variable shape, from spherical to elongated, 6 to 12 μm in diameter. Originally characterised as a myxamoeba or slime mould due to their aggregative volcanic-shaped "sorocarp" formation with a stalk based of Golgi-derived material. This multicellular development has been described only in this species [35]. Molecular data is also available for *Fonticula*-like SCN 57-25 derived from metagenome sequence data of a lab-scale bioreactor used to study cyanide and thiocyanate contaminated wastewater in gold ore processing [36]. No 18S rRNA sequence is available, but its phylogenetic position close to *Fonticula alba* has been validated by phylogenomic analyses [36].

Genus *Parvularia* López-Escardó & Torruella, 2018.

Freshwater minute, almost spherical, cells, from 3 to 5 μm, feeding on small rod-shaped bacteria. Sometimes with a single reflecting vacuole occupying the cytoplasm, never when encysted. Both trophic and encysted cells have mucous coat with one or, occasionally, two

nuclei. Radiating branching actin-based filopodia with variable lengths [10]. Multiple environmental studies have reported sequences related, always from freshwater [32,33,37–39].

Genus *Nuclearia* Cienkowski, 1865.

Freshwater naked filose amoebae, from 10 to 60 µm, with a dozen of described species [40], although few characterized by ultrastructure [11,13] or molecular phylogeny [41]. Cells are highly polymorphic, round when floating, and flattened to elongated when attached to the substrate (Figure S3). Most are covered with a mucus coat [42], but like size and shape, the coat can easily change. Cysts and multinucleate cells have been reported several times. Microtubules have only been observed in *N. moebiusi* nuclear division [13]. 18S rDNA sequences can contain insertions in the V4 and V7 regions [10,41], and phylogenetic analyses show that there are two highly supported clades (Figure S4). One clade contains *N. thermophila* [40], *N. delicatula* [1] and *N. moebiusi* [11], while the second clade contains *N. pattersoni* [43] and the uncharacterised strains NZ, D1 and A1 [41]. *N. simplex* [1], observed and morphologically described multiple times, as other *Nuclearia* isolates, are difficult to characterise due to variable morphology. The only molecular data for this species comes from two German strains obtained independently but no longer available. When *N. simplex* CCAP 1552/4 [11] and CCAP 1552/2 [44] were molecularly analysed [6,9] taxon sampling was limited but, according to Dirren and Posch 2016, CCAP 1552/4 should be renamed as *N. moebiusi* and CCAP 1552/2 as *N. pattersoni*.

Genus *Pompholyxophrys* Archer, 1869.

Planktonic freshwater filose amoebae, always nearly spherical, tightly surrounded by spherical, ovoid, discoid or bone-shaped perforated silica pearls. These hollow scales are formed endogenously as observed in *P. punicea* [14] and embedded in a mucilaginous coat. Regarding, ultrastructure, it resembles *Nuclearia*, with mitochondria with flat cristae, nucleus with prominent nucleolus and perinuclear dyctiosomes, missing microtubule-organizing centre or extrusomes. Seven species described, from 15 to 66 µm [17], which feed on algae and detritus. No species available in culture. We have molecularly characterized three distinct species, although we could only ensure the identity of *P. punicea* (accession number MK547175) by its morphology, with spherical scales variable in size (Figure S2A-D, Video S4). The second sequence MK547174 may correspond to *P. stellata* (Figure S2E) since this

species was the second most abundant one in the sampling site according to several years of observations, whereas the third sequence MK547173 clone PB9 may correspond to *P. exigua* (Figure S2F) which can easily be confused with a small *P. punicea*. See more images in www.penard.de/Explorer/Nucletmycea.

Genus *Lithocolla* Schulze, 1874.

Filose amoebae, from 10 to 50 µm, tightly covered with exogenous material embedded within its mucilaginous envelope. This material can be composed of: small quartz grains, diatom frustules, or even chalk particles depending on the medium conditions. Naked cells are also observed in culture. We have noticed in our *L. globosa* SnP culture that the exogenous material is incorporated into the coat by excretion after phagocytotic ingestion (Figure S1). Cytoplasm contains orange to red globules and greenish digestive vacuoles, as originally observed [22]. Cells have multiple radiating and variable filopodia, sometimes tapering at the base and branching. As for some naked nucleariids, *Lithocolla* can be observed flattened or a bit elongated but it is mostly spherical. No-rolling, smaller freshwater *L. flavescens*, ~18 µm, [45] moves fast in fresh samples, whereas our *L. globosa* SnP moved slowly in culture conditions as observed also for *Nuclearia* (videos S1-3). Multiple observations reported this genus as both marine and freshwater, but our molecular phylogenetic analysis placed it within an exclusively marine clade.

Incertae sedis

Genus *Vampyrellidium* Zopf, 1885.

Freshwater naked filose amoebae, from 4 to 25 µm, planktonic or attached to substrate, with mucous coat. Formation of cysts remains unclear. Plasmodia occasionally observed. Two morphologically described species: *V. vagans* [46] and *V. perforans* [15,47]. Very similar to *Nuclearia*, but has a perinuclear striated band and cytoplasmic microtubules. Unlike other nulceariid amoebae, *Vampyrellidium* feeds by penetration of prokaryotic or eukaryotic algal cell walls followed by uptake of cell contents with a specialized pseudopodium.

Genus *Elaeorhanis* Greeff, 1873.

Covered filose amoebae with branching filopodia, central nucleus and a large amount of mucus in the envelope that includes sparse sand grains and diatom shells. Two species described: the freshwater *E. cincta*, usually 14 to 17 µm [21] but up to 26µm [45], and the

marine *E. tauryanini*, ~25 µm and lacking contractile vacuoles [20]. Mikrjukov distinguished them from *Lithocolla* because of the nature of their cell cover, but we have observed *Lithocolla* cultures with sand grains, diatoms or chalk (as also reported in Penard 1904), and also to the rolling movement never observed in *Lithocolla*; although most authors agree that they are most likely related.

Finally, there are a few genera with silica-based scales whose internal taxonomy, exclusively based on scale morphology, remains under debate. Overall, even if we have limited knowledge on the cell biology of these species, we could speculate that these genera may correspond to the MAFO [29] marine and freshwater/marine environmental clades detected in metabarcoding studies (Figure S4); also according to their microplanktonic cell sizes.

Pinaciophora Greeff, 1873, 1875, sensu Penard 1904.
Scales flat or denticulate. Marine and freshwater. 14 to 80 µm. From five to thirteen species, depending on the authors [18–20].

Rabdiophrys Rainer, 1968.
Marine and freshwater filose amoeba with hollow silica-based scales and spines. Six species described, 25 to 40 µm [3] although up to 14 were proposed [17] and even proposed to unite with *Pinaciophora* genus in a single spicule-based clade, but see [19].

Rabdiaster Mikrjukov, 1999.
Cell size 15 - 25 µm, tubular spine-scales with a solid disc base [20].

Thomseniophora, Nicholls 2012.
Silica-scaled periplasm comprising both plate- and spine-shaped scales. Plate-shaped scales double walled with multiple holes in the distal surface; spine-shaped scales tangential with an elongated shaft and a swollen basal structure. Distinguishable from *Rabdiophrys* by a central large hole in the plate-shaped scales, from *Rabdiaster* because this genus lacks holes in the plate-shaped scales; and from *Pinaciophora* (sensu Penard, 1904) which lacks spine-scales [19].

Lithocolla and *Pompholyxophrys* videos can be downloaded from figshare: https://figshare.com/projects/Rotosphaerida_supplementary_material_datasets_images_and_videos/60932

Video S1. SnPLi_05_10fps_1x.mp4
Example of *Lithocolla* movement, from a culture with *Phaeodactylum* as prey. Normal speed. Phase contrast; scalebar is 20 µm.

Video S2. SnPLi_06_1fps_5x.mp4
Example of *Lithocolla* movement, from a culture with *Phaeodactylum* as prey. Sped up 5x. Phase contrast; scalebar is 20 µm.

Video S3. SnPLi_08_1fps5x.mp4
Example of *Lithocolla* movement, from a culture with *Phaeodactylum* as prey. Sped up 5x. Phase contrast; scalebar is 20µm.

Video S4. Video showing *Pompholyxophrys punicea* LG127 being micromanipulated before WGA using a 110 µm VacuTip microcapillary (Eppendorf), with an inner diameter of 60 µm, using a Eppendorf PatchMan NP2 micromanipulator in an inverted microscope Leica Dlll3000 B.

Detailed material and methods

(a) Biological material and acquisition of molecular data

Lithocolla globosa MK547176 was isolated from a coastal marine water sample at 18 m depth obtained by John O'Brien in Splitnose point, Nova Scotia, Canada (44.478459, -63.545766) [48]. *Lithocolla* cells were grown in culture with distinct prey species, and imaged with Zeiss Axiovert 200M and AxioCam M5 (Figure S1, videos S1-3). *Lithocolla* was fed with *Navicula pseudotenelloides* strain NAVIC33 from SERI microalgae culture collection (provided by Charley O'Kelly) in F/2-Si medium for molecular analyses. Standard PCR with 514F-1498R eukaryotic 18S rRNA gene primers was first used to identify it in a preliminary phylogeny. Total RNA was extracted from a *N. pseudotenelloides* culture without *Lithocolla* to remove its data *in silico* once sequenced. A fully grown *L. globosa* culture was kept in 75 cm^2 vented flask for a month in dark conditions to minimize diatom RNA content and maximize *Lithocolla* transcripts. Flasks were scratched to homogenize cells in the medium, cell pellets were obtained by centrifugation, and RNA extracted with RNeasy Micro (Qiagen, Venlo, Limburg, The Netherlands) including a DNAse treatment. The resulting total RNA from *L. globosa* and *Navicula* was sent to the Centre Nacional d'Anàlisi Genòmica (CNAG, Barcelona, Catalonia) to perform polyA selection, prepare Nextera libraries (to minimize cross-contamination) and sequence with a lane of 2x150 bp Illumina (SAMN10847480 and SAMN11022077). Since the *L. globosa* culture was unstable, single cells were micromanipulated with an Eppendorf PatchMan NP2 micromanipulator using a 110 µm VacuTip microcapillary (Eppendorf) in an inverted microscope Leica Dlll3000 B. Multiple displacement amplification (MDA) of RNA was performed with WTA repli-g Qiagen single-cell kit including polyA selection for two cells (LG146 and LG147) which were also sent to CNAG for sequencing with the same protocols as for total RNA. Finally, DNA was amplified using the WGA repli-g Qiagen single-cell kit for 3 cells LG140, LG144, LG145, and pooled the three outputs into a single tube to minimize amplification bias and sequenced by CNAG after confirming cell identity by PCR using 612F-1389R 18S rRNA gene primers (Table 1).

Pompholyxophrys cells have been collected twice in the same freshwater site, a Lake near Zwönitz (50.641142, 12.868578 Saxony, Germany). *C*ells have been observed during multiple years in the site, and recurrently photographed under optical microscopy (Figure S2A-B). For SEM observations (Figure S2C-F), cells were let to settle before fixation (1% OsO4 and 1% HgCl2 for 45 min). After washing (3 × 10 min in distilled water), samples

were dehydrated in ethanol series (30, 50, 70, 90, 96, and 100%) for 10 min each. After critical-point drying and sputter-coating with platinum, cells were visualized with a Zeiss Sigma FE-SEM at 1 kV acceleration voltage. In summer 2016, *P. punicea* cells were identified by optical microscopy from freshly collected water samples, then manually micromanipulated using a micropipette, deposited in two tubes containing 20 and 30 cells; respectively, and conserved in ethanol at room temperature during transportation. Afterwards, cells were pelleted by centrifugation, ethanol removed and then DNA extracted with PicoPure DNA extraction kit (Applied Biosystems). Standard PCR with general eukaryotic primers 82F-1498R, amplicon cloning with TOPO-TA cloning kit (Invitrogen) and Sanger sequencing were performed to obtain 18S rRNA gene sequences. Sequencing of 12 clones per tube revealed that the 20-cell tube was exclusively composed of *P. punicea* MK547175, whereas the 30-cell tube contained also a single sequence of another *Pompholyxophrys* species - clone PB9 MK547173 (probably *P. exigua*, which is morphologically similar to *P. punicea*, [20]). Therefore, only the 20-cell tube DNA was amplified with WGA repli-g Qiagen single-cell kit. Five distinct reactions were performed, pooled into a single tube to minimize amplification bias, tested for bacterial contamination with duplex PCR (Marron et al. 2013) and, although bacteria were always present, sent for sequencing to CNAG with the same approach as above (SAMN10847136). In April 2017, more freshly sampled water was obtained and single *Pompholyxophrys* cells were individually collected with an Eppendorf PatchMan NP2 micromanipulator using a 110 µm VacuTip microcapillary (Eppendorf) in an inverted microscope Leica Dlll3000 B (video S4) and stored in ethanol for further extraction. Some cells were kept alive in culture plates with available food sources such *Tribonema gayanum* and cyanobacteria. Unfortunately, all cultures died after one week. WTA was obtained for two cells (LG129 corresponding to *P. punicea* SAMN10847027 and LG130 corresponding to an unknown *Pompholyxophrys* sp. MK547174, probably *P. stellata* SAMN10847156) and WGA for 2 cells (LG126, molecularly identical to LG130 SAMN10847165 and LG127 SAMN10847162); once PCR identified with 612F-1389R eukaryotic 18S rRNA gene primers, they were sent for sequencing.

Nuclearia pattersoni XT1 MK547179 was a single cell identified from the content of the intestine of a dissected *Xenopus tropicalis* tadpole provided by Nicolas Pollet (Evolution, Génomes, Comportement & Ecologie, CNRS, IRD, Univ. Paris-Sud, Université Paris-Saclay, 91198, Gif-sur-Yvette, France) and isolated with the before mentioned micromanipulator. RNA was extracted and amplified with WTA repli-g Qiagen single-cell kit, the obtained

cDNA was later PCR-identified with 612F-1389R eukaryotic 18S rRNA gene primers, and sequenced (SAMN10861536).

Nuclearia delicatula MK547177 and *Nuclearia thermophila* MK547178 were acquired from the culture collection Sciento (UK) and maintained with *Oscillatoria*-like filamentous cyanobacteria in freshwater medium BG11. The origin of these strains is uncertain but most likely a freshwater spring near Manchester (UK). We isolated both species from the same initial JP100 culture between 2014 and 2016. *Nuclearia delicatula* was first isolated by culture transfer, although with the presence of a *Poterioochromonas*-like (stramenopile) and an *Echinamoeba*-like (amoebozoan) contaminants (Figure S3A-D). Pictures of the distinct species were obtained with a Coolpix P6000 camera on a Nikon Eclipse TS100 inverted microscope with 10X and 40X objectives (Tokyo, Kantō, Japan). One month-old grown culture in a 75 cm^2 vented flask was scratched to homogenize cells in the medium, cell pellets were obtained by centrifugation and extracted with Qiagen RNeasy Micro including DNAse treatment and sent to Eurofins Genomics (Ebersberg, Bavaria, Germany) for polyA library construction and sequencing with Illumina HiSeq SBS Kit v4 2500 2x125 bp (SAMN10996515). *Nuclearia thermophila* was later isolated from the initial JP100 culture with the Eppendorf PatchMan NP2 micromanipulator using a 110 µm VacuTip microcapillary (Eppendorf) in an inverted Leica Dlll3000 B microscope. Pictures of *N. thermophila* cells were obtained under a Zeiss Axioplan 2 microscope (Jena, Thuringia, Germany) with a NEOFLUAR 100X/1,3 oil DIC objective and pictures taken with an AxiocamMR camera using the Zeiss AxioVision 4.8.2 SP1 suite and edited with FIJI c2.0.0-rc-69/1.52i (Figure S3E-H). Multiple cultures were grown in dark conditions for a month to minimize cyanobacterial contamination and then total RNA was extracted as for *N. delicatula* and sent to Eurofins for sequencing (SAMN10996515-6) (Table 1).

(b) Data assembly, decontamination, and annotation

Sequence reads were screened with FastQC [49] before and after quality/Illumina adapter trimming with Trimmomatic v0.33 Paired End mode [50] with the following parameters: ILLUMINACLIP:adapters.fasta:2:30:10 LEADING:20 TRAILING:20 SLIDINGWINDOW:4:28. Resulting reads were assembled with SPAdes v3.9.1 [51]. For each set of reads the data was assembled accordingly. For RNAseq data the -rna option was used, and for MDA data the -sc -careful options.

In order to provide predicted protein sequences, co-assemblies were performed with all *Lithocolla* libraries. For *Pompholyxophrys* data, only after we first ensured that they

belonged to the same species by 18S rRNA gene phylogenetic analyses. Co-assembly rounds for *Pompholyxophrys* sp., *Pompholyxophrys punicea*, and *Lithocolla globosa* were performed two times, one before and one after decontamination To decontaminate the co-assemblies, BlobTools v0.9.19 [52] was used, generating taxon annotated GC plots. Contaminant prokaryotic reads were identified in the generated plots and removed. Reads belonging to potential eukaryotic food were also removed using the BlobTools taxonomic identity after manual inspected with BLAST [53]. In the case of *L. globosa*, to further decontaminate its proteome, total RNA of *Navicula* was sequenced and the predicted proteome was used to eliminated contigs suspected to belong to the diatom using BLASTp.

Pompholyxophrys sp. originally had 86,851 contigs (168,514,180 reads), after first round of decontamination it was decided to keep all eukaryotic reads and reads without any hits having a new total of 43,957 contigs (9,372,824 reads). A last round of decontamination was performed to filter no-hits and eukaryotic contamination for a total of 1,353 contigs (716,400 reads).

Pompholyxophrys punicea originally had 227,098 contigs (105,077,294 reads), after first round of decontamination it was decided to keep all eukaryotic reads and reads without any hits having a new total of 119,205 contigs (14,845,082 reads). A last round of decontamination was performed to filter no-hits and eukaryotic contamination for a total of 3,950 contigs (450,286 reads). LG127 was discarded from co-assemblies since the content was mainly bacterial.

Lithocolla globosa originally had 70,737 contigs (213,254,801 reads), after first round of decontamination it was decided to keep all eukaryotic reads and reads without any hits having a new total of 22,108 contigs (53,952,834 reads). In addition the genomic data of *Navicula* (*Lithocoll*a's food in culture) was available allowing a third round of decontamination in the proteome of *Lithocolla globosa*. A local database was created using the predicted proteomes of four diatom species (*Fistulifera solaris*, *Phaeodactylum tricornutum*, *Thalassiosira pseudonana* and *Navicula*) and three rotosphaerid species (*Fonticula alba*, *Parvularia atlantis* and *Nuclearia thermophila*). The *L. globosa* proteome was then blasted against this database using BLASTp to retain a total of 9,277 proteins.

Proteins were predicted using Transdecoder v2 (http:transdecoder.github.io) with default parameters and filtered using Cd-hit v4.6 [54] with 100% identity. Annotation of the decontaminated co-assemblies was done using eggNOG-mapper from the EggNOG v4.5 [55] database, DIAMOND as mapping mode, and the taxonomic scope to adjust automatically.

Nuclearia pattersoni XT1 reads were assembled using Spades with -sc –careful options producing 453,169 contigs and 41,060 predicted proteins.

Nuclearia thermophila JP100 reads were assembled using RNASpades with default parameters producing 70,205 transcripts and 65,150 predicted proteins.

Nuclearia delicatula JP100 reads were pre-processed and assembled by Eurofins as follows, producing 56,177 transcripts and 54,191 predicted proteins. The raw reads were quality-clipped before performing the assembly using the software Trimmomatic v0.30 and PRINSEQ v0.20.3 [56]. Adapter sequences were removed. Low quality bases from the start and end of each single read were removed (Phred score < 20). Using a sliding window approach, reads were clipped that did not fulfil the average quality threshold (window size: 4 bp, Phred score < 20). Poly-A and Poly-T tails were clipped if longer than 4 bp. Reads with a GC content of less than 20% or more than 80% were removed. Reads containing ambiguous bases ("N") were removed. Low complexity reads were removed (DUST score threshold 7). Duplicated reads were removed. Finally, clipped reads were removed if they were shorter than a length threshold of 60 bp. The remaining clipped read pairs were corrected using the software Seecer v0.1.3 [57] with a kmer-size of 21. The assembly was conducted using in-house pipeline that incorporates the software tools Velvet v1.2.10 [58] and Oases v0.2.8 [59]. A multi-kmer approach was applied. The idea of this approach is to first conduct separate assemblies with different kmer lengths and then merge the individual assemblies to a final assembly. Here, kmer lengths of 55, 65, 75, and 85 were used. The separate assemblies were merged following the filter1-CD-HIT-EST [60]. This procedure reduces redundancy and number of chimeras by filtering each locus by relative transcript length, read coverage and number of transcripts.

(c) 18S rRNA gene phylogeny

The 18S rRNA gene sequences from the distinct species were obtained from the consensus between PCR and high-throughput sequencing (DNAseq, RNAseq, WTA or WGA depending on the species). Based on the 18S rRNA gene alignment from the recent study on *Parvularia atlantis* [10], the 18S rDNA sequences from the three *Nuclearia*, three *Pompholyxophrys* and *Lithocolla globosa* obtained in this study was added using MAFFT v7 online [61] with -add full sequences and -einsi options. Environmental sequences covering the 18S rRNA gene regions v4 and v8-9 from two recent studies were also compiled using

mafft-add fragments -einsi option; one from coastal soil environments in British Columbia islands [32] and the other one from the Parana river in Argentina [33]. Trimming of the spurious regions of the matrix was performed manually since automatic tools removed too much data (e.g., trimAl v1.2 [62] in -automated1 mode kept less than 300 bp). We obtained a matrix of 207 sequences and 1,756 bp. Phylogenetic trees were inferred using distinct software and parameters. For Maximum Likelihood (ML), the IQTREE online server [63] was used to run the GTR+R8+F0 evolutionary model and assessing branch support with 1,000 ultrafast bootstraps, single branch tests SH-like approximate likelihood ratio test [64], and approximate Bayes test [65] (Figure S4A). For Bayesian inference, MrBayes v3.2.6 [66] was used with the GTR+G+I model, with 4 MCMC chains for 1,000,000 generations, sampling every 100 trees and with a burn-in of the first 2500 trees (Figure S4B). In addition, 1,000 non-parametric bootstraps [67] were obtained using IQTREE v1.6.9 using the CIPRES science gateway server [68] with the TIM3+F+I+G4 mode as the best-fitting one based on the Bayesian information criterion from ModelFinder [69] (Figure S4C).

(e) Bacterial endosymbionts in nucleariids

Using Dirren and Posh (2016) sequences of 16S rRNA gene bacterial endosymbionts form *Nuclearia* sp. as queries, we run local BLASTn searches against our nucleariid assemblies for *Pompholyxophrys* (20cellsWGA; SCT: LG126, LG127, LG129, LG130), *Nuclearia*, *Lithocolla* and other nucleariid assemblies available in public databases (*P. atlantis*, *Fonticula alba*, *Fonticula*-like SCN 57-25, *Nuclearia* CCAP 1552/2, and CCAP 1552/4). These results were then blasted back to the whole GenBank database to assess their taxonomic identity. These sequences along with their closest BLAST hits in the GenBank database were included into the Dirren and Posch's alignment using MAFFT v7.388 with default parameters. Alignments were inspected manually using Geneious v6.0.6 [70], and trimmed from ambiguously aligned regions and gaps using trimAl v1.2 in automated1 mode. We worked with 3 datasets, one complete dataset of 100 sequences and 1,503 bp, and two subsets of this first dataset for the Chlamydiae group (18 sequences and 1,454 bp) and the Rickettsiales group (26 sequences and 1,390 bp). 16S rRNA gene trees were inferred by ML using the IQTREE online server applying the GTR model (for the complete dataset and for the Rickettsiales dataset) or the TIM3 model (for the Chlamydiae dataset) with four gamma categories and empirical base frequencies (F+I+G4), which was the best fit model chosen by BIC [71]. The BI analyses were done with MrBayes v3.2. with the GTR+G+I model, with 4

MCMC chains for 1,000,000 generations, sampling every 100 trees and a burn-in of the first 2500 trees. All trees were visualized using FigTree v1.4.3 (Figures 2A-B, Figure S5)

(d) Phylogenomic analyses

Seven nucleariid taxa were included in two distinct phylogenomic datasets. They included six newly generated assemblies plus a *Fonticula*-like SCN 57-25 from the MEDX01 metagenome project on drainage waters [36] and two *Nuclearia* EST datasets [9] to update the multimarker datasets "GBE" [72] and "SCPD" [73]. Each single marker alignment of each dataset was completed with new sequences to include more than 50 taxa representing most eukaryotic lineages in order to detect possible contaminants, paralogs, or markers with complex evolutionary histories. Sequences were aligned with MAFFT v764 [74] using the L-INS-i algorithm after 1,000 iterations, and spuriously aligned regions were trimmed with TrimAl with the automated1 option. Alignments were visualized and edited with Geneious v6.0.6 and trees visualized with FigTree [75]. Each dataset was assembled into a supermatrix with Alvert.py from the package Barrel-o-Monkeys [76]. Resulting matrices were called SCPD21_23481aa (Figures S6A-B) and GBE22_97918aa (Figures S6C-D), since no orthologous markers were retrieved for *Nuclearia pattersoni* XT1 in the SCPD dataset. BI phylogenies were inferred using PhyloBayes-MPI v1.5 [77] with CAT-Poisson as evolutionary model [78], under the Dirichlet process and without constant sites. Two MCMC chains for each dataset were run for >15,000 generations, saving one every 10 trees. Phylogenetic analyses were stopped once convergence thresholds were reached after a burn-in of 25% (i.e., maximum discrepancy < 0.1 and minimum effective size > 100 estimated using bpcomp). For ML phylogenies IQ-TREE v1.6 [79] was used using the mixture model C60 [80]. Statistical support was obtained with 1,000 ultrafast bootstraps [81] and 1,000 replicates of the SH-like approximate likelihood ratio test [65]. All analyses were carried out locally or using the CIPRES Science Gateway.

All datasets from this article have been uploaded as part of this external supplementary material. New nucleariid 18S rRNA gene sequences have been deposited in GenBank with accession numbers MK547173-MK547179, and *Pompholyxophrys* bacterial endosymbionts 16S rRNA gene sequences with accession number MK616425-MK616429. Illumina reads in NCBI SRA under the Bioproject PRJNA517920, and assemblies, phylogenetic matrices and trees in newick format in Figshare:

https://figshare.com/projects/Rotosphaerida_supplementary_material_datasets_images_and_videos/60932.

References

1. Cienkowski L. 1865 Beiträge zur Kenntnis der Monaden. *Arch. für mikroskopische Anat.* , 203–232.
2. Cann JP, Page FC. 1979 Nucleosphaerium tuckery nov. ge. nov. sp - A New Freshwater Filose Amoeba Without Motile Form in a New Family Nucleariidae (Filosea: Aconchulinida) Feeding by Ingestion Only. *Arch. für Protistenkd.* **122**, 226–240. (doi:10.1016/S0003-9365(79)80034-2)
3. Patterson DJ, Simpson AGB, Rogerson A. 2000 Amoebae of uncertain affinities. In *An illustrated guide to the Protozoa, 2nd ed. Lawrence (KS): Society of Protozoologists. Vol. II* (eds JJ Lee, GF Leedale, P Bradbury), pp. 804–826.
4. Adl SM *et al.* 2018 Revisions to the Classification, Nomenclature, and Diversity of Eukaryotes. *J. Eukaryot. Microbiol.* **66**, 4–119. (doi:10.1111/jeu.12691)
5. Zettler L, Nerad T a, O'Kelly CJ, Sogin ML. 2001 The nucleariid amoebae: more protists at the animal-fungal boundary. *J. Eukaryot. Microbiol.* **48**, 293–7.
6. Steenkamp ET, Wright J, Baldauf SL. 2006 The protistan origins of animals and fungi. *Mol. Biol. Evol.* **23**, 93–106. (doi:10.1093/molbev/msj011)
7. 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**, 2699–709. (doi:10.1093/molbev/msp185)
8. Shalchian-Tabrizi K, Minge M a, 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. (doi:10.1371/journal.pone.0002098)
9. 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**, 1–11. (doi:10.1186/1471-2148-9-272)
10. López-Escardó D, López-García P, Moreira D, Ruiz-Trillo I, Torruella G. 2017 Parvularia atlantis Gen. et Sp. Nov., a Nucleariid Filose Amoeba (Holomycota, Opisthokonta). *J. Eukaryot. Microbiol.* **38**, 42–49. (doi:10.1111/jeu.12450)
11. Patterson DJ. 1984 The genus Nuclearia (Sarcodina, Filosea): Species Composition and Characteristics of the Taxa. *Arch. Mikrobiol.* , 127–139.
12. Rainer H. 1968 Urtiere, Protozoa - Wurzelfuessler, Rhizopoda - Sonnentierchen, Heliozoa. In *Die Tierwelt Deutschlands* (ed Fischer), Jena.
13. Patterson DJ. 1983 On the Organization of the Naked Filose Amoeba, Nuclearia moebiusi Frenzel, 1897 (Sarcodina, Filosea) and Its Implications1. *J. Protozool.* **30**, 301–307.
14. 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**, 241–246.
15. Patterson DJ, Surek B, Melkonian M. 1987 The Ultrastructure of Vampyrellidium perforans Surek & Melkonian and Its Taxonomic Position Among the Naked Filose Amoebae1. *J. Protozool.* **34**, 63–67.

(doi:10.1111/j.1550-7408.1987.tb03133.x)

16. Page FC. 1987 The Classification of 'Naked' Amoebae (Phylum Rhizopoda). *Arch. für Protistenkd.* **133**, 199–217. (doi:10.1016/S0003-9365(87)80053-2)
17. Roijackers RMM, Siemensma FJ. 1988 A study of cristidiscoidid amoebae (Rhizopoda, Filosea), with descriptions of new species and keys to genera and species. *Arch. fur Protistenkd.* **135**, 237–253. (doi:10.1016/S0003-9365(88)80072-1)
18. Thomsen HA. 1978 On the identity of the Heliozoan Pinaciophora fluviatilis and Potamodiscus kalbei; with the description of eight new Pinaciophora species. *Protistologica* , 359–373.
19. Nicholls KH. 2012 New and little-known marine species of Pinaciophora, Rabdiaster and Thomseniophora gen. nov. (Rotosphaerida: Pompholyxophryidae). *J. Mar. Biol. Assoc. United Kingdom* **93**, 1211–1229. (doi:10.1017/S002531541200135X)
20. Mikrjukov KA. 1999 Taxonomic Revision of Scale-bearing Heliozoon-like Amoebae (Pompholyxophryidae, Rotosphaerida). *Acta Potozoologica* **38**, 119–131.
21. Greeff R. 1873 Über Radiolarien und radiolarienartige Rhizopoden des süßen Wassers. In *Sitzungsberichte der gesellschaft zur beforderung der gesammten naturwissenschafte zu Marburg*, pp. 47–64.
22. Schulze FE. 1874 Rhizopodenstudien II. *Arch. für mikroskopische Anat.* , 377–400.
23. Esteban GF, Gooday AJ, Clarke KJ. 2007 Siliceous scales of filose-amoebae (Pompholyxophryidae, Rotosphaerida) from deep Southern Ocean sediments, including first records for the Southern Hemisphere. *Polar Biol.* **30**, 945–950. (doi:10.1007/s00300-007-0280-4)
24. Leonov MM. 2012 Heliozoans (Heliozoa, Sarcodina, Protista) of fresh and marine waters of the European part of Russia: Species composition, morphology, and distribution. *Inl. Water Biol.* **3**, 344–355. (doi:10.1134/s1995082910040073)
25. Nicholls KH. 2012 Structure of Organic Spines in the Rhizarian Protist *Belonocystis tubistella* Rainer, 1968, and a Description of *Belonocystis quadrangularis* n. sp. (Cercozoa, Insertae Sedis). *Acta Protozool.* **51**, 131–142. (doi:10.4467/16890027AP.12.002.0385)
26. Wujek DE. 2015 Freshwater Silica-Scaled Heterotrophic Protista: Heliozoa, Thaumatomonad Flagellates, Amoebae, and Bicosoecids, From the Lake Itasca Region, Minnesota. *J. Minnesota Acad. Sci.* **78**, 1–14. (doi:10.1007/s10773-016-2998-7)
27. Cavalier-Smith T. 1993 Kingdom protozoa and its 18 phyla. *Microbiol. Rev.* **57**, 953–94.
28. del Campo J, Ruiz-Trillo I. 2013 Environmental Survey Meta-analysis Reveals Hidden Diversity among Unicellular Opisthokonts. *Mol. Biol. Evol.* **30**, 802–5. (doi:10.1093/molbev/mst006)
29. del Campo J, Mallo D, Massana R, de Vargas C, Richards TA, Ruiz-Trillo I. 2015 Diversity and distribution of unicellular opisthokonts along the European coast analysed using high-throughput sequencing. *Environ. Microbiol.* **17**, 3195–3207. (doi:10.1111/1462-2920.12759)
30. Torruella G *et al.* 2015 Phylogenomics Reveals Convergent Evolution of Lifestyles in Close Relatives of Animals and Fungi. *Curr. Biol.* **25**, 2404–2410. (doi:10.1016/j.cub.2015.07.053)
31. Grau-Bové X, Torruella G, Donachie S, Suga H, Leonard G, Richards TA, Ruiz-Trillo I. 2017 Dynamics of genomic innovation in the unicellular ancestry of animals. *Elife* **6**. (doi:10.7554/eLife.26036)

32. Heger TJ, Giesbrecht IJW, Gustavsen J, del Campo J, Kellogg CTE, Hoffman KM, Lertzman K, Mohn WW, Keeling PJ. 2018 High-throughput environmental sequencing reveals high diversity of litter and moss associated protist communities along a gradient of drainage and tree productivity. *Environ. Microbiol.* **20**, 1185–1203. (doi:10.1111/1462-2920.14061)
33. Arroyo AS, López-Escardó D, Kim E, Ruiz-Trillo I, Najle SR. 2018 Novel Diversity of Deeply Branching Holomycota and Unicellular Holozoans Revealed by Metabarcoding in Middle Paraná River, Argentina. *Front. Ecol. Evol.* **6**, 1–17. (doi:10.3389/fevo.2018.00099)
34. Worley AC, Raper KB, Hohl M. 1979 *Fonticula alba*: A New Cellular Slime Mold (Acrasiomycetes). *Mycologia* **71**, 746–760.
35. Deasey MC, Olive LS. 1981 Role of Golgi Apparatus in Sorogenesis by the Cellular Slime Mold *Fonticula alba*. *Science.* **213**, 561–3. (doi:10.1126/science.213.4507.561)
36. Kantor RS, van Zyl AW, van Hille RP, Thomas BC, Harrison STL, Banfield JF. 2015 Bioreactor microbial ecosystems for thiocyanate and cyanide degradation unravelled with genome-resolved metagenomics. *Environ. Microbiol.* **17**, 4929–4941. (doi:10.1111/1462-2920.12936)
37. Zettler L, Gómez F, Zettler E. 2002 Eukaryotic diversity in Spain's River of Fire. *Nature* **417**, 137.
38. Lara E, Moreira D, López-García P. 2010 The environmental clade LKM11 and *Rozella* form the deepest branching clade of fungi. *Protist* **161**, 116–21. (doi:10.1016/j.protis.2009.06.005)
39. Simon M, López-García P, Deschamps P, Moreira D, Restoux G, Bertolino P, Jardillier L. 2015 Marked seasonality and high spatial variability of protist communities in shallow freshwater systems. *ISME J.* , 1–13. (doi:10.1038/ismej.2015.6)
40. 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**, 147–55. (doi:10.1016/j.ejop.2008.09.004)
41. Dirren S, Posch T. 2016 Promiscuous and specific bacterial symbiont acquisition in the amoeboid genus *Nuclearia* (Opisthokonta). *FEMS Microbiol. Ecol.* **92**, 1–16. (doi:10.1093/femsec/fiw105)
42. Artari A. 1889 Morphologische und biologische Studien über *Nuclearia delicatula* Cienk. *Zool. Anz.* **12**, 408–416.
43. 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. (Praha).* **50**, 161–70.
44. Janitschke K, Lichy S, Westphal C. 1982 Examination of thermally polluted water for free living amoebae and testing for their possible pathogenic properties. *Zentralbl Bakteriol Mikrobiol Hyg B.* **176**, 160–6.
45. Penard E. 1904 *Les héliozoaires d'eau douce*. Genève: H. Kündig. See https://www.biodiversitylibrary.org/item/16282.
46. Zopf W. 1885 *Die Pilzthiere oder Schleimpilze: nach dem neuesten Standpunkte bearbeitet*. Breslau: Eduard Trewendt. See https://archive.org/details/b22301689.
47. Surek B, Melkonian M. 1980 The Filose Amoeba *Vampyrellidium perforans* nov. sp. (Vampyrellidae, Aconchulinida): Axenic Culture, Feeding Behaviour and Host Range Specifity. *Arch. für Protistenkd.* **123**, 166–191. (doi:10.1016/S0003-9365(80)80003-0)

48. Buchwald R, Scheibling R, Simpson A. 2018 Detection and quantification of a keystone pathogen in a coastal marine ecosystem. *Mar. Ecol. Prog. Ser.* **606**, 79–90. (doi:10.3354/meps12776)
49. Andrews S. 2010 FastQC A Quality Control tool for High Throughput Sequence Data.
50. Bolger AM, Lohse M, Usadel B. 2014 Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120. (doi:10.1093/bioinformatics/btu170)
51. Nurk S *et al.* 2013 Assembling Single-Cell Genomes and Mini-Metagenomes From Chimeric MDA Products. *J. Comput. Biol.* **20**, 714–737. (doi:10.1089/cmb.2013.0084)
52. Laetsch DR, Blaxter ML, Leggett RM. 2017 BlobTools: Interrogation of genome assemblies. *F1000Research 2017* **6**, 1–16. (doi:10.12688/f1000research.12232.1)
53. Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL. 2009 BLAST plus: architecture and applications. *BMC Bioinformatics* **10**, 1–9. (doi:doi:10.1186/1471-2105-10-421)
54. Li W, Godzik A. 2006 Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics* **22**, 1658–1659. (doi:10.1093/bioinformatics/btl158)
55. Huerta-Cepas J *et al.* 2016 EGGNOG 4.5: A hierarchical orthology framework with improved functional annotations for eukaryotic, prokaryotic and viral sequences. *Nucleic Acids Res.* **44**, D286–D293. (doi:10.1093/nar/gkv1248)
56. Schmieder R, Edwards R. 2011 Quality control and preprocessing of metagenomic datasets. *Bioinformatics* **27**, 863–864. (doi:10.1093/bioinformatics/btr026)
57. Le H-S, Schulz MH, McCauley BM, Hinman VF, Bar-Joseph Z. 2013 Probabilistic error correction for RNA sequencing. *Nucleic Acids Res.* **41**, e109–e109. (doi:10.1093/nar/gkt215)
58. Zerbino DR, Birney E. 2008 Velvet: Algorithms for de novo short read assembly using de Bruijn graphs. *Genome Res.* **18**, 821–829. (doi:10.1101/gr.074492.107)
59. Schulz MH, Zerbino DR, Vingron M, Birney E. 2012 Oases: Robust de novo RNA-seq assembly across the dynamic range of expression levels. *Bioinformatics* **28**, 1086–1092. (doi:10.1093/bioinformatics/bts094)
60. Yang Y, Smith SA. 2013 Optimizing de novo assembly of short-read RNA-seq data for phylogenomics. *BMC Genomics* **14**, 11. (doi:10.1186/1471-2164-14-328)
61. Katoh K, Rozewicki J, Yamada KD. 2017 MAFFT online service: multiple sequence alignment, interactive sequence choice and visualization. *Brief. Bioinform.* 1–7. (doi:10.1093/bib/bbx108)
62. Capella-Gutiérrez S, Silla-Martínez JM, Gabaldón T. 2009 trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* **25**, 1972–3. (doi:10.1093/bioinformatics/btp348)
63. Trifinopoulos J, Nguyen LT, von Haeseler A, Minh BQ. 2016 W-IQ-TREE: a fast online phylogenetic tool for maximum likelihood analysis. *Nucleic Acids Res.* **44**, W232–W235. (doi:10.1093/nar/gkw256)
64. Guindon S, Gascuel O. 2003 A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst. Biol.* **52**, 696–704. (doi:10.1080/10635150390235520)
65. Anisimova M, Gil M, Dufayard J-F, Dessimoz C, Gascuel O, Nisimova MAA, Il MAG, Ranc JEAN. 2011 Survey of Branch Support Methods Demonstrates Accuracy, Power, and Robustness of Fast Likelihood-based Approximation Schemes. *Syst. Biol.* **60**, 685–99. (doi:10.1093/sysbio/syr041)
66. Ronquist F, Huelsenbeck JP. 2003 MrBayes 3: Bayesian phylogenetic inference under mixed models.

Bioinformatics **19**, 1572–1574. (doi:10.1093/bioinformatics/btg180)

67. Felsenstein J. 1985 Confidence limits on phylogenies: An approach using the bootstrap. *Evolution (N. Y).* **39**, 783–791.
68. Miller MA, Pfeiffer W, Schwartz T. 2010 Creating the CIPRES Science Gateway for inference of large phylogenetic trees. *2010 Gatew. Comput. Environ. Work. GCE 2010* (doi:10.1109/GCE.2010.5676129)
69. Kalyaanamoorthy S, Minh BQ, Wong TKF, von Haeseler A, Jermiin LS. 2017 ModelFinder: fast model selection for accurate phylogenetic estimates. *Nat. Methods* **14**, 587–589. (doi:10.1038/nmeth.4285)
70. Kearse M *et al.* 2012 Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* **28**, 1647–9. (doi:10.1093/bioinformatics/bts199)
71. Posada D. 2008 jModelTest: Phylogenetic model averaging. *Mol. Biol. Evol.* **25**, 1253–1256. (doi:10.1093/molbev/msn083)
72. Mikhailov K V., Simdyanov TG, Aleoshin V V. 2016 Genomic survey of a hyperparasitic microsporidian Amphiamblys sp. (Metchnikovellidae). *Genome Biol. Evol.* **9**, 454–467. (doi:10.1093/gbe/evw235)
73. Torruella G, Grau-Bové X, Moreira D, Karpov SA, Burns JA, Sebé-Pedrós A, Völcker E, López-García P. 2018 Global transcriptome analysis of the aphelid Paraphelidium tribonemae supports the phagotrophic origin of fungi. *Commun. Biol.* **1**, 231. (doi:10.1038/s42003-018-0235-z)
74. 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**, 3059–66.
75. Rambaut A. 2009 FigTree.
76. Leigh JWJW, Roger AJ. 2008 Barrel-o-Monkeys.
77. Lartillot N, Lepage T, Blanquart S. 2009 PhyloBayes 3: A Bayesian software package for phylogenetic reconstruction and molecular dating. *Bioinformatics* **25**, 2286–2288. (doi:10.1093/bioinformatics/btp368)
78. Lartillot N, Philippe H. 2004 A Bayesian mixture model for across-site heterogeneities in the amino-acid replacement process. *Mol. Biol. Evol.* **21**, 1095–109. (doi:10.1093/molbev/msh112)
79. Nguyen LT, Schmidt HA, Von Haeseler A, Minh BQ. 2015 IQ-TREE: A fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol. Biol. Evol.* **32**, 268–274. (doi:10.1093/molbev/msu300)
80. Le SQ, Lartillot N, Gascuel O. 2008 Phylogenetic mixture models for proteins. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* **363**, 3965–76. (doi:10.1098/rstb.2008.0180)
81. Minh BQ, Nguyen MAT, Von Haeseler A. 2013 Ultrafast approximation for phylogenetic bootstrap. *Mol. Biol. Evol.* **30**, 1188–1195. (doi:10.1093/molbev/mst024)

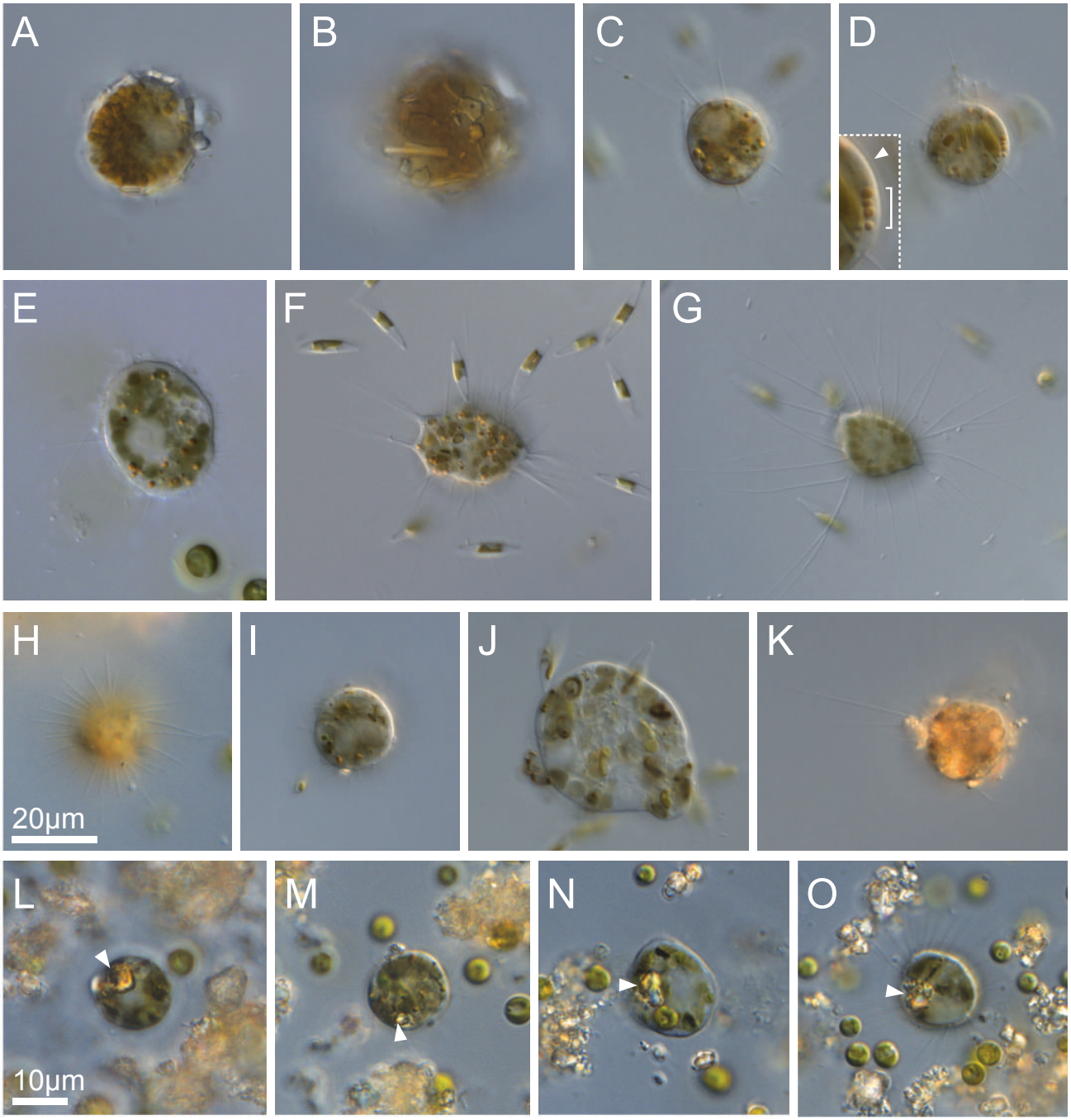


Figure S1. *Lithocolla globosa* SnP. A-B) Agglutinated cells from enrichment with sediment still present. Surface view (B) showing attached mineral particles. C-O) Cells from culture, fed on *Navicula* (C-G), *Isochrysis* (H,L-O) and *Phaeodactylum* (I-K). C-D) General view of same cell. Note nucleus (in C) and ingested diatoms (in D), orange globules (D inset, bracket) and layer (D inset, arrowhead). E) General view of cell. Note nucleus, orange globules and greenish digestive vacuoles. F-H) Examples of intact filopodia. Note that filopodia degrade rapidly under exposure to bright light. I-J) Individuals with 2 (I) and 3 (J) nuclei, relatively commonly observed in culture. K) 'Re-agglutinated' cell one day following addition of chalk dust (as fine calcium carbonate mineral inclusions) to culture. Note mineral particles inside and on the surface of the cell, resembling the agglutinated cells from enrichment in (A-B). L-O) Pseudoseries of different individuals showing mineral inclusions found in cells (arrowheads), 20min (L), 23min (M), 47min (N), 60min (O) following addition of chalk dust to mineral-particle-deprived culture, indicating that mineral particles are ingested by the cell. Differential interference contrast, scalebars are 20µm (A-K, all images at same scale), 10µm (L-O).

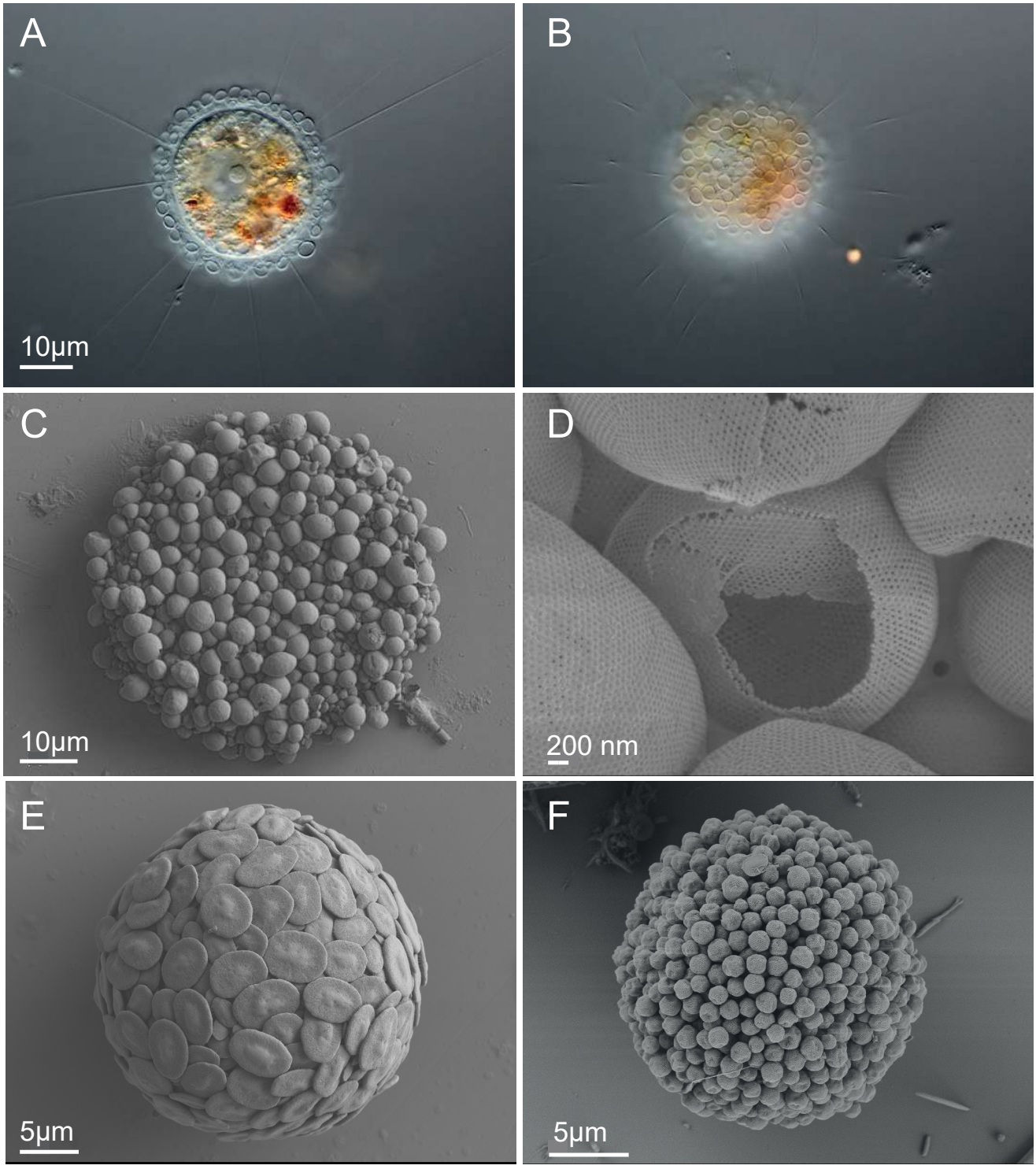


Figure S2. *Pompholyxophrys* morphology from the same sampling site. A-D) *P. punicea*. A-B) Distinct planes of a spherical cell showing radiating filopodia, the typical orange to red food vacuoles and the nucleus with prominent nucleoli. C-D) *P. punicea* spherical pearls with heterogenous size. D) detail of a broken pearl showing the hollow structure. E) *P. stellata*. Note flat scales. F) *P. exigua*. Note small homogeneous spherical scales, compared to *P. punicea*. A-B) Differential interference contrast, scalebars are 10µm. C-F) Scanning Electron Microscopy, scalebars are 10 µm, 200 nm and 5 µm.

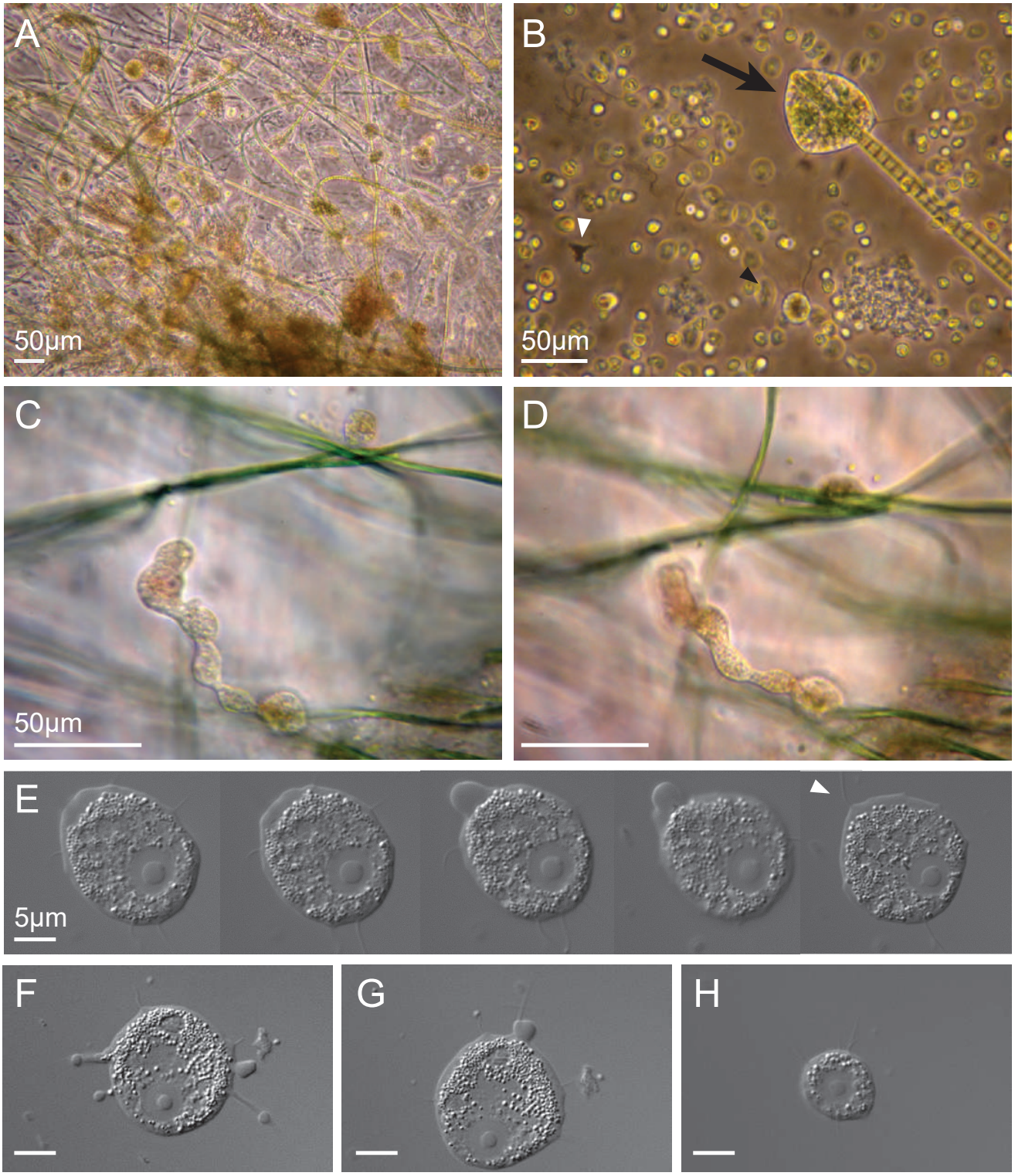


Figure S3. *Nuclearia* JP100. A) General view of a culture of *Nuclearia* JP100 Sciento. B) Details of the distinct species in the original JP100 culture: *Poterioochromonas*-like stramenopile (small round cells), *Echinamoeba*-like amoebozoan (white arrowhead), small *Nuclearia* (black arrowhead), big *Nuclearia* feeding on *Oscillatoria* filamentous cyanobacteria (black arrow). C-D) *Nuclearia* pluricellular form, not clear whether plasmodial or syncytial. E-H) *Nuclearia thermophila* JP100 under cover slip. Note the nucleous with central prominent nucleolus. E) the same cell elongating plasma membrane and creating a filopod (white arrohead). F-H) Different cells showing distinct plasmodial protrusions. A-D) Phase contrast, scalebars are 50µm. E-H) Differential interference contrast, scalebars are 5µm.

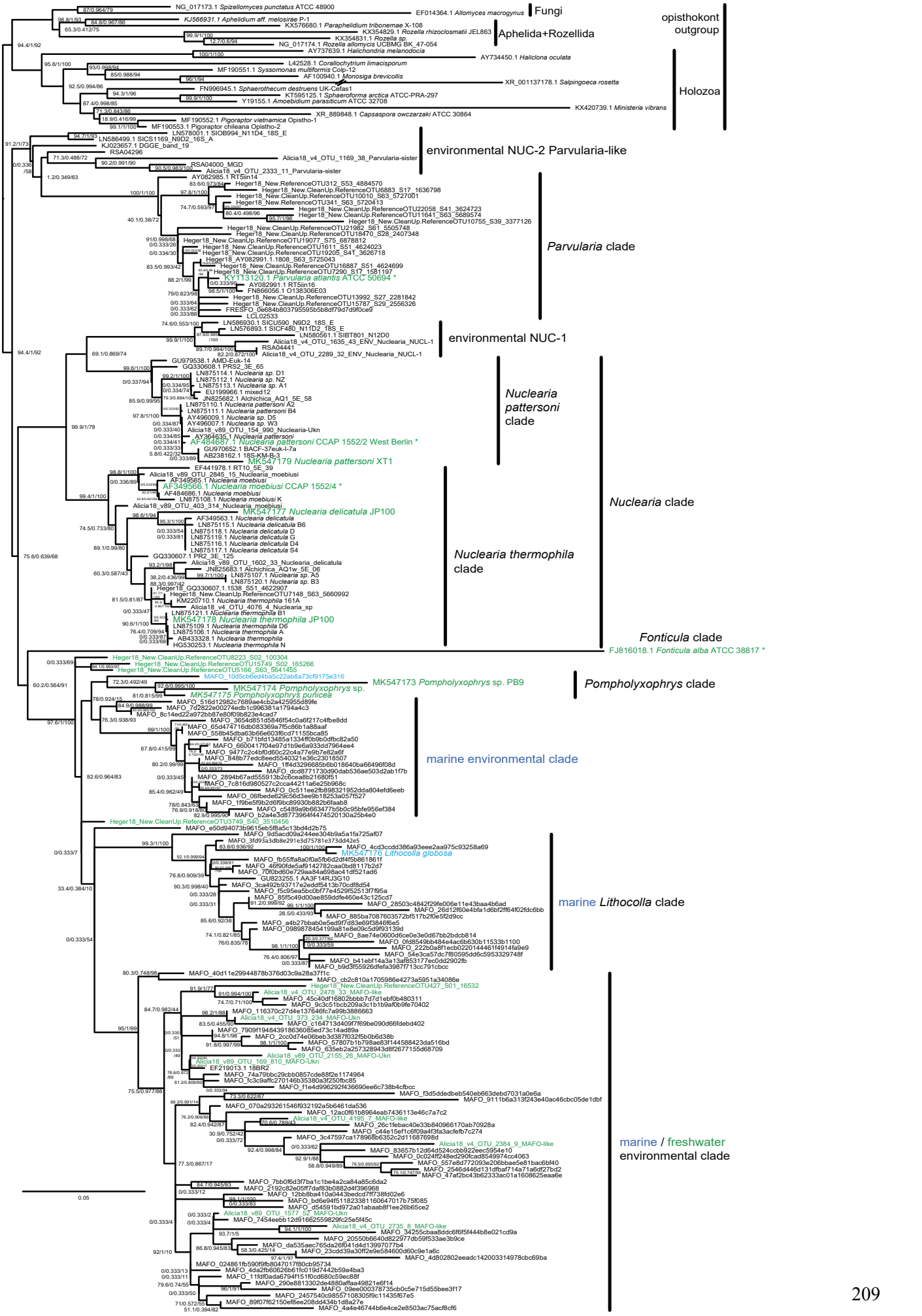


Figure S4A. 18S rRNA gene phylogeny. Maximum Likelihood phylogenetic tree (1,756 conserved nucleotide positions and 207 species) inferred with the GTR+R8 model. Branch supports are from left to right 1,000 ultrafast bootstrap approximation, approximate Bayes test and SH-like approximate likelihood ratio test.

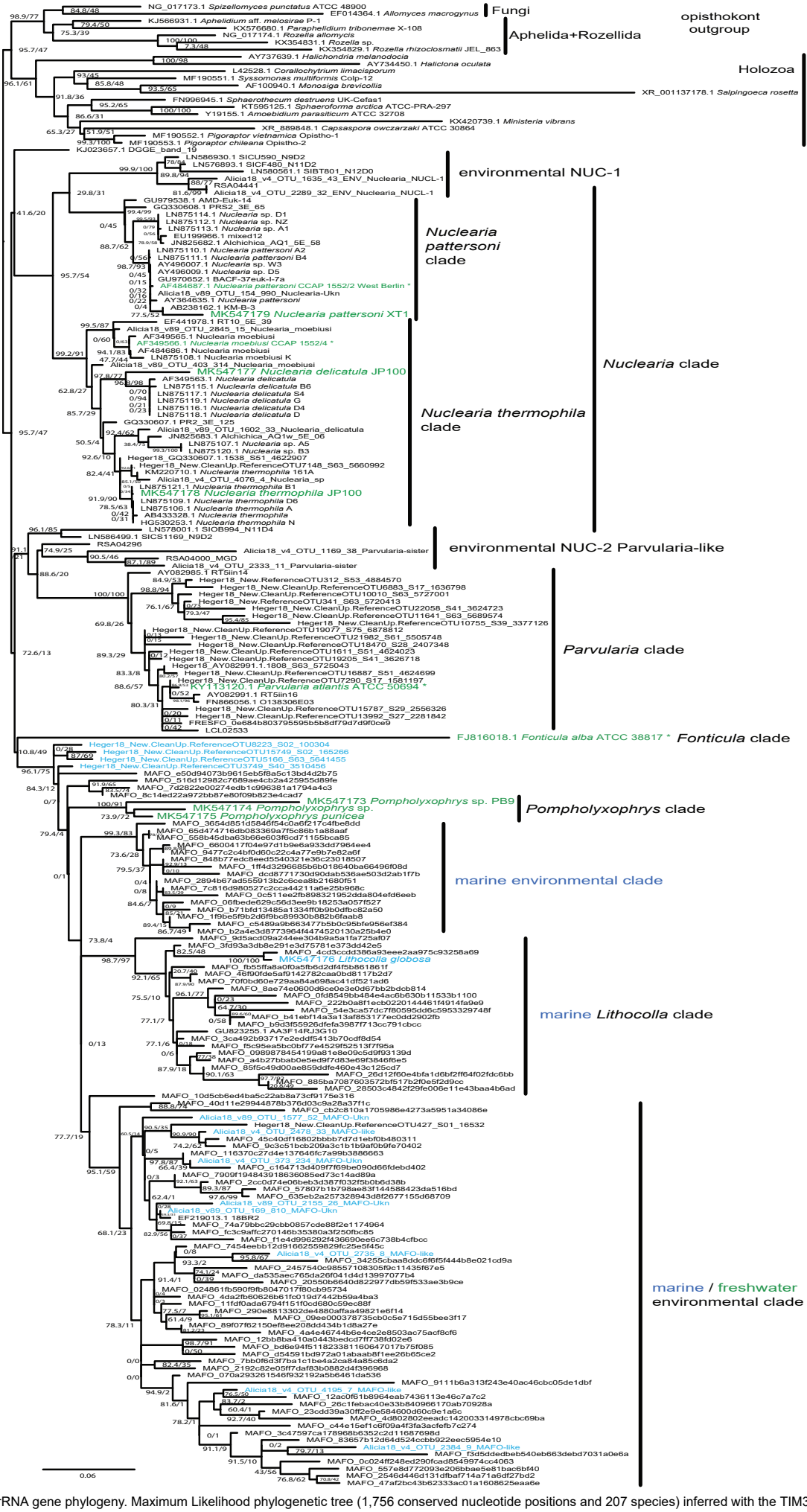


Figure S4B. 18S rRNA gene phylogeny. Maximum Likelihood phylogenetic tree (1,756 conserved nucleotide positions and 207 species) inferred with the TIM3+F+I+G4 model. Branch supports are from left to right SH-like approximate likelihood ratio test and 1,000 non-parametric bootstraps.

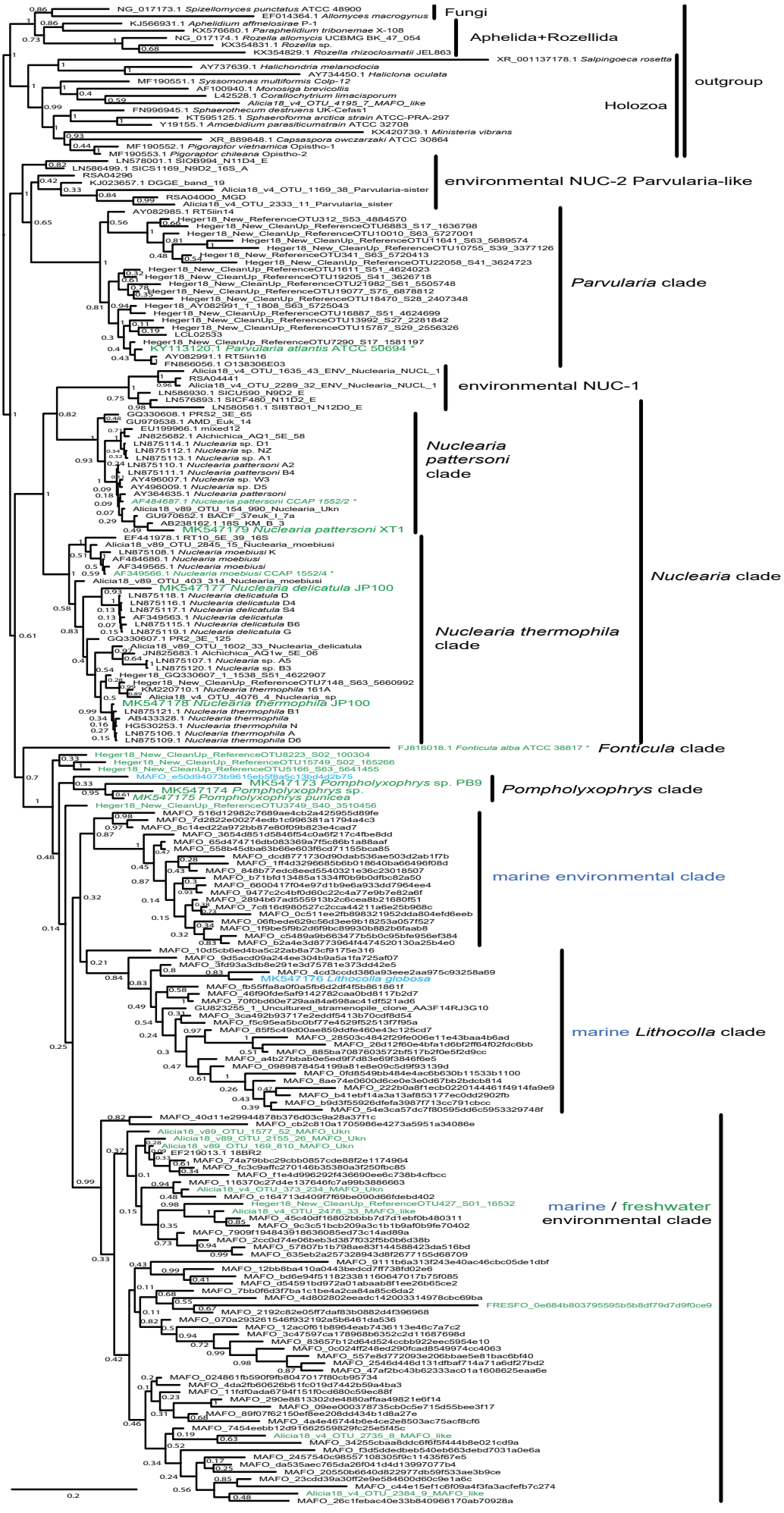


Figure S4C. 18S rRNA gene phylogeny. Bayesian Inference phylogenetic tree (1,756 conserved nucleotide positions and 207 species) inferred with the GTR+G+I model. Branch supports are posterior probabilities.

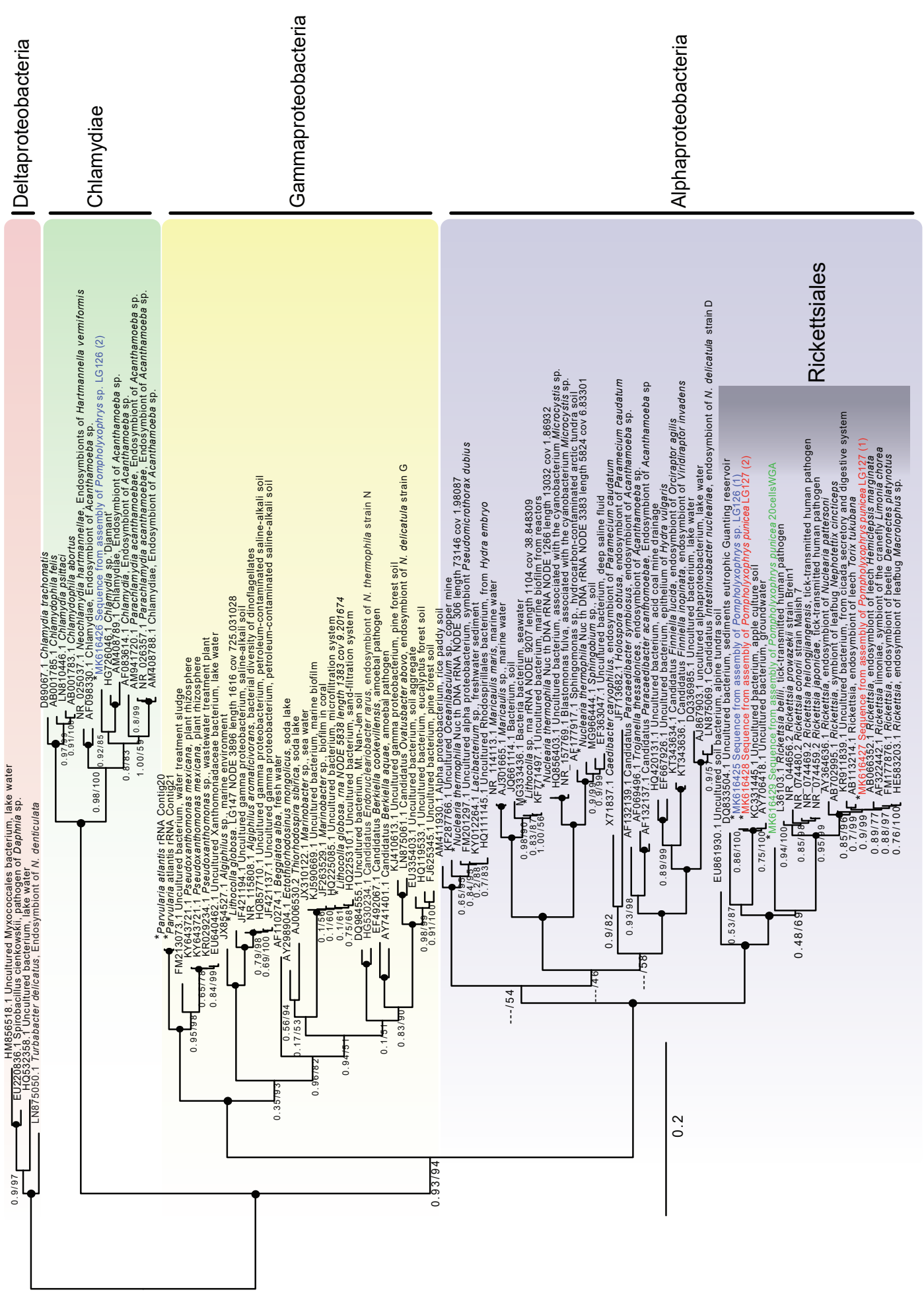


Figure S5. 16S rRNA gene Maximum Likelihood phylogenetic tree showing the known diversity of nucleariid bacterial endosymbionts. The tree contains 87 bacterial sequences retrieved from GenBank and 13 sequences obtained in this study, considered potential endosymbiont candidates (marked with asterisks). Six sequences were finally considered to come from actual endosymbionts: one sequence from *Pompholyxophrys* sp. LG126 (2) shown in blue, two from *Pompholyxophrys punicea* LG 127 (1,2) shown in red, one from *Pompholyxophrys* sp. LG126 (1) shown in blue and one from *Pompholyxophrys punicea* 20cellsWGA shown in green. The tree was inferred using 1,503 bp conserved nucleotide positions. BI tree inference with the model GTR+G+I. ML tree inference with the model GTR+F+I+G4 UFBootstrap. Statistical supports are shown at the nodes; Bayesian posterior probabilities (pp) are shown in the left and Maximum Likelihood ultrafast bootstrap (ufbs) values on the right. Black dots represent supports higher or equal than 0.99 pp and 95% ufbs.

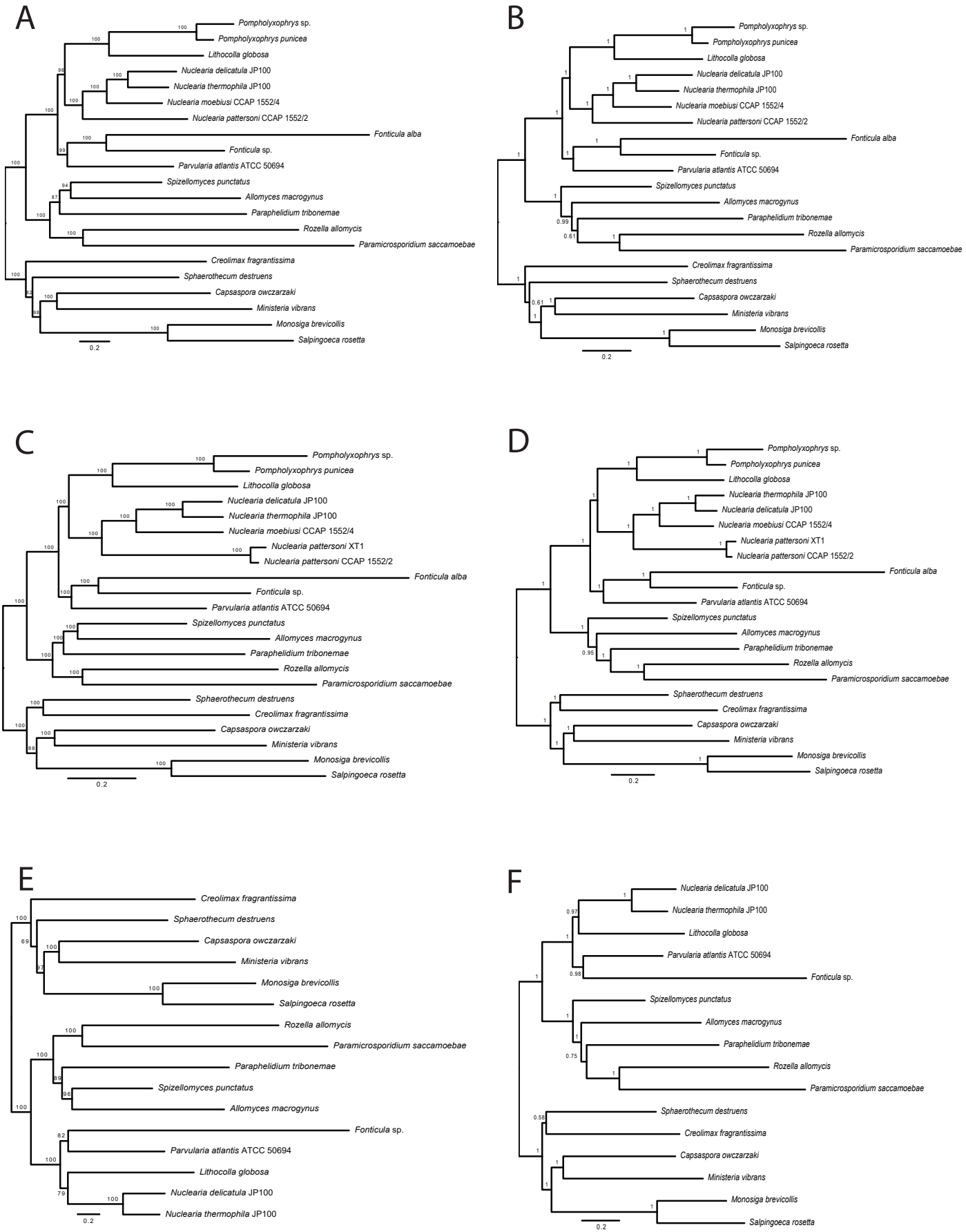


Figure S6. Phylogenomic trees based on the SCPD (A-B,E-F) and GBE (D-E) datasets. Maximum Likelihood trees (A, C, E) were reconstructed using IQ-TREE under the LG+R5+C60 model and ultrafast bootstrap as statistical support. Bayesian inference trees (B, D, F) were inferred using PhyloBayes with the CAT-Poisson model with posterior probability as statistical support. GBE dataset was built using 264 protein alignments with 22 species and 96,276 conserved amino acidic positions (D-E). SCPD dataset was built under two taxon samplings, with 21 species and 23,481 conserved amino acidic positions (A-B) and a second taxon sampling without taxa with >50% of missing data for a total of 16 species and 23,924 conserved amino acidic positions (E-F).

Table S1. Sequencing metadata and statistics of the newly generated molecular data. For more details go to: https://figshare.com/authors/Luis_Javier_Galindo/6432803

Bioproject: PRJNA517920																			
Lithocolla globosa MK547176	**source**	**SSU ribosomal identity**	**design description**	**Volume uL**	**[ ] ng/uL**	**total ug**	**SRA**	**library ID**	**Million read-pairs (or reads)**	**Yield (Gb)**	**Avg PhiX error r1**	**Avg PhiX error r2**	**Trimmomatic HC:10 L:20 T:20 SW:4:28 ML:100**	**Assembly**	**# contigs/scaffolds**	**# proteins**	**# clean proteins**
RNAseq	culture with *Navicula*	82F-1498R	ingek_08: Total RNA from culture using Qiagen RNeasy Micro REF 74004 resuspended in water. PolyA selection. NEXTERA library, 1 lane HS25000 v2, 2x150pb; >14.7Gb/sampl	13	70	0.91	SAMN10847460	HMFYKBBXX_2_N703-S502-NX-xt	47 518	23.854	1.53	1.72	41033000 Both Surviving: 16894123 (41.17%) Forward Only Surviving: 15037527 (36.65%) Reverse Only Surviving: 1063936 (2.59%) Dropped: 8037414 (19.59%)	Co-assembly with RNASpades	70 737	72 580	9 277
WGA	LG140, LG144, LG145 single-cell micromanipulation	612F-1376R	GEN180808_L: WGA repli-g Qiagen single-cell. Library HiSeq SBS Kit v4 2x100 bp	23	652	15	SAMN10847428	GEN180808_L_UT1	lb57490_1163_1	154 626 638	15.462	88.3 %Q30		77313319 Both Surviving: 63383851 (81.98%) Forward Only Surviving: 3686984 (4.77%) Reverse Only Surviving: 5868181 (7.59%) Dropped: 4375323 (5.66%)				
WTA	LG146 single-cell micromanipulation	612F-1376R	Singek_03: WTA repli-g Qiagen single-cell REF150065: polyA selected cDNA resuspended in water. NEXTERA library, 1 lane HS25000 v2, 2x250pb; >14.7Gb/sample	20	18.6	0.37	SAMN10847194	HMLNJBCXY_1_N703-N502-NX					16963855 Both Surviving: 3180053 (16.77%) Forward Only Surviving: 6826421 (36.00%) Reverse Only Surviving: 1072612 (5.66%) Dropped: 7684769 (41.58%)				
WTA	LG147 single-cell micromanipulation	612F-1376R	Singek_03: WTA repli-g Qiagen single-cell REF150065: polyA selected cDNA resuspended in water. NEXTERA library, 1 lane HS25000 v2, 2x250pb; >14.7Gb/sample	20	158	3.16	SAMN10847194	HMLNJBCXY_2_N703-N503-NX	37 213	18.681	1.53	1.72	18248555 Both Surviving: 2366535 (12.97%) Forward Only Surviving: 7300139 (40.00%) Reverse Only Surviving: 750897 (4.11%) Dropped: 7830984 (42.91%)				
Navicula sp.																			
RNAseq	culture	-	ingek_08: Total RNA from culture using Qiagen RNeasy Micro REF 74004 resuspended in water. PolyA selection. NEXTERA library, 1 lane HS25000 v2, 2x150pb; >14.7Gb/sampl	13	150	1.95	SAMN11022077	HMFYKBBXX_2_N704-S503-NX-xt	44 463	13.428	0.41	0.56	44463054 Both Surviving: 21595126 (48.57%) Forward Only Surviving: 14223855 (31.99%) Reverse Only Surviving: 1197652 (2.69%) Dropped: 7446421 (16.75%)	RNASpades	36 618	3 350	-
Pompholyxophrys sp. clone PB9 MK547173																			
	cloned sequence from 30 micromanipulated cells of mixed Pompholyxophrys sp.	82F-1498R	-	-	-	-	-	-											
Pompholyxophrys sp. MK547174																			
WGA	LG126 single-cell micromanipulation	612F-1376R	GEN180808_L: WGA repli-g Qiagen single-cell. Library HiSeq SBS Kit v4 2x100 bp	12	880	10.5	SAMN10847165	GEN180808_L_POM1_lb57488_1163_1	146 215 632	14 621	92.2 %Q30		73107816 Both Surviving: 64090244 (87.67%) Forward Only Surviving: 3008092 (4.11%) Reverse Only Surviving: 3039046 (4.16%) Dropped: 2970434 (4.06%)				
WTA	LG130 single-cell micromanipulation	612F-1376R	Singek_03: WTA repli-g Qiagen single-cell REF150065: polyA selected cDNA resuspended in water. NEXTERA library, 1 lane HS25000 v2, 2x250pb; >14.7Gb/sample	52	298	15.49	SAMN10847156	HMLNJBCXY_1_N701-N517-NX	37 135	18.642	1.53	1.72	18187002 Both Surviving: 3250322 (17.87%) Forward Only Surviving: 9557506 (52.55%) Reverse Only Surviving: 245598 (1.35%) Dropped: 5103586 (28.23%)	Co-assembly with Spades --sc --careful	86 851	39 399	1 094
								HMLNJBCXY_2_N701-N517-NX	-	-	-	-	18948205 Both Surviving: 4353084 (22.97%) Forward Only Surviving: 8897722 (46.96%) Reverse Only Surviving: 378926 (2.00%) Dropped: 5318473 (28.07%)				
Pompholyxophrys punicea MK547175																			
WTA	LG129 single-cell micromanipulation	612F-1376R	Singek_03: WTA repli-g Qiagen single-cell REF150065: polyA selected cDNA resuspended in water. NEXTERA library, 1 lane HS25000 v2, 2x250pb; >14.7Gb/sample	52	236	12.27	SAMN10847027	HMLNJBCXY_1_N702-N502-NX	39 501	19.829	1.53	1.72	18187002 Both Surviving: 3250322 (17.87%) Forward Only Surviving: 9557506 (52.55%) Reverse Only Surviving: 245598 (1.35%) Dropped: 5103586 (28.23%)				
								HMLNJBCXY_2_N702-N502-NX	-	-	-	-	18948205 Both Surviving: 4353084 (22.97%) Forward Only Surviving: 8897722 (46.96%) Reverse Only Surviving: 378926 (2.00%) Dropped: 5318473 (28.07%)	Co-assembly with Spades --sc --careful	227 096	82 091	3 121
WGA	20 cell (20cellsWGA) micromanipulated MDA/6	82F-1498R	Singek_04: WGA repli-g Qiagen single-cell suspended in water. NEXTERA library, 1 lane HS25000 v2, 2x250pb; >14.7Gb/sample	54	550	29.7	SAMN10847136	HMLNJBCXY_1_N706-N502-NX	47 518	23.854	1.53	1.72	23410117 Both Surviving: 12144448 (51.88%) Forward Only Surviving: 7299734 (31.18%) Reverse Only Surviving: 381382 (1.63%) Dropped: 3584653 (15.3%)				
								HMLNJBCXY_2_N706-N502-NX	-	-	-	-	24107543 Both Surviving: 15723815 (65.22%) Forward Only Surviving: 4263317 (17.68%) Reverse Only Surviving: 580162 (2.41%) Dropped: 3540249 (14.69%)				
WGA	LG127 single-cell micromanipulation	612F-1376R	GEN180808_L: WGA repli-g Qiagen single-cell. Library HiSeq SBS Kit v4 2x100 bp	12	875	10.5	SAMN10847162	GEN180808_L_POM2	lb57489_1163_1	137 065 246	13.706	92.2 %Q30		68532623 Both Surviving: 5592 1036 (81.01%) Forward Only Surviving: 2926950 (4.27%) Reverse Only Surviving: 8537672 (12.46%) Dropped: 1547265 (2.26%)	Spades --sc --careful	2 356	-	-
Nucleoria patterseni XT1 MK547179																			
WTA	single-cell micromanipulation, multiplexed with Opalina	612F-1376R	WTA repli-g Qiagen single-cell REF150065: polyA selected cDNA resuspended in water. MiSeq Version3 2x300	12	880	10.5	SAMN10861536	XT1_GATCAG_L001	7 062 454	4 237	29.61 mean Q		7062454 Both Surviving: 1285672 (18.20%) Forward Only Surviving: 4323938 (61.22%) Reverse Only Surviving: 95447 (1.35%) Dropped: 1357827 (19.23%)	Spades --sc --careful	453 189	41 060	-
Nucleoria delicatula JP100 MK547177																			
RNAseq	culture nov14 multiplexed with Rectinomonas	82F-1498R	Total RNA from culture using Qiagen RNeasy Micro REF 74004 resuspended in water. PolyA selection. Illumina HiSeq 2x125bp	12.5	116	14.5	SAMN10996515	RNA_Nuc_CTTGTA_L002	83 127 257	10 390	33.1 mean Q		preprocessed by Eurofins: 67 031 399	Velvet + Oases	56 177	54 191	-
Nucleoria thermophila JP100 MK547178																			
RNAseq	culture cleaned	82F-1498R	Total RNA from culture using Qiagen RNeasy Micro REF 74004 resuspended in water. PolyA selection. Illumina HiSeq SBS Kit v4 2500 2x125bp	29	9.16	265.64	SAMN10996516	NUC_100_RNA_2_GEN150818_Lo_seq15	64 644 125	16 161	84.19 %Q30	32.76 mean Q	64644125 Both Surviving: 17732712 (27.43%) Forward Only Surviving: 9472750 (14.65%) Reverse Only Surviving: 9741130 (15.07%) Dropped: 27697513 (42.85%)	Trinity	70 205	65 150	-
								NUC_100_RNA_2_GEN150818_Lo_nov15	63 908 111	15 977	87.79 %Q30	33.6 mean Q	63908111 Both Surviving: 23303543 (36.46%) Forward Only Surviving: 10702882 (16.75%) Reverse Only Surviving: 7723534 (12.09%) Dropped: 22178152 (34.70%)				

9.2. Supplementary material of manuscript 2

Evolutionary Genomics of Metchnikovella incurvata (Metchnikovellidae): An Early Branching Microsporidium

(Genome Biol. Evol. 10(10):2736–2748)

Luis Javier Galindo[1]*, Guifré Torruella[1], David Moreira[1], Hélène Timpano[1], Gita Paskerova[2], Alexey Smirnov[2], Elena Nassonova[2,3], and Purificación López-García[1]*

[1]Ecologie Systématique Evolution, CNRS, Université Paris-Sud, AgroParisTech, Université Paris-Saclay, Orsay, France.

[2]Department of Invertebrate Zoology, Faculty of Biology, St Petersburg State University, Russia

[3]Laboratory of Cytology of Unicellular Organisms, Institute of Cytology Russian Academy of Sciences, St. Petersburg, Russia

* Corresponding authors

E-mails: puri.lopez@u-psud.fr; luis.galindo@u-psud.fr.

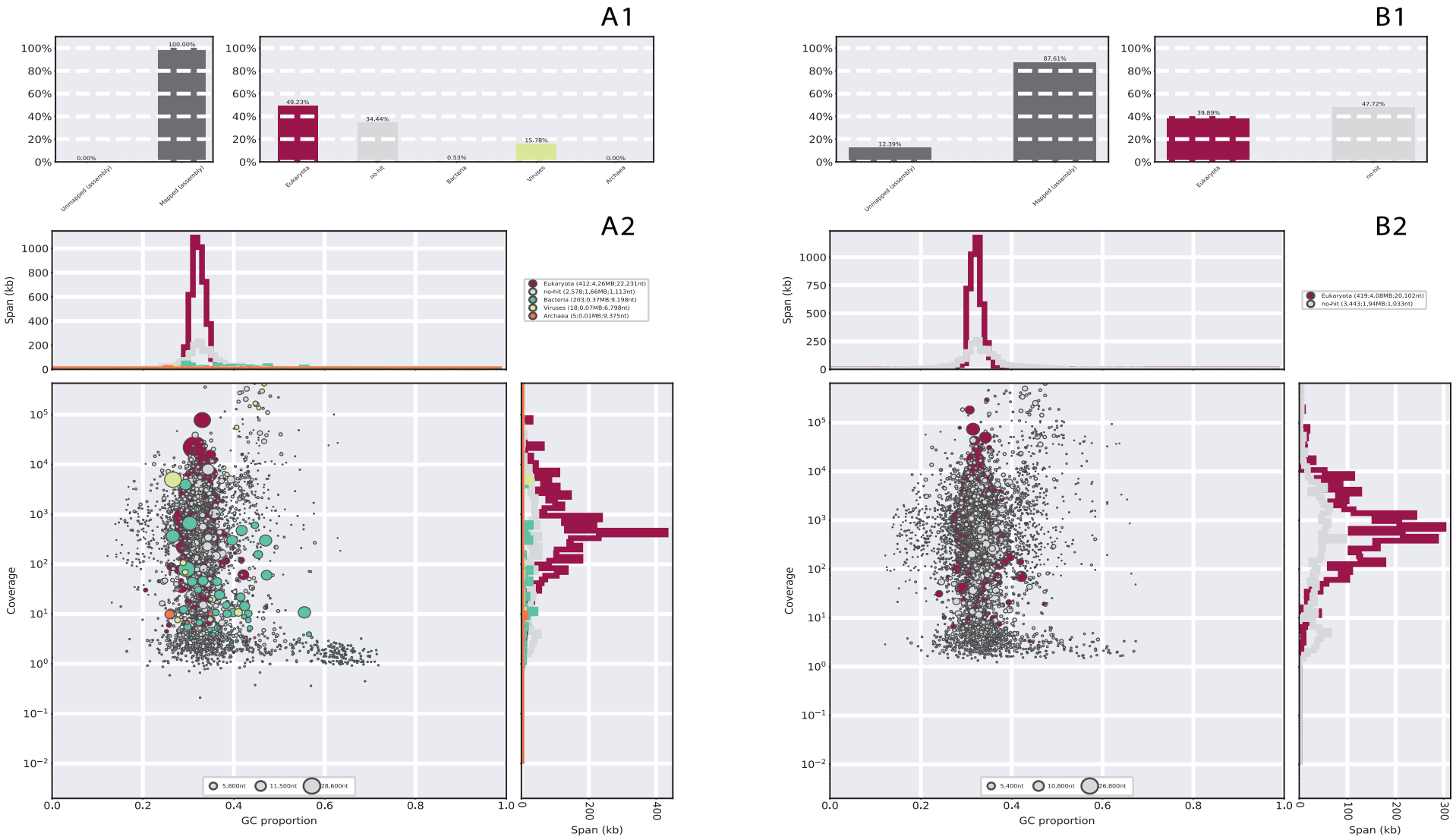


Fig. S1. Visualization of metchnikovellid and contaminant sequences by BlobTools. A1 and B1 are histograms after and before decontamination respectively, representing the proportion of reads of a library that are unmapped or mapped, showing the percentage of mapped reads by taxonomic group. A2 and B2 are taxon-annotated scatter plots decorated with coverage and GC content after and before decontamination respectively, the legend represents the taxonomic affiliation of sequences and lists count, total span and N50 by taxonomic group. Colors represent eukaryotes (purple), bacteria (green), viruses (yellow) and archaea (orange).

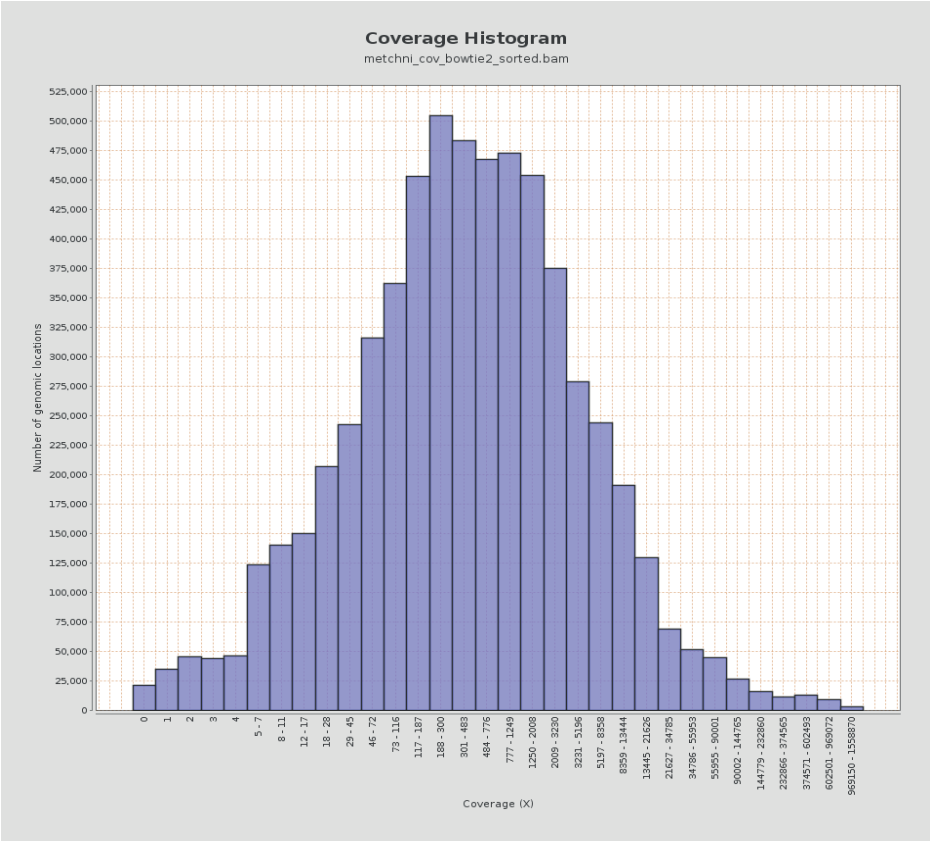


Fig. S2. Coverage histogram made with Qualimap v.2.21 (Okonechnikov et al. 2015) representing the normal distribution of the coverage in the *M. incurvata* genome. The X axis represents the tota coverage across the genome and the Y axis represents the number of genomic locations.

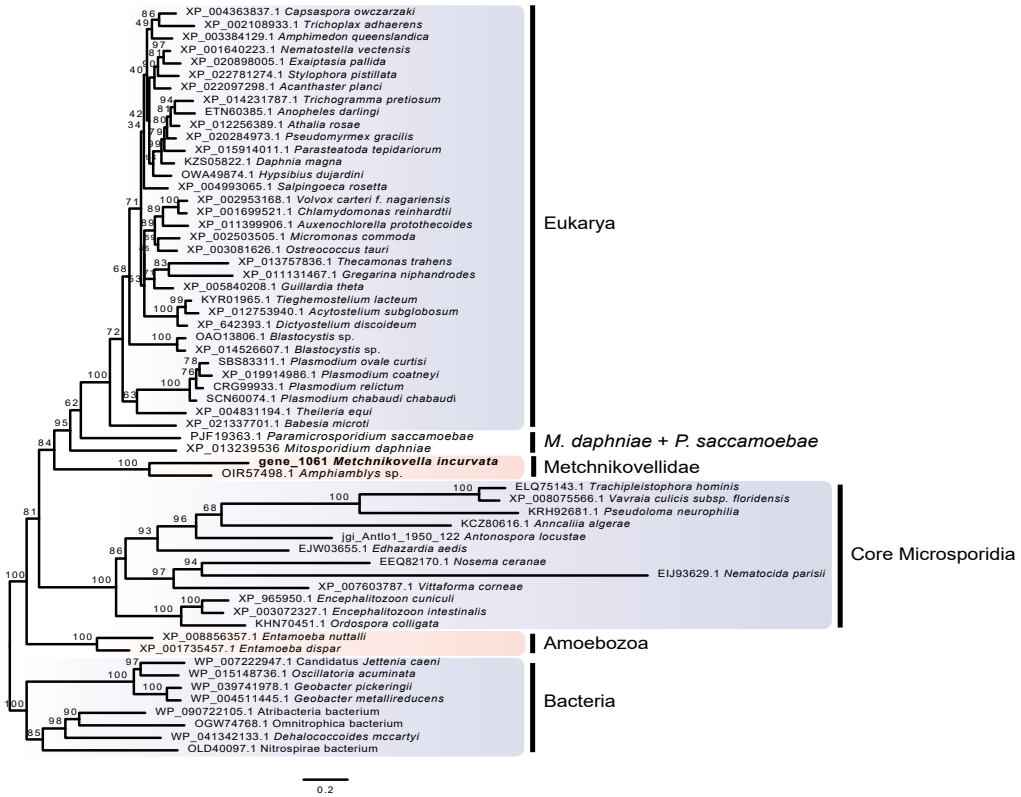


Fig. S3. Maximum Likelihood (ML) phylogenetic tree of RtcB-like ligase proteins. The tree included 61 sequences and was reconstructed with IQ-TREE under the LG+I+G4 evolutionary model. Sequences obtained in this study are highlighted in black. Bootstrap values are represented at nodes.

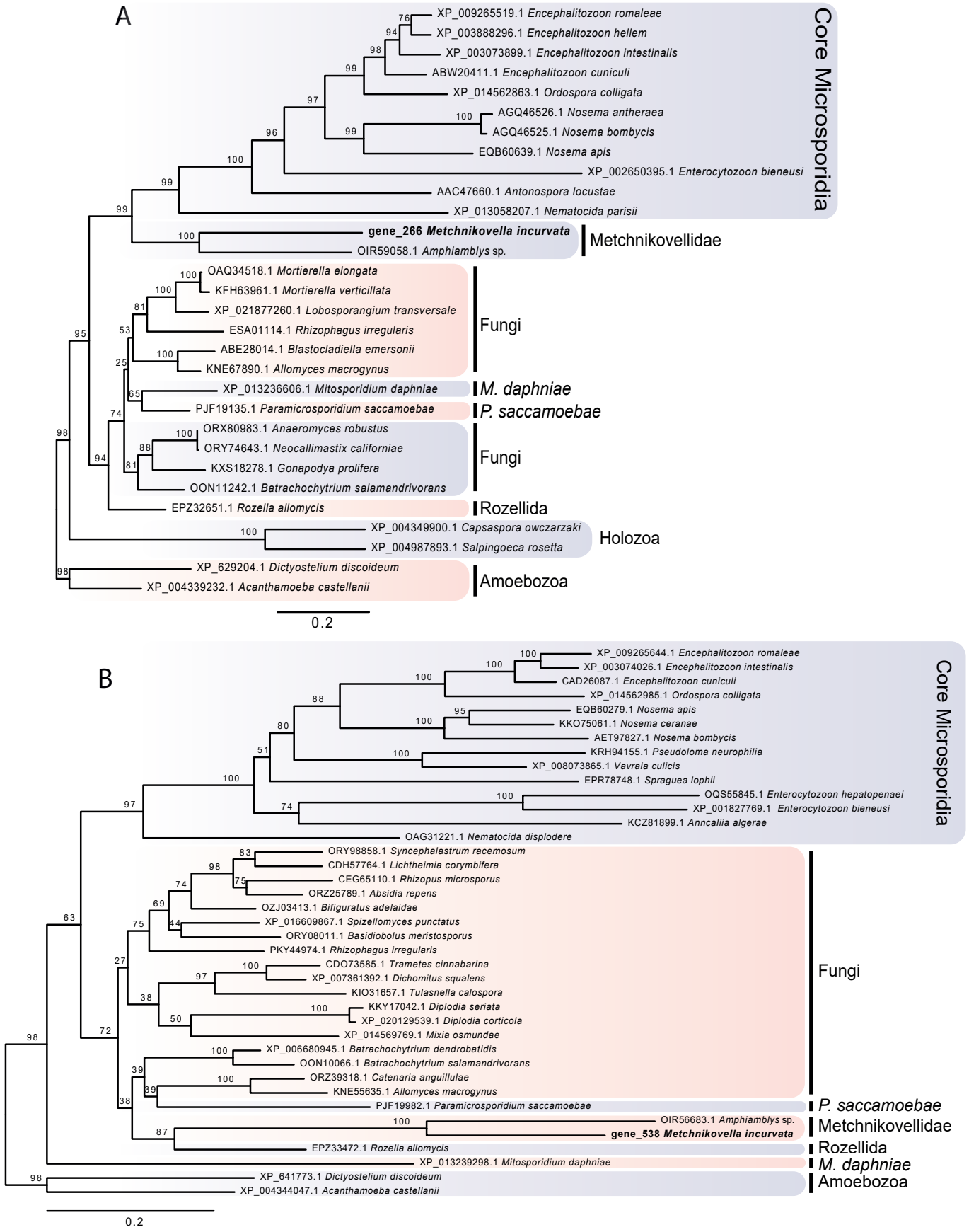


Fig. S4. A. ML tree for the Hsp70 protein. The tree includes 32 protein sequences and was made with IQ-TREE v1.6.2 under the LG+I+G4 evolutionary model. B. ML tree for the Nfs1 protein. The tree includes 39 protein sequences and was made with IQ-TREE v1.6.2 under the LG+I+G4 evolutionary model. Sequences obtained in this study are highlighted in black. Bootstrap values are indicated at nodes.

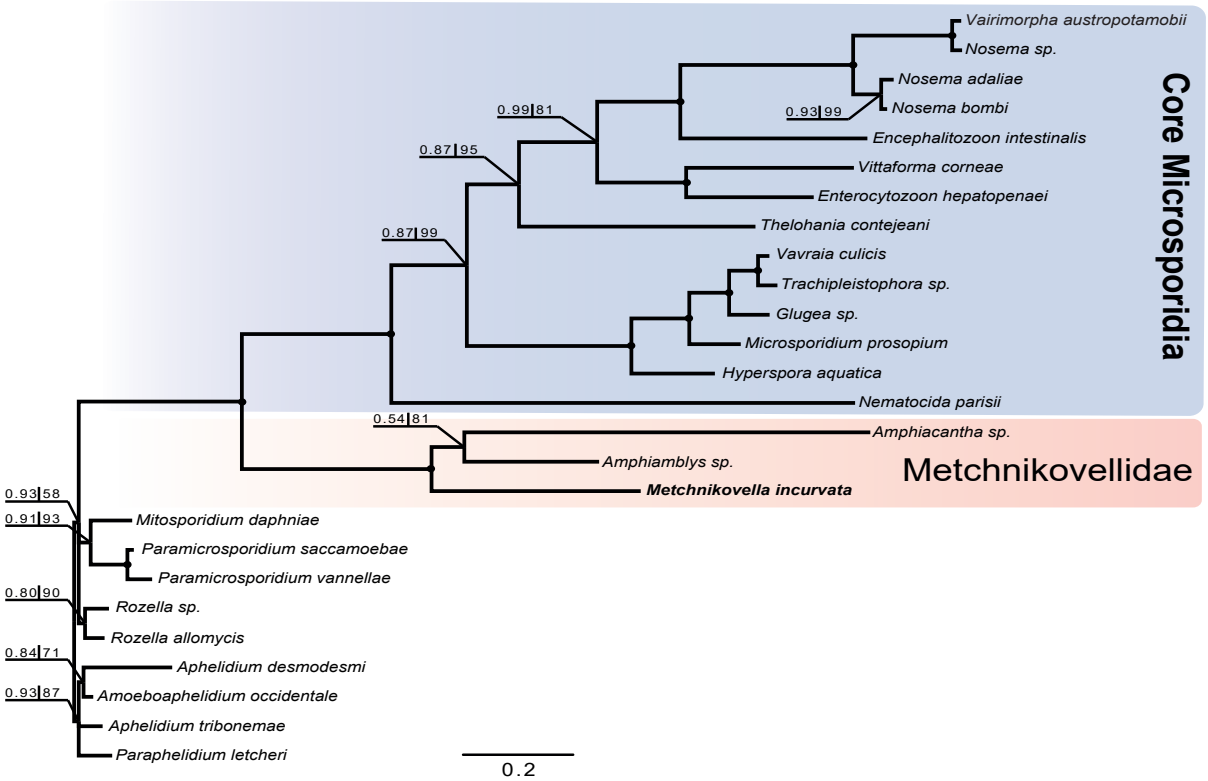


Fig. S5. Bayesian inference (BI) phylogenetic tree of 18S rRNA gene sequences including metchnikovellids and various representatives of the Holomycota clade. 26 species and 879 unambiguously aligned positions were used to reconstruct the tree. Split supports are posterior probabilities (pp) (on the left) and ML bootstrap (bs) values (on the right). Sequences obtained in this study are highlighted in black. Support values > 0.99 pp and > 95% bs are indicated with a black bullet.

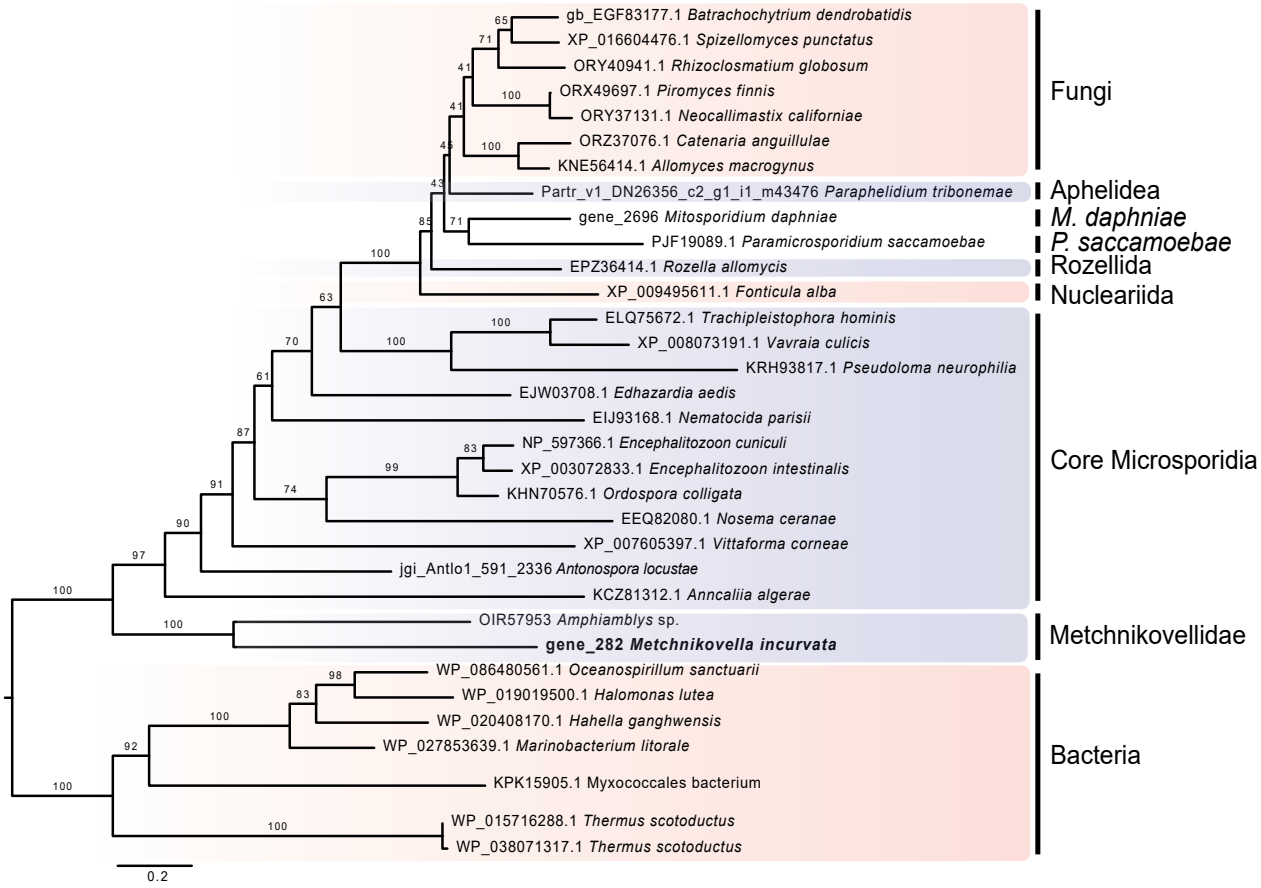


Fig. S6. Maximum Liklehood tree reconstructed for the cytosolic glycerol-3-phosphate dehydrogenase. The tree includes 34 protein sequences and was reconstructed with IQ-TREE under the LG+I+G4 evolutionary model. Sequences obtained in this study are highlighted in black. Bootstrap values are represented at nodes.

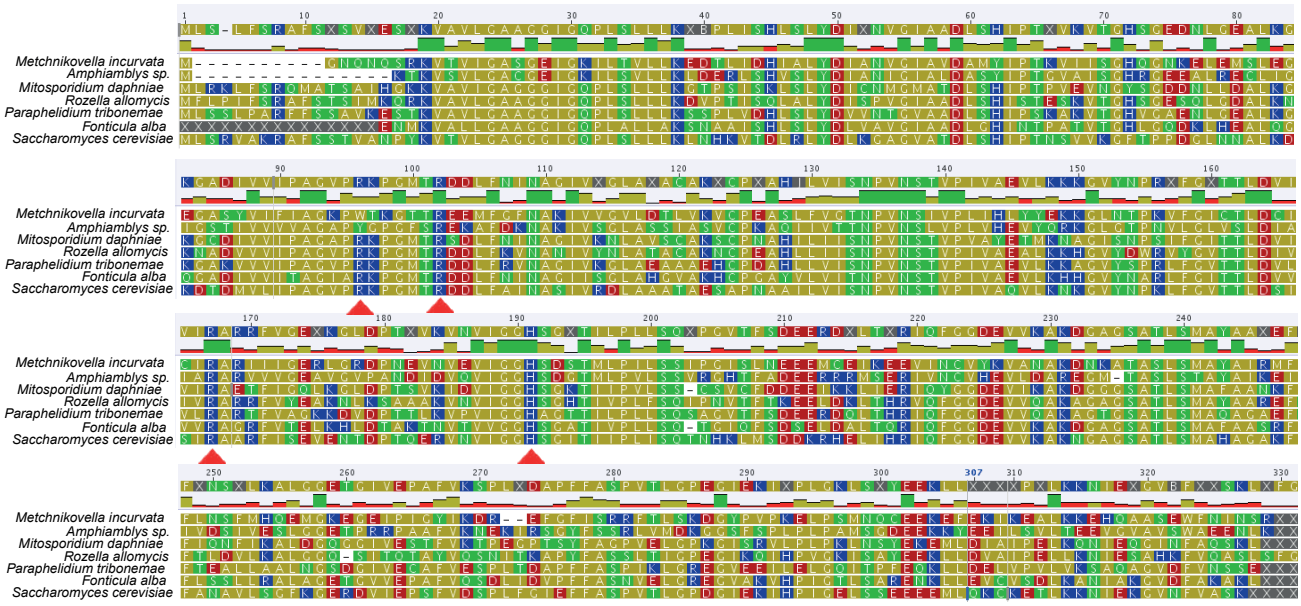


Fig. S7. Alignment of mMDH sequences including the sequence from *Metchnikovella incurvata* and *Amphiamblys* sp. The substrate-binding residues are indicated with a red triangle. Note that Arg (R), responsible for specificity, at position 96 in the alignment is replaced with Trp (W) in *Metchnikovella incurvata*.

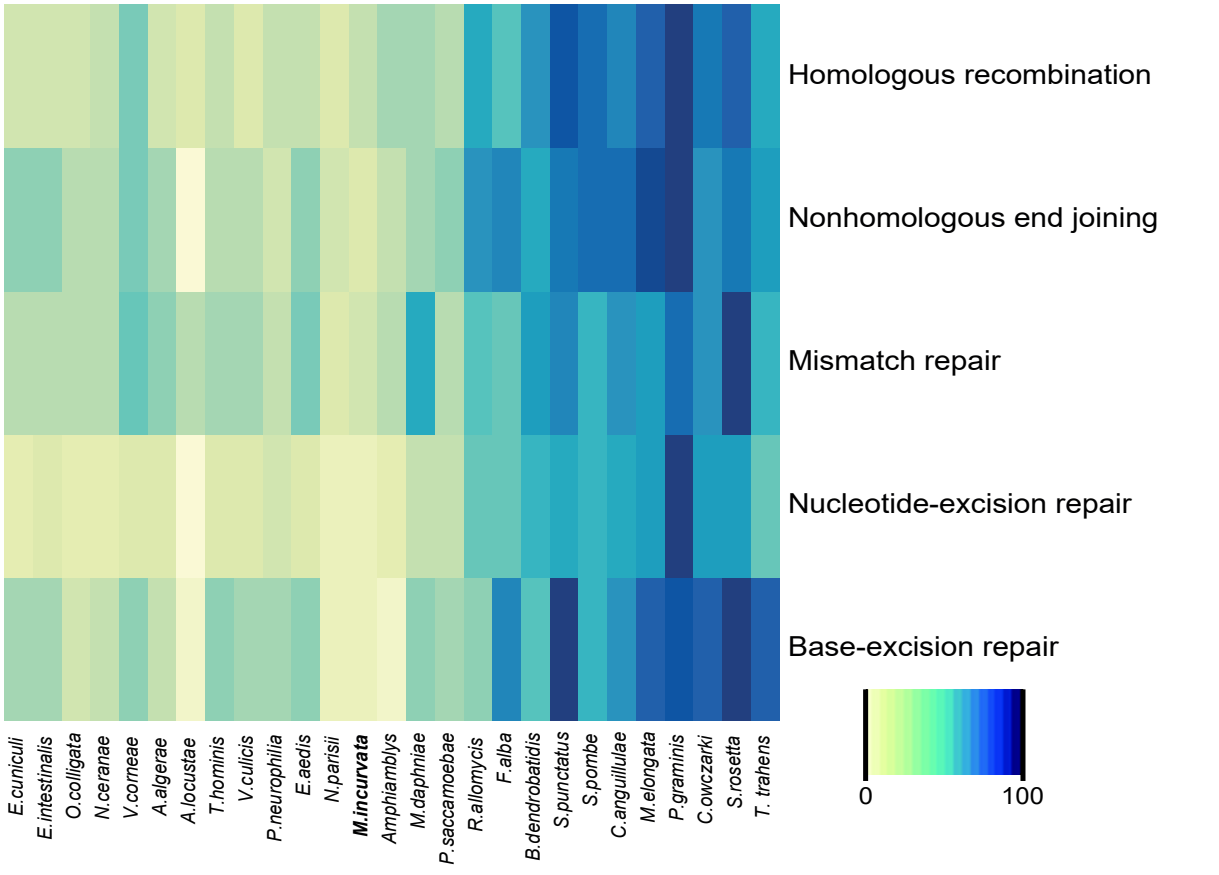


Fig. S8. Heatmap of DNA repair pathways in Microsporidia and other opisthokont lineages. Genes were identified based on 5 Gene Ontology terms (GO) from EggNOG and 25 opisthokont representative proteomes. The genome sequence obtained in this study is highlighted in black. Colors indicate the percentage of annotated genes within a GO term. Detailed gene numbers are provided in Supplementary Table S6.

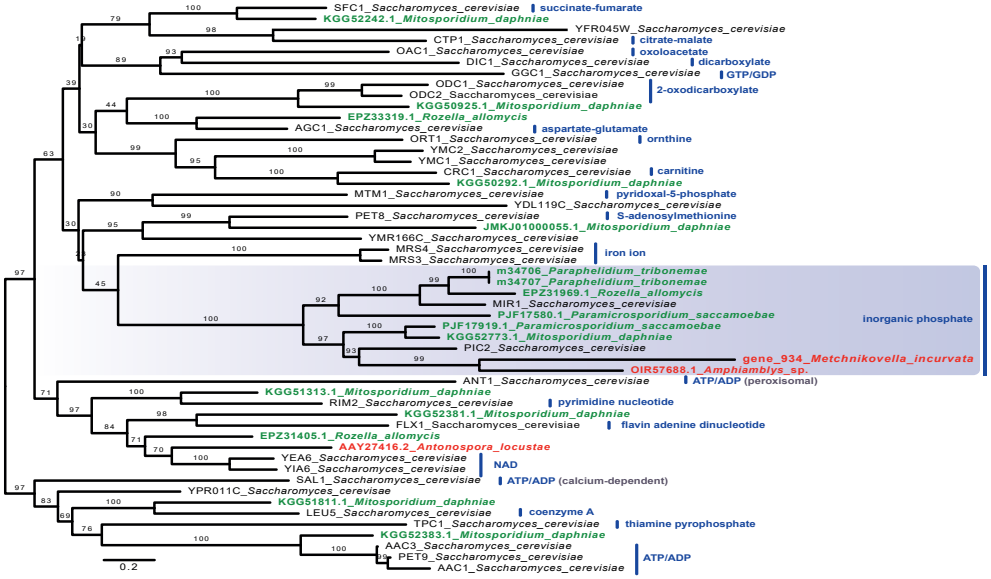


Fig. S9. ML tree for mitochondrial carrier proteins (MCF). The tree includes 51 protein sequences and was reconstructed with IQ-TREE v1.6.2 based on the model made by Mikhailov et al. (2017) under the LG+F+I+G4 evolutionary model. Bootstrap values are indicated at nodes. The specificity of the transporter is next to the names in blue. *Saccharomyces cerevisiae* sequences are indicated in black. *Mitosporidium daphniae*, *Rozella allomycis*, *Paramicrosporidium saccamoebae* and *Paraphelidium tribonemae* sequences are indicated in green. Microsporidia sequences are indicated in red.

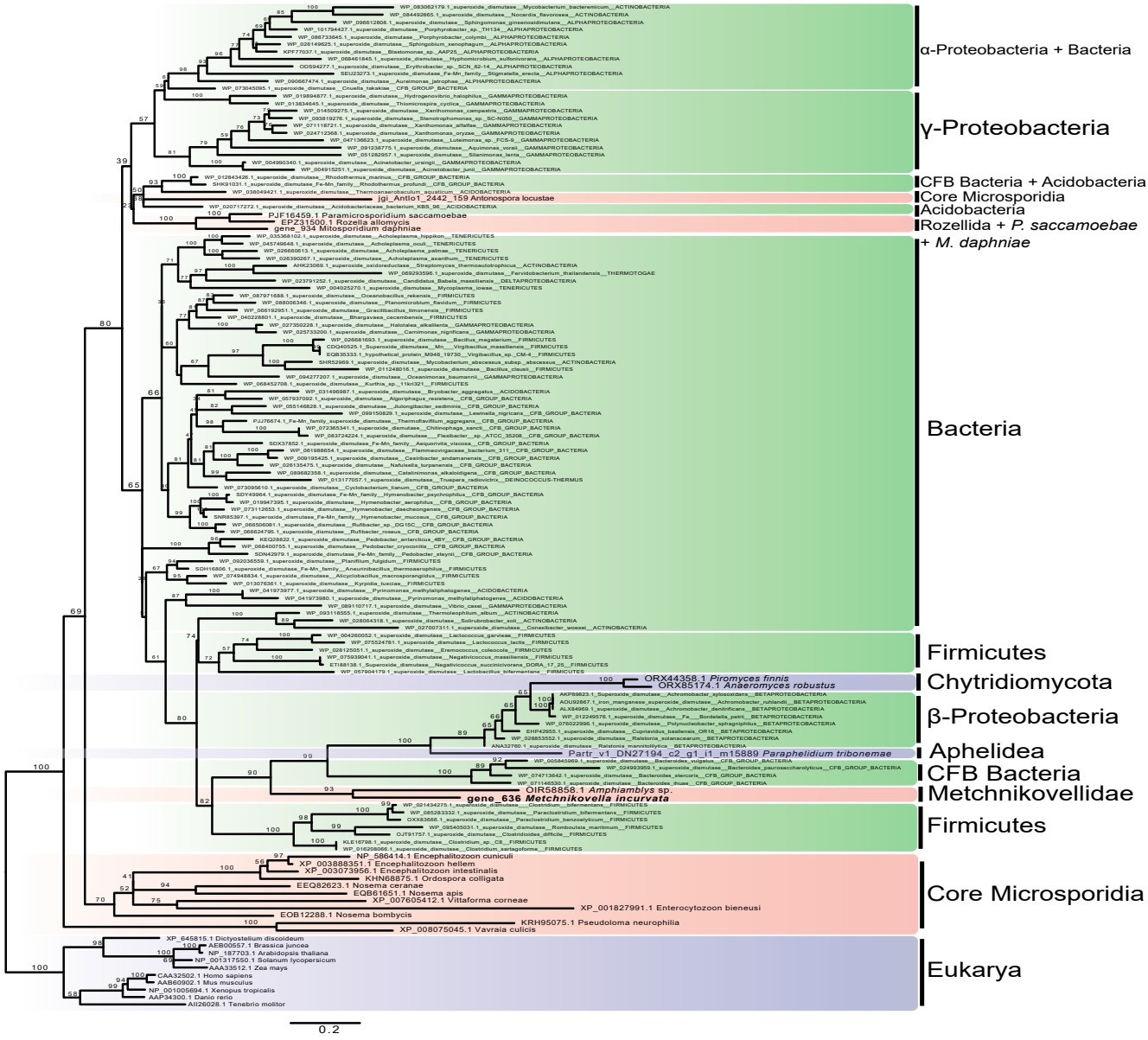


Fig. S10. ML phylogenetic tree for the mitochondrial MnSOD protein. The tree includes 136 protein sequences and was reconstructed with IQ-TREE under the LG+I+G4 evolutionary model. Highlights in green represent bacterial sequences, in blue represent eukaryotic sequences and in red microsporidian + rozellids sequences. Sequences obtained in this study are highlighted in black. Bootstrap values are indicated at nodes.

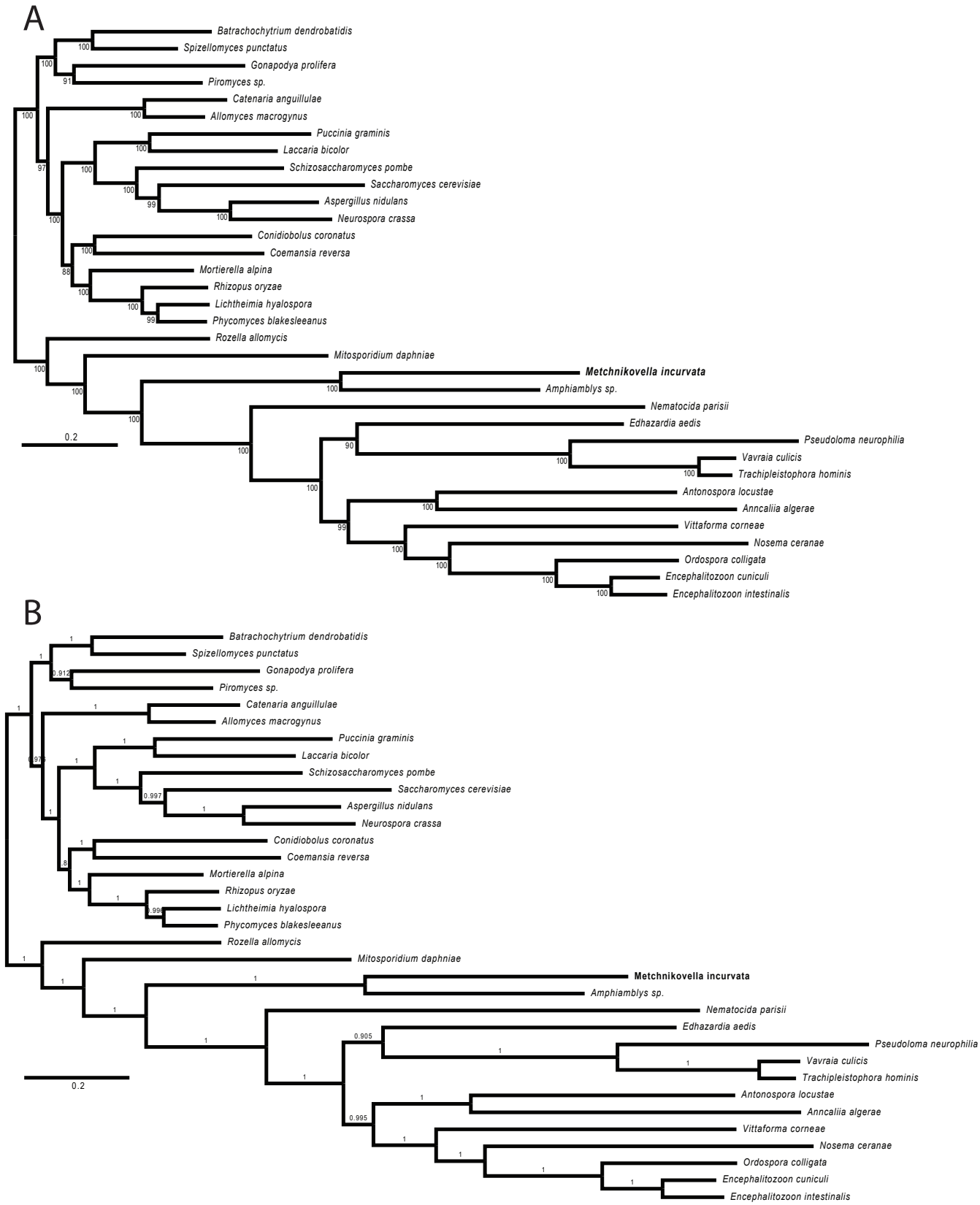


Fig. S11. Phylogenomic tree of 56 single-copy protein domain concatenated for a total of 16,272 conserved amino acidic positions for 32 representatives of the Holomycota clade and 5 other Amorphea species as an outgroup (2 Holozoa, 1 Apusomonadida, and 2 Amoebozoa). Maximum Likelihood tree (A) was reconstructed using IQ-TREE under the LG+F+I+G4 model and ultrafast bootstrap as statistical supportBayesian inference tree (B) was inferred using PhyloBayes with the CAT-Poisson model with posterior probability as statistical support. over the branches.

Table S1. Genome completeness analysis estimated with BUSCO

BUSCO Dataset			
Microsporidia			
	M.incurvata	*Amphiamblys* sp.	
	179	190	Complete BUSCOs (C)
	173	182	Complete and single-copy BUSCOs (S)
	6	8	Complete and duplicated BUSCOs (D)
	41	53	Fragmented BUSCOs (F)
	298	275	Missing BUSCOs (M)
	518	518	Total BUSCO groups searched
	42.47	**46.91**	%
Fungi			
	202	203	Complete BUSCOs (C)
	196	197	Complete and single-copy BUSCOs (S)
	6	6	Complete and duplicated BUSCOs (D)
	31	39	Fragmented BUSCOs (F)
	57	48	Missing BUSCOs (M)
	290	290	Total BUSCO groups searched
	80.34	**83.44**	%

Table S2. Repetitive elements analysis with Repeat Masker

	% percentage of repetitive sequences								number of repetitive elements								length occupied (bp)							
	R. allomycis	*P.saccamoebae*	*M.daphniae*	*M.incurvata*	*Amphiamblys sp.*	*N.parisii (ERTm1)*	*T.hominis*	*E.cuniculis*	*R. allomycis*	*P.saccamoebae*	*M.daphniae*	*M.incurvata*	*Amphiamblys sp.*	*N.parisii (ERTm1)*	*T.hominis*	*E.cuniculis*	*R. allomycis*	*saccamoeba*	*M.daphniae*	*M.incurvata*	*mphiamblys s*	*parisii (ERTm*	*T.hominis*	*E.cuniculis*
SINEs	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
ALUs	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
MIRs	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
LINEs	0.04	0	0	0.37	2.27	0	0.61	0	41	0	0	92	328	0	146	0	5118	0	0	22451	117778	0	51417	0
LINE1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
LINE2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
L3/CR1	0	0	0	0	0,02	0	0	0	0	0	0	0	16	0	0	0	0	0	0	0	1284	0	0	0
LTR elements	0.51	0	0	0	0.53	0	0.29	0	255	0	0	0	37	0	45	0	60293	0	0	0	27312	0	24886	0
ERVL	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
ERVL-MaLRs	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
ERV classI	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
ERV classII	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
DNA elements	0	0	0	0.53	0.23	0	0	0	0	0	0	96	75	0	0	0	0	0	0	31624	11938	0	0	0
hAT-Charlie	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
TcMar-Tigger	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Unclassified	1.94	5.44	3.14	15.3	28.7	9.28	5.77	10.61	1828	1644	410	6630	5415	1094	3525	148	230409	390361	177178	921051	1486402	377916	490702	264865
Total interspersed repeats	2.49	5.44	3.14	16.2	31.74	9.28	6.67	10.61	x	x	x	x	x	x	x	x	295820	390361	177178	0	1643430	377916	567005	264865
Small RNA	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Satellites	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Simple repeats	0.54	0.42	0.42	1.14	0.78	0.99	0.64	0.13	1374	601	575	1071	386	729	903	75	63819	30399	23900	68596	40388	40372	54027	3196
Low complexity	0.2	0.05	0.12	0.2	0.07	0.31	0.12	0.05	476	67	141	214	81	233	213	25	23946	3407	6776	11922	3871	12793	10553	1270

Table S3. Core carbon metabolism for Microsporidia and basal relatives. The illustrated pathways include Glycolysis (blue), Pentose phosphate pathway (yellow), Trehalose metabolism (red), Tricarboxylic acid

Symbol	Name	EC number	*E.cuniculi*	*N.parisii*	*M.incurvata*	*Amphiambly*	*M.daphniae*	*P.saccamoe*	*R.allomycis*
PGM	phosphoglucomutase	5.4.2.2	X		X	X	X	X	X
UGP	UTP--glucose-1-phosphate uridylyltransferase	2.7.7.9	X	X	X	X	X	X	X
TPS	trehalose 6-phosphate synthase	2.4.1.15	X	X		X	X	X	X
TPP	trehalose 6-phosphate phosphatase	3.1.3.12	X	X	X	X	X	X	X
TREH	trehalase	3.2.1.28	X	X	X	X	X	X	X
G6PD	glucose-6-phosphate dehydrogenase	1.1.1.49	X	X	X	X	X	X	X
PGL	6-phosphogluconolactonase	3.1.1.31	X	X	X	X	X	X	X
PGD	6-phosphogluconic carboxylase (Phosphogluconate	1.1.1.44	X	X	X	X	X	X	X
RPI	ribose-5-phosphate isomerase	5.3.1.6	X	X	X	X	X	X	X
RPE	ribulose-phosphate 3-epimerase	5.1.3.1	X	X		X	X	X	X
TKL	transketolase	2.2.1.1	X	X	X	X	X	X	X
TALDO	transaldolase	2.2.1.2					X	X	X
HK	hexokinase	2.7.1.1	X	X	X	X	X	X	X
PGI	glucose-6-phosphate isomerase (phosphoglucose i	5.3.1.9	X	X		X	X	X	X
PFK	6-phosphofructokinase	2.7.1.11	X	X	X	X	X	X	X
ALDO	fructose-bisphosphate aldolase	4.1.2.13	X	X	X	X	X	X	X
TPI	triose-phosphate isomerase	5.3.1.1	X	X	X	X	X	X	X
GAPDH	glyceraldehyde 3-phosphate dehydrogenase	1.2.1.12	X	X	X	X	X	X	X
PGK	phosphoglycerate kinase	2.7.2.3	X	X	X	X	X	X	X
PGAM	phosphoglycerate mutase	5.4.2.12	X	X		X	X	X	X
ENO	phosphopyruvate hydratase (Enolase)	4.2.1.11	X	X	X	X	X	X	X
PK	pyruvate kinase	2.7.1.40	X	X	X	X	X	X	X
cGPD	cytosolic glycerol-3-phosphate dehydrogenase	1.1.1.8	X	X	X	X	X	X	X
mGPD	mitochondrial glycerol-3-phosphate dehydrogenase	1.1.5.3	X	X			X	X	X
AOX	alternative oxidase	1.10.3.11		X			X		X
MPC	mitochondrial pyruvate carrier						X		X
PDC	pyruvate dehydrogenase	1.2.4.1	X	X			X		X
	dihydrolipoyl transacetylase	2.3.1.12					X		X
	dihydrolipoyl dehydrogenase	1.8.1.4					X	X	X
CS	citrate synthase	2.3.3.1				X	X		X
ACO	aconitase	4.2.1.3					X	X	X
IDH	isocitrate dehydrogenase	1.1.1.41					X		X
OGDC	oxoglutarate dehydrogenase	1.2.4.2					X	X	X
	dihydrolipoyl transsuccinylase	2.3.1.61					X		X
	dihydrolipoyl dehydrogenase	1.8.1.4					X	X	X
SCS	succinyl-CoA synthetase	6.2.1.4					X		X
SQR	succinate dehydrogenase	1.3.5.1					X	X	X
FH	fumarate hydratase (Fumarase)	4.2.1.2					X	X	X
MDH	malate dehydrogenase	1.1.1.37			X	X	X	X	X

Table S4. Core fatty acid metabolic pathway by Kegg. Numbers in the green cases represent the numbers of genes in the genome represented by a GO term.

Go term	*E.cuniculi*	*E.intestinalis*	*O.colligata*	*N.ceranae*	*V.corneae*	*A.algerae*	*A.locustae*	*T.hominis*	*V.culicis*	*P.neurophilia*	*E.aedis*	*N.parisii*	*M.incurvata*	*Amphiamblys*	*M.daphniae*	*P.saccamoeb*	*R.allomycis*	*B.dendrobati*	*P.gramis*
GO:0004022															2	2	1	2	2
GO:0003857															3	2	2	3	5
GO:0003985	1	1		1		1	1				1				1	5	2	4	5
GO:0003988	1	1		1		1	1				1				1	5	3	5	4
GO:0003997		1				1					1				2	3	3	2	3
GO:0004029															2	4	2	6	8
GO:0004085															2		2	2	2
GO:0004095																			1
GO:0004165																			
GO:0004300																			
GO:0004361															1	1		1	1
GO:0004466																			
GO:0004467	2	1	1	1	1	4	1	2	2	2	5	1	1	1	4	1	3	5	8
GO:0004497	2	2	1	2	2	4	1	2	2	4	3	3	3	3	10	7	10	18	23
GO:0008692																			
GO:0008860																			
GO:0008922																			
GO:0015044																			
GO:0015045																			
GO:0016508															1	1	1	1	2
GO:0016509																			
GO:0017099																		1	1
GO:0018685						1												3	
GO:0047113																			
GO:0047645																			
GO:0047948																			
GO:0050060																			
GO:0050061																		1	
GO:0070991																			

Table S5. Conserved main endocytic components in the *Metchnikovella incurvata* proteome

Name	PFAM accession		*Absent*	*Present*	Gene	Comments
Clathrin-H-link	PF13838			X	gene_1846	
Clathrin heavy chain	PF00637			X	gene_1846	
ANTH domain protein	PF07651			X	gene_1422	
cytoskeleton assembly control protein Sla2	PF01608			X	gene_1422	
WH1 domain-containing protein	PF00568			X	gene_471	fragment
Actin-binding WH2	PF02205		X		no hit	
Rho-binding domain-domain-containing protein	PF00786			X	gene_471, gene_1003	fragment
SLA1 homology domain 1, SHD1	PF03983			X	gene_41	fragment
ENTH domain-containing protein	PF01417			X	gene_1880	
ARP2/3 complex subunit 21kDa	PF04045			X	gene_34	fragment
ARP2/3 complex subunit ARPC3	PF04062		X		no hit	
ARP2/3 complex subunit ARPC4	PF05856			X	gene_1433	fragment

Table S6. Number of genes involved in 5 diferent DNA repair pathways in 27 Opisthokonta species

Related Function	GO term	*E.cuniculi*	*E.intestinali*	*O.colligata*	*N.ceranae*	*V.corneae*	*A.algerae*	*A.locustae*	*T.hominis*	*V.culicis*	*P.neurophil*	*E.aedis*	*N.parisii*	*M.incurvata*	*Amphiamblys*	*Mitosporidi*	*P.saccamoe*	*R.allomycis*	*F.alba*	*B.dendroba*	*S.punctatus*	*S.pombe*	*C.anguillula*	*M.elongata*	*P.graminis*	*C.owczarki*	*S.rosetta*	*T. trahens*	*#genes/H.sa*
Base-excision repair	GO:0006284	20	20	17	18	21	18	11	21	20	20	21	13	**13**	**12**	22	20	22	33	25	41	28	31	37	38	37	40	37	54
Nucleotide-excision repair	GO:0006289	47	48	47	44	49	49	31	51	51	55	51	40	**41**	**46**	58	56	80	77	91	93	92	94	100	134	98	98	78	126
Mismatch repair	GO:0006298	17	17	17	17	22	20	18	19	19	16	21	14	**15**	**18**	27	18	23	22	28	30	26	29	28	33	29	38	26	47
Nonhomologous end joining	GO:0006303	16	16	14	14	17	15	8	14	14	12	16	12	**11**	**13**	15	16	24	25	22	26	27	27	30	31	24	26	23	76
Homologous recombination	GO:0035825	30	29	29	31	41	29	27	33	28	31	32	28	**32**	**36**	36	34	52	45	58	69	63	59	67	74	61	66	52	60

Table S7. Presence/absence of 15 HGTs identified in the literature for 18 genomes of core Microsporidia and related early-branching lineages. Black "X" represent HGTs found from bacterial origin. Red "X"

Gene	first cited	*E.hallem*	*E.intestinalis*	*E.cuniculi*	*O.colligata*	*No.bombycis*	*No.ceranae*	*Ann.algerae*	*Ant.locustae*	*T.hominis*	*V.culicis*	*P.neurophilia*	*E.aedis*	*N.parisii*	*Amphiamblys*	*M.incurvata*	*M.daphniae*	*P.saccamoebae*	*R.allomycis*
Cytidylate kinase	*E. cuniculi* Marcet-Houben & Gabaldón (2010); Cuomo et al. (2012)	X	X	X	X	X	X	X	-	X	X	X	X	X	-	-	-	-	-
Hemolysin III-like protein	*E. cuniculi* Marcet-Houben & Gabaldón (2010)	X	X	X	X	X	X	X	X	X	X	X	X	X	-	-	-	X	X
Phosphoribosyltransferase	*E. hallem* Pombert et al. (2012)	X	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
GTP cyclohydrolase	*E. hallem* Pombert et al. (2012)	X	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	X
PNPs	*E. hallem* Pombert et al. (2012)	X	-	-	-	-	-	-	-	-	-	-	-	-	-	-	X	X	X
Folylpolyglutamate synthase	*E.hallem* Pombert et al. (2012)	X	-	-	-	-	-	-	-	-	-	-	-	-	-	-	X	X	X
ADP/ATP translocase PF03219	E.cuniculi Tsaousis et al. (2008)	X	X	X	X	X	X	X	X	X	X	X	X	X	-	-	-	-	X
Aspartate-ammonia ligase	*No. ceranae* Heinz et al. (2012)	-	-	-	-	-	X	-	-	X	X	X	-	-	X	-	-	-	-
Cysteine-rich secretory proteins, antigen 5, and pathogenesis-related 1 protein	*E. cuniculi,N. parisii* Nakjang et al. (2013)	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Folic acid synthase	*E. hallem* Pombert et al. (2012)	X	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	X
Dihydrofolate synthase	*E. hallem* Pombert et al. (2012)	X	-	-	-	-	-	-	-	-	-	-	-	-	-	-	X	X	-
Thymidine kinase	Alexander et al. (2016)	X	X	X	X	X	X	X	X	-	-	-	-	-	-	-	X	-	X
Manganese superoxide dismutase	*N. bombycis* (Xiang et al. 2010)	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
Catalase	*A.locustae* Fast et al. (2003)	-	-	-	-	-	-	-	X	-	-	-	-	-	-	-	X	X	-
Purine nucleoside phosphorylase	*E. hallem* Selman et al. (2011)	X	-	-	-	-	-	-	-	-	-	-	-	-	-	-	X	X	X

Table S9. Orthogroup statistics.

Number of genes	330609			
Number of genes in orthogroups	212244			
Number of unassigned genes	118365			
Percentage of genes in orthogroups	64.2			
Percentage of unassigned genes	35.8			
Number of orthogroups	12448			
Number of species-specific orthogroups	792			
Number of genes in species-specific orthogroups	5717			
Percentage of genes in species-specific orthogroups	1.7			
Mean orthogroup size	17.1			
Median orthogroup size	6			
G50 (assigned genes)	39			
G50 (all genes)	16			
O50 (assigned genes)	1170			
O50 (all genes)	3547			
Number of orthogroups with all species present	175			
Number of single-copy orthogroups	0			
Average number of genes per-species in orthogroup	Number of orthogroups	Percentage of orthogroups	Number of genes	Percentage of genes
<1	11120	89.3	100181	47.2
'1	923	7.4	44057	20.8
'2	194	1.6	17115	8.1
'3	87	0.7	11079	5.2
'4	29	0.2	4609	2.2
'5	19	0.2	3881	1.8
'6	17	0.1	4097	1.9
'7	16	0.1	4494	2.1
'8	4	0	1195	0.6
'9	7	0.1	2418	1.1
'10	5	0	1920	0.9
11-15	12	0.1	5802	2.7
16-20	10	0.1	6461	3
21-50	5	0	4935	2.3
51-100	0	0	0	0
101-150	0	0	0	0
151-200	0	0	0	0
201-500	0	0	0	0
501-1000	0	0	0	0
'1001+	0	0	0	0
Number of species in orthogroup	Number of orthogroups			
1	792			
2	4174			
3	1194			
4	682			
5	498			
6	366			
7	323			
8	347			
9	249			
10	236			
11	250			
12	238			
13	219			
14	196			
15	181			
16	197			
17	211			
18	242			
19	212			
20	228			
21	205			
22	139			
23	92			
24	54			
25	39			
26	33			
27	37			
28	25			
29	38			
30	51			
31	41			
32	52			
33	70			
34	98			
35	102			
36	162			
37	175			

Table S11. Metchnikovellids exclusive orthogroups that returned a hit in the HMMER database. In red are those sequences in a orthogroup that have not been found in the database.

Ortholog	genes	name	function
OG0008667	gene_1551_Metchnikovella_incurvata, gene_1890_Metchnikovella_incurvata, gene_1_Amphiamblys_sp._1008	Kinase	Cyclin-Dependent Kinase
OG0010229	gene_1272_Metchnikovella_incurvata, gene_6_Amphiamblys_sp._90	CAMK4	calcium calmodulin-dependent protein kinase IV
OG0010228	gene_1273_Metchnikovella_incurvata, gene_6_Amphiamblys_sp._89		serine threonine-protein kinase
OG0010185	gene_1292_Metchnikovella_incurvata, gene_2_Amphiamblys_sp._20		Filamin C, gamma
OG0010361	gene_1495_Metchnikovella_incurvata, gene_2_Amphiamblys_sp._741		Helicase 1
OG0008670	gene_169_Metchnikovella_incurvata, gene_2_Amphiamblys_sp._171, gene_2_Amphiamblys_sp._811	AP2A2	Adaptor-related protein complex 2, alpha
OG0010363	gene_1951_Metchnikovella_incurvata, gene_2_Amphiamblys_sp._836		vacuolar import and degradation protein
OG0010263	gene_1_Amphiamblys_sp._168, gene_2128_Metchnikovella_incurvata	Hsp70 protein	Chaperone
OG0010181	gene_1_Amphiamblys_sp._47, gene_251_Metchnikovella_incurvata		(ABC) transporter
OG0007496	gene_2_Amphiamblys_sp._399, gene_2_Amphiamblys_sp._469, gene_2_Amphiamblys_sp._604, gene_365_Metchnikovella_in		Ceramidase
OG0010364	gene_2_Amphiamblys_sp._939, gene_37_Metchnikovella_incurvata		Transmembrane amino acid transporter protein
OG0010331	gene_1_Amphiamblys_sp._1019, gene_472_Metchnikovella_incurvata		transforming growth factor beta regulator
OG0010254	gene_516_Metchnikovella_incurvata, gene_9_Amphiamblys_sp._50		ABC transporter transmembrane region
OG0010179	gene_578_Metchnikovella_incurvata, gene_9_Amphiamblys_sp._10	GDI1	rab gdp-dissociation inhibitor
OG0010256	gene_1158_Metchnikovella_incurvata, gene_9_Amphiamblys_sp._80	ORC5	Origin recognition complex subunit
OG0008671	gene_2642_Metchnikovella_incurvata, gene_2646_Metchnikovella_incurvata, gene_2_Amphiamblys_sp._242		Inherit from KOG: enhancer of polycomb homolog
OG0010223	gene_1568_Metchnikovella_incurvata, gene_6_Amphiamblys_sp._15	FANCD2	Fanconi anemia complementation group d2

Tables S8, S10 and S12 and all tables in high quality can be found in:
https://figshare.com/authors/Luis_Javier_Galindo/6432803
https://academic.oup.com/gbe/article/10/10/2736/5098297

9.3. Supplementary material of manuscript 3

A new fungal clade helps reconstructing the tree of Fungi and the evolution of the flagellum in Holomycota

(Manuscript in preparation)

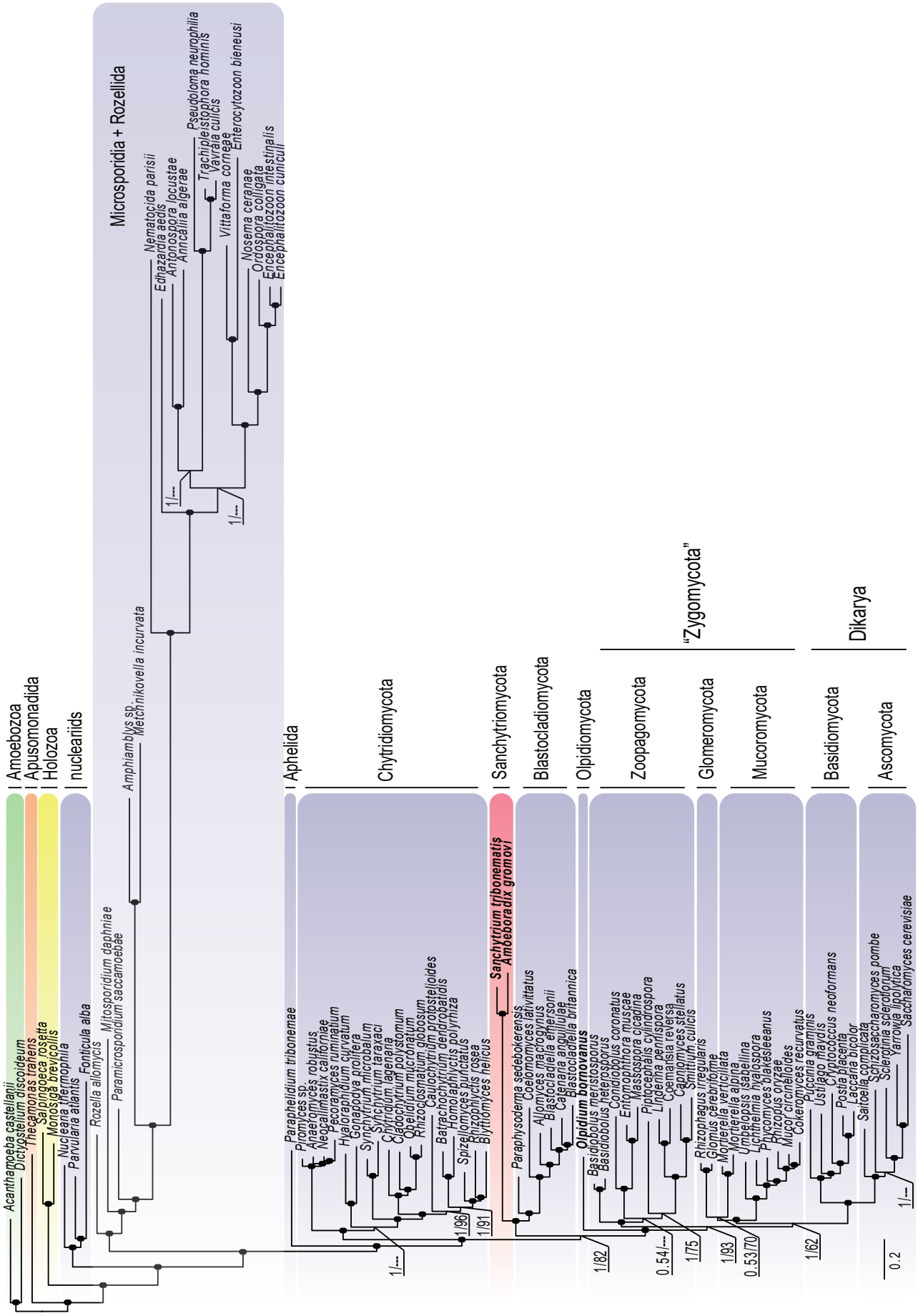
Luis Javier Galindo[1], Guifré Torruella[1], Purificación López-García[1], Sergey Karpov[2] and David Moreira[1]

[1]Ecologie Systématique Evolution, CNRS, Université Paris-Sud, AgroParisTech, Université Paris-Saclay, Orsay, France.

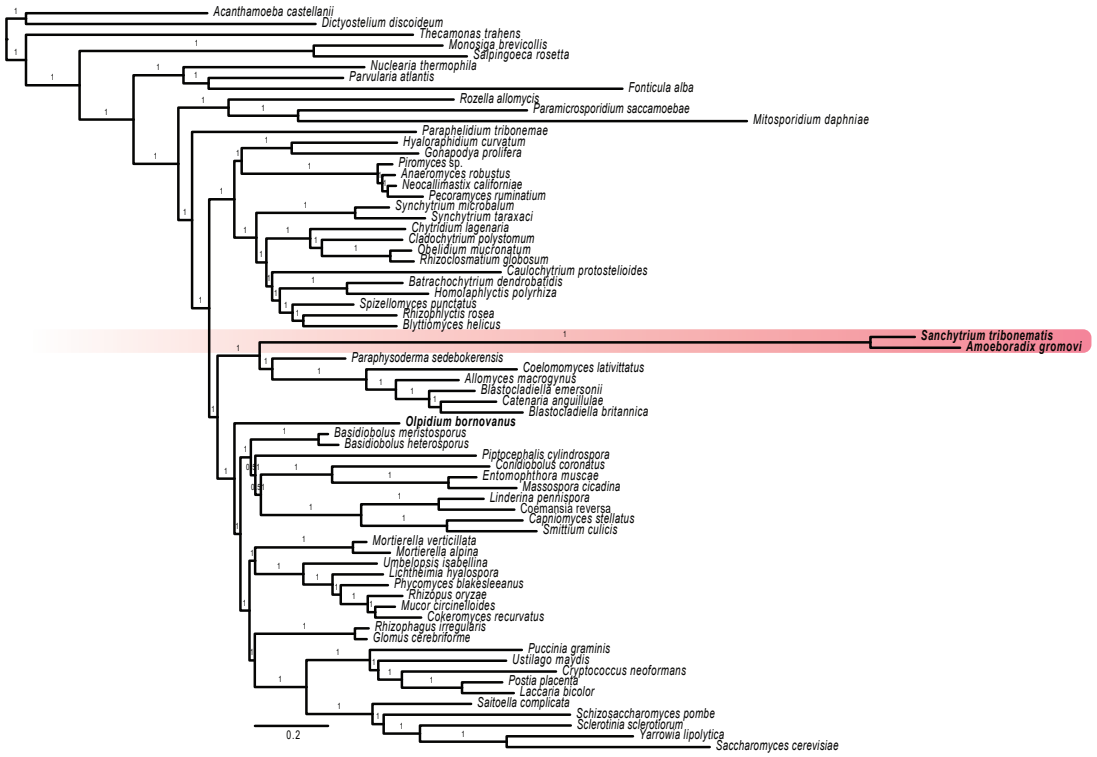
[2]Department of Invertebrate Zoology, Faculty of Biology, St Petersburg State University, Russia

* Corresponding authors

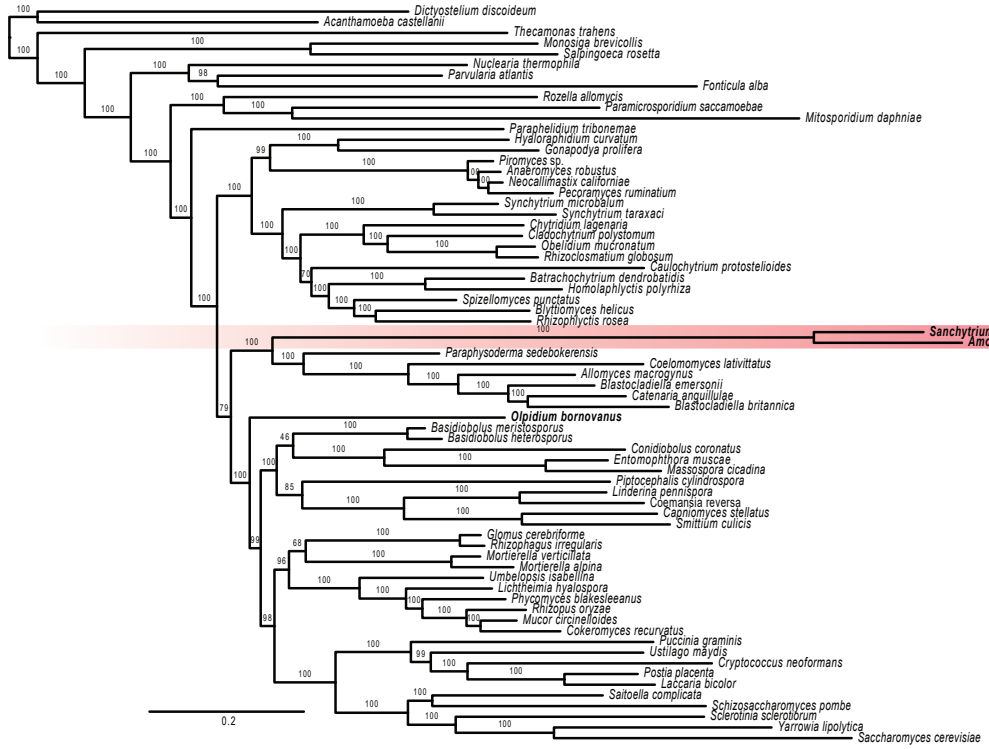
E-mails: david.moreira@u-psud.fr; luis.galindo@u-psud.fr.



Supplementary figure 1A. Bayesian inference phylogenomic tree based on the GBE protein dataset. The tree was reconstructed using 264 conserved proteins, 84 species, and 83,321 conserved amino acid positions with the LG+F+R10+PMSF model for ML (BI: CAT-Poisson). Branches with support values higher or equal to 0.99 pp and 99% ufbs are indicated by black dots.

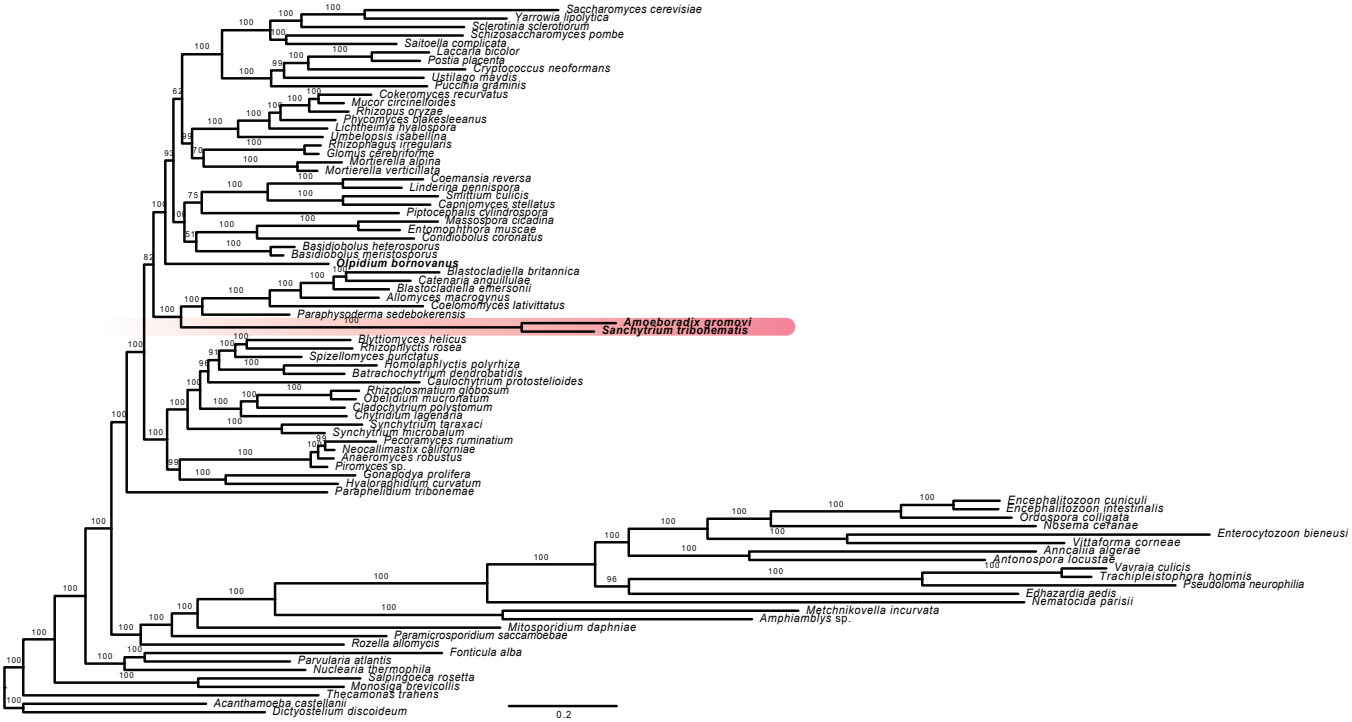


Supplementary Figure 1B. Bayesian phylogenomic tree based on the GBE protein dataset. The tree was reconstructed using 264 conserved proteins, 69 species, and 91,768 conserved amino acid positions, it was inferred using PhyloBayes under the CAT-Poisson model with posterior probability as statistical support.



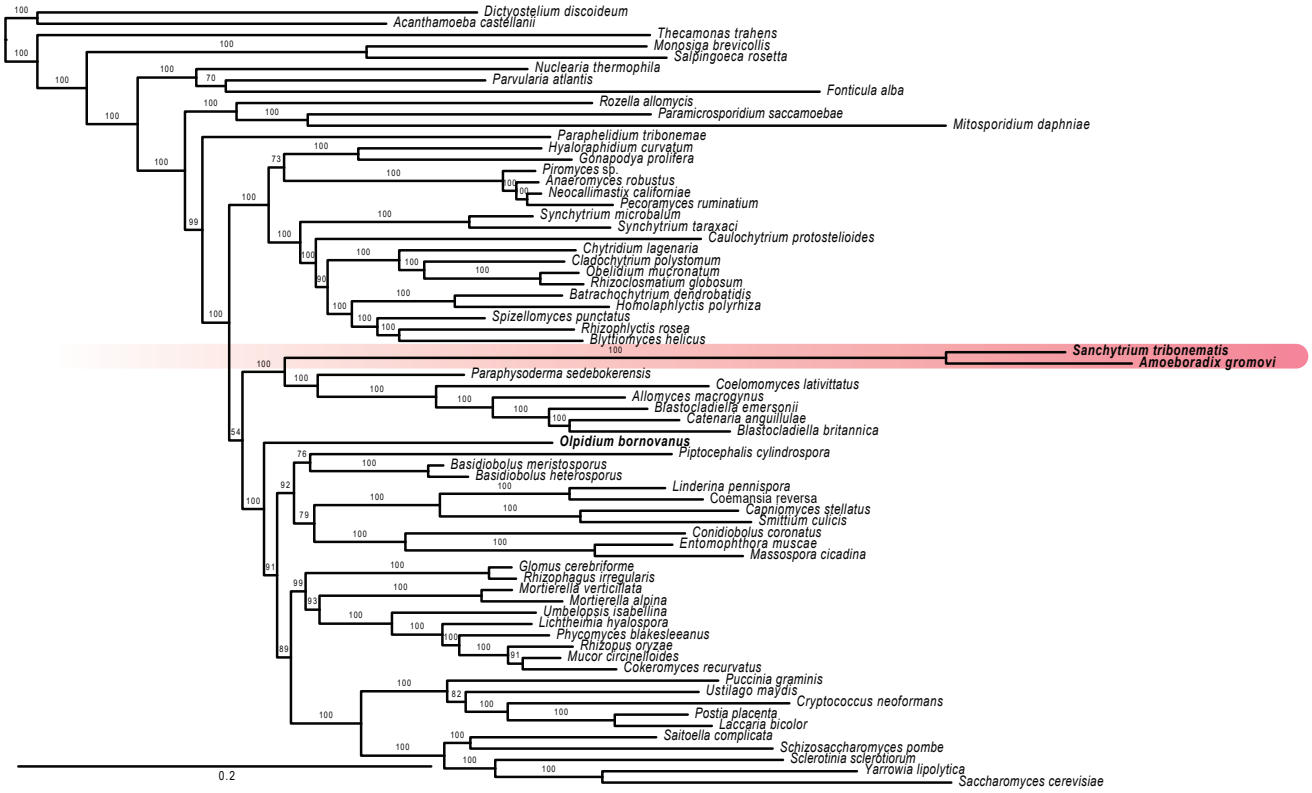
Supplementary Figure 1C. Maximum likelihood tree based on the GBE protein dataset. The tree was reconstructed using 264 conserved proteins, 69 species, and 91,768 conserved amino acid positions, it was inferred with IQ-TREE under the LG+R9+PMSF model and ultrafast bootstrap as statistical support.

Supplementary Figure 1D. Bayesian phylogenomic tree based on the GBE protein dataset. The tree was reconstructed using 264 conserved proteins, 84 species, and 83,321 conserved amino acid positions, it was inferred using PhyloBayes under the CAT-Poisson model with posterior probability as statistical support.



Supplementary Figure 1E. Maximum likelihood tree based on the GBE protein dataset. The tree was reconstructed using 264 conserved proteins, 84 species, and 83,321 conserved amino acid positions, it was inferred with IQ-TREE under the LG+F+R10+PMSF model and ultrafast bootstrap as statistical support.

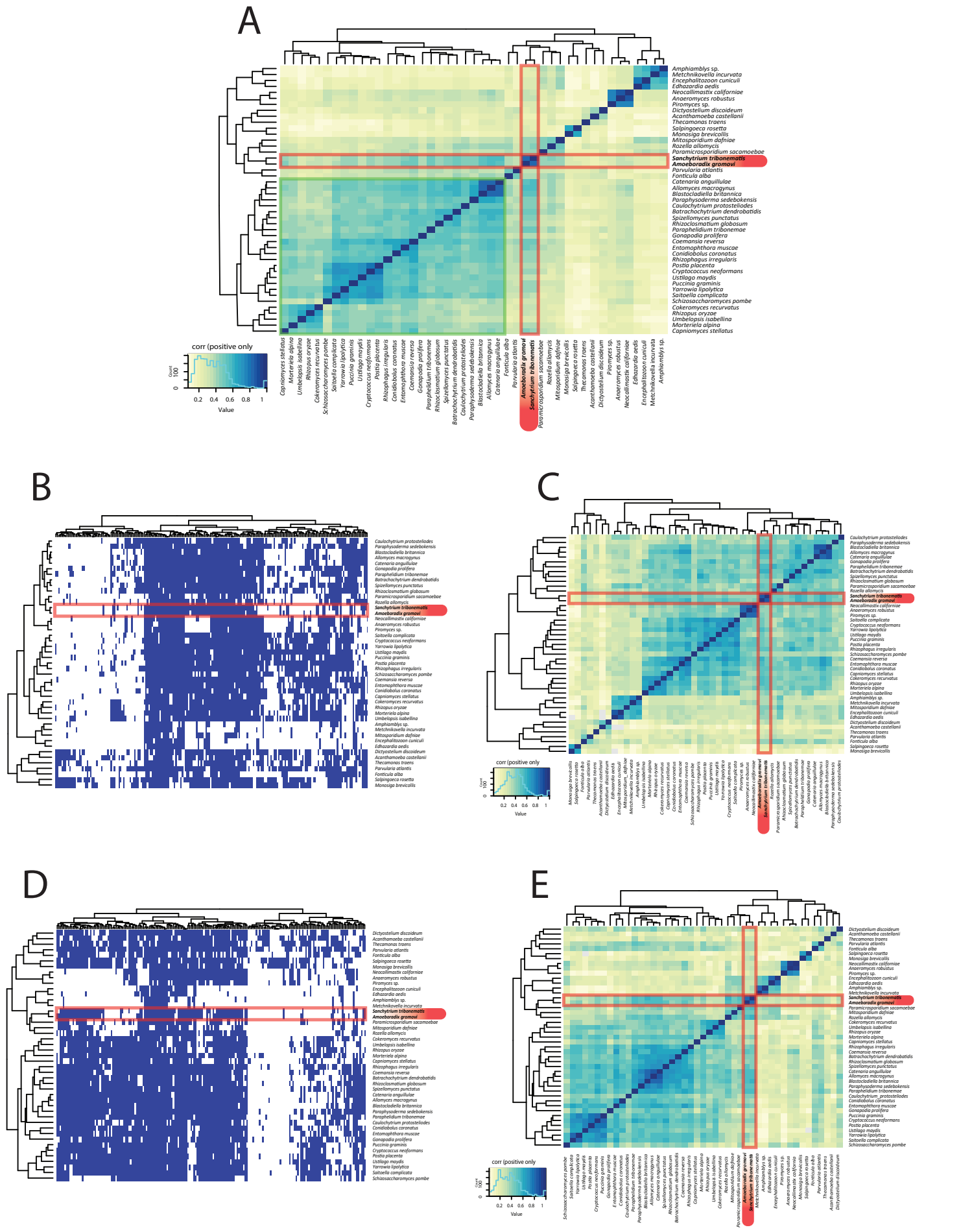
Supplementary Figure 2A. Recoded bayesian phylogenomic tree based on the GBE protein dataset. The tree was reconstructed using 264 nucleotide recoded conserved proteins, 69 species, and 91,768 recoded nucleotide positions, it was inferred using PhyloBayes under the CAT-Poisson model with posterior probability as statistical support.



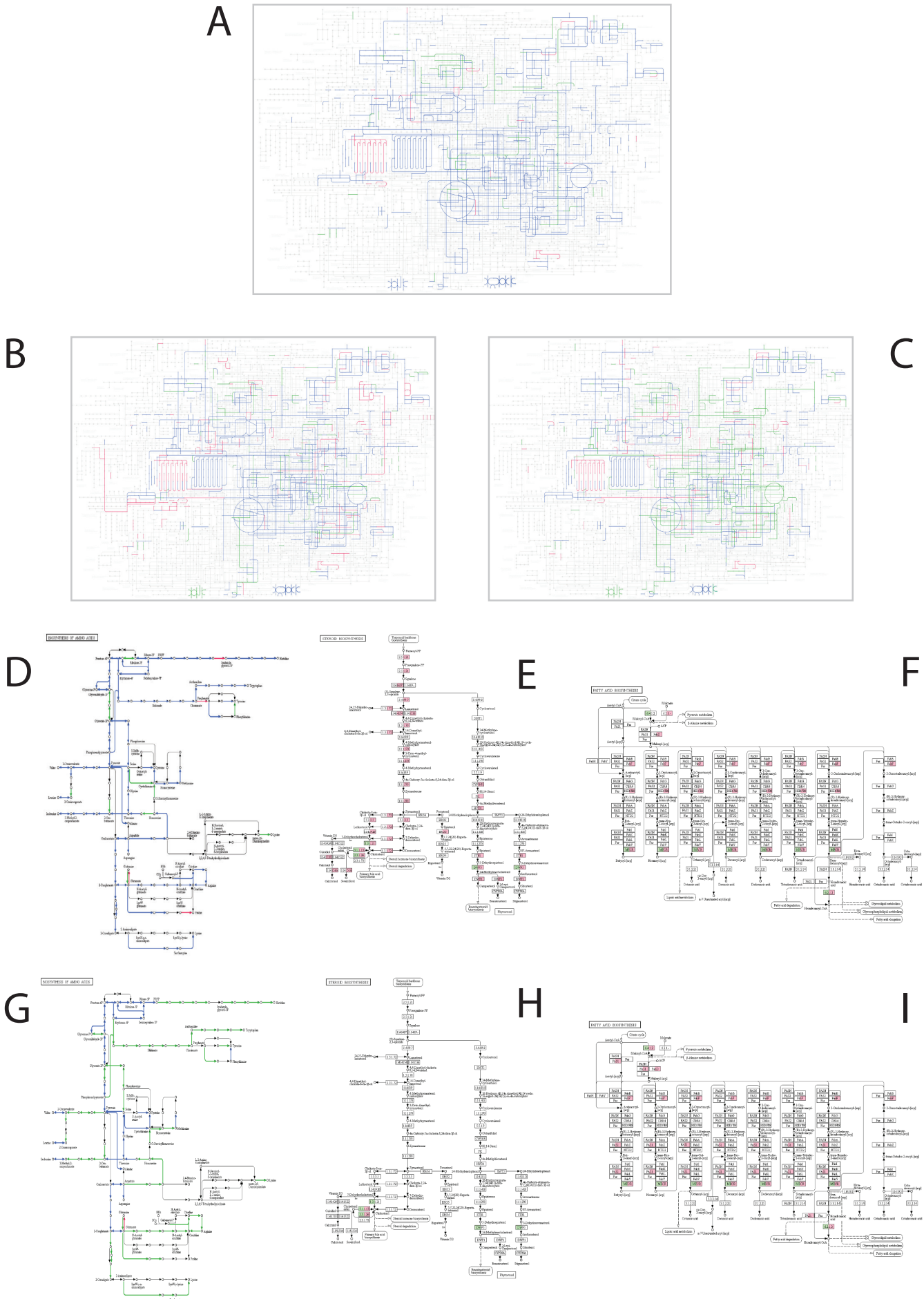
Supplementary Figure 2B. Recoded maximum likelihood tree based on the GBE protein dataset. The tree was reconstructed using 264 nucleotide recoded conserved proteins, 69 species, and 91,768 recoded nucleotide positions, it was inferred with IQ-TREE under the GTR+F+I+G4 model and ultrafast bootstrap as statistical support.

Supplementary Figure 2C. Recoded bayesian phylogenomic tree based on the GBE protein dataset. The tree was reconstructed using 264 conserved proteins, 84 species, and 83,321 recoded nucleotide positions, it was inferred using PhyloBayes under the CAT-Poisson model with posterior probability as statistical support.

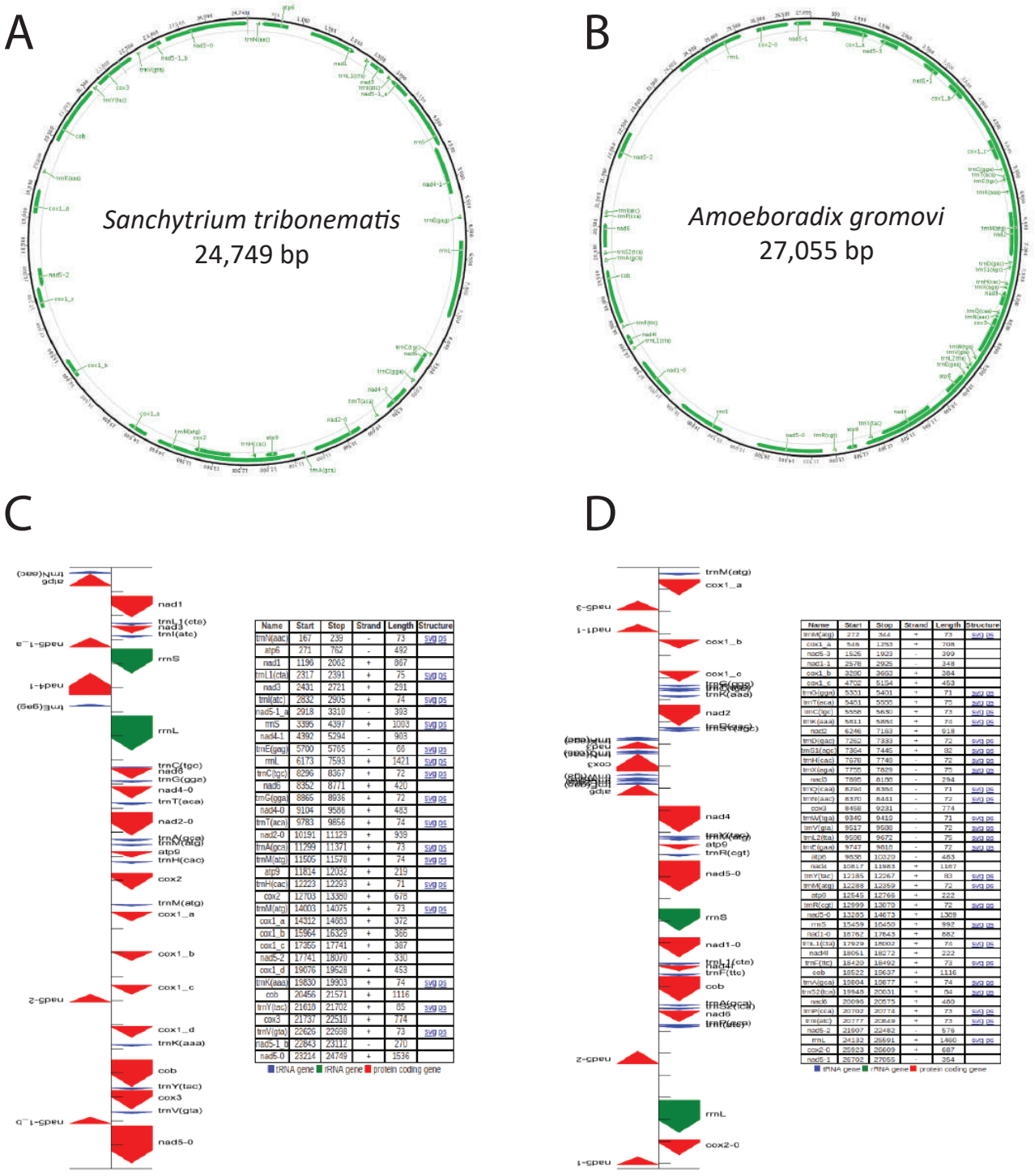
Supplementary Figure 2D. Recoded maximum likelihood tree based on the GBE protein dataset. The tree was reconstructed using 264 nucleotide recoded conserved proteins, 84 species, and 83,321 recoded nucleotide positions, it was inferred with IQ-TREE under the GTR+F+I+G4 model and ultrafast bootstrap as statistical support.



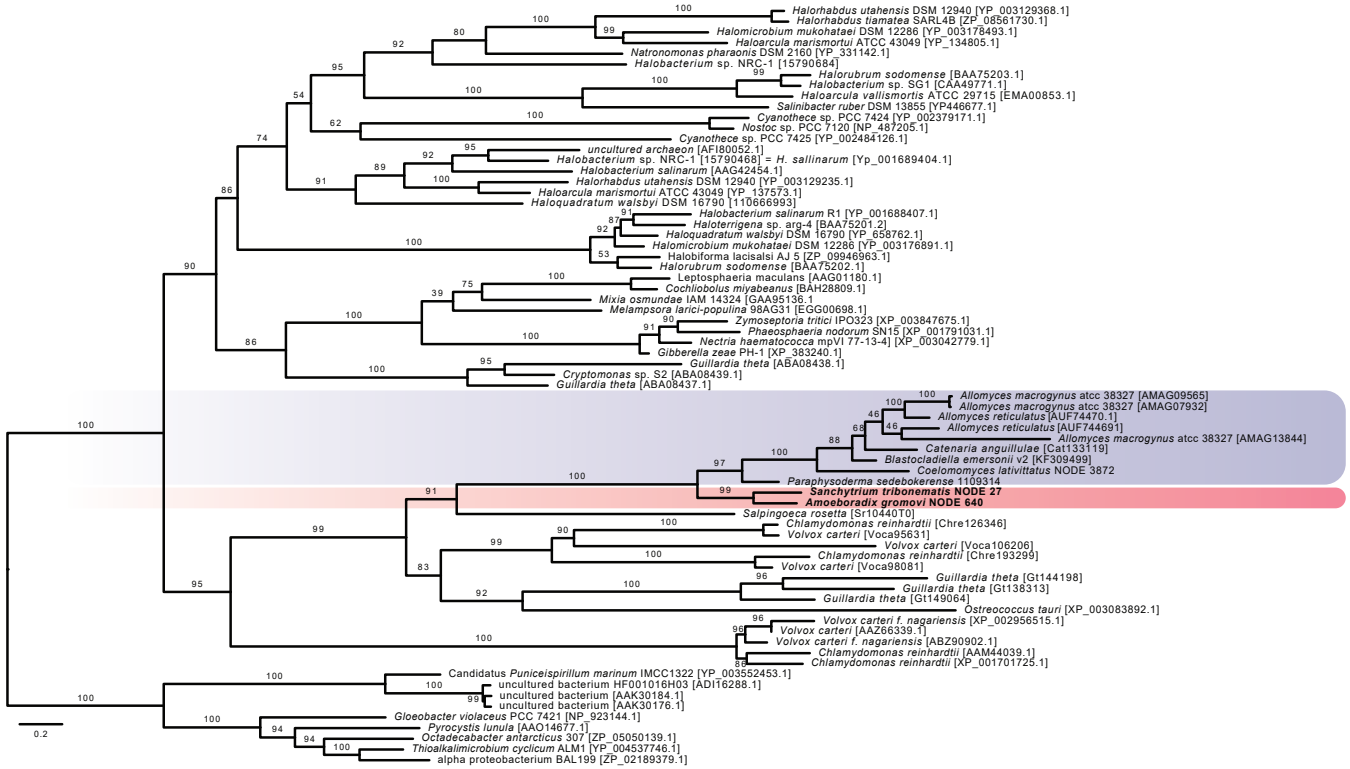
Supplementary Figure 3. Comparative analyses of primary metabolism. (A) Clustering of 43 eukaryotic species according to their profile of presence/absence of the same 8 primary metabolism categories for a total of 1158 COGs. Color code is scaled to display positive Pearson's r (0 to 1). (B and D) Binary heat maps of orthologs (COGs) in 43 eukaryotic genomes/transcriptomes. (C and E) Clustering of 43 eukaryotic genomes/transcriptomes according to their profile of presence/absence color coded to display positive Pearson's r (0 to 1) (B and C) Carbohydrate transport and metabolism (205 OGs, category G). (D and E) Lipid transport and metabolism (201 OGs, category I).



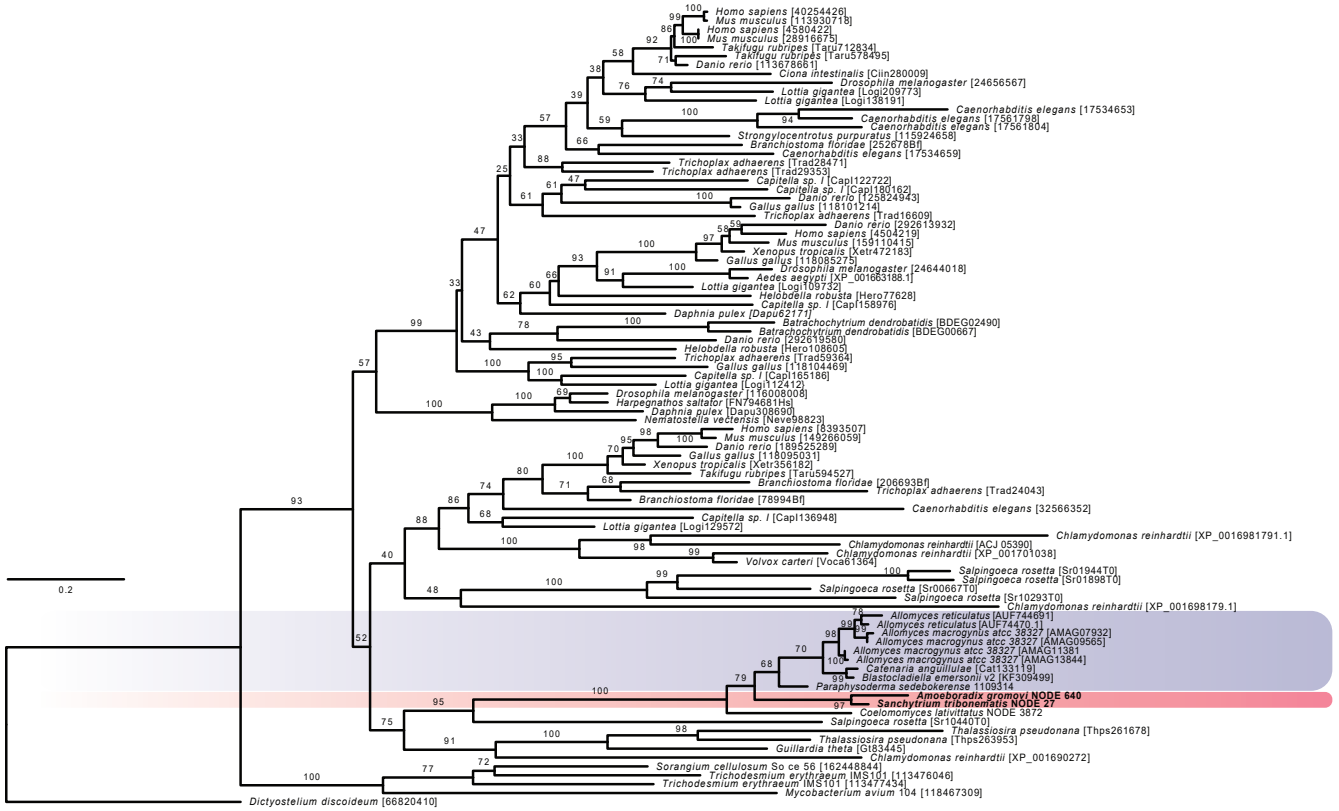
Supplementary Figure 4. KEGG Pathways metabolic comparison. (A-C) KEGG metabolic pathways map01100. (A) *Sanchytrium tribonematis* (green) vs *Amoeboradix gromovi* (pink); both (blue). (B) *Sanchytrium tribonematis (*green) vs *Allomyces macrogynus* (pink); both (blue). (C) *Sanchytrium tribonematis* (green) vs *Rozella allomycis* (pink); both (blue). (D-F) KEEG specific metabolic pathways of *Sanchytrium tribonematis* (green) vs *Allomyces macrogynus* (pink); both (blue). (G-I) KEEG specific metabolic pathways of *Sanchytrium tribonematis* (green) vs *Rozella allomycis* (pink); both (blue). (D and G) KEGG Amino acid synthesis map01230. (E and H) KEGG Steroid biosynthesis map00100. (F and I) KEGG Fatty acid biosynthesis map00061.



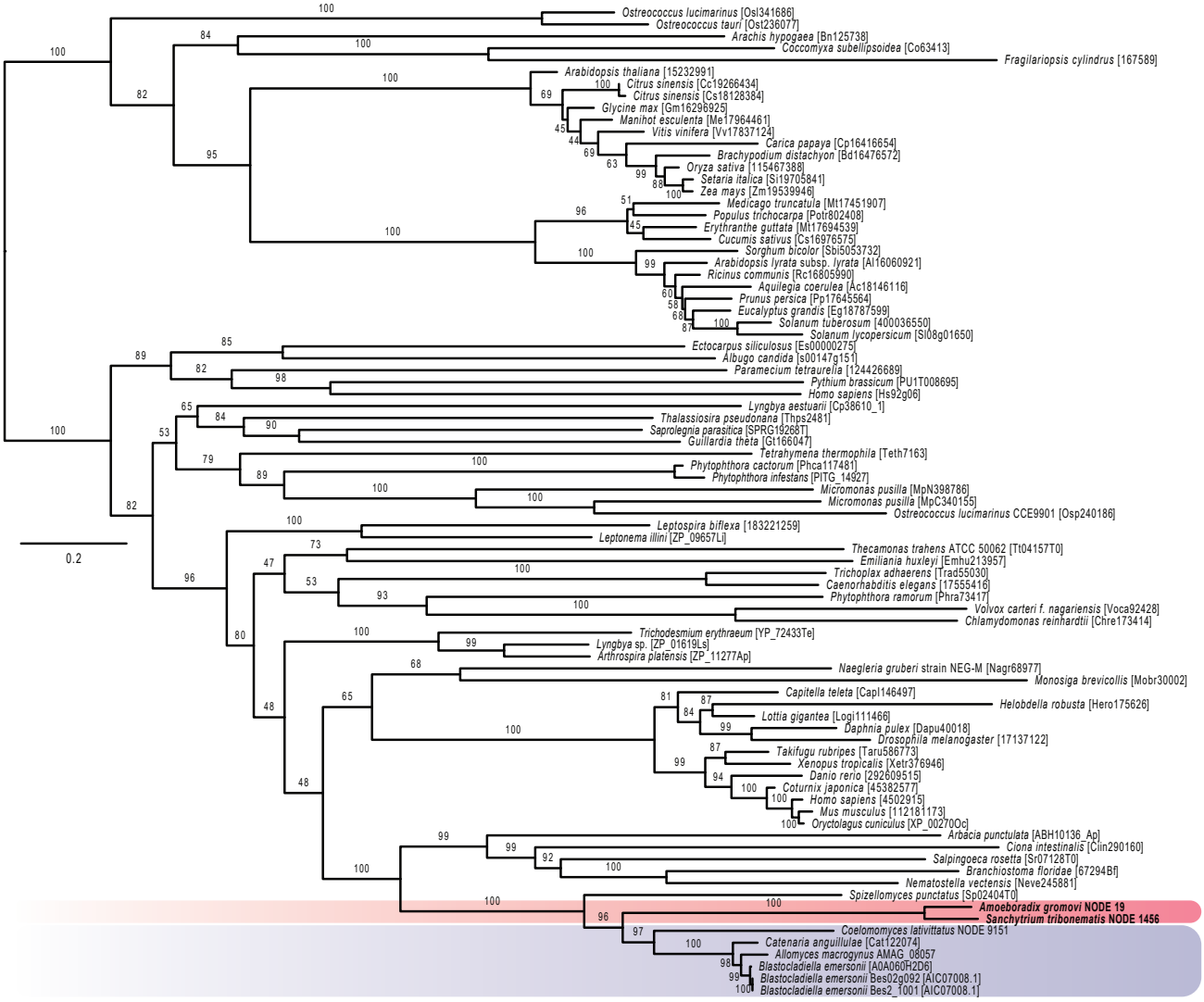
Supplementay Figure 5. MITOS Graphical representation of the mitochondrial genome and gene content of Sanchytrium tribonematis (A and C), and Amoeboradix gromovi (B and C).



Supplementary Figure 6A. Maximum likelihood tree based on the protein dataset of Avelar et al. (2014). Reconstruction of the Type I rhodopsin domain of the BeGC1 gene-fusion, and 416 amino acidic positions, it was inferred with IQ-TREE under the LG+F+I+G4 model and ultrafast bootstrap as statistical support.



Supplementary Figure 6B. Maximum likelihood tree based on the protein dataset of Avelar et al. (2014). Reconstruction of the GC1 guanylyl-cyclase domain of the BeGC1 gene-fusion, and 180 amino acidic positions, it was inferred with IQ-TREE under the LG+G4 model and ultrafast bootstrap as statistical support.



Supplementary Figure 6C. Maximum likelihood tree based on the protein dataset of Avelar et al. (2014). Reconstruction of the cyclic nucleotide gated channel BeCNG1, and 301 amino acidic positions, it was inferred with IQ-TREE under the LG+F+I+G4 model and ultrafast bootstrap as statistical support.

GBE69									
	logL	deltaL	bp-RELL	p-KH	p-SH	p-WKH	p-WSH	c-ELW	p-AU
B+F	-4872140.56	35.265	0.117	0.305	0.484	0.305	0.715	0.118	0.307
C+F	-4872105.29	0.00017871	0.0867	0.498	0.941	0.498	0.938	0.119	0.531
B+F;O+Z	-4872230.71	125.42	0	0.0494	0.11	0.0002	0.0008	2.18E-06	0.0000273
C+F;O+Z	-4872182.49	77.2	0.0001	0.0828	0.25	0.0014	0.006	0.000232	0.000226
B+F;O+ZMD	-4872140.56	35.264	0.119	0.305	0.484	0.305	0.708	0.118	0.307
C+F;O+ZMD	-4872105.29	0.00023738	0.12	0.489	0.927	0.489	0.924	0.119	0.522
O+ZMD	-4872105.29	0	0.145	0.502	1	0.504	0.935	0.119	0.539
O+Z	-4872182.49	77.2	0.0003	0.0828	0.25	0.0014	0.005	0.000233	0.000167
GBE84									
	logL	deltaL	bp-RELL	p-KH	p-SH	p-WKH	p-WSH	c-ELW	p-AU
B+F	-5581490.26	38.89	0.134	0.283	0.374	0.283	0.638	0.136	0.302
C+F	-5581451.37	0	0.163	0.499	1	0.499	0.98	0.182	0.485
B+F;O+Z	-5581588.38	137.01	0.0001	0.027	0.0317	0.0001	0.0003	0.0001	1.14E-07
C+F;O+Z	-5581543.5	92.129	0	0.0008	0.121	0.0008	0.0012	5.00E-05	0.00116
B+F;O+ZMD	-5581490.26	38.889	0.14	0.283	0.374	0.283	0.64	0.136	0.301
C+F;O+ZMD	-5581451.37	0.00028475	0.211	0.501	0.891	0.501	0.888	0.182	0.588
O+ZMD	-5581451.37	0.00071028	0.262	0.495	0.896	0.476	0.891	0.182	0.554
O+Z	-5581543.5	92.13	0.0001	0.0008	0.121	0.0008	0.0013	5.00E-05	0.00122

GBE69						
all_taxa	C+F (%)	B+F (%)	O+ZMD (%)	O+Z (%)	S+B (%)	Dicarya (%)
91,768 aa	79.4	20.6	99	0.2	100	100
87,180 aa	81.5	18.5	99	0	100	100
82,592 aa	93.6	6.4	95.2	0.1	100	100
78,003 aa	90.8	9.2	94.7	0	100	100
73,415 aa	95.7	4.3	96.9	0	100	100
68,826 aa	97.9	2.1	96.7	0	100	100
64,238 aa	95.7	4.3	90.5	0.4	100	100
59,650 aa	98.9	1.1	76.3	1.3	100	100
55,061 aa	99.8	0.1	84.4	0.1	100	100
50,473 aa	97.5	0.7	90.4	0.6	100	100
45,884 aa	95.4	0.6	86.9	5.3	100	100
41,296 aa	77.4	0.1	57.6	39.1	100	100
36,708 aa	37.3	0.5	84.6	6.8	100	100
32,119 aa	26.6	0.5	35.8	40.5	100	100
27,531 aa	6.3	10.5	93.1	0.4	100	100
22,942 aa	0.6	43.2	82.8	3.1	99.8	100
18,354 aa	0	0	0.2	38.5	98	99.1
13,766 aa	0	0	0	66.7	1.5	94.8
9,177 aa	0	0	0	0	0	0
4,589 aa	0	0	0	0	0	0
GBE84						
all_taxa	C+F (%)	B+F (%)	O+ZMD (%)	O+Z (%)	S+B (%)	Dikarya (%)
83,321 aa	82.3	17.7	93.2	2.7	100	100
79,155 aa	86.2	13.8	97.5	0.5	100	100
74,989 aa	94.4	5.6	99.1	0	100	100
70,823 aa	91.3	8.7	99.4	0	100	100
66,657 aa	92.8	7.2	98.9	0	100	100
62,491 aa	93.8	6.2	97.4	0	100	100
58,325 aa	95.6	4.4	95.2	0	100	100
54,159 aa	97	2.9	97.3	0.3	100	100
49,993 aa	98.9	1.1	96.5	0.2	100	100
45,827 aa	98.6	0.1	93.2	1.4	100	100
41,661 aa	96.9	1.4	95.4	0.3	100	100
37,495 aa	98.1	0.2	95.1	2.8	100	100
33,327 aa	94.1	0.1	98.3	0.6	100	100
29,163 aa	55.9	0.5	73.2	25	100	100
24,997 aa	4.6	4.2	28.3	12.3	100	100
20,831 aa	0.7	10.3	82.8	3.9	100	100
16,665 aa	14.1	0.1	21.3	0.8	99.3	98.7
12,499 aa	0	0	0	0	95.6	99.3
8,333 aa	0	0	0	0	93.5	71.4
4,167 aa	0	0	0	0	0	0

Supplementary Table 2. Alternative topology tests of phylogenomic analyses and removal of fast-evolving sites test. IQ-tree output alternative topology tests; and progressive exclusion of fastest evolving sites. B+F = Blastocladiomycota+Sanchytriaceae sister of all other fungi, C+F = Chytridiomycota sister of all other fungi, B+F;O+Z = Blastocladiomycota+Sanchytriaceae sister of all other fungi and Olpidium within Zoopagomycota, C+F;O+Z = Chytridiomycota sister of all other fungi and Olpidium within Zoopagomycota, B+F;O+ZMD = Blastocladiomycota+Sanchytriaceae sister of all other fungi and Olpidium independent lineage sister to all non-flagellated fungi, C+F;O+ZMD = Chytridiomycota sister of all other fungi. and Olpidium independent lineage sister to all non-flagellated fungi, O+ZMD = Olpidium independent lineage sister to all non-flagellated fungi, O+Z = Olpidium within Zoopagomycota, S+B = Sanchytriaceae within Blastocladiomycota, Dicarya = The Dicarya monophyly. Blue-shadowed cells indicate significance. The KH, SH and AU tests return p-values; a tree is rejected if its p-value < 0.05.

Supplementary Tables S1 and S3 and all tables in high quality can be found in:
https://figshare.com/authors/Luis_Javier_Galindo/6432803

Résumé en français

L'arbre phylogénétique des eucaryotes comprend plusieurs grands supergroupes monophylétiques, dont les Opisthokonta. Ce groupe comprend deux branches, les Holozoa, qui inclut les animaux et quelques lignées unicellulaires, et les Holomycota, qui regroupe les champignons et leurs parents unicellulaires. Le terme Holomycota a été proposé pour la première fois par Liu et al. (2009) pour désigner les nucleariidés plus les champignons ; un synonyme est Nucletmycea qui a été proposé par Brown et al. (2009).

Les Holomycota ont été reconnus comme formant un clade à partir des analyses phylogénétiques moléculaires, qui regroupaient des organismes qui, par leurs caractères morphologiques, n'avaient jamais été considérés comme apparentés. En raison de la grande diversité morphologique et de modes de vie de ses membres, la reconstruction des traits de leur dernier ancêtre commun est une tâche difficile (Richards *et al.*, 2017b).

D'un point de vue phylogénétique, les Holomycota comprennent à la fois des champignons multicellulaires, bien connus, et leurs parents phylogénétiques, dont la véritable diversité reste largement inconnue. La fraction unicellulaire connue des Holomycota comprend plusieurs lignées zoosporiques (par exemple Chytridiomycota et Blastocladiomycota) au sein des champignons, mais aussi une variété de lignées apparentées aux champignons classiques : nucleariidés, rozellides, aphélides et microsporidies. Les relations phylogénétiques de ces lignées entre elles et avec les champignons classiques restent à établir solidement.

Le développement des méthodes moléculaires a permis aussi l'exploration de la diversité directement à partir de l'environnement, notamment par des approches de (meta)barcoding de gènes d'ARNr 18S et des régions intergéniques adjacentes (ITS). Ces études ont montré une grande diversité de lignées chez les Holomycota, dont certaines sont nouvelles et restent inexplorées (Lilleskov *et al.*, 2002; Cox *et al.*, 2010; Tedersoo *et al.*, 2014; Yahr *et al.*, 2016; Bass *et al.*, 2018). Cependant, les analyses du gène de l'ARNr 18S sont insuffisantes pour résoudre les relations profondes entre de nombreuses clades Holomycota.

Au cours des dernières années, les techniques à haut débit ont permis le séquençage de centaines de nouveaux génomes et transcriptomes (Spatafora *et al.*, 2017). Cela a permis de réaliser des études phylogénomiques multigéniques, qui augmentent le signal disponible pour résoudre les relations évolutives. Il a ainsi été possible de reconstruire l'arbre phylogénétique des Holomycota

avec une résolution sans précédent (par exemple Chang *et al.*, 2015; Torruella *et al.*, 2018). Néanmoins, la plupart de génomes séquencés correspondent à des espèces fongiques faciles à cultiver, qui présentent souvent un intérêt particulier pour l'homme (par exemple les parasites, les symbiotes végétaux, les levures).

Ainsi, si de nombreuses relations entre des lignées importantes ont été résolues, de nombreuses nouvelles questions se posent (Chang *et al.*, 2015; Mikhailov *et al.*, 2017; Torruella *et al.*, 2018). Pour résoudre ces nouvelles questions et reconstruire des arbres phylogénétiques robustes, il est essentiel d'obtenir des données génomiques et transcriptomiques des nouveaux clades d'Holomycota. Cependant, en particulier pour la fraction unicellulaire non cultivée des Holomycota, isoler les organismes et obtenir ensuite suffisamment de matériel pour les séquencer peut être une tâche difficile. En effet, beaucoup de ces organismes appartiennent à une fraction de la diversité souvent désignée comme la « matière noire » microbienne. Un terme qui définit la grande fraction des micro-organismes (certains aventurent des chiffres de ~99%) qui ne peuvent pas être cultivés et étudiés au laboratoire (Ishoey *et al.*, 2008; Lasken & McLean, 2014; Wang & Navin, 2015).

Récemment, les méthodes de type ‘omiques’ à partir des cellules uniques se sont imposés comme l'une des meilleures approches pour dévoiler cette fraction unicellulaire incultivable, en générant des données génomiques/transcriptomiques (pour obtenir un signal phylogénétique élevé) à partir d'une diversité autrement inaccessible. Dans le cadre de mon doctorat, j'appliquerai des approches "omiques" de « cellule unique » pour générer des données génomiques et transcriptomiques afin de mieux résoudre l'arbre phylogénétique des Holomycota et de mieux comprendre leur évolution en étudiant trois de ses principales branches : les nucleariidés, les microsporidies et les champignons. Dans ce contexte général, les objectifs spécifiques de mon doctorat sont les suivants :

- Résoudre les relations phylogénétiques internes des nucleariidés. Pour ce faire, nous combinerons des approches basées sur cellule unique et des cultures pour obtenir des données génomiques et transcriptomiques sur les nucleariidés. Nous générerons et analyserons des données provenant du genre nucleariidé *Nuclearia* et de deux espèces présumées de nucléariidés à thèque pour lesquelles aucune donnée moléculaire n’est disponible, des genres *Pompholyxophrys* et *Lithocolla*.
- Déterminer la position phylogénétique des metchnikovellides au sein des Microsporidia et étudier les synapomorphies du clade Microsporidia + Rozellida. Nous séquencerons le génome obtenu à

partir de « cellule unique » de *Metchnikovella incurvata*, une espèce de metchnikovellidé dont la taxonomie a été vérifiée par une approche classique. Nous chercherons à étudier si les deux metchnikovellides (*M. incurvata* et *Amphiamblys* sp.) se regroupent et nous confirmerons éventuellement leur position en tant que clade sœur de core Microsporidia dans des analyses phylogénomiques poussées. Les analyses du contenu génétique devraient également fournir des indications sur l'évolution du génome le long de la branche des Microsporidia et éventuellement mettre en évidence le gain et/ou la perte de certaines fonctions.

- Résoudre le placement phylogénétique des sanchytrides (*Amoeboradix gromovi* et *Sanchytrium tribonematis*) au sein des champignons et étudier leurs traits (spécifiques et partagés) d'histoire de vie. Notre troisième et dernier objectif est de résoudre la position phylogénétique de cet clade fongique énigmatique et d'améliorer potentiellement la phylogénie fongique globale générant en analysant des données génomiques de « cellule unique » obtenues à partir des sanchytrides. En particulier, nous comparerons leurs génomes à ceux d'autres champignons zoosporiques pour : (i) déterminer le nombre de pertes indépendantes du flagelle chez les Holomycota à partir du contenu en gènes associés à ce trait, (ii) étudier le profil métabolique ancestrale, et (iii) comprendre les déterminants moléculaires du flagelle sanchytridien, réduit mais très particulier.

1) Combinaison de la culture et des approches de « cellule unique » pour l'étude phylogénomique des amibes nucleariidées, proches parents des champignons

Les nucleariidés sont des amibes filoses non flagellées, de vie libre et phagotrophes (Patterson, 1984), connues depuis 1865 (Cienkowski, 1865). Elles présentent des caractéristiques morphologiques que l'on retrouve largement chez différents lignées d'eucaryotes, ce qui a conduit à leur classification historique comme faisant partie de différents taxons d'amibes (Cavalier-Smith, 1993a; Patterson *et al.*, 2000).

Les phylogénies moléculaires du gène ARNr 18S ont d'abord placé *Nuclearia* comme groupe sœur des champignons (Brown *et al.*, 2009; Liu *et al.*, 2009), ce qui a été ensuite corroboré des analyses phylogénomiques (Schalchian-Tabrizi *et al.*, 2008; Liu *et al.*, 2009). Les nucleariidés constituent la première branche à émerger du clade Holomycota, la lignée sœur de tous les autres membres du clade (Torruella *et al.*, 2015, 2018).

Toutefois, on sait que les nucleariidés constituent un groupe diversifié d'après les analyses de metabarcoding environnemental (Del Campo & Ruiz-Trillo, 2013; Arroyo *et al.*, 2018), mais la quantité de données génomiques/transcriptomiques disponibles pour le clade est faible. On ne dispose des données que pour les membres des fonticulides (génome de *Fonticula alba* et un métagénome de fonticulide), *Nuclearia* (données de séquence exprimées (EST) pour *N. pattersoni* et *N. moebiusi*) et, récemment, un transcriptome pour *Parvularia atlantis*.

En raison du bas signal phylogénétique du gène de l'ARNr 18S, et du fait que presque aucune donnée génomique sur les nucleariidés n'est disponible, les relations entre les membres du clade restent non résolues. Une partie de cette diversité non résolue englobe les nucleariidés à thèque, dont la principale caractéristique morphologique est la présence d'une couverture siliceuse recouvrant la membrane cellulaire. Plusieurs candidats nucleariides putatifs dont l'affiliation phylogénétique reste à valider par des données moléculaires.

Les efforts combinés de notre équipe et des laboratoires partenaires ont permis d'isoler des échantillons pour les nucleariidés à thèque *Lithocolla* (une souche marine cultivée) et *Pompholyxophrys* (à partir d'échantillons d'eau douce d'un lac). En outre, pour les nucleariidés canoniques, nous avons pu obtenir plusieurs espèces de *Nuclearia* à partir des collections de cultures et des micromanipulations à partir de l'intestin d'un têtard. Les transcriptomes, les génomes de cellule unique (SCT et SCG) et les transcriptomes provenant de la souche en culture ont été séquencés pour ces nouveaux échantillons de nucléariidés. Nous avons fixé les objectifs suivants :

- Résoudre les relations des différentes lignées au sein du clade des nucleariidés. Pour ce faire, nous réaliserons une analyse phylogénomique multigénique. Étant donné l'origine mixte de nos données, nous voulons atteindre cet objectif en utilisant à la fois des approches de « cellule unique » et des techniques basées sur la culture.
- En outre, étant donné que nous utiliserons des approches de culture et de cellule unique pour le séquençage des données transcriptomiques et génomiques des espèces de nucleariidés, nous avons voulu évaluer quelle technique est la plus performante, et voir si différentes approches peuvent donner des aperçus différents sur certaines caractéristiques du groupe (par exemple, les aspects écologiques).

Nos analyses phylogénomiques ont été effectuées avec deux ensembles de données de 264 protéines conservées (ensemble de données GBE) et 74 marqueurs de protéines à copie unique

(ensemble de données SCPD) utilisés précédemment pour étudier la phylogénie des Holomycota (Torruella et al., 2012 ; Mikhailov et al., 2016). Nos résultats ont montré que les organismes à thèque des genres *Lithocolla* et *Pompholyxophrys* appartiennent bien aux nucleariidés et formant un groupe monophylétique sœur du clade *Nuclearia*, avec lequel ils forment un groupe sœur de la lignée des petites amibes filoses *Parvularia* et *Fonticula* (Figure 25).

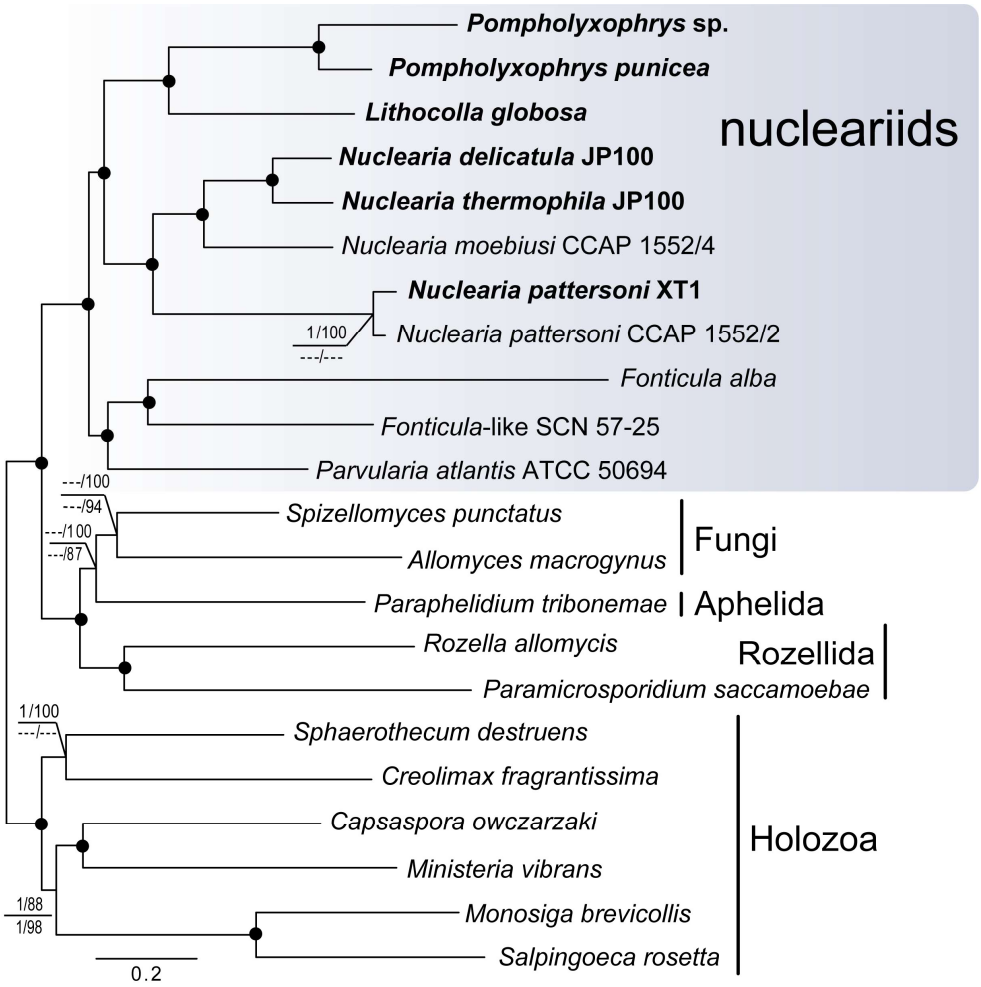


Figure 25. Arbre phylogénomique de ML basé sur l'ensemble des données des protéines GBE. L'arbre a été reconstruit en utilisant 264 protéines conservées, 22 espèces et 96 276 positions d'acides aminés conservées avec le modèle d'évolution de séquence LG + R5 + C60. Les valeurs supérieures correspondent aux supports obtenus à partir de l'ensemble de données GBE et les valeurs inférieures à ceux obtenus à partir de l'ensemble de données du domaine protéique à copie unique (SCPD21 ; sans *N. pattersoni* XT1). Les PP bayésiens sous le modèle CAT-Poisson sont indiqués à gauche et les supports ML UFBS sont indiqués à droite. Les branches dont les valeurs de support sont supérieures ou égales à 0,99 PP et 95 % UFBS sont indiquées par des points noirs. Les noms d'espèces en gras correspondent à ceux pour lesquels nous avons obtenu des séquences de transcriptome et/ou de génome dans cette étude.

Ces résultats nous ont permis de reconstituer les caractéristiques de l'ancêtre commun le plus récent (MRCA) des nucleariidés et ont montré qu'il s'agissait très probablement d'une amibe d'eau douce, bacterívore, non flagellée, filose et mucilagineuse. A partir de cet ancêtre, deux groupes ont évolué pour atteindre des tailles cellulaires plus petites (*Parvularia* et *Fonticula*) et plus grandes (*Nuclearia* et les genres porteurs de couverture), ce qui a conduit à une spécialisation écologique différente. La clade *Lithocolla* + *Pompholyxophrys* a développé des couvertures cellulaires exogènes ou endogènes à partir d'un ancêtre nu semblable au *Nuclearia*.

Nous avons également étudié les génomes de cellule unique et décrit la présence de séquences d'endosymbionts bactériens pour *Pompholyxophrys*. L'utilisation de techniques de génomique de cellule unique a permis d'identifier de nouveaux endosymbionts bactériens des clades Rikettsiales et Chlamydiae.

Finalement, nous évaluons comment les différentes techniques de culture et de cellule unique ont fonctionné dans notre étude. Nous avons confirmé que les approches de cellule unique permettaient de récupérer suffisamment de marqueurs conservés pour les études phylogénomiques sur les nucleariidés. Le pourcentage de marqueurs protéiques phylogénétiques conservés récupérés pour nos études phylogénétiques était plus élevé pour les SCT que pour les SCG dans les cellules des nucleariidés, ce qui est plus que suffisant pour effectuer des analyses phylogéniques robustes. Il convient de souligner que les techniques basées sur la culture ont permis d'obtenir plus de données que toute autre méthode.

Données supplémentaires provenant des nouveaux représentants des nucleariidés permettront de dévoiler de nouvelles relations au sein de la lignée et contribueront à continuer à façonner l'identité de l'ancêtre nucleariide.

2) Génomique évolutive de *Metchnikovella incurvata* (Metchnikovellidae) : Une microsporidie basale

La prochaine branche à diverger après les nucleariidés dans l'arbre phylogénétique des Holomycota est celle composée par les Microsporidia et les rozellides. Les relations entre les différents lignées de ce clade restent non résolues. Par exemple, les Rozellida apparaissent comme un groupe paraphylétique qui semble inclure les Microsporidia (James *et al.*, 2013b; Mikhailov *et*

al., 2016; Quandt *et al.*, 2017). L'un de ces clades ayant des relations non résolues est le metchnikovellids (Metchnikovellidae). Les metchikovellids sont un groupe de microsporidiens avec des caractéristiques historiquement considérées comme "primitives", notamment un tube polaire court et l'absence d'un polaroplaste et d'un stade de prolifération mérogoniale (Larsson, 2000; Larsson & Køie, 2006; Sokolova *et al.*, 2013). Récemment, le génome du premier metchnikovellide, *Amphiamblys* sp. a été séquencé (Mikhailov *et al.*, 2016), montrant que les metchnikovellides sont le groupe frère de tous les autres core Microsporidia. Cependant, cette espèce n'était pas caractérisée morphologiquement et son seul génome ne suffisait donc pas pour valider la position des metchnikovellides avec certitude. Il était donc important de séquencer davantage de génomes de metchnikovellids (si possible caractérisés morphologiquement) pour vérifier s'ils forment un groupe cohérent frère des core Microsporidia. En outre, cela permettrait d'étudier si le contenu génétique des metchnikovellides ressemble davantage aux Microsporidia canoniques ou aux rozellides.

Pour aborder ces questions, nous avons séquencé à partir de cellule unique le génome d'une deuxième espèce de metchnikovellide, *Metchnikovella incurvata*, qui a été caractérisée morphologiquement. Nous avons établi les objectifs suivants :

- Confirmer ou infirmer l'ordre d'embranchement des metchnikovellides comme groupe sœur des core Microsporidia. Le séquençage du génome de *M. incurvata* nous a permis d'effectuer des analyses phylogénomiques en incluant *Amphiamblys* sp. pour voir s'ils forment un groupe cohérent, et si les metchnikovellides sont bien les sœurs de core Microsporidia.
- Déterminer si le contenu en gènes et les principales voies métaboliques de *M. incurvata* et *Amphiamblys* sp. (si leur relation est confirmée) ressemblent plus à ceux des core Microsporidia ou bien à des microsporidies basales apparentés aux rozellidees ou à des rozellids.

Nos études phylogénomiques multigéniques utilisant l'ensemble des données du jeu de données SCPD avec les génomes de *Metchnikovella incurvata* et *Amphiamblys* sp. ont confirmées que *M. incurvata* appartient au Metchinokovellidae, et forment un groupe monophylétique sœur de toutes les autres core Microsporidia à longue branche (Figure 26).

Nous avons utilisé les catégories métaboliques définies dans les listes d'orthologues « GO terms » pour caractériser et comparer les génomes des différents membres du clade Microsporidia + Rozellida. Nous avons montré que le contenu du génome des metchnikovellides ressemble plus à

celui des core Microsporidia à longue branche qu'à des Microsporidia branchant précocement dans l'arbre (*Mitosporidium*) ou à des rozellides (*Paramicrosporidium* et *Rozella*).

Nos analyses ont montré la présence de gènes acquis par transfert horizontal dans le génome de *M. incurvata*, notamment le gène de la superoxyde dismutase de manganèse (*MnSOD*). Très probablement cette enzyme protège la cellule des metchnikovellides, qui se développent dans des conditions anaérobies, des effets délétères de l'oxygène

Finalement, nous avons inféré les gains et les pertes de gènes orthologues dans les génomes des différents représentants à travers l'arbre des Microsporidia + Rozellida, y compris les metchnikovellides. Nos résultats suggèrent que les événements de réduction du génome et l'évolution de nouveaux gènes ont eu lieu en même temps que l'évolution et l'adaptation des Microsporidia à leurs hôtes.

Le séquençage de plus de génomes des Microsporidia et des rozellideds (par exemple chytridiopsides et *Nucleophaga*) pourrait aider à clarifier les relations entre les deux clades et à comprendre les mécanismes de leur évolution réductive.

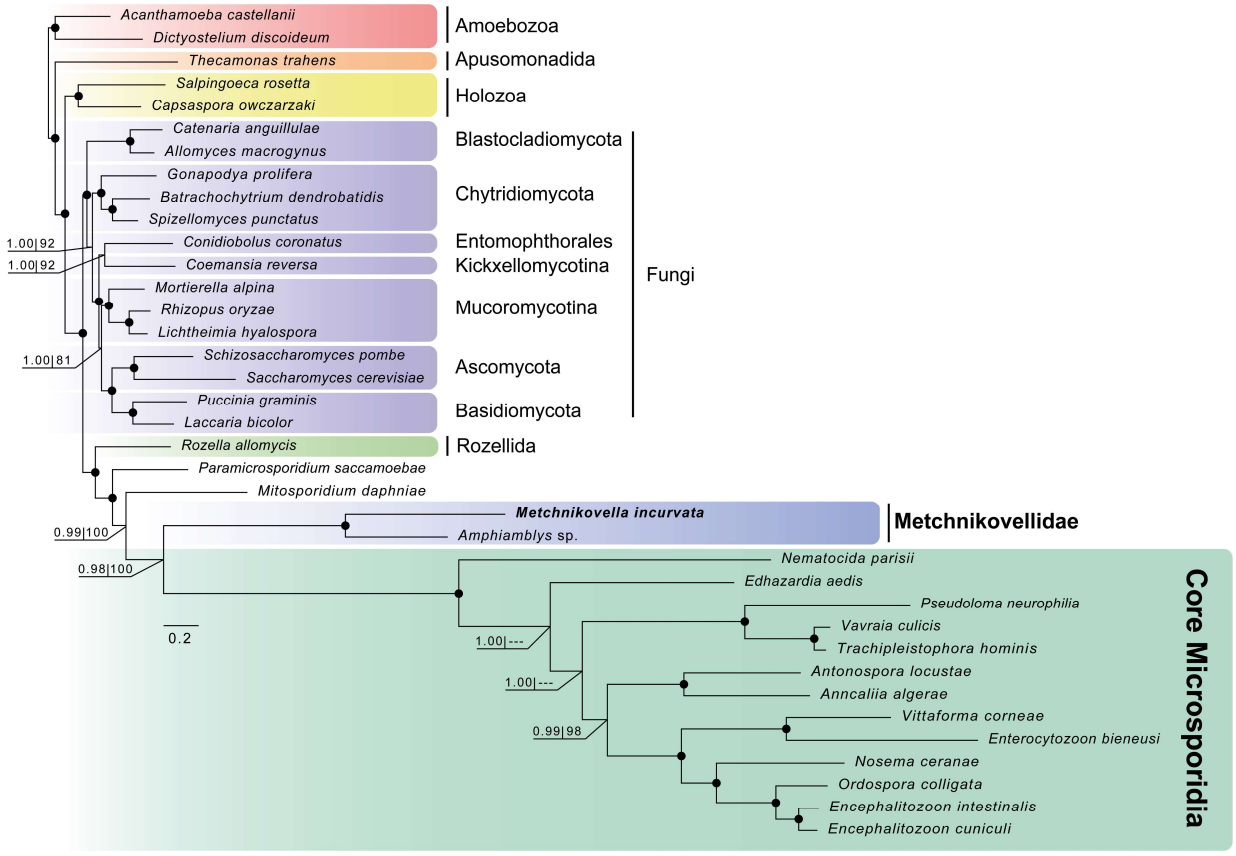


Figure 26. Arbre phylogénomique bayésien montrant la position des metchnikovellidés. L'arbre a été reconstruit en utilisant une concaténation de 56 domaines protéiques à copie unique (SCPD) pour 32 représentants du clade

Holomycota et 5 autres espèces d'Amorphea comme groupe extérieur (2 Holozoa, 1 Apusomonadida, et 2 Amoebozoa). Les supports statistiques sur les nœuds correspondent à des probabilités postérieures (pp) (valeurs sur la gauche) et les valeurs de bootstrap (bs) du maximum de vraisemblance (ML) (sur la droite). Les séquences obtenues dans le cadre de cette étude sont surlignées en noir. Les valeurs de support >0,99 pp et >95% bs sont indiquées par un cerclée noire.

3) Un nouveau clade fongique permet de raffiner l'arbre des champignons et de reconstruire l'évolution du flagelle chez les Holomycota

Le dernier grand groupe de la branche Holomycota correspond aux célèbres champignons. Les premières branches de l'arbre des champignons sont composées de deux lignées zoosporiques : Blastocladiomycota et Chytridiomycota (James *et al.*, 2006a, 2006b). Tant Blastocladiomycota (Chang *et al.*, 2015; McCarthy & Fitzpatrick, 2017; Ahrendt *et al.*, 2018; Torruella *et al.*, 2018) comme Chytridiomycota (Sekimoto *et al.*, 2011; Letcher *et al.*, 2013; Torruella *et al.*, 2015; Spatafora *et al.*, 2016b; Mikhailov *et al.*, 2017; Tedersoo *et al.*, 2018) ont été respectivement récupérés comme la lignée sœur de tous les autres champignons dans les études phylogénétiques et phylogénomiques. Ainsi, la position globale des Blastocladiomycota et des chytrides au sein des champignons reste incertaine.

La relation de plusieurs groupes zoosporiques *incertae sedis* reste également non résolue. L'un de ces groupes est celui des Sanchytriaceae, une famille composé de deux espèces connues, *Amoeboradix gromovi* et *Sanchitrium tribonematis* (Karpov *et al.*, 2017a, 2018, 2019). Les sanchytrides sont des champignons zoosporiques très atypiques, ayant une ultrastructure de flagelle très réduite, qui semble non fonctionnelle, et en même temps l'un des plus longs kinétosomes de tous les eucaryotes. Les deux espèces de sanchytrides branchent ensemble dans des phylogénies de gènes 18S + 28S rRNA (Karpov *et al.*, 2018), mais leur affinité avec toute autre lignée fongique reste incertaine.

En plus des sanchytrides, *Olpidium* est un autre genre zoosporique dont la filiation au sein des champignons est non résolue. Les analyses phylogénétiques du gène de l'ARNr 18S des espèces d'*Olpidium* semblent indiquer une relation avec les Zoopagomycota non flagellés (James *et al.*, 2006a, 2006b; Sekimoto *et al.*, 2011).

Pour résoudre les relations phylogénétiques entre les champignons zoosporiques, deux approches seront nécessaires : 1) améliorer l'échantillonnage autour des taxons divergents branchant

précocement autour de la division Blastocladiomycota-Chytridiomycota (par exemple à partir des génomes des sanchytrides, *Olpidium*, Nephridiophagidae, chytrid-like-clade-1) et (2) améliorer les méthodes de reconstruction phylogénétique associées au génome et au contenu et à la composition des gènes (Spatafora *et al.*, 2017).

Nous avons décidé de suivre la première approche et de séquencer les génomes de cellule unique d'*A. gromovi* et de *S. tribonematis*. Nous avons établi les objectifs suivants :

- Reconstruire les relations phylogénétiques des sanchytrides et d'autres lignées zoosporiques des champignons. En séquençant les génomes de ces deux espèces de sanchytrides, nous allons reconstruire un arbre phylogénétique des Holomycota dans lequel nous pourrions potentiellement résoudre la position des sanchytrides. En outre, nous inclurons les données génomiques disponibles d'*Olpidium bornovanus*, afin de confirmer leur ordre d'embranchement au sein des champignons non flagellés. Enfin, l'ajout des sanchytrides, d'*Olpidium* et d'un large échantillonnage de taxons fongiques permettra de construire un arbre phylogénomique plus solide permettant de clarifier les relations évolutives entre les chytrides et les Blastocladiomycota.
- Comparer les génomes des sanchytrides, d'autres champignons zoosporiques et d'autres représentants des Holomycota. Le but est d'estimer le nombre de pertes de flagelle indépendantes dans les Holomycota, par l'étude des gènes impliqués dans la formation du flagelle. Nous analyserons également le potentiel métabolique primaire de ces organismes. Finalement, nous voulons comprendre quels sont les déterminants moléculaires du flagelle réduit et atypique des sanchytrids.

Nos analyses phylogénomiques réalisés en utilisant l'ensemble de données GBE de 264 gènes avec deux échantillonnages de taxons de 84 et 69 espèces eucaryotes, respectivement, ont placé les sanchytrides comme lignée sœur de Blastocladiomycota, dans un clade bien supporté (Figure 27). Les sanchytrides présentent suffisamment de caractéristiques pour faire potentiellement partie de leur propre phylum, les Sanchytriomycota. En plus du flagelle réduit avec un long kinétosome, le sanchytrides ont une longue branche, indiquant un génome qui évolue rapidement.

Ces analyses phylogénomiques et d'autres tests sur la position des chytrides ont également indiqué que les chytrides pourraient être la lignée sœur du reste des champignons. En outre, pour la première fois, nous avons confirmé dans un cadre phylogénomique que le champignon zoosporique *Olpidium* fait partie d'une nouvelle lignée sœur des champignons non flagellés, formant son propre phylum Olpidiomycota.

Notre évaluation de la composition métabolique primaire des Holomycota, incluant les sanchytrides, en utilisant les catégories de gènes orthologues COG, suggère un potentiel métabolique très atypique et différent "champignons canoniques". Certaines voies métaboliques des sanchytrids semblent particulièrement réduites, notamment celles du métabolisme et le transport des lipides et des glucides.

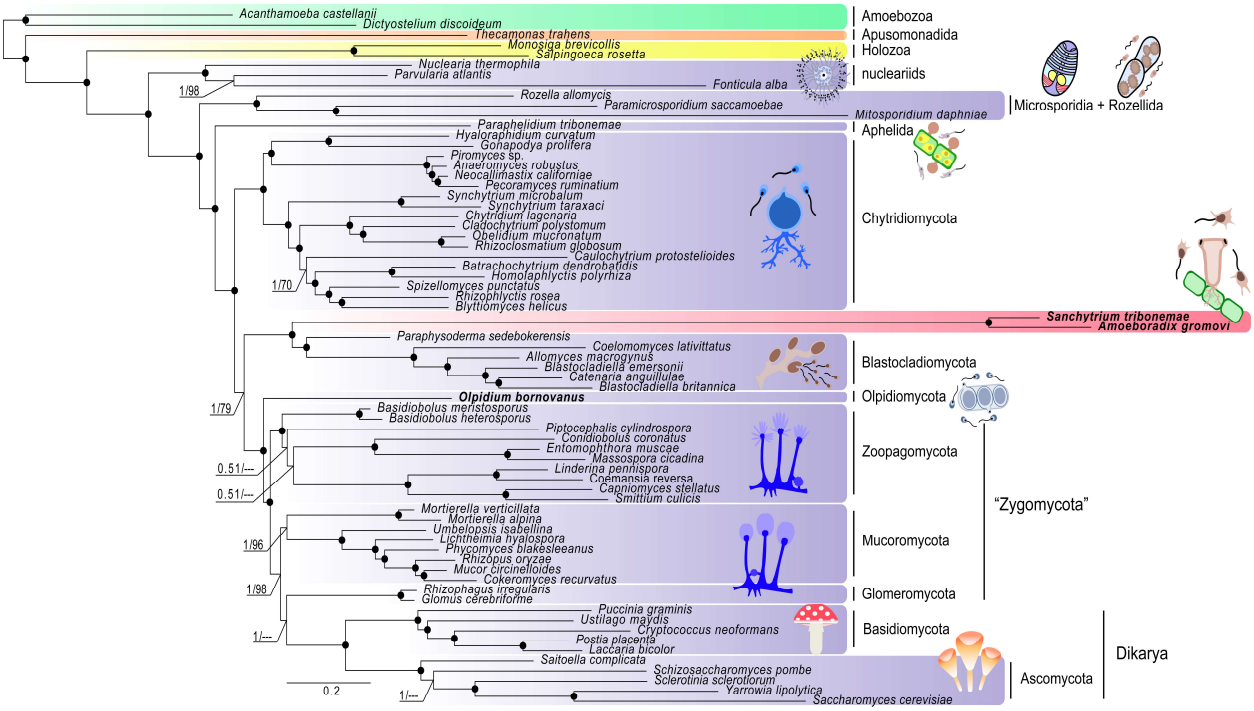


Figure 27. Arbre phylogénomique d'inférence bayésienne (BI) basé sur 264 protéines conservées (GBE). L'arbre a été reconstruit en utilisant 69 espèces et 91 768 positions d'aminoacides avec le modèle d'évolution de séquences CAT-Poisson et le modèle LG + R9 + PMSF pour le maximum de vraisemblance (ML). Les branches dont les valeurs de support sont supérieures ou égales à 0,99 BI de probabilité postérieure et 99% ML de bootstrap sont indiquées par des points noirs.

Les zoospores de sanchytrides portent un flagelle non mobile structurellement réduit (pseudocilium) dont l'ultrastructure a été caractérisée (Karpov *et al.*, 2018, 2019). Des analyses comparatives pour plus de 60 protéines spécifiques du flagelle indiquent que l'ensemble de gènes impliqués dans la formation et la fonction du flagelle manquent dans les deux génomes de sanchytrides, ce qui explique leur ultrastructure réduite. En effectuant cette analyse dans le cadre de notre échantillonnage actuel des taxons, nous constatons 4 événements indépendants de perte de flagelles chez les Holomycota.

Nos analyses ont montré que les génomes des sanchytrides portent le gène de fusion *BeGC1* et le gène de canal *BeCNG1*, deux gènes impliqués dans une cascade de perception de la lumière chez le Blastocladiomycota *B. emersonii,* qui semble étroitement apparenté, (Avelar *et al.*, 2014). Il a également été démontré la présence d'organites lipidiques proéminents dans les sanchytrides. Cet organite lipidique chez *B. emersonii* est l'endroit où les produits de ces gènes sont localisés, fonctionnant comme un " œil ". S'il est confirmé que les sanchytrides peuvent percevoir la lumière, cela pourrait expliquer le maintien et la sélection d'un flagelle très réduit mais en même temps d'un long kinétosome. Nous avons émis l'hypothèse que la pression évolutive vers la perception de la lumière pourrait avoir conduit au maintien d'une structure de soutien flagellaire (le kinétosome) pour « l'œil » lipidique, ce qui expliquerait ce trait atypique dans les sanchytrides.

En conclusion, dans mon travail de doctorat, j'ai enrichi les données génomiques disponibles pour les Holomycota unicellulaires, notamment par le biais d'approches de cellule unique. Les analyses phylogénomiques basées sur ces données clarifient les relations entre et au sein des clades des Holomycota, y compris les nucleariidés, les microsporidies et les champignons et. Ensemble, ces résultats constituent une contribution modeste mais significative à la tâche ambitieuse de résoudre l'arbre de la vie.

References

ADL, S.M., BASS, D., LANE, C.E., LUKES, J., SCHOCH, C.L., SMIRNOV, A., AGATHA, S., BERNEY, C., BROWN, M.W., BURKI, F., CARDENAS, P., ČEPICKA, I., CHISTYAKOVA, L., DEL CAMPO, J., DUNTHORN, M., ET AL. (2019) Revisions to the Classification, Nomenclature, and Diversity of Eukaryotes. *Journal of Eukaryotic Microbiology* **66**, 4–119.

ADL, S.M., SIMPSON, A.G.B., FARMER, M.A., ANDERSEN, R.A., ANDERSON, O.R., BARTA, J.R., BOWSER, S.S., BRUGEROLLE, G., FENSOME, R.A., FREDERICQ, S., JAMES, T.Y., KARPOV, S., KUGRENS, P., KRUG, J., LANE, C.E., ET AL. (2005) The new higher level classification of eukaryotes with emphasis on the taxonomy of protists. *Journal of Eukaryotic Microbiology*. .

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., HEISS, A., HOPPENRATH, M., LARA, E., GALL, L. LE, LYNN, D.H., ET AL. (2012) The revised classification of eukaryotes. *Journal of Eukaryotic Microbiology* **59**, 429–493.

AGIZZIO, A.P., DA CUNHA, M., CARVALHO, A.O., OLIVEIRA, M.A., RIBEIRO, S.F.F. & GOMES, V.M. (2006) The antifungal properties of a 2S albumin-homologous protein from passion fruit seeds involve plasma membrane permeabilization and ultrastructural alterations in yeast cells. *Plant Science*.

AHRENDT, S.R., QUANDT, C.A., CIOBANU, D., CLUM, A., SALAMOV, A., ANDREOPOULOS, B., CHENG, J.F., WOYKE, T., PELIN, A., HENRISSAT, B., REYNOLDS, N.K., BENNY, G.L., SMITH, M.E., JAMES, T.Y. & GRIGORIEV, I. V (2018) Leveraging single-cell genomics to expand the fungal tree of life. *Nature Microbiology* **3**, 1417–1428. Springer US.

AINSWORTH, G.C. (1976) Introduction to the History of Mycology. *Cambridge University Press*.

ALEXOPOULOS, C.J., MIMS, C.W. & BLACKWELL, M. (1996) Introductory Mycology. 4th Edition. *Wiley, New York.*

ANDERSON, S.A., STEWART, A. & ALLEN, G.T. (1995) Pseudosphaerita euglenae, a fungal parasite of euglena spp. in the mangere oxidation ponds, auckland, New Zealand. *New Zealand Journal of Marine and Freshwater Research*.

ANDREWS, S. (2010) FastQC: A quality control tool for high throughput sequence data. *Http://Www.Bioinformatics.Babraham.Ac.Uk/Projects/Fastqc/*. .

ARCHER, W. (1869) Memoirs: On some Freshwater Rhizopoda, New or Little-Known. *Journal of Cell Science* **s2-9**, 250–271. The Company of Biologists Ltd.

ARCHIBALD, J.M., SIMPSON, A.G.B., SLAMOVITS, C.H., MARGULIS, L., MELKONIAN, M., CHAPMAN, D.J. & CORLISS, J.O. (eds) (2017) *Handbook of the Protists*. In *Handbook of the Protists* p. Springer International Publishing, Cham.

ARROYO, A.S., LÓPEZ-ESCARDÓ, D., KIM, E., RUIZ-TRILLO, I. & NAJLE, S.R. (2018) Novel diversity of deeply branching holomycota and unicellular holozoans revealed by metabarcoding in Middle Paraná River, Argentina. *Frontiers in Ecology and Evolution* **6**, 1–17.

ASATRYAN, A., BOUSSIBA, S. & ZARKA, A. (2019) Stimulation and isolation of Paraphysoderma sedebokerense(Blastocladiomycota) propagules and their infection capacity toward their host under different physiological and environmental conditions. *Frontiers in Cellular and Infection Microbiology*.

AVELAR, G.M., GLASER, T., LEONARD, G., RICHARDS, T.A., ULRICH, H. & GOMES, S.L. (2015) A cyclic GMP-dependent K+ channel in the blastocladiomycete fungus Blastocladiella emersonii. *Eukaryotic Cell* **14**, 958–963.

AVELAR, G.M., SCHUMACHER, R.I., ZAINI, P.A., LEONARD, G., RICHARDS, T.A. & GOMES, S.L. (2014) A Rhodopsin-Guanylyl cyclase gene fusion functions in visual perception in a fungus. *Current Biology* **24**, 1234–1240. Elsevier.

BALBIANI, E.G. (1882) Sur les microsporidies ou psorospermies des articules. *Compt. Rend. Acad. Sci. Paris* **95**, 1168–1171.

BALDAUF, S. & STEENKAMP, E. (2004) Origin and Evolution of Animals, Fungi and Their Unicellular Allies (Opisthokonta).

BALDAUF, S.L. & PALMERT, J.D. (1993) Animals and fungi are each other ' s closest relatives : Congruent evidence from multiple proteins **90**, 11558–11562.

BALDAUF, S.L., ROGER, A.J., WENK-SIEFERT, I. & DOOLITTLE, W.F. (2000) A kingdom-level phylogeny of eukaryotes based on combined protein data. *Science*.

BANKEVICH, A., NURK, S., ANTIPOV, D., GUREVICH, A.A., DVORKIN, M., KULIKOV, A.S., LESIN, V.M., NIKOLENKO, S.I., PHAM, S., PRJIBELSKI, A.D., PYSHKIN, A. V., SIROTKIN, A. V., VYAHHI, N., TESLER, G., ALEKSEYEV, M.A., ET AL. (2012) SPAdes: A New Genome Assembly Algorithm and Its Applications to Single-Cell Sequencing. *Journal of Computational Biology* **19**, 455–477.

BAPTESTE, E., BRINKMANN, H., LEE, J.A., MOORE, D. V., SENSEN, C.W., GORDON, P., DURUFLÉ, L., GAASTERLAND, T., LOPEZ, P., MÜLLER, M. & PHILIPPE, H. (2002) The analysis of 100 genes supports the grouping of three highly divergent amoebae: Dictyostelium, Entamoeba, and Mastigamoeba. *Proceedings of the National Academy of Sciences of the United States of America*.

BARR, D.J.S. (1980) An outline for the reclassification of the Chytridiales, and for a new order, the Spizellomycetales. *Canadian Journal of Botany* **58**, 2380–2394.

BARR, D.J.S. (1990) Phylum chytridiomycota. *Handbook of Protoctista*, 454–466. Jones and Bartlett.

BARR, D.J.S. (1992) Evolution and Kingdoms of Organisms from the Perspective of a Mycologist. *Mycologia*.

BARR, D.J.S. (2001) Chytridiomycota. In *Systematics and Evolution* pp. 93–112. Springer.

BARR, D.J.S. & HADLAND-HARTMANN, V.E. (1977) Zoospore ultrastructure of Olpidium cucurbitacearum (Chytridiales) . *Canadian Journal of Botany*.

BARRON, G.L. & SZIJARTO, E. (1986) A New Species of Olpidium Parasitic in Nematode Eggs. *Mycologia* **78**, 972–975. Mycological Society of America.

BARSTOW, W.E. & LOVETT, J.S. (1974) Apical vesicles and microtubules in rhizoids of Blastocladiella emersonii: Effects of actinomycin D and cycloheximide on development during germination. *Protoplasma*.

BARTNICKI-GARCIA, S. (1970) Cell wall composition and other biochemicel markers in fungal phylogeny. *Phytochemical phylogeny*.

BASS, D., CZECH, L., WILLIAMS, B.A.P., BERNEY, C., DUNTHORN, M., MAHÉ, F., TORRUELLA, G., STENTIFORD, G.D. & WILLIAMS, T.A. (2018) Clarifying the Relationships between Microsporidia and Cryptomycota.

Journal of Eukaryotic Microbiology **65**, 773–782.

BAUER, L.S. & PANKRATZ, H.S. (1993) Nosema scripta N. Sp. (Microsporida: Nosematidae), A Microsporidian Parasite of the Cottonwood Leaf Beetle, Chrysomela scripta (Coleoptera: Chrysomelidae). *Journal of Eukaryotic Microbiology*.

BENITES, L.F., POULTON, N., LABADIE, K., SIERACKI, M.E., GRIMSLEY, N. & PIGANEAU, G. (2019) Single cell ecogenomics reveals mating types of individual cells and ssDNA viral infections in the smallest photosynthetic eukaryotes. *Philosophical Transactions of the Royal Society B: Biological Sciences*.

BENJAMIN, R. (1965) Addenda to 'The Merosporangiferous Mucorales' III Dimargaris. *Aliso*.

BENNY, G.L., HO, H.M., LAZARUS, K.L. & SMITH, M.E. (2016a) Five new species of the obligate mycoparasite Syncephalis (Zoopagales, Zoopagomycotina) from soil. *Mycologia*.

BENNY, G.L., HUMBER, R.A. & VOIGT, K. (2014) Zygomycetous Fungi: Phylum Entomophthoromycota and Subphyla Kickxellomycotina, Mortierellomycotina, Mucoromycotina, and Zoopagomycotina. In *Systematics and Evolution: Part A: Second Edition* p.

BENNY, G.L., SMITH, M.E., KIRK, P.M., TRETTER, E.D. & WHITE, M.M. (2016b) Challenges and Future Perspectives in the Systematics of Kickxellomycotina, Mortierellomycotina, Mucoromycotina, and Zoopagomycotina.

BERBEE, M.L., JAMES, T.Y. & STRULLU-DERRIEN, C. (2017) Early Diverging Fungi: Diversity and Impact at the Dawn of Terrestrial Life. *Annual Review of Microbiology* **71**, 41–60.

BEZNOUSSENKO, G. V., DOLGIKH, V. V., SELIVERSTOVA, E. V., SEMENOV, P.B., TOKAREV, Y.S., TRUCCO, A., MICARONI, M., DI GIANDOMENICO, D., AUINGER, P., SENDERSKIY, I. V., SKARLATO, S.O., SNIGIREVSKAYA, E.S., KOMISSARCHIK, Y.Y., PAVELKA, M., DE MATTEIS, M.A., ET AL. (2007) Analogs of the Golgi complex in microsporidia: structure and avesicular mechanisms of function. *Journal of Cell Science* **120**, 1288–1298.

BIDARTONDO, M.I., READ, D.J., TRAPPE, J.M., MERCKX, V., LIGRONE, R. & DUCKETT, J.G. (2011) The dawn of symbiosis between plants and fungi. *Biology Letters*.

BLACKWELL, E. (1940) A life cycle of Blastocladia pringsheimii reinsch. *Transactions of the British Mycological Society* **24**, 68–86.

BLACKWELL, W.H., LETCHER, P.M. & POWELL, M.J. (2016) Reconsideration of the inclusiveness of genus Plasmophagus (Chytridiomycota, posteris traditus) based on morphology.

BLACKWELL, W.H., LETCHER, P.M. & POWELL, M.J. (2017) The taxa of Dictyomorpha (Chytridiomycota, in praesens tempus). *Phytologia* **99**, 74–82.

BLAKESLEE, A.F. (1904) Sexual Reproduction in the Mucorineae. *Proceedings of the American Academy of Arts and Sciences*.

BLANK, C.E. & SÁNCHEZ-BARACALDO, P. (2010) Timing of morphological and ecological innovations in the cyanobacteria - A key to understanding the rise in atmospheric oxygen. *Geobiology*.

BŁASZKOWSKI, J., CHWAT, G., GÓRALSKA, A. & BOBROWSKA-CHWAT, A. (2015) Glomus tetrastratosum, a new species of arbuscular mycorrhizal fungi (Glomeromycota). *Mycoscience*.

BOHNE, W., FERGUSON, D.J.P., KOHLER, K. & GROSS, U. (2000) Developmental expression of a tandemly repeated,

glycine- and serine- rich spore wall protein in the microsporidian pathogen Encephalitozoon cuniculi. *Infection and Immunity*.

BOLGER, A.M., LOHSE, M. & USADEL, B. (2014) Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120.

BONFANTE, P. & GENRE, A. (2010) Mechanisms underlying beneficial plant - Fungus interactions in mycorrhizal symbiosis. *Nature Communications*. .

BOUSSIBA, S. (2000) Carotenogenesis in the green alga Haematococcus pluvialis: Cellular physiology and stress response. *Physiologia Plantarum*. .

BRACKER, C.E. (1967) Ultrastructure of Fungi. *Annual Review of Phytopathology*.

BRAUN, A. (1851) Betrachtungen über die Erscheinung der Verjüngung in der Natur. *Leipzig*.

BROSSON, D., KUHN, L., PRENSIER, G., VIVARÈS, C.P. & TEXIER, C. (2005) The putative chitin deacetylase of Encephalitozoon cuniculi: A surface protein implicated in microsporidian spore-wall formation. *FEMS Microbiology Letters*.

BROWN, M.W., SPIEGEL, F.W. & SILBERMAN, J.D. (2009) Phylogeny of the 'forgotten' cellular slime mold, Fonticula alba, reveals a key evolutionary branch within Opisthokonta. *Molecular Biology and Evolution*.

BRUNS, T.D., VILGALYS, R., BARNS, S.M., GONZALEZ, D., HIBBETT, D.S., LANE, D.J., SIMON, L., STICKEL, S., SZARO, T.M., WEISBURG, W.G. & SOGIN, M.L. (1992) Evolutionary relationships within the fungi: Analyses of nuclear small subunit rRNA sequences. *Molecular Phylogenetics and Evolution*.

BURKI, F., ROGER, A.J., BROWN, M.W. & SIMPSON, A.G.B. (2019) The New Tree of Eukaryotes. *Trends in Ecology & Evolution* **xx**, 1–13. Elsevier.

BURKI, F., SHALCHIAN-TABRIZI, K., MINGE, M., SKJÆVELAND, Å., NIKOLAEV, S.I., JAKOBSEN, K.S. & PAWLOWSKI, J. (2007) Phylogenomics reshuffles the eukaryotic supergroups. *PLoS ONE*.

CAMACHO, C., COULOURIS, G., AVAGYAN, V., MA, N., PAPADOPOULOS, J., BEALER, K. & MADDEN, T.L. (2009) BLAST+: Architecture and applications. *BMC Bioinformatics*.

DEL CAMPO, J., MALLO, D., MASSANA, R., DE VARGAS, C., RICHARDS, T.A. & RUIZ-TRILLO, I. (2015) Diversity and distribution of unicellular opisthokonts along the European coast analysed using high-throughput sequencing. *Environmental Microbiology* **17**, 3195–3207.

DEL CAMPO, J. & RUIZ-TRILLO, I. (2013) Environmental survey meta-analysis reveals hidden diversity among unicellular opisthokonts. *Molecular Biology and Evolution* **30**, 802–805.

DEL CAMPO, J., SIERACKI, M.E., MOLESTINA, R., KEELING, P., MASSANA, R. & RUIZ-TRILLO, I. (2014) The others: Our biased perspective of eukaryotic genomes. *Trends in Ecology and Evolution* **29**, 252–259.

CANN, J.P. & PAGE, F.C. (1979) Nucleosphaerium tuckery nov. ge. nov. sp - A New Freshwater Filose Amoeba Without Motile Form in a New Family Nucleariidae (Filosea: Aconchulinida) Feeding by Ingestion Only. *Archiv für Protistenkunde* **122**, 226–240. VEB Gustav Fischer Verlag Jena.

CANNON, P.F. & HAWKSWORTH, D.L. (1995) The diversity of fungi associated with vascular plants: The known, the unknown and the need to bridge the knowledge gap. *Advances in Plant Pathology*.

CAPELLA-GUTIÉRREZ, S., MARCET-HOUBEN, M. & GABALDÓN, T. (2012) Phylogenomics supports microsporidia as

the earliest diverging clade of sequenced fungi. *BMC Biology* **10**.

CAPELLA-GUTIÉRREZ, S., SILLA-MARTÍNEZ, J.M. & GABALDÓN, T. (2009) trimAl: A tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* **25**, 1972–1973.

CARR, M., LEADBEATER, B.S.C., HASSAN, R., NELSON, M. & BALDAUF, S.L. (2008) Molecular phylogeny of choanoflagellates , the sister group to Metazoa, 7–9.

CAULLERY, M. & MESNIL, F. (1897) Sur un type nouveau (Metchnikovella n.g.) d'organismes parasites des Grégarines. *Comptes Rendus de Seances de la Societe de Biologie* **49**, 960–962. Paris.

CAULLERY, M. & MESNIL, F. (1914) Metchnikovellidae et autres Protistes parasites des Grégarines d'Annélides. *CR Soc. biol* **77**, 527–532.

CAULLERY, M. & MESNIL, F. (1919) Metchnikovellidae et autre Protistes des Grégarines d'Annélides. *Annales de l'Institut Pasteur* **33**, 209–240. Paris.

CAVALIER-SMITH, T. (2013) Early evolution of eukaryote feeding modes , cell structural diversity , and classification of the protozoan phyla Loukozoa , Sulcozoa , and Choanozoa. *European Journal of Protistology* **49**, 115–178. Elsevier GmbH.

CAVALIER-SMITH, T. (1983) A 6-Kingdom Classification And A Unified Phylogeny. In *Intracellular space as oligogenetic ecosystem. Proceedings* p.

CAVALIER-SMITH, T. (1987) The origin of Fungi and pseudofungi. *Evolutionary Biology of the FUngi, British Mycological Society Symposium* **12**, 339–353. Cambridge Univ. Press.

CAVALIER-SMITH, T. (1993a) Kingdom protozoa and its 18 phyla. *Microbiological Reviews*. .

CAVALIER-SMITH, T. (1993b) Kingdom protozoa and its 18 phyla. *Microbiological reviews* **57**, 953–994.

CAVALIER-SMITH, T. (1998) A revised six-kingdom system of life. *Biological Reviews of the Cambridge Philosophical Society* **73**, 203–266.

CAVALIER-SMITH, T. & CHAO, E.E.Y. (2003) Phylogeny of choanozoa, apusozoa, and other protozoa and early eukaryote megaevolution. *Journal of Molecular Evolution*.

CHAMBOUVET, A., MONIER, A., MAGUIRE, F., ITOÏZ, S., DEL CAMPO, J., ELIES, P., EDVARDSEN, B., EIKREIM, W. & RICHARDS, T.A. (2019) Intracellular Infection of Diverse Diatoms by an Evolutionary Distinct Relative of the Fungi. *Current Biology*.

CHANG, Y., WANG, S., SEKIMOTO, S., AERTS, A.L., CHOI, C., CLUM, A., LABUTTI, K.M., LINDQUIST, E.A., NGAN, C.Y., OHM, R.A., SALAMOV, A.A., GRIGORIEV, I. V., SPATAFORA, J.W. & BERBEE, M.L. (2015) Phylogenomic analyses indicate that early fungi evolved digesting cell walls of algal ancestors of land plants. *Genome Biology and Evolution* **7**, 1590–1601.

CIENKOWSKI, L. (1865) Beiträge zur Kenntnis der Monaden. *Archiv für mikroskopische Anatomie.*, 203–232.

CLARK, H.J. (1866) Conclusive proofs of the animality of the ciliate sponges, and of their affinities with the Infusoria flagellata. *American Journal of Science*.

COLE, F. . (1926) The History of Protozoology. *Nature* **117**, 650–651.

CORNU, M. (1872) Monographie des Saprolegniées. *Annales des Sciences Naturelles Botanique* **15**, 1–198.

CORRADI, N., POMBERT, J.F., FARINELLI, L., DIDIER, E.S. & KEELING, P.J. (2010) The complete sequence of the

smallest known nuclear genome from the microsporidian Encephalitozoon intestinalis. *Nature Communications* **1**.

CORRADI, N. & SLAMOVITS, C.H. (2011) The intriguing nature of microsporidian genomes. *Briefings in Functional Genomics*.

CORSARO, D., MICHEL, R., WALOCHNIK, J., VENDITTI, D., MÜLLER, K.D., HAURÖDER, B. & WYLEZICH, C. (2016) Molecular identification of Nucleophaga terricolae sp. nov. (Rozellomycota), and new insights on the origin of the Microsporidia. *Parasitology Research*.

CORSARO, D., WALOCHNIK, J., VENDITTI, D., MÜLLER, K., HAURÖDER, B. & MICHEL, R. (2014a) Rediscovery of Nucleophaga amoebae , a novel member of the Rozellomycota.

CORSARO, D., WALOCHNIK, J., VENDITTI, D., STEINMANN, J., MÜLLER, K.D. & MICHEL, R. (2014b) Microsporidia-like parasites of amoebae belong to the early fungal lineage Rozellomycota. *Parasitology Research* **113**, 1909–1918.

CORSARO, D., WYLEZICH, C., VENDITTI, D., MICHEL, R., WALOCHNIK, J. & WEGENSTEINER, R. (2019) Filling gaps in the microsporidian tree: rDNA phylogeny of Chytridiopsis typographi (Microsporidia: Chytridiopsida). *Parasitology Research* **118**, 169–180. Parasitology Research.

COUCH, J.N. (1945) Observations on the Genus Catenaria. *Mycologia*.

COUCH, J.N. & WHIFFEN, A.J. (1942) Observations on the Genus Blastocladiella. *American Journal of Botany*.

COX, F., BARSOUM, N., LILLESKOV, E.A. & BIDARTONDO, M.I. (2010) Nitrogen availability is a primary determinant of conifer mycorrhizas across complex environmental gradients. *Ecology Letters*.

CUOMO, C.A., DESJARDINS, C.A., BAKOWSKI, M.A., GOLDBERG, J., MA, A.T., BECNEL, J.J., DIDIER, E.S., FAN, L., HEIMAN, D.I., LEVIN, J.Z., YOUNG, S., ZENG, Q. & TROEMEL, E.R. (2012) Microsporidian genome analysis reveals evolutionary strategies for obligate intracellular growth. *Genome Research*.

DANGEARD, P.-A. (1895) Mémoire sur les parasites du noyau et du protoplasme. *Le Botaniste* **4**, 199–248.

DAS, S., VAN DELLEN, K., BULIK, D., MAGNELLI, P., CUI, J., HEAD, J., ROBBINS, P.W. & SAMUELSON, J. (2006) The cyst wall of Entamoeba invadens contains chitosan (deacetylated chitin). *Molecular and Biochemical Parasitology*.

DAVIS, W.J., AMSES, K.R., BENNY, G.L., CARTER-HOUSE, D., CHANG, Y., GRIGORIEV, I., SMITH, M.E., SPATAFORA, J.W., STAJICH, J.E. & JAMES, T.Y. (2019a) Genome-scale phylogenetics reveals a monophyletic Zoopagales (Zoopagomycota, Fungi). *Molecular Phylogenetics and Evolution*.

DAVIS, W.J., AMSES, K.R., JAMES, E.S. & JAMES, T.Y. (2019b) A new 18S rRNA phylogeny of uncultured predacious fungi (Zoopagales). *Mycologia*.

DEARNALEY, J.D.W., MARTOS, F. & SELOSSE, M.A. (2012) Orchid mycorrhizas: Molecular ecology, physiology, evolution and conservation aspects. In *Fungal Associations, 2nd Edition* p.

DEE, J.M., MOLLICONE, M., LONGCORE, J.E., ROBERSON, R.W. & BERBEE, M.L. (2015) Cytology and molecular phylogenetics of Monoblepharidomycetes provide evidence for multiple independent origins of the hyphal habit in the Fungi. *Mycologia*.

DEGAWA, Y. (2014) Verrucocephalum, a new nematophagous genus in the Helicocephalidaceae (Zoopagales).

Mycoscience.

DELWICHE, C.F. & COOPER, E.D. (2015) The evolutionary origin of a terrestrial flora. *Current Biology*. .

DERELLE, R., LÓPEZ-GARCÍA, P., TIMPANO, H. & MOREIRA, D. (2016) A phylogenomic framework to study the diversity and evolution of stramenopiles (=heterokonts). *Molecular Biology and Evolution* **33**, msw168.

DICK, M.W. (1999) FUNGI | Classification of the Peronosporomycetes. In *Encyclopedia of Food Microbiology* p.

DICK, M.W. (2001) *Straminipilous fungi : systematics of the peronosporomycetes, including accounts of the marine straminipilous protists, the plasmodiophorids, and similar organisms*. Kluwer Academic Publishers, Dordrecht; Boston.

DIRREN, S., PITSCH, G., SILVA, M.O.D. & POSCH, T. (2017) Grazing of Nuclearia thermophila and Nuclearia delicatula (Nucleariidae , Opisthokonta) on the toxic cyanobacterium Planktothrix rubescens. *European Journal of Protistology* **60**, 87–101. Elsevier GmbH.

DIRREN, S. & POSCH, T. (2016) Promiscuous and specific bacterial symbiont acquisition in the amoeboid genus Nuclearia (Opisthokonta), 1–16.

DOBELL, C. (1911) *The Principles of Protistology*. G. Fischer.

DOBELL, C.C. (1932) Antony van Leeuwenhoek and his "Little Animals": being some Account of the Father of Protozoology and Bacteriology and his Multifarious Discoveries in these Disciplines. *Nature* **130**, 679–680.

DÖFLEIN, F. (1901) Die Protozoen als Parasiten und Krankheitserreger nach biologischen Gesichtspunkten dargestellt. *Botanische Zeitung*, 296. Jena,Fischer.

DOGGETT, M.S. & PORTER, D. (1996) Sexual reproduction in the fungal parasite, Zygorhizidium planktonicum. *Mycologia*.

DONACHIE, S., SUGA, H. & GRAU-BOVE, X. (2017) Dynamics of genomic innovation in the unicellular ancestry of animals Dynamics of genomic innovation in the unicellular ancestry of animals.

DOUZERY, E.J.P., SNELL, E.A., BAPTESTE, E., DELSUC, F. & PHILIPPE, H. (2004) The timing of eukaryotic evolution: Does a relaxed molecular clock reconcile proteins and fossils? *Proceedings of the National Academy of Sciences of the United States of America*.

DOWELD, A.B. (2001) *Prosyllabus Tracheophytorum. Tentamen systematis plantarum vascularium (Tracheophyta)*. GEOS, Moscow.

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. *Molecular Biology and Evolution* **29**, 1319–1334.

ELIAS, M., BRIGHOUSE, A., GABERNET-CASTELLO, C., FIELD, M.C. & DACKS, J.B. (2012) Sculpting the endomembrane system in deep time: High resolution phylogenetics of Rab GTPases. *Journal of Cell Science*.

EMBLEY, T.M. & HIRT, R.P. (1998) Early branching eukaryotes? *Current Opinion in Genetics and Development*.

EMBLEY, T.M. & MARTIN, W. (2006) Eukaryotic evolution, changes and challenges. *Nature* **440**, 623–630.

EME, L., SHARPE, S.C., BROWN, M.W. & ROGER, A.J. (2014) On the Age of Eukaryotes: Evaluating Evidence from Fossils and Molecular Clocks. *Cold Spring Harbor Perspectives in Biology* **6**, 1–16.

EMERSON, R. (1941) An experimental study of the life cycles and taxonomy of allomyces. *Lloydia*. .

EMMS, D.M. & KELLY, S. (2015) OrthoFinder: solving fundamental biases in whole genome comparisons

dramatically improves orthogroup inference accuracy. *Genome Biology* **16**.

ESTEBAN, G.F., GOODAY, A.J. & CLARKE, K.J. (2007) Siliceous scales of filose-amoebae (Pompholyxophryidae, Rotosphaerida) from deep Southern Ocean sediments, including first records for the Southern Hemisphere. *Polar Biology* **30**, 945–950.

FABEL, P., RADEK, R. & STORCH, V. (2000) A new spore-forming protist, Nephridiophaga blaberi sp. nov., in the Death's head cockroach Blaberus craniifer. *European Journal of Protistology*.

FEENEY, D.M. & TRIEMER, R.E. (1979) Cytochemical localization of Golgi maker enzymes in Allomyces. *Experimental Mycology*.

FELSENSTEIN, J. (1981) Evolutionary trees from DNA sequences: A maximum likelihood approach. *Journal of Molecular Evolution* **17**, 368–376.

FISCHER, W.M. & PALMER, J.D. (2005) Evidence from small-subunit ribosomal RNA sequences for a fungal origin of Microsporidia. *Molecular Phylogenetics and Evolution*.

FISHER, K.E., ROMBERGER, I., LOWRY, D., SHANGE, P. & ROBERSON, R.W. (2018) Hyphal tip growth and cytoplasmic characters of Conidiobolus coronatus (Zoopagomycota, Entomophthoromycotina). *Mycologia*.

FITZPATRICK, H.M. (1930) *The lower fungi Phycomycetes*. New York,McGraw-Hill,.

FORGET, L., USTINOVA, J., WANG, Z., HUSS, V.A.R. & LANG, B.F. (2002) Hyaloraphidium curvatum: A linear mitochondrial genome, tRNA editing, and an evolutionary link to lower fungi. *Molecular Biology and Evolution*.

FORSTER, H., COFFEY, M.D., ELWOOD, H. & SOGIN, M.L. (1990) Sequence Analysis of the Small Subunit Ribosomal RNAs of Three Zoosporic Fungi and Implications for Fungal Evolution. *Mycologia*.

FOUST, F.K. (1937) A new species of Rozella parasitic on Allomyces. *Journal of the Elisha Mitchell Scientific Society* **53**, 197–204. Temporary Publisher.

FREEMAN, K.R., MARTIN, A.P., KARKI, D., LYNCH, R.C., MITTER, M.S., MEYERA, A.F., LONGCORE, J.E., SIMMONS, D.R. & SCHMIDT, S.K. (2009) Evidence that chytrids dominate fungal communities in high-elevation soils. *Proceedings of the National Academy of Sciences of the United States of America*.

FREEMAN, M.A., KASPER, J.M. & KRISTMUNDSSON, Á. (2013) Nucleospora cyclopteri n. sp., an intranuclear microsporidian infecting wild lumpfish, Cyclopterus lumpus L., in Icelandic waters. *Parasites and Vectors*.

FRISVAD, J.C., HUBKA, V., EZEKIEL, C.N., HONG, S.B., NOVÁKOVÁ, A., CHEN, A.J., ARZANLOU, M., LARSEN, T.O., SKLENÁŘ, F., MAHAKARNCHANAKUL, W., SAMSON, R.A. & HOUBRAKEN, J. (2019) Taxonomy of Aspergillus section Flavi and their production of aflatoxins, ochratoxins and other mycotoxins. *Studies in Mycology*.

GAWAD, C., KOH, W. & QUAKE, S.R. (2016) Single-cell genome sequencing: current state of the science. *Nat Rev Genet* **17**, 175–188. Nature Publishing Group.

GAWRYLUK, R.M.R., DEL CAMPO, J., OKAMOTO, N., STRASSERT, J.F.H., LUKEŠ, J., RICHARDS, T.A., WORDEN, A.Z., SANTORO, A.E. & KEELING, P.J. (2016) Morphological Identification and Single-Cell Genomics of Marine Diplonemids. *Current Biology*.

VAN GELDER, R.N., VON ZASTROW, M.E., YOOL, A., DEMENT, W.C., BARCHAS, J.D. & EBERWINE, J.H. (1990) Amplified RNA synthesized from limited quantities of heterogeneous cDNA. *Proceedings of the National*

Academy of Sciences of the United States of America.

GLEASON, F.H., CARNEY, L.T., LILJE, O. & GLOCKLING, S.L. (2012) Ecological potentials of species of Rozella (Cryptomycota). *Fungal Ecology* **5**, 651–656. Elsevier Ltd.

GLOCKLING, S.L., MARSHALL, W.L. & GLEASON, F.H. (2013) Phylogenetic interpretations and ecological potentials of the Mesomycetozoea (Ichthyosporea). *Fungal Ecology*, 1–11. Elsevier Ltd.

GOFFEAU, A., BARRELL, G., BUSSEY, H., DAVIS, R.W., DUJON, B., FELDMANN, H., GALIBERT, F., HOHEISEL, J.D., JACQ, C., JOHNSTON, M., LOUIS, E.J., MEWES, H.W., MURAKAMI, Y., PHILIPPSEN, P., TETTELIN, H., ET AL. (1996) Life with 6000 genes. *Science*.

GONÇALVES, V.N., CANTRELL, C.L., WEDGE, D.E., FERREIRA, M.C., SOARES, M.A., JACOB, M.R., OLIVEIRA, F.S., GALANTE, D., RODRIGUES, F., ALVES, T.M.A., ZANI, C.L., JUNIOR, P.A.S., MURTA, S., ROMANHA, A.J., BARBOSA, E.C., ET AL. (2016) Fungi associated with rocks of the Atacama Desert: Taxonomy, distribution, diversity, ecology and bioprospection for bioactive compounds. *Environmental Microbiology*.

GORDETSKAYA, O., MESENTSEV, Y., KAMYSHATSKAYA, O., MICHEL, R., WALOCHNIK, J., ALEXEY, S. & NASSONOVA, E. (2019) Real-time observations on the development of intra- nuclear parasite Nucleophaga amoebae (Rozello- mycota) in the culture of Thecamoeba quadrilineata. *Protistology* **13**, 236–245.

GREEFF, R. (1873) Über Radiolarien und radiolarienartige Rhizopoden des süßen Wassers. In *Sitzungsberichte der gesellschaft zur beforderung der gesammten naturwissenschafte zu Marburg* pp. 47–64.

GROMOV, B. V. (2000) Algal parasites of the genera Aphelidium, Amoeboaphelidium, and Pseudaphelidium from the cienkovski's 'monadinea' group as representatives of a new class. *Zoologicheskii Zhurnal*.

GROSSART, H.P., WURZBACHER, C., JAMES, T.Y. & KAGAMI, M. (2016) Discovery of dark matter fungi in aquatic ecosystems demands a reappraisal of the phylogeny and ecology of zoosporic fungi. *Fungal Ecology*.

GRUNINGER, R.J., PUNIYA, A.K., CALLAGHAN, T.M., EDWARDS, J.E., YOUSSEF, N., DAGAR, S.S., FLIEGEROVA, K., GRIFFITH, G.W., FORSTER, R., TSANG, A., MCALLISTER, T. & ELSHAHED, M.S. (2014) Anaerobic fungi (phylum Neocallimastigomycota): Advances in understanding their taxonomy, life cycle, ecology, role and biotechnological potential. *FEMS Microbiology Ecology*.

GRYGANSKYI, A.P., HUMBER, R.A., SMITH, M.E., MIADLIKOVSKA, J., WU, S., VOIGT, K., WALTHER, G., ANISHCHENKO, I.M. & VILGALYS, R. (2012) Molecular phylogeny of the Entomophthoromycota. *Molecular Phylogenetics and Evolution*.

GUREVICH, A., SAVELIEV, V., VYAHHI, N. & TESLER, G. (2013) QUAST: Quality assessment tool for genome assemblies. *Bioinformatics* **29**, 1072–1075.

HAAG, K.L., JAMES, T.Y., POMBERT, J.-F., LARSSON, R., SCHAER, T.M.M., REFARDT, D. & EBERT, D. (2014) Evolution of a morphological novelty occurred before genome compaction in a lineage of extreme parasites. *Proceedings of the National Academy of Sciences* **111**, 15480–15485.

HACKER, C., HOWELL, M., BHELLA, D. & LUCOCQ, J. (2014) Strategies for maximizing ATP supply in the microsporidian Encephalitozoon cuniculi: Direct binding of mitochondria to the parasitophorous vacuole and clustering of the mitochondrial porin VDAC. *Cellular Microbiology*.

HAECKEL, E. (1866) *Generelle morphologie der organismen. Allgemeine grundzüge der organischen formen-*

wissenschaft, mechanisch begründet durch die von Charles Darwin reformirte descendenztheorie,. Berlin,G. Reimer,.

HAECKEL, E. (1874) Die Gastraea-Theorie, die phylogenetische Classification des Thierreichs und die Homologie der Keimblatter. *Jenaische Z f Naturwiss* **8**, 1–55.

HAECKEL ERNST (1862) *Die Radiolarien (Rhizopoda radiaria) : eine Monographie /*. Berlin :G. Reimer,.

HALL, R.P. (1953) *Protozoology*. New York,Prentice-Hall,.

HANSON, A.M. (1946) A Morphological, Developmental and Cytological Study of Four Saprophytic chytrids. IV. Phlyctorhiza endogena Gen. Nov. Sp. Nov. *American Journal of Botany*.

HARRIS, S.D. (2011) Hyphal morphogenesis: An evolutionary perspective. *Fungal Biology*.

HATCH, W.R. (1935) Gametogenesis in allomyces arbuscula. *Annals of Botany*.

HAUSNER, G., BELKHIRI, A. & KLASSEN, G.R. (2000) Phylogenetic analysis of the small subunit ribosomal RNA gene of the hyphochytrid Rhizidiomyces apophysatus. *Canadian Journal of Botany*.

HAWKSWORTH, D.L. (1991) The fungal dimension of biodiversity: magnitude, significance, and conservation. *Mycological Research*.

HAWKSWORTH, D.L. & LÜCKING, R. (2017) Fungal Diversity Revisited: 2.2 to 3.8 Million Species. *Microbiology Spectrum* **5**.

HAYMAN, J.R., HAYES, S.F., AMON, J. & NASH, T.E. (2001) Developmental expression of two spore wall proteins during maturation of the microsporidian Encephalitozoon intestinalis. *Infection and Immunity*.

HEGER, T.J., GIESBRECHT, I.J.W., GUSTAVSEN, J., DEL CAMPO, J., KELLOGG, C.T.E., HOFFMAN, K.M., LERTZMAN, K., MOHN, W.W. & KEELING, P.J. (2018) High-throughput environmental sequencing reveals high diversity of litter and moss associated protist communities along a gradient of drainage and tree productivity. *Environmental Microbiology* **20**, 1185–1203.

HEHENBERGER, E., TIKHONENKOV, D. V, KOLISKO, M., CAMPO, J., ESAULOV, A.S., MYLNIKOV, A.P., KEELING, P.J., HEHENBERGER, E., TIKHONENKOV, D. V, KOLISKO, M., CAMPO, J., ESAULOV, A.S., MYLNIKOV, A.P. & KEELING, P.J. (2017) Novel Predators Reshape Holozoan Phylogeny and Reveal the Presence of a Two-Component Signaling System in the Ancestor of Animals Report Novel Predators Reshape Holozoan Phylogeny and Reveal the Presence of a Two-Component Signaling System in the Ancestor. *Current Biology*, 1–8. Elsevier Ltd.

HEILBRON, J.L. (2003) *The oxford companion to the history of modern science*. In *The Oxford Companion to the History of Modern Science* p.

HEINZ, E., HACKER, C., DEAN, P., MIFSUD, J., GOLDBERG, A. V., WILLIAMS, T.A., NAKJANG, S., GREGORY, A., HIRT, R.P., LUCOCQ, J.M., KUNJI, E.R.S. & EMBLEY, T.M. (2014) Plasma Membrane-Located Purine Nucleotide Transport Proteins Are Key Components for Host Exploitation by Microsporidian Intracellular Parasites. *PLoS Pathogens*.

HEINZ, E., WILLIAMS, T.A., NAKJANG, S., NOËL, C.J., SWAN, D.C., GOLDBERG, A. V., HARRIS, S.R., WEINMAIER, T., MARKERT, S., BECHER, D., BERNHARDT, J., DAGAN, T., HACKER, C., LUCOCQ, J.M., SCHWEDER, T., ET AL. (2012) The Genome of the Obligate Intracellular Parasite Trachipleistophora hominis: New Insights into

Microsporidian Genome Dynamics and Reductive Evolution. *PLoS Pathogens* **8**.

HELD, A.A. (1975) The zoospore of Rozella allomycis : ultrastructure . *Canadian Journal of Botany*.

HELD, A.A. (1981) Rozella and Rozellopsis: Naked endoparasitic fungi which dress-up as their hosts. *The Botanical Review*.

HERR, R.A., AJELLO, L., TAYLOR, J.W. & ARSECULERATNE, S.N. (1999) Phylogenetic Analysis of Rhinosporidium seeberi ' s 18S Small- Subunit Ribosomal DNA Groups This Pathogen among Members of the Protoctistan Mesomycetozoa Clade **37**, 2750–2754.

HERTEL, L.A., BAYNE, C.J. & LOKER, E.S. (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 q **32**, 1183–1191.

HIBBETT, D.S., BINDER, M., BISCHOFF, J.F., BLACKWELL, M., CANNON, P.F., ERIKSSON, O.E., HUHNDORF, S., JAMES, T., KIRK, P.M., LÜCKING, R., THORSTEN LUMBSCH, H., LUTZONI, F., MATHENY, P.B., MCLAUGHLIN, D.J., POWELL, M.J., ET AL. (2007) A higher-level phylogenetic classification of the Fungi. *Mycological Research*.

HIBBETT, D.S., BLACKWELL, M., JAMES, T.Y., SPATAFORA, J.W., TAYLOR, J.W. & VILGALYS, R. (2018) Phylogenetic taxon definitions for Fungi, Dikarya, Ascomycota and Basidiomycota. *IMA Fungus*.

HIRT, R.P., LOGSDON, J.M., HEALY, B., DOREY, M.W., DOOLITTLE, W.F. & EMBLEY, T.M. (1999) Microsporidia are related to fungi: Evidence from the largest subunit of RNA polymerase II and other proteins. *Proceedings of the National Academy of Sciences of the United States of America*.

HO, Y.W. & BARR, D.J.S. (1995) Classification of Anaerobic Gut Fungi from Herbivores with Emphasis on Rumen Fungi from Malaysia. *Mycologia*.

HOFF, J.A., KLOPFENSTEIN, N.B., MCDONALD, G.I., TONN, J.R., KIM, M.S., ZAMBINO, P.J., HESSBURG, P.F., ROGERS, J.D., PEEVER, T.L. & CARRIS, L.M. (2004) Fungal endophytes in woody roots of Douglas-fir (Pseudotsuga menziesii) and ponderosa pine (Pinus ponderosa). *Forest Pathology*.

HOFFMANN, K., PAWŁOWSKA, J., WALTHER, G., WRZOSEK, M., DE HOOG, G.S., BENNY, G.L., KIRK, P.M. & VOIGT, K. (2013) The family structure of the Mucorales: A synoptic revision based on comprehensive multigene-genealogies. *Persoonia: Molecular Phylogeny and Evolution of Fungi*.

HONIGBERG, B.M., BALAMUTH, W., BOVEE, E.C., CORLISS, J.O., GOJDICS, M., HALL, R.P., KUDO, R.R., LEVINE, N.D., LOBBLICH, A.R., WEISER, J. & WENRICH, D.H. (1964) A Revised Classification of the Phylum Protozoa. *The Journal of Protozoology*.

HOOKE, R. (1665) *Micrographia, or, Some physiological descriptions of minute bodies made by magnifying glasses :with observations and inquiries thereupon /*. London :Printed by Jo. Martyn and Ja. Allestry, printers to the Royal Society ... ,.

HUANG, Y.H., HUANG, C.T. & HSEU, R.S. (2005) Effects of dockerin domains on Neocallimastix frontalis xylanases. *FEMS Microbiology Letters*.

HUELSENBECK, J.P. & RONQUIST, F. (2001) MR BAYES: Bayesian inference of phylogenetic trees. *Bioinformatics* **17**, 754–755.

HUERTA-CEPAS, J., SZKLARCZYK, D., FORSLUND, K., COOK, H., HELLER, D., WALTER, M.C., RATTEI, T., MENDE,

D.R., SUNAGAWA, S., KUHN, M., JENSEN, L.J. & MERING, C. VON (2016) eggNOG 4 . 5 : a hierarchical orthology framework with improved functional annotations for eukaryotic , prokaryotic and viral sequences **44**, 286–293.

HUMBER, R.A. (2012) Entomophthoromycota: A new phylum and reclassification for entomophthoroid fungi. *Mycotaxon*.

HUTCHINSON, G.E. (1961) The Paradox of the Plankton. *The American Naturalist*.

IMACHI, H., NOBU, M.K., NAKAHARA, N., MORONO, Y., OGAWARA, M., TAKAKI, Y., TAKANO, Y., UEMATSU, K., IKUTA, T., ITO, M., MATSUI, Y., MIYAZAKI, M., MURATA, K., SAITO, Y., SAKAI, S., ET AL. (2020) Isolation of an archaeon at the prokaryote–eukaryote interface. *Nature* **577**, 519–525.

ISHOEY, T., WOYKE, T., STEPANAUSKAS, R., NOVOTNY, M. & LASKEN, R.S. (2008) Genomic sequencing of single microbial cells from environmental samples. *Current Opinion in Microbiology*. .

JAMES, T.Y. & 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.

JAMES, T.Y., KAUFF, F., SCHOCH, C.L., MATHENY, P.B., HOFSTETTER, V., COX, C.J., CELIO, G., GUEIDAN, C., FRAKER, E., MIADLIKOWSKA, J., LUMBSCH, H.T., RAUHUT, A., REEB, V., ARNOLD, A.E., AMTOFT, A., ET AL. (2006a) Reconstructing the early evolution of Fungi using a six-gene phylogeny. *Nature* **443**, 818–822.

JAMES, T.Y., LETCHER, P.M., LONGCORE, J.E., MOZLEY-STANDRIDGE, S.E., PORTER, D., POWELL, M.J., GRIFFITH, G.W. & VILGALYS, R. (2006b) A molecular phylogeny of the flagellated fungi (Chytridiomycota) and description of a new phylum (Blastocladiomycota). *Mycologia* **98**, 860–871.

JAMES, T.Y., PELIN, A., BONEN, L., AHRENDT, S., SAIN, D., CORRADI, N. & STAJICH, J.E. (2013a) Report Shared Signatures of Parasitism and Phylogenomics Unite Cryptomycota and Microsporidia. *Current Biology*, 1–6. Elsevier Ltd.

JAMES, T.Y., PELIN, A., BONEN, L., AHRENDT, S., SAIN, D., CORRADI, N. & STAJICH, J.E. (2013b) Shared signatures of parasitism and phylogenomics unite cryptomycota and microsporidia. *Current Biology* **23**, 1548–1553. Elsevier Ltd.

JAMES, T.Y., PORTER, D., LEANDER, C.A., VILGALYS, R. & JOYCE, E. (2000) Molecular phylogenetics of the Chytridiomycota supports the utility of ultrastructural data in chytrid systematics **350**, 336–350.

JAMES, T.Y., PORTER, T.M. & MARTIN, W.W. (2014) Blastocladiomycota. *Systematics and Evolution*, 177–207.

JEFFRIES, P. & YOUNG, T.W.K. (1979) Ultrastructure of Septa in Dimargaris cristalligena R. K. Benjamin. *Journal of General Microbiology*.

JONES, E.B.G. (2012) *Marine Fungi and Fungal-like Organisms*. In *Marine Fungi and Fungal-like Organisms* p.

JONES, M.D.M., FORN, I., GADELHA, C., EGAN, M.J., BASS, D., MASSANA, R. & RICHARDS, T.A. (2011a) Discovery of novel intermediate forms redefines the fungal tree of life. *Nature*.

JONES, M.D.M., RICHARDS, T.A., HAWKSWORTH, D.L. & BASS, D. (2011b) Validation and justification of the phylum name Cryptomycota phyl. nov. *IMA Fungus* **2**, 173–175.

KAGAMI, M., MIKI, T. & TAKIMOTO, G. (2014) Mycoloop: Chytrids in aquatic food webs. *Frontiers in Microbiology*.

KAMAISHI, T., HASHIMOTO, T., NAKAMURA, Y., MASUDA, Y., NAKAMURA, F., OKAMOTO, K. -I., SHIMIZU, M. & HASEGAWA, M. (1996) Complete Nucleotide Sequences of the Genes Encoding Translation Elongation Factors 1 and 2 from a microsporidian parasite, Glugea plecoglossi: Implications for the Deepest Branching of Eukaryotes. *Journal of Biochemistry* **120**, 1095–1103.

KANEHISA, M. (2000) KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Research*.

KANTOR, R.S., ZYL, A.W. VAN, HILLE, R.P. VAN, THOMAS, B.C., HARRISON, S.T.L., BANFIELD, J.F., TOWN, C., TOWN, C. & AFRICA, S. (2015) Bioreactor microbial ecosystems for thiocyanate and cyanide degradation unravelled with genome-resolved metagenomics **17**, 4929–4941.

KARLING, J.S. (1972) The Present Status of Sphaerita, Pseudosphaerita, Morella and Nucleophaga. *Bulletin of the Torrey Botanical Club*.

KARPOV, S.A., LÓPEZ-GARCÍA, P., MAMKAEVA, M.A., KLIMOV, V.I., VISHNYAKOV, A.E., TCVETKOVA, V.S. & MOREIRA, D. (2018) The Chytrid-like Parasites of Algae Amoeboradix gromovi gen. et sp. nov. and Sanchytrium tribonematis Belong to a New Fungal Lineage. *Protist* **169**, 122–140. Elsevier GmbH.

KARPOV, S.A., MAMANAZAROVA, K.S., POPOVA, O. V., ALEOSHIN, V. V., JAMES, T.Y., MAMKAEVA, M.A., TCVETKOVA, V.S., VISHNYAKOV, A.E. & LONGCORE, J.E. (2017a) Monoblepharidomycetes diversity includes new parasitic and saprotrophic species with highly intronized rDNA. *Fungal Biology* **121**, 729–741. Elsevier Ltd.

KARPOV, S.A., MAMKAEVA, M.A., ALEOSHIN, V. V., NASSONOVA, E., LILJE, O. & GLEASON, F.H. (2014) Morphology, phylogeny, and ecology of the aphelids (Aphelidea, Opisthokonta) and proposal for the new superphylum Opisthosporidia. *Frontiers in Microbiology*. .

KARPOV, S.A., MIKHAILOV, K. V & MIRZAEVA, G.S. (2013) Obligately Phagotrophic Aphelids Turned out to Branch with the Earliest-diverging Fungi. *Annals of Anatomy* **164**, 195–205. Elsevier GmbH.

KARPOV, S.A., TCVETKOVA, V.S., MAMKAEVA, M.A. & MAMANAZAROVA, K.S. (2016) Morphological and Genetic Diversity of Opisthosporidia : New Aphelid Paraphelidium tribonemae gen . et sp . nov ., 1–9.

KARPOV, S.A., TCVETKOVA, V.S., MAMKAEVA, M.A., TORRUELLA, G., TIMPANO, H., MOREIRA, D., MAMANAZAROVA, K.S. & LÓPEZ-GARCÍA, P. (2017b) Morphological and Genetic Diversity of Opisthosporidia: New Aphelid Paraphelidium tribonemae gen. et sp. nov. *Journal of Eukaryotic Microbiology*.

KARPOV, S.A., TORRUELLA, G., MOREIRA, D., MAMKAEVA, M.A. & LÓPEZ-GARCÍA, P. (2017c) Molecular Phylogeny of Paraphelidium letcheri sp. nov. (Aphelida, Opisthosporidia). *Journal of Eukaryotic Microbiology* **64**, 573–578.

KARPOV, S.A., VISHNYAKOV, A.E., MOREIRA, D. & LÓPEZ-GARCÍA, P. (2019) The Ultrastructure of Sanchytrium tribonematis (Sanchytriaceae, Fungi incertae sedis) Confirms its Close Relationship to Amoeboradix. *Journal of Eukaryotic Microbiology*, 0–2.

KATINKA, M.D., DUPRAT, S., CORNILLOTT, E., MÉTÉNLER, G., THOMARAT, F., PRENSIER, G., BARBE, V., PEYRETAILLADE, E., BROTTIER, P., WINCKER, P., DELBAC, F., EL ALAOUL, H., PEYRET, P., SAURIN, W., GOUY, M., ET AL. (2001) Genome sequence and gene compaction of the eukaryote parasite Encephalitozoon cuniculi. *Nature* **414**, 450–453.

KATOH, K. & STANDLEY, D.M. (2013) MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. *Molecular Biology and Evolution* **30**, 772–780.

KEARSE, M., MOIR, R., WILSON, A., STONES-HAVAS, S., CHEUNG, M., STURROCK, S., BUXTON, S., COOPER, A., MARKOWITZ, S., DURAN, C., THIERER, T., ASHTON, B., MEINTJES, P. & DRUMMOND, A. (2012) Geneious Basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics*.

KEELING, P.J. (2003) Congruent evidence from α-tubulin and β-tubulin gene phylogenies for a zygomycete origin of microsporidia. *Fungal Genetics and Biology*.

KEELING, P.J., BURGER, G., DURNFORD, D.G., LANG, B.F., LEE, R.W., PEARLMAN, R.E., ROGER, A.J. & GRAY, M.W. (2005) The tree of eukaryotes **20**.

KEELING, P.J. & BURKI, F. (2019) Review Progress towards the Tree of Eukaryotes. *Current Biology* **29**, R808–R817. Elsevier Ltd.

KEELING, P.J., LUKER, M.A. & PALMER, J.D. (1999) Evidence from Beta-Tubulin Phylogeny that Microsporidia Evolved from Within the Fungi, 23–31.

KERWIN, J.L. & PETERSEN, E.E. (1997) Fungi: oomycetes and chytridiomycetes. In *Manual of techniques in insect pathology* pp. 251–268. Elsevier.

KIMURA, M. (1983) *The neutral theory of molecular evolution*. Cambridge University Press.

KING, N., WESTBROOK, M.J., YOUNG, S.L., KUO, A., ABEDIN, M., CHAPMAN, J., FAIRCLOUGH, S., HELLSTEN, U., ISOGAI, Y., LETUNIC, I., MARR, M., PINCUS, D., PUTNAM, N., ROKAS, A., WRIGHT, K.J., ET AL. (2008) The genome of the choanoflagellate Monosiga brevicollis and the origin of metazoans **451**, 783–788.

KING, R., URBAN, M., LAUDER, R.P., HAWKINS, N., EVANS, M., PLUMMER, A., HALSEY, K., LOVEGROVE, A., HAMMOND-KOSACK, K. & RUDD, J.J. (2017) A conserved fungal glycosyltransferase facilitates pathogenesis of plants by enabling hyphal growth on solid surfaces. *PLoS Pathogens*.

KISS, E., HEGEDÜS, B., VIRÁGH, M., VARGA, T., MERÉNYI, Z., KÓSZÓ, T., BÁLINT, B., PRASANNA, A.N., KRIZSÁN, K., KOCSUBÉ, S., RIQUELME, M., TAKESHITA, N. & NAGY, L.G. (2019) Comparative genomics reveals the origin of fungal hyphae and multicellularity. *Nature Communications* **10**.

KOLISKO, M., BOSCARO, V., BURKI, F., LYNN, D.H. & KEELING, P.J. (2014) Single-cell transcriptomics for microbial eukaryotes. *Current Biology*. .

KOSUGE, T. & HOSHINO, T. (1999) The α-aminoadipate pathway for lysine biosynthesis is widely distributed among Thermus strains. *Journal of Bioscience and Bioengineering*.

KOUMANDOU, V.L., DACKS, J.B., COULSON, R.M.R. & FIELD, M.C. (2007) Control systems for membrane fusion in the ancestral eukaryote; Evolution of tethering complexes and SM proteins. *BMC Evolutionary Biology*.

KRABBERØD, A.K., ORR, R.J.S., BRÅTE, J., KRISTENSEN, T., BJØRKLUND, K.R. & SHALCHIAN-TABRIZI, K. (2017) Single Cell Transcriptomics, Mega-Phylogeny, and the Genetic Basis of Morphological Innovations in Rhizaria. *Molecular biology and evolution*.

KRINGS, M., DOTZLER, N., LONGCORE, J.E. & TAYLOR, T.N. (2010) An unusual microfungus in a fungal spore from the Lower Devonian Rhynie chert. *Palaeontology*.

KU, C. & SEBÉ-PEDRÓS, A. (2019) Using single-cell transcriptomics to understand functional states and interactions in microbial eukaryotes. *Philosophical Transactions of the Royal Society B: Biological Sciences*.

KUBICEK, C. & DRUZHININA, I. (2007) Fungi in Extreme Environments. In *Environmental and Microbial Relationships* p.

KUBICEK, C.P., STARR, T.L. & GLASS, N.L. (2014) Plant Cell Wall – Degrading Enzymes and Their Secretion in Plant-Pathogenic Fungi.

KUDO, R. (1918) Experiments on the Extrusion of Polar Filaments of Cnidosporidian Spores. *The Journal of Parasitology*.

KWON-CHUNG, K.J. (2012) Taxonomy of fungi causing mucormycosis and entomophthoramycosis (zygomycosis) and nomenclature of the disease: Molecular mycologic perspectives. *Clinical Infectious Diseases*. .

LAETSCH, D.R. & BLAXTER, M.L. (2017) BlobTools: Interrogation of genome assemblies. *F1000Research* **6**, 1287.

LALLEMAND, F., GAUDEUL, M., LAMBOURDIÈRE, J., MATSUDA, Y., HASHIMOTO, Y. & SELOSSE, M.A. (2016) The elusive predisposition to mycoheterotrophy in Ericaceae. *The New phytologist*. .

LANG, B.F., KELLY, C.O., NERAD, T., GRAY, M.W., BURGER, G., CENTRE-VILLE, S. & HARBOR, W.B. (2002) The Closest Unicellular Relatives of Animals **12**, 1773–1778.

LANGE, L. (1979) *Uniflagellate phycomycete zoospore*. Dansk botanisk forening.

LANGE, L. & OLSON, L.W. (1976) The zoospore of Olpidium brassicae. *Protoplasma*.

LANGE, L. & OLSON, L.W. (1980) Germination of the resting sporangia of Physoderma maydis, the causal agent of Physoderma disease of maize. *Protoplasma*.

LARA, E., MOREIRA, D. & LÓPEZ-GARCÍA, P. (2010) The Environmental Clade LKM11 and Rozella Form the Deepest Branching Clade of Fungi. *Protist* **161**, 116–121.

LARSSON, J.I.R. (2000) The hyperparasitic microsporidium Amphiacantha longa Caullery et Mesnil, 1914 (Microspora: Metchnikovellidae) - Description of the cytology, redescription of the species, emended diagnosis of the genus Amphiacantha and establishment of the new family Amph. *Folia Parasitologica*.

LARSSON, J.I.R., STEINER, M.Y. & BJØRNSON, S. (1997) Intexta acarivora gen. et sp. n. (Microspora: Chytridiopsidae)-ultrastructural study and description of a new microsporidian parasite of the forage mite Tyrophagus putrescentiae (Acari: Acaridae). *Acta Protozoologica*.

LARSSON, R.J.I. & KØIE, M. (2006) The ultrastructure and reproduction of Amphiamblys capitellides (Microspora, Metchnikovellidae), a parasite of the gregarine Ancora sagittata (Apicomplexa, Lecudinidae), with redescription of the species and comments on the taxonomy. *European Journal of Protistology*.

LARTILLOT, N., LEPAGE, T. & BLANQUART, S. (2009) PhyloBayes 3: A Bayesian software package for phylogenetic reconstruction and molecular dating. *Bioinformatics*.

LASKEN, R.S. & MCLEAN, J.S. (2014) Recent advances in genomic DNA sequencing of microbial species from single cells. *Nature Reviews Genetics*.

LASKEN, R.S. & STOCKWELL, T.B. (2007) Mechanism of chimera formation during the Multiple Displacement Amplification reaction. *BMC Biotechnology* **7**, 1–11.

LAWREY, J.D. & DIEDERICH, P. (2003) Lichenicolous Fungi: Interactions, Evolution, and Biodiversity. *The*

Bryologist.

LAZARUS, K.L. & JAMES, T.Y. (2015) Surveying the biodiversity of the Cryptomycota using a targeted PCR approach. *Fungal Ecology* **14**, 62–70. Elsevier Ltd.

LEADBEATER, B.S.C. (2014) *The choanoflagellates: Evolution, biology and ecology*. In *The Choanoflagellates: Evolution, Biology and Ecology* p.

LEE, S.C., WEISS, L.M. & HEITMAN, J. (2009) Generation of genetic diversity in microsporidia via sexual reproduction and horizontal gene transfer. *Communicative and Integrative Biology*. .

LEFÈVRE, E., BARDOT, C., NOËL, C., CARRIAS, J.F., VISCOGLIOSI, E., AMBLARD, C. & SIME-NGANDO, T. (2007) Unveiling fungal zooflagellates as members of freshwater picoeukaryotes: Evidence from a molecular diversity study in a deep meromictic lake. *Environmental Microbiology*.

LEIPE, D.D., GUNDERSON, J.H., NERAD, T.A. & SOGIN, M.L. (1993) SMALL SUBUNIT RIBOSOMAL RNA+ OF HEXAMITA-INFLATA AND THE QUEST FOR THE 1ST BRANCH IN THE EUKARYOTIC TREE. *Molecular And Biochemical Parasitology*. .

LENE LANGE, L. & OLSON, L.W. (1978) The zoospore of Olpidium radicale. *Transactions of the British Mycological Society*.

LEONOV, M.M. (2012) Heliozoans (Heliozoa, Sarcodina, Protista) of fresh and marine waters of the European part of Russia: Species composition, morphology, and distribution. *Inland Water Biology* **3**, 344–355.

LETCHER, P.M., LEE, P.A., LOPEZ, S., BURNETT, M., MCBRIDE, R.C. & POWELL, M.J. (2016) An ultrastructural study of Paraphysoderma sedebokerense (Blastocladiomycota), an epibiotic parasite of microalgae. *Fungal Biology* **120**, 324–337. Elsevier Ltd.

LETCHER, P.M., LOPEZ, S., SCHMIEDER, R., LEE, P.A., BEHNKE, C., POWELL, M.J. & MCBRIDE, R.C. (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**.

LETCHER, P.M. & POWELL, M.J. (2018) A taxonomic summary and revision of Rozella (Cryptomycota). *IMA Fungus* **9**, 383–399.

LETCHER, P.M., POWELL, M.J., CHAMBERS, J.G., LONGCORE, J.E., CHURCHILL, P.F. & HARRIS, P.M. (2005) Ultrastructural and molecular delineation of the Chytridiaceae (Chytridiales). *Canadian Journal of Botany*.

LETCHER, P.M., VE, C.G., POWELL, M.J., CHURCHILL, P.F., WAKEFIELD, W.S. & HAWKSWORTH, D.L. (2008) Ultrastructural and molecular analyses of Rhizophydiales (Chytridiomycota) isolates from North America and Argentina **112**, 759–782.

LI, W. & GODZIK, A. (2006) Cd-hit: A fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics*.

LILLESKOV, E.A., FAHEY, T.J., HORTON, T.R. & LOVETT, G.M. (2002) Belowground Ectomycorrhizal Fungal Community Change Over a Nitrogen Deposition Gradient in Alaska. *Ecology*.

LIN, C.C. & ARONSON, J.M. (1970) Chitin and cellulose in the cell walls of the oomycete, Apodachlya sp. *Archiv für Mikrobiologie*.

LIN, K., LIMPENS, E., ZHANG, Z., IVANOV, S., SAUNDERS, D.G.O., MU, D., PANG, E., CAO, H., CHA, H., LIN, T.,

ZHOU, Q., SHANG, Y., LI, Y., SHARMA, T., VAN VELZEN, R., ET AL. (2014) Single Nucleus Genome Sequencing Reveals High Similarity among Nuclei of an Endomycorrhizal Fungus. *PLoS Genetics*.

LINNAEUS, C. (1758) *Systema naturae per regna tria naturae: secundum classes, ordines, genera, species, cum characteribus, differentiis, synonymis, locis. Editio duodecima. 1. Regnum Animale. 1 & 2 Holmiae, Laurentii Salvii*. In *Holmiae [Stockholm], Laurentii Salvii* p.

LIU, Y., STEENKAMP, E.T., BRINKMANN, H., FORGET, L., PHILIPPE, H. & LANG, B.F. (2009) Phylogenomic analyses predict sistergroup relationship of nucleariids and Fungi and paraphyly of zygomycetes with significant support. *BMC Evolutionary Biology* **9**, 1–11.

LIU, Y.J., HODSON, M.C. & HALL, B.D. (2006a) Loss of the flagellum happened only once in the fungal lineage : phylogenetic structure of Kingdom Fungi inferred from RNA polymerase II subunit genes **13**, 1–13.

LIU, Y.J., HODSON, M.C. & HALL, B.D. (2006b) Loss of the flagellum happened only once in the fungal lineage: Phylogenetic structure of Kingdom Fungi inferred from RNA polymerase II subunit genes. *BMC Evolutionary Biology* **6**, 1–13.

LIU, Z., HU, S.K., CAMPBELL, V., TATTERS, A.O., HEIDELBERG, K.B. & CARON, D.A. (2017) Single-cell transcriptomics of small microbial eukaryotes: Limitations and potential. *ISME Journal*.

LOMAN, N.J., CONSTANTINIDOU, C., CHAN, J.Z.M., HALACHEV, M., SERGEANT, M., PENN, C.W., ROBINSON, E.R. & PALLEN, M.J. (2012) High-throughput bacterial genome sequencing: An embarrassment of choice, a world of opportunity. *Nature Reviews Microbiology*.

LONGCORE, J.E., PESSIER, A.P. & NICHOLS, D.K. (1999) Batrachochytrium dendrobatidis gen. et sp. nov., a chytrid pathogenic to amphibians. *Mycologia*.

LONGCORE, J.E. & SIMMONS, D.R. (2012) The Polychytriales ord . nov . contains chitinophilic members of the rhizophlyctoid alliance **104**, 276–294.

LÓPEZ-ESCARDÓ, D., GRAU-BOVÉ, X., GUILLAUMET-ADKINS, A., GUT, M., SIERACKI, M.E. & RUIZ-TRILLO, I. (2017) Evaluation of single-cell genomics to address evolutionary questions using three SAGs of the choanoflagellate Monosiga brevicollis. *Scientific Reports*.

LÓPEZ-ESCARDÓ, D., LÓPEZ-GARCÍA, P., MOREIRA, D., RUIZ-TRILLO, I. & TORRUELLA, G. (2018) Parvularia atlantis gen. et sp. nov., a Nucleariid Filose Amoeba (Holomycota, Opisthokonta). *Journal of Eukaryotic Microbiology* **65**, 170–179.

LÓPEZ-GARCÍA, P. & MOREIRA, D. (2015) Open Questions on the Origin of Eukaryotes. *Trends in Ecology and Evolution*. .

LORON, C.C., FRANÇOIS, C., RAINBIRD, R.H., TURNER, E.C., BORENSZTAJN, S. & JAVAUX, E.J. (2019) Early fungi from the Proterozoic era in Arctic Canada. *Nature*. .

LOVETT, J.S. (1975) Growth and differentiation of the water mold Blastocladiella emersonii: cytodifferentiation and the role of ribonucleic acid and protein synthesis. *Bacteriological Reviews* **39**, 345–404.

LUBINSKY, G. (1955) On some parasites of parasitic protozoa. II. Sagittospora cameroni gen. n., sp. n. -a phycomycete parasitizing Ophryoscolecidae. *Canadian journal of microbiology* **1**, 675–684.

LUCAROTTI, C.J. & FEDERICI, B.A. (1984) Gametogenesis in Coelomomyces dodgei Couch (Blastocladiales,

Chytridiomycetes). *Protoplasma*.

LÜCKING, R., HUHNDORF, S., PFISTER, D.H., PLATA, E.R. & LUMBSCH, H.T. (2009) Fungi evolved right on track. *Mycologia* **101**, 810–822.

LUTZONI, F., KAUFF, F., COX, C.J., MCLAUGHLIN, D., CELIO, G., DENTINGER, C., PADAMSEE, M., HIBBETT, D., JAMES, T.Y., BALOCH, E., GRUBE, M., REEB, V., HOFSTETTER, V., SCHOCH, C., ARNOLD, A.E., ET AL. (2004) Assembling the fungal tree of life: Progress, classification, and evolution of subcellular traits. *American Journal of Botany*. .

LUTZONI, F., NOWAK, M.D., ALFARO, M.E., REEB, V., MIADLIKOWSKA, J., KRUG, M., ARNOLD, A.E., LEWIS, L.A., SWOFFORD, D.L., HIBBETT, D., HILU, K., JAMES, T.Y., QUANDT, D. & MAGALLÓN, S. (2018) Contemporaneous radiations of fungi and plants linked to symbiosis. *Nature Communications* **9**, 1–11. Springer US.

MA, L., CHEN, Z., HUANG, D.W., KUTTY, G., ISHIHARA, M., WANG, H., ABOUELLEIL, A., BISHOP, L., DAVEY, E., DENG, R., DENG, X., FAN, L., FANTONI, G., FITZGERALD, M., GOGINENI, E., ET AL. (2016) Genome analysis of three Pneumocystis species reveals adaptation mechanisms to life exclusively in mammalian hosts. *Nature Communications*.

MACHLIS, L. (1969) Zoospore Chemotaxis in the Watermold Allomyces. *Physiologia Plantarum* **22**, 126–139.

MAJOR, P., MAJOR, P., SENDRA, K.M., DEAN, P., WILLIAMS, T.A., WATSON, A.K., THWAITES, D.T., EMBLEY, T.M. & HIRT, R.P. (2019) A new family of cell surface located purine transporters in Microsporidia and related fungal endoparasites.

MANGOT, J.F., LOGARES, R., SÁNCHEZ, P., LATORRE, F., SEELEUTHNER, Y., MONDY, S., SIERACKI, M.E., JAILLON, O., WINCKER, P., VARGAS, C. DE & MASSANA, R. (2017) Accessing the genomic information of unculturable oceanic picoeukaryotes by combining multiple single cells. *Scientific Reports*.

MARDIS, E.R. (2011) A decade's perspective on DNA sequencing technology. *Nature*. .

MARGULIS, L., SCHMID, R., CORLISS, J.O., MELKONIAN, M. & CHAPMAN, D.J. (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 Mold, 914.

MARGULIS, L., SOYER-GOBILLARD, M.O. & CORLISS, J.O. (1984) *Evolutionary Protistology. The Organism as Cell*. D. Reidel, Dordrecht, Holland.

MAROSKE, S. & MAY, T.W. (2018) Naming names: the first women taxonomists in mycology. *Studies in Mycology*.

MARTÍN-DURÁN, J.M. & VELLUTINI, B.C. (2019) Old Questions and Young Approaches to Animal Evolution. *Springer Nature Switzerland AG 2019*, 1–12.

MARTIN, W.W. (1969) A morphological and cytological study of development in Coelomomyces punctatus parasitic in Anopheles quadrimaculatus. *Journal of the Elisha Mitchell Scientific Society*, 59–72. JSTOR.

MARTIN, W.W. (1987) Zoosporic parasites of aquatic insects: collection, identification, and culture. *In Zoosporic Fungi in Teaching and Research*, 137–142. M.S. Fuller and A.Jaworski. Southeastern Publ. Corp., Athens,Georgia.

DE MATTOS-SHIPLEY, K.M.J., FORD, K.L., ALBERTI, F., BANKS, A.M., BAILEY, A.M. & FOSTER, G.D. (2016) The good, the bad and the tasty: The many roles of mushrooms. *Studies in Mycology*.

MAXIME CORNU, M. (1871) Note Sur Deux Genres Nouveaux De La Famille Des Saprolégniées. *Bulletin de la Société Botanique de France* **18**, 58–59. Taylor & Francis.

MAYR, E. (1982) *The growth of biological thought: Diversity, evolution, and inheritance*. Harvard University Press.

MCCARTHY, C.G.P. & FITZPATRICK, D.A. (2017) *Multiple Approaches to Phylogenomic Reconstruction of the Fungal Kingdom*. In *Advances in Genetics* p., 1st edition. Elsevier Inc.

MCKERRACHER, L.J. & HEATH, I.B. (1985) The Structure and Cycle of the Nucleus-Associated Organelle in Two Species of Basidiobolus. *Mycologia*.

MCLAUGHLIN, D.J., HEALY, R.A., CELIO, G.J., ROBERSON, R.W. & KUMAR, T.K.A. (2015) Evolution of zygomycetous spindle pole bodies: Evidence from Coemansia reversa mitosis. *American Journal of Botany*.

MEDINA, M., COLLINS, A.G., TAYLOR, J.W., VALENTINE, J.W., LIPPS, J.H., AMARAL-ZETTLER, L. & SOGIN, M.L. (2003) Phylogeny of Opisthokonta and the evolution of multicellularity and complexity in Fungi and Metazoa. *International Journal of Astrobiology*.

MEIRINHO, P. DO A., NISHIMURA, P.Y., PIRES-ZOTTARELLI, C.L.A., MOCHINI-CARLOS, V. & POMPÊO, M.L.M. (2013) Olpidium gregarium, a chytrid fungus affecting rotifers populations in Rio Grande Reservoir, São Paulo State, Brazil. *Biota Neotropica*.

MÉLIDA, H., SANDOVAL-SIERRA, J. V., DIÉGUEZ-URIBEONDO, J. & BULONE, V. (2013) Analyses of extracellular carbohydrates in oomycetes unveil the existence of three different cell wall types. *Eukaryotic Cell*.

MENDOZA, L., TAYLOR, J.W. & AJELLO, L. (2002) T HE C LASS M ESOMYCETOZOEA : A Heterogeneous Group of Microorganisms at the Animal-Fungal Boundary ∗.

MICHEL, R. (2019) Intranuclear parasites of free-living amoebae: the Rozellomycota and the origins of the Microsporidia.

MICHEL, R., SCHMID, E.N., BÖKER, T., HAGER, D.G., MÜLLER, K.D., HOFFMANN, R. & SEITZ, H.M. (2000) Vannella sp. harboring Microsporidia-like organisms isolated from the contact lens and inflamed eye of a female keratitis patient. *Parasitology Research*.

MIKHAILOV, K. V., SIMDYANOV, T.G. & ALEOSHIN, V. V. (2017) Genomic survey of a hyperparasitic microsporidian Amphiamblys sp. (Metchnikovellidae). *Genome Biology and Evolution*.

MIKHAILOV, K. V, SIMDYANOV, T.G., ALEOSHIN, V. V & BELOZERSKY, A.N. (2016) Genomic survey of a hyperparasitic microsporidian Amphiamblys sp. (Metchnikovellidae) Genome Biology and Evolution Advance Access. *at University of Library and Information Science on*.

MIKRJUKOV, K.A. (1999a) Taxonomic Revision of Scale-bearing Heliozoon-like Amoebae (Pompholyxophryidae, Rotosphaerida). *Acta Potozoologica* **38**, 119–131.

MIKRJUKOV, K.A. (1999b) Acta protozoologica.

MILLER, C.E. & DYLEWSKI, D.P. (1981) Syngamy and Resting Body Development in Chytriomyces hyalinus (Chytridiales). *American Journal of Botany*.

MISRA, J.K., TEWARI, J.P. & DESHMUKH, S.K. (2012) *Systematics and evolution of fungi*. In *Systematics and*

Evolution of Fungi p.

MITCHELL, R.T. & DEACON, J.W. (1986) Selective accumulation of zoospores of chytridiomycetes and oomycetes on cellulose and chitin. *Transactions of the British Mycological Society*.

MOHAMED, D.J. & MARTINY, J.B.H. (2011) Patterns of fungal diversity and composition along a salinity gradient. *ISME Journal*.

MONCHY, S., SANCIU, G., JOBARD, M., RASCONI, S., GERPHAGNON, M., CHABÉ, M., CIAN, A., MELONI, D., NIQUIL, N., CHRISTAKI, U., VISCOGLIOSI, E. & SIME-NGANDO, T. (2011) Exploring and quantifying fungal diversity in freshwater lake ecosystems using rDNA cloning/sequencing and SSU tag pyrosequencing. *Environmental Microbiology*.

MOORE, R.T. (1980) Taxonomic proposals for the classification of marine yeasts and other yeast-like fungi including the smuts. *Botanica marina*. .

MULKIDJANIAN, A.Y., BYCHKOV, A.Y., DIBROVA, D. V., GALPERIN, M.Y. & KOONIN, E. V. (2012) Origin of first cells at terrestrial, anoxic geothermal fields. *Proceedings of the National Academy of Sciences of the United States of America*.

NADIMI, M., BEAUDET, D., FORGET, L., HIJRI, M. & FRANZ LANG, B. (2012) Group i intron-mediated trans-splicing in mitochondria of gigaspora rosea and a robust phylogenetic affiliation of arbuscular mycorrhizal fungi with mortierellales. *Molecular Biology and Evolution*.

NAGAHAMA, T., SATO, H., SHIMAZU, M. & SUGIYAMA, J. (1995) Phylogenetic divergence of the entomophthoralean fungi: Evidence from nuclear 18S ribosomal RNA gene sequences. *Mycologia*.

NAGAHAMA, T., TAKAHASHI, E., NAGANO, Y., ABDEL-WAHAB, M.A. & MIYAZAKI, M. (2011) Molecular evidence that deep-branching fungi are major fungal components in deep-sea methane cold-seep sediments. *Environmental Microbiology*.

NÄGELI, K. (1857) Über die neue Krankheit der Seidenraupe und verwandte Organismen. *Botanische Zeitung* **15**, 760–761.

NAGY, L.G., KOVÁCS, G.M. & KRIZSÁN, K. (2018) Complex multicellularity in fungi: evolutionary convergence, single origin, or both? *Biological Reviews* **93**, 1778–1794.

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., LANG, B.F., SPATAFORA, J.W., MARTIN, F.M., GRIGORIEV, I. V. & HIBBETT, D.S. (2014) Latent homology and convergent regulatory evolution underlies the repeated emergence of yeasts. *Nature Communications*.

NARANJO-ORTIZ, M.A. & GABALDÓN, T. (2019a) Fungal evolution: diversity, taxonomy and phylogeny of the Fungi. *Biological Reviews* **94**, 2101–2137.

NARANJO-ORTIZ, M.A. & GABALDÓN, T. (2019b) Fungal evolution: major ecological adaptations and evolutionary transitions. *Biological Reviews* **94**, 1443–1476.

NAVIN, N., KENDALL, J., TROGE, J., ANDREWS, P., RODGERS, L., MCINDOO, J., COOK, K., STEPANSKY, A., LEVY, D., ESPOSITO, D., MUTHUSWAMY, L., KRASNITZ, A., MCCOMBIE, W.R., HICKS, J. & WIGLER, M. (2011) Tumour evolution inferred by single-cell sequencing. *Nature*.

NEEDHAM, D.M., POIRIER, C., HEHENBERGER, E., JIMÉNEZ, V., SWALWELL, J.E., SANTORO, A.E. & WORDEN, A.Z. (2019) Targeted metagenomic recovery of four divergent viruses reveals shared and distinctive characteristics of giant viruses of marine eukaryotes. *Philosophical Transactions of the Royal Society B: Biological Sciences*.

NGUYEN, L.T., SCHMIDT, H.A., VON HAESELER, A. & MINH, B.Q. (2015) IQ-TREE: A fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Molecular Biology and Evolution* **32**, 268–274.

NGUYEN, T.T., LEE, S.H., BAE, S., JEON, S.J., MUN, H.Y. & LEE, H.B. (2016) Characterization of two new records of zygomycete species belonging to undiscovered taxa in korea. *Mycobiology*.

NICHOLLS, K.H. (2012a) New and little-known marine species of Pinaciophora, Rabdiaster and Thomseniophora gen. nov. (Rotosphaerida: Pompholyxophryidae). *Journal of the Marine Biological Association of the United Kingdom* **93**, 1211–1229.

NICHOLLS, K.H. (2012b) Structure of Organic Spines in the Rhizarian Protist Belonocystis tubistella Rainer, 1968, and a Description of Belonocystis quadrangularis n. sp. (Cercozoa, Insertae Sedis). *Acta Protozoologica* **51**, 131–142.

NICHOLLS, K.H. (2013) New and little-known marine species of Pinaciophora , Rabdiaster and Thomseniophora gen . nov . (Rotosphaerida : Pompholyxophryidae) **93**, 1211–1229.

NITSCHE, F., CARR, M., ARNDT, H. & LEADBEATER, B.S.C. (2011) Higher level taxonomy and molecular phylogenetics of the Choanoflagellatea. *Journal of Eukaryotic Microbiology*.

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 Protozoologica*.

NURK, S., BANKEVICH, A., ANTIPOV, D., GUREVICH, A.A., KOROBEYNIKOV, A., LAPIDUS, A., PRJIBELSKI, A.D., PYSHKIN, A., SIROTKIN, A., SIROTKIN, Y., STEPANAUSKAS, R., CLINGENPEEL, S.R., WOYKE, T., MCLEAN, J.S., LASKEN, R., ET AL. (2013) Assembling single-cell genomes and mini-metagenomes from chimeric MDA products. *Journal of Computational Biology* **20**, 714–737.

O'MALLEY, M.A., WIDEMAN, J.G. & RUIZ-TRILLO, I. (2016) Losing Complexity: The Role of Simplification in Macroevolution. *Trends in Ecology and Evolution*. .

OLIVE, L.S. & WHITTAKER, R.H. (1969) Reassignment of gymnomycota. *Science*.

OLSON, D.H., AANENSEN, D.M., RONNENBERG, K.L., POWELL, C.I., WALKER, S.F., BIELBY, J., GARNER, T.W.J., WEAVER, G. & FISHER, M.C. (2013) Mapping the Global Emergence of Batrachochytrium dendrobatidis, the Amphibian Chytrid Fungus. *PLoS ONE*.

OLSON, L.W. (1984) *Allomyces, a different fungus*. Copenhagen : Council for Nordic Publications in Botany.

ONSBRING, H., TICE, A.K., BARTON, B.T., BROWN, M.W. & ETTEMA, T.J.G. (2019) An efficient single-cell transcriptomics workflow to assess protist diversity and lifestyle. *bioRxiv*. Cold Spring Harbor Laboratory.

ORCHARD, S., HILTON, S., BENDING, G.D., DICKIE, I.A., STANDISH, R.J., GLEESON, D.B., JEFFERY, R.P., POWELL, J.R., WALKER, C., BASS, D., MONK, J., SIMONIN, A. & RYAN, M.H. (2017) Fine endophytes (Glomus tenue) are related to Mucoromycotina, not Glomeromycota. *New Phytologist*. .

OWCZARZAK, A., STIBBS, H.H. & BAYNE, C.J. (1980) The destruction of Schistosoma mansoni mother sporocysts in vitro by amoebae isolated from Biomphalaria glabrata: an ultrastructural study. *Journal of Invertebrate*

Pathology.

PAGE, F.C. (1987) The Classification of 'Naked' Amoebae (Phylum Rhizopoda). *Archiv für Protistenkunde* **133**, 199–217.

PAGE, F.C. & SIEMENSMA, F.J. (1991) Nackte Rhizopoda und Heliozoea.

PAPS, J., MEDINA-CHACÓN, L.A., 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**, 2–12.

PAPS, J. & RUIZ-TRILLO, I. (2010) Animals and Their Unicellular Ancestors. *Encyclopedia of Life Sciences*, 1–8.

PARFREY, L.W., LAHR, D.J.G., KNOLL, A.H. & KATZ, L.A. (2011) Estimating the timing of early eukaryotic diversification with multigene molecular clocks. *Proceedings of the National Academy of Sciences of the United States of America*.

PARKER, T.J. & HASWELL, W.A. (1900) A Manual of Zoology. *The Macmillan Company*.

PATTERSON, D.J. (1983a) On the Organization of the Naked Filose Amoeba, Nuclearia moebiusi Frenzel, 1897 (Sarcodina, Filosea) and Its Implications1. *Journal of Protozoology* **30**, 301–307.

PATTERSON, D.J. (1983b) Nuclearia moebiusi **30**, 301–307.

PATTERSON, D.J. (1984) The Genus Nuclearia (Sarcodina, Filosea): Species Composition and Characteristics of the Taxa. *Archiv fur Protistenkunde* **128**, 127–139.

PATTERSON, D.J. (1985) On the Organization and Affinities of the Amoeba, Pompholyxophrys punicea Archer, Based on Ultrastructural Examination of Individual Cells from Wild Material. *Journal of Eukaryotic Microbiology* **32**, 241–246.

PATTERSON, D.J. (1999) The diversity of eukaryotes. *American Naturalist*. .

PATTERSON, D.J., SIMPSON, A.G.B. & ROGERSON, A. (2000) Amoebae of uncertain affinities. In *An illustrated guide to the Protozoa, 2nd ed. Lawrence (KS): Society of Protozoologists. Vol. II* (eds J.J. LEE, G.F. LEEDALE & P. BRADBURY), pp. 804–826, 2nd edition.

PATTERSON, D.J., SUREK, B. & MELKONIAN, M. (1987) The Ultrastructure of Vampyrellidium perforans Surek & Melkonian and Its Taxonomic Position Among the Naked Filose Amoebae1. *The Journal of Protozoology* **34**, 63–67.

VAN DE PEER, Y. & DE WACHTER, R. (1997) Evolutionary relationships among the eukaryotic crown taxa taking into account site-to-site rate variation in 18S rRNA. *Journal of Molecular Evolution*.

PERSOON, C.H. (1801) *Synopsis methodica fungorum*. Gottingae : Apvd Henricvm Dieterich.

PETERSEN, H.E. (1909) Studier over Ferskvands-Phycomyceter. *Botanisk tidsskrift* **29**, 345–440. København :G.E.C. Gads Forlag,1866-1981.

PETKOVITS, T., NAGY, L.G., HOFFMANN, K., WAGNER, L., NYILASI, I., GRIEBEL, T., SCHNABELRAUCH, D., VOGEL, H., VOIGT, K., VÁGVÖLGYI, C. & PAPP, T. (2011) Data partitions, bayesian analysis and phylogeny of the zygomycetous fungal family mortierellaceae, inferred from nuclear ribosomal DNA sequences. *PLoS ONE*.

PHILIPPE, H., LOPEZ, P., BRINKMANN, H., BUDIN, K., GERMOT, A., LAURENT, J., MOREIRA, D., MÜLLER, M. & LE GUYADER, H. (2000) Early-branching or fast-evolving eukaryotes? An answer based on slowly evolving

positions. *Proceedings of the Royal Society B: Biological Sciences* **267**, 1213–1221.

PINEVICH, A., GROMOV, B., MAMKAEVA, K. & NASONOVA, E. (1997) Study of Molecular Karyotypes in Amoeboaphelidium protococcarum , the Endotrophic Parasite of Chlorophycean alga Scenedesmus **34**, 122–126.

POMBERT, J.-F., SELMAN, M., BURKI, F., BARDELL, F.T., FARINELLI, L., SOLTER, L.F., WHITMAN, D.W., WEISS, L.M., CORRADI, N. & KEELING, P.J. (2012) Gain and loss of multiple functionally related, horizontally transferred genes in the reduced genomes of two microsporidian parasites. *Proceedings of the National Academy of Sciences* **109**, 12638–12643.

PORTER, T.M., MARTIN, W., JAMES, T.Y., LONGCORE, J.E., GLEASON, F.H., ADLER, P.H., LETCHER, P.M. & VILGALYS, R. (2011) Molecular phylogeny of the Blastocladiomycota (Fungi) based on nuclear ribosomal DNA. *Fungal Biology* **115**, 381–392. Elsevier Ltd.

POWELL, M.J. (1978) Phylogenetic implications of the microbody-lipid globule complex in zoosporic fungi. *BioSystems*.

POWELL, M.J. (1984) Fine Structure of the Unwalled Thallus of Rozella polyphagi in Its Host Polyphagus euglenae. *Mycologia*.

POWELL, M.J. (1993) Looking at mycology with a Janus face: a glimpse at Chytridiomycetes active in the environment. *Mycologia*.

POWELL, M.J. (2017a) Chytridiomycota. In *Handbook of the Protists: Second Edition* p.

POWELL, M.J. (2017b) Blastocladiomycota. In *Handbook of the Protists: Second Edition* p.

POWELL, M.J. & LETCHER, P.M. (2014) Chytridiomycota, Monoblepharidomycota, and Neocallimastigomycota. In *Systematics and Evolution: Part A: Second Edition* p.

POWELL, M.J. & LETCHER, P.M. (2019) Ultrastructure of early stages of Rozella allomycis (Cryptomycota) infection of its host, Allomyces macrogynus (Blastocladiomycota). *Fungal Biology* **123**, 109–116. Elsevier Ltd.

POWELL, M.J., LETCHER, P.M., JAMES, T.Y. & BEAKES, G.W. (2017) Ultrastructural characterization of the host e parasite interface between Allomyces anomalus (Blastocladiomycota) and Rozella allomycis (Cryptomycota). *Fungal Biology* **121**, 561–572. Elsevier Ltd.

PRESSEL, S., BIDARTONDO, M.I., FIELD, K.J., RIMINGTON, W.R. & DUCKETT, J.G. (2016) Pteridophyte fungal associations: Current knowledge and future perspectives. *Journal of Systematics and Evolution*. .

PRICE, M.N., DEHAL, P.S. & ARKIN, A.P. (2009) Fasttree: Computing large minimum evolution trees with profiles instead of a distance matrix. *Molecular Biology and Evolution*.

PROHASKA, J.R. (1981) Role of Golgi Apparatus in Sorogenesis by the Cellular Slime Mold Fonticula alba **213**, 561–563.

PURRINI, K. & WEISER, J. (1985) Ultrastructural study of the microsporidian chytridiopsis typographi (Chytridiopsida: Microspora) infecting the bark beetle, Ips typographus (Scolytidae: Coleoptera), with new data on spore dimorphism. *Journal of Invertebrate Pathology*.

QUANDT, C.A., BEAUDET, D., CORSARO, D., WALOCHNIK, J., MICHEL, R., CORRADI, N. & JAMES, T.Y. (2017) The genome of an intranuclear parasite, Paramicrosporidium saccamoebae, reveals alternative adaptations to

obligate intracellular parasitism. *eLife* **6**.

RADEK, R., KARITON, M., DABERT, J. & ALBERTI, G. (2015) Ultrastructural characterization of Acarispora falculifera n.gen., n.sp., a new microsporidium (Opisthokonta: Chytridiopsida) from the feather mite Falculifer rostratus (Astigmata: Pterolichoidea). *Acta Parasitologica*.

RADEK, R., WURZBACHER, C., GISDER, S., NILSSON, R.H., OWERFELDT, A., GENERSCH, E., KIRK, P.M. & VOIGT, K. (2017) Morphologic and molecular data help adopting the insect-pathogenic nephridiophagids (Nephridiophagidae) among the early diverging fungal lineages, close to the Chytridiomycota. *MycoKeys*.

RAGAN, M.A., II, C.L.G., CAWTHORN, R.J., CERENIUS, L., JAMIESONTT, A.V.C., PLOURDE, S.M., RANDTT, T.G., SDERHALL, K. & GUTELL, R.R. (1996) A novel clade of protistan parasites near the animal- fungal divergence **93**, 11907–11912.

RAGHUNATHAN, A., FERGUSON, H.R., BORNARTH, C.J., SONG, W., DRISCOLL, M. & LASKEN, R.S. (2005) Genomic DNA amplification from a single bacterium. *Applied and Environmental Microbiology*.

RAINER, H. (1968) Urtiere, Protozoa - Wurzelfuessler, Rhizopoda - Sonnentierchen, Heliozoa. In *Die Tierwelt Deutschlands* (ed FISCHER), p. Jena.

RAMBAUT, A. (2016) FigTree v1.4.3. *Molecular evolution, phylogenet ics and epidemiology*. .

REDECKER, D. & SCHÜßLER, A. (2014) Glomeromycota. In *Systematics and Evolution: Part A: Second Edition* p.

REFARDT, D. & EBERT, D. (2012) The impact of infection on host competition and its relationship to parasite persistence in a Daphnia microparasite system. *Evolutionary Ecology*.

RICHARDS, T.A. & GOMES, S.L. (2015) How to build a microbial eye. *Nature* **523**, 166–167.

RICHARDS, T.A., LEONARD, G., MAHÉ, F., DEL CAMPO, J., ROMAC, S., JONES, M.D.M., MAGUIRE, F., DUNTHORN, M., DE VARGAS, C., MASSANA, R. & CHAMBOUVET, A. (2015) Molecular diversity and distribution of marine fungi across 130 european environmental samples. *Proceedings of the Royal Society B: Biological Sciences*.

RICHARDS, T.A., LEONARD, G. & WIDEMAN, J.G. (2017a) What Defines the "Kingdom" Fungi? *Microbiology Spectrum* **5**, 1–21.

RICHARDS, T.A., LEONARD, G.U.Y. & WIDEMAN, J.G. (2017b) What De fi nes the " Kingdom " Fungi ?, 1–21.

RICHARDSON, E., BASS, D., SMIRNOVA, A., PAOLI, L., DUNFIELD, P. & DACKS, J.B. (2019) Phylogenetic Estimation of Community Composition and Novel Eukaryotic Lineages in Base Mine Lake: An Oil Sands Tailings Reclamation Site in Northern Alberta. *Journal of Eukaryotic Microbiology*.

RICHTER, D.J. & KING, N. (2013) The Genomic and Cellular Foundations of Animal Origins.

RIETHMÜLLER, A., WEIß, M. & OBERWINKLER, F. (1999) Phylogenetic studies of Saprolegniomycetidae and related groups based on nuclear large subunit ribosomal DNA sequences. *Canadian Journal of Botany*.

RINKE, C., LEE, J., NATH, N., GOUDEAU, D., THOMPSON, B., POULTON, N., DMITRIEFF, E., MALMSTROM, R., STEPANAUSKAS, R. & WOYKE, T. (2014) Obtaining genomes from uncultivated environmental microorganisms using FACS-based single-cell genomics. *Nature Protocols*.

RINKE, C., SCHWIENTEK, P., SCZYRBA, A., IVANOVA, N.N., ANDERSON, I.J., CHENG, J.F., DARLING, A., MALFATTI, S., SWAN, B.K., GIES, E.A., DODSWORTH, J.A., HEDLUND, B.P., TSIAMIS, G., SIEVERT, S.M., LIU, W.T., ET AL. (2013) Insights into the phylogeny and coding potential of microbial dark matter. *Nature*.

ROBERSON, R.W., SAUCEDO, E., MACLEAN, D., PROPSTER, J., UNGER, B., ONEIL, T.A., PARVANEHGOHAR, K., CAVANAUGH, C. & LOWRY, D. (2011) The hyphal tip structure of Basidiobolus sp.: A zygomycete fungus of uncertain phylogeny. *Fungal Biology*.

ROBERTSON, J.A. (1972) Phototaxis in a new Allomyces. *Archiv für Mikrobiologie*.

RODRÍGUEZ-MARTÍNEZ, R., LEONARD, G., MILNER, D.S., SUDEK, S., CONWAY, M., MOORE, K., HUDSON, T., MAHÉ, F., KEELING, P.J., SANTORO, A.E., WORDEN, A.Z. & RICHARDS, T.A. (2020) Controlled sampling of ribosomally active protistan diversity in sediment-surface layers identifies putative players in the marine carbon sink. *The ISME Journal*.

ROIJACKERS, R.M.M. & SIEMENSMA, F.J. (1988) A study of cristidiscoidid amoebae (Rhizopoda, Filosea), with descriptions of new species and keys to genera and species. *Archiv fur Protistenkunde* **135**, 237–253. VEB Gustav Fischer Verlag Jena.

ROTARI, Y.M., PASKEROVA, G.G. & SOKOLOVA, Y.Y. (2015) Diversity of metchnikovellids (Metchnikovellidae, Rudimicrosporea), hyperparasites of bristle worms (Annelida, Polychaeta) from the White Sea. *Protistology* **9**, 50–59.

ROTHSCHILD, L.J. (1989) Protozoa, Protista, Protoctista: What's in a Name? *Journal of the History of Biology* **22**, 277–305. Springer.

RUIZ-TRILLO, I., INAGAKI, Y., DAVIS, L.A., SPERSTAD, S., LANDFALD, B. & ROGER, A.J. (2004) Capsaspora owczarzaki is an independent opisthokont lineage [2]. *Current Biology*. .

RUIZ-TRILLO, I., ROGER, A.J., BURGER, G., GRAY, M.W. & LANG, B.F. (2008) A phylogenomic investigation into the origin of Metazoa. *Molecular Biology and Evolution*.

SANDERS, W.B. (2002) *The Mycota*. In *The Bryologist* p.

SCHALCHIAN-TABRIZI, K., MINGE, M.A., ESPELUND, M., ORR, R., RUDEN, T., JAKOBSEN, K.S. & CAVALIER-SMITH, T. (2008) Multigene phylogeny of Choanozoa and the origin of animals. *PLoS ONE*.

SCHERFFEL, A. (1925) Endophytische phycomyceten-parasiten der Bacillariaceen und einige neue monadinen. *Arch. Protistenkd* **52**, 1–141.

SCHNEPF, E., DEICHGRÄBER, G., HEGEWALD, E. & SOEDER, C.-J. (1971) Elektronenmikroskopische Beobachtungen an Parasiten aus Scenedesmus-Massenkulturen. *Archiv für Mikrobiologie* **75**, 230–245.

SCHOCH, C.L., SUNG, G.H., LÓPEZ-GIRÁLDEZ, F., TOWNSEND, J.P., MIADLIKOWSKA, J., HOFSTETTER, V., ROBBERTSE, B., MATHENY, P.B., KAUFF, F., WANG, Z., GUEIDAN, C., ANDRIE, R.M., TRIPPE, K., CIUFETTI, L.M., WYNNS, A., ET AL. (2009) The ascomycota tree of life: A phylum-wide phylogeny clarifies the origin and evolution of fundamental reproductive and ecological traits. *Systematic Biology*.

SCHULZE, F.E. (1874) Rhizopodenstudien II. *Archiv für mikroskopische Anatomie.*, 377–400. Max Cohen & Son, Bonn.

SCHÜßLER, A., SCHWARZOTT, D. & WALKER, C. (2001) A new fungal phylum, the Glomeromycota: Phylogeny and evolution. *Mycological Research*.

SEBÉ-PEDRÓS, A., DEGNAN, B.M. & RUIZ-TRILLO, I. (2017) The origin of Metazoa : a unicellular. *Nature Publishing Group*. Nature Publishing Group.

SEBÉ-PEDRÓS, A., DE MENDOZA, A., LANG, B.F., DEGNAN, B.M. & RUIZ-TRILLO, I. (2011) Unexpected repertoire of metazoan transcription factors in the unicellular holozoan capsaspora owczarzaki. *Molecular Biology and Evolution*.

SEELEUTHNER, Y., MONDY, S., LOMBARD, V., CARRADEC, Q., PELLETIER, E., WESSNER, M., LECONTE, J., MANGOT, J.F., POULAIN, J., LABADIE, K., LOGARES, R., SUNAGAWA, S., DE BERARDINIS, V., SALANOUBAT, M., DIMIER, C., ET AL. (2018) Single-cell genomics of multiple uncultured stramenopiles reveals underestimated functional diversity across oceans. *Nature Communications*.

SEKIMOTO, S., ROCHON, D., LONG, J.E., DEE, J.M. & BERBEE, M.L. (2011) A multigene phylogeny of Olpidium and its implications for early fungal evolution. *BMC Evolutionary Biology* **11**, 331. BioMed Central Ltd.

SERRIS, A., DANION, F. & LANTERNIER, F. (2019) Disease entities in mucormycosis. *Journal of Fungi*.

SETO, K., KAGAMI, M. & DEGAWA, Y. (2017) Phylogenetic Position of Parasitic Chytrids on Diatoms: Characterization of a Novel Clade in Chytridiomycota. *Journal of Eukaryotic Microbiology* **64**, 383–393.

SEWALL, T.C., ROBERSON, R.W. & POMMERVILLE, J.C. (1989) Identification and characterization of Golgi equivalents from Allomyces macrogynus. *Experimental Mycology*.

SHALCHIAN-TABRIZI, K., MINGE, M.A., ESPELUND, M., ORR, R., RUDEN, T., KJETILL, S. & CAVALIER-SMITH, T. (2008) Multigene Phylogeny of Choanozoa and the Origin of Animals **3**.

SIERACKI, M.E., POULTON, N.J., JAILLON, O., WINCKER, P., DE VARGAS, C., RUBINAT-RIPOLL, L., STEPANAUSKAS, R., LOGARES, R. & MASSANA, R. (2019) Single cell genomics yields a wide diversity of small planktonic protists across major ocean ecosystems. *Scientific reports*.

SIMÃO, F.A., WATERHOUSE, R.M., IOANNIDIS, P., KRIVENTSEVA, E. V. & ZDOBNOV, E.M. (2015) BUSCO: Assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* **31**, 3210–3212.

SIMON, M., LÓPEZ-GARCÍA, P., DESCHAMPS, P., MOREIRA, D., RESTOUX, G., BERTOLINO, P. & JARDILLIER, L. (2015) Marked seasonality and high spatial variability of protist communities in shallow freshwater systems, 1–13.

SIMPSON, A.G.B., INAGAKI, Y. & ROGER, A.J. (2004) Comprehensive Multigene Phylogenies of Excavate Protists Reveal the Evolutionary Positions of ‘“ Primitive ”’ Eukaryotes.

SMITH, M.E., GRYGANSKYI, A., BONITO, G., NOUHRA, E., MORENO-ARROYO, B. & BENNY, G. (2013) Phylogenetic analysis of the genus Modicella reveals an independent evolutionary origin of sporocarp-forming fungi in the Mortierellales. *Fungal Genetics and Biology*.

SMITH, S. & READ, D. (2008) *Mycorrhizal Symbiosis*. In *Mycorrhizal Symbiosis* p.

SOKOLOVA, Y.Y., PASKEROVA, G.G., ROTARI, Y.M., NASSONOVA, E.S. & SMIRNOV, A. V. (2013) Fine structure of Metchnikovella incurvata Caullery and Mesnil 1914 (microsporidia), a hyperparasite of gregarines Polyrhabdina sp. from the polychaete Pygospio elegans. *Parasitology* **140**, 855–867.

SPANG, A., SAW, J.H., JØRGENSEN, S.L., ZAREMBA-NIEDZWIEDZKA, K., MARTIJN, J., LIND, A.E., VAN EIJK, R., SCHLEPER, C., GUY, L. & ETTEMA, T.J.G. (2015) Complex archaea that bridge the gap between prokaryotes and eukaryotes. *Nature* **521**, 173–179.

SPARROW, F.K. (1960) Aquatic Phycomycetes. *University of Michigan Press* **132**.

SPATAFORA, J.W., AIME, M.C., GRIGORIEV, I. V., MARTIN, F., STAJICH, J.E. & BLACKWELL, M. (2017) The Fungal Tree of Life: from Molecular Systematics to Genome-Scale Phylogenies. *Microbiology Spectrum* **5**, 1–32.

SPATAFORA, J.W., CHANG, Y., BENNY, G.L., LAZARUS, K., SMITH, M.E., BERBEE, M.L., BONITO, G., CORRADI, N., GRIGORIEV, I., GRYGANSKYI, A., JAMES, T.Y., O'DONNELL, K., ROBERSON, R.W., TAYLOR, T.N., UEHLING, J., ET AL. (2016a) A phylum-level phylogenetic classification of zygomycete fungi based on genome-scale data. *Mycologia* **108**, 1028–1046.

SPATAFORA, J.W., STAJICH, J.E., BENNY, G.L., SMITH, M.E., BERBEE, M.L., CORRADI, N., GRIGORIEV, I., JAMES, T.Y., DONNELL, K.O. & ROBERSON, R.W. (2016b) A phylum-level phylogenetic classification of zygomycete fungi based on genome-scale data **108**, 1028–1046.

SPRAGUE, V. (1977) Annotated List of Species of Microsporidia BT - Comparative Pathobiology: Volume 2 Systematics of the Microsporidia. In (eds L.A. BULLA & T.C. CHENG), pp. 31–334. Springer US, Boston, MA.

STAIRS, C.W., LEGER, M.M. & ROGER, A.J. (2015) Diversity and origins of anaerobic metabolism in mitochondria and related organelles. *Philosophical Transactions of the Royal Society B: Biological Sciences*. .

STAJICH, J.E., BERBEE, M.L., BLACKWELL, M., HIBBETT, D.S., JAMES, T.Y., SPATAFORA, J.W. & TAYLOR, J.W. (2009) The Fungi. *Current Biology*. .

STEENKAMP, E.T., WRIGHT, J. & BALDAUF, S.L. (2006) The protistan origins of animals and fungi. *Molecular Biology and Evolution*.

STEINBERG, G., NYGAARD, K. & STEINBERG, G. (1993) Heterotrophic flagellates and other protists associated with oceanic detritus throughout the water column in the mid north atlantic. *Journal of the Marine Biological Association of the United Kingdom*.

STENTIFORD, G.D., BATEMAN, K.S., LONGSHAW, M. & FEIST, S.W. (2007) Enterospora canceri n. gen., n. sp., intranuclear within the hepatopancreatocytes of the European edible crab Cancer pagurus. *Diseases of Aquatic Organisms*.

STENTIFORD, G.D., RAMILO, A., ABOLLO, E., KERR, R., BATEMAN, K.S., FEIST, S.W., BASS, D. & VILLALBA, A. (2017) Hyperspora aquatica n.gn., n.sp. (Microsporidia), hyperparasitic in Marteilia cochillia (Paramyxida), is closely related to crustacean-infecting microspordian taxa. *Parasitology*.

STERN, A.I., SCHIFF, J.A. & KLEIN, H.P. (1960) Isolation of Ergosterol from Euglena gracilis; Distribution Among Mutant Strains. *The Journal of Protozoology*.

STRITTMATTER, M., GUERRA, T., SILVA, J. & GACHON, C.M.M. (2016) A new flagellated dispersion stage in Paraphysoderma sedebokerense, a pathogen of Haematococcus pluvialis. *Journal of Applied Phycology* **28**, 1553–1558. Journal of Applied Phycology.

STROTHER, P.K., BATTISON, L., BRASIER, M.D. & WELLMAN, C.H. (2011) Earth's earliest non-marine eukaryotes. *Nature*.

STRULLU-DERRIEN, C., GORAL, T., LONGCORE, J.E., OLESEN, J., KENRICK, P. & EDGECOMBE, G.D. (2016) A new chytridiomycete fungus intermixed with crustacean resting eggs in a 407-millionyear-old continental freshwater environment. *PLoS ONE*.

STRULLU-DERRIEN, C., SELOSSE, M.A., KENRICK, P. & MARTIN, F.M. (2018a) The origin and evolution of mycorrhizal symbioses: from palaeomycology to phylogenomics. *New Phytologist*. .

STRULLU-DERRIEN, C., SPENCER, A.R.T., GORAL, T., DEE, J., HONEGGER, R., KENRICK, P., LONGCORE, J.E. & BERBEE, M.L. (2018b) New insights into the evolutionary history of fungi from a 407 Ma blastocladiomycota fossil showing a complex hyphal thallus. *Philosophical Transactions of the Royal Society B: Biological Sciences* **373**.

STUMM, C., HERMANS, J.M.H., CROES, A.F. & BUCKS, J.H. (1976) Chemotaxis and transport of amino acids in Allomyces arbuscula. *Antonie van Leeuwenhoek*.

SUMMERBELL, R.C. (2005) Root endophyte and mycorrhizosphere fungi of black spruce, Picea mariana, in a boreal forest habitat: Influence of site factors on fungal distributions. *Studies in Mycology*.

SUN, X., CHEN, W., IVANOV, S., MACLEAN, A.M., WIGHT, H., RAMARAJ, T., MUDGE, J., HARRISON, M.J. & FEI, Z. (2018) Genome and evolution of the arbuscular mycorrhizal fungus Diversispora epigaea (formerly Glomus versiforme) and its bacterial endosymbionts. *New Phytologist*.

TAKISHITA, K., TSUCHIYA, M., KAWATO, M., OGURI, K., KITAZATO, H. & MARUYAMA, T. (2007) Genetic Diversity of Microbial Eukaryotes in Anoxic Sediment of the Saline Meromictic Lake Namako-ike (Japan): On the Detection of Anaerobic or Anoxic-tolerant Lineages of Eukaryotes. *Protist*.

TANABE, Y., SAIKAWA, M., WATANABE, M.M. & SUGIYAMA, J. (2004) Molecular phylogeny of Zygomycota based on EF-1α and RPB1 sequences: Limitations and utility of alternative markers to rDNA. *Molecular Phylogenetics and Evolution* **30**, 438–449.

TANABE, Y., WATANABE, M.M. & SUGIYAMA, J. (2005) Evolutionary relationships among basal fungi (Chytridiomycota and Zygomycota): Insights from molecular phylogenetics **276**, 267–276.

TANG, F., BARBACIORU, C., WANG, Y., NORDMAN, E., LEE, C., XU, N., WANG, X., BODEAU, J., TUCH, B.B., SIDDIQUI, A., LAO, K. & SURANI, M.A. (2009) mRNA-Seq whole-transcriptome analysis of a single cell. *Nature Methods*.

TAYLOR, T.N., HASS, H. & REMY, W. (1992) Devonian fungi: interactions with the green alga Palaeonitella. *Mycologia*.

TAYLOR, T.N. & OSBORN, J.M. (1996) The importance of fungi in shaping the paleoecosystem. *Review of Palaeobotany and Palynology*.

TEDERSOO, L., BAHRAM, M., PÕLME, S., KÕLJALG, U., YOROU, N.S., WIJESUNDERA, R., RUIZ, L.V., VASCO-PALACIOS, A.M., THU, P.Q., SUIJA, A., SMITH, M.E., SHARP, C., SALUVEER, E., SAITTA, A., ROSAS, M., ET AL. (2014) Global diversity and geography of soil fungi. *Science*.

TEDERSOO, L., BAHRAM, M., PUUSEPP, R., NILSSON, R.H. & JAMES, T.Y. (2017) Novel soil-inhabiting clades fill gaps in the fungal tree of life. *Microbiome*.

TEDERSOO, L., SÁNCHEZ-RAMÍREZ, S., KÕLJALG, U., BAHRAM, M., DÖRING, M., SCHIGEL, D., MAY, T., RYBERG, M. & ABARENKOV, K. (2018) High-level classification of the Fungi and a tool for evolutionary ecological analyses. *Fungal Diversity* **90**, 135–159.

TELENIUS, H., CARTER, N.P., BEBB, C.E., NORDENSKJÖLD, M., PONDER, B.A.J. & TUNNACLIFFE, A. (1992)

Degenerate oligonucleotide-primed PCR: General amplification of target DNA by a single degenerate primer. *Genomics*.

THOMARAT, F., VIVARÈS, C.P. & GOUY, M. (2004) Phylogenetic analysis of the complete genome sequence of Encephalitozoon cuniculi supports the fungal origin of microsporidia and reveals a high frequency of fast-evolving genes. *Journal of Molecular Evolution* **59**, 780–791.

THOMSEN, H.A. (1978) On the identity of the Heliozoan Pinaciophora fluviatilis and Potamodiscus kalbei; with the description of eight new Pinaciophora species. *Protistologica*, 359–373.

TISCH, D. & SCHMOLL, M. (2010) Light regulation of metabolic pathways in fungi. *Applied Microbiology and Biotechnology*. .

TORRUELLA, G., DERELLE, R., PAPS, J., LANG, B.F., ROGER, A.J., SHALCHIAN-TABRIZI, K. & RUIZ-TRILLO, I. (2012) Phylogenetic relationships within the Opisthokonta based on phylogenomic analyses of conserved single-copy protein domains. *Molecular Biology and Evolution* **29**, 531–544.

TORRUELLA, G., GRAU-BOVÉ, X., MOREIRA, D., KARPOV, S.A., BURNS, J.A., SEBÉ-PEDRÓS, A., VÖLCKER, E. & LÓPEZ-GARCÍA, P. (2018) Global transcriptome analysis of the aphelid Paraphelidium tribonemae supports the phagotrophic origin of fungi. *Communications Biology* **1**, 1–11.

TORRUELLA, G., DE MENDOZA, A., GRAU-BOVÉ, X., ANTÓ, M., CHAPLIN, M.A., DEL CAMPO, J., EME, L., PÉREZ-CORDÓN, G., WHIPPS, C.M., NICHOLS, K.M., PALEY, R., ROGER, A.J., SITJÀ-BOBADILLA, A., DONACHIE, S. & RUIZ-TRILLO, I. (2015) Phylogenomics Reveals Convergent Evolution of Lifestyles in Close Relatives of Animals and Fungi. *Current Biology* **25**, 2404–2410.

TORRUELLA, G., SUGA, H., RIUTORT, M., PERETÓ, J. & RUIZ-TRILLO, I. (2009) The evolutionary history of lysine biosynthesis pathways within eukaryotes. *Journal of Molecular Evolution*.

TSAI, C.F., QIU, X. & LIU, J.H. (2003) A comparative analysis of two cDNA clones of the cellulase gene family from anaerobic fungus Piromyces rhizinflata. *Anaerobe*.

TSAOUSIS, A.D., KUNJI, E.R.S., GOLDBERG, A. V., LUCOCQ, J.M., HIRT, R.P. & EMBLEY, T.M. (2008) A novel route for ATP acquisition by the remnant mitochondria of Encephalitozoon cuniculi. *Nature* **453**, 553–556.

UEBELMESSER, E. RUTH (1956) Über einige neue Chytridineen aus Erdboden - Olpidium, Rhizophidium, Phlyctochytrium und Rhizophlyctis. *Archiv für Mikrobiologie*.

USTINOVA, I., KRIENITZ, L. & HUSS, V.A.R. (2000) Hyaloraphidium curvatum is not a green alga but a lower fungus; Amoebidium parasiticum is not a fungus, but a member of the DRIPs. *Protist*.

VALLE, L.G. & CAFARO, M.J. (2008) First report of zygospores in Asellariales and new species from the Caribbean. *Mycologia*.

VARGAS, M.M., ARONSON, J.M. & ROBERSON, R.W. (1993) The cytoplasmic organization of hyphal tip cells in the fungus Allomyces macrogynus. *Protoplasma*.

VÁVRA, J. & LARSSON, J.I.R. (2014) Structure of Microsporidia. In *Microsporidia: Pathogens of Opportunity: First Edition* p.

VÁVRA, J. & LUKEŠ, J. (2013) Microsporidia and 'The Art of Living Together'. In *Advances in Parasitology* pp. 253–319.

VIERECK, R. (2000) *‘ Zwar sind es weibliche Hände’: die Botanikerin und Pädagogin Catharina Helena Dörrien (1717-1795)*. Campus Verlag.

VISVESVARA, G.S., LEITCH, G.J., MOURA, H., WALLACE, S., WEBER, R. & BRYAN, R.T. (1991) Culture, electron microscopy, and immunoblot studies on a microsporidian parasite isolated from the urine of a patient with AIDS. *The Journal of protozoology*.

VIVARÈS, C.P. & MÉTÉNIER, G. (2000) Towards the minimal eukaryotic parasitic genome. *Current Opinion in Microbiology*. .

VIVIER, E. (1975) Vivier1975_The Microsporidia of the Protozoa.pdf. *Parazitologica*, 361.

VOGEL, H.J. (1964) Distribution of Lysine Pathways Among Fungi: Evolutionary Implications. *The American Naturalist*.

VREDENBURG, V.T., KNAPP, R.A., TUNSTALL, T.S. & BRIGGS, C.J. (2010) Dynamics of an emerging disease drive large-scale amphibian population extinctions. *Proceedings of the National Academy of Sciences of the United States of America*.

WAINWRIGHT, M., WICKRAMASINGHE, N.C., NARLIKAR, J. V. & RAJARATNAM, P. (2003) Microorganisms cultured from stratospheric air samples obtained at 41 km. *FEMS Microbiology Letters*.

WALLROTH, C.F.W. (1833) Flora Cryptogamica Germaniae. *Journal of the Elisha Mitchell Scientific Society* **2**, 1–923. J.L. Schrag, Nürnberg.

WANG, Y. & NAVIN, N.E. (2015) Advances and Applications of Single-Cell Sequencing Technologies. *Molecular Cell* **58**, 598–609. Elsevier Inc.

WATSON, J.D. & CRICK, F.H.C. (1953) Molecular structure of nucleic acids: A structure for deoxyribose nucleic acid. *Nature* **171**, 737–738.

WEBER, R., BRYAN, R.T., SCHWARTZ, D.A. & OWEN, R.L. (1994) Human microsporidial infections. *Clinical Microbiology Reviews*. .

WEETE, J.D., ABRIL, M. & BLACKWELL, M. (2010) Phylogenetic distribution of fungal sterols. *PLoS ONE*.

WEIDNER, E. (1972) Ultrastructural study of microsporidian invasion into cells. *Zeitschrift fur Parasitenkunde*.

WHISLER, H.C., ZEBOLD, S.L. & SHEMANCHUK, J.A. (1975) Life history of Coelomomyces psorophorae. *Proceedings of the National Academy of Sciences of the United States of America*.

WHITE, M.M., JAMES, T.Y., O’DONNELL, K., CAFARO, M.J., TANABE, Y. & SUGIYAMA, J. (2006) Phylogeny of the Zygomycota based on nuclear ribosomal sequence data. *Mycologia*.

WHITTAKER, R.H. (1969) New concepts of kingdoms of organisms. *Science*.

WICKETT, N.J., MIRARAB, S., NGUYEN, N., WARNOW, T., CARPENTER, E., MATASCI, N., AYYAMPALAYAM, S., BARKER, M.S., BURLEIGH, J.G., GITZENDANNER, M.A., RUHFEL, B.R., WAFULA, E., DER, J.P., GRAHAM, S.W., MATHEWS, S., ET AL. (2014) Phylotranscriptomic analysis of the origin and early diversification of land plants. *Proceedings of the National Academy of Sciences of the United States of America*.

WIDEMAN, J.G., MONIER, A., RODRÍGUEZ-MARTÍNEZ, R., LEONARD, G., COOK, E., POIRIER, C., MAGUIRE, F., MILNER, D.S., IRWIN, N.A.T., MOORE, K., SANTORO, A.E., KEELING, P.J., WORDEN, A.Z. & RICHARDS, T.A. (2020) Unexpected mitochondrial genome diversity revealed by targeted single-cell genomics of heterotrophic

flagellated protists. *Nature Microbiology*.

WIJAYAWARDENE, N.N., PAWŁOWSKA, J., LETCHER, P.M., KIRK, P.M., HUMBER, R.A., SCHÜßLER, A., WRZOSEK, M., MUSZEWSKA, A., OKRASIŃSKA, A., ISTEL, Ł., GĘSIORSKA, A., MUNGAI, P., LATEEF, A.A., RAJESHKUMAR, K.C., SINGH, R. V., ET AL. (2018) *Notes for genera: basal clades of Fungi (including Aphelidiomycota, Basidiobolomycota, Blastocladiomycota, Calcarisporiellomycota, Caulochytriomycota, Chytridiomycota, Entomophthoromycota, Glomeromycota, Kickxellomycota, Monoblepharomycota, Mortierellomyc*. In *Fungal Diversity* p.

WILLIAMS, B.A.P., HIRT, R.P., LUCOCQ, J.M. & EMBLEY, T.M. (2002) A mitochondrial remnant in the microsporidian Trachipleistophora hominis. *Nature* **418**, 865–869.

WILLIAMS, T.A., COX, C.J., FOSTER, P.G., SZÖLLŐSI, G.J. & EMBLEY, T.M. (2020) Phylogenomics provides robust support for a two-domains tree of life. *Nature Ecology & Evolution* **4**, 138–147.

WILLIAMS, T.A., FOSTER, P.G., COX, C.J. & EMBLEY, T.M. (2013) An archaeal origin of eukaryotes supports only two primary domains of life. *Nature*. .

WITTNER, M. & WEISS, L. (1999) *The Microsporidia and Microsporidiosis*. American Society of Microbiology.

WOESE, C.R. & FOX, G.E. (1977) Phylogenetic structure of the prokaryotic domain : The primary kingdoms **74**, 5088–5090.

WOESE, C.R., KANDLER, O. & WHEELIS, M.L. (1990) Towards a natural system of organisms: Proposal for the domains Archaea, Bacteria, and Eucarya. *Proceedings of the National Academy of Sciences of the United States of America*.

WORLEY, A.C., RAPER, K.B. & HOHL, M. (1979) Fonticula alba : A New Cellular Slime Mold (Acrasiomycetes). *Mycologia* **71**, 746–760.

WUJEK, D.E. (2015) Freshwater Silica-Scaled Heterotrophic Protista: Heliozoa, Thaumatomonad Flagellates, Amoebae, and Bicosoecids, From the Lake Itasca Region, Minnesota. *Journal of the Minnesota Academy of Science* **78**, 1–14.

WYLEZICH, C., KARPOV, S.A., MYLNIKOV, A.P., ANDERSON, R. & JÜRGENS, K. (2012) Ecologically relevant choanoflagellates collected from hypoxic water masses of the Baltic Sea have untypical mitochondrial cristae.

XU, Y. & WEISS, L.M. (2005) The microsporidian polar tube: A highly specialised invasion organelle. *International Journal for Parasitology*. .

YAHR, R., SCHOCH, C.L. & DENTINGER, B.T.M. (2016) Scaling up discovery of hidden diversity in fungi: Impacts of barcoding approaches. *Philosophical Transactions of the Royal Society B: Biological Sciences*. .

YAMAMOTO, K., ENDO, N., DEGAWA, Y., FUKUDA, M. & YAMADA, A. (2017) First detection of Endogone ectomycorrhizas in natural oak forests. *Mycorrhiza*.

YOON, H.S., PRICE, D.C., STEPANAUSKAS, R., RAJAH, V.D., SIERACKI, M.E., WILSON, W.H., YANG, E.C., DUFFY, S. & BHATTACHARYA, D. (2011) Single-cell genomics reveals organismal interactions in uncultivated marine protists. *Science*.

YOSHIDA, M., NAKAYAMA, T. & INOUYE, I. (2009) Nuclearia thermophila sp. nov. (Nucleariidae), a new nucleariid species isolated from Yunoko Lake in Nikko (Japan). *European Journal of Protistology* **45**, 147–155.

YOUSSEF, N.H., COUGER, M.B., STRUCHTEMEYER, C.G., LIGGENSTOFFER, A.S., PRADE, R.A., NAJAR, F.Z., ATIYEH, H.K., WILKINS, M.R. & ELSHAHED, M.S. (2013) The genome of the anaerobic fungus orpinomyces sp. strain c1a reveals the unique evolutionary history of a remarkable plant biomass degrader. *Applied and Environmental Microbiology*.

ZETTLER, L., GÓMEZ, F. & ZETTLER, E. (2002) Eukaryotic diversity in Spain's River of Fire. *Nature* **417**, 137.

ZETTLER, L.A.A., NERAD, T.A. & SOGIN, M.L. (2001) The Nucleariid Amoebae : More Protists at the Animal-Fungal Boundary **48**, 293–297.

ZOPF, W. (1885a) *Zur Morphologie und Biologie der niederen Pilzthiere (Monadinen), zugleich ein Beitrag zur Phytopathologie*. Leipzig :Veit & Comp.,.

ZOPF, W.F. (1885b) Die Pilzthiere oder Schleimpilze. *Handbuch der Botanik (Encykl Naturwiss)* **3**, 1–174. Breslau.

ZUCKERKANDL, E. & PAULING, L. (1965) Molecules as documents of evolutionary history. *Journal of theoretical biology* **8**, 357–366. Elsevier.

Additional work

Annex 1: Manuscript of annex 1

Ancient Adaptive Lateral Gene Transfers in the Symbiotic Opalina–Blastocystis Stramenopile Lineage

(Mol Biol Evol. 2019 Nov 6. pii: msz250. doi: 10.1093/molbev/msz250)

Naoji Yubuki,[1] Luis Javier Galindo,[1] Guillaume Reboul,[1] Purificación López-García,[1] Matthew W. Brown,[2,3] Nicolas Pollet,[4] and David Moreira*,[1]

[1]Unité d'Ecologie Systématique et Evolution, CNRS, Université Paris-Sud, AgroParisTech, Université Paris-Saclay, Orsay, France

[2]Department of Biological Sciences, Mississippi State University, Mississippi State, MS

[3]Institute for Genomics, Biocomputing and Biotechnology, Mississippi State University, Mississippi State, MS

[4]Laboratoire Evolution Génomes Comportement et Ecologie, CNRS, IRD, Université Paris-Sud, Université Paris-Saclay, Gif-sur-Yvette, France

*Corresponding author: E-mail: david.moreira@u-psud.fr.

Ancient Adaptive Lateral Gene Transfers in the Symbiotic Opalina - Blastocystis Stramenopile Lineage

Naoji Yubuki, Luis Xavier Galindo, Guillaume Reboul, Purificacion Lopez-Garcia, Matthew Brown, Nicolas Pollet, David Moreira

Ancient Adaptive Lateral Gene Transfers in the Symbiotic *Opalina* - *Blastocystis* Stramenopile Lineage

Naoji Yubuki[1], Luis Javier Galindo[1], Guillaume Reboul[1], Purificación López-García[1], Matthew W. Brown[2,3], Nicolas Pollet[4], and David Moreira[1]*

[1]Unité d'Ecologie Systématique et Evolution, CNRS, Université Paris-Sud, AgroParisTech, Université Paris-Saclay, Orsay, France
[2]Department of Biological Sciences, Mississippi State University, MS Mississippi State, MS, USA
[3]Institute for Genomics, Biocomputing and Biotechnology, Mississippi State University, Mississippi State, MS, USA
[4]Laboratoire Evolution Génomes Comportement et Ecologie, CNRS, IRD, Université Paris-Sud, Université Paris-Saclay, Gif-sur-Yvette, France

***Corresponding author:** E-mail: david.moreira@u-psud.fr.

Abstract

Lateral gene transfer (LGT) is a very common process in bacterial and archaeal evolution, playing an important role in the adaptation to new environments. In eukaryotes, its role and frequency remain highly debated, although recent research supports that gene transfer from bacteria to diverse eukaryotes may be much more common than previously appreciated. However, most of this research focused on animals and the true phylogenetic and functional impact of bacterial genes in less-studied microbial eukaryotic groups remains largely unknown. Here, we have analyzed transcriptome data from the deep-branching stramenopile Opalinidae, common members of frog gut microbiomes and distantly related to the well-known genus *Blastocystis*. Phylogenetic analyses suggest the early acquisition of several bacterial genes in a common ancestor of both lineages. Those LGTs most likely facilitated the adaptation of the free-living ancestor of the Opalinidae-*Blastocystis* symbiotic group to new niches in the oxygen-depleted animal gut environment.

Key words: Opalinids, *Blastocystis*, lateral gene transfer, gut microbiome.

Lateral gene transfer (LGT) plays an important role in prokaryotic evolution. LGT provides bacteria and archaea with the possibility to adapt, sometimes very rapidly, to new environments by obtaining genes from organisms already living in those environments. Although the significance of this phenomenon is widely recognized in prokaryotes, LGT-mediated gene acquisition from distant donors remains a contentious issue in eukaryotes (Martin 2017; Leger et al. 2018). Nevertheless, there is increasing evidence for LGT in eukaryotes from prokaryotes as well as from other eukaryotes (e.g., Keeling and Palmer 2008; Karnkowska et al. 2016; Eme et al. 2017; Husnik and McCutcheon 2017). A recent example concerns the stramenopile *Blastocystis,* which experienced LGTs from both eukaryotic and prokaryotic donors (Denoeud et al. 2011; Eme et al. 2017).

Blastocystis is recognized as the most widespread human gut eukaryotic parasite (Clark et al. 2013). This strict anaerobic and single-celled protist displays some unique and interesting biological features, such as the presence of unusual mitochondrion-related organelles (MRO) that display functions of mitochondria, hydrogenosomes and mitosomes (Stechmann et al. 2008). Some *Blastocystis* enzymes crucial for life in oxygen-depleted conditions were acquired by LGT from prokaryotes. For instance, the sulfur-mobilization (SUF) machinery involved in Fe-S protein maturation in the cytoplasm appears to have been acquired from archaeal Methanomicrobiales (Tsaousis et al. 2012). Furthermore, Eme et al. (2017) reported 74 purported cases of LGT mostly from prokaryotes to various subtypes of *Blastocystis* and suggested that several of the new LGT-acquired functions facilitated the metabolic adaptation of *Blastocystis* to the human gut in terms of metabolism but also to escape the immune defense mechanisms. The origins of those 74 gene families were very diverse. Although many of them were already present in the common ancestor of several *Blastocystis* subtypes, the time of their acquisition remained unclear due to the poor taxon sampling available for closely related stramenopile lineages.

Together with Alveolata and Rhizaria, Stramenopiles (or Heterokonta) constitute one of the main clades of the eukaryotic super-group SAR (Burki et al. 2007; Adl et al. 2019). Stramenopiles mostly encompass free-living phagotrophs or photosynthetic algae, but some are well-known parasites, such as the oomycetes and *Blastocystis*, or commensals, such as the Opalinidae (Patterson 1989; Andersen 2004). Ribosomal RNA phylogenetic analyses suggested a close relationship

between *Blastocystis* and Opalinidae, supporting the existence of a deep-branching symbiotic (parasitic/commensal) clade adapted to live in the gut of very diverse vertebrates (Silberman et al. 1996; Kostka et al. 2004; Li et al. 2018). However, despite the phylogenetic affinity of *Opalina* and *Blastocystis*, their morphological characteristics and lifestyles are very different. *Blastocystis* is characterized by a round unflagellated cell largely filled by a large vacuole. The cytoplasm and organelles are concentrated in the thin peripheral area between the vacuole and the cell membrane. Members of the genus *Blastocystis* live in the intestines of humans, birds, cows and pigs, most likely as parasites (Tan 2004). By contrast, members of the genus *Opalina* have a leaf-like cell shape with numerous nuclei and hundreds of short flagella on the cell surface, which is reminiscent of the cellular organization of ciliates. They live mainly in the intestine of anurans (frogs and toads) but seem to be innocuous to their hosts being therefore most often reported as commensal symbionts (Kostka 2016). Using the numerous flagella, *Opalina* members actively move in the intestine. All other known Opalinidae species are also commensal symbionts (Kostka 2016). Phylogenetic analyses have supported the monophyly of the Opalinidae-*Blastocystis* clade with the Placidida, a lineage of small free-living marine flagellates such as *Wobblia* and *Placidia* (Li et al. 2018; Shiratori et al. 2015, 2017; Derelle et al. 2016). Another free-living marine flagellate, *Cantina marsupialis*, is an anaerobic deep-branching relative that also possesses MROs (Yubuki et al. 2015). Since the closest relatives of Opalinidae and *Blastocystis* are all free-living, their ancestor was most likely free-living as well.

Here, we report the first transcriptome sequences from two Opalinidae strains, *Opalina* sp. OP10 and Opalinidae sp. Opal32, from two different continents (Europe and North America). OP10 and Opal32 cells were collected manually from the intestine of a *Xenopus tropicalis* frog and a *Lithobates sphenocephalus* tadpole, respectively. After transcriptome sequencing and assembly, we decontaminated the translated protein sequences inferred from the two transcriptomes to remove host and bacterial sequences (see Materials and Methods) and kept 7,232 and 18,765 proteins for OP10 and Opal32, respectively. Using BUSCO (Simão et al. 2015), we determined 33.3% transcriptome completeness for OP10 and 57.4% for Opal32. For comparison, we also applied BUSCO on the near-complete genome of *Blastocystis hominis* and determined a completeness of 75.2%, indicating a reduced genome as expected for a derived symbiont. We found in our datasets 44.3% (OP10) and 76.3%

(Opal32) of the *Blastocystis* proteome (supplementary table S1, Supplementary Material online), suggesting a rather good coverage especially for Opal32.

After transcriptome decontamination we searched with BLAST (Camacho et al. 2009) Opalinidae homologues of the 74 gene families likely acquired by LGT in *Blastocystis* (Eme et al. 2017). We recovered 37 and 38 of those LGT candidates in OP10 and Opal32, respectively. Thirty genes were common in both OP10 and Opal32, and seven and eight genes were unique in OP10 and Opal32, respectively. In total, 45 different candidate LGT genes were found in the two Opalinidae species. To verify that these sequences did not derive from the genomes of other gut microbes, we carried out two types of analyses. First, we investigated the codon usage of the coding sequences of both decontaminated transcriptomes and those of the LGT candidates and measured the frequency of optimal (F_{OP}) codons, which indicates the ratio of optimal (most frequent) codons to synonymous codons. The proportion of synonymous codons is unique to each genome and often results in a unimodal distribution of the F_{OP} score (Ikemura 1985), whereas the presence two F_{OP} peaks has been linked to contamination with bacterial sequences (Heinz et al. 2012). We obtained single-peak F_{OP} plots for our transcriptomes, indicating homogeneous codon usage and absence of contamination. All our LGT candidates fitted into these unimodal distributions supporting that they represent bona fide opalinid genes (supplementary figure S1, Supplementary Material online). Furthermore, their fit into the unimodal distribution supports an ancient integration of these LGT genes since they have adapted to the codon usage of the recipient genome. Second, we conducted phylogenetic analyses for all the LGT protein sequences. Phylogenetic trees showed that 29 of these proteins clustered robustly with their respective *Blastocystis* homologues (supplementary table S2 and supplementary figures S2-S30, Supplementary Material online). Those 29 proteins belonged to different functional families including carbohydrate metabolism, lipid metabolism, amino acid metabolism, and transporters. The phylogenetic analyses also allowed the identification of the donors of these sequences. Most of them had prokaryotic donors belonging to the Archaea, Proteobacteria and Actinobacteria, which are major components of frog gut microbiomes (Colombo et al. 2015). In some cases, the two Opalinidae species grouped with other eukaryotes belonging to the Amoebozoa, Excavata and Metazoa, suggesting eukaryote-to-eukaryote LGT, although it was impossible to infer from these trees whether the Opalinidae species were donors or recipients. Several of the LGT proteins most likely play important functions in the adaptation of Opalinidae to the anaerobic gut environment. One example is the mitochondrial iron-sulfur cluster (ISC) biogenesis system, essential for the assembly of iron-sulfur-containing proteins. These proteins are involved in a variety of metabolisms, including electron transport, nitrogen

fixation, and photosynthesis. In some protists living in low-oxygen environments, the canonical eukaryotic ISC machinery has been replaced by alternative bacterial machineries acquired via LGT, such as the nitrogen fixation (NIF) system and the bacterial sulfur mobilization (Suf) machinery. For instance, *Entamoeba histolytica* has a bacterial NIF system (van der Giezen et al. 2004), whereas *Monocercomonoides exilis*, which has completely lost mitochondria and the mitochondrial ISC pathway, contains a bacterial Suf system (Karnkowska et al. 2015). By contrast, *Blastocystis* has an archaeal-like SufC+SufB fused protein (Tsaousis et al. 2012). Similar fused *sufCB* genes related to Methanomicrobiales homologues were also identified in anaerobic flagellates such as the jakobid *Stygiella incarcerate* and the breviate *Pygsuia biforma* (Leger et al. 2016; Stairs et al. 2014). In prokaryotes, the *suf* operon is upregulated under oxidative stress (Outten et al. 2004), suggesting that the Suf machinery can be important for living in oxygen-depleted environments. We only identified an incomplete *sufB* gene in *Opalina*, which lacked a mitochondrial target signal. Similarly, SufCB is inferred to function in the cytosol in *Blastocystis*, *Pygsuia* and *Stygiella* (Tsaousis et al. 2012; Stairs et al. 2014; Leger et al. 2016). Our phylogenetic analysis showed that *Opalina* was closely related to these other anaerobic protists within a clade of Methanomicrobiales with robust support (fig. 1). These eukaryotes belong to three unrelated supergroups (*Opalina* and *Blastocystis* to SAR, *Pygsuia* to Breviatea, and *Stygiella* to Excavata). Therefore, one parsimonious explanation for this uneven distribution of SufCB is that one of these eukaryotic lineages first obtained the *sufC* and *sufB* genes from Methanomicrobiales, then both genes fused and, finally, the fused gene was transferred by eukaryote-to-eukaryote LGT to the other eukaryotic lineages. Since we only identified the *sufB* part in *Opalina*, it seems that it secondarily lost *sufC* after branching off from the lineages with fused *sufCB*. In fact, the well-supported separation of *Opalina* and *Blastocystis* in our tree (fig. 1) suggests that they have followed different evolutionary histories for the *sufCB* gene. Interestingly, the SufB and SufC proteins of *M. exilis* and *Paratrimastix pyriformis* are not related with the clade of *Opalina*, *Blastocystis*, *Pygsuia,* and *Stygiella*, indicating that they acquired these genes by independent LGT events from other prokaryotic donors. These genes were not identified in *C. marsupialis.*

In anoxic conditions, some eukaryotes use rhodoquinone instead of ubiquinone to receive electrons from NADH in the mitochondrial complex I of the electron

transport chain (ETC) and generate rhodoquinol (Castro-Guerrero et al. 2005; Sakai et al. 2012; Takamiya et al. 1999). Rhodoquinol is then reoxidized by the mitochondrial complex II catalyzing the reverse reaction as a fumarate reductase (van Hellemond and Tielens 1994; Tielens et al. 2002). This pathway helps to produce ATP and to reduce the respiratory chain without using the mitochondrial complexes III to V. The putative methyltransferase RquA is required for rhodoquinone biosynthesis (Lonjers et al. 2012) and its distribution among eukaryotes suggests that it is important for the adaptation of the mitochondrial metabolism to low-oxygen environments. In *Blastocystis,* RquA was suggested to be targeted to the MRO (Eme et al. 2017). We identified RquA homologues in both OP10 and Opal32 that also contained predicted mitochondrial-targeting sequences. By contrast, this protein seemed to be absent in *C. marsupialis.* RquA is not very common in eukaryotes and previous phylogenetic analyses demonstrated that RquA-containing eukaryotes are scattered among prokaryotic lineages, mostly Proteobacteria. Stairs et al. (2018) proposed that LGT of *rquA* genes from bacteria to eukaryotes occurred at least twice before subsequent multiple independent LGTs among eukaryotes. Our updated RquA phylogeny (fig. 2) is consistent with this proposal. We retrieved two major clades, A and B: *Opalina* spp. branched together with *Proteromonas* and *Blastocystis* in clade A, composed mostly of alpha- and beta-proteobacteria, and several other eukaryotes (Breviata, Amoebozoa and Euglenida). Group B also contained some eukaryotes (choanoflagellates, diatoms, and ciliates) embedded among bacteria, again mostly alpha- and beta-proteobacteria. The presence of alphaproteobacteria close to the eukaryotic sequences opens the possibility of a mitochondrial origin by endosymbiotic gene transfer (EGT). Nevertheless, several observations argue against this hypothesis: (i) the eukaryotic sequences are not monophyletic, (ii) several eukaryotic sequences appear to be closer to betaproteobacteria than to alphaproteobacteria, and (iii) if *rquA* was present in the last eukaryotic common ancestor (which already had mitochondria), it must have been lost independently many times to result in its current patchy distribution. Thus, the available data so far rather support the origin of eukaryotic *rquA* by LGT from bacteria followed by subsequent LGTs among eukaryotes.

In most mitochondria, coenzyme A is transferred from acetyl-CoA to succinate by two types of acetate:succinate CoA-transferases (ASCT1B and ASCT1C). The resulting succinyl-CoA is used for ATP production by succinyl-CoA synthetase

(SCS). This ASCT/SCS system plays a crucial role in MROs of protists living in anoxic environments, such the human urogenital parasite *Trichomonas vaginalis*, for the production of ATP by substrate-level phosphorylation independent of the mitochondrial Krebs cycle (van Grinsven et al. 2008). In the case of the free-living amoeboflagellate *Naegleria gruberi*, which contains classical mitochondria and transiently experiences low-oxygen conditions, ASCT was predicted to function in mitochondria (Fritz-Laylin et al. 2010). We identified an ASCT/SCS system in our *Opalina* transcriptomes. In contrast with the *Blastocystis* ASCT, which has an MRO-targeting sequence, the *Opalina asct1C* and *asct1B* were incomplete ORFs and did not contain any recognizable mitochondrial targeting signal. The ASCT1C phylogenetic tree (fig. 3) recovered *Opalina* and *Blastocystis* grouped within a large clade also containing trichomonads, *Naegleria*, fungi, and dictyostelid cellular slime molds (Amoebozoa). This eukaryotic clade was closely related to Deltaproteobacteria and Firmicutes. As in the previous cases described above, this tree suggests a bacterial origin of the gene followed by eukaryote-to-eukaryote LGT.

To carry out a more comprehensive comparison of the mitochondrial metabolism of *Opalina* with that of other MRO-containing anaerobic stramenopiles (the parasitic *Blastocystis* and the free-living *C. marsupialis* (Stechmann et al. 2008; Noguchi et al. 2015)), we used BLAST to search for homologues of MRO proteins of these organisms in *Opalin*a. We also manually annotated the *Opalina* mitochondrial proteins involved in major energy metabolism pathways. As shown above, *Opalina* obtained many genes for typical MRO anaerobic metabolism by LGT from either prokaryotes or other eukaryotes, but it also contains typical mitochondrial genes vertically inherited (supplementary tables S2 and S3, Supplementary Material online). *Blastocystis* spp. and *C. marsupialis* completely lack complexes III and IV, and F1Fo ATPase (complex V) (Gentekaki et al. 2017; Noguchi et al. 2015). *Opalina* possesses some genes of the tricarboxylic acid (TCA) cycle, complex I (NADH:ubiquinone oxidoreductase), and complex II (succinate dehydrogenase) of the ETC, but does not seem to encode any other recognizable canonical components such as complexes III and IV or the F1Fo ATPase (supplementary table S4, Supplementary Material online). This suggests that *Opalina* has a partial ETC that does not appear to function in energy generation. Data from *Blastocystis* and *Pygsuia* suggest that complex II functions in reverse as a fumarate reductase to regenerate the quinone pool under anaerobic conditions without using complex III, IV

and F1Fo ATPase to conduct oxidative phosphorylation. RquA, acquired by LGT in *Opalina* (see above), is the crucial enzyme for this alternative electron transport machinery. *Opalina* also possesses genes involved in classical mitochondrial activities, including transporters, fatty acid metabolism, amino acid metabolism, pyruvate metabolism, and [2Fe-2S] ferredoxin for FeS cluster assembly, some of which are lost in *Blastocystis*. (supplementary table S2, Supplementary Material online). By contrast, we did not identify in *Opalina* some essential mitochondrial proteins, such as those involved in the eukaryotic iron-sulfur cluster (ISC) synthesis system and several enzymes (pyruvate:ferredoxin oxidoreductase (PFO), [FeFe] hydrogenase (HydA), the HydA hydrogenase maturases HydE, HydF and HydG, and two subunits of the NADH:ubquinone oxidoreductase (NuoE and NuoF)) that are hallmarks of the MROs found in many anaerobic protists, including *Blastocystis and Cantina*. In those organisms, PFO oxidizes pyruvate to acetyl-CoA and CO_2. The reduced ferredoxin is reoxidized by HydA that reduces protons to H_2 gas. In *Opalina*, which lacks HydA, the pyruvate:NADP+ oxidoreductase (PNO), instead of PFO, presumably oxidizes pyruvate to acetyl-CoA and, then, acetyl-CoA can be utilized by the ASCT/SCS system to generate ATP by substrate-level phosphorylation. Since PFO and HydA are present in *Blastocystis*, we can propose two evolutionary scenarios: First, these two enzymes were present in the common ancestor of *Opalina* and *Blastocystis* and secondarily lost in the *Opalina* lineage or, second, they were obtained in *Blastocystis* independently after it diverged from the *Blastocystis-Opalina* common ancestor. As in the case of *Blastocystis* and *Cantina*, we did not identify a pyruvate carrier in *Opalina*. Glycolysis is described as a cytosolic process in eukaryotes and its product, pyruvate, is imported into the mitochondrion by the pyruvate carrier. However, the second half of glycolysis in some stramenopiles has been predicted to occur in both the cytosol and mitochondria/MRO (Abrahamian et al. 2017). Moreover, in *Blastocystis* this second half of the glycolysis is solely localized in the MRO (Rártulos et al. 2018). Similarly, we identified in *Opalina* several enzymes of the second half of the glycolsis (glyceraldehyde phosphate dehydrogenase (GAPDH), phosphoglycerate kinase (PGK), and enolase (ENO)) with mitochondria-targeting signals (supplementary table S3, Supplementary Material online). Despite these similarities and other shared key adaptations to the oxygen-depleted gut environment, *Opalina* appears to have kept a less derived version of

the mitochondrial metabolism than its sister lineage *Blastocystis* and the stramenopile relative *Cantina*.

Conclusion

Our examination of two *Opalina* transcriptomes based on sequence similarity searches and phylogenetic analyses identified 29 genes likely acquired by LGT by a common ancestor of both *Blastocystis* and Opalinidae (supplementary table S1, Supplementary Material online). Among these genes, those coding for the Suf, RquA and ASCT proteins play important roles in anaerobic metabolism in MROs. It is unclear when a common ancestor of these organisms entered the animal gut but some of the LGTs most likely facilitated the adaptation to this new oxygen-deprived environment before the divergence of these two lineages. *Blastocystis* MROs combine metabolic properties of both mitochondria and hydrogenosomes and contain PFO and [FeFe] hydrogenase as well as incomplete TCA cycle and the complexes I and II (Gentekaki et al. 2017; Stechmann et al. 2008). Although *Opalina* shares with *Blastocystis* many enzymes involved in anaerobic metabolisms acquired via LGT and both lineages have several metabolic modifications in common (incomplete TCA cycles and absence of complexes III and IV and F1Fo ATPase), our data suggest the absence of the typical hydrogenosomal enzymes PFO and [FeFe] hydrogenase. This important difference indicates that *Blastocystis* has achieved a more derived adaptation to hypoxic condition than Opalinidae. *Opalina* represents therefore an excellent model of intermediate adaptation between conventional aerobic mitochondria and derived anaerobic MROs and can help to understand the initial steps in the evolutionary path between both types of organelles.

Materials and Methods

Isolation of *Opalina* sp. Cells

For OP10 strains, the gut content of a *Xenopus tropicalis* frog was collected and resuspended in sterile PBS buffer. Eight *Opalina* cells were manually isolated under an inverted Leica DMI3000 microscope equipped with an Eppendorf TransferMan 4r micromanipulator. The cells were rinsed twice in sterile PBS and finally resuspended in 1.5 µl of sterile water. For Opal32 strain, a smear of ca. 100 µl of *Lithobates sphenocephalus* tadpole gut contents was placed onto a sterile Petri dish and 500 µl of sterile amphibian Ringer's solution (ARS: in 1 L distilled water, 6.6 g NaCl, 0.15 g KCl, 0.15 g CaCl2, and 0.2 g NaHCO3) was added to the drop of gut content. Roughly 10 µl of this solution was examined under a Zeiss AxioSkop Plus upright microscope, and cells were imaged. A single cell was manually isolated using a micropipetter and washed six times in 100 µl of fresh and sterile ARS. The cell was then transferred to a 0.5 µl to nuclease-free PCR tube and processed as below.

Opalina sp. Transcriptome Sequencing and Assembly

For Opalina sp. OP10, RNA extraction, cDNA synthesis and amplification were done using the REPLI-g WTA Single Cell kit following the manufacturer's protocol (Qiagen). The resulting cDNA was sequenced using Illumina HiSeq 2500 paired-end sequencing (2x125 bp). For Opal32, the cell was subjected to a modified version of SmartSeq-2 (Picelli et al. 2014, Kang et al. 2017) and full-length cDNA was constructed. This cDNA was then sheared using a Covaris focused-ultrasonicator (Duty% 10, Intesity 5, Burst Cycle 200, Time 30s, Frequency Sweeping Mode). This sheared cDNA was prepped using NEBnext Ultra DNA library kit for Illumina (New England Biolabs) and sequenced on an Illumina MiSeq paired-end (2x300 bp) sequencing run. For both datasets, Illumina adapters were removed using Trimmomatic v. 0.36 (Bolger et al. 2014) and paired-end sequences were assembled using Trinity v.2.2.0 (Haas et al. 2013) with default parameters. A total of 24,170 assembled transcripts were obtained from OP10 and 16,943 from Opal32.

Transcriptome Decontamination and Completeness

The decontamination of the two transcriptomes was carried out by a three-step process. First, the transcriptome sequences were subjected to two rounds of assembly, before and after bacterial sequence removal by BlobTools v0.9.19 (Laetsch et al. 2017). Second, open-reading frames were predicted and translated from the assembled transcripts using Transdecoder v2 (http:transdecoder.github.io) to produce protein sequences for OP10 and Opal32. Finally, to remove possible host sequences, the predicted protein sequences were searched by BLASTp (Camacho et al. 2009) against two predicted anuran proteomes. We used *Xenopus tropicalis* v9.1 for OP10 and, because of the lack of a proteome from the host species of Opal32 (*Lithobates sphenocephalus*) we used *Rana catesbeiana* RCv2.1, which is the closest member of the same Ranidae family with available sequence data. At the end, we obtained 8,432 and 11,480 protein sequences from OP10 and Opal32, respectively.

To assess transcriptome completeness, we used BUSCO v2.0.1 (Simão et al. 2015) on the decontaminated predicted proteins with the eukaryote_odb9 dataset of 303 near-universal single-copy orthologs. As an additional step of quality completeness comparison, we calculated the completeness value of the near-complete genome of *Blastocystis hominis* (ASM15166v1) and compared it with the opalinid data.

Codon usage for the coding sequences of both transcriptomes and their LGT candidates were measured using the index of frequency of optimal (F_{OP}) codons (Ikemura 1985). We calculated F_{OP} values using CodonW (Peden 2005) with default settings and generated F_{OP} plots using R (http://www.r-project.org).

Identification of LGT Candidates and Phylogenetic Analysis

We used the 74 LGT proteins of *Blastocystis* sp. ST1 Nand II (Eme et al. 2017) as queries to identify Opalinidae homologs using BLASTp searches (Camacho et al. 2009) with an e-value cutoff of 1e-05. 37 and 38 proteins yielded hits in the OP10 and Opal32 protein databases, respectively. Of these, 30 were found in both transcriptomes and 7 and 8 were unique to OP10 and Opal32, respectively. In total, 45 proteins were recovered from the two strains as LGT candidates. To reconstruct their phylogenies, we searched these proteins by BLASTp against the non-redundant GenBank database with an e-value cutoff of 1e-05 and maximum of 2,000

hits. To reduce the dataset size for subsequent phylogenetic analysis, hit sequences were clustered by CD-HIT (Limin et al. 2012) at 95% similarity. The resulting 45 protein sequence datasets were aligned using MAFFT v7.388 with default settings (Katoh and Stanley 2013). Ambiguously aligned sites were removed using trimAl v1.4.rev15 (Capella-Gutierrez et al. 2009) with -automated1 setting prior to phylogenetic analyses. Preliminary phylogenies were reconstructed using FastTree 2.1.7 (Price et al. 2010) and inspected manually to reduce the size of the data set by keeping only a few representatives for the prokaryotic clades distantly related to the eukaryotic sequences. We thus identified 29 proteins from the two Opalinidae strains as LGT candidates. The final datasets were aligned and trimmed as described above. Maximum likelihood phylogenetic trees for each dataset were constructed using IQ-TREE (Nguyen et al. 2015) with the best fitting model determined by applying the Bayesian Information Criterion (BIC) with the -m MFP (model selection) with default settings for each dataset. Branch supports were calculated with 1,000 ultrafast bootstrap replicates.

Protein cellular localization was predicted using TargetP 1.1 (Emanuelesson et al. 2000), MitoFates (Fukasawa et al. 2015) and TPpred 2.0 (Savojardo et al. 2014) with default settings. Homologs of mitochondrial proteins in *Opalina* sp. OP10 were searched with BLASTp using MRO sequences from two close relatives: *Blastocystis* (Stechmann et al. 2008) and *Cantina marsupialis* (Noguchi et al. 2015) (supplementary table S3, Supplementary Material online).

Data Availability

Protein sequence data sets used in this work, including complete and trimmed alignments and phylogenetic trees, are available for download at figshare (10.6084/m9.figshare.9746360). *Opalina* sequences have been submitted to GenBank (for accession numbers, see supplementary tables S2 and S3, Supplementary Material online).

Supplementary Material

Supplementary data are available at Molecular Biology and Evolution online.

Acknowledgments

This study was supported by European Research Council grants ProtistWorld (P.L.-G., agreement no. 322669) and Plast-Evol (D.M. agreement no. 787904), the Agence Nationale de la Recherche (D.M., project ANR-15-CE32-0003 “ANCESSTRAM”) and the Institut Diversité Ecologie et Evolution du Vivant (D.M. and N.P.). This project was supported in part by the United States National Science Foundation (NSF) Division of Environmental Biology (DEB) grant 1456054 (http://www.nsf.gov), awarded to M.W.B. We thank Prof. Ronald Altig (Mississippi State University) for collection of tadpoles and Andrew J. Roger, Courtney Stairs, and Marlena Dlutek (Dalhousie University, Canada) for *Opalina* sp. Opal32 transcriptome sequencing, which was supported by grant MOP-142349 from the Canadian Institutes of Health Research awarded to A.J. Roger. We thank two anonymous reviewers for constructive comments.

References

Abrahamian M, Kagda M, Ah-Fong, AMV, Judelson HS 2017. Rethinking the evolution of eukaryotic metabolism: novel cellular partitioning of enzymes in stramenopiles links serine biosynthesis to glycolysis in mitochondria. *BMC Evol Biol*. 17:241.

Adl SM, Bass D, Lane CE, Lukeš J, Schoch CL, Smirnov A, Agatha S, Berney C, Brown MW, Burki F, et al. 2019. Revisions to the classification, nomenclature, and diversity of eukaryotes. *J Eukaryot Microbiol*. 66(1):4-119.

Andersen RA. 2004. Biology and systematics of heterokont and haptophyte algae. *Am J Bot*. 91:1508–1522.

Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*. 30:2114-2120.

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

Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL. 2009. BLASTþ: architecture and applications. *BMC Bioinformatics* 10:421.

Capella-Gutierrez S, Silla-Martinez JM, Gabaldon T. 2009. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* 25:1972-1973

Castro-Guerrero NA, Jasso-Chávez R, Moreno-Sánchez R. 2005. Physiological role of rhodoquinone in *Euglena gracilis* mitochondria. *Biochim Biophys Acta.* 1710:113-121.

Clark CG, van der Giezen M, Alfellani MA, Stensvold CR. 2013. Recent developments in *Blastocystis* research. *Adv Parasitol.* 82:1-32

Colombo BM, Scalvenzi T, Benlamara S, Pollet N. 2015. Microbiota and mucosal immunity in amphibians. *Front Immunol*. 6:111.

Denoeud F, Roussel M, Noel B, Wawrzyniak I, Da Silva C, Diogon M, Viscogliosi E, Brochier-Armanet C, Couloux A, Poulain J, et al. 2011. Genome sequence of the stramenopile *Blastocystis*, a human anaerobic parasite. *Genome Biol*. 12:R29.

Derelle R, López-García P, Timpano H, Moreira D. 2016 A phylogenomic framework to study the diversity and evolution of stramenopiles (=heterokonts). *Mol Biol Evol*. 33:2890-2898.

Emanuelsson O, Nielsen H, Brunak S, von Heijne G. 2000. Predicting subcellular localization of proteins based on their N-terminal amino acid sequence. *J Mol Biol*. 300:1005-1016.

Eme L, Gentekaki E, Curtis B, Archibald J, Roger A. 2017. Lateral gene transfer in the adaptation of the anaerobic parasite *Blastocystis* to the gut. *Curr Biol*. 27:807–820.

Fritz-Laylin LK, Prochnik SE, Ginger ML, Dacks JB, Carpenter ML, Field MC, Kuo A, Paredez A, Chapman J, Pham J, et al. 2010. The genome of *Naegleria gruberi* illuminates early eukaryotic versatility. *Cell* 140:631–642.

Fukasawa Y, Tsuji J, Fu S, Tomii K, Horton P, Imai K. 2015. MitoFates: improved prediction of mitochondrial targeting sequences and their cleavage sites. *Mol Cell Proteomics*.14:1113-1126.

Gentekaki E, Curtis BA, Stairs CW, Klimeš V, Elias M, Salas-Leiva DE, Herman EK, Eme L, Arias MC, Henrissat B, et al. 2017. Extreme genome diversity in the hyper-prevalent parasitic eukaryote *Blastocystis*. *PLoS Biol*. 15:e2003769–42.

Haas BJ, Papanicolaou A, Yassour M, Grabherr M, Blood PD, Bowden J, Couger MB, Eccles D, Li B, Lieber M, MacManes MD, Ott M, Orvis J, Pochet N, Strozzi F, Weeks N, Westerman R, William T, Dewey CN, Henschel R, LeDuc RD, Friedman N, Regev A. 2013. De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nat Protoc*. 8:1494-1512.

Heinz E, Williams TA, Nakjang S, Noël CJ, Swan DC, Goldberg AV, Harris SR. 2012. The genome of the obligate intracellular parasite *Trachipleistophora hominis*: new insights into microsporidian genome dynamics and reductive evolution. *PLoS Pathog*. 8:e1002979.

Husnik F, McCutcheon JP. 2018. Functional horizontal gene transfer from bacteria to eukaryotes. *Nat Rev Micro.* 16:67–79.

Ikemura T. 1985. Codon usage and tRNA content in unicellular and multicellular organisms. *Mol Biol Evol*. 2:13–34.

Kang S, Tice AK, Spiegel FW, Silberman JD, Pánek T, Čepička I, Kostka M, Kosakyan A, Alcântara DM, Roger AJ, Shadwick LL, Smirnov A, Kudryavstev A, Lahr DJG, Brown MW. 2017. Between a pod and a hard test: the deep evolution of amoebae. *Mol Biol Evol*. 34:2258-2270.

Karnkowska A, Vacek V, Zubáčová Z, Treitli SC, Petrželková R, Eme L, Novák L, Žárský V, Barlow LD, Herman EK, Soukal P, Hroudová M, Dolezal P, Stairs CW, Roger AJ, Elias M, Dacks JB, Vlček C, and Hampl V. 2016. A eukaryote without a mitochondrial organelle. *Curr Biol* 26:1274–1284.

Katoh K, Standley DM. 2013. MAFFT Multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol.* 30:772–780

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

Kostka M. 2016. Opalinidae. In: Archibald JM, Simpson AGB, Slamovits CH, Margulis L, editors. Handbook of the Protists. Springer, Charm, Switzerland.

Kostka M, Hampl V, Cepicka I, Flegr J. 2004. Phylogenetic position of *Protoopalina intestinalis* based on SSU rRNA gene sequence. *Mol Phylogenet Evol.* 33:220–224.

Laetsch DR, Blaxter ML, Leggett RM. 2017. BlobTools: Interrogation of genome assemblies. *F1000Research.* 6:1287.

Leger MM, Eme L, Hug LA, Roger AJ. 2016. Novel hydrogenosomes in the microaerophilic jakobid *Stygiella incarcerata*. *Mol Biol Evol*. 33:2318–2336.

Leger MM, Eme L, Stairs CW, Roger AJ. 2018. Demystifying eukaryote lateral gene transfer. *Bioessays*. 40:e1700242.

Li M, Ponce-Gordo F, Grim JN, Li C, Zou H, Li W, Wu S, Wang G. 2018. Morphological redescription of *Opalina undulata* Nie 1932 from *Fejervarya limnocharis* with molecular phylogenetic study of Opalinids (Heterokonta, Opalinea). *J Eukaryot Microbiol.* 65:783-791

Limin F, Beifang N, Zhengwei Z, Sitao W, Weizhong L, 2012. CD-HIT: accelerated for clustering the next generation sequencing data. *Bioinformatics* 28:3150-3152.

Lonjers ZT, Dickson EL, Chu TPT, Kreutz JE, Neacsu FA, Anders KR, Shepherd JN. 2012. Identification of a new gene required for the biosynthesis of rhodoquinone in *Rhodospirillum rubrum*. *J Bacteriol*. 194:965–971.

Martin WF. 2017. Too much eukaryote LGT. *Bioessays*. 39:1700115.

Noguchi F, Shimamura S, Nakayama T, Yazaki E, Yabuki A, Hashimoto T, Inagaki Y, Fujikura K, Takishita K. 2015. Metabolic capacity of mitochondrion-related organelles in the free-living anaerobic stramenopile *Cantina marsupialis*. *Protist* 166:534–550.

Nguyen LT, Schmidt HA, von Haeseler A, Minh BQ. 2015. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol Biol Evol*. 32:268-274.

Outten FW, Djaman O, Storz G. 2004. A suf operon requirement for Fe-S cluster assembly during iron starvation in *Escherichia coli*. *Mol Microbiol.* 52:861–872.

Peden J 2005. CodonW version 1.4.2. http://codonw.sourceforge.net/

Picelli S, Faridani OR, Björklund ÅK, Winberg G, Sagasser S,Sandberg R. 2014. Full-length RNA-Seq from single cells using Smart-seq2. *Nat. Protocols* 9:171-181.

Price MN, Dehal PS, Arkin AP (2010) FastTree 2 – Approximately maximum-likelihood trees for large alignments. *PLoS One* 5(3):e9490.

Patterson DJ. 1989. Stramenopiles: Chromophytes from a protistan perspective. In: Green JC, Leadbeater BSC, Diver WL, editors. The chromophyte algae: problems and perspectives. Clarendon Press, Oxford. p357–379.

Río Bártulos C, Rogers MB, Williams TA, Gentekaki E, Brinkmann H, Cerff R, Liaud M-F, Hehl AB, Yarlett NR, Gruber A, Kroth PG, van der Giezen M. 2018. Mitochondrial glycolysis in a major lineage of eukaryotes. *Genome Biol Evol*. 10:2310-2325.

Sakai C, Tomitsuka E, Esumi H, Harada S, Kita K. 2012. Mitochondrial fumarate reductase as a target of chemotherapy: from parasites to cancer cells. *Biochim Biophys Acta.* 1820:643-651.

Savojardo C, Martelli PL, Fariselli P, Casadio R. 2014. TPpred2: improving the prediction of mitochondrial targeting peptide cleavage sites by exploiting sequence motifs. *Bioinformatics* 30:2973-2974.

Shiratori T, Nakayama T, Ishida K-I. 2015. A new deep-branching stramenopile, *Platysulcus tardus* gen. nov., sp. nov. *Protist* 166:337–348.

Shiratori T, Thakur R, Ishida K-I. 2017. *Pseudophyllomitus vesiculosus* (Larsen and Patterson 1990) Lee, 2002, a poorly studied phagotrophic biflagellate is the first characterized member of stramenopile environmental clade MAST-6. *Protist* 168:439–451.

Silberman JD, Sogin ML, Leipe DD, Clark CG. 1996. Human parasite finds taxonomic home. *Nature* 380:398–398.

Simão FA, Waterhouse RM, Ioannidis P, Kriventseva E V., Zdobnov EM. 2015. BUSCO: Assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics.* 31:3210–3212.

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*. 24:1176-1186.

Stairs CW, Eme L, Muñoz-Gómez SA, Cohen A, Dellaire G, Shepherd JN, Fawcett JP, Roger AJ. 2018. Microbial eukaryotes have adapted to hypoxia by horizontal acquisitions of a gene involved in rhodoquinone biosynthesis. *eLife* 7:e34292

Stechmann A, Hamblin K, Perez-Brocal V, Gaston D, Richmond GS, van der Giezen M, Clark CG, Roger AJ. 2008. Organelles in *Blastocystis* that blur the distinction between mitochondria and hydrogenosomes. *Curr Biol.* 18:580–585.

Takamiya S, Matsui T, Taka H, Murayama K, Matsuda M, Aoki T. 1999. Free-living nematodes *Caenorhabditis elegans* possess in their mitochondria an additional rhodoquinone, an essential component of the eukaryotic fumarate reductase system. *Arch Biochem Biophys*. 371:284-289.

Tan KSW. 2004. *Blastocystis* in humans and animals: new insights using modern methodologies. *Vet Parasitol*. 126:121–144.

Tielens AGM, Rotte C, van Hellemond JJ, Martin W. 2002. Mitochondria as we don't know them. *Trends Biochem Sci*. 27:564–572.

Tsaousis AD, de Choudens SO, Gentekaki E, Long S, Gaston D, Stechmann A, Vinella D, Py B, Fontecave M, Barras F. 2012. Evolution of Fe/S cluster biogenesis in the anaerobic parasite *Blastocystis*. *Proc Natl Acad Sci USA* 109:10426–10431.

van der Giezen M, Cox S, Tovar J. 2004. The iron-sulfur cluster assembly genes iscS and iscU of *Entamoeba histolytica* were acquired by horizontal gene transfer. *BMC Evol Biol.* 4:7.

van Grinsven KWA, Rosnowsky S, van Weelden SWH, Pütz S, van der Giezen M, Martin W, van Hellemond JJ, Tielens AGM, Henze K. 2008. Acetate:Succinate CoA-transferase in the hydrogenosomes of *Trichomonas vaginalis. J Biol Chem*. 283:1411–1418.

van Hellemond JJ, Tielens AGM 1994. Expression and functional properties of fumarate reductase. *Biochem J.* 304:321–331.

Yubuki N, Pánek T, Yabuki A, Cepicka I, Takishita K, Inagaki Y, Leander BS. 2015. Morphological identities of two different marine stramenopile environmental sequence clades: *Bicosoeca kenaiensis* (Hilliard, 1971) and *Cantina marsupialis* (Larsen and Patterson, 1990) gen. nov., comb. nov. *J Eukaryot Microbiol*. 62:532–542.

Figure Legends

FIG. 1. Maximum likelihood phylogenetic tree of SufCB (188 sequences). Bootstrap values <50% are not shown. The long branch of *Paratrimastix* and *Monocercomonoides* has been shortened to 1/4. For the complete tree see supplementary figure S2, Supplementary Material online.

FIG. 2. Maximum likelihood phylogenetic tree of RquA (102 sequences). Bootstrap values <50% are not shown. Groups A and B are defined according to Stairs et al. 2018. For the complete tree see supplementary figure S3, Supplementary Material online.

FIG. 3. Maximum likelihood phylogenetic tree of ASCT1C (96 sequences). Bootstrap values <50% are not shown. The branch of *Schizosaccharomyces cryophilus* has been shortened to 1/2. For the complete tree see supplementary figure S4, Supplementary Material online.

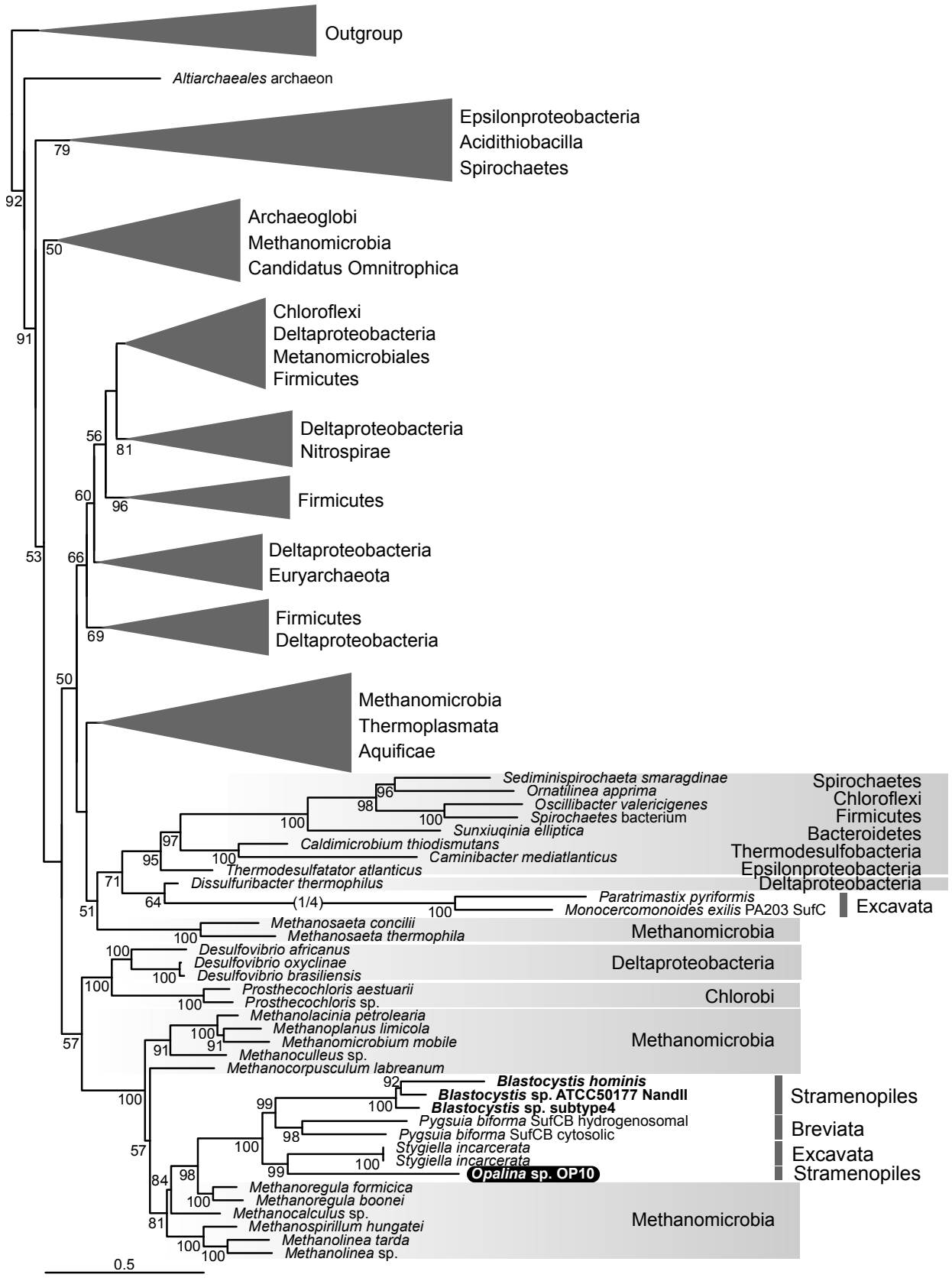

Outgroup
Altiarchaeales archaeon
Epsilonproteobacteria
Acidithiobacilla
Spirochaetes
Archaeoglobi
Methanomicrobia
Candidatus Omnitrophica
Chloroflexi
Deltaproteobacteria
Metanomicrobiales
Firmicutes
Deltaproteobacteria
Nitrospirae
Firmicutes
Deltaproteobacteria
Euryarchaeota
Firmicutes
Deltaproteobacteria
Methanomicrobia
Thermoplasmata
Aquificae
Sediminispirochaeta smaragdinae
Ornatilinea apprima
Oscillibacter valericigenes
Spirochaetes bacterium
Sunxiuqinia elliptica
Caldimicrobium thiodismutans
Caminibacter mediatlanticus
Thermodesulfatator atlanticus
Dissulfuribacter thermophilus
Paratrimastix pyriformis
Monocercomonoides exilis PA203 SufC
Methanosaeta concilii
Methanosaeta thermophila
Desulfovibrio africanus
Desulfovibrio oxyclinae
Desulfovibrio brasiliensis
Prosthecochloris aestuarii
Prosthecochloris sp.
Methanolacinia petrolearia
Methanoplanus limicola
Methanomicrobium mobile
Methanoculleus sp.
Methanocorpusculum labreanum
Blastocystis hominis
Blastocystis sp. ATCC50177 NandII
Blastocystis sp. subtype4
Pygsuia biforma SufCB hydrogenosomal
Pygsuia biforma SufCB cytosolic
Stygiella incarcerata
Stygiella incarcerata
Opalina sp. OP10
Methanoregula formicica
Methanoregula boonei
Methanocalculus sp.
Methanospirillum hungatei
Methanolinea tarda
Methanolinea sp.
0.5
Spirochaetes
Chloroflexi
Firmicutes
Bacteroidetes
Thermodesulfobacteria
Epsilonproteobacteria
Deltaproteobacteria
Excavata
Methanomicrobia
Deltaproteobacteria
Chlorobi
Methanomicrobia
Stramenopiles
Breviata
Excavata
Stramenopiles
Methanomicrobia

Figure 1

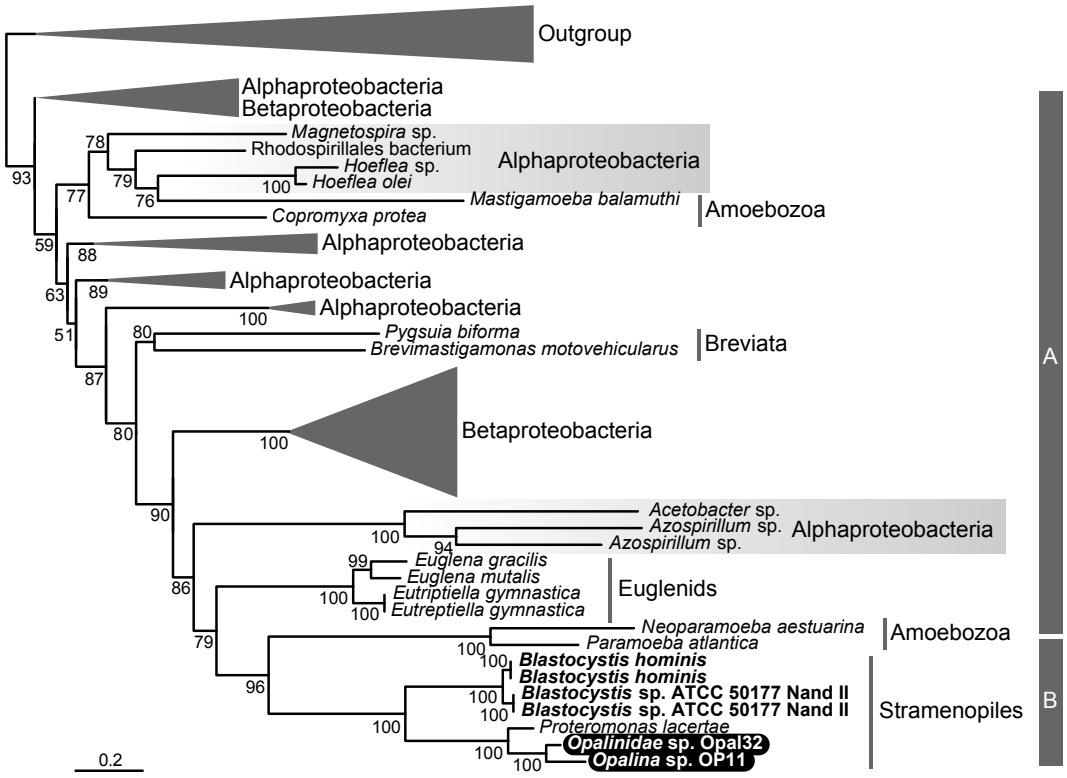

Outgroup
Alphaproteobacteria
Betaproteobacteria
Magnetospira sp.
Rhodospirillales bacterium
Hoeflea sp.
Hoeflea olei
Alphaproteobacteria
Mastigamoeba balamuthi
Copromyxa protea
Amoebozoa
Alphaproteobacteria
Alphaproteobacteria
Alphaproteobacteria
Pygsuia biforma
Brevimastigamonas motovehicularus
Breviata
Betaproteobacteria
Acetobacter sp.
Azospirillum sp.
Azospirillum sp.
Alphaproteobacteria
Euglena gracilis
Euglena mutalis
Eutriptiella gymnastica
Eutreptiella gymnastica
Euglenids
Neoparamoeba aestuarina
Paramoeba atlantica
Amoebozoa
Blastocystis hominis
Blastocystis hominis
Blastocystis sp. ATCC 50177 Nand II
Blastocystis sp. ATCC 50177 Nand II
Proteromonas lacertae
Opalinidae sp. Opal32
Opalina sp. OP11
Stramenopiles
A
B
0.2

Figure 2

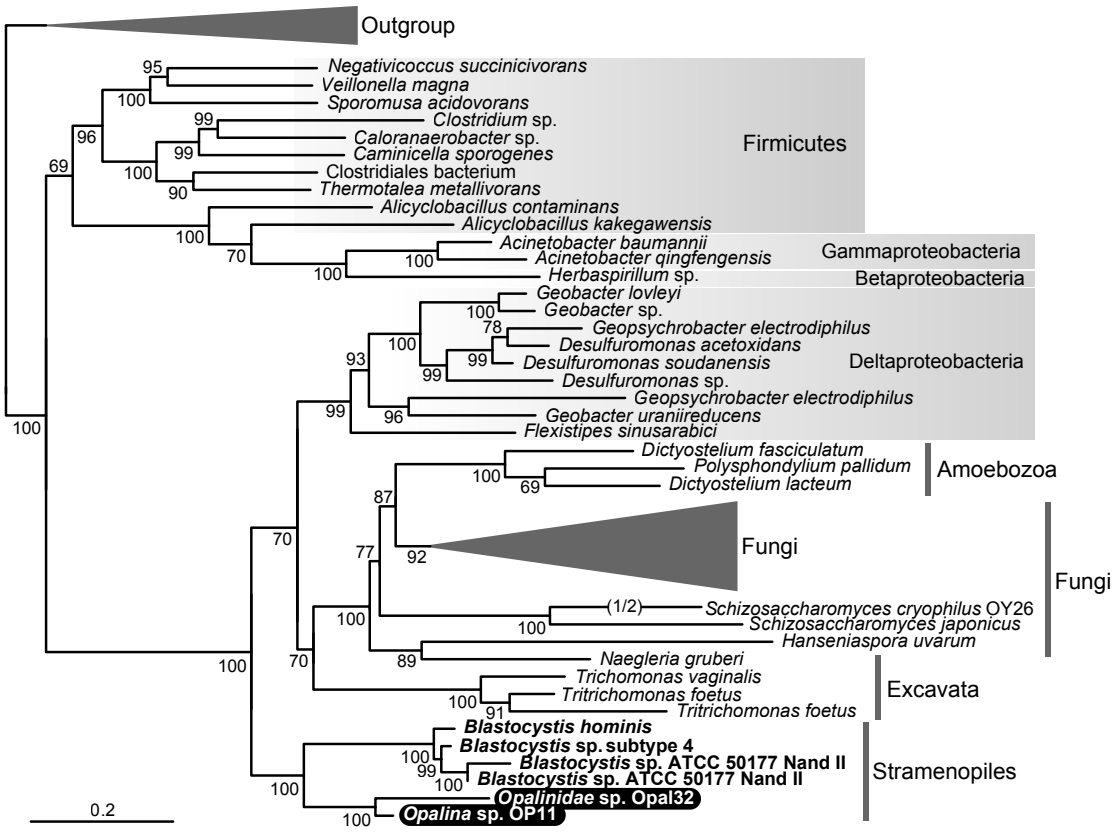

Outgroup
Negativicoccus succinicivorans
Veillonella magna
Sporomusa acidovorans
Clostridium sp.
Caloranaerobacter sp.
Caminicella sporogenes
Clostridiales bacterium
Thermotalea metallivorans
Alicyclobacillus contaminans
Alicyclobacillus kakegawensis
Firmicutes
Acinetobacter baumannii
Acinetobacter qingfengensis
Herbaspirillum sp.
Gammaproteobacteria
Betaproteobacteria
Geobacter lovleyi
Geobacter sp.
Geopsychrobacter electrodiphilus
Desulfuromonas acetoxidans
Desulfuromonas soudanensis
Desulfuromonas sp.
Deltaproteobacteria
Geopsychrobacter electrodiphilus
Geobacter uraniireducens
Flexistipes sinusarabici
Dictyostelium fasciculatum
Polysphondylium pallidum
Dictyostelium lacteum
Amoebozoa
Fungi
Fungi
(1/2)
Schizosaccharomyces cryophilus OY26
Schizosaccharomyces japonicus
Hanseniaspora uvarum
Naegleria gruberi
Trichomonas vaginalis
Tritrichomonas foetus
Tritrichomonas foetus
Excavata
Blastocystis hominis
Blastocystis sp. subtype 4
Blastocystis sp. ATCC 50177 Nand II
Blastocystis sp. ATCC 50177 Nand II
Opalinidae sp. Opal32
Opalina sp. OP11
Stramenopiles
0.2

Figure 3

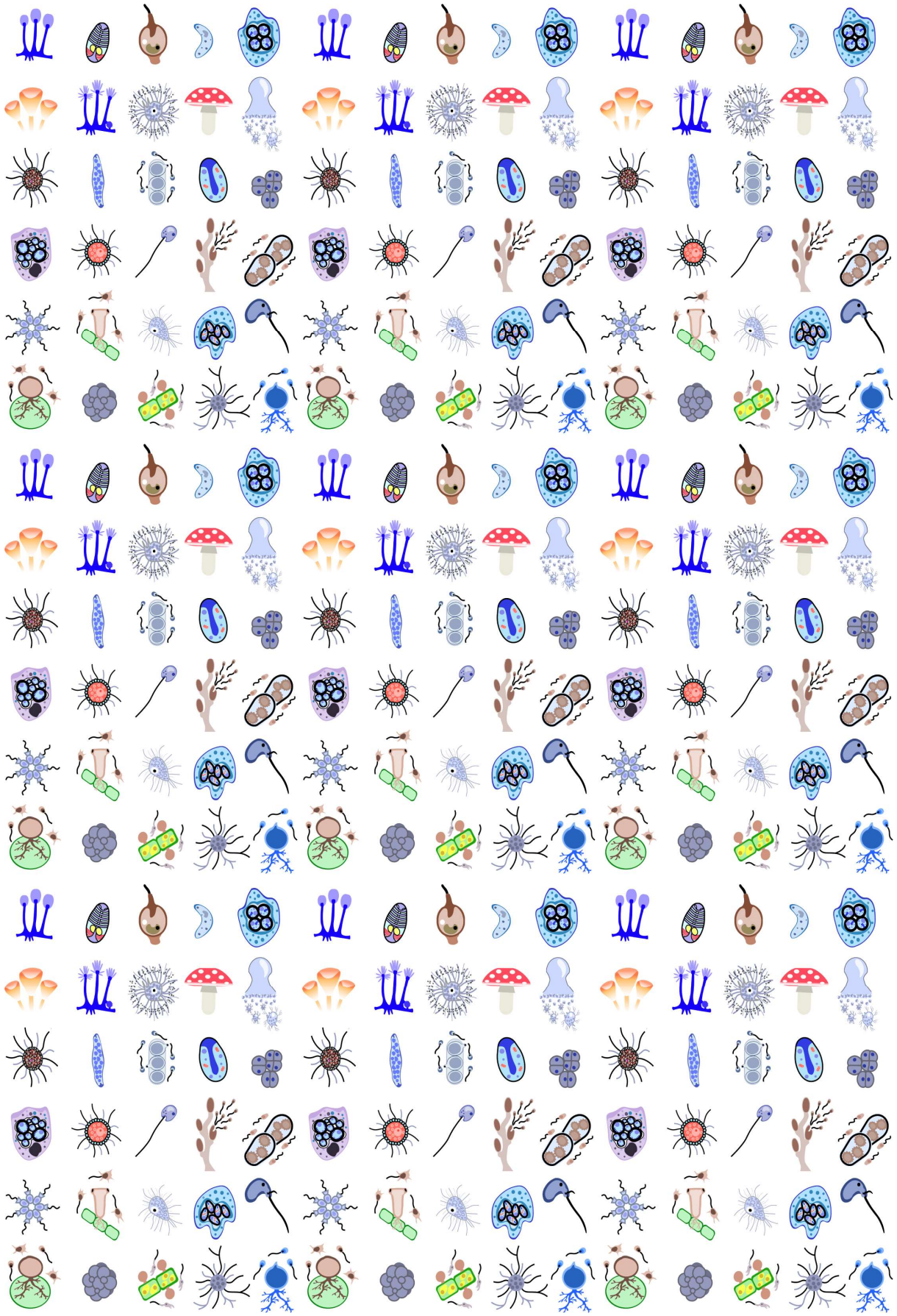
Annex 2: Congress communications

- **Oral communications**

1) **Galindo LJ.,** Torruella G., Moreira M., Karpov S., López-García P. Phylogenomics and comparative genomics of a new fungal clade helps reconstructing fungal evolution and suggests a flagellar reduction path. Comparative genomics of eukaryotic microbes: Genomes in flux, and flux between genomes. Sant Feliu de Guíxols, Spain. October 2019.
2) **Galindo LJ.,** Torruella G., Moreira M., Eglit Y., Simpson AGB., Völcker E., Clauß S., López-García P. Phylogenomic analysis of the Nucleariid amoebae, the earliest-diverging lineage of Holomycota. VIII ECOP-ISOP Joint Meeting. Rome, Italy. July 2019.
3) **Galindo LJ.,** Torruella G., Moreira M., Timpano H., Paskerova G, Smirnov A., Nassonova E., López-García P. Evolutionary genomics of Metchnikovella incurvata (Metchnikovellidae), an early branching microsporidian. International Society for Evolutionary Protistology (ISEP). Paphos, Cyprus. May 2018.

- **Posters communications**

1) **Galindo LJ.,** Torruella G., Moreira M., Eglit Y., Simpson AGB., Völcker E., Clauß S., López-García P. Don't judge a cell by its cover: A single-cell study of nucleariids. Royal Society Discussion Meetings - Single cell ecology/Developing new approaches for single cell manipulations and experimentation. London/ Chicheley Hall, UK. December 2018.
2) **Galindo LJ.,** Torruella G., Moreira M., Timpano H., Paskerova G, Smirnov A., Nassonova E., López-García P. Evolutionary genomics of Metchnikovella incurvata (Metchnikovellidae), an uncultured early branching microsporidian. International Society for Microbial Ecology 17 (ISME17). Leipzig, Germany. August 2018.
3) **Galindo LJ.,** Torruella G., Moreira M., Timpano H., Paskerova G, Smirnov A., Nassonova E., López-García P. Evolutionary genomics of Metchnikovella incurvata (Metchnikovellidae), an uncultured early branching microsporidian. International Congress of Protistology (ICOP). Prague, Czech Republic. July 2017.



université PARIS-SACLAY | ÉCOLE DOCTORALE Structure et dynamique des systèmes vivants (SDSV)

Titre : Phylogénomique des eucaryotes : la branche holomycota

Mots clés : cellule unique génomique, microorganismes, phylogénomique, Holomycota, Fungi

Résumé : La plupart de la diversité biologique est en réalité microbienne. L'arbre phylogénétique des eucaryotes comprend plusieurs grands supergroupes monophylétiques, dont les Opisthokonta. Ce groupe comprend deux branches, les Holozoa, qui inclut les animaux, et les Holomycota, qui regroupe les champignons et leurs parents unicellulaires. Bien que les champignons multicellulaires soient bien connus, nos connaissances sur la diversité des champignons unicellulaires et de leurs parents phylogénétiques restent limitées. Cette fraction unicellulaire comprend plusieurs lignées zoosporiques (par exemple chytrids) au sein des champignons, mais aussi une variété de lignées liées aux champignons classiques : les nucleariids, les rozellids, les aphelids et les microsporidies. Cependant, les relations phylogénétiques de ces lignées entre elles et avec les champignons restent à établir solidement. Les arbres phylogénétiques des gènes d'ARNr 18S environnementaux montrent une grande diversité d'Holomycota unicellulaires dans la plupart des écosystèmes terrestres. Cependant, le signal phylogénétique de ce gène est limité et ne permet pas de résoudre la plupart des relations phylogénétiques profondes. Au cours des dernières années, les techniques à haut débit ont permis de séquencer des centaines de nouveaux génomes et transcriptomes. Cela a permis de réaliser des études phylogénomiques multi-gènes, qui augmentent le signal disponible pour résoudre les relations évolutives. Néanmoins, la plupart de ces génomes correspondent à des espèces fongiques faciles à cultiver, souvent avec un intérêt particulier pour l'homme. Actuellement, les approches de type « omique » à partir des cellules uniques se révèlent comme potentiellement utiles pour étudier les eucaryotes unicellulaires non cultivés, en permettant de reconstruire des analyses phylogénétiques robustes d'une grande diversité environnementale à l'aide de données génomiques et transcriptomiques. Au cours de mon travail de doctorat, j'ai appliqué des approches de « cellule unique » pour obtenir des informations phylogénétiques à partir de lignées Holomycota divergentes, clarifier les relations phylogénétiques entre les champignons et ses proches parents et inférer l'évolution de leurs traits. Plus précisément, j'ai utilisé cette approche pour : 1) Générer des données génomiques et transcriptomiques pour les nucleariids et mieux reconstruire les relations internes dans le clade et les caractères présents dans leur ancêtre. Nos résultats confirment que les genres de protistes à thèque Pompholyxophrys et Lithocolla sont en effet des nucleariids et branchent avec Nuclearia, Parvularia et Fonticula. La reconstruction d'une phylogénie robuste de ce groupe nous a permis d'inférer les traits (par exemple pas de flagelle) ancestraux du groupe. 2)Séquencer et analyser de manière comparative le génome de Metchnikovella incurvata, pour confirmer sa position relativement basale dans Microsporidia et déterminer les synapomorphies du clade. L'analyse phylogénomique du metchnikovellid Metchnikovella incurvata a confirmé que des Metchnikovellidae branchent à la base des Core-Microsporidia. Nous avons également confirmé que leur profil métabolique était plus similaire à celui des Core-microsporidia, tous deux ayant réduit de manière similaire leurs gènes / fonctions. 3) Générer des données génomiques pour Amoeboradix gromovi et Sanchytrium tribonematis, qui forment le clade des sanchytrides, une nouvelle lignée de champignons zoosporiques identifiée récemment, et résoudre leur position phylogénétique. L'étude des deux génomes de sanchytrids a clarifié leur placement au sein des Fungi en tant que nouvelle groupe frère des Blastocladiomycota. Des analyses génomiques comparatives montrent que leur métabolisme est réduit par rapport aux lignées apparentées. En particulier, le système flagellaire est fortement réduit par rapport à d'autres Holomycota, avec 4 événements indépendants de perte de flagelle dans le clade.

Title: Deep eukaryotic phylogenenomics: the holomycota branch

Keywords: single cell genomics, microeukaryotes, phylogenomics, Holomycota, Fungi

Abstract: Despite the astonishing diversity of plants, animals and macroscopic fungi, most eukaryotic diversity is actually microbial. The eukaryotic tree comprises several large monophyletic supergroups. One of these groups is the Opisthokonta, which encompasses two branches, Holozoa, including animals, and Holomycota, grouping Fungi and their unicellular relatives. While multicellular fungi are well known, knowledge on the diversity of unicellular Fungi and their phylogenetic relatives is still poor. This unicellular fraction includes several zoosporic lineages (e.g. Chytridiomycota and Blastocladiomycota) within Fungi, but also a variety of lineages related to the classical core Fungi: nucleariids, rozellids, aphelids and Microsporidia. However, the phylogenetic relationships of these lineages among them and with classical Fungi remain to be solidly established. Molecular phylogenetic trees of 18S rRNA genes retrieved from environmental studies have showed a wide diversity of unicellular holomycotans in almost all environments on Earth. However, the phylogenetic signal of this gene is limited and does not allow robustly resolving most deep phylogenetic relationships. During past years, high-throughput techniques have allowed sequencing hundreds of new genomes and transcriptomes. This has made possible to carry out multi-gene phylogenomic studies, which increase the available signal to resolve evolutionary relationships. Nevertheless, most sequenced genomes correspond to easy-to-culture fungal species, often with particular interest for humans (e.g. parasites, plant symbionts, yeast). Recently, single-cell omics has become a potential useful approach to study uncultured unicellular eukaryotes, making it possible to reconstruct robust phylogenetic analyses of a wide environmental diversity using genomic and transcriptomic data. During my PhD work, I have applied single-cell techniques to get phylogenetic information from divergent holomycotan lineages, clarify phylogenetic relationships among fungi and their close relatives and infer trait evolution. More specifically, I have used this approach to: 1) Generate genomic and transcriptomic data for nucleariids and better reconstruct inner relationships in the clade and the characters present in the nucleariid ancestor. Our results confirm that the cover-bearing unicellular genera Pompholyxophrys and Lithocolla are indeed nucleariids and branch together with Nuclearia, Parvularia and Fonticula. The reconstruction of a robust phylogeny for the group allowed us to infer the traits (e.g. no flagellum, glycocalyx, no cover) already present in their ancestor. 2) Sequence and comparatively analyze the genome of Metchnikovella incurvata, to confirm its relatively basal position within Microsporidia, and determine synapomorphies for the clade. Phylogenomic analysis of the metchnikovellid Metchnikovella incurvata confirmed that Metchnikovellidae branch at the base of Core-Microsporidia. We also confirmed their metabolic profile to be more similar to Core-microsporidia, being both similarly reduced in genes/functions. 3) Generate genomic data for Amoeboradix gromovi and Sanchytrium tribonematis, which form the newly described zoosporic fungal clade of sanchytrids and resolve their phylogenetic position. The study of the two sanchytrid genomes clarified their placement within Fungi as a new clade sister to Blastocladiomycota. Comparative genomics showed that their metabolic composition was reduced in comparison with related lineages. This reduction was especially important in their flagellar toolkit when compared with other Holomycota, confirming 4 independent flagellum loss events in the clade.