Unveiling new molecular Opisthokonta diversity: A perspective from evolutionary genomics

David López Escardó

TESI DOCTORAL UPF / ANY 2017

DIRECTOR DE LA TESI

Dr. Iñaki Ruiz Trillo

DEPARTAMENT DE CIÈNCIES EXPERIMENTALS I DE LA SALUT

upf. Universitat Pompeu Fabra Barcelona

Acknowledgements:

Aquesta tesi va començar el dia que, mirant grups on poguer fer el projecte de màster, vaig trobar la web d'un grup de recerca que treballaven amb uns microorganismes, desconeguts per mi en aquell moment, per entendre l'origen dels animals. Tres o quatre assignatures de micro a la carrera, i cap d'elles tractava a fons amb protistes i menys els parents unicel·lulars dels animals. En fi, "Bitxos" raros, origen dels animals... em va semblar interessant. Així que em vaig presentar al despatx del Iñaki vestit amb el tratge de comercial de Tecnocasa, per veure si podia fer les pràctiques del Màster amb ells. El primer que vaig volguer remarcar durant l'entrevista és que no pretenia anar amb tratge a la feina, que no es pensés que jo de normal vaig tan seriós... Sigui com sigui, després de rumiar-s'ho, em va dir que endavant i em va emplaçar a fer una posterior entrevista amb els diferents membres del lab perquè tries un projecte de màster. Per tant, aquí el meu primer agraïment i molt gran, al Iñaki, per permetre'm, no només fer el màster, sinó també permetre'm fer aquesta tesis al seu laboratori. També per la llibertat alhora de triar projectes i per donar-li al MCG un ambient cordial on hi dóna gust treballar. Gràcies, doncs, per deixar-me entrar al món científic per la porta de la protistologia, la genòmica i l'evolució, que de ben segur m'acompanyaran sempre.

Dels possibles projectes de màster, el Javi va ser el més seductor (Xavi, les ubiquitin, d'entrada, no em van semblar atractives). High-throughput sequencing, FISH, cultivar protist, semblava l'hòstia. Llàstima, que Minorisa no va voler crèixer, no hi havia filtres de FISH disponibles, i només va quedar l'opció de Metamarks. Segon agraïment, doncs pel Javi, que malgrat els problemes inicials vam tirar endevant un projecte novedós, i que de ben segur publicarem aviat, per fi. Gràcies també pel mentoratge i la paciència intentant explicar conceptes d'ecologia microbiana i protistologia a un biotecnòleg / bioquímic, que la seva única relació prèvia amb protistes es concentrava només amb *Saccharomyces cerevisae*, i no era una relació exclusivament científica, precisament.

A banda de metamarks, un doctorand de la vella del guàrdia del lab que venia de Canadà, París o ves a saber d'on em va dir que hi havia un bitxo pel laboratori que no era el que semblava, que potser era un

llinatge nou. I ves per on la Nucelaria sp. es va convertir en la *Parvularia*, tot i els patiments inicials per fer una simple PCR del 18S. Guifré, moltes gràcies per aquest projecte, aparanment esborrejat, però que va anar a prenent forma fins convertir-se en la meva primera publicació. Tot un puntàs. A part d'això, gràcies pel teu bon-rollisme, m'ho he passat molt bé treballant amb tu i fent uns bons riures dins i fora de la feina. Esperem que continuïn.

Ja que hem citat a la vella guàrdia doncs, continuem, gràcies Arnau i Mendo, per omplir el laboratori d'un ambient intel·lectual i no només en l'àmbit ciéntific. Sovint pels éssers normalets com un jo, podia semblar abrumador, però he après un munt només d'estar compartint espai de feina amb vosaltres, moltes gràcies per haver-hi sigut.

Meri, Xavi i Helena, hem sigut més o menys coetanis al MCG lab, i això ens ha fet partíceps de moltes experiències compartides i de veure la transformació del lab. Xavi, gràcies per la paciència en l'ajuda i el suport als meus problemes bioinformàtics trivials. Meri, ja ho he oblidat, però al seu dia vas posar una mica d'ordre a la meva manera de fer poiata. Helena, sin ti, esto hubiera sido muy aburrido, además muchas gracias por la paciencia y ayuda con el Zeiss del lab i el confocal. Moltes gràcies als tres per co-transitar amb mi aquests últims 5 anys!

Als nous doctorands del MCG, Núria (tu no ets tant nova eh!!), Alicia, Edu, Alberto, Konstantina i Aleksandra?. Moltes gràcies per la vostra empenta, i les vostres ganes, sempre és gratificant veure savia fresca amb noves inquietuds. Només animar-vos a continuar igual i agraïr-vos de ser com sou.

Als post-docs i altres membres que han passat pel laboratori, gràcies pel bon ambient que sempre hi ha hagut el lab, pels consells i les sempre intressant converses de ciència o de qualsevol tema. Hiroshi, Núria (Ponsy), Matija, Sebas, Cristina, Michelle, Omaya, Andrej, Xavi F., Maries F. i R. Moltes gràcies.

Agraïr també a la gent de París que em va acollir durant 3 mesos, ja fa quasi 4 anys. No vaig trobar filastereans, però em vaig convertir en un master del clonning i vaig poder refrescar el meu francés

macarrònic. Puri, David, Aurelien i demés gent que corria pel lab Merci beacoup!

Finalment, amics i familia que heu estat suportant la meva absència en molts moments i que m'heu acompanyat a fer d'aquesta tesis, moltes gràcies per està sempre allà. Gent del carrasco, limadores, i gent de la UAB i demés gent que tenim contacte i complicitats moltes gràcies pel vostre suport i estima.

Menció especial per los dos persones que sense elles no estaria aquí el pare i la mare. Biòleg i matemàtica, que són els grans responables indirectes del meu interès per la natura, les ciències i del saber en general. Sobretot al mama per aguantar com una jabata un episodi greu de salut. Va ser dur compaginar la tesi i estar per tu, però més dur hagués sigut si la cosa hagués anat pitjor. Moltes gràcies als dos per tot el que m'heu donat. Evidentment també a l'àvia i als avis que ja no hi són que sempre estaran en mi.

I com no moltes gràcies Roger per haver compartit junts la vida, haver-me fet de germà gran i pel suport desde la distància londinenca en aquesta tesis.

I per acabar infinites gràcies a tu Júlia, per tot el suport durant aquesta tesi i especialment en aquests últims mesos. Gràcies, per portar-me a la muntanya a la mínima que tenim oportunitat, pel teu amor, companyia i comprensió. Tu sí que ets la millor! En definitiva gràcies per tot, i per ser la millor companya de vida que podria tenir.

Abstract

Opisthokonta is an eukaryotic supergroup that contains Metazoa, Fungi and their unicellular relatives. Therefore, this group provides an ideal framework to study distinct unicellular-multicellular transitions, among them, the transition towards animal multicellularity. This thesis aims to describe new Opisthokonta diversity at many different levels, a very neeeded starting point to better understand the evolution of opisthokonts and the origin of multicellularity in animals. In particular, we here described a new genus and species at the onset of Holomycota (*Parvularia atlantis*), detected new molecular metazoan diversity through a metabarcoding approach, and evaluated the power of single-cell genomics to increment the genomic diversity of choanoflagellates. Overall, this thesis provided new insights into the diversity of Opisthokonta and into the power of single-cell genomics technologies. Finally, our results also have reshaped the phylogeny of choanoflagellates and increased the knowledge of the pre-metazoan genetic tool-kit.

Resum

Els Opistoconts són el supergroup d'eucariotes que engloba els animals, els fongs i els seus respectius parents unicel·lulars. Ofereixen, per tant, un marc ideal on estudiar la transició d'organismes unicel·lulars a pluricel·lulars, entre elles, la transició a la pluricel·lularitat animal. Aquesta tesis pretén entendre millor la diversitat dels Opistoconts a diferents nivells, com a primer pas imprescindible, per poder entendre millor la seva evolució i obtenir noves dades que ajudin a compendre els processos que van precedir l'origen de la pluricel·lularitat animal. Concretament, els resultats obtinguts corresponen a: la descripció d'un nou génere i una nova espècie a la base dels holomycots (*Parvularia atlantis*), la detecció de nova diversitat animal gràcies a un estudi metagenètic (metabarcoding) i l'avaluació de l'ús de técniques de genòmica unicel·lular (Single-cell genomics) per poder expandir la diversitat genòmica dels coanoflagel·lats. Amb tot, els resultats donen una millor comprensió de la diversitat dels Opistoconts i del potencial de la genòmica unicel·lular. Finalment, els resultats han permés

també remodelar la filogènia dels coanoflagel·lats i incrementar el conexiement sobre el contingut genètic que va precedir l'aparició dels animals.

Preface

The interest for the living creatures that co-habites with us on earth dates back from the beginning of our species. In the first human representations, 30,000-40,000 years ago, hunter-gathered humans painted different animals in caves like Altamira (Cantabria, Spain) or Chauvet (Auvergne-Rhône-Alpes, France). The reason for this early fascination, and the fact that we still are curious on other living beings is quite simple. Our survival and needs relay on other living organisms. Animals, plants and Fungi have provided us along human history, and still provide, food, medicines, clothes, even services like transport, company or protection. Therefore, the curiosity and interest for other living beings it is inherent to our species. This curiosity had been translated in different scientific disciplines that pretend to unveil the mysteries and questions regarding living beings. These are the motivations behind this work, in which I pretend to increase the knowledge of Opisthokonta diversity at a species and at a molecular level, providing new insights into Opisthokonta evolution and the origins of animals. Opisthokonts are the eukaryotic supergroup that contains Animal, Fungi and their unicellular relatives, which can be divided in two main branches: Holomycota and Holozoa (Adl et al. 2012).

Holomycota (Liu et al. 2009) or Nucletmycea (Matthew W. Brown, Spiegel, and Silberman 2009) contains Fungi (Spatafora et al. 2016), the parasitic Opisthosporidia (S. A. Karpov et al. 2014), and the filopodiated nucleariid amoeba (Tom Cavalier-Smith 1993). On the other hand, Holozoa (Lang et al. 2002) is composed by animals and their closest unicellular relatives: choanoflagellates, filastereans and teretosporeans. The choanoflagellates are flagellated organisms surrounded by a collar of microvilli that are able to form colonies (Leadbeater 2015). The Filasterea (Shalchian-Tabrizi et al. 2008), until recently, only composed by two described species (*Capsaspora owczarzaki* and *Ministeria vibrans*), are filose amoeba, one of them (*C. owczazarki*) with aggregative capabilities (Sebé-Pedrós, Irimia, et al. 2013). Finally, Teretosporea (Torruella et al. 2015), includes taxa with coenocytic development and a palintomic division (Glockling, Marshall, and Gleason 2013). Therefore, within opisthokonts there is a wide range of organisms, with different life cycles and different forms of multicellularity,

providing an ideal framework to study distinct unicellular-multicellular transitions.

In this regard, the origins of animal multicellularity has been suggested to be particularly unique in the eukaryotic world (Thomas Cavalier-Smith 2017). Cavalier-Smith argues that mechanistically, is much harder for a single-celled organism whose feeding mode is based in intracellular phagocytosis, neither osmotrophy or photosynthesis, to move towards eating through a mouth and a gut. This requires different cells playing different roles for the common entity that demands cooperation at a higher organizational level (Thomas Cavalier-Smith 2017). To regulate this and other functions, animals have developed a wide range of different cell types – Bilateria can have >50 different cell types (Carroll 2001) –. In addition, Metazoa is the eukaryotic kingdom with more species described (del Campo et al. 2014). This implies that animal multicellularity would have been very difficult to achieve but, once accomplished, turned out to be very successful. That makes the transition towards animal multicellularity one of the most interesting questions for evolutionary biologists.

The first clues about the origin of animals started with Henry James-Clark observations 150 years ago. He found a structural similarity between choanoflagellates and the choanocytes of sponges (James-Clark 1866a), pointing out a potential link between sponges and choanoflagellates. This discovery inspired the Choanoblastea theory, in which animals would had appeared from choanoflagellates colonies and, after successive generations, became more complex and developed different cell types (the choanoblastea). The choanoblastea then increased its complexity until the formation of the first sponge, before the Cambrian explosion and the diversification of animals into their aprox. 35 different phyla (Nielsen 2008).

Later on, this association between choanoflagellates and sponges received support from molecular phylogenies that showed that choanoflagellates are the sister-group to Metazoa (Thomas Cavalier-Smith and Chao 2003; Medina et al. 2003; Ruiz-Trillo et al. 2004; Ruiz-Trillo et al. 2006; Steenkamp, Wright, and Baldauf 2006; Ruiz-Trillo et al. 2008). More recently, the genomes of *Monosiga brevicollis* (Nicole King et al. 2008) and *Salpingoeca rosetta* (Fairclough et al. 2013) revealed that genes related in

multicellular functions were encoded on those genomes. This implied that some genes involved in multicellularity, such as tyrosine kinases and cadherins, were already present in the shared ancestor of Metazoa and Choanoflagellatea.

The finding of new unicellular lineages and the sequencing of their genomes showed that the unicellular ancestor of animals was even more complex (Sebé-Pedrós et al. 2010; Sebé-Pedrós et al. 2011; Hiroshi Suga et al. 2013; A. De Mendoza, Sebé-Pedrós, and Ruiz-Trillo 2014; A. de Mendoza et al. 2013) and that some genes had been secondarily lost in the sequenced choanoflagellates (Hiroshi Suga et al. 2013). Overall, all these examples show that the discovery of new Opisthokonta diversity, together with their genomic data, is crucial to better understand the transition towards animal multicellularity from their single-celled ancestors.

In this work I pursued to unravel new Opisthokonta diversity at many different levels: 1) by describing a new unicellular Opisthokonta lineage and species within Nucleariids, *Parvularia atlantis* (section 3.1); 2) taking advantage of metabarcoding data to unravel new metazoan molecular diversity (section 3.2); 3) evaluating the use and potential of single-cell genomics (SCG) technologies (section 3.3); and 4) by expanding the genomic diversity of choanoflagellates using single-cell genomics (section 3.4). All these new Opisthokonta diversity (at molecular and at species level) offers new perspectives on Opisthokonta evolution, and also, concretely, on the transition towards animal multicellularity.

Table of contents

1. INTRODUCTION

1.1. The eukaryotic tree of life: A Historical framework

The first documented human initiatives to classify and understand living beings date from the beginning of civilizations. There are available ancient books and documents that reflect the primary ideas and observations regarding biological diversity of human beings. Early hunter-gathered or Neolithic populations had, surely an enormous knowledge of the plants, and animals that were co-habiting with them. However, without scripture, those early human populations could not shape their knowledge to far future generations (Magner 2002).

The first attempts to classify and understand living beings come from ancient Indian, China and Egyptian civilizations which already possessed practical information for medicine and plants. However, the first propositions of more systematic systems to classify the living beings were proposed during the ancient Greek and Roman civilizations (Magner 2002). For example, Aristotle's Scala natura, a.k.a the Great Chain of Being, was an hierarchichal classification of organisms from inferior/simple to superior/complex. This classification, that was rather arbitrary, continued during medieval ages until 18th century, when Linné proposed the Systema Naturae, the biological classification based on taxonomy and systematics, and precursor of modern classification of living beings.

Linné concepts, the advances in anatomy comparison during the 19th century, fossil record evidences and Lamarck's theory shaped the intellectual framework in which Darwin and Wallace proposed their theory: the evolution by natural selection as the mechanism for species origin from a common ancestry.

Thus, Darwin ideas changed the concept of homology, used for functional similarity between taxa, to be interpreted with a signal of common ancestry. Since that moment, organisms were not longer classified among arbitrary criteria, organisms were grouped according to their evolutionary history or phylogeny. Thus, since then, tree diagrams were populated to represent the evolution from the common ancestor to the extant diversity. For example, six years after the publication of the Origin of Species, Ernst Haeckel

proposed a tree of organization of living beings that was based in three maing branches: Plants Animals and Protista (Fig.1a).

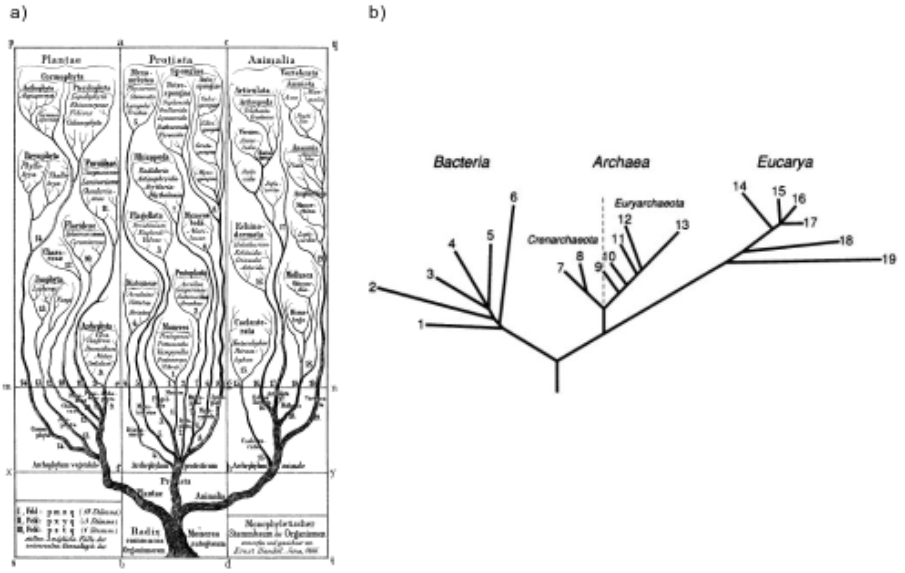


Figure 1. Early tree of life representations. a) The tree of life proposed by Ernst Haeckel published in the General Morphology of Organisms (1866). b) Phylogenetic inference based on the SSU of ribosomal gene (Carl R Woese, Kandlert, and Wheelis 1990).

In this regard, the discovery in the 1950's of the DNA molecule by Watson and Crick as a the carrier of genetic information, provided a molecular target to address the evolutionary relationships among living beings. DNA sequencing technologies allowed already in the 1970's to perform the first phylogenetic approach based on the sequences of the ribosomal small subunit gene (SSU rDNA) (C. R. Woese and Fox 1977) with eukaryotic and prokaryotic members. In this work, Woese and Fox showed by the first time that prokaryotes were splitted in two distinct groups: Bacteria and Arceha, which together with eukaryotes, compose the three domains of life (Fig. 1b).

Molecular phylogenies provided significant advantages in comparison to morphology-based phylogenies, specially in microbial organisms, in which there are less morphological characters to identify and are more prone to convergent evolution

processes (Baldauf 2003; Parfrey et al. 2006). DNA or protein sequences allow to have more characters (nucleotidic or amino acidic positions) to compare objectively in terms of orthology, among properly aligned sequences. Thus, molecular phylogenies provided a powerful tool to address evolutionary questions.

Regarding eukaryotic organisms, molecular phylogenies of SSU rDNA, revealed that protistan organisms are paraphyletic and distinctly related with multicellular lineages (Tom Cavalier-Smith 1993; Parfrey et al. 2006). Therefore, the classical classification of eukaryotes based in Animals, Plants, Fungi and Protista had no longer meaning (Parfrey et al. 2006). As a result of this, several eukaryotic supergroups were established: Opisthokonta, Amoebozoa, Excavata, Archaeplastida, Alveolata, Rhizaria and Stramenopila, being multicellular eukaryotes, an small fraction of eukaryotic diversity (Finet et al. 2010; F. Burki et al. 2012; M. W. Brown et al. 2013; Derelle et al. 2016; Torruella et al. 2015; Fabien Burki et al. 2016) (Fig. 2).

Interestingly, these more recent phylogenies were not based in the 18S ribosomomal gene, because SSU rDNA alone had not enough resolution to resolve some phylogenetic questions, like deep eukaryotic relationships among supergroups or problems with long-branch taxa (Baldauf 2000). For instance, in first of eukaryotic phylogenies, microsporidians –single-celled animal parasites that do not have mitochondria– appeared as the earliest branching eukaryotic lineage. At that moment such position of microsporidians was not strange. Scientist thought that there might be an ancestral amitochondrial eukaryotic group that diverged previous to the acquisition of the mitochondria –Archeozoa theory– (Cavalier-Smith 1983; Tom Cavalier-Smith 1993). However this phylogenetic placement of the microsporidians was a long-branch attraction artifact (Baldauf 2000; Baldauf 2003). The use of more genes (multi genes approaches), or the whole genome sequences thanks to high-throughput sequencing technologies, allowed to increase the resolution and provided a more accurate view of the eukaryotic diversity. Thus, and thanks to phylogenomic analyses (Finet et al. 2010; F. Burki et al. 2012; M. W. Brown et al. 2013; Derelle et al. 2016; Torruella et al. 2015; Fabien Burki et al. 2016) (Fig. 2), mircrosporidians were shown to belong to Opisthokonts, as close relatives of Fungi. Cell biology and molecular analyses also showed that they had a mitochondria-derived organelle as and their

lost their mitochondria as a consequence of adaptation to parasitism (Keeling 2009).

However, there are still some uncertainties regarding the eukaryotic tree of life. For instance, there are early-branching taxa like *Malawimonas* or *Collodyction*, that represent an early divergence from the rest of eukaryotic groups. There is also a lack of other close related species that would help to solve their position in the Eukaryotic Tree of Life (EToL) and also in the establishment of the root of the EToL (Derelle, Torruella, and Klime 2015). Therefore, to fully solve the EToL we need to have the widest possible genomic taxon sampling, also accompanied with improved phylogenetic datasets algorithms and models (Fabien Burki et al. 2014).

1.2. Expanding the known eukayotic diversity: molecular-based approaches

a) Environmental sequencing of 18S ribosomal gene

Most eukaryotic lineages correspond to single-celled organisms (Fig. 2). Thus, if we want to fully understand the eukaryotic diversity and its phylogenetic relationship among different lineages, it is crucial to know and understand such microbial eukaryotic diversity. However, exploring microbial diversity is not easy. Due to their small sizes, these organisms are difficult to identify and isolate. In addition, it also very laborious culturing them, as well as study them at morphological and molecular level, specially pico-nano- planktonic (orgasims smaller than 20μm) heterotrophic organisms (Heywood et al. 2011).

Therefore, to circumvent those issues, scientists started to use molecular techniques, a couple of decades ago, to unveil eukaryotic microbial diversity, in the same way that previous studies had targeted prokaryotic diversity (Carl R. Woese 1996). The method consisted in amplifying, from environmental DNA, the 18S ribosomal gene by PCR, using eukaryotic universal primers (Medlin et al. 1988). The PCR product is cloned, and the clone library represents the molecular eukaryotic diversity found in the environmental sample.

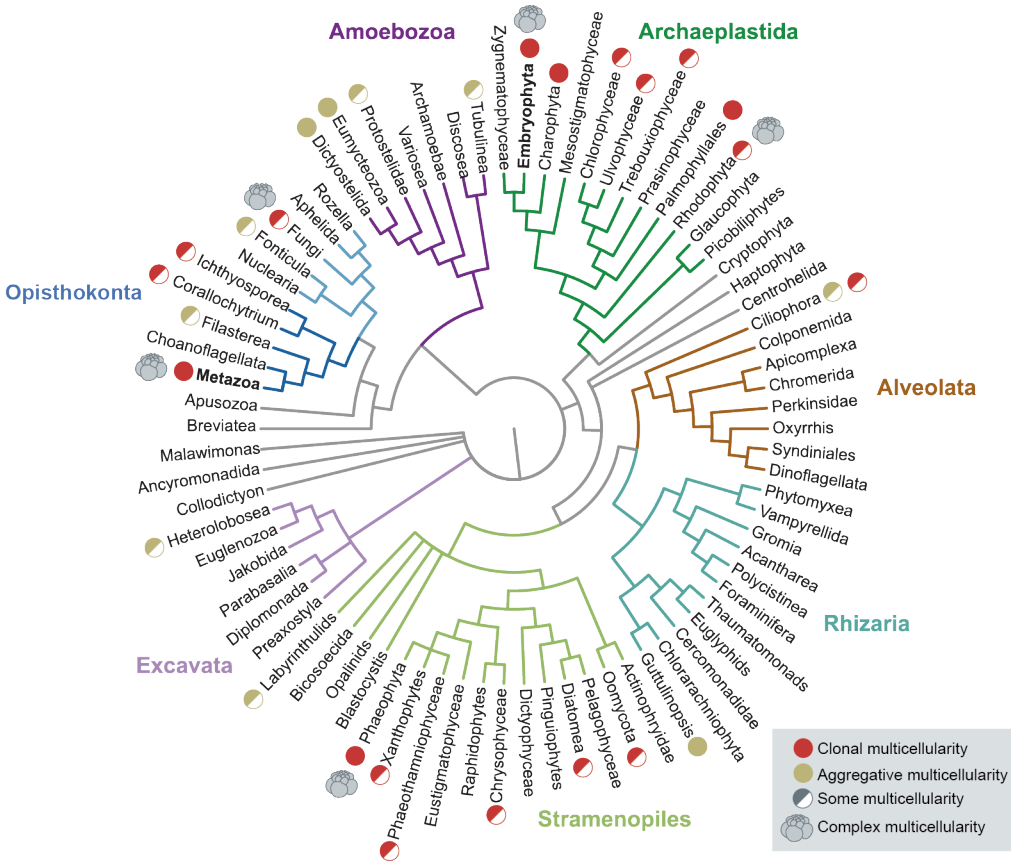


Figure 2. Overview of the eukaryotic tree of life, highlighting the multiple origins of multicellularity and their modalities: aggregative, clonal and complex based in recent phylogenomic analysis (Finet et al. 2010; F. Burki et al. 2012; M. W. Brown et al. 2013; Derelle et al. 2016; Torruella et al. 2015; Fabien Burki et al. 2016). Adapted from (Grau-Bové 2017) (Sebé-Pedrós et al., 2017).

This technique allowed to unveil the eukaryotic diversity present in the environments and the first studies using this approach revealed an unsuspected hidden eukaryotic diversity (Dìez, Pedrós-alió, and Massana 2001; López-García et al. 2001; Moon-van der Staay, de Wachter, and Vaulot 2001; Moreira and López-García 2002).

Thus, this molecular data started to provide a lot of sequences that were phylogenetically related between them, but not related with any molecular data in which there are cultures available. Therefore, scientist started to describe environmental groups, such as MAST, a

novel group of marine Stramenopiles (Massana et al. 2004). The 18S sequences of these new groups allowed also to design oligonucleotide probes to be applied by fluorescence in situ hybridization (FISH) to assess the distribution and abundance of these new eukaryotic groups and also obtain hints regarding their size and morphology (Massana et al. 2006). These data already showed that some environmental lineages like MAST, were very abundant in the marine environment and could play important roles in marine ecosystems (Zhu et al. 2005). Furthermore, high-throughput sequencing technologies allowed to move from tens of sequences in cloning libraries until million of reads in metabarcoding analysis, increasing the sensitivity of these molecular approaches (Massana et al. 2011; Logares et al. 2012; Logares et al. 2014; del Campo et al. 2015; Richards et al. 2015; de Vargas et al. 2015). Thus, we have now a powerful tool to address the distribution and the diversity of many eukaryotic taxa, including the one that is yet not culturable.

However, these new approaches only allow the researchers to get insights on ecological and diversity patterns of eukaryotic taxa, not to improve the eukaryotic tree of life or/and reconstruct gene families evolution. For those questions, you need full genomic sequences. Precisely, non-cultured single-celled eukaryotes are the most underepresented fraction of eukaryotic genomic diversity (del Campo et al. 2014). In this regard, single-cell genomics technologies might help to circumvent this limitation.

b) Single-cell genomics

Single-cell genomics (SCG) appeared to be a promising technique to get the genomes of uncultured taxa directly from the environment (Stepanauskas and Sieracki 2007; Stepanauskas 2012). In contrast to metagenomics data, SCG allows to recover genomic DNA from one single cell. Single cells from the environment can be isolated using different techniques such as micromanipulation (Woyke et al. 2009), microfluidics (Ciuffi, Rato, and Telenti 2016) and by using a Fluorescence Activated Cell Sorting (FACS) (Stepanauskas 2012). Cell isolation is then followed by cell lysis and a whole genome amplification (WGA) step (Stepanauskas 2012). The discovery of chemolithoautotrophy pathways in uncultured Proteobacteria (Swan et al. 2011) , or the proposition of two new prokaryotic superphyla

(Rinke et al. 2013) , are two examples of promising findings obtained thanks to single-cell genomics in prokaryotes.

However, single-cell genomics have also some important drawbacks that challenge their use in all microbial forms, including eukaryotes. For example, the sample can suffer an amplification bias at the WGA, as well as the appearance of artefacts or genome loss (de Bourcy et al. 2014; Gawad, Koh, and Quake 2016). Multiple Displacement Amplification (MDA) (Dean et al. 2002; Jiang et al. 2005), which uses a high fidelity phi29-polymerase (J. A. Esteban, Salas, and Blanco 1993), is the standard method used for microbial applications (Rinke et al. 2014). However, there is another WGA method more recent than MDA. It is called multiple annealing and looping-based amplification cycles (MALBAC). This method, based in an MDA and a PCR-based phase, has been tested in SW480 cancer human cells and it presents more uniformity and better detection of different copy number variants (CNVs) than MDA (Zong et al. 2012). The chemistry of the reaction provokes a loop formation of the more amplified genomic areas to force reaction to amplify other genomic regions (Zong et al. 2012). However, for low amounts of starting DNA, such as bacteria, this method seems to present a greater amplification error rate and greater contamination reads (de Bourcy et al. 2014; Gawad, Koh, and Quake 2016). Therefore, MDA is still the standard method used for microbial applications (Rinke et al. 2014), which provides between 5–100% of genome completeness in bacteria, with an average of around 40% (Rinke et al. 2013).

Fewer studies have been done in unicellular eukaryotes and some of them appeared during the course of this thesis. One example is the recent work done in the parasite *Cryptosporidium*. The authors recovered almost the full genome sequence (Troell et al. 2016). However, this case might be not representative of the works done with SCG in eukaryotes, given that *Cryptosporidium* can be purified from fecal samples and has a rather small genome with low GC content compared to most eukaryotes, factors that favour a more uniform WGA reaction (Stepanauskas et al. 2017) SCG works done in environmental samples targeting different organisms like picobiliphytes (Yoon et al. 2011), *Paulinella* (Bhattacharya et al. 2012) or MAST (Roy et al. 2014; Mangot et al. 2017), reveal a different story. The genome recovery on those studies varied widely and was not high (between 9–55%). Interestingly, one of the studies

focused in MAST (Mangot et al., 2017), showed that by co-assembling different SAGs from different cells the genome recovered increased substantially (Mangot et al., 2017). Thus, it remains yet unclear the full potential of this methodology, however it should be taken into account in any attempt to expand the eukaryotic genomic information from organisms that currently are not culturable.

1.3. Opisthokonta diversity

Opisthokonta is an eukaryotic supergroup (Fig. 2) that comprises, besides protistan taxa, animals and fungi, that is two (out of five) of the complex multicellular lineages (Andrew H. Knoll 2011). Opisthokonts are divided in two branches: Holozoa and Holomycota (Fig. 3). Opisthokonts could have appeared between 1,579-904 Mya (Eme et al. 2014) and the first evidences of animals date from 635 Mya (Love et al. 2008) (Fig. 4), on the other hand multicellular fungi might have appeared later round 500Mya (Berbee and Taylor 2010) (Fig. 4).

Besides fungi and animals, there is also a great diversity of morphologies and lifestyles within single-celled opisthokonts. Organisms can be free-living flagellated phagotrophs, filopodiated amoebas, cell-walled osmotrophic parasites or saprotrophs, even eukaryotic predators. Opisthokonta clade was initially proposed by Thomas Cavalier-Smith in a symposium of the British Mycological Society in 1986 (published in the book Evolutionary Biology of the Fungi Ed. Cambridge University Press, 1987). There, TCS proposed that Fungi and Metazoa share a more recent common ancestor than plants. Choanoflagellatea (see section 1.3e) was also included within the Opisthokonta clade, whose major morphological synapomorphy was the single emerging flagellum that is located at the posterior end of the cell. This classification proposal was soon confirmed by SSU rDNA phylogenetic analyses, and later on corroborated by protein-coding gene trees (Baldauf 2000; Lang et al. 2002).

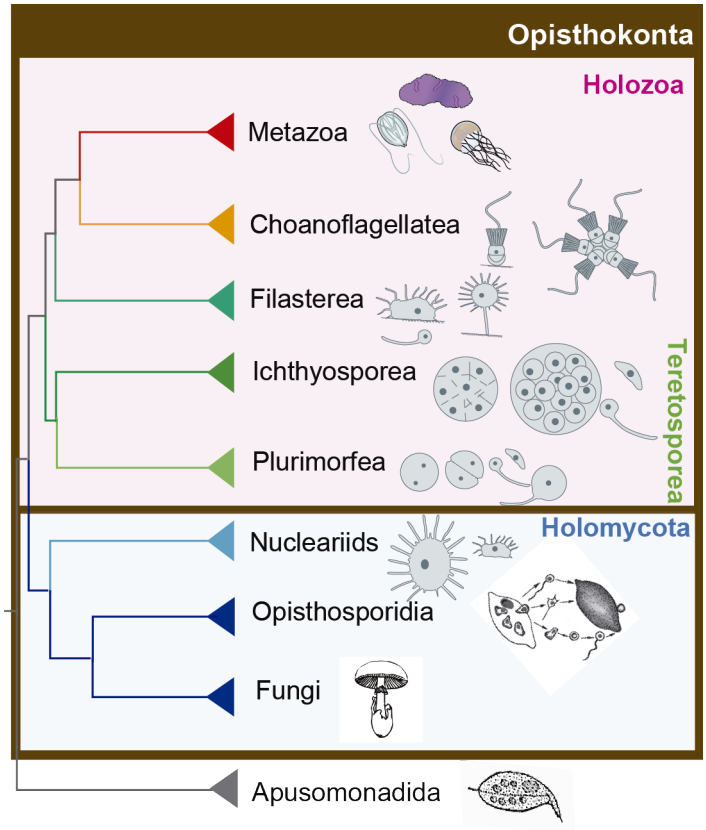


Figure 3. Schematic representation of the phylogenetic relationships among Opishtokonta diversity based in (Karpov, et el. 2014; Torruella et al., 2015; Spatofora 2016).

The list of Opisthokonta protists have been progressively increasing during the past three decades. One of the new opisthokont lineages is the Ichthyosporea (T. Cavalier-Smith 1998) –a.k.a. Mesomycetozoa (Herr et al. 1999)–, which was confidently placed as sister group to Choanoflagellata and Metazoa, forming the Holozoa clade (Lang et al. 2002; Ruiz-Trillo et al. 2004; Steenkamp, Wright, and Baldauf 2006). Another example is the enigmatic free living fungi-like *Corallochytrium limacisporum*, a species originally classified as a thraustochytrid (Raghu-Kama, Chandramohan, and Ramaiah 1987), but molecular phylogenies placed it within the Opisthokonta (Thomas Cavalier-Smith and Allsopp 1996). Currently, *Corallochytrium* cluster with *Syssomonas multiformis* a newly described unicellular holozoan, which together

with *Corallochytrium* comprise the lineage Plurimorfea (Hehenberger et al. 2017). Recent phylogenomic analysis suggest that plurimorfeans and ichthyosporeans might form a monophyletic group called Teretosporea (Torruella et al. 2015; Grau-Bové et al. 2017), although another suggest that they might be splitted in different holozoan lineages (Hehenberger et al. 2017).

Finally, among holozoans, there is another lineage, Filastera, which is sister to Choanozoa[1] (animals plus choanoflagellates) and was comprised, at the beginning, only by two filose amoeba species: *Ministeria vibrans* and *Capsaspora owczarzaki* (Shalchian-Tabrizi et al. 2008). Now, there are two additional filasterean species described *Prigoraptor chileana* and *Pigoraptor vietnamica* (Hehenberger et al. 2017) (see section 1.3d).

On the other side of the Opisthokonta tree, there are filose amoebae called Nuclearia (Patterson 1984), which were positioned as sister group to Fungi (Medina et al. 2003). Curiously, the filasterean *C. owczarzaki* was initially described as a Nulcearia (Zettler LAA et al. 2001) until it was properly placed as an unicellular holozoan relative of animals (Hertel, Loker, and Bayne 2002; Ruiz-Trillo et al. 2004). Together with Nuclearia genus, *Fonticula alba* was positioned as sister to Nuclearia at the root of Holomycota, forming the nucleariids amoeba clade (Matthew W. Brown, Spiegel, and Silberman 2009). Recently, the Aphelida (S. a Karpov et al. 2013) and the Rozella (or Cryptomycota) (James and Berbee 2012) groups were proposed to cluster with Microsporidia in a clade called Opisthosporidia (S. A. Karpov et al. 2014) sister to Fungi (Fig. 3) (see section 1.3a). In the next section, I will proceed to further explain the different Opisthokonta lineages.

1 Choanozoa was firts defined by Tom Cavalier-Smith as the unicellular lineages closely related to animals and fungi (Tom Cavalier-Smith 1987). However, I rather used the nomenclature proposed by Thibaut and King (Brunet and King 2017), in which Choanozoa is the group comprised by animals and choanoflagellates

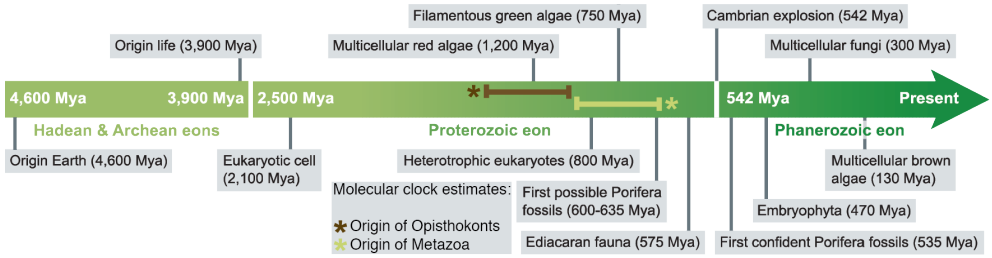


Figure 4. Time-line of origin of major multicellular eukaryotic lineages, adapted from (Grau-Bové 2017).

a) Fungi and Opisthosporidia

Fungi is the third richest of traditional eukaryotic kingdoms, with round of 330,000 species described (del Campo et al. 2014) and presents a wide range of morphologies, from unicellular to multicellular organism, with complex life-cycles often accompanied with a sexual reproduction (Lee et al. 2010). Fungi traditional kingdom can be divided between: chytrids, Mucoromycota, Zoopagomycota and Dikarya (Spatafora et al. 2016). Dykaria is the lineage that include the complex multicellular forms that have appeared independently within Basidiomycota and Ascomycota (Andrew H. Knoll 2011). However within these groups there are examples of regression to unicellular forms like the ascomycotan *Saccharmyces cerevisiae* (Nguyen et al. 2017). In addition, some Dykaria fungi are still poorly studied and they are exclusively represented by environmental surveys (Richards et al. 2012). All the Fungi, including the early-branching are osmotrophic and present a cell-wall. However there are not clear synapomorphies that define the clade Fungi, since those characters are also found in other eukaryotic lineages (Richards, Leonard, and Wideman 2017). Fungi, perhaps, can instead be defined through its losses rather than unique acquisitions. For instance all fungi lineages have lost phagotrophy and most of them also, the posterior flagellum, typical from Opisthokonts (Richards, Leonard, and Wideman 2017). However, the early-branching fungi, the chytrids, comprise species with flagellated dispersal forms like *Allomyces macrogynus* (Fig. 5b). There are two lineages included within chytrids:

blastocladiomycetes and chytridiomycta. It is not clear yet, however, if chytryds are monopyhyletic or paraphyletic. (Fig. 5).

Similarly, Opisthosporidia, the sister clade of Fungi, can be paraphyletic as well, because its monophyly was proposed based solely on ribosomal trees, which comprises Cryptomycota (a.k.a. Rozellida or Rozellomycota) (James and Berbee 2012) Aphelida (S. A. Karpov et al. 2014) and the Microsporidia (Vávra and Lukeš 2013) (Fig. 5). They are all intracellular parasites with a naked amoeboid vegetative stage, a cystic stage and a specialized apparatus for penetration into host cell. Aphelida is represented by at least three cultured species, all parasites of marine or fresh-water algae, as well as some environmental sequences (S. A. Karpov et al. 2014).

Cryptomycota (Jones et al. 2011) –a.k.a. Rozellida (Lara, Moreira, and López-García 2010) or Rozellomycota (Corsaro et al. 2014) – is a large group that includes many diversified environmental sequences (Jones, Forn, et al. 2011), among them the clade LKM11, the first environmental sequences to be proposed as sister group to the filamentous fungi (Lara et al. 2010). There is only one characterized genus and one culturable species (*Rozella allomycis*), whose genome has been recently sequenced (James and Berbee 2012). Cryptomycota share many similarities with aphelids, such as the intracellular amoeba or the dispersal flagellated cell types, as well as the parasitic life cycle. The main difference is the host range, fungi-like organisms in the case of Rozella, and its lack of amoeboid dispersal stage (S. A. Karpov et al. 2014)

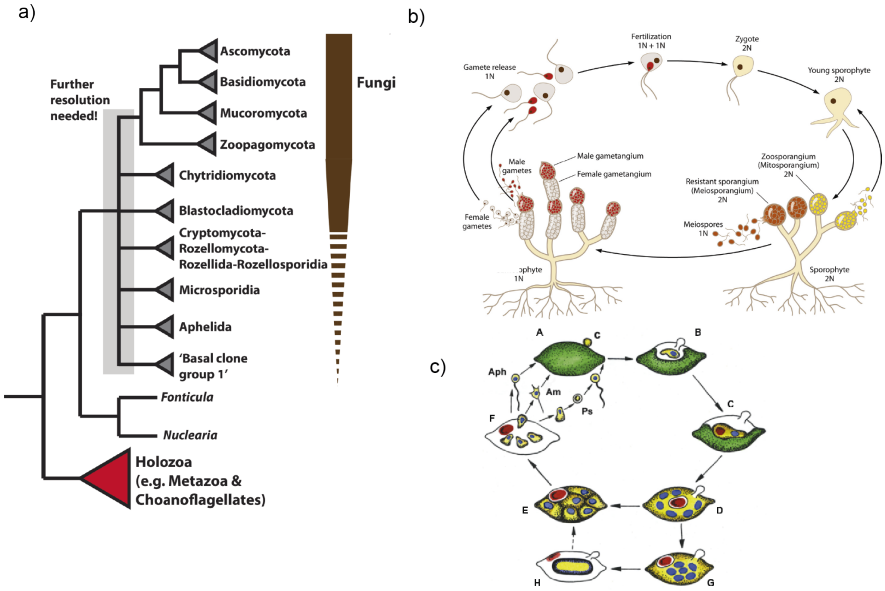


Figure 5. Fungi and Opisthosporidia. a) Schematic phylogenetic tree illustrating additional groups branching proximate to the origin of the fungal clade and the phylogenetic uncertainty among the deep branches of the Fungi and associated groups. Adapted from (Richards et al. 2017). b) Life cycle of *Allomyces macrogynus* from (Lee et al. 2010). c) Life cycle of aphelids from (Karpov et al. 2014).

Finally, Microsporidia represents an extremely derived group of intracellular parasites. They principally parasite animals, for which they seem to have an strong phylogenetic association (Smith 2009). They have many simplified characters such as reduced genomes, ribosomes, and endomembrane system. They even lack dictyosomes, and organelles like peroxisomes and canonical mitochondria. Such characteristics, together with their special structure polar tube they use to penetrate the host cytoplasm, made these organisms a clear monophyletic group, but with no easy classification among other eukaryotes, as explained before (see section 1.1) (Keeling 2009).

b) Nucleariids

The nucleariid amoebae comprise only two genera confirmed by molecular phylogeny Nucleria Cienkowski, 1865 (Patterson 1984) and Fonticula. Nucleria comprises species with typical spherical

filopodiated cells that feed on filamentous cyanobacteria in freshwater environments (Fig. 6a). Half a dozen Nucleariа species have been properly described, and are relatively easy to find and grow in culture conditions (Yoshida, Nakayama, and Inouye 2009). Recently, several additional Nucleariа strains have been isolated and thoroughly studied both from morphological and molecular phylogeny standpoints. These strains harbor endo- and/or ectosymbionts, which might facilitate the feeding of Nucleariа species (Dirren and Posch 2016; Dirren et al. 2017). Nucleariа can contain an extracellular matrix which might be the a reservoir to accommodate bacteria ectosymbionts (Dirren and Posch 2016; Dirren et al. 2017). Furthermore, cell sizes are often bigger than other unicellular Opishtokonta lineages (from 10-25 μm up to 60thμm). Thus, Nucleariа species are all free-living phagotrophes, and they share some common morphological traits: spherical or flattened protoplasm with radiating thin hyaline filopodia, usually with a central prominent nucleus and presence of a contractile vacuole (Mikrjukov and Mylnikov 2001). However, each species have some particularities, some are multinucleated –such as *N. delicatula* (Blanc-Brude, Skreb, and Dragesco 1955)–, some have branching filopodia –like *N. moebiusi* (Patterson 1983)–, and some have cystic stages –*N. simplex* (Patterson 1984)–.

On the other hand, *Fonticula alba* (Worley, Raper, and Hohl 1979) (Fig. 6b) has been positioned as sister group to Nucleariа (Matthew W. Brown, Spiegel, and Silberman 2009). *F. alba* is a small filose amoeba (5-10 μm), that feeds from bacteria and forms aggregative multicellular fruiting bodies using a stalk formed with golgi derived extracellular matrix.

Finally, there are groups that were classically were related to Nucleariids, regarding its morphological similarities with the genus Nucleariа. One example is the genus of filose amoebae Vampyrellidium Zopf, 1885 (Surek and Melkonian 1980). It has only two species described, but without a culture or molecular data available. Thus it is difficult to further investigate or compare with other nucleariids species. In addition, Vampyrellidium amoebae are able to penetrate the cell wall of algae using a specialized flattened pseudopodium.

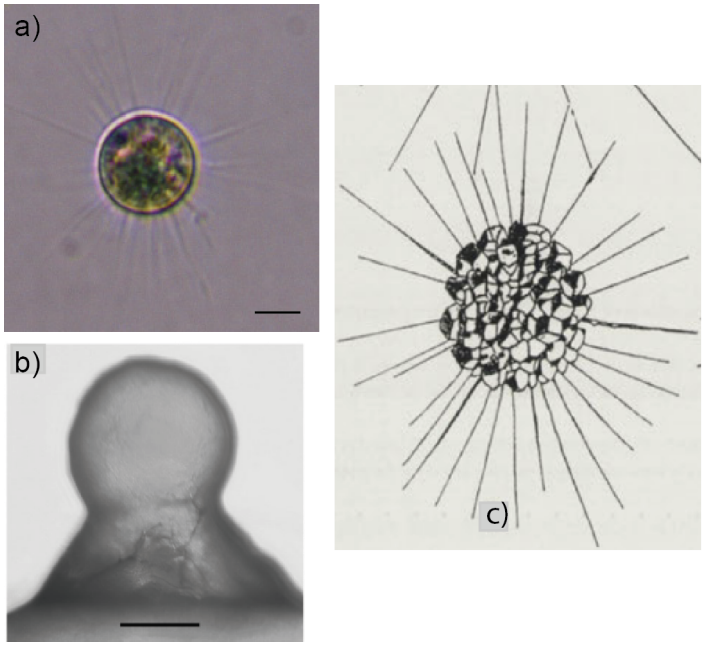


Figure 6. Nucleariids or nuclareiid related taxa. a) Micrograph of *Nuclearia thermophila*. Scale bar = 10µm. b) *Fonticula alba* fruiting body (Brown et al. 2009). Scale bar= 100µm. c) Lithocolla drawing representation from (Mirkjukov et al. 1999).

Other genera morphologically associated to nucleariids (no sequences available) are the Pinaciophora Greef, 1869, Pompholyxophrys Archer, 1869 and Lithocolla Shulze, 1874 (Mikrjukov 1999). All them described as Rotosphareids, which besides, the similarities regarding Nuclearia morphology, they have the particularity to present Silica coverings outside the cell body (Fig. 6c). Rotosphaerids have been found within marine sediments (G. F. Esteban, Gooday, and Clarke 2007).

c) Teretosporea: Plurimorfea and Ichthyosporea

Ichthyosporea a.k.a Mesomycetozoea (L. Mendoza, Taylor, and Ajello 2002), were formerly named as the DRIP clade (an acronym of the original species that composed this clase: Dermocystidium, the 'rosette agent', Ichthyophonus, and Psorospermium (T. Cavalier-Smith 1998). They are a group of osmotrophic/saprotrophic protists, frequently multinucleated, and

sometimes with a single posterior flagellum in dispersal cell forms. Almost all known ichthyosporeans have been isolated from animal tissues, where they live either as parasites, mutualists or commensals (Glockling, Marshall, and Gleason 2013); but a few free-living species have been identified as well (Hassett, López, and Gradinger 2015). Furthermore, it has been defined some environmental clades thanks to 18S ribosonal gene sequences from marine and freshwater environments, known as MAIP and FRESHIP respectively (Del Campoand Ruiz-Trillo 2013; del Campo et al. 2015).

Ichthyosporea are divided in two groups: Ichthyophonida and Dermocystida (T. Cavalier-Smith 1998; L. Mendoza, Taylor, and Ajello 2002; Adl et al. 2012; Glockling, Marshall, and Gleason 2013; Torruella et al. 2015; Grau-Bové et al. 2017). This division is shown by phylogenetic analyses, according to which both groups are monophyletic (Marshall and Berbee 2013; Grau-Bové et al. 2017), and is consistent with phenotypic traits related to morphology and life cycle (L. Mendoza, Taylor, and Ajello 2002; Glockling, Marshall, and Gleason 2013).

The Ichthyophonida is the most species-rich clade according to environmental surveys (del Campo and Ruiz-Trillo 2013) and includes lineages that were previously related with Trichomycetes fungi, Eccrinales and Amoebodiales (Lichtwardt, Cafaro, and White 2001). Within Amoebodiales there are the genus Amoebidium (6 described species) and Paramoebidium (13 species) and both are found as ecto- or endocommensals of freshwater arthropods.
The Eccrinales, with more than 50 species described (Lichtwardt, Cafaro, and White 2001), are found inside arthropod guts in any environment (terrestrial, marine or freshwater) (Cafaro 2005). Eccrinales show a diverse and complex polarized maturation with no amoeboid stage, but contain septate tips, used to release the offspring.

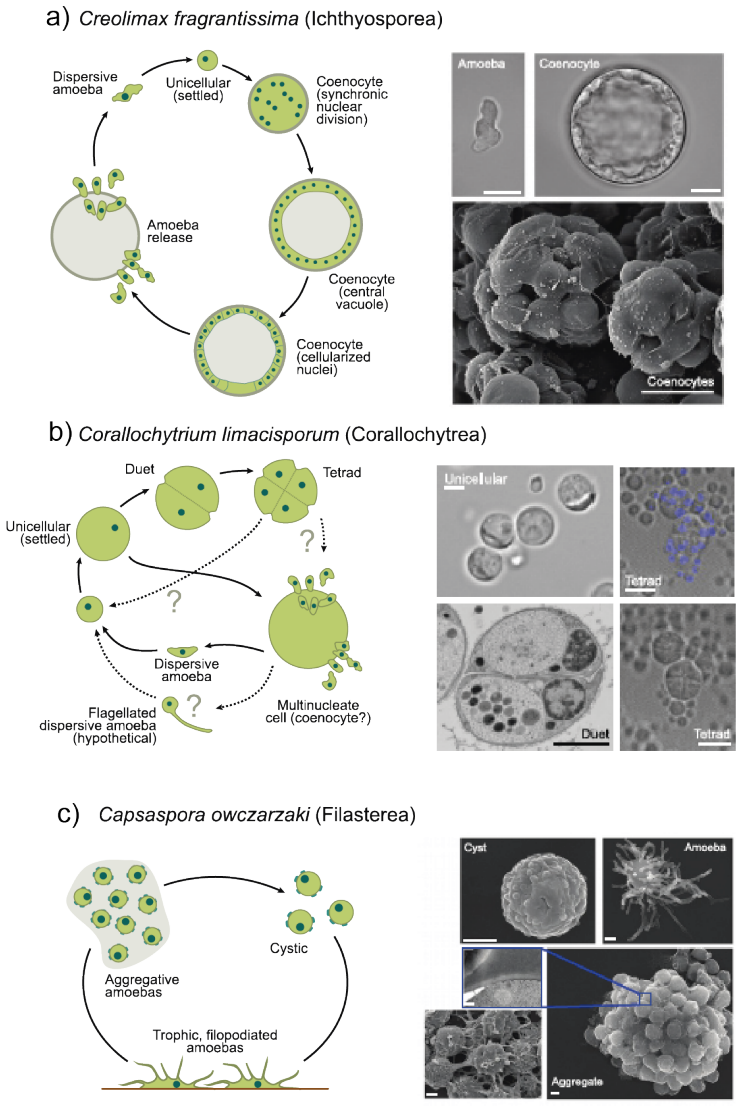


Figure 7. a) Life cycle of *Creolimax fragrantissima*, an ichthyosporean. Single-celled motile amoebas settle and start a coenocytic outgrowth with synchronized nuclear division. The nuclei are gradually displaced towards the cell periphery as a central vacuole grows. Then, individual nuclei are cellularized and released as dispersive amoebas. Scale bars are 10µm, except in lower picture, where it is 50µm. b) Life cycle of *Capsaspora owczarzaki*, an aggregative amoeba within the Filasterea. The proliferative stage consists of filopodiated, surface-adherent amoebas that can form aggregates by extracellular matrix segregation (composition unknown). Amoebas can encyst (resistance form). Scale bars are 1µm, except in the aggregate panel, where it is 200nm. c) Life cycle of *Corallochytrium limacisporum*, sister group to ichthyosporeans. Clonal outgrowths from settled amoebas are similar to *C. fragrantissima*'s, but the

existence of a multinucleate, vacuolated coenocyte is unclear. Sometimes, individual cells undergo (confocal microscopy) serial binary palintomic division to form cell duets (TEM picture), tetrads (pictured with confocal microscopy and DAPI nuclear staining; upper right), etc. A flagellated stage (possibly dispersive) has been hypothesized. Scale bars are 1μm. Adapted from (Sebé-Pedrós, Degnan, and Ruiz-Trillo 2017; Grau-Bové 2017).

The other Ichthyoponida lineages comprise around 40 described species including *Ichthyophonus hoferi*, *Creolimax fragrantissima*, *Pirum gemmata, Abeoforma whisleri, Sphaeroforma tapetis, Sphaeroforma arctica*, *Sphaeroforma sirkka*, *Sphaeroforma napiecek* (L. Mendoza, Taylor, and Ajello 2002; Glockling, Marshall, and Gleason 2013; Hassett, López, and Gradinger 2015). Many ichthyophonids have a broadly conserved developmental mode consisting of large, multinucleated, spherical coenocytes with a central vacuole, that release a dispersive amoeboid stage – sometimes referred to as spores, zoospores, endospores or schizonts– by cellularization of the internal nuclei; amoebas will then disperse and establish a new colony (Figure 7a) (L. Mendoza, Taylor, and Ajello 2002). Ichthyophonid amoebas are frequently spherical or limax-shaped and lack a flagellum. However, some species exhibit fungal-like features: *I. hoferi* can develop hyphal structures (L. Mendoza, Taylor, and Ajello 2002). Others, like *A. whisleri*, exhibit a wide range of phenotypes: cells with pseudopodia, hyphal and plasmodial structures, and amoeboid cell types that can divide without reaching the coenocytic stage (Marshall and Berbee 2011).

The order Dermocystida (sometimes known as Rhinosporideacae) is historically composed of strictly parasitic species, a notable example being the 'rosette agent' *Sphaerothecum destruens*, a well-known fish pathogen (L. Mendoza, Taylor, and Ajello 2002; Glockling, Marshall, and Gleason 2013). Their developmental mode is roughly conserved with ichthyophonids: a spherical sporangium that releases dispersive zoospores. However, the zoospores are often uni-flagellated; and the sporangia can lack the central vacuole. Due to their strictly parasitic nature and difficulties in establishing monoaxenic cultures, they are less well characterized than ichthyophonids from the molecular point of view (Glockling, Marshall, and Gleason 2013). Recently, it has been described a new dermocystid species *Chromosphaera perkinsii* (Grau-Bové et al. 2017).

Several ichthyosporeans have had their genome sequences: *C. fragrantissima* (A. de Mendoza et al. 2015), *I. hoferi* (Torruella et al. 2015) and *S. arctica* (Ruiz-Trillo et al. 2007), *Pirum gemmata*, *Abeoforma whisleri*, *Chromosphaera perkinsii* (Grau-Bové et al. 2017). Transcriptomic data aslo exists for Dermocystid species *Sphaerothecum destruens* (Torruella et al. 2015).

Finally, the transcriptomic profile of *C. fragrantissima* developmental cell types has been investigated in a comparative analysis with other holozoans, and demonstrated that it has a program of transcriptionally regulated cell type specification (A. de Mendoza et al. 2015). Unexpectedly, they identified an up-regulation of animal-like gene tool-kits in the amoeboid dispersive stage, and not in the coenocytic growth phase: this pattern includes developmental transcription factors and adhesion genes involved in the integrin adhesome. The multinucleated coenocytes, instead, appear to have transcriptomic profiles analogous to the proliferative, undifferentiated animal cell types, like stem cells. In parallel, they also demonstrated that *C. fragrantissima* has co-opted ancestral gene regulatory programs to develop a novel osmotrophic feeding mode (absent in non-ichthyosporean holozoans). Overall, they provide direct evidence of the plasticity of cell type evolution across holozoan lineages, supporting a scenario of recurrent recruitment of co-regulated expression programs to support the emergence of novel cell types and developmental programs (Newman 2012).

Plurimorfea is a newly described clade that includes the enigmatic *Corallochytrium limacisporum* and the recently described *Syssomonas multiformis*. *C. limacisporum* is a small (4.5-20 μm) osmotrophic spherical protist. Its life cycle starts with a uninucleated cell that undergoes with a number of rounds of binary cell division during which the daughter cells remain attached to each other, until the release of amoeboid limax-like cells that settle and form new colonies (Fig. 7b) (Raghu-Kama, Chandramohan, and Ramaiah 1987). It is not clear whether it goes through a coenocytic stage like ichthtyosporeans. Interestingly, cell division sometimes occur by palintomic cleavage (i.e., originating Y-shaped junctions and without/little cytoplasmic growth between divisions), a feature that has otherwise been used to classify unclear micro-fossils as animals (Xiao et al. 2012) . Therefore it can be speculated that this

division might be homologous to animal embryonic division (Cunningham et al. 2017), although it can fuel further speculation as to a possible.

Furthermore, it has been proposed that *Corallochytrium* lost its flagellum secondarily (T. Cavalier-Smith 1998), although a recent comparative transcriptomic analysis revealed that it expresses most of the required flagellar genetic tool-kit (Torruella et al. 2015). Those findings are in agreement with its recently grouping with *Syssomonas multiformis*. *Syssomonas* is a freshwater-dwelling predator that can present amoeboflagellate or amoeboid stages. It presents an unknown complex life cycle which include the ability to form a syncytium before dividing into progeny, similarly to *Corallochytrium*.

Finally, *Corallochytrium*, without *Syssomonas*, clustered together with ichthyosporeans in phylogenomics analyses, forming a monophyletic group called Teretosporea (Torruella et al. 2015; Grau-Bové et al. 2017). However, with the addition of *Syssomonas*, using a different phylogenomic dataset with reduced dermocystid and outgrup taxa, plurimorfeans form an independent clade within holozoa between Ichthyosporea and Filasterea. Therefore, it is not clear yet whether Teretosporea are a monophyletic group.

d) Filasterea

Filasterea is the holozoan lineage sister-group to animals and choanoflagellates, that was comprised during almost a decade by only two filopodiated amoebas: *Capsaspora owczarzaki* and *Ministeria vibrans* (Shalchian-Tabrizi et al. 2008) without any related 18S ribosomal environmental sequences (del Campo and Ruiz-Trillo 2013; del Campo et al. 2015). *Capsaspora owczarzaki* was isolated from the hemolymph of the freshwater snail *Biomphalaria glabrata* (Stibbs et al. 1979). It grows anexically in a rich media and presents three different life stages: ameboid, cystic and aggregative (Sebé-Pedrós, Irimia, et al. 2013) (Fig. 7c). It has been found that these three cell stages are regulated at transcriptomically (Sebé-Pedrós, Irimia, et al. 2013), epigenetic (Sebé-Pedrós et al. 2015) and proteomic level (Sebé-Pedrós et al. 2016). The genome of *Capsaspora owczazarki* revealed genes thought to be metazoan specific that were posteriorly lost in the

choanoflagellate species (Hiroshi Suga et al. 2013) *Salpingoeca rosetta* and *Monosiga brevicollis* (see section 1.4b), like the complete integrin-adhesome (Sebé-Pedrós et al. 2010; Hiroshi Suga et al. 2013).

Integrin proteins in animals allow the cells to be attached and interact with the extracellular matrix of tissues. Precisely, integrins proteins are up-regulated in the aggregative stage of *Capsaspora*. On the other hand, *Capsaspora* genome also encodes the transcription factor *Brachyury*, a key protein in the regulation of animal development. Sebé-Pedrós and co-workers showed that *Capsaspora brachyury* protein mimics the function of the *Xenopus* (clawed frog) *Brachyury*. That is, overexpression of *Capsaspora Brachyury* recovers the embryo phenotype under a double mutant of the *Xenopus Brachyury*, showing a high degree of functional conservation (Sebé-Pedrós, Ariza-Cosano, et al. 2013). Indeed, it has been found that the TF network of *Brachyury* is conserved between animals and *Capsaspora* (Sebé-Pedrós et al. 2015). Furthermore, this epigenetic study showed distal regulatory regions –enhancers– are metazoan innovations. Thus, *C. owczarkaki* has been deeply studied and revelead that many genes and molecular functions, key to animal multicellularity, predated the origin of animals.

Ministeria vibrans it is a free-living, marine, filopodiated amoeba with around 4-5 μm (similar size to *Capsaspora*) that preys on bacteria. It conserves an stalked flagella, even though it is a filopodiated amoeba, while *Capsaspora* has lost it completely. Recently, two additional filasterean species have been described: *Pigoraptor vitetnamica* and *Pigorapotor chileana*. Both Pigoraptor species were isolated from freshwater environments. They are flagellated predators of eukarytoes, even though they can feed on bacteria or detritus. Pigoraptors can adopt cystic forms like *C. owczarzaki* and can present multicellular clusters of a few cells (Hehenberger et al. 2017). Thanks to the addition of Pigoraptors into ribosomal SSU phylogenies, it seems that the previously defined environmental holozoan group MAOP1 is also related to filasterans (Hehenberger et al. 2017). Pigorapotor and Ministeria genus have transcriptomal data available but their genomes have not yet been sequenced.

e) Choanoflagellatea

Choanoflagellates are a well-known group of unicellular eukaryotes that have long been associated with animals due to their apparent cytological similarities with choanocytes, a cell type of sponges (James-Clark 1866). The position of choanoflagellates as the closest unicellular relatives of animals was later confirmed with molecular phylogenies (Wainright et al. 1993; Zettler LAA et al. 2001; Lang et al. 2002; Ruiz-Trillo et al. 2004; Ruiz-Trillo et al. 2008).

Choanoflagellates is a diverse protist group, with approximately 250 described species (M Carr et al. 2008; B.S.C Leadbeater 2015). They are aquatic, either marine or freshwater, heterotrophic organisms that prey on bacteria, being important players in microbial food webs. Morphologically, choanoflagellates are characterized by an ovoid to spherical cell body containing a single anterior flagella surrounded by a collar of microvilli (B.S.C Leadbeater 2015).

Molecular phylogenies based in a few genes showed that choanoflagellates are divided into two major clades, known as Craspedida and Acanthoecida (Martin Carr et al. 2008; Paps et al. 2013; Martin Carr et al. 2017). Craspedida includes the choanoflagellates with organic coverings, that can be thecated (Salpingoecidae morphology) with cup-, tube-, flask- shaped investments; or non-thecated with non-restrictive coverings like glycocalix or sheath (Codosigidae morphology) (Fig. 8c). It has been found that species are not phylogenetically clustered in different theca morphologies (Martin Carr et al. 2017). Within Craspedia, three clades have been described by a phylogenetic approach based in 6 different genes and 47 choanoflagellate taxa (Martin Carr et al. 2017): Clades 1, 2, 3 (Fig. 9). Only Clade 3 presents species with the same tube shaped theca morphology (Fig. 8b) (*Salpingoeca tuba* and *Salpingoeca dolichothecata*).

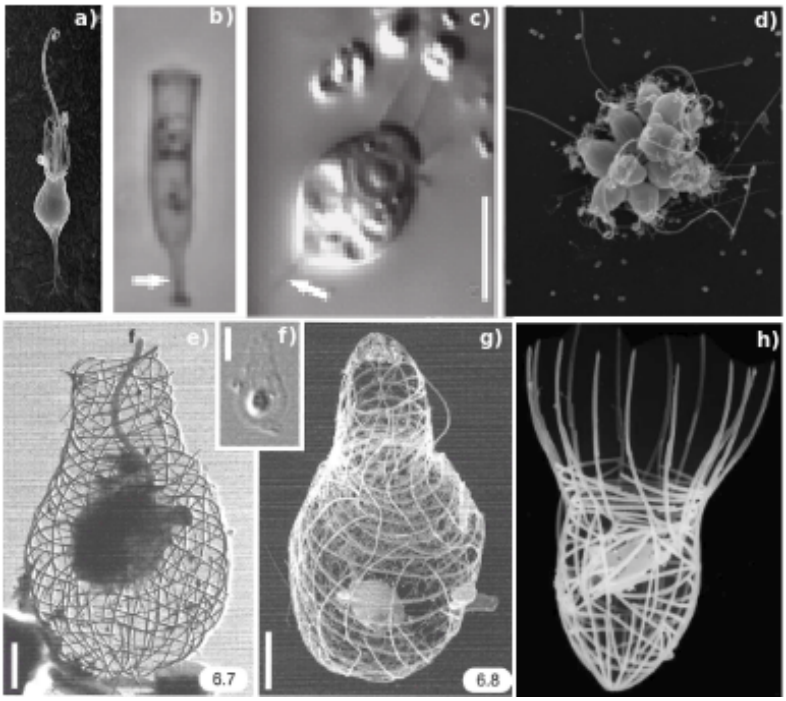


Figure 8. Choanoflagellates morphology. a) SEM picture from an stalked *S. rosetta* cell (www.pinterest.com). b) Empty tube theca of *S. dolichothecata* , phase contrast (Martin Carr et al. 2017). c) *Codosiga hollandica* cell on peduncle, DIC (Martin Carr et al. 2017). d) *S. rosetta* colony, SEM (www.pinterest.com). e,f, g) Nudiform *Savillea parva* adopting the parva form (source (B.S.C Leadbeater 2015)), TEM, Interference contrast micrograph and SEM, respectively (Scale bars e) and g) = 2µm, f) = 1µm). h) Tectiform lorica *Acanthocorbis unguiculata* (www.pinterest.com).

Among the other clades, flask theca and non-thecated choanoflagellates are paraphyletic, however cup theca seems exclusive from Craspedida clade 1 (Martin Carr et al., 2017). The most common life cycle of craspedid choanoflagellates comprises a sedentary interphase, whose major function is feeding, followed by cell division, that give rise to a transitory motile phase, whose function is dispersal. Craspedids also can produce cell colonies. Choanoflagellates with codosigid morphology present an stalked colony, where daughter cells resulting from division remain attached to the parent stalk, thereby forming a head of cells. A mature head may contain 10-20 cells and, eventually, the colony can be dislodged and swim to another place, being indistinguishable

from other naked craspedida genera like *Sphaeroca* (Leadbeater, 2015). However, the best-known colonial choanoflagellate is *Salpingoeca rosetta,* which together with the non-colonial *M. brevicollis* are the only two choanoflagellate species with their genome sequenced (Fairclough et al., 2013; King et al., 2008).

S. rosetta colonies are produced when the cells are floating (Fig. 8d), although the cells can be attached to the surface with the help of the cup theca (Fig. 8a). It has been proved that, transcriptomically, *S.rosetta* colonies present a differential gene expression patterns, showing a transcriptionally regulated life cycle through different stages (Fairclough et al., 2013). In addition, it has been shown that colony formation in *S. rosetta* is triggered by the presence of a bacterial sphingolopid[2] (Alegado et al., 2012). It has also been identified a gene crucial for rosette development[3], the rosetteless, which is a C-type lectin protein that localizes in the center of rosettes. Thus rosetteless, can help to bind the cells among each others or to the extracellular-matrix. It is not clear if these proteins have an homologous functions that the ones found in animals or it is a case of convergent evolution.

S. rosetta has been found to have cell reproduction within swimming cells that can undergo meiosis and fuse afterwards (Levin & King, 2013). Recently, it has been found that sex it is also regulated by the presence of bacteria (Woznica et al. 2017). Thus, overall *S. rosetta* presents a complex life-cycle with sexual reproduction regulated through differential cell expression (Fig. 10). Life cycle development is dependent on the bacterial environmental stimuli, that facilitates the molecular mechanism to allow the colony formation or sexual reproduction.

[2] The sphingolipid is from the bacteriodetes *Algoriphagus machipongonensis.*
[3] Rosette is the name recived by the colonies of *S.rosetta*

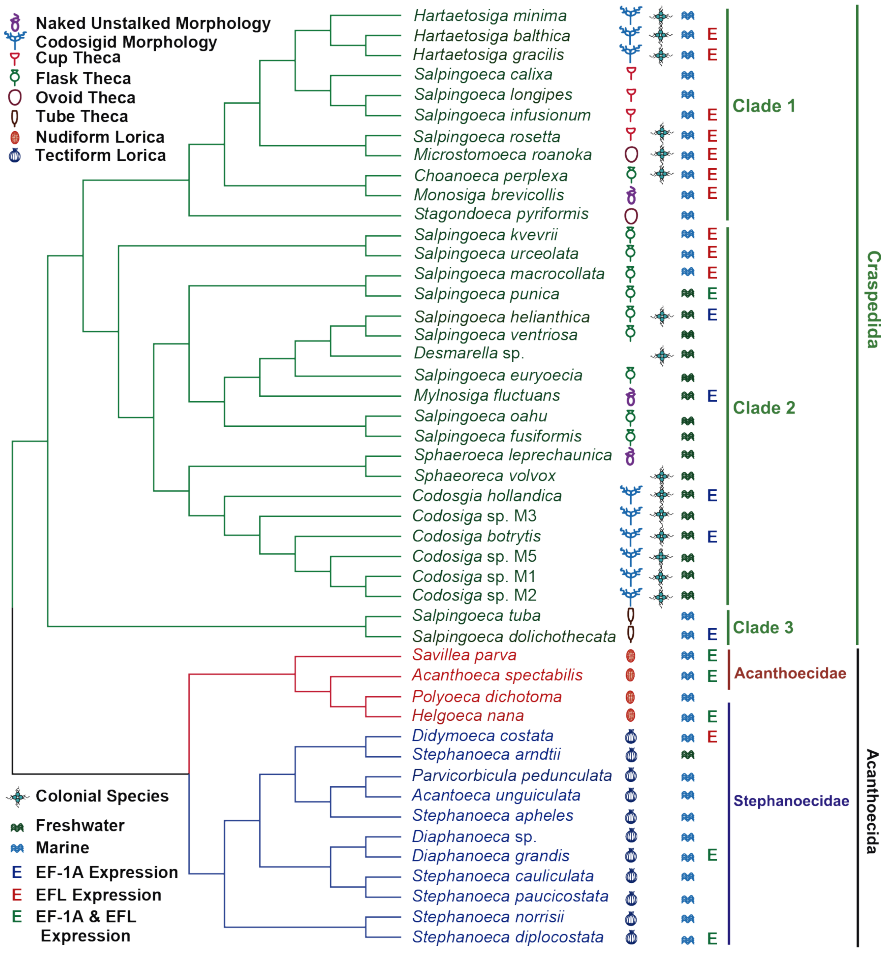


Figure 9. Morphological, ecological and genomic traits of the choanoflagellates Adapted from (Martin Carr et al., 2017). The representative phylogeny highlights traits for each species present in the main phylogeny. Periplast morphology is shown in the first column and the identity of known colonial species shown in the second column. The third column identifies species which are found in a freshwater or saline environment and the final column shows which species are known to express EF-1A and EFL. Gaps represent missing data

On the other hand, Acanthoecida is composed by choanoflagellates with a siliceous loricae, being most of the described species marine and with a tectiform lorica (Fig.8h) (Stephanocidae) (around 150 species) (Martin Carr et al. 2008), although there are 5-6 species described with nudiform lorica (Acanthoecidae) (Fig.8efg) (B.S.C Leadbeater 2015). The lorica is formed with rod-shaped units attached to each other end-to-end to form costae that combine in two layers to produce a rigid basket-like cage in which is located

the choanoflagellate cell. Kent (1878-1880), was the first author in illustrate and described this choanoflaegelalte structure. The main.

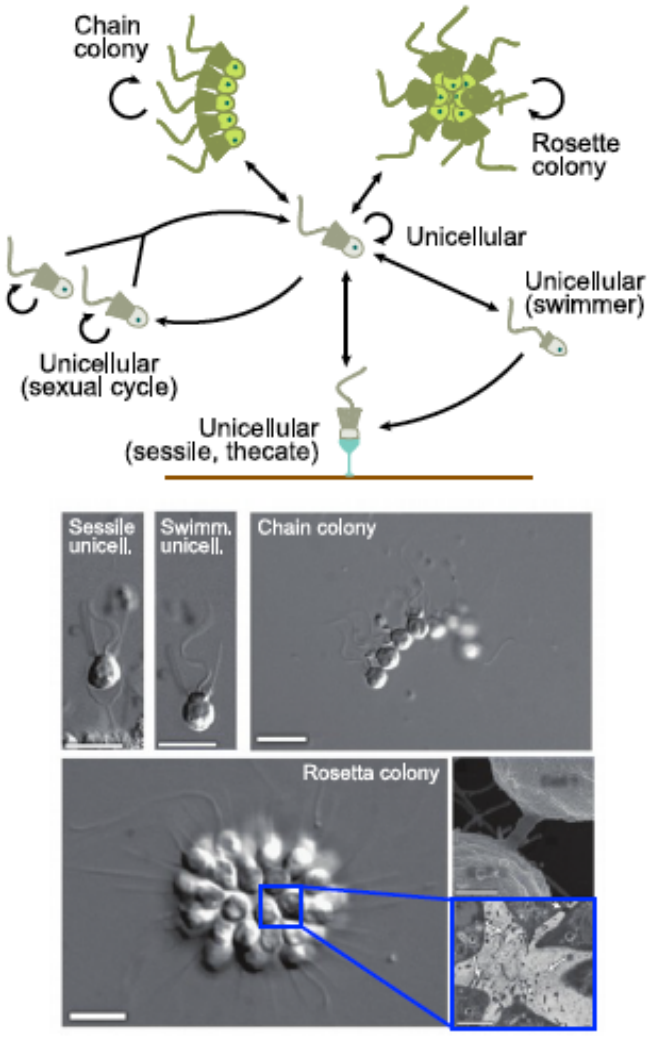


Figure 10. Life cycle of Salpingoeca rosetta, adapted from (Grau-Bové 2017; Levin et al. 2014) . Circular arrows indicate the proliferative stages. It comprises two kinds of colonies (chain and rosette-like), three unicellular stages (fast- and slow-swimming and a thecate, sessile flagellate) and a unicellular sexual cycle (meiosis). Scale bars are 5µm, except in the 'rosette' right panel, where it is 1µm. (Sebé-Pedrós, Degnan, and Ruiz-Trillo 2017; Grau-Bové 2017).

differences between tectiform and nudiform species are: 1) after cell division, nudiform juvenile cells swims away from parent lorica, while juvenil tectiform cells are not motile and receives a full set of strips from parent cell; 2) nudiform cells produce the outer layer first and tectiform species start for the inner layer. It has been done a lot of work in the understanding of the formation of these loricate structures for a complete view can be found in Barry S. C. Leadbeater Book, the Choanofagellates (Leadbeater 2015)

Finally, there are other environmental clades of choanoflagellates defined by environmental 18S ribosomal sequences, such as Clade L (Weber et al. 2012), FRESCHOs and MACHOs (del Campo and Ruiz-Trillo 2013).

f) Metazoa

Metazoa or animals comprise the multicellular heterotrophic organism that have their cells differentiated in cell types, tissues and organs. They are the best known Opisthokonta lineage, with 1.3 million species described (del Campo et al. 2014) and, it is estimated that this number can increase up to 10 million (del Campo et al. 2014). Metazoa is comprised of 35 phyla (Fig.11). The phylogenetic relationships among the early-branching metazoan phyla (Ctenophora, Porifera, Cnidaria and Placozoa) is still controversial.

Classically, it was thought that the earliest-branching metazoan lineage were the sponges, because they are simpler –sponge lack of neurvous system and present a simpler morphology (Dunn et al. 2013)–. Moreover, earlier phylogenies had shown sponges to be the sister-group to the rest of animals. However, a phylogenomic approach based in EST data placed Ctenophora at the base of Metazoa (Dunn et al. 2008). Since then, many phylogenomic studies have appeared either denying or accepting this hypothesis and their position is on debate (Philippe et al. 2009; Nosenko et al. 2013; Whelan et al. 2015; Giribet 2015; Simion et al. 2017; X.-X. Shen, Hittinger, and Rokas 2017). On the other hand, bilaterian animals –the ones that have bilateral symmetry– include the early-branching Xenoacelomorpha (Cannon et al. 2016), Deuterostomia and Protostomia. Within those clades there are also phyla whose phylogenetic position are not fully clear. Therefore, there are still many open questions regarding animal evolution that will require to

obtain genomic data of more organisms coupled with improved phylogenomics dataset and algorithms (Giribet 2015; X.-X. Shen, Hittinger, and Rokas 2017). Hopefully, in the near future, we will have a more clear vision regarding the evolution of animals from the Urmetazoa[4] to the extant phyla.

Animals are multicellular heterotrophic organisms that prey on bacteria or other eukaryotic organisms. Multicellularity allowed animals to cover a great variety of sizes and morphologies. Actually, animals cover from microbial sizes (smaller than 2mm) (Blaxter et al. 2005) to animals up to 30m long, as is the case of *Balaenoptera musculus* (the blue whale). Animal cells are in contact with each others through junction molecules (Adell et al. 2004), some of which make transport of nutrients between the cells possible. This enables the division of labor of the different cell types, leading just a group of cells in charge of obtaining the nutrients, and the others free of doing other functions like to digest, sense, contract, secrete etc. This is one of the key differences between animal and choanoflagellate transient multicellularity (Nielsen 2012). This division of labor, has allowed animals to develope a wide variety of cell types distributed in different tissues and organs –Bilaterea animals can have more than 50 different cell types (Carroll 2001)–.

Another characteristic of metazoans is their life cycle, based in sexual reproduction. An haploid sperm fuses with an haploid egg forming the zygote. Besides the gametes (sperm and egg), the rest of metazoan cells are diploid. Once the gametes are fused in a zygote, it divides forming the embryo. The embryo divides, starting a program based in successive cell divisions –embryogenesis–, accompanied by cell differentiation. During the embryogenesis cells become organized in functional units, usually forming layers that give rise to tissues and organs. Many animals extend its development in larval/juvenil stages until they become a new adult with fully reproductive capacities (Nielsen 2012).

[4] The metazoan last common ancestor

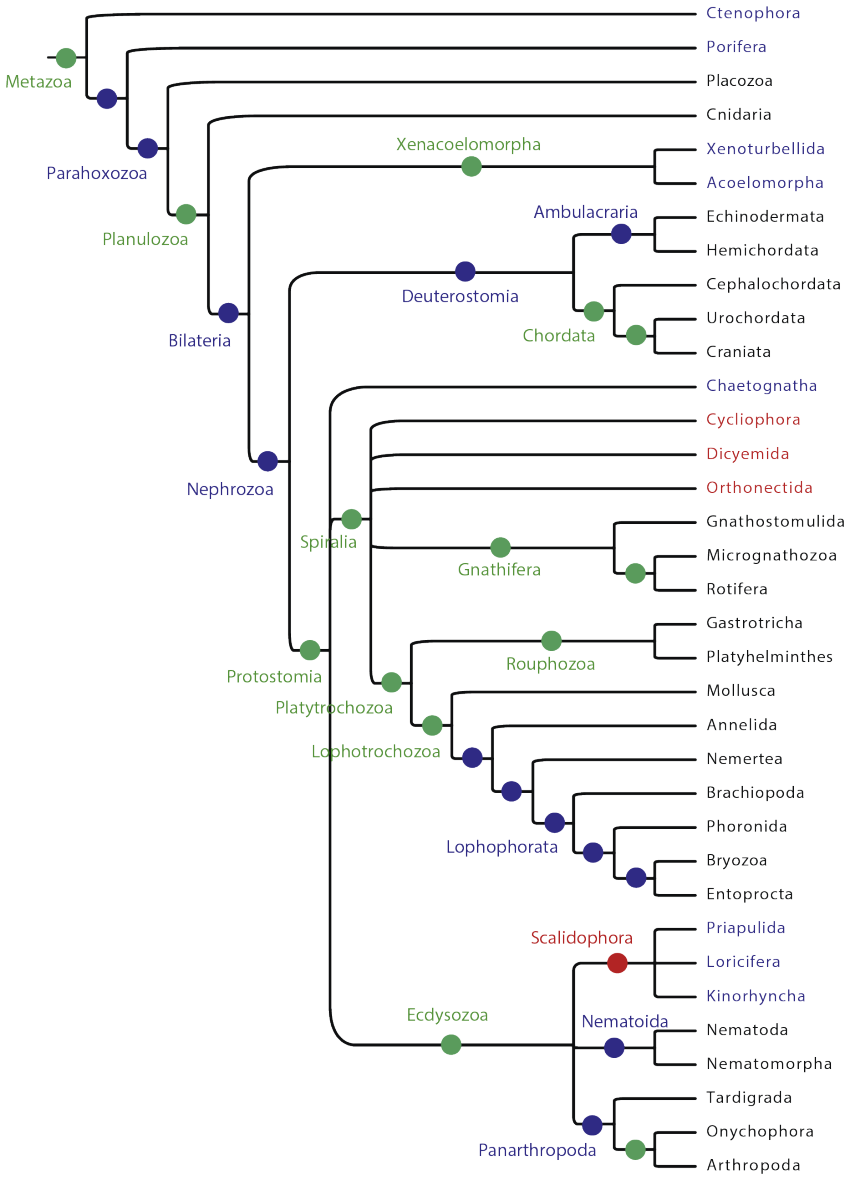


Figure 11. Hypothesis of animal phylogeny derived from multiple phylogenomic sources. Adapted from (Giribet 2015). Taxa in red indicate unstable taxa, taxa with deficient genomic/transcriptomic data, or taxa for which no phylogenomic analysis is available. Taxa in blue indicate conflict between some studies, but with a relatively stable position. Green circles indicate clades supported across most well-sampled studies; blue circles indicate clades that are contradicted in some studies, especially due to the position of some rough taxa; red circle indicates a putative clade not thoroughly tested in phylogenomic analyses

Many phyla have a blastula stage during the development, and it has been considered one of the most important apomorphies of the Metazoa (Nielsen, 2012). In fact, one of the earliest stages of animal development, the Blastula, inspired Ernst Haeckel to propose the Gastrea theory.

In that theory, Haeckel suggest that the first step in the evolution of Metazoa was a hollow ball of identical and flagellated cells, which he called the blastea (Haeckel 1874). Modern authors have adapted this idea, to use instead the concept of choanoblastea theory, in which the flagellated blastea, comes from a choanoflagellate colony (Nielsen 2008).

Choanoblastea theory argues that at the early phase of the transition to multicellularity, all the cells that compose the blastea were undifferentiated, and by successive generations they started to develop different cell types. Those differentiated cell-types lead to blastula formation and the colony was becoming highly complex until the formation of the Urmetazoa (Fig.12a) (Nielsen 2008).

This is a nice view. However, the unicellular relatives of animals like the ichthyosporean, *Creolimax fragantissima*, the filasterean *Capsaspora owczarzaki*, and the choanoflagellate *Salpingoeca rosetta* have shown to present transcriptionally differentiated cell types (Sebé-Pedrós, Burkhardt, et al. 2013; Fairclough et al. 2013; A. de Mendoza et al. 2015) (see sections 1.3cde). In the case of *Capsaspora* these differences have been proven to be also at epigenetic (Sebé-Pedrós et al. 2015) and proteomic level (Sebé-Pedrós et al. 2016). Therefore, these evidences, make the authors Sebé-Pedrós et al. propose a new model regarding the origin of animal multicellularity (Sebé-Pedrós, Degnan, and Ruiz-Trillo 2017), inspired in the Synzoospore theory originally proposed by Zakhvatkin in 1949, and more recently developed by Mikhailov et al. (Mikhailov et al. 2009). Sebé-Pedrós et al. propose that a the single-celled ancestor of metazoans had a complex life cycle influenced by different environmental stimuli, likely, with colonial and aggregative stages and also, with sexual reproduction. Different environmental stimuli provoked the progressive introgression in one entity, of all temporary different cell types. More cell types were

developed given rise to the Urmetazoa (Sebé-Pedrós, Degnan, and Ruiz-Trillo 2017) (Fig. 12b). Therefore, the authors suggest that the origin of metazoans consisted in the transformation from an organism with temporally differentiated cell types into a single entity with spatial cell differentiation.

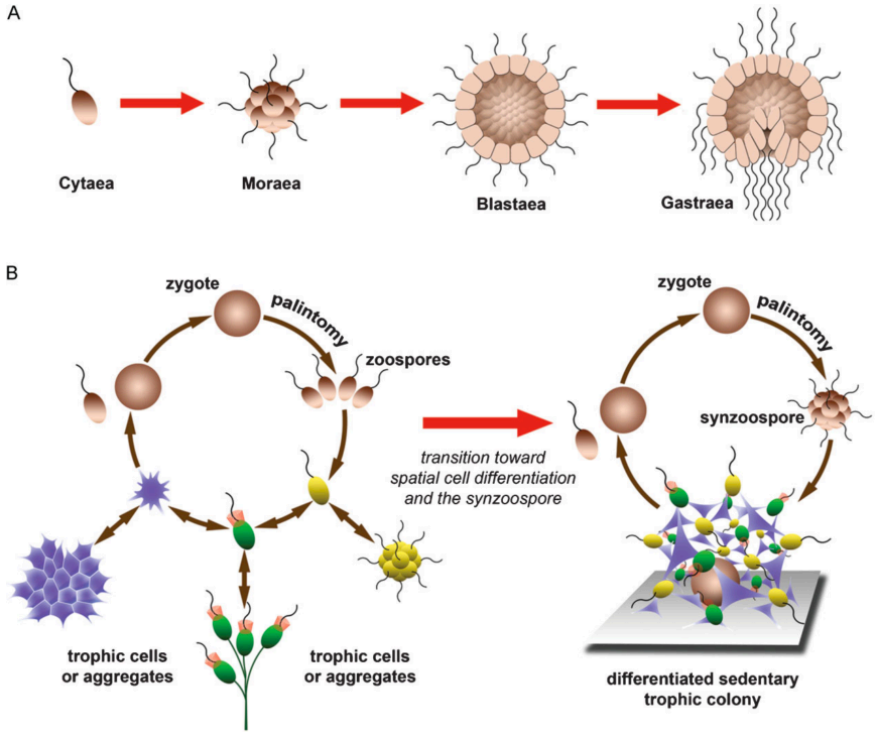


Figure 12. The Gastraea and Synzoospore scenarios, from (Mikhailov et al. 2009). a): The Gastraea theory assumes gradual modification of a colony of uniform cells. Primary cell differentiation occurs with the formation of functional primary gut, the evolutionary precursor of endoderm. b): The Synzoospore theory envisions the metazoan ancestor as a protist with a complex life cycle that includes monotomously dividing trophic cells (or cellular aggregates), hypertrophic growth of gametes, and their subsequent palintomic cleavage producing non-feeding dispersal zoospores. The transition to multicellularity occurs with (i) integration of trophic cells into a differentiated colonial body and (ii) integration of zoospores into the uniform synzoospore, the primary lecitotrophic dispersal larva of the animals. Red arrows mark hypothetical evolutionary transitions, brown arrows designate the life cycle.

Overall, Metazoa is the eukaryotic group with most species described, spread in 35 different phyla, some of the phylogenetic relationships among animal phyla are still on debate. In addition, animals present an extremely regulated development, which has allowed the appearance of organisms with many different cell types and morphologies. Therefore, understanding the origins of animal multicellularity and their evolution is one of the main biological questions. In this regard, studies on unicellular Opisthokonta diversity have been key in the formulation of hypothesis regarding the origin of animals. However to fully understand these processes we need to know when that important transition could had happenned and which were the molecular mechanisms that could facilitate this transition. In the following section, I will explain what is known so far on those topics, thanks to the animal fossil record and to the genome sequences of unicellular holozoans.

1.4 The Origins of Metazoa

a) Fossil record and biogeochemical context

To determine when the first animals appeared on earth we have to look at the fossil record. However, fossil record regarding the origin of animals has always been unclear. In his book *On the Origin of Species,* Charles Darwin already suggested that one of the greatest challenges to his ideas was the "sudden appearance of groups of allied species in the lowest known fossilliferous strata" (Darwin, 1859). Darwin could not trace the ancestor of trilobites, brachiopods, molluscs and other lower Cambrian species, because they were not preserved in the previous rock records. This sudden appearance of different animal phyla precursors, it is known as the "Cambrian explosion". Nowadays, there are ancient fossil records of pre-Cambrian fauna, belonging to Ediacarian period and before. Although, the fossil record is not conclusive, and those forms cannot be easily assigned to any eukaryotic affiliation.

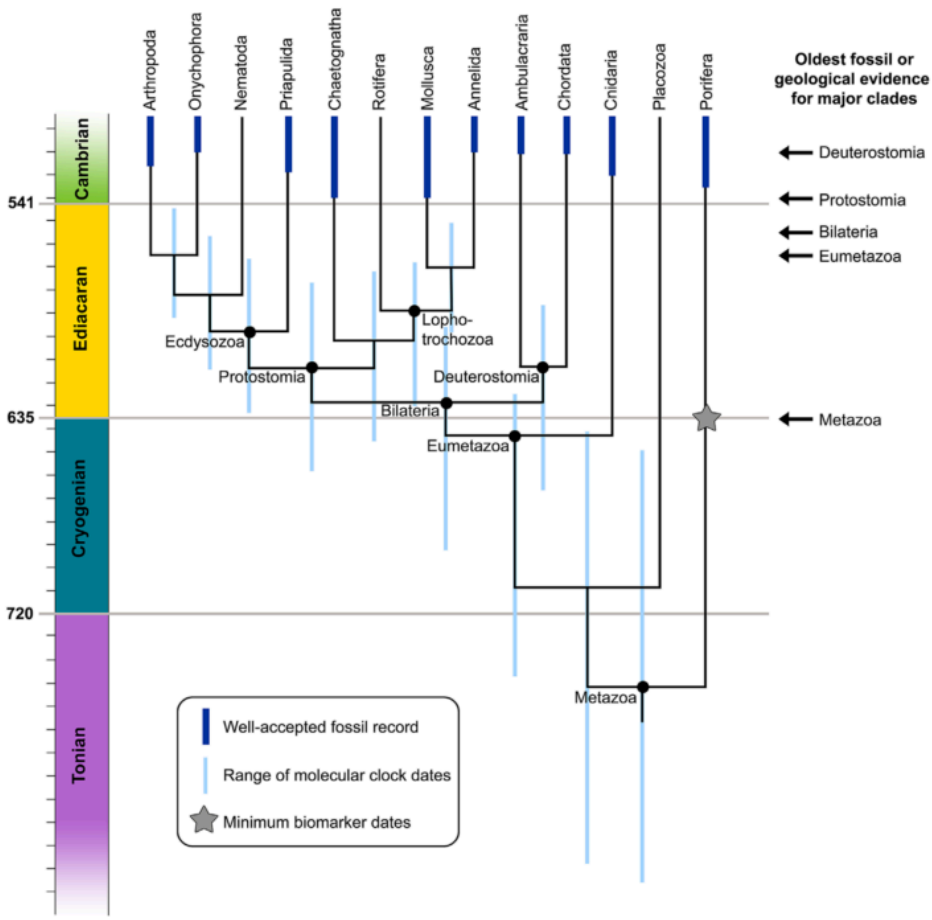


Figure 13. The mismatch between the fossil and molecular clock records of early animal evolution, from (Cunningham et al. 2017). The phylogeny follows (Dos Reis et al. 2015); note that ctenophores, the phylogenetic position of which is contentious, were not included. Dark blue bars represent well-accepted reports of fossils that can be assigned to extant animal phyla, which are limited to the Cambrian; ranges mainly follow (Erwin et al. 2011). Pale blue bars represent the range of molecular clock estimates for the origins of major clades obtained in (Dos Reis et al. 2015); note that the origin of eumetazoans is always inferred to predate the Ediacaran and the origins of bilaterians, protostomes, deuterostomes, ecdysozoans and lophotrochozoans are always inferred to predate the Cambrian. The righthand column shows the first evidence, as interpreted here, for major clades in the geological record: Metazoa = 635Ma, possible biomarker evidence, alternatively 565Ma eumeta- zoan trace fossils; Eumetazoa = 565Ma, trace fossils; Bilateria = 555 Ma, trace fossils; Protostomia = 540Ma, helcionellids, protoconodonts; Deuterostomia = 530 Ma, echinoderm plates.

The first clear animal evidence, thanks to the presence of biomarkers dates from 635 Mya, and the first eumetazoan fossil from 565 Mya (Cunningham et al. 2017). Molecular clocks estimates, place the origin of animals between 850-640 Mya (Cunningham et al. 2017) (Fig 4. and Fig. 13). The average of this estimate (~750Mya) corresponds to the second burst of atmosphere oxygenation (Lyons, Reinhard, and Planavsky 2014). The first event started at 2.3 billion years ago and was carried out by photosynthetic cyanobacteria, that lead oxygen levels at ~1% of present atmospheric levels (Alegado and King 2014) (Fig. 4), a value that is at the threshold of the minimum requirements for metazoan life (Sperling et al. 2013; A. H. Knoll and Sperling 2014).

During the Cryogenian (720Mya), the levels started to rise again until the Cambrian period, in which oxygen levels remained stable since our days (Lyons, Reinhard, and Planavsky 2014). This is the reason that many authors have linked the availability of oxygen with the emergence of animals. Firstly, because high oxygen concentrations were thought to be needed in order to synthesize the collagen-based extracellular matrices that sustain multicellular tissues in Metazoa (Towe 1970). However, this is not entirely true. For example, the demosponge *Halichondria panicea* can grow at 0.5-4% of the present atmospheric oxygen levels (Mills et al. 2014) (Mills et al. 2014); some bilaterians as low as 0.3% (Mills and Canfield 2014). Moreover, collagen synthesis can occur at low oxygen concentrations, albeit with lower efficiency (Mills and Canfield 2014).

Finally, another factor that has been linked to the origin of animals is the bacterivory capacity of the earlier animals. Stromatolites[5] diminish drastically among Cambrian rocks indicating that they could be the source of food for first animals (Alegado and King 2014). In addition, the fact that colony formation of choanoflagellate *S. rosetta* is triggered by the presence of a bacterial sphingolipid (Alegado et al. 2012), and that a bacteria regulate its

[5] Stromatolites are layered biochemical structures formed in shallow water by by the trapping, binding and cementation of sedimentary grains by biofilms of microorganisms especially cyanobacteria.

sex reproduction (Woznica et al. 2017) suggest that bacteria could have played a role in origins of animal multicellularity.

b) Genomic changes in the transition towards animal multicellularity

The genome sequences of key early-branching animals and their unicellular relatives over the last decade (Putnam et al. 2007; Nicole King et al. 2008; Srivastava et al. 2010; Hiroshi Suga et al. 2013; Fairclough et al. 2013; Simakov and Kawashima 2016; Ryan et al. 2013; Moroz et al. 2014; Grau-Bové et al. 2017) has open a new window in the understanding of the genetic content present in the Urmetazoa, and in the unicellular ancestor of animals. This offers a new vision of the changes that might have occurred at a molecular level to originate the transition towards animal multicellularity. In this section I will briefly summarize some of these discoveries.

The genomic comparisons between animals and their unicellular relatives (choanoflagellates, filastereans, teretosporeans, see sections 1.3cde) has shown that the unicellular ancestor of animals was already equipped with a rich repertoire of genes involved in multicellular functions, including developmental transcription factors –like *Brachyury*, MYC, Runx or even P53–, cell adhesion proteins –ECM elements, integrins, cadherins and C-type lectins– and cell signaling receptors and transducers (Nicole King et al. 2008; Sebé-Pedrós et al. 2010; Sebé-Pedrós et al. 2011; H. Suga et al. 2012; Fairclough et al. 2013; Richter and King 2013; Sebé-Pedrós, Degnan, and Ruiz-Trillo 2017) (Fig. 14). These findings suggests that co-option of ancestral genes into new functions was an important mechanism that occured in the transition from the unicellular ancestor of animals to the Urmetazoa (Sebé-Pedrós, Degnan, and Ruiz-Trillo 2017). Although, interestingly, the function could be partially conserved, as it has been shown for the TF *Brachyury* (Sebé-Pedrós, Ariza-Cosano, et al. 2013) or the integrins of *Capsaspora owczarzaki* (Sebé-Pedrós, Irimia, et al. 2013) (see section 1.3d).

However, not only gene co-option was the molecular driver in the transitions towards animal multicellularity. There was a process of gene innovation with around 300-400 novel genes (Srivastava et al. 2010; Tautz and Domazet-los 2010; Richter and King 2013;

Simakov and Kawashima 2016). Some of these animal specific genes are product of domain shuffling events, as it is the case of the Notch and Hedgehog proteins (Nicole King et al. 2008; Fairclough et al. 2013), which are involved in key signalling pathways for animal devolpment. Actually, cell signalling pathways are a nice example in which gene co-option and the innovation of new proteinic players conformed the establishment of metazoan-specific cell-to-cell communication systems.

The establishment of diverse signal transduction pathways is essential in order to coordinate the functions of a multicellular body (Bonner 1998; N. King 2003). One of the major signaling systems of eukaryotes is protein phosphorylation, by which protein products can be labeled with phosphate groups in specific residues. These phosphorylation systems are involved many signaling and processes: cell-to-cell and cell-to-matrix adhesion, proliferation, development, or differentiation (Nicole King 2004). Tyrosine-specific kinases, together with serine/threonine kinases, are the dominant phosphorylation systems of eukaryotes (Choi et al. 2008) and consist of a wide array of highly diverse gene families that are thoroughly conserved in Metazoa. However, recent studies have also identified important enrichments in their closest unicellular relatives, like *M. brevicollis*, *C. owczarzaki*, *M. vibrans* (Suga et al. 2012), and teretosporeans (H. Suga et al. 2012; Hiroshi Suga et al. 2014; Grau-Bové et al. 2017).

Interestingly, this holozoan-wide expansion of phosphotyrosine signaling was due to a dual evolutionary trend by which the cytoplasmic enzymes tend to be conserved across holozoan genomes, but the membrane-bound receptor enzymes are largely lineage- or species-specific (Hiroshi Suga et al. 2014).
Furthermore, It has been shown in *Capsapora owczarzaki* that proteins related in signaling processes such tyrosine kinases, presents different phosphorylation patterns among the different cell stages. Therefore, the phosphorylation of those proteins modulate temporal cell differentiation in *Capsaspora*, unraveling a molecular mechanisms that could have been used as well by the single-celled ancestor of animals (Sebé-Pedrós et al. 2016), and later transformed to spatial regulation in the origins of animals (Sebé-Pedrós et al., 2017) (see section 1.3f).

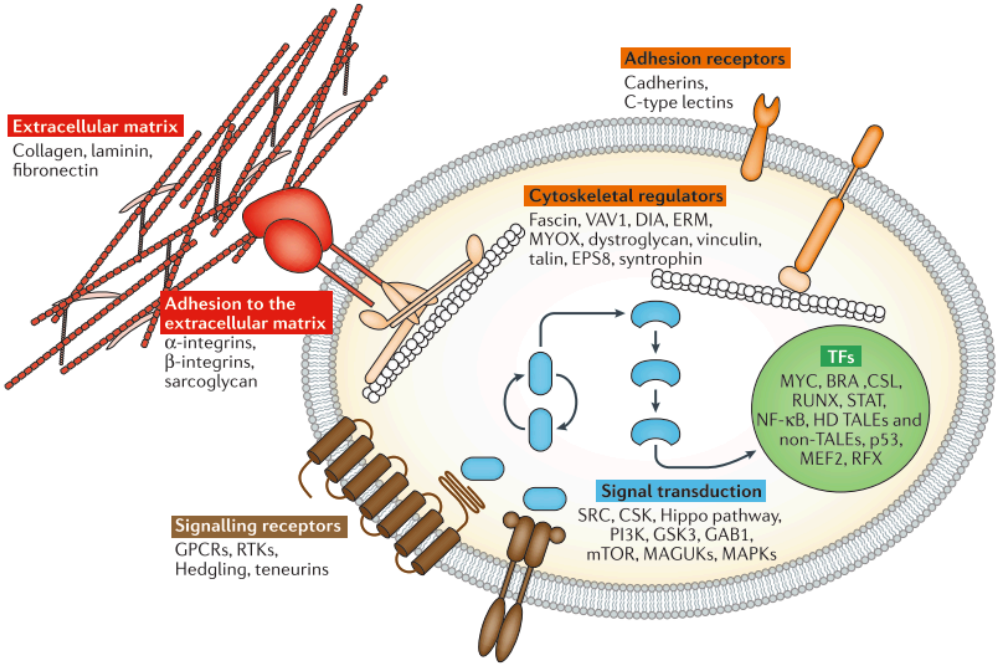


Figure 14. The pre-metazoan genetic tool-kit. Adapted from (Sebé-Pedrós, Degnan, and Ruiz-Trillo 2017). EPS8, epidermal growth factor receptor kinase substrate 8; GAB, GRB2-associated binding protein; GPCRs, G protein-coupled receptors; GSK3, glycogen synthase kinase 3; HD, homeodomain; MAGUKs, membrane-associated guanylate kinases; MAPKs, mitogen-activated protein kinases; MEF2, myocyte-specific enhancer factor 2; mTOR, mechanistic target of rapamycin; MYOX, myosin X; NF-κB, nuclear factor-κB; PI3K, phos phatidylinositol 3 -kinase; RTKs, receptor tyrosine kinases; STAT, signal transducer and activator of transcription; TALEs, three amino acid loop extensions; TF, transcription factor.

It is frequent that some ligands of a metazoan signalling pathway were present in the unicellular ancestor but the receptors or other components of the pathway have been gained at the stem of Metazoa, as it is the case of the Hippo pathway (Sebe-Pedrs et al. 2012). Thus, as comented, metazoan specific signalling pathways are an example of co-option of ancient proteins and acquisition of new ligands and receptors at the onset of metazoa.

Finally, another type of cell-to-cell signaling are the molecular mechanisms that are involved in neural functions. It is indeed, an important case of cell-to-cell communication in animals. It has been described many components related with neural functions that

predated the origins of animals like sodium (Liebeskind 2011) and calcium channels (Cai 2008), Neuroglobulins (Lechauve et al. 2013) proteins related in synapsis (Alié et al. 2011; Fairclough et al. 2013) and postsynaptic functions like Homer and processes related in neural secretion (Burkhardt et al. 2011).

2. OBJECTIVES

The general framework of my thesis is unravel new Opisthokonta diversity at many different levels: from describing new species to reveal the genomic sequences of uncultured choanoflagellate taxa. The aim was to have a better understanding of Opisthokonta evolution and the origins of animal multicellularity. Thus, to this end, I have focused in six main objectives:

1. The characterization at morphological level of the incertae sedis species Nuclearia sp. ATCC 50694 and determine its phylogenetic position within nucleariids molecular diversity by sequencing its 18S ribosomal gene.

2. Analysis of the microbial metazoan diversity of European marine coastal environments through a metabarcoding dataset in order to determine putative new molecular metazoan diversity.

3. Evaluation of the utility of single-cell genomics techniques to address evolutionary questions by sequencing three single-cell amplified genomes of the choanoflagellate *Monosiga brevicollis*.

4. Sequencing and analysing the SAGs of 4 undescribed choanoflagellate taxa using the experience acquired in the third point.

5. Resolution of the phylogenetic relationships of choanoflagellates taxa using the sequenced SAGs.

6. Comparative genomics of our SAGs and other unicellular holozoans with animals, in order to update the pre-metazoan genetic toolkit, taking into account as well genomes of the rest of eukaryotic diversity.

3. RESULTS

3.1 Parvularia atlantis gen. et sp. nov., a Nucleariid Filose Amoeba (Holomycota, Opisthokonta).

López-Escardó D, López-García P, Moreira D, Ruiz-Trillo I, Torruella G. Parvularia atlantis gen. et sp. nov., a Nucleariid Filose Amoeba (Holomycota, Opisthokonta). J Eukaryot Microbiol. 2018 Mar;65(2):170–9. DOI: 10.1111/jeu.12450

3.2 Metabarcoding analysis on European coastal samples reveals new molecular metazoan diversity

López-Escardó D, Paps J, de Vargas C, Massana R, Ruiz-Trillo I, Del Campo J. Metabarcoding analysis on European coastal samples reveals new molecular metazoan diversity. Sci Rep. 2018 Jun 14;8(1):9106. DOI: 10.1038/s41598-018-27509-8

3.3 Evaluation of single-cell genomics to address evolutionary questions using three SAGs of the choanoflagellate Monosiga brevicollis

López-Escardó D, Grau-Bové X, Guillaumet-Adkins A, Gut M, Sieracki ME, Ruiz-Trillo I. Evaluation of single-cell genomics to address evolutionary questions using three SAGs of the choanoflagellate Monosiga brevicollis. Sci Rep. 2017 Dec 8;7(1):11025. DOI: 10.1038/s41598-017-11466-9

3.4. Phylogenomics reshapes choanoflagellate evolution and reveals new insights into the pre-metazoan genetic tool-kit

López-Escardó D, Grau-Bové X, Guillaumet-Adkins A, Gut M, Sieracki M, Ruiz-Trillo, I. Phylogenomics reshapes choanoflagellate evolution and reveals new insights into the pre-metazoan genetic tool-kit. Unpublished.

Phylogenomics reshapes choanoflagellate evolution and reveals new insights into the pre-metazoan genetic tool-kit.

David López-Escardó[1*], Xavier Grau-Bové[1,2] Amy Guillaumet-Adkins[3,4], Marta Gut[3,4], Michael E. Sieracki[5] & Iñaki Ruiz-Trillo[1,2,6*]

[1]Institut de Biologia Evolutiva (CSIC-Universitat Pompeu Fabra), Passeig Marítim de la Barceloneta 37-49, 08003 Barcelona, Catalonia, Spain.

[2]Departament de Genètica, Microbiologia i Estadística, Universitat de Barcelona, Barcelona, Catalonia, Spain.

[3]CNAG-CRG, Centre for Genomic Regulation (CRG), Barcelona Institute of Science and Technology (BIST), Barcelona, Spain.

[4]Universitat Pompeu Fabra (UPF), Barcelona, Spain.

[5]National Science Foundation; Arlington, VA, USA

[6] ICREA, Pg. Lluís Companys 23, 08010 Barcelona.

Abstract

Choanoflagellates are the closest unicellular relatives of animals. Therefore, the internal phylogeny of choanoflagellates, as well as their genomic content are crucial to better understand the early origins of animal multicellularity. So far, there are only two choanoflagellate taxa with whole genome sequences (*Monosiga brevicollis* and *Salpingoeca rosetta)*, representing a narrow fraction of choanoflagellates diversity. In this work, we have expanded the available genomic information of choanoflagellates by sequencing four single-cell amplified genomes (SAGs) collected during the TARA Oceans expedition. The SAGs were chosen to expand phylogenetically our previous knowledge, with one of them being and early-branching acanthoecid and the third most abundant choanoflagellate in TARA Oceans. This SAG and an early-branching clade 1 craspedidan were complete enough to be used in our phylogenomics analysis. Our newly updated choanoflagellate tree, that includes these new SAGs and all the available transcriptomic data of choanoflagellates, breaks the monophyly of Craspedida and establishes *Codosiga hollandica* as the earliest-branching choanoflagellate. This suggests a non-thecated colonial and freshwater ancestor of choanoflagellates, opening new hypotheses regarding the ecological context in which the ancestors of choanoflagellates and animals could have emerged. Finally, a comparative genomic analysis revealed a pre-metazoan origin of protein domains that are involved in the organization of animal multicellularity, such as transcription factors related to development (Nucleophosmin and Smad), protein domains related to immunological and sperm functions (IRF and TILa respectively), and genes that expand the neural pre-metazoan toolkit (NKAIN and Praxilin). Overall, our new choanoflagellate genomes have provided a new phylogenetic tree of this group, as well as expanded the list of genes and protein domains with a pre-metazoan origin.

Introduction

Choanoflagellates are a well-known group of unicellular eukaryotes that have long been associated with animals due to their apparent cytological similarities with choanocytes, a cell type of sponges (James-Clark 1866). The position of choanoflagellates as the closest unicellular relatives of animals was later confirmed with molecular phylogenies (Wainright et al. 1993; Zettler LAA et al. 2001; Lang et al. 2002; Steenkamp, Wright, and Baldauf 2006; Ruiz-Trillo et al. 2008).

Choanoflagellates is a diverse protist group, with approximately 250 described species (M Carr et al. 2008; B.S.C Leadbeater 2015). They are aquatic heterotrophic organisms, either marine or freshwater, being important players in microbial food webs. Morphologically, choanoflagellates are characterized by an ovoid to spherical cell body containing a single anterior flagella surrounded by a collar of microvilli (B.S.C Leadbeater 2015).

Molecular phylogenies based in a few genes showed that choanoflagellates are divided in two major clades, known as Craspedida and Acanthoecida (Martin Carr et al. 2008; Paps et al. 2013; Martin Carr et al. 2017). Craspedida includes the choanoflagellates with organic coverings, that can be thecated (Salpingoecidae morphology); or non-thecated with non-restrictive coverings like glycocalix or sheath (Codosigidae morphology) (B.S.C Leadbeater 2015). On the other hand, Acanthoecida is composed by choanoflagellates with a siliceous loricae, being most of the described species marine and with a tectiform lorica (around 150 species) (Martin Carr et al. 2008), although there are 5-6 species described with nudiform lorica (B.S.C Leadbeater 2015). Furthermore, there other clades of choanoflagellates, such as Clade L (Weber et al. 2012), FRESCHOs and MACHOs (del Campo and Ruiz-Trillo 2013), that had been defined by environmental sequences of the 18S rDNA gene. Thus, choanoflagellates are a rich group of single-celled organisms with a great variety of cell morphologies.

Given their phylogenetic position as sister-group to animals, analyses of their gene can provide important insights into animal origins. Indeed, the genomes of the two so-far sequenced choanoflagellate taxa (King et al. 2008; Fairclough et al. 2013) already showed that choanoflagellates already had some genes key for animal multicellularity, involved in processes such as cell-to-cell signalling, cell adhesion or animal development (King et al. 2008; Fairclough et al. 2013; Richter and King 213, Sebé-Pedrós et al. 2017). Interestingly, however, the genomes of other more ancient lineages of protists closely related to animals, encoded genes key to animal multicellularity that were potentially secondarily lost in choanoflagellates (Sebé-Pedrós et al. 2011 MBE; Sebé-Pedrós et al. PNAS 2010; Suga et al. 2013; Sebé-Pedrós et al. 2017). These findings emphasized the importance of taxon sampling when reconstruction the genome content of ancestral nodes.

In this regard, the two choanoflagellate genomes so far sequenced belong to the same craspedidan clade 1 (Martin Carr et al. 2017). This represents a narrow diversity of choanoflagellates with genomes available, even though the recently sequenced transcriptomes of 19 other choanoflagellates (Martin Carr et al. 2017; Simion et al. 2017) should help to have a more wider representation of the choanoflagellate genetic landscape.

To wider our genomic knowledge of the choanoflagellates, we here sequenced, assembled and annotated four single-cell amplified genomes belonging to distinct taxa collected during the TARA Oceans expedition (Tara Oceans Consortium, Coordinators; Tara Oceans Expedition 2014). With this new molecular data, we have re-defined the choanoflagellate phylogeny , basically breaking the monophyly of Craspedida, which opens new perspectives on the ecology of early choanoflagellates. Finally, we have also further expanded the list of genes already present in the unicellular ancestor of animals, with the finding of some genes that were previously thought to be animal-specific in the genomes of these uncultured choanoflagellates.

Results and discussion

Expanding the genomic diversity of choanoflagellates

We sequenced four single-cell amplified genomes corresponding to uncultured choanoflagellates cells collected during the TARA oceans expedition (Vargas et al. 2015) (see Table S1 for collection environmental details). The four cells belonged to different choanoflagellate taxa and they do not appear related with any previous described species with transcriptomic or genomic information available (Fig.1).

In particular, and to place the different SAGs within the choanoflagellates, we first performed a phylogeny of the 18S ribosomal subunit that included the SAGs and the known 18S molecular diversity of unicellular holozoans, including environmental sequences (del Campo and Ruiz-Trillo 2013). UC1 appears as an early-branching clade 1 craspedidan that groups with *Lagenoeca antartica* (Nitsche et al. 2007) (Fig.1). Its 18S sequence is completely identical to the environmental NCBI sequence AY426842 (100% of pairwise identity). UC2 forms a monophyletic clade with the rest of Acanthoecidae (nudiform loricates) (Fig.1), and UC3 clusters with the tectiform loricates *Stephanoeca paucicostata* and *Stephanoeca cauliculata*. Finally, UC4 falls as the earliest-branching acanthoecid, together with the environmental sequence JQ223245. Thus, the four cells belonged to different choanoflagellate taxa and were not related to any previously described species (Fig.1), thus expanding the genomic information so far available for choanoflagellates.

Metabarcoding data from TARA Oceans database (de Vargas et al. 2015), allowed us to reveal the geographical distribution of these uncultured choanoflagellates. The craspedidan UC1 is specially present in Mediterranean samples, although not exclusively (Fig. 2A). Interestingly, the environmental sequence AY426842 was also sampled in the Mediterranean (Massana et al. 2004), therefore, it can be an interesting area for future attempts of isolating this species. The basal acanthoecid UC4 is the third most abundant choanoflagellate in TARA Oceans, and it is of cosmopolitan distribution (46 samples out of 47) (Fig. 2A). The nudiform UC2 and the tectiform UC3 are also widely distributed (45 samples out of 47), albeit less abundant than UC4 (Fig. 2B). Since most of the TARA Oceans reads associated to our SAGs appear in the picoplanktonic fraction (Fig. 2B) our SAGs' cell size likely ranges between 0.8 and 5mm – in agreement with the typical size range of described choanoflagellate species (B.S.C Leadbeater 2015). Furthermore, our four SAGs are relatively more abundant in surface waters than in deeper sampling points such as the depth chlorophyll maximum (Fig. 2B).

Genome completeness and statistics of the SAGs

Once deciphered the taxonomy and the ecological distributions of our SAGs, we sequenced their genome by MiSeq Illumina 2X250pb as in Lopez-Escardó 2017. We then checked the genome recovery and the genome statistics of our final assemblies. Unfortunately, single-cell genomics is a technique that leads fragmented genomes and sometimes with extreme low genome completeness(López-Escardó et al. 2017). UC1 and UC4 presented a significant genome recovery (7.74 MB and 31.68% BUSCO for UC1; 7.25 Mb and 13.53% of BUSCO for UC4) (Table 1, Supplementary Table 2).

However, SAGs from UC2 and UC3 were largely incomplete and fragmented (Table 1), and for this reason were not included in most of the subsequent analyses, except the eight-gene based phylogeny (Fig. S1). Interestingly, we were able to recover the mitochondrial genome of UC2 (Table 1), which is the first available mitochondrial genome of an acanthoecid choanoflagellate. We could annotate 59 mitochondrial genes (Supplementary Table 3) that revealed a high degree of conservation with the mitochondrial genome of *M. brevicollis* (J. Yang et al. 2017).

Thanks to the predictions of core eukaryotic genes Busco and CEGMA, we could extrapolate the real genome size of the SAGs UC1 and UC4. The craspedidan UC1 (29.4 Mb, see Table 2) would potentially contain the smallest genome among the so far sequenced choanoflagellates; *S. rosetta* (55.4 Mb) and *M. brevicollis* (41.6 Mb). The predicted genome size of the UC4 Acanthoecida (52.5 Mb) is similar than the size of *S. rosetta*.

Next, we predicted the number of genes that might contain the full genomic sequences of UC1 and UC4 taxa, by extrapolating the numbers of genes annotated with

the Busco/CEGMA values, removing the potential contamination and taking into account pfam protein domain predictions (see methods). The difference in estimated size is proportional to the number of estimated genes. UC1 has less number of genes (6,039) according to the predicted reduced genome size and UC4 would present a similar number of genes than the previous choanoflagellate genomes (10,075) (Table 2).

Finally, we characterized our SAGs by screening for genes linked to morphological structures, such as the microvilli or the lorica of acanthoecids, in order to speculate among the potential morphology of these two choanoflagellate taxa. Therefore, we searched for the presence of Ezrin/Radixin/Moesin (ERM) protein (Sebé-Pedrós et al. 2013; Peña et al. 2016), which is known to be involved in microvilli elongation processes; as well as for the presence of Si transporters (SITs)(Marron et al. 2016), needed for the lorica formation in Acanthoecida. The microvilli-related ERM protein was only found in the UC4 genome, and not detected in UC1. We also failed to detect any SIT in any of those taxa, including the early branching loricate UC4 (Marron et al. 2016). This could suggest that either those taxa do not have lorica. However, as we have partial genomes, a negative result cannot be considered as evidence of gene absence.

Phylogenomics reshapes the phylogeny of choanoflagellates

We built a phylogenomic matrix based on 87 single-copy proteins domains over 79 taxa including animals and all its unicellular relatives with available transcriptomic or genomic data, including our SAGs, the recently described genera *Pigoraptor*, *Syssomonas* (Hehenberger et al. 2017) and *Chromospharea* (Grau-Bové et al. 2017)] (Fig. 3), as well as the transcriptomes from 19 choanoflagellates (Simion et al. 2017; Martin Carr et al. 2017). In addition, we included an extensive outgroup composed by holomycotans (18 taxa), apusomonads (2 taxa), breviates (3 taxa) and amoebozoans (4 taxa).

Our results recovered the Teretosporea monophyly (Torruella et al. 2015) with a high statistical support , with 99% of ultra-fast bootstrap from maximum likelihood (UFBS) and 1 of posterior probability of Bayesian inference (BI) (Fig 3). Apparently, the addition of *Syssomonas* and *Chromosphaera* together in the same phylogeny allows better statistical supports than obtained in previous studies.

Our tree also recovers monophyly for choanoflagellates with maximum support. Our SAGs UC1 and UC4 were confidently placed within the choanoflagellates. UC1 it is confirmed to be a clade 1 crapedidan, as in the 18S rRNA tree (Figure 1), appearing as sister-group to the previous described craspedidan clade 1 (Martin Carr et al. 2017) and UC4 it is confirmed

to be the earliest-branching acanthoecid so far described. However, and somehow unexpectedly, our tree recovered some important topological differences compared to previous choanoflagellate phylogenies that were based on a few genes (M Carr et al. 2008; Martin Carr et al. 2017).

The first incongruence between our topology and the previously defined is that we recovered a paraphyletic clade 2 of craspedidans. In particular, *Salpingoeca urceolata* and *Salpingoeca kvevrii*, appear sister to clade 1 and not within clade 2 as previously described (Martin Carr et al. 2017). The same topology is obtained without UC1 and UC4 (Fig. S1). Therefore our results redefine these two groups of craspedidans (Fig. 3). On the other hand, our data shows that nudiforms cluster within tectiforms, meaning that they are clearly not two independent lineages within Acanthoecida. These two situations are highly supported in our phylogeny (*Salpingoeca urceolata* and *Salpingoeca kvevrii* sister to clade 1 of craspedidans with a node support of 94% ML UFBS and 1 pp BI, and maximum support in all Achantoecida nodes). However, in deeper nodes the supports are not that high (Fig. 3). For example, *Salpingoeca dolicothecata* appears as sister-group to Craspedida and Acanthoecida. However its position is weakly supported and, the Maximum likelihood (ML) topology places the species as early branching within Acanthoecida (Fig.S2). *S. dolicothecata* together with *S. tuba* had been described as members of the craspedidans clade 3 (Martin Carr et al. 2017). The lack of taxon sampling in our analysis is likely the responsible for the incongruences between the different phylogenetic positions for *S. dolicothecata* (ML and Bayes). Interestingly, however, when we built an 8 house-keeping genes phylogeny based on Carr and co-workers (Martin Carr et al. 2017), *S. dolicothecata* and *S. tuba* appeared as sister-group to Craspedida and Acanthoecida in the Bayesian inference topology (Fig. S3). Thus, we can interpret that both *S.dolicothecata* and *S. tuba* have some attraction to the root of choanoflagellates. In the ML inference the topology recovered is the same than in (Martin Carr et al. 2017) (Supplementary material).

Finally, one of the most surprising results is that we recovered *Codosiga hollandica* as the sister-group to the rest of the choanoflagellates prior to the split of *S. dolicothecata* and the divergence of Craspedida and Acanthoecida. This basal position for *Codosiga hollandica* is highly supported (75% UFBS ML / 0.99 pp BI), specially when we removed the faster-evolving sites (Fig. S4). Interestingly, this same topology with *C. hollandica* as the sister-group to the rest of choanoflagellates was obtained in the recent phylogenomic analysis of Simion and co-workers, which aimed to reconstruct animal phylogeny using choanoflagellates as an outgroup (Simion et al. 2017). When we reanalysed the dataset of Simion et al. with a reduced sampling of animals to maximise the resolution at the internal choanoflagellate nodes, we recovered the same early-branching position of *C. hollandica* (Fig.S5). Therefore, we suggest that *C.*

hollandica might be the sister-group to the rest of choanoflagellates, and that craspedid species are paraphyletic. To confirm this topology, however, future studies would need to consider the addition of more taxa and the construction of choanoflagellate-specific phylogenomic dataset. For example, the inclusion of species from *Sphaeroeca* genus, which is related to *Codosiga* in the 18S rRNA (Fig. 1) and in the eight-gene phylogenies (Fig. S3), would surely help to increment the statistical support and get a more reliable phylogeny. Other clades, like the environmental described Clade L (Weber et al. 2012) or FRESCHO3-4 groups (del Campo and Ruiz-Trillo 2013) would be as well key to break ancestral nodes and get a more complete topology, given that they fall in early branching positions in the 18S rRNA phylogeny.

Our new phylogenetic framework for the choanoflagellates have some important implications with regards the evolution of choanoflagellates. If we plot the morphological and environmental characters over the new topology (Fig. 4), the earliest-branching choanoflagellate lineage (*C. hollandica*) is a non-thecate, fresh-water, species with a peduncle to attach to the surface (codosigid morphology) and would be able to form colonies. This result is in disagreement with the idea of a marine choanoflagellate ancestor (M Carr et al. 2008; Martin Carr et al. 2017). Our results rather suggest that the first choanoflagellate had colony-forming capabilities and could be either marine or fresh-water and probably with codosigid morphology. The subsequent evolution of choanoflagellates led to the development of different theca structures including the lorica.

Finally, our results could have larger implications with regards animal origins. For example, if the first choanoflagellate was living in a fresh-water environment, and considering that most of the filasterean species (the sister-group to choanoflagellates and animals) were isolated from freshwater environments (Stibbs et al. 1979; Hehenberger et al. 2017), this could indicate that the ancestor of animals and choanoflagellates might had been living in freshwaters environments. If so, one could speculate that the origin of animals could as well, had taken place in a fresh-water environment. In any case, more data is needed to confirm this scenario.

Updating the pre-metazoan genetic tool-kit

Given that the SCG annotation can potentially detect the presence of a considerable percentage of protein domains than (López-Escardó et al. 2017), we decided to check our choanoflagellate SAGs in order to update the genetic repertoire of the unicellular ancestor of animals. To perform this analysis we built a database of 116 proteomes of different eukaryotic taxa (Supplementary Table 4), representing the entire eukaryotic diversity. We then predicted the protein domain architectures and produced a matrix of presence/absence of each of protein domain across all the eukaryotic taxa (see methods). We then

inferred the gains and loses among each node of the eukaryotic tree of life (summarized in Fig.4 for Opisthokonts).

Our results show that there was an important acquisition of protein domains at the stem of opisthokonts and that choanoflagellates are the unicellular holozoan clade that have lost less protein domains after diverge from the evolutinoary path that lead to animals (Grau-Bové et al. 2017) (Fig. 4). Our new dataset allow us to have a detailed view of the protein domains that originated at the Choanozoa (choanoflagellates + Metazoa(Brunet and King 2017)). We then, screened the literature to link each protein domain gained at Choanozoa with a biochemical role or biological process. The results are depicted at Figure 4B and show that most of the protein domains in which it is known a biochemical role belong to transcription factors or epigenetic regulators, such as the protein domains MH1 and zf-C4, which until now were described as metazoan specific (A. de Mendoza et al. 2013). MH1 is the DNA binding protein of Smad transcription factors that together with MH2 conform the canonical Smad proteins (Attisano and Lee-hoeflich 2001). Both, have a choanozoan origin even though the canonical Smad architecture remains metazoan-specific, as previously described (de Mendoza et al. 2013; Sebé-Pedrós, Degnan, and Ruiz-Trillo 2017). Thus, Smad proteins are a product of a domain shufling event at the stem of metazoa, as suggested for Notch and Hedgehog proteins (King et al. 2008; Fairclough et al. 2013).

Development is the biological process in which most of the proteins domains gained at Choanozoa are involved (Fig. 4B). This is logical given that development requires many other biological processes from control of cellular growth to cell signaling. There are, however, protein domains acting in transcription factors that are related to other biological functions rather than development. Such is the case, for example, of IRF, an interferon regulatory factor that binds to interferon and to DNA, and controls the expression of genes related in immune response (Weisz et al. 1992). Finally, we also identified a protein domain related in sperm function: TILa, a cysteine rich domain that binds specifically to egg extracellular matrix (Hardy and Garbers 1995).

Therefore, at the origin of animals and choanoflagellates appeared protein domains that were involved in crucial functions to maintain animal multicellularity. Those protein domains were involved not only in development, cell-to-cell signaling or adhesion, but also in other multicellular functions (Richter and King 2013), such as neural functions, immunologic response (King et al. 2008; Fairclough et al. 2013), cell cycle control, or the control of cell polarity and division. It is worth mentioning that these results are based on the taxon sampling used, and that future analyses with an extended taxon sampling (specially from filastereans, in which only 4 taxa are sampled) may change the origin of some of those protein domains.

Those "multicellular" protein domains are more retaind in animals species than in choanoflagellates (Fig. 4), however one may wonder whether these protein domains are involved in colony formation in choanoflagellates . In this regard, it is interesting to see that colonial choanoflagellates do not particularly have kept more (16.5 in average) of those protein domains related in multicellular functions than non-colonial taxa (18 in average). This may suggest that the molecular mechanisms involved in animal and choanoflagellates multicellularity, might require different protein players. Furthermore, we found around 30 proteins domains acquired in the origin of animals and choanoflagellates with undescribed functions (Fig.4B). These domains might play important roles in extant animal taxa, being interesting candidates for functional studies in Metazoa.

Among the protein domains that originated in Choanozoa, 9 protein domains were recovered in our SAGs and also most of them in other chonaflagellate taxa (Fig.4A). Among them, it is specially relevant the C-terminal protein domain of the transcription factor Nucleophosmin, named NPM1-C, and present in the SAG UC4 (Fig. 4A). This was confirmed by pfam domain analysis and by phylogenetic inferences (Fig. S6). Nucleophosmin is a transcription factor thought to be specific to vertebrates (Box et al. 2016). Nucleophosmin is a transcription factor, key in the regulation of DNA replication malfunctions, and it is involved in p53 mediated pathways to promote apoptosis in case of DNA damage (Box et al. 2016). Nucleophosmin is composed by a Nucleoplasmin protein domain followed by NPM1-C in all animals. The domain Nucleoplasmin is paneukaryotic, and the architecture Nucleoplasmin plus NPM1-C is animal specific according to our results. In the SAG UC4 we identified the domain NPM1-C, while the Nucleoplasmin domain was missing. As SAGs genomes are partial, we can not rule out that the Nucleoplasmin domain is indeed present in the SAG. However, we believe the most likely explanation is that the animal Nucleophosmin, as Smad proteins, appeared as a product of a domain shuffling between the two more ancient Nucleophosmin domains (Nucleoplasmin and NMP1-C) as had been suggested for other animal proteins like Notch (King et al. 2008).

Another protein domain linked to a transcription factor found in our SAGs is the domain Fanconi_A_N (Fig. 4). It is the N-terminal domain of the Fanconi anemia complementation grup A protein (FANCA human protein) that acts in DNA damage-repair processes and also in the differentiation of blood cells(de Winter and Joenje 2009). Mutations in these gene causes the Fanconi anemia (FA) in humans (de Winter and Joenje 2009). Our data show a pre-metazoan origin for the Fanconi_A_N domain, and a vertebrate acquisition for the Fanconi_A domain that, together with Fanconi_A_N, conform the canonical FANCA protein.

Our SAGs also contain two protein domains involved in extracellular

receptors or transmembrane proteins that participate in neural functions in animals. These are NKAIN, a sodium dependent ATPase interacting protein (Gorokhova et al. 2007) and Sema, which conform Plexin proteins, both found to be encoded in the UC1 SAG. Plexin proteins are neural semaphorin receptors that guide axon formation in neural development (Winberg et al. 1998). The UC1 Plexin contained the same protein domain architecture than in vertebrates, which implies that the full protein was already present in the unicellular ancestor of animals. Therefore, Plexin and NKAIN are components of the pre-metazoan genetic toolkit of genes related to neural functions together with the already described sodium (Liebeskind 2011) and calcium channels (Cai 2008), Neuroglobulins (Lechauve et al. 2013) and proteins related in synapsis (Alié et al. 2011; Fairclough et al. 2013), postsynaptic functions like homer(Burkhardt et al. 2014) and neural secretion (Burkhardt et al. 2011).

Overall, our data have allowed us to have a more detailed view of the gene or protein domain content of the unicellular ancestor of animals. We have further expanded the list of genes and protein domains present in the ancestor, including some that were thought to be animal-specific. This demonstrates the importance of having a good taxon sampling when inferring ancestral states.

Conclusions

We expanded the choanoflagellate genomic information available thanks to single-cell genomics from environmental cells. In particular, we recovered meaningful information from two taxa, the UC1, a clade 1 craspedidan and UC4, an early branching acanthoecid, which is also the third most abundant choanoflagellate from TARA Oceans database. We could recover as well the first mitochondrial sequence of an Acanthoecida thanks to the SAG UC2.

Our phylogenomics analysis re-shape the phylogeny of the choanoflagellates. Our results break the monophyly of Craspedidans and bring to the earliest branching position of choanoflagellates the species *Codosiga hollandica*. This suggests a non-thecated and freshwater ancestor of choanoflagellates, opening new hypothesis among the ecological context in which choanoflagellates and animals could have emerged.

Finally, our comparative genomics show that most of the protein domains related to multicellular functions and innovated at the Choanozoa, were retained in animals, and less in choanoflagellates. Our data was also key to better define the protein domain composition of the unicellular ancestor of animals, that now includes some additional protein domains, previously thought to be animal-specific. Our results show that new genomic data is still needed to clarify the evolutionary history of animal genes and understand the genetic content present in pre-metazoan lineages.

Methods

Cell collection and whole genome amplification

Cells for single-cell genomics were collected from the Mediterranean sea and different places of Indian Ocean during the Tara Oceans expedition (Karsenti et al. 2011) and cryopreserved as described before (Heywood et al. 2011). Flow cytometry cell sorting, single cell lysis and whole genome amplification by Multiple Displacement Amplification (MDA) (Dean et al. 2002) were performed at Bigelow Single-cell genomics facility (Boothbay, Maine US), as previously described (Stepanauskas and Sieracki 2007; Martinez-Garcia et al. 2012; Mangot et al. 2017) (Table. S1). The SAGs obtained were screened by PCR using universal eukaryotic 18S 350 rDNA primers (Mangot et al. 2017). The 4 SAGs were placed in interesting phylogenetic positions (Fig. 1). Associated environmental data is summarized in Supplementary Table 1 and more details can be found in PANGAEA (Tara Oceans Consortium, Coordinators; Tara Oceans Expedition 2014; Pesant et al. 2015).

Library preparation and genome sequencing

Four SAGs (UC1, UC2, UC3 and UC4) were sent for sequencing at CNAG (Barcelona, Spain). The libraries were constructed with the TruSeq Nano DNA Library Preparation Kit according to manufactures protocol. Briefly, aiming for an insert size of 550 bp, 200ng of gDNA was sheared by sonication using Covaris E210 (Covaris). Fragmented DNA was purified with Agencourt AMPure XP beads. Afterwards, end repair and size selection were performed, following 3'adenylation reaction and ligation of the Illumina adapter indexes. DNA fragments were enriched by 8 cycles of PCR, and then purified with Agencourt AMPure XP beads. The Agilent Technologies 2100 Bioanalyzer DNA 1000 assay was used for library quality control and quantification.

Each library was sequenced using one lane of MiSeq reagent kit v2 (Illumina). The sequencing run was performed according to standard Illumina operation procedures in paired-end mode, with a read length of 2x251bp and the yield of >11Gb. Primary data analysis, the image analysis, base calling and quality scoring of the run, was processed using the manufacturer's software Real Time Analysis (RTA 1.18.54) and followed by generation of FASTQ sequence files by CASAVA. The reads obtained were used to perform a downsampling analysis as described in (López-Escardó et al. 2017), UC1 and UC4 presented longer assemblies and the curve was still not saturated. Thus, we decided to apply more sequencing depth for them. At the end, UC1 and UC4 were sequenced in 3 Miseq lanes with a yield of >34Gb.

Genome assembly and annotation

Raw reads obtained were trimmed with Trimmomatic v3.0 (Bolger, Lohse, and Usadel 2014) using the following options:

ILLUMINACLIP:/adapters/NexteraPE-PE.fa:2:40:15 HEADCROP:10 CROP:240 SLIDINGWINDOW:6:20 MINLEN:50. A range between 42-45 million reads were obtained from the low quality SAGs UC2 and UC3, and for the SAGs with more sequencing applied, UC1 and UC4, the range moves between 110-125 millions reads (Supplementary Table S2). Next, we performed the genome assembly with SPAdes v3.6.1 (Bankevich et al. 2012) with the options --sc --careful and -k 21,33,55,77,99. The final genome statistics were obtained with QUAST (Gurevich et al. 2013). The percentage of core eukaryotic conserved proteins was calculated with CEGMA (Parra, Bradnam, and Korf 2007) and BUSCO (Simão et al. 2015). We screened for mitochondrial genomic sequences in our SAGs by performing a tBLASTn v.2.2.31+ (Camacho et al. 2009) using as query the mitochondrial proteins of *Andalucia godoy* (Burger et al. 2013). Only the SAG UC2 presented three scaffolds with mitochondrial proteins. We mapped SAG UC2 reads over the scaffolds selected using the program Bowtie2 v.21.0 (Langmead and Salzberg 2012), in order to perform a re-assembly with SPAdes v3.6.1 (Bankevich et al. 2012), and see if we could recover in one scaffold the full mitochondrial genomic sequence. The assembly yielded a more fragmented output, and the two of the previous selected scaffolds contained proteins that came from bacterial contamination. Thus, we decided to keep the first scaffolds obtained (32 Kb length). UC2 partial mitochondrial genome is available at NCBI (accession number XXXX). We annotated the mitochondrial genes with Mfannot (Beck and Lang 2010) and are available at Supplementary Table 3.

As the genome completeness of the SAGs UC2 and UC3 was very low, we decided to continue the genome annotation only for the SAGs UC1 and UC4. We annotated the genome with Augustus (Stanke and Morgenstern 2005) trained with CEGMA proteins (Parra, Bradnam, and Korf 2007) as explained in (López-Escardó et al. 2017). To predict the number of genes that may contain the full genome sequence of UC1 and UC4, we first performed a BLASTp v.2.2.31+ (Camacho et al. 2009) using as a query our predicted proteins against a database that includes all non redundant proteins from Uniprot (Wasmuth and Lima 2016), in order to identify the potential contaminant proteins. We removed the proteins that had the first hit from a bacterial or archaeal origin. However, as they were many genes without blast match and the annotation process can overpredict gene content (López-Escardó et al. 2017), we decided to take into account only proteins in which, thanks to Pfam scan, we could find protein domains. The proteomes of *M. brevicollis* and *S. rosetta* contain approximately 70% of their proteins with a described Pfam protein domain. Therefore, we took this fact into consideration in our calculations together with the average between the genomic completeness obtained by Busco and CEGMA, to infer the total number of proteins that UC1 and UC4 may have in their complete genomes. SAGs assembly and annotation is available at Figshare

(XXXX), including the list of protein classification.

Ecological distribution of our SAGs

We performed a BLASTn v.2.2.31 (Camacho et al. 2009) using as query the 18S sequences of our SAGs against the OTUs from TARA oceans database (de Vargas et al. 2015). We found 4 OTUs that correspond to our SAGs with 100% or 99.2% identity (only one mismatch) (Supplementary Table 5). We plotted the read distribution according to geographical locations using R (R Core Team 2013).

18S ribosomal gene phylogeny

We collected 18S rDNA ribosomal sequences from representatives of all known 18S rDNA molecular diversity of unicellular holozoans, including the uncultured lineages Clade L (Weber et al. 2012), FRESCHOs, MACHO and MAOPs (del Campo and Ruiz-Trillo 2013) (Supplementary Table 6). We ended up with a dataset of 117 18S rDNA sequences. Next, we aligned them using MAFFT (Katoh et al. 2002) with the E-INS-i algorithm. After manually trimming sequence ends, indels and spuriously aligned sites we ended up with a total of 1,754 sites. We inferred phylogenetic trees from this alignment using a Maximum Likelihood (ML) inference. The best substitution model for phylogenetic inference was selected using IQ-TREE (Nguyen et al. 2015), using the TESTNEW model selection procedure and following the BIC criterion. In all four cases, the GTR substitution matrix with a 5-categories free-rate distribution (Z. Yang 1995) (a modification of the standard Γ distribution) was selected as the best-fitting model. Maximum likelihood inferences were performed with IQ-TREE, and statistical supports were drawn from 1,000 ultrafast bootstrap values with a 0.99 minimum correlation as convergence criterion (Minh, Nguyen, and Von Haeseler 2013), and 1,000 replicates of the SHlike approximate likelihood ratio test (Guindon et al. 2010). The phylogenetic tree in nexus file and the Alignments before and after the trimming are available at Figshare (XXXX).

Eight-gene phylogeny

Similar to the recent published choanoflagellates phylogeny (Martin Carr et al. 2017), we built a phylogenetic matrix with the nucleotidic sequences of eight house-keeping genes, to infer the choanoflagellates phylogeny with a wider diversity than our phylogenomics approach. The genes used are the ribosomal SSU (18S) and LSU (28S) genes, actin, beta tubulin, hsp90, hsp70, EF and EF1A. In the supplementary table 7 is summarized the presence of each gene in each taxa. All the sequences used in the analysis are available at Figshare. The analysis was performed over 66 taxa, 57 of them being choanoflagellates. To build the final matrix, we aligned each gene separately with MAFFT (Katoh et al. 2002) using E-INS-i algorithm, and next we trimmed manually the spurious positions. Finally, we concatenated the trimmed alingments for each gene, building a phylogenetic matrix composed by 12,884 nucleotide

positions. To run the phylogenetic analysis we partitioned in three parts our dataset, to run in each an evolution model with different rate distributions separating: the ribosomal genes (partition 1), the 1st and 2nd codon positions of the non-ribosomal genes (partition 2), and the third codon position of the non-ribosomal genes (partition 3). The best substitution model for each partition was selected, again, using IQ-TREE (Nguyen et al. 2015), with the TESTNEW model selection procedure and following the BIC criterion. The Maximum Likelihood analysis run with GTR substitution matrix with a 5-categories free-rate distribution(Z. Yang 1995) (a modification of the standard Γ distribution) was selected as the best-fitting model in the partition 1, with 3-categories in the partition 2 and with 4-categories in the partition 3. Statistical supports were drawn from 1,000 ultrafast bootstrap values with a 0.99 minimum correlation as convergence criterion(Minh, Nguyen, and Von Haeseler 2013). Bayesian inference was performed with MrBayes 3.2.6 (Ronquist and Huelsenbeck 2003) using the GTR+ Γ model of nucleotide substitution in all partitions, running at different distribution according to the model given by IQ tree (Γ5, Γ3, Γ4 respectively for each partition). Four chains ran for 4,400,000 generations and were analyzed after a burn-in of 25%. The trimmed concatenated alignment, the partition information and the phylogenetic trees from ML and BI are available at Fighsare (XXX).

Phylogenomic analysis of Amorphea using 87 single-copy protein domains

We updated the phylogenomic dataset developed in (Torruella et al. 2015; Grau-Bové et al. 2017) , consisting of 87 single-copy protein domains from 57 amorphean taxa, with new data from SAGs UC1 and UC4. We used a custom script (Torruella et al. 2015) which uses tBLASTn alignments (Camacho et al. 2009) to search protein domains over the assembled genome. We recovered 32 and 20 proteins domains for the SAGs UC1 and UC4 respectively, which accounted for 6,844 and 6,132 ungapped positions out of 22,201 ungapped positions of the consensus sequences of the final alignment. The final alignment contained 23,364 amino acid positions. In addition, we included new transcriptomes from 19 choanoflagellate taxa (Martin Carr et al. 2017; Simion et al. 2017), plus three species from the recently described holozoan genera *Pigoraptor* and *Syssomonas* (Hehenberger et al. 2017). Most of the protein domains were found in each of these new added taxa.

We built ML phylogenetic trees using IQ-TREE v1.5.1, under the LG model with a 7-categories free-rate distribution, and a frequency mixture model with 60 frequency component profiles based on CAT (LG+R7+C60) (Minh, Nguyen, and Von Haeseler 2013). LG+R7 was selected as the best-fitting model according to the IQ-TREE TESTNEW algorithm as per the Bayesian information criterion (BIC), and the C60 CAT approximation was added because of its higher rate of true topology inference (Quang, Gascuel, and Lartillot 2008). Statistical supports were drawn from 1,000 ultrafast

bootstrap values with a 0.99 minimum correlation as convergence criterion (Minh, Nguyen, and Von Haeseler 2013), and 1,000 replicates of the SHlike approximate likelihood ratio test (Guindon et al. 2010).

The same alignment was used to build a Bayesian inference tree with Phylobayes MPI 755 v1.5, using the LG exchange rate matrix with a 7-categories gamma distribution and the non-parametric CAT model (LG+Γ7+CAT) (Lartillot and Philippe 2004). A Γ7 distribution was considered to be the closest approximation to the free-rates R7 distribution of the IQ-TREE ML analysis (as free-rates distributions are not implemented in Phylobayes). We removed constant sites to reduce computation time. We ran two independent chains for 5660 and 5685 generations, respectively, until convergence was achieved (maximum discrepancy = 0.0851376) with a burn-in value of 13% (739 burnt-in trees). The adequate burn-in value was selected by sequentially increasing the number of burn-in trees, until the the maximum discrepancy statistic reached the <0.01 threshold, and 2) the highest effective size for the log-likelihood parameter. The bpcomp analysis of the sampled trees yielded a maximum discrepancy = 0.0851376 and a mean discrepancy = 0.00130004. The tracecomp parameter analysis gave a minimum effective size for the log-likelihood parameter = 4. The trimmed alignment, and the phylogenetic trees from ML and BI analysis are available at Figshare (XXXX).

Phylogenomic analysis of choanoflagellates using 1719 gene markers from Simion et al.

We performed a second phylogenomic analysis using a subset of holozoan from the gene marker-rich alignment matrix analysed in (Simion et al. 2017),, which included 97 holozoan species and 1719 gene markers and 401,632 alignment positions. We retrieved the alignments corresponding to *M. brevicollis*, *S. rosetta*, the 19 choanoflagellate transcriptomes, 10 representative animal genomes and transcriptomes (*Alatina alata, Nematostella vectensis, Branchiostoma floridae, Saccoglossus kowalevskii, Trichoplax adhaerens, Oscarella* sp., *Plakina jani*, *Clathrina coriacea, Amphimedon queenslandica* and *Kirkpatrickia variolosa*), and 2 filastereans (*Capsaspora owczarzaki* and *Ministeria vibrans*). A reciprocal-best-BLAST-hit search for orthologs of the 1719 gene markers in UC1 and UC4 retrieved no putative hits. Therefore, UC1 and UC4 were not included in this analysis. The resulting alignment was analysed with IQ-TREE using the same parameters used for the 87-single copy protein domain matrix (LG+R7+C60 selected with TESTNEW, statistical supports from 1,000 ultrafast bootstrap replicates).

Comparative genomics by looking Protein domain gains and loses

116 different eukaryotic taxa with proteomic information available was selected to perform an analysis of protein gains and loses over the

eukaryotic tree of life focusing on holozoans (56 taxa) (Supplementary Table 4). Protein domain annotations of each proteome were computed using Pfamscan and the 29th release of the Pfam database (Bateman et al. 2004). We used a custom script to build a matrix containing the eukaryotic taxa and the number of copies presence of each protein domain. In order, to reduce noise and eliminate possible contaminants, we removed all the protein domains that >95% of the sequences found in the Pfam database belong to Bacteria or Archaea. We ended up with a matrix of 116 taxa and 8,920 protein domains. Next, we produced a tree nexus file according to the topology of eukaryotes (Derelle, Torruella, and Klime 2015), and for unicellular holozoans we incorporated the topology of our phylogenomic analysis. With the protein domain matrix, and the consensus taxa tree we used Count (Csurös 2010) to infer the gains and loses of each node of tree by Dollo parsimony. Thanks to Count, the domains gained at the different ancestral nodes of holozoans could be retrieved. The functional annotation of the 120 protein domains gains at Choanozoa was done manually by checking the literature available of each protein domain. The list of proteins domains gained at the ancestral nodes of Opisthokonta (Opisthokonta, Holozoa, Filozoa, Choanozoa and Metazoa) are available at Figshare (XXXX) together with the protein domain matrix used.

References

Alié, Alexandre, Lucas Leclère, Muriel Jager, Cyrielle Dayraud, Patrick Chang, Hervé Le Guyader, Eric Quéinnec, and Michaël Manuel. 2011. "Somatic Stem Cells Express Piwi and Vasa Genes in an Adult Ctenophore: Ancient Association of 'germline Genes' with Stemness." *Developmental Biology* 350 (1). Elsevier Inc.: 183–97. doi:10.1016/j.ydbio.2010.10.019.

Attisano, Liliana, and Si Tuen Lee-hoeflich. 2001. "Protein Family Review The Smads," 1–8.

Bankevich, Anton, Sergey Nurk, Dmitry Antipov, Alexey A. Gurevich, Mikhail Dvorkin, Alexander S. Kulikov, Valery M. Lesin, et al. 2012. "SPAdes: A New Genome Assembly Algorithm and Its Applications to Single-Cell Sequencing." *Journal of Computational Biology* 19 (5): 455–77. doi:10.1089/cmb.2012.0021.

Bateman, Alex, Lachlan Coin, Richard Durbin, Robert D Finn, Volker Hollich, Sam Grif, Ajay Khanna, et al. 2004. "The Pfam Protein Families Database." *Nucleic Acids Res* 32 (Database issue): 138D–41. doi:10.1093/nar/gkh121.

Beck, N, and B Franz Lang. 2010. "Mfannot, Organelle Genome Annotation Webserver." http://megasun.bch.umontreal.ca/cgi-bin/mfannot/mfannotInterface.pl.

Bolger, Anthony M., Marc Lohse, and Bjoern Usadel. 2014. "Trimmomatic: A Flexible Trimmer for Illumina Sequence Data." *Bioinformatics* 30 (15): 2114–20. doi:10.1093/bioinformatics/btu170.

Box, Joseph K, Nicolas Paquet, Mark N Adams, Didier Boucher, Emma Bolderson, Kenneth J O Byrne, and Derek J Richard. 2016. "Nucleophosmin : From Structure and Function to Disease Development." *BMC Molecular Biology*. BioMed Central, 1–12. doi:10.1186/s12867-016-0073-9.

Brunet, Thibaut, and Nicole King. 2017. "The Origins of Animal Multicellularity and Cell Differentiation." *BioRxiv*.

Burger, Gertraud, Michael W. Gray, Lise Forget, and B. Franz Lang. 2013. "Strikingly Bacteria-like and Gene-Richmitochondrial Genomes throughout Jakobid Protists." *Genome Biology and Evolution* 5 (2): 418–38. doi:10.1093/gbe/evt008.

Burkhardt, Pawel, Mads Grønborg, Kent McDonald, Tara Sulur, Qi Wang, and Nicole King. 2014. "Evolutionary Insights into Premetazoan Functions of the Neuronal Protein Homer." *Molecular Biology and Evolution* 31 (9): 2342–55. doi:10.1093/molbev/msu178.

Burkhardt, Pawel, Christian M Stegmann, Benjamin Cooper, Tobias H Kloepper, and Cordelia Imig. 2011. "Primordial Neurosecretory Apparatus Identi Fi Ed in the Choano Fl Agellate Monosiga Brevicollis," 2–7. doi:10.1073/pnas.1106189108/-/DCSupplemental.www.pnas.org/cgi/doi/10.1073/pnas.1106189108.

Cai, Xinjiang. 2008. "Unicellular Ca2+ Signaling 'Toolkit' at the Origin of Metazoa." *Molecular Biology and Evolution* 25 (7): 1357–61. doi:10.1093/molbev/msn077.

Camacho, C, G Coulouris, V Avagyan, N Ma, J Papadopoulos, K Bealer, and T L Madden. 2009. "BLAST plus: Architecture and Applications." *BMC Bioinformatics* 10 (421): 1. doi:Artn 421\nDoi 10.1186/1471-2105-10-421.

Carr, M, B S C Leadbeater, R Hassan, M Nelson, and S L Baldauf. 2008. "Molecular Phylogeny of Choanoflagellates, the Sister Group to Metazoa." *Proceedings of the National Academy of Sciences of the United States of America* 105 (43): 16641–46. doi:10.1073/pnas.0801667105.

Carr, Martin, Barry S. C. Leadbeater, Ruhana Hassan, Michaela Nelson, and Sandra L. Baldauf. 2008. "Molecular Phylogeny of Choanoflagellates, the Sister Group to Metazoa." *Proceedings of the National Academy of Sciences* 105 (43): 16641–46. doi:10.1073/pnas.0801667105.

Carr, Martin, Daniel J. Richter, Parinaz Fozouni, Timothy J. Smith, Alexandra Jeuck, Barry S.C. Leadbeater, and Frank Nitsche. 2017. "A Six-Gene Phylogeny Provides New Insights into Choanoflagellate Evolution." *Molecular Phylogenetics and*

Evolution 107. The Authors: 166–78. doi:10.1016/j.ympev.2016.10.011.

Csurös, Miklós. 2010. "Count: Evolutionary Analysis of Phylogenetic Profiles with Parsimony and Likelihood." *Bioinformatics* 26 (15): 1910–12. doi:10.1093/bioinformatics/btq315.

de Mendoza, Alex, Arnau Sebé-Pedrós, Martin Sebastijan Šestak, Marija Matejcic, Guifré Torruella, Tomislav Domazet-Loso, and Iñaki Ruiz-Trillo. 2013. "Transcription Factor Evolution in Eukaryotes and the Assembly of the Regulatory Toolkit in Multicellular Lineages." *Proceedings of the National Academy of Sciences of the United States of America* 110 (50): 1–9. doi:10.1073/pnas.1311818110.

Dean, Frank B, Seiyu Hosono, Linhua Fang, Xiaohong Wu, a Fawad Faruqi, Patricia Bray-Ward, Zhenyu Sun, et al. 2002. "Comprehensive Human Genome Amplification Using Multiple Displacement Amplification." *Proceedings of the National Academy of Sciences of the United States of America* 99 (8): 5261–66. doi:10.1073/pnas.082089499.

de Vargas, Colomban, Stéphane Audic, Nicolas Henry, Johan Decelle, F. Mahe, Ramiro Logares, Enrique Lara, et al. 2015. "Eukaryotic Plankton Diversity in the Sunlit Ocean." *Science (New York, N.Y.)* 348 (6237): 1261605–1261605. doi:10.1126/science.1261605.

de Winter, Johan P, and Hans Joenje. 2009. "The Genetic and Molecular Basis of Fanconi Anemia." *Mutation Research* 668 (1–2): 11–19. doi:10.1016/j.mrfmmm.2008.11.004.

del Campo, Javier, and Iñaki Ruiz-Trillo. 2013. "Environmental Survey Meta-Analysis Reveals Hidden Diversity among Unicellular Opisthokonts." *Molecular Biology and Evolution* 30 (4): 802–5. doi:10.1093/molbev/mst006.

Derelle, Romain, Guifré Torruella, and Vladimír Klime. 2015. "Bacterial Proteins Pinpoint a Single Eukaryotic Root," 1–7. doi:10.1073/pnas.1420657112.

Fairclough, Stephen R, Zehua Chen, Eric Kramer, Qiandong Zeng, Sarah Young, Hugh M Robertson, Emina Begovic, et al. 2013. "Premetazoan Genome Evolution and the Regulation of Cell Differentiation in the Choanoflagellate Salpingoeca Rosetta." *Genome Biology* 14 (2): R15. doi:10.1186/gb-2013-14-2-r15.

Frank, Nitsche, Thomsen Helge Abuldhauge, and J. Richter Daniel. 2017. "Bridging the Gap between Morphological Species and Molecular Barcodes ??? Exemplified by Loricate Choanoflagellates." *European Journal of Protistology* 57. Elsevier GmbH: 26–37. doi:10.1016/j.ejop.2016.10.006.

Gorokhova, Svetlana, Stéphanie Bibert, Käthi Geering, and Nathaniel Heintz. 2007. "A Novel Family of Transmembrane Proteins Interacting with β Subunits of the Na,K-ATPase." *Human Molecular*

Genetics 16 (20): 2394–2410. doi:10.1093/hmg/ddm167.

Grau-Bové, Xavier, Guifré Torruella, Stuart Donachie, Hiroshi Suga, Guy Leonard, Thomas A Richards, and Iñaki Ruiz-Trillo. 2017. "Dynamics of Genomic Innovation in the Unicellular Ancestry of Animals." *eLife* 6. doi:10.7554/eLife.26036.

Guindon, Stéphane, Jean Franois Dufayard, Vincent Lefort, Maria Anisimova, Wim Hordijk, and Olivier Gascuel. 2010. "New Algorithms and Methods to Estimate Maximum-Likelihood Phylogenies: Assessing the Performance of PhyML 3.0." *Systematic Biology* 59 (3): 307–21. doi:10.1093/sysbio/syq010.

Gurevich, Alexey, Vladislav Saveliev, Nikolay Vyahhi, and Glenn Tesler. 2013. "QUAST: Quality Assessment Tool for Genome Assemblies." *Bioinformatics* 29 (8): 1072–75. doi:10.1093/bioinformatics/btt086.

Hardy, D M, and D L Garbers. 1995. "A Sperm Membrane-Protein That Binds in a Species-Specific Manner to the Egg Extracellular-Matrix Is Homologous to Von-Willebrand-Factor." *Journal of Biological Chemistry* 270 (44): 26025–28.

Hehenberger, Elisabeth, Denis V. Tikhonenkov, Martin Kolisko, Javier del Campo, Anton S. Esaulov, Alexander P. Mylnikov, and Patrick J. Keeling. 2017. "Novel Predators Reshape Holozoan Phylogeny and Reveal the Presence of a Two-Component Signaling System in the Ancestor of Animals." *Current Biology*. Elsevier Ltd., 1–8. doi:10.1016/j.cub.2017.06.006.

Heywood, Jane L, Michael E Sieracki, Wendy Bellows, Nicole J Poulton, and Ramunas Stepanauskas. 2011. "Capturing Diversity of Marine Heterotrophic Protists: One Cell at a Time." *The ISME Journal* 5 (4). Nature Publishing Group: 674–84. doi:10.1038/ismej.2010.155.

James-Clark, H. 1866. "Conclusive Proofs on the Animality of the Ciliate Sponges, and Their Affinities with the Infusoria Flagellata." *Am J Sci Ser. 2 42:320–325* 2 (42): 320–325.

Karsenti, Eric, Silvia G Acinas, Peer Bork, Chris Bowler, Colomban De Vargas, Jeroen Raes, Matthew Sullivan, et al. 2011. "A Holistic Approach to Marine Eco-Systems Biology." *PLoS Biology* 9 (10): e1001177. doi:10.1371/journal.pbio.1001177.

Katoh, Kazutaka, Kazuharu Misawa, Kei-ichi Kuma, and Takashi Miyata. 2002. "MAFFT: A Novel Method for Rapid Multiple Sequence Alignment Based on Fast Fourier Transform." *Nucleic Acids Research* 30 (14): 3059–66. http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=135756&tool=pmcentrez&rendertype=abstract.

King, Nicole, M Jody Westbrook, Susan L Young, Alan Kuo, Monika Abedin, Jarrod Chapman, Stephen Fairclough, et al. 2008. "The Genome of the Choanoflagellate Monosiga Brevicollis and the

Origin of Metazoans." *Nature* 451 (7180): 783–88. doi:10.1038/nature06617.

Lang, B. F., C. O'Kelly, T. Nerad, M. W. Gray, and G. Burger. 2002. "The Closest Unicellular Relatives of Animals." *Current Biology* 12 (20): 1773–78. doi:10.1016/S0960-9822(02)01187-9.

Langmead, Ben, and Steven L Salzberg. 2012. "Fast Gapped-Read Alignment with Bowtie 2." *Nat Methods* 9 (4): 357–59. doi:10.1038/nmeth.1923.

Lartillot, Nicolas, and Hervé Philippe. 2004. "A Bayesian Mixture Model for across-Site Heterogeneities in the Amino-Acid Replacement Process." *Molecular Biology and Evolution* 21 (6): 1095–1109. doi:10.1093/molbev/msh112.

Leadbeater, B.S.C. 2015. *The Choanoflagellates: Evolution, Biology, and Ecology*. Cambridge: Cambridge University Press.

Leadbeater, Barry S C, QiBin Yu, Joyce Kent, and Dov J Stekel. 2009. "Three-Dimensional Images of Choanoflagellate Loricae." *Proceedings. Biological Sciences / The Royal Society* 276 (1654): 3–11. doi:10.1098/rspb.2008.0844.

Lechauve, Christophe, Muriel Jager, Laurent Laguerre, Laurent Kiger, Gaëlle Correc, Cédric Leroux, Serge Vinogradov, Mirjam Czjzek, Michael C. Marden, and Xavier Bailly. 2013. "Neuroglobins, Pivotal Proteins Associated with Emerging Neural Systems and Precursors of Metazoan Globin Diversity." *Journal of Biological Chemistry* 288 (10): 6957–67. doi:10.1074/jbc.M112.407601.

Liebeskind, Benjamin J. 2011. "Evolution of Sodium Channels and the New View of Early Nervous System Evolution." *Evolution*, no. December: 679–83. doi:10.4161/cib.4.6.17069.

López-Escardó, David, Xavier Grau-Bové, Amy Guillaumet-Adkins, Marta Gut, Michael E. Sieracki, and Iñaki Ruiz-Trillo. 2017. "Evaluation of Single-Cell Genomics to Address Evolutionary Questions Using Three SAGs of the Choanoflagellate Monosiga Brevicollis." *Scientific Reports* 7:11025. doi:10.1038s41598-017-11466-9.

Mangot, Jean-françois, Ramiro Logares, Pablo Sanchez, and Fran Latorre. 2017. "Accessing to the Genomic Information of Unculturable Oceanic Picoeukaryotes by Combining Multiple Single Cells." *Scientific Reports* 7. doi:10.1038/srep41498.

Marron, Alan O., Sarah Ratcliffe, Glen L. Wheeler, Raymond E. Goldstein, Nicole King, Fabrice Not, Colomban De Vargas, and Daniel J. Richter. 2016. "The Evolution of Silicon Transport in Eukaryotes." *Molecular Biology and Evolution* 33 (12): 3226–48. doi:10.1093/molbev/msw209.

Martinez-Garcia, Manuel, David Brazel, Nicole J Poulton, Brandon K Swan, Monica Lluesma Gomez, Dashiell Masland, Michael E Sieracki, and Ramunas Stepanauskas. 2012. "Unveiling in

Situ Interactions between Marine Protists and Bacteria through Single Cell Sequencing." *The ISME Journal* 6 (3). Nature Publishing Group: 703–7. doi:10.1038/ismej.2011.126.

Massana, Ramon, Vanessa Balagué, Laure Guillou, and Carlos Pedrós-Alió. 2004. "Picoeukaryotic Diversity in an Oligotrophic Coastal Site Studied by Molecular and Culturing Approaches." *FEMS Microbiology Ecology* 50 (3): 231–43. doi:10.1016/j.femsec.2004.07.001.

Minh, Bui Quang, Minh Anh Thi Nguyen, and Arndt Von Haeseler. 2013. "Ultrafast Approximation for Phylogenetic Bootstrap." *Molecular Biology and Evolution* 30 (5): 1188–95. doi:10.1093/molbev/mst024.

Nguyen, Lam Tung, Heiko A. Schmidt, Arndt Von Haeseler, and Bui Quang Minh. 2015. "IQ-TREE: A Fast and Effective Stochastic Algorithm for Estimating Maximum-Likelihood Phylogenies." *Molecular Biology and Evolution* 32 (1): 268–74. doi:10.1093/molbev/msu300.

Nitsche, Frank, Claudia Wylezich, Friedrich Loeffler Institut, and Greifswald-insel Ri. 2007. "Deep Sea Records of Choanoflagellates with a Description of Two New Species." *Acta Protozoologica* 46 (July 2017): 99–106.

Paps, Jordi, Luis a Medina-Chacón, Wyth Marshall, Hiroshi Suga, and Iñaki Ruiz-Trillo. 2013. "Molecular Phylogeny of Unikonts: New Insights into the Position of Apusomonads and Ancyromonads and the Internal Relationships of Opisthokonts." *Protist* 164 (1): 2–12. doi:10.1016/j.protis.2012.09.002.

Parra, Genis, Keith Bradnam, and Ian Korf. 2007. "CEGMA: A Pipeline to Accurately Annotate Core Genes in Eukaryotic Genomes." *Bioinformatics* 23 (9): 1061–67. doi:10.1093/bioinformatics/btm071.

Peña, Jesús F., Alexandre Alié, Daniel J. Richter, Lingyu Wang, Noriko Funayama, and Scott Anthony Nichols. 2016. "Conserved Expression of Vertebrate Microvillar Gene Homologs in Choanocytes of Freshwater Sponges." *EvoDevo* 7 (1). BioMed Central: 13. doi:10.1186/s13227-016-0050-x.

Pesant, Stéphane, Fabrice Not, Marc Picheral, Stefanie Kandels-Lewis, Noan Le Bescot, Gabriel Gorsky, Daniele Iudicone, et al. 2015. "Open Science Resources for the Discovery and Analysis of Tara Oceans Data." *Scientific Data* 2 (Lmd): 150023. doi:10.1038/sdata.2015.23.

Quang, Le Si, Olivier Gascuel, and Nicolas Lartillot. 2008. "Empirical Profile Mixture Models for Phylogenetic Reconstruction." *Bioinformatics* 24 (20): 2317–23. doi:10.1093/bioinformatics/btn445.

R Core Team. 2013. "R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing." Vienna (Austria). http://www.r-project.org/.

Richter, Daniel, and Nicole King. 2013. “The Genomic and Cellular Foundations of Animal Origins.” *Annual Review of Genetics*, 1–31. doi:10.1146/annurev-genet-111212-133456.

Ronquist, F., and J. P. Huelsenbeck. 2003. “MrBayes 3: Bayesian Phylogenetic Inference under Mixed Models.” *Bioinformatics* 19 (12): 1572–74. doi:10.1093/bioinformatics/btg180.

Ruiz-Trillo, Iñaki, Andrew J. Roger, Gertraud Burger, Michael W. Gray, and B. Franz Lang. 2008. “A Phylogenomic Investigation into the Origin of Metazoa.” *Molecular Biology and Evolution* 25 (4): 664–72. doi:10.1093/molbev/msn006.

Sebé-Pedrós, Arnau, Pawel Burkhardt, Núria Sánchez-Pons, Stephen R. Fairclough, B. Franz Lang, Nicole King, and Iñaki Ruiz-Trillo. 2013. “Insights into the Origin of Metazoan Filopodia and Microvilli.” *Molecular Biology and Evolution* 30 (9): 2013–23. doi:10.1093/molbev/mst110.

Sebé-Pedrós, Arnau, Bernard M Degnan, and Iñaki Ruiz-Trillo. 2017. “The Origin of Metazoa, a Unicellular Perspective.” *Nature Reviews Genetics* in press. doi:10.1038/nrg.2017.21.

Simão, Felipe A., Robert M. Waterhouse, Panagiotis Ioannidis, Evgenia V. Kriventseva, and Evgeny M. Zdobnov. 2015. “BUSCO: Assessing Genome Assembly and Annotation Completeness with Single-Copy Orthologs.” *Bioinformatics* 31 (19): 3210–12. doi:10.1093/bioinformatics/btv351.

Simion, Paul, Hervé Philippe, Denis Baurain, Muriel Jager, Daniel J. Richter, Arnaud Di Franco, Béatrice Roure, et al. 2017. “A Large and Consistent Phylogenomic Dataset Supports Sponges as the Sister Group to All Other Animals.” *Current Biology*, 1–10. doi:10.1016/j.cub.2017.02.031.

Stanke, Mario, and Burkhard Morgenstern. 2005. “AUGUSTUS: A Web Server for Gene Prediction in Eukaryotes That Allows User-Defined Constraints.” *Nucleic Acids Research* 33 (SUPPL. 2): 465–67. doi:10.1093/nar/gki458.

Steenkamp, Emma T, Jane Wright, and Sandra L Baldauf. 2006. “The Protistan Origins of Animals and Fungi.” *Molecular Biology and Evolution* 23 (1): 93–106. doi:10.1093/molbev/msj011.

Stepanauskas, Ramunas, and Michael E Sieracki. 2007. “Matching Phylogeny and Metabolism in the Uncultured Marine Bacteria, One Cell at a Time.” *Proceedings of the National Academy of Sciences of the United States of America* 104 (21): 9052–57. doi:10.1073/pnas.0700496104.

Stibbs, Henry H., Alfred Owczarzak, Christopher J. Bayne, and Peggy DeWan. 1979. “Schistosome Sporocyst-Killing Amoebae Isolated from Biomphalaria Glabrata.” *Journal of Invertebrate Pathology* 33 (2): 159–70. doi:10.1016/0022-2011(79)90149-6.

Tara Oceans Consortium, Coordinators; Tara Oceans Expedition, Participants. 2014. "Registry of Selected Samples from the Tara Oceans Expedition (2009-2013)." doi:10.1594/PANGAEA.842197 553.

Torruella, Guifré, Alex De Mendoza, Xavier Grau-Bové, Meritxell Antó, Mark A. Chaplin, Javier Del Campo, Laura Eme, et al. 2015. "Phylogenomics Reveals Convergent Evolution of Lifestyles in Close Relatives of Animals and Fungi." *Current Biology* 25 (18): 2404–10. doi:10.1016/j.cub.2015.07.053.

Vargas, Colomban De, Stéphane Audic, Nicolas Henry, Johan Decelle, Frédéric Mahé, Ramiro Logares, Enrique Lara, et al. 2015. "Eukaryotic Plankton Diversity in the Sunlit Ocean." *Science*, no. May: 1–12.

Wainright, PO, G Hinkle, ML Sogin, and SK Stickel. 1993. "Monophyletic Origins of Metazoa: An Evolutionary Link with Fungi." *Science* 260 (5106): 340–42.

Wasmuth, Elizabeth V, and Christopher D Lima. 2016. "UniProt: The Universal Protein Knowledgebase." *Nucleic Acids Research* 45 (November 2016): 1–12. doi:10.1093/nar/gkw1152.

Weber, Felix, Javier del Campo, Claudia Wylezich, Ramon Massana, and Klaus Jürgens. 2012. "Unveiling Trophic Functions of Uncultured Protist Taxa by Incubation Experiments in the Brackish Baltic Sea." *PLoS ONE* 7 (7). doi:10.1371/journal.pone.0041970.

Weisz, A., P. Marx, R. Sharf, E. Appella, P. H. Driggers, K. Ozato, and B. Z. Levi. 1992. "Human Interferon Consensus Sequence Binding Protein Is a Negative Regulator of Enhancer Elements Common to Interferon-Inducible Genes." *Journal of Biological Chemistry* 267 (35): 25589–96.

Winberg, Margaret L., Jasprina N. Noordermeer, Luca Tamagnone, Paolo M. Comoglio, Melanie K. Spriggs, Marc Tessier-Lavigne, and Corey S. Goodman. 1998. "Plexin A Is a Neuronal Semaphorin Receptor That Controls Axon Guidance." *Cell* 95 (7): 903–16. doi:10.1016/S0092-8674(00)81715-8.

Yang, Jiwon, Tommy Harding, Ryoma Kamikawa, Alastair G.B. Simpson, and Andrew J. Roger. 2017. "Mitochondrial Genome Evolution and a Novel RNA Editing System in Deep-Branching Heteroloboseids." *Genome Biology and Evolution* 9 (5): 1161–74. doi:10.1093/gbe/evx086.

Yang, Z. 1995. "A Space-Time Process Model for the Evolution of DNA Sequences." *Genetics* 139 (2): 993–1005.

Zettler LAA, T a Nerad, C J O'Kelly, and M L Sogin. 2001. "The Nucleariid Amoebae: More Protists at the Animal-Fungal Boundary." *The Journal of Eukaryotic Microbiology* 48 (3): 293–97. http://www.ncbi.nlm.nih.gov/pubmed/11411837.

Figures and Tables:

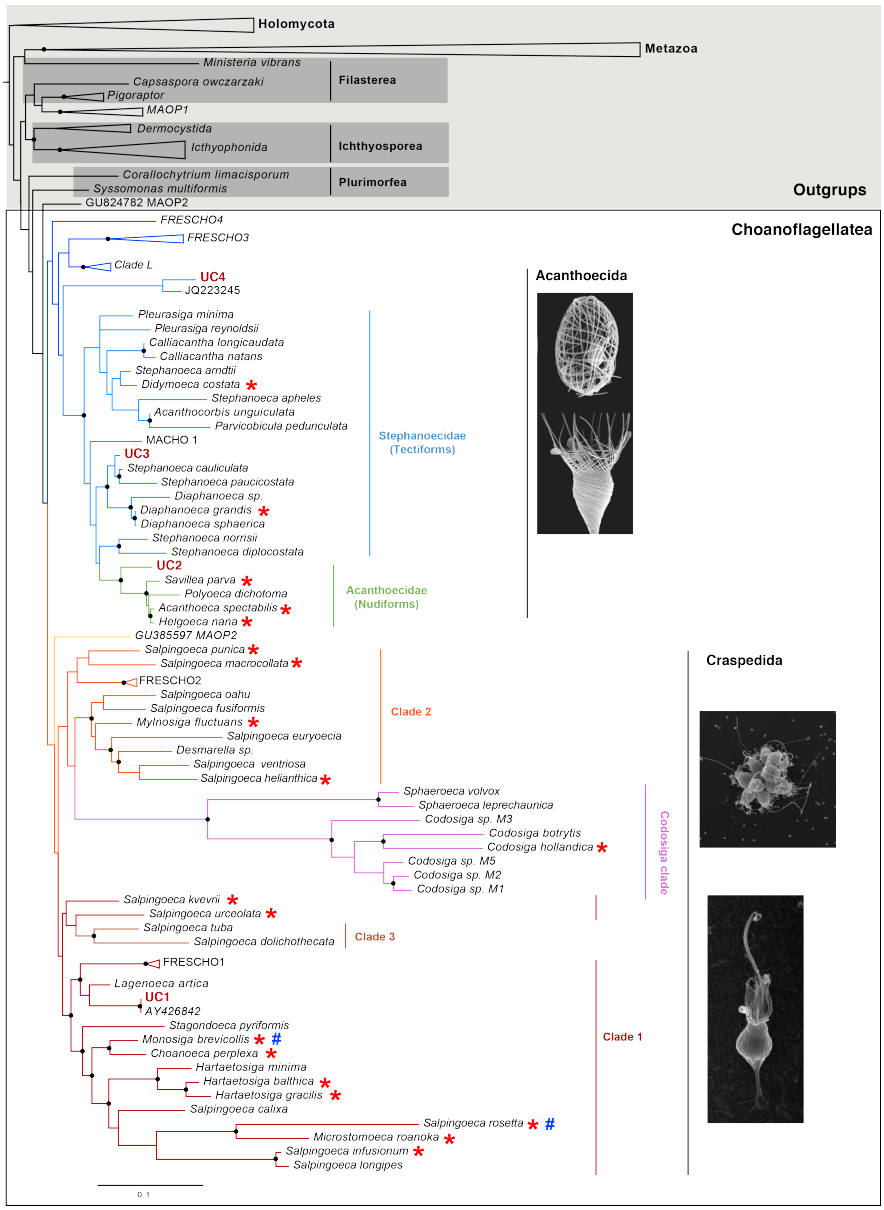


Figure 1. Phylogenetic position of the new choanoflagellate SAGs. Phylogenetic tree based in the 18S rDNA gene of 117 sequences representing all known molecular diversity of choanoflagellates and unicellular holozoans, including environmental lineages. The phylogenetic analysis was inferred by

Maximum likelihood under the
GTR+ G free rate with 6 categories model. Split supports were calculated: bootstr aps of single branch test (SH-aLRT) and ultrafast bootstraps calculated with IQ-TREE. Split support values >80 of SH-aLRT bootstraps and >95 of ultra fast bootstrap computed with IQ-tree are indicated by a bullet (•). Choanoflagellates with transcriptomic data available are depicted with a red asterisk with genomic data available are depicted with a blue hash. Choanoflagellates craspedidan clades were named according to our phylogenomic analysis (Fig.3). Clade 3 nomenclature, and nomenclature within Acanthoecida is the same than. Acanthoecida picture was taken from (Barry S C Leadbeater et al. 2009) and Craspedida pictures were taken from the web page (www.pinterest.com).

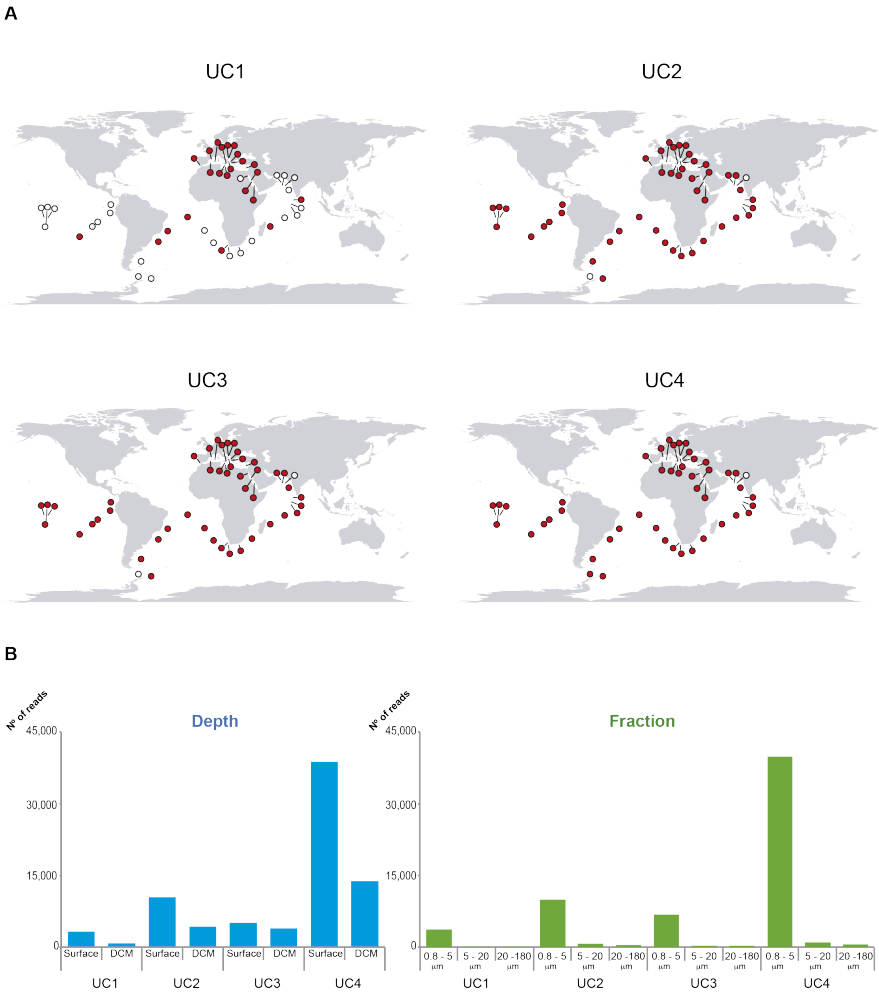


Figure 2. Ecological distribution of our SAGs. (**A**) Geographical location of our SAGs according to the metabarcoding data from TARA oceans expedition (Vargas et al. 2015) (see methods). Red circles mark TARA ocean stations with reads detected from each of our SAGs. White circles represent stations without signal of our SAGs. (**B**) Read distribution according to Depth and size fraction of our SAGs. In blue (left) appears the distribution of reads among different depths: the surface, and the depth clorophyl maximum (DCM). In green (right) it is shown the read distribution according to different size fractions.

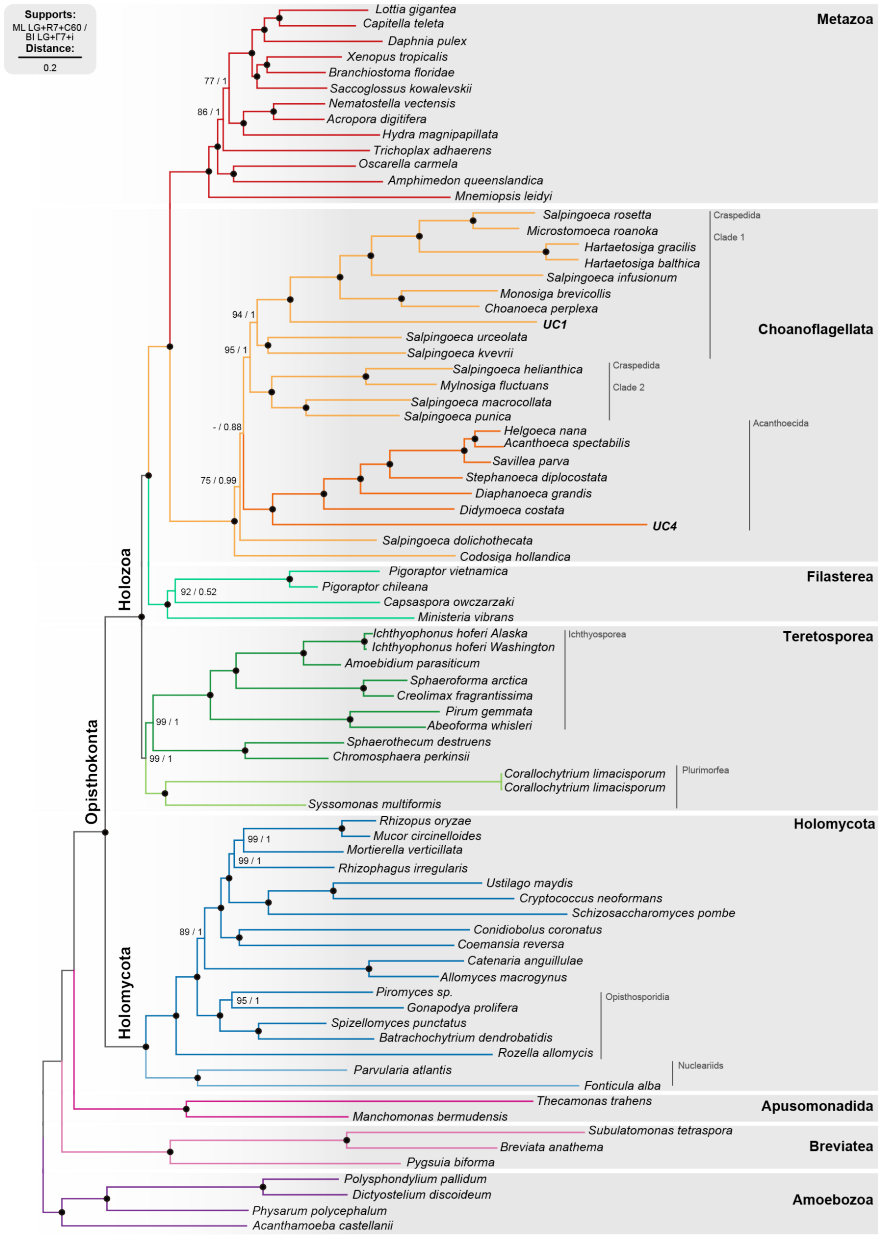


Figure 3. Phylogenomic tree of holozoans. Phylogenomic analysis of 87 single-copy protein domains (Grau-Bové et al. 2017) accounting for 23,364 aminoacid positions. Tree topology is the consensus of two Markov chain Monte Carlo chains run for XXX generations, saving every XXX trees and XXX after a burn-in of XX%. Statistical supports are indicated at each node: on the left, non-parametric maximum likelihood ultrafast-bootstrap (UFBS) values obtained from 1,000 replicates using IQ-TREE and the LG+R7+C60 model; on the right,

Bayesian posterior probabilities (BPP) under the LG+Γ7+CAT model as implemented in Phylobayes. Nodes with maximum support values (BPP = 1 and UFBS = 100) are indicated with a black bullet. Raw trees are available on Figshare (XXX) are available and Figure S5 shows the topology and the supports of the ML inference.

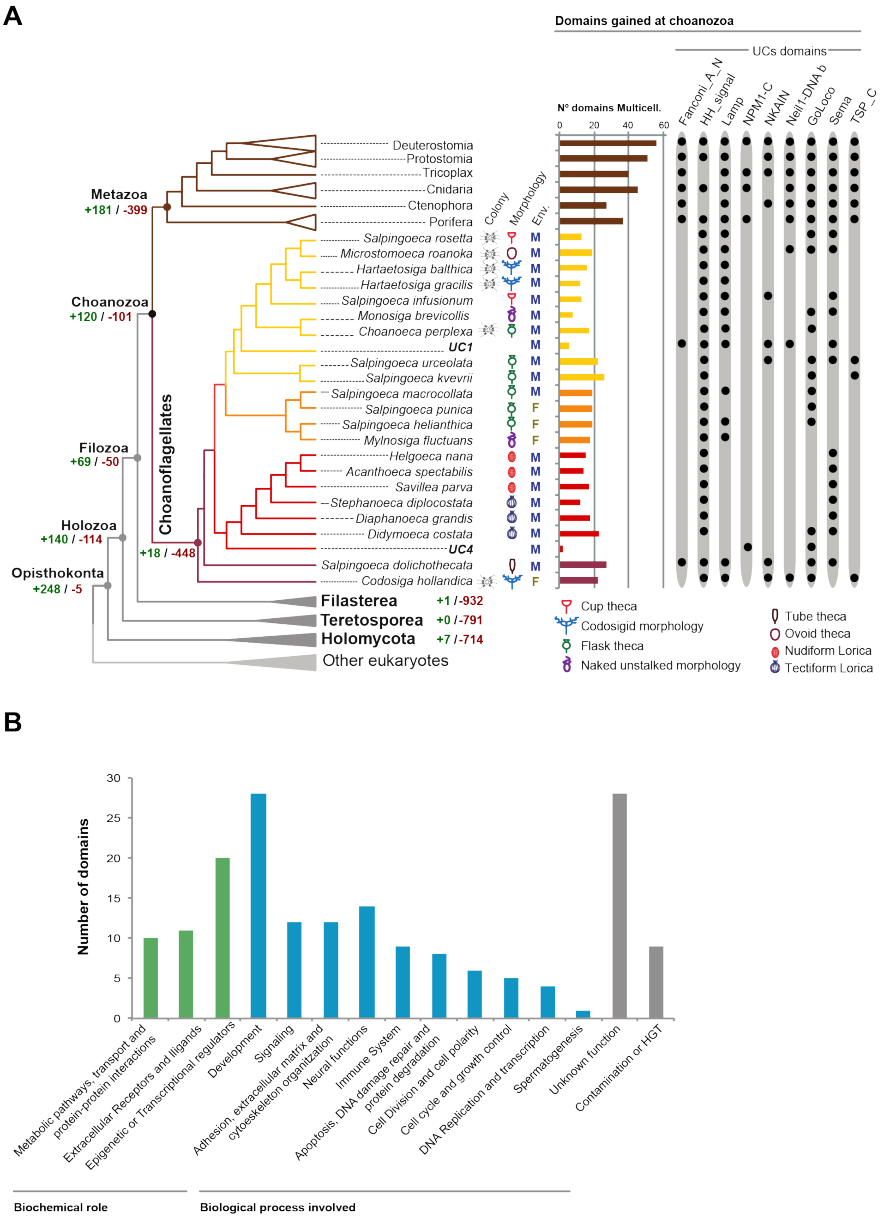


Figure 4. Summary of proteins gains and losses in Opisthokonta, focusing on Choanozoa gains. (A) Schematic representation of the choanoflagellate phylogeny obtained, including the number of protein domains gains and loses in each Opisthokonta clade (depicted in green and red respectively). Protein domains from potential bacterial or archea contamination were excluded from the analysis (see methods). The theca morphology (morphology legends are rigth-down), the ability to form colonies (marked with a colony drawing) and the environment of isolation (marine or freshwater represented by M or F

respectively), are shown in the right, and has been adapted from (Martin Carr et al. 2017). Our SAGs (UC1 and UC4) are marked in bold. Next to the tree, there is a bar chart indicating the number of protein domains gained at Choanozoa and described to be involved in animal multicellular processes (a total of 69 domains out of 120), retained in each choanoflagellate taxa. As animal are represented by lineages instead of species, it is shown the average of domains kept in each species of the metazoan group. On the right the protein domains gained at Choanoza and present in the sequenced SAGs UC1 and UC4. A black dot indicates the presence of each domain in the different taxa/clade. (**B**) Function of the protein domains gained at Choanoza. In green, the biochemical roles in which the protein domain are involved. In blue, the biological processes that the domain has been shown to participate. These two classifications are not exclusive; one protein domain can appear in one or multiple categories. In grey, protein domains with unknown function, or contaminants or a product of an horizontal gene transfer event.

Table 1. Summary of the genome statistics of each SAG assembly

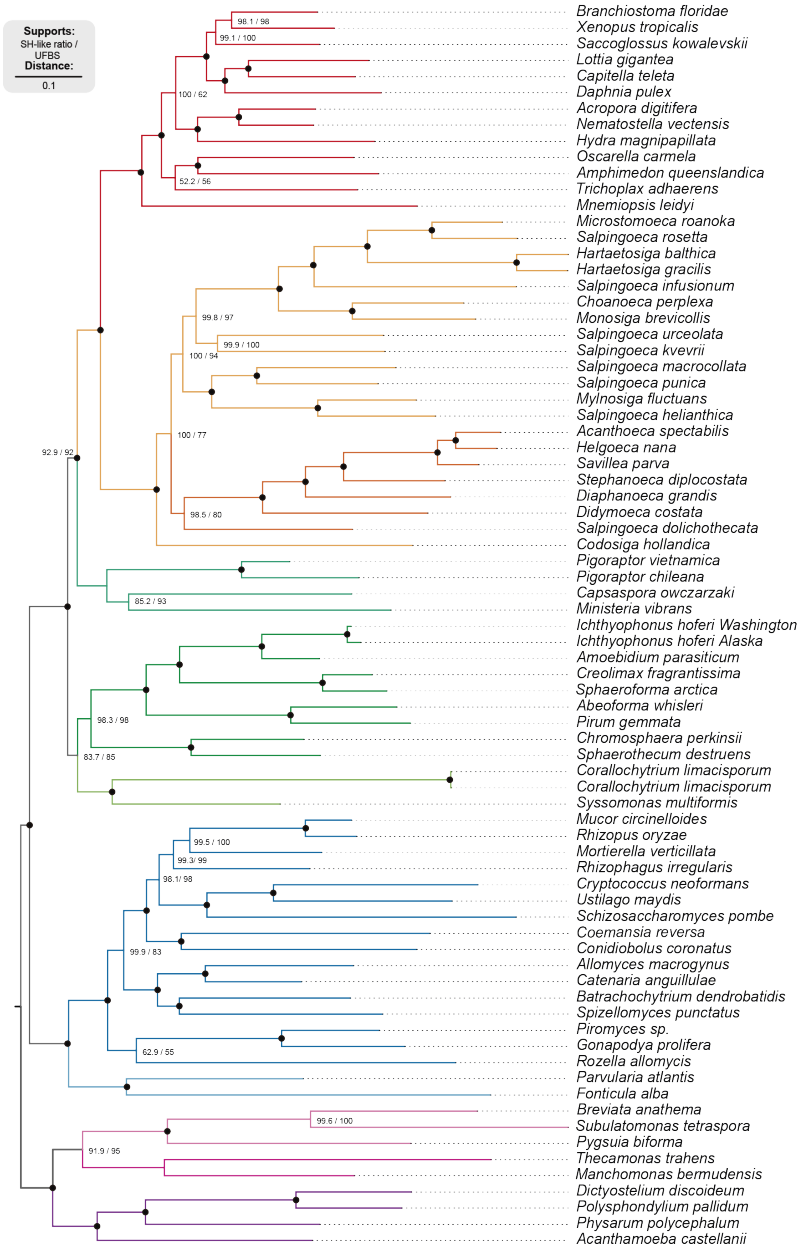
SAG	Taxonomy	#*	Largest Scaffold (bp)	N50	Total length (Mb)	GC (%)	CEGMA (%)	Busco (%)
UC1	Craspedida clade 1	3276	41,637	4928	7.74	49.8	20.1	31.7
UC2	Acanthoecidae	746	32,186	1499	1.00	30.8	0.8	0.7
UC3	Stephanocidae	819	111,87	2197	1.31	33.5	-	0.3
UC4	Basal Acanthoecida	2527	72,672	11360	7.25	40.0	14.1	13.5

*Number of Scaffolds bigger than 500bp

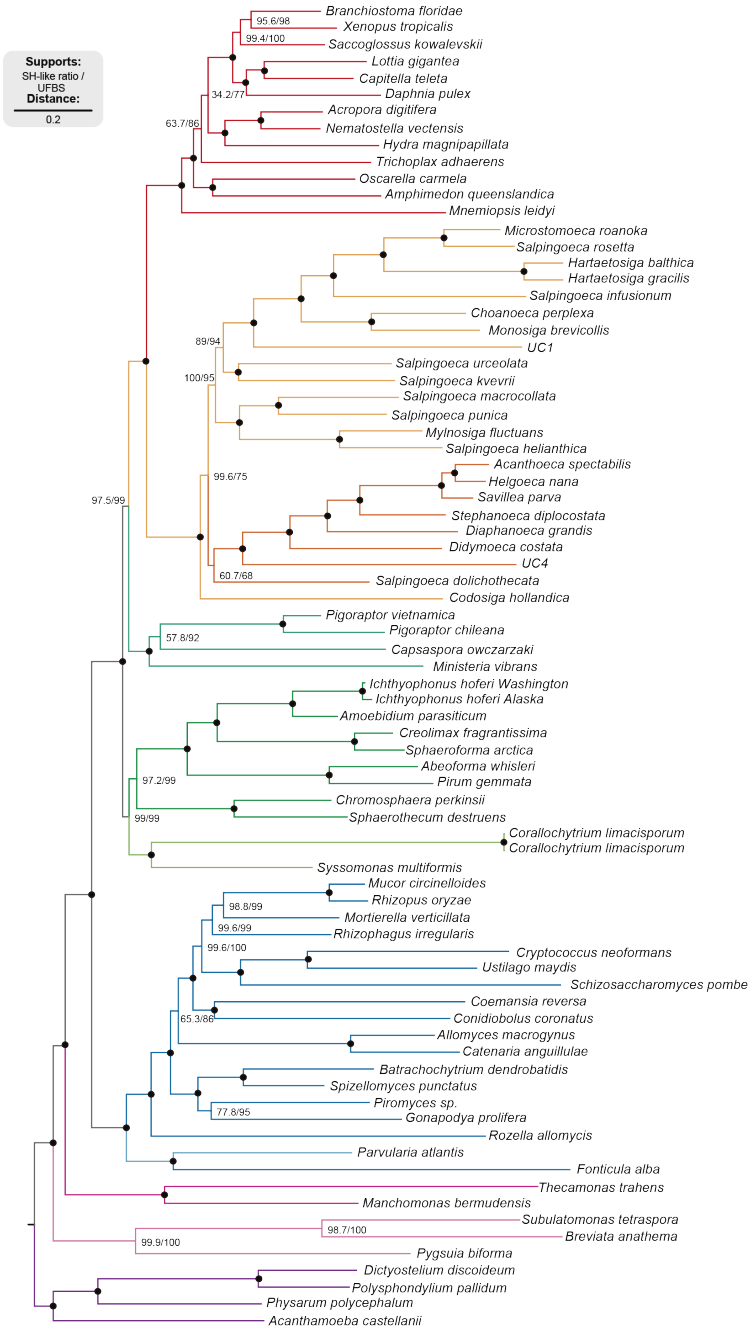
Table 2. Genome estimation of our SAGs† within choanoflagellate context

Genome	Assembly size (Mb)	Genome size (Mb)	Nº of annotated genes	Total Nº of genes
UC1	7.74	29.4†	3,025	6,039†
UC4	7.25	52.5†	2,518	10,075†
Salpingoeca rosetta	-	55.4	-	11,624
Monosiga brevicollis	-	41.6	-	9,172

Supplementary Figures and tables

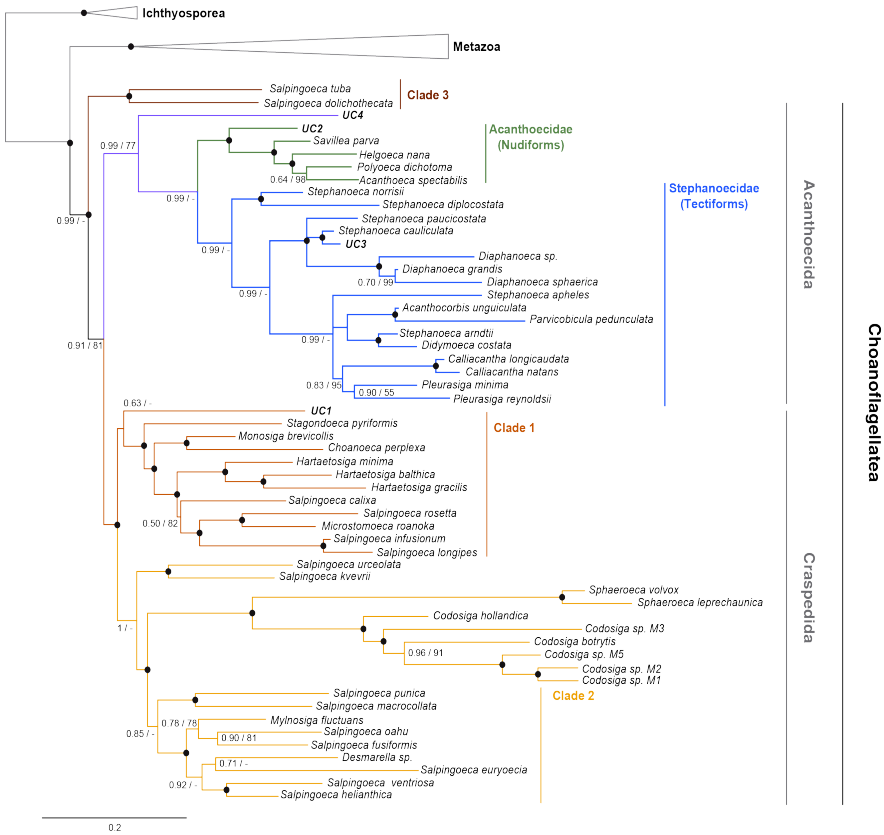


Supplementary Figure 1. Phylogenomic tree of holozoans without our SAGs. Maximum likelihood inference using the same dataset from Figure 3, but without our SAGs. Calculated with IQtree with LG+R7+PMSF model (supports are SH-like approximate likelihood ratio test / UFBS, respectively). Bullets indicate maximum nodal support (100/100).

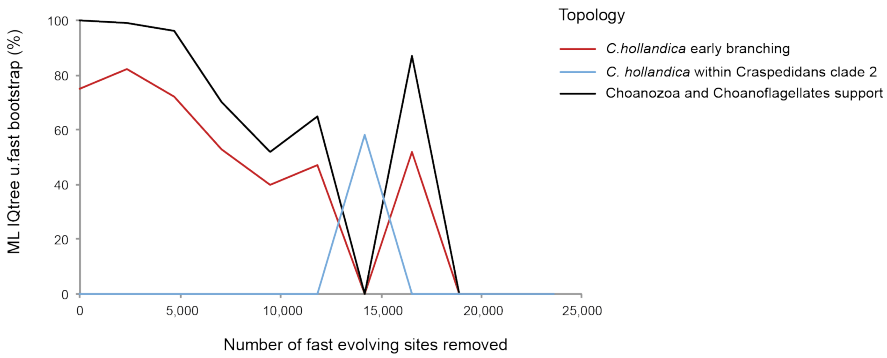


Supplementary Figure 2. Phylogenomic tree of holozoans, ML inference. Maximum likelihood inference of the UFBS displayed at Figure 3. Calculated with IQtree with LG+R7+C60 model (supports

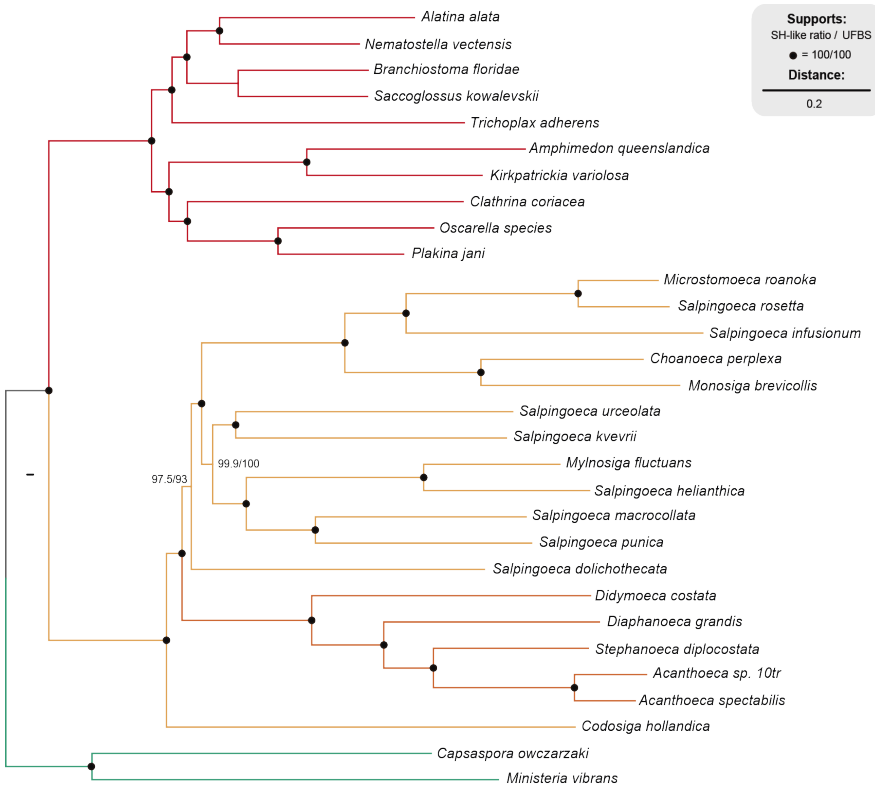
are SH-like approximate likelihood ratio test / UFBS, respectively). Bullets indicate maximum nodal support (100/100).



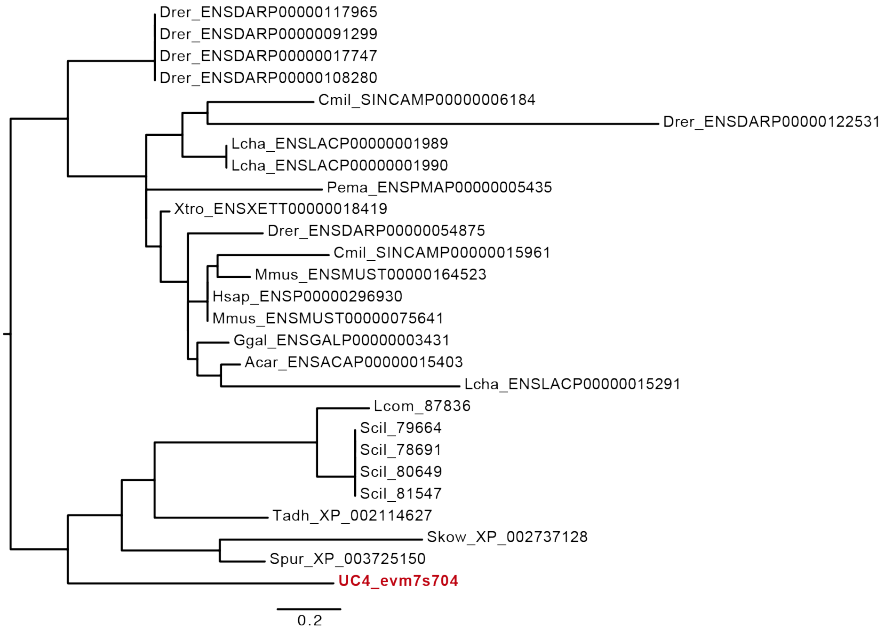
Supplementary Figure 3. Eight-gene phylogeny of choanoflagellates. Bayesian inference of an eight-gene phylogeny, adapted from (Martin Carr et al. 2017) and including the following genes: 18S rDNA, 28S rDNA, hsp90, alpha tubulin, EFL, EF-1A, actin and hsp70 . The choanoflagellate taxa includes our SAGs, the species from (Martin Carr et al. 2017) and as well the some choanoflagellate taxa described in (Frank, Helge Abuldhauge, and Daniel 2017). A summary of taxa is available at Supplementary Table 7. Nodal supports indicates the bayesian posterior probability (right) and the ultra fast bootstrap (UFBS) computed with IQtree (left) in a Maximum Likelihood analysis. The tree produced with ML is available at Figshare (XXXX).



Supplementary Figure 4. Fast-evolving site removal of alignment positions sorted by their rates of evolution. Sites were sequentially removed from fastest to slowest positions, 2,310 sites at a time, generating alternative datasets at each step. Ultra-fast bootstrap values were generated and are plotted the ones that support *Codosiga hollandica* as sister-group to the rest of choaonoflagellates (red) and the node of the topology that includes *C. hollandica* within Craspedida clade 2 (blue)**.** As a control are depicted the supports for Choanozoa and choanoflagellate monophyly (black).



Supplementary Figure 5. Choanflagellates phylogeny. Maximum likelihood inference tree of choanoflagellates using another dataset from (Simion et al. 2017). Our SAGs an another choanoflagellate taxa are not included. Calculated with IQtree with LG+R7+C60 model (supports are SH-like approximate likelihood ratio test / UFBS, respectively). Bullets indicate maximum nodal support (100/100).



Supplementary Figure 6. **Phylogeny of the domain NPM1-C.** Maximum likelihood analysis run with IQtree of the NPM1-C domain, including all the sequences found in our proteome database (List of species available at Supplementary Table 4).

Supplementary Table 1. Main environmental features of the SAG samples:

SAG	SAG TARA ID	TARA station	Coordinates	Date	Depth (m)	Temp (ºC)	Oxygen (µmol/kg)	Salinity (psu)	Chloroph-yll (mg Chl/m3)
UC1	AAA538_N18_CHOA	23	42.1735ºN 17.7252ºE	18-11-09	55	16.9	226.2	38.3	0.19
UC2	AB242_J22_CHOA	51	21.5043º S 42.1735ºE	11-5-10	5.0	27.3	194.3	35.1	0.22
UC3	AB537_J22	41	14.5536º N 70.0128ºE	30-3-10	60	29.1	185.5	36.2	0.45
UC4	AB240_J14	41	14.6059º N 69.9776ºE	30-3-10	5	29.1	187.7	36.2	0.32

Supplementary Table 2. Read information and genome statistics from the assemblies performed in each SAG

SAG	Number of reads	Assembly length (Mb)	GC (%)	N50	L75	Largest scaffold (bp)	Number of scaffolds*
UC1	1.24E+08	7.75	49.8	4,928	1,061	41,637	3,276
UC2	4.48E+07	1.00	30.8	1,499	351	32,189	746
UC3	4.25+07	1.32	33.5	2,197	309	111,870	819
UC4	1.15E+08	7.25	40.0	11,360	496	72,672	2,527

Supplementary Table 3. Sumary of the genes annotated in the UC2 mitochondrial genome

UC2 mitochondrial genes	
atp9	rps12
cob	rps13
cox1	rps3
cox2	rps4
cox3	rps8
nad1	rrn5
nad2	trnA(ugc)
nad3	trnC(gca)
nad4	trnD(guc)
nad4L	trnE(uuc)
nad5	trnF(gaa)
nad6	trnG(ucc)
orf109	trnH(gug)
orf130	trnI(gau)
orf139	trnK(uuu)
orf143	trnL(uaa)
orf154	trnL(uag)
orf181	trnM(cau)
orf194	trnN(guu)
orf212	trnP(ugg)
orf386	trnQ(uug)
orf408	trnR(ucg)
orf717	trnR(ucu)
rnl	trnS(gcu)
rns	trnS(uga)
rpl14	trnT(ugu)
rpl16	trnV(uac)
rpl2	trnW(uca)
rpl5	trnY(gua)
rpl6	

Supplementary Table 4. Summary of eukaryotic species used for the comparative genomics analysis

Species	Taxonomy	Abbrevation
Homo sapiens	Metazoa	Hsap
Mus musculus	Metazoa	Mmus
Xenopus tropicalis	Metazoa	Xtro
Branchiostoma floridae	Metazoa	Bflo
Ciona intestinalis	Metazoa	Cint
Oikopleura dioica	Metazoa	Odio
Daphnia pulex	Metazoa	Dpul
Drosophila melanogaster	Metazoa	Dmel
Tribolium castaneum	Metazoa	Tcas
Capitella teleta	Metazoa	Ctel
Saccoglossus kowalevskii	Metazoa	Skow
Lottia gigantea	Metazoa	Lgig
Trichoplax adhaerens	Metazoa	Tadh
Nematostella vectensis	Metazoa	Nvec
Aiptasia	Metazoa	Aipt
Hydra magnipapillata	Metazoa	Hmag
Mnemiopsis leidyi	Metazoa	Mlei
Acropora digitifera	Metazoa	Adig
Amphimedon queenslandica	Metazoa	Aque
Oscarella carmela	Metazoa	Ocar
Sycon ciliatum	Metazoa	Scil
Codosiga hollandica	Choanoflagellata	Chol
UC4	Choanoflagellata	UC4
Helgoeca nana	Choanoflagellata	Hnan
Didymoeca costata	Choanoflagellata	Dcos
Savillea parva	Choanoflagellata	Sepa
Acanthoeca spectabilis	Choanoflagellata	Aspe
Stephanoeca diplocostata	Choanoflagellata	Sdip
Diaphanoeca grandis	Choanoflagellata	Dgra
Salpingoeca dolichothecata	Choanoflagellata	Sdol
Salpingoeca rosetta	Choanoflagellata	Sros
Salpingoeca roanoka	Choanoflagellata	Sroa
Hartaetosiga balthica	Choanoflagellata	Hbal
Hartaetosiga gracilis	Choanoflagellata	Hgra
Salpingoeca infusionum	Choanoflagellata	Sinf
Monosiga brevicollis	Choanoflagellata	Mbre
Choanoeca perplexa	Choanoflagellata	Cper
UC1	Choanoflagellata	UC1
Salpingoeca kvevrii	Choanoflagellata	Skve
Salpingoeca urceolata	Choanoflagellata	Surc
Salpingoeca macrocollata	Choanoflagellata	Smac
Salpingoeca punica	Choanoflagellata	Sapu
Salpingoeca helianthica	Choanoflagellata	Shel
Mylnosiga fluctuans	Choanoflagellata	Mflu
Pigoraptor vietnamica	Filasterea	Pvie
Pigoraptor chilena	Filasterea	Pchi
Capsaspora owczarzaki	Filasterea	Cowc
Ministeria vibrans	Filasterea	Mvib
Abeoforma whisleri	Teretosporea	Awhi
Creolimax fragrantissima	Teretosporea	Cfra
Pirum gemmata	Teretosporea	Pgem
Sphaeroforma arctica	Teretosporea	Sarc
Sphaerothecum destruens	Teretosporea	Sdes

Chromosphaera perkinsii	Teretosporea	Nk52
Syssomonas multiformis	Teretosporea	Smul
Corallochytrium limacisporum	Teretosporea	Clim
Fonticula alba	Discicristoidea	Falb
Nuclearia spp.	Discicristoidea	Nspp
Spizellomyces punctatus	Chytridiomycota	Spun
Batrachochytrium dendrobatidis	Chytridiomycota	Bden
Mortierella verticillata	incertae sedis	Mver
Rozella allomycis	Cryptomycota	Rall
Encephalitozoon cuniculi	Microsporidia	Ecun
Nematocida parisii	Microsporidia	Npar
Piromyces sp. E2	Neocallimastigomycota	Pisp
Catenaria anguillulae	Blastocladiomycota	Cang
Allomyces macrogynus	Blastocladiomycota	Amac
Rhizophagus irregularis	Glomeromycota	Rirr
Coemansia reversa	Kickxellomycotina	Crev
Conidiobolus coronatus	Entomophthoromycota	Ccor
Rhizopus oryzae	Mucoromycotina	Rory
Gonapodya prolifera	Monoblepharidomyocta	Gpro
Coprinopsis cinerea	Basidiomycota	Ccin
Cryptococcus neoformans	Basidiomycota	Cneo
Ustilago maydis	Basidiomycota	Umay
Neurospora crassa	Ascomycota	Ncra
Saccharomyces cerevisiae	Ascomycota	Scer
Schizosaccharomyces pombe	Ascomycota	Spom
Thecamonas trahens	Apusozoa	Ttra
Pygsuia biforma	Breviatea	Pbif
Acanthamoeba castellanii	Amoebozoa	Acas
Dictyostelium discoideum	Amoebozoa	Ddis
Entamoeba histolytica	Amoebozoa	Ehis
Polysphondylium pallidum	Amoebozoa	Ppal
Physarum polycephalum	Amoebozoa	Ppol
Perkinsus marinus	Alveolata	Pmar
Paramecium tetraurelia	Alveolata	Ptet
Symbiodinium minutum	Alveolata	Smin
Toxoplasma gondii	Alveolata	Tgon
Tetrahymena thermophila	Alveolata	Tthe
Aplanochytrium kerguelense	Heterokonta	Aker
Aurantiochytrium limacinum	Heterokonta	Alim
Ectocarpus siliculosus	Heterokonta	Esil
Phytophthora infestans	Heterokonta	Pinf
Thalassiossira pseudonana	Heterokonta	Tpse
Arabidopsis thaliana	Viridiplantae	Atha
Brachypodium distachyon	Viridiplantae	Bdis
Cyanidioschyzon merolae	Viridiplantae	Cmer
Cyanophora paradoxa	Viridiplantae	Cpar
Chlamydomonas reinhardtii	Viridiplantae	Crei
Chlorella variabilis	Viridiplantae	Cvar
Micromonas pusilla	Viridiplantae	Mpus
Ostreococcus tauri	Viridiplantae	Otau
Physcomitrella patens	Viridiplantae	Ppat
Selaginella moellendorffii	Viridiplantae	Smoe
Volvox cartieri	Viridiplantae	Vcar
Bigelowiella natans	Rhizaria	Bnat
Reticulomyxa filosa	Rhizaria	Rfil
Emiliania huxleyi	Haptophyta	Ehux
Guillardia theta	Cryptophyta	Gthe
Bodo saltans	Excavata	Bsal

Leishmania major	Excavata	Lmaj
Naegleria gruberi	Excavata	Ngru
Trypanosoma cruzi	Excavata	Tcru
Trichomonas vaginalis	Excavata	Tvag
Nutomonas longa	Ancyromonadida	Nlon

Supplementary Table 5. Blast identity of SAGs 18S ribosomal sequences againts TARA oceans OTUs

SAG	OTU_code	Id (%)	N° of missmatches
UC1	9b38ebc15ad3400e51b8fdcb3be290e6	100	0
UC2	ee0099358d26055b2c572ea2605cdc05	100	0
UC3	8da960f4921d51c0e3c8b2fdea2f6330	99.2	1
UC4	1e98f1edca3c7471c2f8fa0b3a12cee8	99.2	1

Supplementary Table 6. Sumary of the sequences used for 18S ribosomal gene phylogeny

Nickname	Taxa	SSU
	Choanoflagellatea	
CHO_Aung	*Acanthocorbis unguiculata* (Thomsen) Hara et Takahashi	HQ026764
CHO_Aspe	*Acanthoeca spectabilis* Ellis (ATCC PRA-103)	KT757415
CHO_Cper	*Choanoeca perplexa* Ellis (ATCC 50453)	KT757437
CHO_Cbot	*Codosiga botrytis* (Ehrenberg 1838) Stein 1878	JF706243
CHO_Cnat	*Calliacantha natans*	KU587842
CHO_Clon	*Calliacantha longicaudata*	KU587840
CHO_Chol	*Codosiga hollandica Carr, Richter and Nitsche (ATCC PRA-388)*	KT757430
CHO_Csp1	*Codosiga sp. M1/pIIp*	JF706237
CHO_Csp2	*Codosiga sp. M2/Morocco*	JF706236
CHO_Csp3	*Codosiga sp. M3/Mvid*	JF706242
CHO_Csp5	*Codosiga sp. M5/Iceland*	JF706239
CHO_Desp	*Desmarella sp.*	AF084231
CHO_Dgra	*Diaphanoeca grandis Ellis (ATCC 50111)*	KT757448
CHO_Dsph	*Diaphanoeca sphaerica*	KU587846
CHO_Disp	*Diaphanoeca sp.*	HQ237460
CHO_Dcos	*Didymoeca costata (Valkanov) Doweld (ATCC PRA-389)*	KT757444
CHO_Hbal	*Hartaetosiga balthica (Wylezich et Karpov) Carr, Richter and Nitsche (ATCC 50964)*	KT757421
CHO_Hgra	*Hartaetosiga gracilis (Kent) Carr, Richter and Nitsche (ATCC 50454)*	KT757426
CHO_Hmin	*Hartaetosiga minima (Wylezich et Karpov) Carr, Richter and Nitsche*	JQ034422
CHO_Hnan	*Helgoeca nana Leadbeater (ATCC 50073)*	KT757452
CHO_Mroa	*Microstomoeca roanoka (ATCC 50931) Carr, Richter and Nitsche*	KT757502
CHO_Mbre	*Monosiga brevicollis Ruinen (ATCC 50154)*	AF084618
CHO_Mflu	*Mylnosiga fluctuans Carr, Richter and Nitsche (ATCC 50635)*	AF084230
CHO_Pmin	*Pleurasiga minima*	KU587849
CHO_Pped	*Parvicobicula pedunculata* Leadbeater	HQ026765
CHO_Prey	*Pleurasiga reynoldsii*	KU587851
CHO_Pdic	*Polyoeca dichotoma* Kent (*Calliacantha* sp. CEE-2003)	AF272000
CHO_Scal	*Salpingoeca calixa* Carr, Richter and Nitsche	KT757470
CHO_Sdol	*Salpingoeca dolichothecata* (ATCC 50959) Carr, Richter and Nitsche	KT757472
CHO_Seur	*Salpingoeca euryoecia* Jeuck, Arndt & Nitsche	KJ631038

CHO_Sfus	*Salpingoeca fusiformis* Kent	KJ631044
CHO_Shel	*Salpingoeca helianthica* (ATCC 50153) Carr, Richter and Nitsche	KT757487
CHO_Sinf	*Salpingoeca infusionum* Kent (ATCC 50559)	KT757477
CHO_Slon	*Salpingoeca longipes* Kent	KJ631046
CHO_Smac	*Salpingoeca macrocollata* (ATCC 50938) Carr, Richter and Nitsche	KT757482
CHO_Soah	*Salpingoeca oahu* Carr, Richter and Nitsche	KT757492
CHO_Spun	*Salpingoeca punica* (ATCC 50788) Carr, Richter and Nitsche	KT757460
CHO_Skve	*Salpingoeca kvevrii* (ATCC 50929) Carr, Richter and Nitsche	KT757494
CHO_Sros	*Salpingoeca rosetta* King (ATCC 50818)	EU011924
CHO_Stub	*Salpingoeca tuba* Kent	HQ026774
CHO_Surc	*Salpingoeca urceolata* Kent (ATCC 50560)	KT757514
CHO_Sven	*Salpingoeca ventriosa* Jeuck, Arndt and Nitsche	KJ631041
CHO_Spar	*Savillea parva* Norris (ATCC PRA-391)	KT757467
CHO_Slep	*Sphaeroeca leprechaunica* Jeuck, Arndt & Nitsche	KJ631047
CHO_Svol	*Sphaeroeca volvox* Lauterborn	Z34900
CHO_Spyr	*Stagondoeca pyriformis* Carr, Richter and Nitsche	KT757499
CHO_Sarn	*Stephanoeca arndtii* Nitsche	JX069943
CHO_Saph	*Stephanoeca apheles* Thomsen	EF523336
CHO_Scau	*Stephanoeca cauliculata* Leadbeater	HQ026766
CHO_Sdip	*Stephanoeca diplocostata* Ellis (ATCC PRA-392)	KT757508
CHO_Snor	*Stephanoeca norrisii* Thomsen	HQ026768
CHO_Spau	*Stephanoeca paucicostata* Throndsen	HQ026769
DQ995807_Lagenoeca_artica	*Lagenoeca artica* Nitsche 2007	DQ995807
	Uncultured Holozoans	
HQ219444_FRESCHO1	FRESCHO1 del Campo 2013	HQ219444
AY821948_FRESCHO4	FRESCHO4 del Campo 2013	AY821948
AY821949_Ukn	FRESCHO3 del Campo 2013	AY821949
GU647170_CladeL	Clade L Weber 2012	GU647170
EF024885_CladeL	Clade L Weber 2012	EF024885
GU825407_ChoanoflagellateE	MACHO1 del Campo 2013	GU825407
JQ223245_Ukn	Unassinged Acanthoecida del Campo 2013	JQ223245
DQ104587_FRESCHO1	FRESCHO1 del Campo 2013	DQ104587
FJ410610_FRESCHO2	FRESCHO2 del Campo 2013	FJ410610
GU647190_FRESCHO2	FRESCHO2 del Campo 2013	GU647190
AY426842_Lagenoeca_ENV	Lagenoeca del Campo 2013	AY426842
FJ176220_1_MAOP_1	MAOP1 del Campo 2013	FJ176220
GU825148_MAOP1	MAOP1 del Campo 2013	GU825148
GU385597_1_MAOP_2	MAOP2 del Campo 2013	GU385597

GU824782_1_MAOP2	MAOP2 del Campo 2013	GU824782
AB191435_1_MAIP1_Ukn	MAIP1 del Campo 2013	AB191435
HQ219425_1_FRESHIP1_Ukn	FRESHIP1 del Campo 2013	HQ219425
DQ244007_1_FRESHIP2_Ukn	FRESHIP2 del Campo 2013	DQ244007
UC1	UC1 Clade 1 Craspedida Lopez-Escardo 2017	XXXXX
UC2	UC2 Acanthoecidae Lopez-Escardo 2017	XXXXX
UC3	UC3 Stephanocidae Lopez-Escardo 2017	XXXXX
UC4	UC4_Early branching Acanthoecida Lopez-Escardo 2017	XXXXX
	Filasterea	
Opistho_2_18S	*Pigoraptor chileana* Hehenberger 2017	MF190553
LAB0002_Opistho1	*Pigoraptor vitenamica* Hehenberger 2017	MF190552
AF436888_1_Capsaspora_owczarzaki	*Capsaspora owczarzaki*	AF436888
AF271998_1_Ministeria_vibrans	*Ministeria vibrans*	AF271998
	Ichthyosporea	
Apar	*Amoebidium parasiticum* Cienkowski	Y19155
Ihof	*Ichthyophonus hoferi* Plehn & Mulsow	U25637
LAB0004_NK52	*Chromosphaera perkinsii* Grau-Bové 2017	XXXXX
FN996945_1_Sphaerothecum_destruens	*Sphaerothecum destruens*	FN996945
AY363958_1_Anurofeca_sp_LAH_2003	*Anurofeca sp.*	AY363958
AY336701_1_Eccrinales	*Eccrinales*	AY336701
EU124916_1_Creolimax_fragrantissima	*Creolimax fragrantissima*	EU124916
GU810144_1_Pirum_gemmata	*Pirum gemmata*	GU810144
GU810145_1_Abeoforma_whisleri	*Abeoforma whisleri*	GU810145
Y16260_2_Sphaeroforma_arctica	*Sphaeroforma arctica*	Y16260_2
AF533941_1_Dermocystidium_percae	*Dermocystidium percae*	AF533941
AY372365_1_Rhinosporidium_sp	*Rhinosporidium sp.*	AY372365
	Plurimorfea	
L42528_1_Corallochytrium_limacisporum	*Corallochytrium limacisporum*	L42528
LAB0003_Colp12	*Syssomonas multiformis* Hehenberger 2017	MF190551
	Metazoa	
Bova	*Beroe ovata* Mayer	AF293694
Hasp	*Halichondria* sp.	AY737639
Hlsp	*Haliclona* sp.	AY734450
Lesp	*Leucosolenia* sp.	AF100945
Nvec	*Nematostella vectensis* Stephenson	AF254382
Susp	*Suberites* sp.	AF100947
Sysp	*Sycon* sp.	AM180970
Tadh	*Trichoplax adhaerens* von Schultze	AY652581
AF102892_Acoela_Paratomella_rubra	*Paratomella rubra*	AF102892
AY040680_Acanthobdella_peledina	*Acanthobdella peledina*	AY040680
U49909_Milnesium_tardigradum	*Milnesium tardigradum*	U49909
AJ228794_Pseudoceros_tritriatus	*Pseudoceros tritriatus*	AJ228794
EU368616_Ctenolepisma_longicaudata	*Ctenolepisma longicaudata*	EU368616
D14357_Antedon_serrata	*Antedon serrata*	D14357

D14359_Balanoglossus_carnosus	*Balanoglossus carnosus*	D14359
AF120533_Mollusca_Lima_lima	*Lima lima*	AF120533
AY049861_Urobatis_jamaicensis	*Urobatis jamaicensis*	AY049861
	Holomycota	
LAB0001_Parvularia_sp_ATCC50694	*Parvularia atlantis* López-Escardó 2017	KY113120
AB433328_Nuclearia_thermophila	*Nuclearia thermophila* Yoshida 2009	AB433328
FJ816018_Fonticula_alba	*Fonticula alba*	FJ816018
AY546684_Spizellomyces_punctatus	*Spizellomyces punctatus*	AY546684
DQ536481_1_Cyllamyces_aberensis	*Cyllamyces aberensis*	DQ536481
KC673103_1_Malassezia_globosa	*Malassezia globosa*	KC673103

Supplementary Table 7. Sumary of the taxa and the genes used in the eight-gene phylogeny. All the sequences are available at Figshare.

Nickname	Taxa	SSU	LSU	hsp90	tubA	EFL	EF-1A	Actin	hsp70
	Choanoflagellatea								
Aung	*Acanthocorbis unguiculata* (Thomsen) Hara et Takahashi	HQ026764	-	-	-	-	-	-	-
Aspe	*Acanthoeca spectabilis* Ellis (ATCC PRA-103)	KT757415	KT757416	KT757419	KT757420	KT757418	KT757417	comp16397_c1_seq5_1_7.8e257	comp16977_c0_seq1_1_2.7e292
Cper	*Choanoeca perplexa* Ellis (ATCC 50453)	KT757437	KT757438	KT757435	KT757439	KT757434	-	comp25775_c0_seq1_1_3.5e195	comp12883_c0_seq1_1_0
Cbot	*Codosiga botrytis* (Ehrenberg 1838) Stein 1878	JF706243	KT757422	-	-	-	HQ896019	HQ896017.1	HQ896020
Chol	*Codosiga hollandica* Carr, Richter and Nitsche (ATCC PRA-388)	KT757430	KT757431	KT757433	KT757436	-	KT757432	comp52078_c1_seq1_1_9.6e172	comp56372_c0_seq1_1_0
Csp1	*Codosiga* sp. M1/pIIp	JF706237	KT757440	-	-	-	-	-	-
Csp2	*Codosiga* sp. M2/Morocco	JF706236	KT757441	-	-	-	-	-	-
Csp3	*Codosiga* sp. M3/Mvid	JF706242	KT757442	-	-	-	-	-	-
Csp5	*Codosiga* sp. M5/Iceland	JF706239	KT757443	-	-	-	-	-	-
Desp	*Desmarella* sp.	AF084231	-	-	-	-	-	-	-
Dgra	*Diaphanoeca grandis* Ellis (ATCC 50111)	KT757448	EU011939	KT757450	KT757451	KT757449	KT768098	comp33416_c0_seq1_1_4.2e238	comp25564_c0_seq1_1_5.7e268
Disp	*Diaphanoeca* sp.	HQ237460	-	-	-	-	-	-	-
Dcos	*Didymoeca costata* (Valkanov) Doweld (ATCC PRA-389)	KT757444	EU011938	KT757446	KT757447	KT757445	-	comp16066_c1_seq2_1_1.5e138	comp16248_c0_seq2_1_3.1e245 1920
Hbal	*Hartaetosiga balthica* (Wylezich et Karpov) Carr, Richter and Nitsche (ATCC 50964)	KT757421	KT988065	KT757424	KT757425	KT757423	-	comp2182_c0_seq1_1_6.5e253	comp9438_c2_seq1_1_4.7e296

Hgra	*Hartaetosiga gracilis* (Kent) Carr, Richter and Nitsche (ATCC 50454)	KT757426	EU011935	KT757428	KT757429	KT757427	-	comp11731_c0_seq1_1_1.3e257	comp13062_c2_seq4_substitute
Hmin	*Hartaetosiga minima* (Wylezich et Karpov) Carr, Richter and Nitsche	JQ034422	JQ034423	-	-	-	-	-	-
Hnan	*Helgoeca nana* Leadbeater (ATCC 50073)	KT757452	KT757453	KT768096	KT768097	KT757455	KT757454	comp25775_c0_seq1_1_3.5e195	comp27130_c3_seq9_1_9.3e285
Mroa	*Microstomoeca roanoka* (ATCC 50931) Carr, Richter and Nitsche	KT757502	KT757503	KT757505	KT757506	KT757504	-	comp15624_c0_seq2_1_2.5e257	comp16714_c3_seq1_1_0
Mbre	*Monosiga brevicollis* Ruinen (ATCC 50154)	AF084618	KT757456	AY226081	AY026070	AY026073	-	37852_1_5.2e269	35238_1_0
Mflu	*Mylnosiga fluctuans* Carr, Richter and Nitsche (ATCC 50635)	AF084230	EU011940	KT757458	KT757459	-	KT757457	comp10041_c0_seq1_1_4.9e219	comp10254_c0_seq1_1_0
Pped	*Parvicobicula pedunculata* Leadbeater	HQ026765	-	-	-	-	-	-	-
Pdic	*Polyoeca dichotoma* Kent (*Calliacantha* sp. CEE-2003)	AF272000	-	-	-	-	-	-	-
Scal	*Salpingoeca calixa* Carr, Richter and Nitsche	KT757470	KT757471	-	-	-	-	-	-
Sdol	*Salpingoeca dolichothecata* (ATCC 50959) Carr, Richter and Nitsche	KT757472	KT757473	KT757475	KT757476	-	KT757474	comp26491_c0_seq1_1_2.9e179	comp26705_c0_seq1_1_2.8e275
Seur	*Salpingoeca euryoecia* Jeuck, Arndt & Nitsche	KJ631038	-	-	-	-	-	-	-
Sfus	*Salpingoeca fusiformis* Kent	KJ631044	-	-	-	-	-	-	-
Shel	*Salpingoeca helianthica* (ATCC 50153) Carr, Richter and Nitsche	KT757487	KT757488	KT757490	KT757491	-	KT757489	comp15346_c0_seq1_1_1.9e266	comp15440_c5_seq1_1_5e192
Sinf	*Salpingoeca infusionum* Kent (ATCC 50559)	KT757477	KT757478	KT757480	KT757481	KT757479	-	comp14544_c0_seq3_1_3.8e262	comp11421_c0_seq1_1_0

Slon	*Salpingoeca longipes* Kent	KJ631046	-	-	-	-	-	-	-
Smac	*Salpingoeca macrocollata* (ATCC 50938) Carr, Richter and Nitsche	KT757482	KT757483	KT757485	KT757486	KT757484	-	comp22230_c1_seq2_1_1.6e254	comp14111_c0_seq1_1_0
Soah	*Salpingoeca oahu* Carr, Richter and Nitsche	KT757492	KT757493	-	-	-	-	-	-
Spun	*Salpingoeca punica* (ATCC 50788) Carr, Richter and Nitsche	KT757460	KT757461	KT757464	KT757469	KT757463	KT757462	comp16274_c1_seq2_1_9.8e253	comp16816_c2_seq2_1_1.5e238
Skve	*Salpingoeca kvevrii* (ATCC 50929) Carr, Richter and Nitsche	KT757494	EU011926	KT757497	KT757498	KT757496	-	comp10800_c0_seq1_1_3.3e247	comp11028_c1_seq14_1_0
Sros	*Salpingoeca rosetta* King (ATCC 50818)	EU011924	EU011941	KT757501	AY226048	XM_004996684	-	PTSG_01553T0_1_3.1e259	PTSG_10966T0_1_0
Stub	*Salpingoeca tuba* Kent	HQ026774	KT757507	-	-	-	-	-	-
Surc	*Salpingoeca urceolata* Kent (ATCC 50560)	KT757514	KT757515	KT757517	KT757516	KT757518	-	comp19597_c0_seq4_1_1.3e249	comp17673_c0_seq6_1_1.3e298
Sven	*Salpingoeca ventriosa* Jeuck, Arndt and Nitsche	KJ631041	KT757519	-	-	-	-	-	-
Spar	*Savillea parva* Norris (ATCC PRA-391)	KT757467	KT757495	KT757465	KT757466	KT757468	-	comp11563_c0_seq1_1_7.6e253	comp11748_c0_seq1_1_5.2e297
Slep	*Sphaeroeca leprechaunica* Jeuck, Arndt & Nitsche	KJ631047	-	-	-	-	-	-	-
Svol	*Sphaeroeca volvox* Lauterborn	Z34900	-	-	-	-	-	-	-
Spyr	*Stagondoeca pyriformis* Carr, Richter and Nitsche	KT757499	KT757500	-	-	-	-	-	-
Sarn	*Stephanoeca arndtii* Nitsche	JX069943	-	-	-	-	-	-	-
Saph	*Stephanoeca apheles* Thomsen	EF523336	-	-	-	-	-	-	-
Scau	*Stephanoeca cauliculata* Leadbeater	HQ026766	-	-	-	-	-	-	-

Sdip	*Stephanoeca diplocostata* Ellis (ATCC PRA-392)	KT757508	KT757509	KT757512	KT757513	KT757511	KT757510	comp19908_c2_seq1_1_8.2e131	comp18193_c0_seq1_1_1.4e275
Snor	*Stephanoeca norrisii* Thomsen	HQ026768	-	-	-	-	-	-	-
Spau	*Stephanoeca paucicostata* Throndsen	HQ026769	-	-	-	-	-	-	-
	Undescribed Choanoflagellatea								
UC1		XXXXX	XXXXX	-	-	-	XXXXX	XXXXX	XXXXX
UC2		XXXXX	XXXXX	-	-	-	-	XXXXX	-
UC3		XXXXX	XXXXX	-	-	-	-	-	-
UC4		XXXXX	XXXXX	-	-	XXXXX	-	XXXXX	XXXXX
	Ichthyosporea								
Apar	*Amoebidium parasiticum* Cienkowski	Y19155	EU011932	Apar_MIRA_nonfilt_contigs.fasta Apar_ReadsTotals_rep_c25415	-	-	AY582828	-	-
Ihof	*Ichthyophonus hoferi* Plehn & Mulsow	U25637	AY026370	-	-	-	AF450116 (*I. irregularis*)	-	-
	Metazoa								
Bova	*Beroe ovata* Mayer	AF293694	AY026369	-	-	-	-	-	-
Hasp	*Halichondria* sp.	AY737639	-	AY226083	AY226049	-	GQ330929	-	-
Hlsp	*Haliclona* sp.	AY734450	AF441340	AY226084	AY226050	-	JQ606691	-	-
Lesp	*Leucosolenia* sp.	AF100945	AY026372	AY226087	AY226053	-	DQ087468	-	-

Nvec	*Nematostella vectensis* Stephenson	AF254382	Genome GCA_000209225.1 contig NEMVEscaffold_301 dna:supercontig:GCA000209225.1:NEMVEscaffold_310:1:244201:1	AY226090	AY226056	-	XM_001625261	-	195315_1_3.4e258
Susp	*Suberites* sp.	AF100947	AY026381	AY226085	AY226051	-	GQ330984	-	-
Sysp	*Sycon* sp.	AM180970	AM181000	AY226088	AY226054	-	DQ087461	-	-
Tadh	*Trichoplax adhaerens* von Schultze	AY652581	AY303975	XM_002113141	XM_002110626	-	NW_002060962	63375_1_6.9e197	38294_1_1.5e216

4. DISCUSSION

4.1 Seeking new Metazoa diversity: A metabarcoding approach

Metazoa is the eukaryotic kingdom with more species described so far, around 1.3 million (del Campo et al. 2014), and there are estimations that suggest that on Earth it might be up to 10 million of animal species (Blaxter et al. 2005). This indicates that there is an extensive unknown metazoan diversity, mostly from microbial animals (animals smaller than 2mm) (Blaxter et al., 2005). In the chapter 3.2 I described new molecular metazoan diversity by analysing metabarcoding data from six different locations across the European coast. The sampling in each location included both water column and sediments, oxic and anoxic environments, and both DNA and RNA templates through different size fractions (picoplankton 0.8-3 um; nanoplankton 3-20µm; micro-mesoplankton 20-2000µm).

a) Metabarcoding, a potential approach to obtain micrometazoans diversity patterns

The results showed that the general read abundance and richness patterns partially corroborate previous morphological (Humes 1994; Snelgrove 1999; Grégory Beaugrand et al. 2003; Bouquet et al. 2009) and molecular studies (Chariton et al. 2010; Fonseca et al. 2010; Lindeque et al. 2013; Pearman et al. 2014; Pearman and Irigoien 2015). For instance, Copepods were the most abundant organisms within the water column followed by the Appendicularians (tunicates) and Chaetognaths. In addition, the richest clades were Crustacea and Nematoda and, within the benthic environments, the relative abundance was more distributed among these two groups (Crustacea and Nematoda) together with Polychatea, Platyhelminthes and Mollusca. Overall, in agreement with many studies based on morphological identification of zooplankton and benthic communities (Humes 1994; Snelgrove 1999; G Beaugrand, Brander, and Lindley 2003; Bouquet et al. 2009).

Nevertheless, metabarcoding analyses have some drawbacks. For example, it is well known that it may be misleading to directly translate reads and OTU numbers into biomass and number of species, respectively. In particular, the use of amplicon data as a

proxy for metazoan biomass abundance has been disputed, also with RNA data (Pawlowski et al. 2014). Different number of rRNA copies in the genomes of different taxa, PCR primer mismatches, problems in DNA/RNA extraction of some taxa, biases that affect the correlation between morphological and molecular data (Porazinska et al. 2010; Hirai et al. 2015).
Despite these methodological limitation, some studies have indeed shown positive correlations between read abundances and biomass patterns in bivalve and decapod larvae (Lindeque et al. 2013) and within copepod groups (Hirai et al. 2015). My work, as well, have shown a relative abundance patterns similar to what has been found in morphological studies, specially on zooplanktonic communities. However, we do not have morphological data accompanying the metabarcoding analysis, thus we can not compare both techniques, and have a control of the metabarcoding approach.
Therefore, our results confirmed that, despite some caveats, HTES is a powerful tool to assess microbial metazoan diversity. I consider that the findings of our study as well as others (Hirai et al. 2015) should encourage the scientific community to work in standarizing protocols to monitor, in an easy way, changes of metazoan microbial diversity in diverse environments, specially for those taxa that have a wider impact on the ecosystem. For instance, in our analysis, 20 OTUs represented round of 80% of planktonic metazoan reads. Monitoring these 20 OTUs would provide lots of valuable ecological information in a faster and easier way. This is very relevant nowadays given that the increasing of pollution and carbon emissions are affecting marine microbial fauna (Uriarte and Villate 2004).

b) A potential ecological role of metazoan gametes?

Another interesting insight from my results is that metabarcoding data from studies that pretended to unveil the diversity of unicellular eukaryotes, as it is the case of ours, can also be used to reveal patterns of diversity of microbial animals. This is important given that there are tons of metabarcoding information being generated, especially thanks to megaprojects such as the TARA oceans expedition (de Vargas et al. 2015) or the Biomarks project. In our case, this kind of data allowed us to reveal a potential hidden ecological role of metazoan gametes, concretely among

ctenophores, cnidarians and polychaetes. In particular, we found that their eukaryotic relative abundance of RNA reads in pico/nanoplanktonic fractions (smaller than 5μm) within oxic and anoxic environments were respectively of 3.5% and a 33% of the total reads. Thus, their numbers are comparable to those from unicellular heterotrophic flagellates, which usually reach abundances of up to the 40% of eukaryotic RNA reads in pico/nanoplanktonic fractions (Logares et al. 2012) (Logares et al., 2012). Obviously, further research is needed to assess the effect of animal sperm in microbial nutrient fluxes, but our results suggest that these effects may be particularly relevant during spawning events, according to the high relative abundance of the ctenophore *Pleurobrachia pileus* on the Black sea (33%), and the fact that the samples were collected during *Pleurobrachia* reproductive season (Mutlu Bingel, F. 1999). Therefore, metazoan gametes may represent passive members of the community eaten by other metazoans or protists and participating in the carbon fluxes of microbial communities. A good way to validate this hypothesis is by designing specific metazoan oligonucleotide proves to be used in FISH experiments (see section 1.2a), in order to see the abundance and distribution of metazoan gametes across pico- or nano-planktonic marine samples.

c) Unveiling new metazoan diversity and its potential to better understand animal evolution

I determined the levels of novelty in our dataset by looking the BLAST identities of our OTUs against the NCBI database. I considered that identity values below 97% represented «novel» sequences compared to the ones available at Genbank, as I applied a clustering threshold of 97% to generate our OTUs. The OTUs clustering at 97% is likely a conservative approach for metazoans (Tang et al. 2012), and some of our OTUs may indeed represent more than one species. This largely depends on each metazoan lineage and its specific 18S rRNA evolution rate. Moreover, primer bias can affect the detection of some groups, meaning that some taxa can be present in the environment but missing in our dataset (Creer et al. 2010). However, by clustering at 97% I can directly compare the results with the rest of the eukaryotes and get a more

stringent output avoiding polymorphisms effects (Stoeck et al. 2010) and an overrepresentation of the retrieved diversity.
Following this novelty criterion, there were 36% of novel OTUs, with 10% of our OTUs having a BLAST identity under 90%. This means that a third part of our OTUs were new and a 10% were very divergent compared to the reference sequences and perhaps, those sequences belong to new metazoan clades. Interestingly, this new metazoan diversity was not randomly distributed. Novel OTUs were mostly present in benthic habitats and were often located only in one sampling site. Also they tend to be as well less abundant (35.5% of the OTUs representing 10.5% of the number of reads). Thus, benthic environments can be a hot-spot to identify new metazoan diversity, which appears to not be very abundant and restricted to local areas.

The animal groups with the higher levels of novelty were crustaceans, nematodes, platyhelminthes, gastrotrichs and acoelomorphs. With the exception of crustaceans, these groups occupied early branching phylogenetic positions within the Ecdysozoa or the Lophotrochoa/Spiralia, or even within the Bilateria (Telford 2013). Moreover, the high genetic diversity in often neglected groups such as Acoelomorpha (Arroyo et al. 2016) and Gastrotricha (Chariton et al. 2010) revealed that these groups are needed for a deeper exploration.

Thus, taking into account that the evolutionary relationships among animal phyla are not fully solved (Giribet 2015) (see introduction, section 1.3f), the discovery of new hidden metazoan diversity occupying key phylogenetic positions, might be the first exploratory step to end up with the isolation, characterization (Markmann and Tautz 2005), and genome sequencing of key animal species, which can help to better reconstruct the metazoan tree of life (Arroyo et al., 2016).

In most of the phylogenetic questions regarding animal evolution, it seems easier to obtain metazoan genomic information from already described animal species that fall into key phylogenetic positions. However, metabarcoding analysis permits to know how well-sampled are the different metazoa phyla, and may unravel interesting novel metazoan taxa. If these interesting hidden

metazoan taxa can solve important phylogenetic issues, the efforts of identification and isolation of those species will be rewarded.

Regarding our data, the novel OTUs from the groups mentioned above (crustaceans, nematodes, platyhelminthes, gastrotrichs and acoelomorphs), if someday they are isolated and characterized, they would help to better solve the internal phylogenies of these groups. In addition, new Acoelomorpha and Nematoda diversity might also help to address the root of bilaterians (Arroyo et al. 2016) and the evolutionary relationships among Ecdysozoa. On the other hand, our results suggest that chaetognaths are a well-sampled group in terms of 18S ribosomoal gene sequences available, because its novelty values were very low (most of the OTUS a blast identity >97%). Thus, in order to solve its controversial phylogenetic position within Protostomia (Giribet 2015; X. Shen et al. 2016), the scientific community would have to rely only in the current described species and its genomic sequences.

d) A new molecular defined group of metazoans: MAME 1, a case of reverse taxonomics? The need to link morphological and molecular data

Finally, I described a new metazoan group within Tunicates, the MAME 1 clade, composed by 69 OTUs (three of them found in our dataset) and that contains high genetic diversity. In addition, this potential new group of tunicates is abundant and widespread around the world and phylogenetically related with thaliaceans, according to our results.

The identification of MAME 1 group might be the first step in a reverse taxonomic process (Markmann and Tautz 2005), potentially leading to the isolation and detailed description of a new metazoan lineage. However, before considering different strategies to isolate and morphologically described this new group, we have to interrogate ourselves if this molecular diversity truly belongs to a new animal clade. The grouping could be artifactual, however. For example, it can be the result of very rare events of intragenomic variability from a known Tunicate clade, although I consider this unlikely because there available at Genbank round of 1,000 tunicate

18S ribosomal sequences, and it has not been reported yet any evidence of deep intragenomic variability within tunicate ribosomal sequences. In addition, the identity values of MAME 1 sequences compared with available tunicate species is much lower than the two types of 18S rRNA sequences found in the same individuals of Platyhelminthes and Chaetognatha taxa, which poses events of deep intragenomic ribosomal variability (Carranza et al. 1996; Gasmi et al. 2014). In addition, this molecular diversity can also belong to a group of tunicates previously described but without molecular data available.

In any case, I believe further studies should be done on this potential new group of tunicates because molecularly it is quite different from other tunicate species. The 18S phylogenic tree shows them as a long-branch thaliacean, which breaks the monophyly of the group (Tsagkogeorga et al. 2009). Therefore having their full genomic sequence together with the genomic information of other tunicates will increase our comprehension of tunicate evolution and diversity.

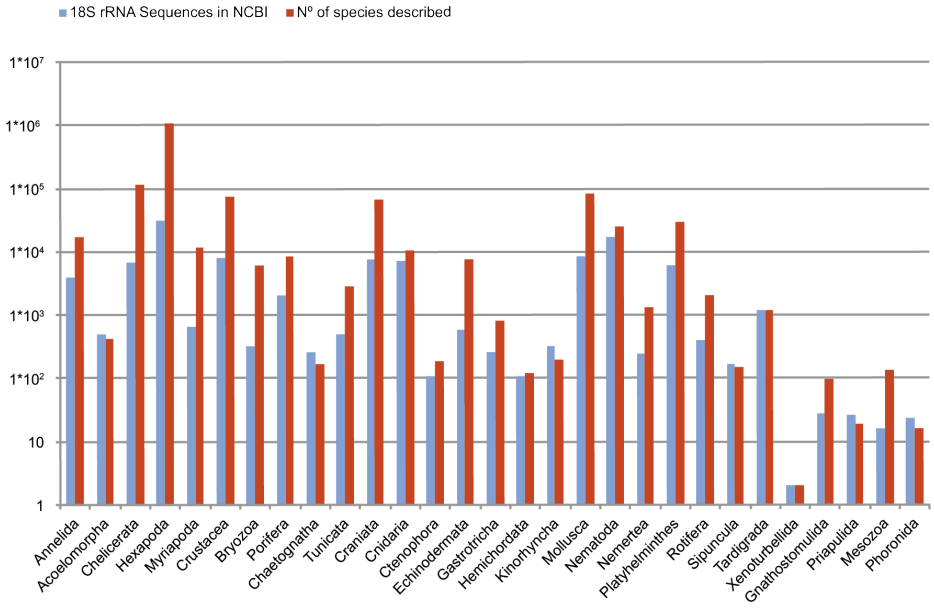


Figure 15. Comparison of the number of described metazoan species and the number of 18S rRNA metazoan sequences in NCBI. Total number of OTUs from 18S rRNA retrieved from Genbank (blue bars) compared to the number of

described species for each metazoan phylum (red bars). The number of 18S rRNA sequences from NCBI was obtained from the following search for each phylum: "*txid33208[Organism:exp] (18S OR SSU) NOT (mitochondrial OR mitochondria)"*.

It is worth taking into account that there are more morphologically described species than 18S ribosomal genes sequences available in animals (Fig. 15). Thus, this fact limits our proper interpretation of the results and shows the need to have linked morphological information with at least a molecular marker, as the 18S ribosomal gene. Accomplishing that is a major challenge for the scientific community, but I hope that my works and others will help to encourage scientist to move in that direction. Actually in the conference Biogenomics2017 (biogenomics2017.org) it was proposed to sequence the genomes of all living beings on earth, even though the proposal mostly concerned macroscopic organisms like bigger animals, plants and fungi.

Finally, the results obtained in the chapter 3.1 are an example of the problems by classifiying microbial taxa based solely in morphological characters. *Parvularia atlantis* (chapter 3.1) was formerly classified within the genus *Nuclearia*. On the other hand, previous to our results, it was described an uncultured nucleariid lineage based in environmental 18S ribosomal sequences, the FRESHOP (del Campo and Ruiz-Trillo 2013). Our results show that *Parvularia* is not a *Nuclearia* species, and its 18S sequence clusters with the previously described FRESHOP group. Thus now, FRESHOP sequences belong to *Parvularia* genus. Showing how environmental sequencing is often the first step prior to morphological classification in protists (Gómez et al. 2011; Shiratori, Thakur, and Ishida 2017). In animals, there are not reported yet such cases. Hopefully, our metabarcoding analysis will predate the formally description of new metazoan taxa and will encourage the scientific community to do efforts in linking morphological and molecular information of microbial animals.

4.2. *Parvularia atlantis*: a new nucleriid taxa and its importance for Opisthokonta evolution

In section 3.1 we described a new Nucleariid genus and species: *Parvularia atlantis.* Nucleariids (Tom Cavalier-Smith 1993) are the earlier branching lineage of Holomycota (Liu et al. 2009), the group of Opisthokonts that includes also Fungi (Spatafora et al. 2016) and Opisthosporidia (S. A. Karpov et al. 2014) (see section 1.3b). Therefore, nucleariid amoebae are in a pivotal position as the sister-group to the rest of Holomycota, a lineage close to the root of Opisthokonts, which makes this group key for understanding the evolution of Opisthokonta from its last common ancestor to the different extant lineages. In this regard, it is essential to have a wide representation of nucleariids diversity with it is morphological characters described and classified behind a robust phylogenetic framework. Therefore, our work represents an starting point to better understanding Nucleariids and Opisthokonta evolution.

a) A newly described nucleariid genera

Parvularia is the third genera confirmed by molecular data comprised within nulcearíids amoeba together with *Nuclearia* and *Fonticula* (Matthew W. Brown, Spiegel, and Silberman 2009). We found within *Parvularia* more molecular diversity besides the sequenced 18S ribosomal subunit of *Parvularia atlantis*. Some of these environmental 18S rDNA sequences were described before as a potential nucleariid clade, named as FRESHOP (del Campo and Ruiz-Trillo 2013). Showing that, often, environmental 18S rDNA sequencing predates the formal description of a protistan group (see section 4.2d). Our results clearly show that *Parvularia* sequences are phylogenetically distinct from the genus *Nuclearia* and *Fonticula*. In addition, *Parvularia* sequences do not contain the insertions described in the V4 and V7 region of the 18S ribosomal gene, found on Nuclearia species. Besides, these phylogenetic differences, there are as well important morphological differences that distinguish *P. atlantis* from previous described nucleariid or nucleariid-related species. *P. atlantis* has an smaller size (round 4 μm) , can form cysts and can contain binucleated cells. All these combination of characters makes *Parvularia* unique among the currently described nucleariid filose amoeba species (Yoshida, Nakayama, and Inouye 2009).

b) Nucleariids diversity, many questions remain to be answered

Our phylogenetic inference of Nucleariid diversity show –besides *Fonticula*, *Nuclearia* and *Parvularia*– three environmental clades: MAFO (del Campo et al. 2015), env-NUC1 and env-NUC2. This means that there is an important fraction of Nucleariid diversity awaiting to be discovered and characterized at a morphological level. But, could this novel diversity correspond to already described protists? For example, there are many protistan genera without a clear taxonomic affiliation that morphologically are associated with Nulcearid amoebas, like *Vampyrellidium* (Surek and Melkonian 1980) or the Rotosphaerids Pinaciophora Greef, 1869, Pompholyxophrys Archer, 1869 and Lithocolla Shulze, 1874 (Mikrjukov 1999) (see section 1.3b). Thus, the nucleariid environmental clades could potentially correspond to these amoebas (*Vampyrellidium* and Rotosphaerids) that lack molecular information. For example, it is known that members of Pompholyxophryidae family, (Pinaciophora, Pompholyxophrys and Lithocolla) life on marine sediments (G. F. Esteban, Gooday, and Clarke 2007) and, interestingly, the environmental group MAFO was found to be particularly abundant in marine sediment samples (del Campo et al. 2015). Therefore, this can be also a hint that MAFO sequences correspond to members of Pompholyxophryidae family. However, without molecular data from Pompholyxophryidae members this question will remain unanswered. Thus, any attempts to better understand at all levels nucleariids diversity should start first to obtain molecular data from members of Pompholyxophryidae family.

c) A genomic approach to better understand Nucleariids and Opisthokonta evolution. A case of convergent evolution.

The phylogenetic inference based in the 18S ribosomal gene, although useful to analyse nuclariids diversity, fails in determining the specific relationships between the different nucleariid clades. Thus in order, to have better resolved the phylogenetic tree, a phylogenomic analysis is needed. To this, it is key to have genomic information from nucleariid lineages. So far, the only available genome among nucleariids is that from *Fonticula alba*

(http://protist.ensembl.org/Fonticula_alba/Info/Index), sequenced within UNICORN project initiative (Ruiz-Trillo et al. 2007). There is also the transcriptome of *P. atlantis*. However, there is still no genomic or transcriptomic data from any Nuclearia species. Therefore, it is still not possible right now to properly tackle the phylogenetic relationship between the three nucleariid described genera.

One of the major challenges to obtain genomic information from nucleariids and rotosphaerids is that that there are not cultures available in public repositories, given the complexity of isolating them from the environment (del Campo et al. 2014). Recently, some Nuclearia species have been isolated from the Swiss lakes (Dirren and Posch 2016; Dirren et al. 2017). However, *Fonticula* and *Parvularia* cultures grow with the presence of rod-shaped-bacteria and *Nuclearia* cultures feed on cyanobacteria, difficulting the recovery of non-contaminated DNA material for further genomic sequencing. On the positive side, nucleariids and rotosphaerids, except for *Parvularia* and *Fonticula*, have big cell sizes (between 10-65µm) (Mikrjukov 1999; Yoshida, Nakayama, and Inouye 2009), making these organisms easier to identify, micromanipulate and isolate from the environment compared to other unicellular opisthokonts. This easier micromanipulation of single-cells might facilitate a single-cell genomics approach (see section 4.4).

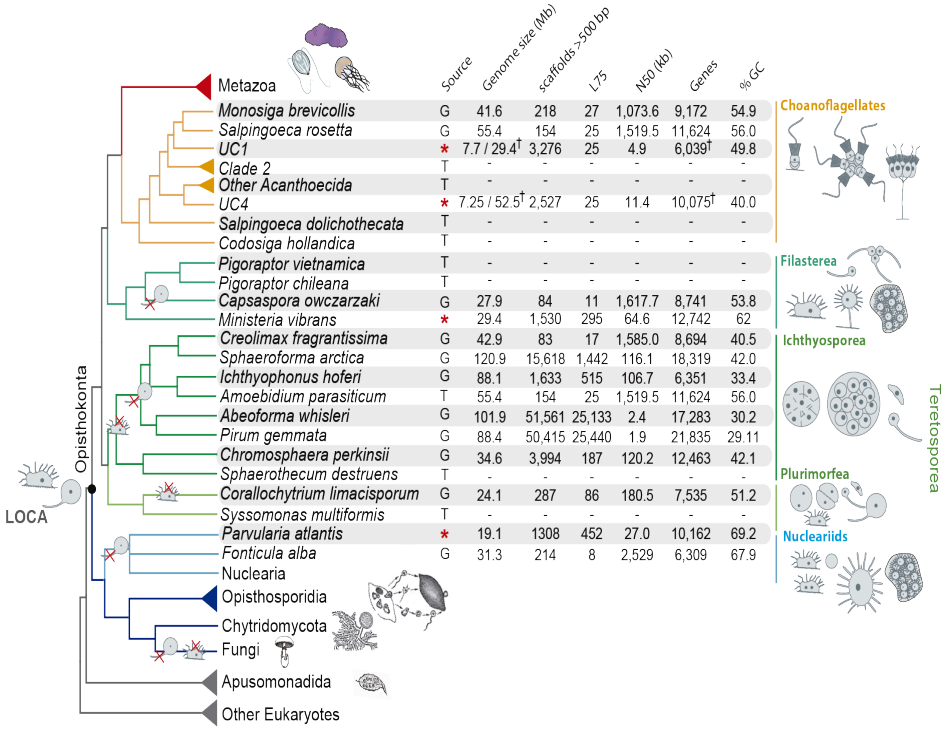


Figure 16. Genomic diversity available of unicellular Opisthokonts and their morphological characters. Schematic phylogenetic tree of eukaryotes, with a focus in Holozoa and Nucleariidae. The adjacent table summarize the genome statistics. Data sources: T, transcriptome; G, genome and red asterisk, a genome sequenced by the author of these thesis in collaboration with other researchers. The symbol † indicates that the value is a prediction of the total genome size and number of genes, only applied in the SAGs UC1 and UC4, due to their partial genomic sequences. On the right side, there are the morphological traits of the different Opisthokonta lineages, within the tree are indicated the losses of flagellum and filopodia from the Last Opisthokonta Common Ancestor (LOCA).

Anyway, it is clear that genomic information is not only important to solve the phylogenetic relationships among nucleariids. The genomes of nucleariids will also provide invaluable information in their biology, while allowing a more complete comparative genomics studies of the opisthokonts. Thus, nucleariids genomes together with the rest of Opisthokonta genomes will allow to trace the evolution of gene families from the Last Common Opisthokonta Ancestor (LOCA) to the extant species (Grau-Bové et al. 2017)

(Grau-Bové et al., 2017). The result of those analyses will provide candidate genes involved in the evolution of the vast phenotypic traits present within Opisthokonta (Fig. 16), including animal multicellularity.

Regarding the evolution of phenotypes, the ancestor of nucleariids lost the flagellum present in the LOCA, and adopted a filopodiated amoebae morphology. Nucleariids are able to predate on bacteria and perform their life-cycle without the need of a flagellum (Thomas Cavalier-Smith 2012). Interestingly, on the other side of the Opisthokonta tree, within Holozoa, there is a similar morphological adaptation. Within Filasterea, there are two filopodiated amoebas that have lost (*Capsaspora owczarzaki*) or are in process of losing (*Ministeria vibrans*) their flagellum (Torruella et al., 2015). In addition, those organisms also prey on bacteria as nucleariids. Indeed, this is also the case of *Capsaspora*, which was isolated from the hemolimph of snail and grows in a rich media (Stibbs, Owczarzak, Bayne, & DeWan, 1979), but can also grow in a poor media with rod-shaped bacteria (López-Escardó et al. Unpublished data). In addition, *C. owczarzaki* (Holozoa) has an aggregative behavior as the nucleariid *Fonticula alba* (Holomycota) (Sebé-Pedrós, Irimia, et al. 2013). Therefore, nucleariids and filastereans share many phenotypical characters, even though they are very distant from a phylogenetic point of view (Fig. 16).

Thus, and given this similarity between nucleariids and filastereans, it would be very interesting to search on their genomes for retention or losses of similar genes, as well as parallel expansions of gene families. Indeed, this is not new. It has already been described in Opisthokonta convergent evolution between holozoans (Teretosporea) and holomycotans (Fungi) (Torruella et al. 2015). In this regard, it is worth mentioning that during this thesis I have collaborated with other members of the lab in trying to obtain the genomes of *Parvularia* and *Ministeria* (Fig. 16). We succeeded, and we already have the whole genome sequenced, assembled and annotated. Those genomes will not only provide clues into convergent evolution, but also may lead to the identification of gene candidates for filose amoeba morphology, aggregative behaviour or bacterial feeding.

4.3. Further challenges in the expansion of Opisthokonta genomic diversity. The role of single-cell genomic technologies

At this point of the thesis, I hope that I have already convinced the readers about the importance of obtaining the widest possible morphological and genomic information of Opisthokonta. As commented before, many nucleariid genomes remain to be sequenced, as well as many other genomes within Opisthokonta that also remain unknown. In particular, within Holozoa there is no data on some environmental groups. We have, for example, FRESHIP and MACHO that belong to ichthyosporeans and choanoflagellates respectively (del Campo and Ruiz-Trillo 2013). However, there are other environmental groups that, in my opinion, deserve more attention because they fall in more interesting phylogenetic positions. In this section, I will discuss the groups of unicellular holozoans that I believe are key to significantly increase our understanding of Opisthokonta evolution, and also how single-cell genomics can help to unravel their genomic sequences.

a) Key Holozoan genomic diversity that remain unknown

The first lineage that I would like to mention is the MAOP, that stands for MArine OPisthokonts, and it is a paraphyletic clade divided in two groups, MAOP1 and MAOP2. They do not have a clear phylogenetic affiliation within holozoans (del Campo and Ruiz-Trillo 2013). Thanks to the addition of pigorapotors within filastereans, it seems that MAOP1 can be related with filastereans (Hehenberger et al. 2017) (Fig.1 within section 3.4). The phylogenetic position of MAOP2 is even more enigmatic, as it appears sister to choanoflagellates on 18S ribosomal gene phylogenies, in a similar position than *Syssomonas* or *Corallochytrium*. Thus MAOP2 may belong to plurimorfeans. What is clear is that MAOPs represent either a new clade within holozoan or they belong to filastereans and plurimorfeans, the holozoan lineages with less species described (4 and 2 species described, respectively), making MAOP an specially interesting group.

Other interesting environmental groups within holozoans are the choanoflagellate FRESCHO4, FRESCHO3 (del Campo and Ruiz-Trillo 2013) and Clade L (Weber et al. 2012), that fall in early-

branching positions sister to Acanthoecida (Fig.1 within section 3.4). Thus, they could help resolve the deeper nodes of choanoflagellates phylogeny. In addition, the sequences of those groups were obtained on freshwater or brackish water environments (del Campo and Ruiz-Trillo 2013; Weber et al. 2012). Given that practically all Acanthaoecida species were found in marine environments (Martin Carr et al. 2008), those groups might provide support to a putative freshwater choanoflagellate ancestor (see section 4.5). In addition, they might not contain the lorica of Acanthoecids and present a different new choanoflagellate covering. Thus, hopefully one day we will be able to know their morphologies and their genomic sequences.

Besides the environmental clades, there are several described species without genomic data that can be specially interesting given their phylogenetic position. This is the case of the choanoflagellates *Salpingoeca tuba* or the genera *Sphaeroca.* Both are related with the earliest branching choanoflagellate taxa found in our phylogenomic approach (section 3.4) –*Salpingoeca dolichothecata* and *Codosiga hollandica* respectively–. Therefore, as the deep nodes of of choanoflagellates phylogeny appear less supported (section 3.4), the addition of more taxa in these phylogenetic position would help to better reconstruct choanoflagellates phylogeny.

On the other hand, the ichthyosporean lineage Eccrinales has many species described spread in more than 17 genera and none of them have yet a genomic sequence available. Eccrinales live within the gut of arthropods in a wide variety of environments (Cafaro 2005). They present a similar morphology to fungi, with protusions that resemble fungal hyphal growth (see introduction 1.3c), therefore its genomic sequences can be interesting also to study another putative new case of convergent evolution betweem Eccrinales and Fungi.

b) Expanding the genomic diversity of microbial Opisthokonts: the role of single-cell genomics

Thus, after summarizing the most interesting candidates to expand the genomic information of Opisthokonts we can ask ourselves, how can be get those genomes? There is not an easy answer for that. The ideal situation would be isolating them from the environment

and establishing a culture, as has been done with the newly described *Pigoraptors* and *Syssomonas* (Hehenberger et al. 2017). However, this process is not straightforward and might not be possible for most taxa. Thus, the alternative is isolating the cells from an environmental sample through micromanipulation, FACS (Fluorescence activated cell sorting) or microfluidics devices and then, take advantage of single-cell genomics technologies.

To facilitate the use of single-cell genomics, metabarcoding studies should collect cells as a back up, for further cell isolation through FACS and single-cell genomics approaches. This was done in some samples of the TARA oceans expedition (de Vargas et al. 2015) (Tara Oceans Consortium, Coordinators; Tara Oceans Expedition, 2014). Because one of the main problems in working with unicellular opisthokonts is that they are not very abundant in the environment compared to the rest of eukaryotes (del Campo et al. 2015). Thus, the probability to find, for instance, a MAOP, in a random marine sample, is very low. However, if from any sample we also froze down cells, we can later, sent only to cell-sorting the sample in which metabarcoding data revealed the group of interest.

For instance, MAOPs, are not highly abundant. The maximum relative abundance that we found from a MAOP OTU on TARA oceans database was around 1% of all eukaryotes. That means, that even though, we could avoid photosynthetic cells (pigmented cells can be detected by FACS and discarded), we will need to generate many single-cell amplified genomes to be able to obtain enough number of MAOP cells, in order to obtain the full genomic sequence. I say enough number of cells, because a Single-cell amplified genome alone provides a fragmented genome and, in the best case, show a 36% of genome recovery (see sections 3.3 or in 3.4), whereas we have seen that SAGs co-assembly can increase the genome recovery up to 55% in our three SAGs –see also (Mangot et al. 2017)–. Overall this seems a very expensive process that includes the generation of many data just to tackle Opisthokonta diversity. Thus, it is key the collaboration from different scientists of different fields, in order to join efforts and take profit altogether from the data generated.

Another option rather than isolating the cells through FACS, is micromanipulating the cells. This requires a high expertise to be able to identify the cells of interest, and collect them for further

whole genome amplification. Actually, micromanipulation was used recently to obtain ribosomal gene sequences (both small and large subunits) of previously described choanoflagellate genera without molecular data available. Interestingly, those choanoflagellates correspond to some of the most abundant choanoflagellates OTUs found within TARA oceans database (Frank, Helge Abuldhauge, and Daniel 2017). In this case, if the authors had performed a SCG approach they would have got genomic data of those choanoflagellates. However, in my opinion, micromanipulation is probably better suited for single-cell transcriptomics (SCT).

SCT provides a similar percentage of gene recovery than SCG, but with the advantage that the genes are not fragmented and you can avoid the annotation process. Additionally, SCT amplifies only the RNA with poly-A tails, avoiding bacterial contamination and making the technique more suitable for any lab conditions. The only drawback of SCT, is that it has not been tested yet over pico/nano sized cells (<20μm). However, you can circumvent this by adding more than one cell of your taxa of interest in the PCR tube for further transcriptomics amplification. Thus, this method can be ideal for bigger cells like members of Pompholyxophryidae family, which are easy to identify. Furthermore, SCT can also be used to obtain transcriptomic data of microbial animals that contain few number of cells and there is not available genomic data like loriciferans, for example.

Therefore, why just not moving directly to SCT even in cells isolated through FACS? The thing is that genomic data provides much more information beyond gene sequences. Moreover, with SCT we miss ribosomal genes data, which makes problematic the identification of undescribed taxa. In addition, SCG can take profit of metagenomics and metatranscriptomics data to improve the final assemblies (Mende et al. 2016).

Metagenomes from samples in which your organim of interest is in high proportion can also be a good approach to obtain uncultured genomic data. Specially nowadays, because the are available many bioinformatic tools that permit the classification (a.k.a binning) of metagenomic scaffolds in different taxa according to GC content, coverage and other features (Alneberg et al. 2014; Kang, Egan, and Wang 2015).

Thus, the combination of both SCG approaches with the help of metagenomics will be key to expand the genomic information of Opisthokonts. Even though the field is moving faster –recently, an improved method of WGA based in MDA was published (Stepanauskas et al. 2017)–, there are still limitations on the single-cell genomics technologies. However, what is clear is that, even with these limitations they allow us to have access to genomic/transcriptomic information that will otherwise remain unknown. Information that will help us to better reconstruct and understand Opisthokonta evolution and provide new insights regarding the transition towards animal multicellularity, as shown in section 3.4 and discussed in the next chapter.

4.4. New insights on Choanoflagellates evolution and the origin of animal multicellularity

a) Discrepancy regarding Teretosporea monophyly

Thanks to our SAGs –UC1 and UC4–, the transcriptomes of 19 choanoflagellate species, and the data from newly described holozoan genera (*Chromosphaera*, *Syssomonas* and *Pigoraptor),* I could infer Choanoflagellata and Holoza evolution through a phylogenomics approach. Before discussing the implications on choanoflagellates evolution I would briefly discuss the issue of the monophyly of teretosporeans.

Teretosporea was established in a recent phylogenomics analysis based in a dataset of 78 single-copy protein domains (Torruella et al. 2012), and was defined as the earliest-branching holozoan clade that clusters together the enigmatic *Corallochytrium* with Ichthyosporeans (Torruella et al. 2015). The clade was also recovered –with higher statistical support– when another early-branching Ichthyosporean, the Dermocystid *Chromosphaera perkinsii* was included (Grau-Bové et al. 2017). In that occasion the same dataset had been expanded up to 87 single-copy protein domains (Grau-Bové et al., 2017). However, when the authors that isolated and described *Syssomonas* and *Pigoraptor* lineages, added those organisms to Holozoa phylogeny, they got a different result resulting in a paraphyletic Teretosporea. In particular, *Syssomonas* clustered with *Corallochytrium,* forming the Plurimorfea clade.

Plurimorfea instead of grouping together with Ichthyosporeans, formed an independent clade, breaking Teretosporea monophyly (Hehenberger et al., 2017). However, Hehenberger an co-workers had not included the early-branching Ichthyosporea, *Chromosphaera perkinsii* (Hehenberger et al., 2017). Moreover, they used another phylogenomic dataset, previously used to infer the eukaryotic tree of life (Fabien Burki et al. 2016).
My phylogenomic analysis included data from all single-celled holozoans with transcriptomic/genomic data available. Thus *Chromosphaera*, *Syssomonas*, *Corallochytrium* and the rest of ichthyosporeans were present in the phylogeny. I also included a more extensive outgroup than Hehenberger and co-workers. The monophyly of Teretospoera was highly supported (99 UFBS / 1 pp). I consider that is strong indicative of the monophyly of Teretosporea. However the fact that in another dataset the topology is different demands further work comparing both datasets, in order to understand the reasons behind the discrepancies.

b) Reshaping choanoflagellates evolution: A freshwater non-thecated colonial ancestor?

My phylogenomics analysis recovered important topological differences compared to a previous choanoflagellates phylogeny (Martin Carr et al. 2008; Martin Carr et al. 2017). The inferred tree breaks the monophyly of Craspedida; and, more importantly, *Codosiga hollandica* appears to be the earliest-branching lineage of choanoflagellates. Thus, this new phylogenetic framework has important implications with regards the evolution of choanoflagellates. The early-branching lineage –*Codosiga hollandica*– is a non-thecated, freshwater species with a peduncle to attach to the surface (codosigid morphology). Therefore, it is possible that the choanoflagellate ancestor was non-thecated, although could contain organic coverings based in Glycocalix as other *Codosiga* species. After *Codosiga* divergence, choanoflagellates developed many types of organic coverings, and also the silica based lorica of acanthoecids. I here summarize and discuss the evolutionary morphological changes of choanoflagellates regarding extracellular coverings.

Our phylogenomic analysis show, that *S. dolichothecata* branches before Acanthoecida and Craspedida clade 1/2 split. In addition, *S. dolichothecata* is related with *S. tuba* and both species have a tube

shaped theca and comprise Craspedida clade 3. Tube theca morphology, with the current data, appears as a synapormophy of the group (see section 1.3e) (Martin Carr et al. 2017). It is worth mentioning, however, that our phylogenetic reconstruction fails to confidently place the position of clade 3 (Fig. 3 within section 3.4). The tree weakly supports *S. dolichothecata* within an early-branching position prior to Acanthoecida and Craspedida clade 1/2 split, it could can also fall either sister to clade 1/2 of craspedidans (Martin Carr et al., 2017) or sister to Acanthoecida –the last option was recovered in the maximum likelihood topology in our analysis– . I expect that in the future a wider taxon sampling including clade L, FRESCHO3-4, *S. tuba*, other *Codosiga* species and species from the genus *Sphaeroeca*, together with a choano-specific phylogenomic dataset, would help to better solve choanoflagellates phylogeny, and to confirm the paraphyly of Craspedida.

Whithin choanoflagellates it appeared also the loricae of Acanthoecida, which was an important achievement that permitted acanthoecids to conquer pelagic environments, being able to predate without the need to be attached to a surface (Leadbeater 2015). It seems that this morphological innovation was, indeed, an evolutionary success. Because acanthoecids are the choanoflagellates with the highest relative abundance of unicellular Opisthokonta in marine environments (del Campo et al. 2015). In addition, they are the most diversified group of choanoflagellates, with 150 species described (Leadbeater, 2015). Furthermore, my results also solve one the debates regarding Acanthoecida phylogeny. Nudiform loricates appear derived from tectiforms. Although many authors have suggested that this topology is in disagreement with the experimental results found on *Stephenoeca diplocostata* (Martin Carr et al., 2017).

Those experiments consisted in depriving of silica *S. diplocostata* cells. The cells become ‘naked’ and loss their lorica covering. When the silica is added again, a new lorica is produced but in a nudiform-like manner. This fact makes some authors suggests that if there is the capacity in a tectiform species to produce, in certain conditions, a lorica in the same way than nudiform species, nudiform lorica must be ancestral to all loricate species (Leadbeater, 2015).

My results contradict this interpretation, and perhaps lorica formation of *S. diplocostata* have a similarity with nudiform species

because *S. diplocostata* is the sister group of nudiform species, rather a sign of ancestral character. In this regard, the morphology of the SAG UC2, as early-branching nudiforms can be very interesting to shed light on this issue (Fig. 1 within section 3.4).

Finally, there is within the rest of Craspedidans –clade 1 and clade 2–, a great variety of organic thecas and extracellular coverings, Although within clade 2 most of choanoflagellate species present a flask theca, argued to be the morphologically most complex, and perhaps ancestral, of both clades (Martin Carr et al. 2017). Within Clade 1 we find also the cup and ovoid theca of *S. rossetta* and *M. roanoka* respectively. In addition, there are also the non-thecated species of *Monosiga brevicollis* and the species of the genera *Harteosiga,* which were formely members of *Codosiga* genus (see section 1.3e) (Martin Carr et al. 2017).

From our phylogenetic reconstruction we can infer that the choanoflagellate ancestor was colonial, probably adopting a *Codosiga* morphology, which was subsequently lost in craspedidans clade 3, acanthoecids, and some craspedidan species of the clade 1 and the clade 2 like *Monosiga brevicollis* and *Salpingoeca macrocollata* respectively.

Finally, overall, our results show that the earliest-branching choanoflagellate lineage (*C. hollandica*) is a non-thecate colonial choanoflagellate, with a peduncle to attach to the surface (codosigid morphology) and that lives on freshwater environments. This result would be in disagreement with the idea of a marine choanoflagellate ancestor (Martin Carr et al. 2008; Martin Carr et al. 2017). Most of choanoflagellates habit marine waters. However, all the species related with *Codosiga* –members of *Codosiga* and *Sphaeroeca* genera– belong to freshwater environments (section 1.3e Fig. 9). Of course, it could had happen that the ancestor was marine and *Codosiga* related species, afterwards invaded freshwater environments. However I consider it less likely, taking into account that most of the filasterean species (the sister-group to choanoflagellates and animals) were isolated from freshwater environments (Hehenberger et al., 2017; Stibbs et al., 1979). However, MAOP1 might be a filasterean lineage (Fig.1 within section 3.4) (see section 1.3d), and its a marine environmental group. Therefore, if further analysis show MAOP1 as a filasterean lineage, this argument will lose some credit. Anyway an being

aware of the limitations, in the section 4.4d I will discuss how a freshwater unicellular ancestor could contribute to the transition towards animal multicellularity.

c) The genomic basis of animal origins

Understanding the origin of animal to multicellularity requires knowledge of the pre-metazoan genomic information that preceded and accompanied the transition from a single-celled organism to the formation of the first animal (Urmetazoa). In this regard, we need to elucidate the gene content of the unicellular ancestor of Metazoa; and the only way is by comparing the genomes of the extant unicellular relatives of animals, the single-celled holozoans. Before, and during the course of this thesis, several genomic and transcriptomic information from unicellular holozoans became available, providing novel insights into the gene content of the unicellular ancestor of animals. The results have been quite surprising. Although there was some gene innovation at the onset of Metazoa, the unicellular ancestor of animals already had a rich repertoire of genes that are required for cell adhesion, cell signaling, transcriptional regulation and neural functions in modern animals (Nicole King et al. 2008; Sebé-Pedrós et al. 2010; Sebé-Pedrós et al. 2011; H. Suga et al. 2012; Fairclough et al. 2013; Richter and King 2013; Sebé-Pedrós, Degnan, and Ruiz-Trillo 2017) (see section 1.4b). The results obtained in section 3.4 expand the single-celled holozoan information and show that some genes or protein domains previously thought to be metazoan specific were present in the unicellular ancestor of animals. Therefore, in this section, I will briefly provide the current view of the pre-metazoan genetic tool-kit highlighting the genes found in the section 3.4.

The animal unicellular ancestor already encoded in its genome genes of cell adhesion proteins, which are necessary for cell-to-cell and cell-to-extracellular matrix (ECM) interactions in animals (Fig. 17). One of the most interesting is the full presence of the integrin adhesome found in the filasterean *C. owzczarkaki*, being part of its components posteriorly lost in choanoflagellates (Sebé-Pedrós et al. 2010). Thus, this would mean that the ancestor of animals already had the ability to interact with the ECM. Concretely, integrins were found to be upregulated during the aggregative stage of *Capsaspora*

owczarzaki, reinforcing this hypothesis (Sebé-Pedrós, Irimia, et al. 2013). Furthermore, the unicellular ancestor of animals possessed ECM proteins like fibronectin, collagen domains and also the Colagen IV, thought to be metazoan specific and cell-to-cell adhesion proteins like c-type lectins and cadherins.

On the other hand, the single-celled metazoan ancestor also possessed components of signaling pathways related to cell-to-cell communication, like the cytoplasmatic transducers Src, MAPKs or components of the hippo pathway, although not all the pathway was complete (H. Suga et al. 2012; Reports 2012). In addition, it encoded signaling receptors like GPCR's (A. De Mendoza, Sebé-Pedrós, and Ruiz-Trillo 2014) and Tyrosine kinase receptors. In addition, the unicellular metazoan ancestor lacked of Notch and Hedgehog proteins, although the protein domains were already there forming different proteinic architectures. Meaning, that Notch and Hedgehog proteins were formed thanks to a domain shuffling event occurred in the transitions towards the Urmetazoa (Nicole King et al. 2008; Hiroshi Suga et al. 2013; Grau-Bové et al. 2017), which permitted the appearance of Notch and Hedgehog pathways (see section 1.4b). Our results (section 3.4) also revealed another domain shuffling event at the stem of Metazoa. Particularly, within Smad and Nucleophosmin transcription factors.

Smad proteins are composed by the domains MH1 and MH2 (Attisano and Lee-hoeflich 2001), the DNA-binding domain MH1 was described as animal specific (A. de Mendoza et al. 2013). I found that both domains were present already in the Choanozoa ancestor, although not in the same architecture. Therefore, Smad proteins were formed as Notch and Hedgehog thanks to a domain shuffling events in the transition between the unicellular ancestor and the last Metazoa common ancestor. Regarding Nucleophosmin, it occurred the same, being the c-terminal domain NPM1-C, a choanozoa innovation (as MH1 in the case of Smads), only found until now in a single-celled holozoan within the SAG UC4. Overall, this shows that domain shuffling was an important source of gene innovation at the origins of animals.

Besides Smad and Nucleophosmin, there are many transcription factors related with multicellular functions with a pre-metazoan origin like T-box, Runix, p53, Myc (Sebé-Pedrós et al. 2011; A. de

Mendoza et al. 2013; Sebé-Pedrós, Degnan, and Ruiz-Trillo 2017), the homeobox LIM (Grau-Bové et al. 2017) or the protein domain Fanconi_A_N found in our SAGs. All these transcription factors with pre-metazoan origin were related with development, cell differentiation or the control of cell cycle.

Finally, comparative genomics analysis also revealed that many components of metazoan neural functions preceded the origins of animals such as sodium (Liebeskind 2011) and calcium channels (Cai 2008), Neuroglobulins (Lechauve et al. 2013) and proteins related in synapsis (Alié and Manuel 2010; Fairclough et al. 2013), in postsynaptic density like Homer (Burkhardt et al. 2014) and neural secretion (Burkhardt et al. 2011). I found two more proteins related with neural function with pre-metazoan origin; the NKAIN, a sodium dependent ATPase interacting protein (Gorokhova et al. 2007) and the Plexin protein, which are neural semaphorin receptors that guide axon formation in neural development (Winberg et al. 1998). In addition, our comparative genomic approach unveiled genes related in other multicellular functions like Immune system response, as the case of the transcription factor IRF, or the domain TILa that allows sperm cells to bind specifically to egg extracellular matrix (Hardy and Garbers 1995).

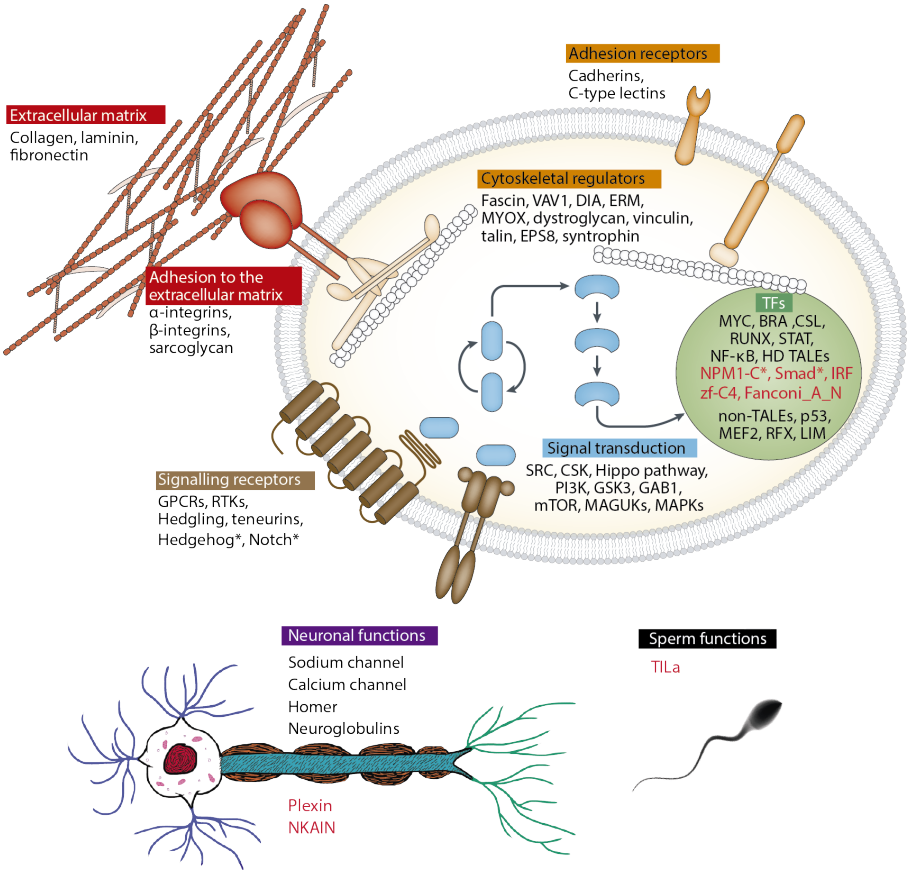


Figure 17. The pre-metazoan genetic tool-kit. Adapted from (Sebé-Pedrós, Degnan, and Ruiz-Trillo 2017) (Sebé-Pedrós et al., 2017b) and including some recent results from (Grau-Bové et al., 2017) and this PhD thesis. In red there are depicted the pre-metazoan genetic toolkit inferred in this work. Genes with an asterisk indicate, domain presence in the unicellular metazoan ancestor but not the protein domain structure. EPS8, epidermal growth factor receptor kinase substrate 8; GAB, GRB2-associated binding protein; GPCRs, G protein-coupled receptors; GSK3, glycogen synthase kinase 3; HD, homeodomain; LIM, homeobox domain LIM; MAGUKs, membrane-associated guanylate kinases; MAPKs, mitogen-activated protein kinases; MEF2, myocyte-specific enhancer factor 2; mTOR, mechanistic target of rapamycin; MYOX, myosin X; NF-κB, nuclear factor-κB; NPM1-C, C-terminal domain Nuclophosmin transcription factor; PI3K, phosphatidylinositol-3 kinase; RTKs, receptor tyrosine kinases; STAT, signal transducer and activator of transcription; TALEs, three amino acid loop extensions; TF, transcription factor

Overall, this shows new genes related to key multicellular functions that were already present in the unicellular ancestor of animals. This findings increases the genetic material in which co-option of ancestral genes into new functions, were used in the transition towards animal multicellularity. That is, many of these genes that currently function within multicellular animals evolved within a unicellular context and were subsequently adapted for a multicellular life. This co-option of the ancestral genes, together with the evolution of novel animal genes –some of them produced by domain shuffling events (like Notch and Smads)– and with the expansion and diversification of some ancestral gene families (Grau-Bové et al., 2017), configured the genomic basis for animal multicellularity (see introduction 1.4b)

Therefore, taking into account the complexity of the pre-metazoan genetic toolkit, together with the fact that unicellular holozoans have complex life-cycles, including different multicellular stages – coloniality (choanoflagellates), aggregative behaviour (*C. owczarzaki*) and coenocytic growth (Teretospora)– that have been proven to be regulated at transciptional (Fairclough et al. 2013; Sebé-Pedrós, Irimia, et al. 2013; A. de Mendoza et al. 2015) epigenetic and proteomic level (Sebé-Pedrós et al. 2015; Sebé-Pedrós et al. 2016). This makes some authors (Sebé-Pedrós, Degnan, and Ruiz-Trillo 2017) to challenge the Choanoblastea theory and embrace the Synzoospore hypothesis for the origins of animals, a theory that was formulated by Zakhvatkin in 1949 and later further developed by Mikhailov et al. (Mikhailov et al. 2009) (see section 1.3f).

In this scenario, Metazoa arose from an ancestral protist with a complex life cycle that involved multiple temporally regulated life stages. This life cycle was dependent on environmental stimuli, as observed in some extant unicellular holozoans (Alegado et al. 2012). These temporally regulated cell types would become spatially integrated into the first animals (Sebé-Pedrós, Degnan, and Ruiz-Trillo 2017).

All of the aspects commented in this section includes intrinsic elements, mainly genomic but also phenotypic, that explain the biological mechanisms allowing the transition towards animal multicellularity. However, what about the extrinsic factors? Which

were the environmental conditions that triggered a protist with a complex life-cycle became an animal? In the next section I will speculate among that question considering the results obtained in the chapter 3.4.

d) The environmental and ecological context that could facilitate the origins of animals

Our phylogenomic analysis revealed the possibility that the ancestor of choanoflagellates could have inhabited freshwater environments. In addition, most of the filasterean species described were isolated also from freshwaters environments (Stibbs et al. 1979; Hehenberger et al. 2017). Thus, these evidences suggest that the Choanozoa ancestor, that lead to the origins of animals, could had been a freshwater organism. Assuming that perhaps in the future the discovery of more Choanoflagellatea or Filasterea taxa could discard such scenario, I here expose some arguments in favour of a freshwater environment being key to facilitate the transitions towards animal multicellularity. Additionally, I propose a hypothesis for the origins of animals.

I will start with the paradigm of the Synzoospore hypothesis (see section 1.3f). Thus, I envision an unicellular ancestor of animals with a complex life-cycle, with one of the life stages with a morphology similar to *Codosiga* genera that could form colonies and also, perhaps, aggregates. Different environmental stimuli could trigger changes among its life stages –as it has been described in the choanoflagellate *S. rosetta* that a bacterial sphingolipid triggers colony formation (Alegado et al. 2012)–. Furthermore, likely this single-celled ancestor had sex capabilities, as described in many unicellular eukaryotes (Woodland 2016) and in *S. rosetta* (Levin and King 2013). It is worth mentioning that *Codosiga* species can adopt sedentary forms with long stalks, and in some cases highly branched (Leadbeater, 2015). The cells produced a extracellular covering based on carbohydrates which in *C. botrytis* have been shown to be more complex than in *Harteosiga*, the other genera of choanoflagellates with codosigid morphology (Leadbeater, 2015).

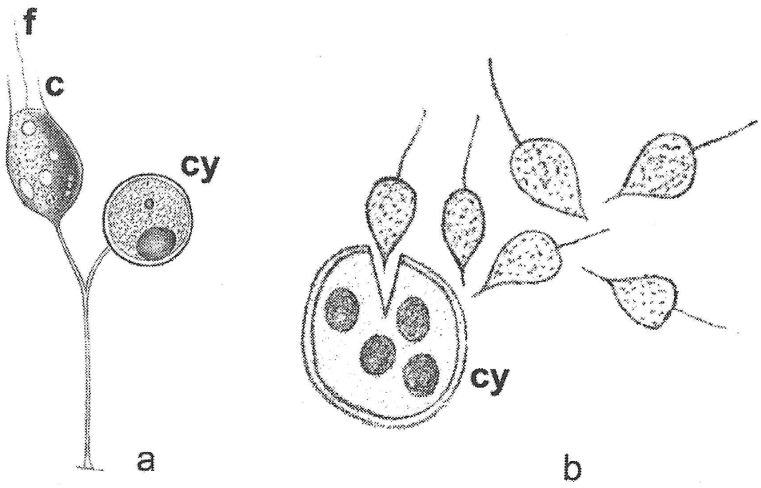


Figure 18. *Codosiga botrytis* representation from (Leadbeater, 2015). (a) Stalked colony of two cells, one of which has encysted with sphaerical cyst wall (cy). Flagellum (f), collar (c). (b) Excystment apparently consisted of division of the cyst contents (cy) and release of many small flagellated cells. Reproduced from Fisch (1885).

Codosiga, besides stalked colonies, also can release a colony of floating cells to colonize other areas (Leadbeater, 2015). It has been described also cystic forms within *C. botrytis* that, in my opinion, resembles the coenocytical division of ichthyosporeans and the release of swimmer flagellated cells to colonize other environments (Fig. 18) (Leadbeater, 2015). Thus, indeed, the unicellular ancestor of animals was, under my theory, a relatively complex organism holding a codosigid morphology and being able to switch between different cell types and between benthic and pelagic colonial forms.
Now let's place the ancestor in a geochemical context. The first biomarker evidences for animals are from nearly 635 Mya, thanks to the discovery of Desmospongiae steroids in Cryogenian rocks (Love et al. 2008). In addition, molecular clock estimates situate the origin of animals between 850-650 Mya (Cunningham et al. 2017), and the origin of Opisthokonts between 1,579-904 Mya (Eme et al. 2014) (see section 1.4a, Fig.4 and Fig.13). Therefore we can speculate that a single-celled animal ancestor was living around 900-700 Mya, within the Tonian period at Neoproterozoic (Fig. 19).

It was an era of many geological changes on Earth. The supercontinent Rodinia was being fragmented leading towards two glacial periods that end up nearly 635Mya (Moczydlowska et al. 2017). The concentration of oxygen was lower than today (~1% of the present levels) (Alegado and King 2014) and the deep ocean was anoxic (Lyons, Reinhard, and Planavsky 2014). Thus, in that environment, the location in which there was a minimum oxygen concentration to sustain heterotrophic life was at the marine coastal areas or at terrestrial-freshwater environments (Alegado and King 2014; Lyons, Reinhard, and Planavsky 2014). Therefore, we may wonder which were, specifically, the characteristics that freshwater environments could have had to be more appealing for the single-celled ancestors of animals. And also, how these environment could facilitate the transition towards multicellularity. Well, the in the following paragraph I propose some explanations for this.

The first thing that we might consider is that freshwater environments contain less osmotic pressure than marine waters. Such chemical conditions could had been more suitable for the cell physiology and biochemistry of the choanozoan ancestor. Additionally, Choanoflagellatea freshwater species, are by far the ones that can adopt larger colonies compared with marine species. The maximum number of cells in marine choanoflagellate colonies are 120, and it is an out-layer. Freshwater choanoflagellates, in contrast, can have colonies with up to 10,000 cells, as is the case of *Proteorospongia pedicellata* Oxley (1884) and Skuja (1932) (Leadbeater, 2015), *Sphaeroca volvox*, sister of Codosiga clade can reach up to several hundred cells. Thus, freshwater environments, seems to present certain conditions that facilitate the formation of bigger colonies. Actually, it has proposed that these differences in colonial size among choanoflagellates, are dependent on environmental bacterial concentrations (Leadbeater, 2015.

According to Leadbeater, in freshwater environments bacterial concentration is higher and this is the reason why freshwater colonies are bigger, which will be in the line of the studies that link the presence of bacterial *A .machipongonensis* sphingolipid with the colony formation in *S. rosetta* (Alegado et al. 2012).
Therefore, in freshwater environments there were the biological, and probably chemical, conditions that could have facilitated the formation of bigger colonies to increase the feeding capabilities.

The colonies started to be more complex and concentrated the different cell types that occurred during the life cycle in a common entity, forming a very complex colony that could still switch from benthic to pelagic forms. The complexity increased until appeared cell differentiation between the soma and the germline (Woodland 2016), leading to a primitive Urmetazoa.

There is, however, a challenge. We do not know whether the bacterial concentration was higher on freshwater environments than in marine. Little is known about the freshwater-terrestrial environments of the Neoprotezoric era. Although it was claimed recently, that the eukaryotic supergroup Archeoplastida could have emerged as well during the Proterozoic within freshwater environments (Ponce-Toledo et al. 2016). Thus, this can be an indirect evidence that Neoproterozoic freshwater-environments were populated of bacteria and eukaryotic life.

Another weak point of my hypothesis is that all early-branching metazoan lineages (Porifera, Ctenophora, Cnidaria, Trichoplax), are mostly marine. There are, for instance, freshwater sponges, but are considered to be an adaptation from marine environments (Meixner et al. 2007). Therefore, this fact might disagree with a transition towards animal multicellularity on freshwater environments. However geological events that occurred during late Proteozoic era provide a plausible explanation for a primitive transition towards animal multicellularity in freshwater environments, and also a marine Urmetazoa ancestor.

During the Tovonian period (1000-700 Mya), Rodinia continent was fragmented and rifted along newly formed continental margins, creating seaways with active circulation (Moczydlowska et al. 2017). This could facilitate the primitive freshwater Urmetazoa, which could posses a very simple multicellularity, to enter in contact with marine environments. This geological event was followed by a collapse of many ecosystems during the Cryogenian period (720-635Mya), due to severe ice ages. Therefore in that extreme conditions only the primitive Urmetazoa that had colonized marine environments could have survived.

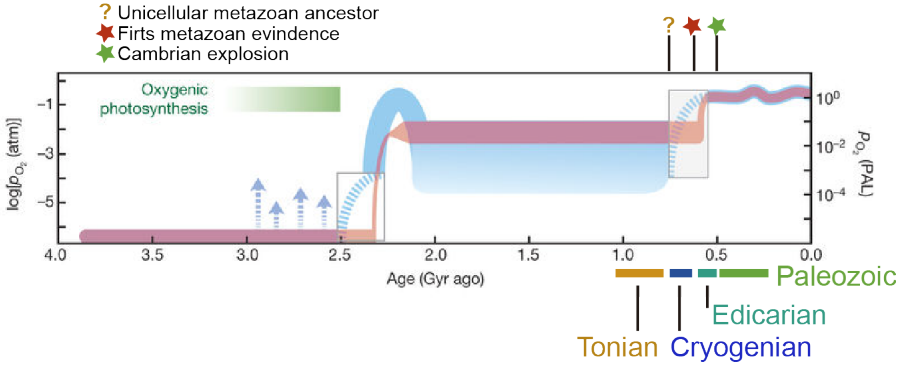


Figure. 19: Evolution of Earth's atmospheric oxygen content through time and animal origins. Adapted from (Lyons et al., 2014). The faded red curve shows a classical, two-step view of atmospheric evolution, while blue curve shows the emerging model (pO_2, atmospheric partial pressure of O_2). Right axis, pO_2, relative to the present atmospheric level (PAL); left axis, log pO_2. Arows denote possible 'whiffs' of O_2 late in the Archaean; their duration and magnitude are poorly understood. An additional frontier lies in reconstructing the detailed fabric of 'state changes' in atmospheric pO_2, such as occurred at transitions from late part of Archean to the early Proterozoic and from late Proterozoic to early Phanerozoic (blue boxes). Below the graph are depicted the NeoProterozoic eras and Palezoic. Above there are indicated events related to the animal origins taken from (Cunningham et al., 2017), and also and the speculated appearance of the unicellular metazoan ancestor.

Thanks to the increase in Oxygen levels, the deep-sea became oxic (Fig. 19) (Lyons, Reinhard, and Planavsky 2014) and animals could colonize all marine environments and evolved into a great variety of forms present in the Cambrian fossil records (Sperling et al. 2013; Mills et al. 2014; A. H. Knoll and Sperling 2014), also known as Cambrian explosion. In Cambrian rocks are preserved the fossils of the ancestors of most of metazoa phyla (see section 1.4a), which were the precursors of the vast and wonderful animal diversity that populates the Earth today, including ourselves.

Thus, the results from the section 3.4 allow us to open new questions and perspectives regarding the origin of animals, as the hypothesis here formulated which demands further studies on the properties of freshwater environments and bacterivory as potential triggers for the development of complex colonies and multicellularity.

5. CONCLUSIONS

The main conclusions of the present work are the following:

1. The former Nuclearia sp. ATCC 50694, *Parvularia atlantis* gen. et sp. nov. represents a distinct phylogenetic lineage from previous described nucleariids genera (Nuclearia and Fonticula). *Parvularia* is a small filose amoeaba (~4μm of cell diameter), with a cystic stage that can adopt binucleated forms; this also differentiates *Parvularia*, with the rest of nucleariid, or nucleariid-related described species.

2. Besides Parvularia, there are other environmental clades within nucleariids: MAFO, env-NUC 1, env-NUC2 which might be related with nucleariid related species that lack of available molecular information.

3. The metabarcoding analysis of marine micrometazoan diversity in European coastal samples shows that metazoan diversity patterns partially corroborates previous morphological and molecular studies. Copepods appear as the most abundant organisms within the water column followed by the Appendicularians (tunicates) and Chaetognaths. In addition, the richest clades were Crustacea and Nematoda and, within the benthic environments, the relative abundance was more distributed among these two groups (Crustacea and Nematoda) together with Polychatea, Platyhelminthes and Mollusca.

4. There is a high relative abundance of metazoan RNA reads within pico-nano size fractions (0.8-20μm) in our dataset, suggesting that the sperm of Ctenophores and Cnidarians plays a relevant ecological role as part of the microbial food network.

5. In some animal phyla there is considerable genetic novelty that is yet to be unravelled, including novelty in several well-sampled groups such as Crustacea, Platyhelminthes or Nematoda.

6. We found a potential new group of widespread tunicates related with thalacean species, named as MAME 1.

7. Single-cell genomics (SCG) techniques produces a high variable results (6-33% of genome completeness), although even in the best case, the genome recovery is low (30-40%). As each individual SAG has different biases, it is important selecting promising SAGs before performing a high-depth sequencing. The more depth sequencing, the longer the final genome assembly whether the downsampling curves are not saturated.

8. Genome assemblies from SCG data are highly fragmented and affect the completeness of the genic regions, difficulting even more the annotation process. Co-assembly of different SAGs improves the genome recovery.

9. Besides the limitations, SCG can still provide interesting insights onto evolutionary questions. SAGs can provide enough information to perform phylogenomic analysis and some gene-family evolution by analysing protein domains.

10. We expanded the choanoflagellate genomic information available thanks to single-cell genomics from environmental cells. In particular, we recovered meaningful information from two taxa, the UC1, a clade 1 craspedidan and UC4, an early branching acanthoecid, which is also the third most abundant choanoflagellate from TARA Oceans database. We could recover as well the first mitochondrial sequence of an Acanthoecida thanks to the SAG UC2.

11. Our phylogenomics analysis re-shaped the phylogeny of the choanoflagellates. Our results break the monophyly of Craspedidans and bring the species *Codosiga hollandica* to the earliest branching position of choanoflagellates. This suggests a non-thecated and freshwater ancestor of choanoflagellates, opening new hypothesis among the ecological context in which choanoflagellates and animals could have emerged

12. Comparative genomics shows that most of the protein domains related to multicellular functions and innovated at the Choanozoa, were further retained in animals than in choanoflagellates.

13. We have now a broader view regarding the protein domain composition of the unicellular ancestor of animals, that now

includes some additional protein domains, previously thought to be animal-specific, like Plexin, NKAIN, Smad domains or the C-terminal domain of the Nucleophosmin protein. The two last metazoan proteins, are the product of a domain shuffling event occurred in the transition towards animal multicellularity.

References

Adell, Teresa, Vera Gamulin, Sanja Perović-Ottstadt, Matthias Wiens, Michael Korzhev, Isabel M. Müller, and Werner E. G. Müller. 2004. “Evolution of Metazoan Cell Junction Proteins: The Scaffold Protein MAGI and the Transmembrane Receptor Tetraspanin in the Demosponge Suberites Domuncula.” *Journal of Molecular Evolution* 59 (1): 41–50. doi:10.1007/s00239-004-2602-2.

Adl, Sina M., Alastair G B Simpson, Christopher E. Lane, Julius Lukeš, David Bass, Samuel S. Bowser, Matthew W. Brown, et al. 2012. “The Revised Classification of Eukaryotes.” *Journal of Eukaryotic Microbiology* 59 (5): 429–93. doi:10.1111/j.1550-7408.2012.00644.x.

Alegado, Rosanna A., Laura W. Brown, Shugeng Cao, Renee K. Dermenjian, Richard Zuzow, Stephen R. Fairclough, Jon Clardy, and Nicole King. 2012. “A Bacterial Sulfonolipid Triggers Multicellular Development in the Closest Living Relatives of Animals.” *eLife* 2012 (1): 1–16. doi:10.7554/eLife.00013.

Alegado, Rosanna A, and Nicole King. 2014. “Bacterial Influences on Animal Origins.” *Cold Spring Harb Prespect Biol.* 6: 1–16.

Alié, Alexandre, Lucas Leclère, Muriel Jager, Cyrielle Dayraud, Patrick Chang, Hervé Le Guyader, Eric Quéinnec, and Michaël Manuel. 2011. “Somatic Stem Cells Express Piwi and Vasa Genes in an Adult Ctenophore: Ancient Association of ‘germline Genes’ with Stemness.” *Developmental Biology* 350 (1). Elsevier Inc.: 183–97. doi:10.1016/j.ydbio.2010.10.019.

Alié, Alexandre, and Michaël Manuel. 2010. “The Backbone of the Post-Synaptic Density Originated in a Unicellular Ancestor of Choanoflagellates and Metazoans.” *BMC Evolutionary Biology* 10 (1): 34. doi:10.1186/1471-2148-

10-34.

Alneberg, Johannes, Ino De Bruijn, Melanie Schirmer, Joshua Quick, Umer Z Ijaz, Leo Lahti, Nicholas J Loman, Anders F Andersson, and Christopher Quince. 2014. “Binning Metagenomic Contigs by Coverage and Composition” 11 (11). doi:10.1038/nmeth.3103.

Arroyo, Alicia S, David López-Escardó, Colomban de Vargas, and Iñaki Ruiz-Trillo. 2016. “Hidden Diversity of Acoelomorpha Revealed through Metabarcoding.” *Biology Letters* 12 (9): 20160674. doi:10.1098/rsbl.2016.0674.

Attisano, Liliana, and Si Tuen Lee-hoeflich. 2001. “Protein Family Review The Smads,” *Genome Biology*, 2(8):reviews3010.1–3010.8

Baldauf, S. L. Roger, A. J. Wenk-Siefer, I. Doolittle, W.F. 2000. “A Kingdom-Level Phylogeny of Eukaryotes Based on Combined Protein Data.” *Science* 290 (5493): 972–77. doi:10.1126/science.290.5493.972.

Baldauf, S. L. 2003. “The Deep Roots of Eukaryotes.” *Science* 300 (5626): 1703–6. doi:10.1126/science.1085544.

Beaugrand, G. , Brander, K. M. and Lindley, J.A. 2003. “Plankton Effect on Cod Recruitment in the North Sea.” *Nature* 426 (2003): 661-664

Berbee, Mary L., and John W. Taylor. 2010. “Dating the Molecular Clock in Fungi - How Close Are We?” *Fungal Biology Reviews* 24 (1–2). Elsevier Ltd: 1–16. doi:10.1016/j.fbr.2010.03.001

.

Bhattacharya, D, DC Price, HS Yoon, EC Yang, NJ Poulton, RA Andersen, and SP Das. 2012. “Single Cell Genome Analysis Supports a Link between Phagotrophy and Primary Plastid Endosymbiosis.” *Scientific Reports* 2: 1–8. doi:10.1038/srep00356.

Blanc-Brude, R., Y. Skreb, and J. Dragesco. 1955. “Sur La Biologie de Nuclearia Delicatula.”

Blaxter, Mark, Jenna Mann, Tom Chapman, Fran Thomas, Claire Whitton, Robin Floyd, and Eyualem Abebe. 2005. “Defining Operational Taxonomic Units Using DNA Barcode Data.” *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences* 360 (1462): 1935–43. doi:10.1098/rstb.2005.1725.

Bonner, J T. 1998. “The Origins of Multicellularity.” *Integrative Biology:* 1 (1): 27–36. doi:10.1002/(SICI)1520-6602(1998)1:1.

Bouquet, Jean-Marie, Endy Spriet, Christofer Troedsson, Helen Otterå, Daniel Chourrout, and Eric M Thompson. 2009. “Culture Optimization for the Emergent Zooplanktonic Model Organism Oikopleura Dioica.” *Journal of Plankton Research* 31 (4): 359–70. doi:10.1093/plankt/fbn132.

Brown, M. W., S. C. Sharpe, J. D. Silberman, A. A. Heiss, B. F. Lang, A. G. B. Simpson, and A. J. Roger. 2013. “Phylogenomics Demonstrates That Breviate Flagellates Are Related to Opisthokonts and Apusomonads.” *Proceedings of the Royal Society B: Biological Sciences* 280 (1769): 20131755–20131755. doi:10.1098/rspb.2013.1755

Brown, Matthew W., Frederick W. Spiegel, and Jeffrey D. Silberman. 2009. “Phylogeny of The ‘forgotten’ cellular Slime Mold, Fonticula Alba, Reveals a Key Evolutionary Branch within Opisthokonta.” *Molecular Biology and Evolution* 26 (12): 2699–2709. doi:10.1093/molbev/msp185.

Brunet, Thibaut, and Nicole King. 2017. “The Origins of Animal Multicellularity and Cell Differentiation.” *BioRxiv*.

Burkhardt, Pawel, Mads Grønborg, Kent McDonald, Tara Sulur, Qi Wang, and Nicole King. 2014. “Evolutionary

Insights into Premetazoan Functions of the Neuronal Protein Homer." *Molecular Biology and Evolution* 31 (9): 2342–55. doi:10.1093/molbev/msu178.

Burkhardt, Pawel, Christian M Stegmann, Benjamin Cooper, Tobias H Kloepper, and Cordelia Imig. 2011. "Primordial Neurosecretory Apparatus Identi Fi Ed in the Choano Fl Agellate Monosiga Brevicollis," 2–7. doi:10.1073/pnas.1106189108/-/DCSupplemental.www.pnas.org/cgi/doi/10.1073/pnas.1106189108.

Burki, F., N. Okamoto, J.-F. Pombert, and P. J. Keeling. 2012. "The Evolutionary History of Haptophytes and Cryptophytes: Phylogenomic Evidence for Separate Origins." *Proceedings of the Royal Society B: Biological Sciences* 279 (1736): 2246–54. doi:10.1098/rspb.2011.2301.

Burki, Fabien, Maia Kaplan, Denis V. Tikhonenkov, Vasily Zlatogursky, Bui Quang Minh, Liudmila V. Radaykina, Alexey Smirnov, Alexander P. Mylnikov, and Patrick J. Keeling. 2016. "Untangling the Early Diversification of Eukaryotes: A Phylogenomic Study of the Evolutionary Origins of Centrohelida, Haptophyta and Cryptista." *Proceedings of the Royal Society B: Biological Sciences* 283 (1823): 20152802. doi:10.1098/rspb.2015.2802.

Burki, Fabien, Nick Lane, Geoffrey I Mcfadden, Michael W Gray, and Angela E Douglas. 2014. "The Eukaryotic Tree of Life from a Global Phylogenomic Perspective The Eukaryotic Tree of Life from a Global Phylogenomic Perspective," 1–17. doi:10.1101/cshperspect.a016147.

Cafaro, Matías J. 2005. "Eccrinales (Trichomycetes) Are Not Fungi, but a Clade of Protists at the Early Divergence of Animals and Fungi." *Molecular Phylogenetics and Evolution* 35 (1): 21–34. doi:10.1016/j.ympev.2004.12.019.

Cai, Xinjiang. 2008. “Unicellular Ca2+ Signaling ‘Toolkit’ at the Origin of Metazoa.” *Molecular Biology and Evolution* 25 (7): 1357–61. doi:10.1093/molbev/msn077.

Cannon, Johanna Taylor, Bruno Cossermelli Vellutini, Julian Smith, Fredrik Ronquist, Ulf Jondelius, and Andreas Hejnol. 2016. “Xenacoelomorpha Is the Sister Group to Nephrozoa.” *Nature* 530 (7588). Nature Publishing Group: 89–93. doi:10.1038/nature16520.

Carr, M, B S C Leadbeater, R Hassan, M Nelson, and S L Baldauf. 2008. “Molecular Phylogeny of Choanoflagellates, the Sister Group to Metazoa.” *Proceedings of the National Academy of Sciences of the United States of America* 105 (43): 16641–46. doi:10.1073/pnas.0801667105.

Carr, Martin, Daniel J. Richter, Parinaz Fozouni, Timothy J. Smith, Alexandra Jeuck, Barry S.C. Leadbeater, and Frank Nitsche. 2017. “A Six-Gene Phylogeny Provides New Insights into Choanoflagellate Evolution.” *Molecular Phylogenetics and Evolution* 107. The Authors: 166–78. doi:10.1016/j.ympev.2016.10.011.

Carranza, Salvador, Gonzalo Giribet, C Ribera, Jaume Baguñà, and Marta Riutort. 1996. “Evidence That Two Types of 18S rDNA Coexist in the Genome of Dugesia (Schmidtea) Mediterranea (Platyhelminthes, Turbellaria, Tricladida).” *Molecular Biology and Evolution* 13 (6): 824–32.

Carroll, S B. 2001. “Chance and Necessity: The Evolution of Morphological Complexity and Diversity.” *Nature* 409 (6823): 1102–9. doi:10.1038/35059227.

Cavalier-Smith, T. 1983. *A 6-Kingdom Classification and Unified Phylogeny*. Edited by HEA Schenk and WS Schwemmier. Endocytobi. Berlin: Walter de Gruyter & Co.

Cavalier-Smith, T. 1998. “A Revised Six-Kingdom System of Life.” *Biological Reviews* 73 (3): 203–66. doi:10.1111/j.1469-185X.1998.tb00030.x.

Cavalier-Smith, T. 2012. “Early Evolution of Eukaryote Feeding Modes, Cell Structural Diversity, and Classification of the Protozoan Phyla Loukozoa, Sulcozoa, and Choanozoa.” *European Journal of Protistology*, October. Elsevier GmbH, 1–64. doi:10.1016/j.ejop.2012.06.001.

Cavalier-Smith, T. 2017. “Origin of Animal Multicellularity: Precursors, Causes, Consequences—the Choanoflagellate/sponge Transition, Neurogenesis and the Cambrian Explosion.” *Philosophical Transactions of the Royal Society B: Biological Sciences* 372 (1713): 20150476. doi:10.1098/rstb.2015.0476.

Cavalier-Smith, T., and M. T. E. P. Allsopp. 1996. “Corallochytrium, an Enigmatic Non-Flagellate Protozoan Related to Choanoflagellates.” *European Journal of Protistology* 32 (3). Gustav Fischer Verlag · Stuttgart · Jena · New York: 306–10. doi:10.1016/S0932-4739(96)80053-8.

Cavalier-Smith, T., and E. E Y Chao. 2003. “Phylogeny of Choanozoa, Apusozoa, and Other Protozoa and Early Eukaryote Megaevolution.” *Journal of Molecular Evolution* 56 (5): 540–63. doi:10.1007/s00239-002-2424-z.

Cavalier-Smith, T. 1987. “The Origin of Fungi and Pseudo-Fungi.” In *Evolutionary Biology of the Fungi: Symposium of British Mycological Society*, edited by A.D.M. Rayner, C.M. Brasier, and D. Moore, Cambridge, 13:339–353. Cambridge.

Cavalier-Smith, T. 1993. “Kingdom Protozoa and Its 18 Phyla.” *Microbiological Reviews* 57 (4): 962–66.

Chariton, Anthony A., Leon N. Court, Diana M. Hartley, Matthew J. Colloff, and Christopher M. Hardy. 2010. “Ecological Assessment of Estuarine Sediments by Pyrosequencing Eukaryotic Ribosomal DNA.” *Frontiers in Ecology and the Environment* 8 (5): 233–38. doi:10.1890/090115.

Choi, Hyung Seok, Jeong Rae Kim, Sang Woo Lee, and Kwang Hyun Cho. 2008. “Why Have Serine/threonine/tyrosine Kinases Been Evolutionarily Selected in Eukaryotic Signaling Cascades?” *Computational Biology and Chemistry* 32 (3): 218–21. doi:10.1016/j.compbiolchem.2008.02.005.

Ciuffi, Angela, Sylvie Rato, and Amalio Telenti. 2016. “Single-Cell Genomics for Virology.” *Viruses* 8 (5): 1–10. doi:10.3390/v8050123.

Corsaro, Daniele, Julia Walochnik, Danielle Venditti, Karl Dieter Müller, Bärbel Hauröder, and Rolf Michel. 2014. “Rediscovery of Nucleophaga Amoebae, a Novel Member of the Rozellomycota.” *Parasitology Research* 113 (12): 4491–98. doi:10.1007/s00436-014-4138-8.

Creer, S, V G Fonseca, D L Porazinska, R M Giblin-Davis, W Sung, D M Power, M Packer, et al. 2010. “Ultrasequencing of the Meiofaunal Biosphere: Practice, Pitfalls and Promises.” *Molecular Ecology* 19 (March): 4–20. doi:10.1111/j.1365-294X.2009.04473.x.

Cunningham, John A., Alexander G. Liu, Stefan Bengtson, and Philip C.J. Donoghue. 2017. “The Origin of Animals: Can Molecular Clocks and the Fossil Record Be Reconciled?” *BioEssays* 39 (1): 1–12. doi:10.1002/bies.201600120.

de Bourcy, Charles F a, Iwijn De Vlaminck, Jad N Kanbar, Jianbin Wang, Charles Gawad, and Stephen R Quake. 2014. “A Quantitative Comparison of Single-Cell Whole Genome Amplification Methods.” *PloS One* 9 (8): e105585. doi:10.1371/journal.pone.0105585.

De Mendoza, Alex, Arnau Sebé-Pedrós, and Iñaki Ruiz-Trillo. 2014. “The Evolution of the GPCR Signaling System in Eukaryotes: Modularity, Conservation, and the Transition to Metazoan Multicellularity.” *Genome Biology and Evolution* 6 (3): 606–19. doi:10.1093/gbe/evu038.

de Mendoza, Alex, Arnau Sebé-Pedrós, Martin Sebastijan Šestak, Marija Matejcic, Guifré Torruella, Tomislav Domazet-Loso, and Iñaki Ruiz-Trillo. 2013. “Transcription Factor Evolution in Eukaryotes and the Assembly of the Regulatory Toolkit in Multicellular Lineages.” *Proceedings of the National Academy of Sciences of the United States of America* 110 (50): 1–9. doi:10.1073/pnas.1311818110.

de Mendoza, Alex, Hiroshi Suga, Jon Permanyer, Manuel Irimia, and Iñaki Ruiz-Trillo. 2015. “Complex Transcriptional Regulation and Independent Evolution of Fungal-like Traits in a Relative of Animals.” *eLife* 4 (OCTOBER2015): 1–26. doi:10.7554/eLife.08904.001.

Dean, Frank B, Seiyu Hosono, Linhua Fang, Xiaohong Wu, a Fawad Faruqi, Patricia Bray-Ward, Zhenyu Sun, et al. 2002. “Comprehensive Human Genome Amplification Using Multiple Displacement Amplification.” *Proceedings of the National Academy of Sciences of the United States of America* 99 (8): 5261–66. doi:10.1073/pnas.082089499.

de Vargas, Colomban, Stéphane Audic, Nicolas Henry, Johan Decelle, F. Mahe, Ramiro Logares, Enrique Lara, et al. 2015. “Eukaryotic Plankton Diversity in the Sunlit Ocean.” *Science (New York, N.Y.)* 348 (6237): 1261605–1261605. doi:10.1126/science.1261605.

del Campo, Javier, Diego Mallo, Ramon Massana, Colomban de Vargas, Tom A. Richards, and Inaki Ruiz-Trillo. 2015. “Diversity and Distribution of Unicellular Opisthokonts along the European Coast Analyzed Using High-Throughput Sequencing.” *Environmental Microbiology* 17 (9): 3195–3207. doi:10.1111/1462-2920.12759.

del Campo, Javier, and Iñaki Ruiz-Trillo. 2013. "Environmental Survey Meta-Analysis Reveals Hidden Diversity among Unicellular Opisthokonts." *Molecular Biology and Evolution* 30 (4): 802–5. doi:10.1093/molbev/mst006.

del Campo, Javier, Michael E. Sieracki, Robert Molestina, Patrick Keeling, Ramon Massana, and Iñaki Ruiz-Trillo. 2014. "The Others: Our Biased Perspective of Eukaryotic Genomes." *Trends in Ecology & Evolution* 29 (5): 252–59. doi:10.1016/j.tree.2014.03.006.

Derelle, Romain, Purificación López-García, Hélène Timpano, and David Moreira. 2016. "A Phylogenomic Framework to Study the Diversity and Evolution of Stramenopiles (= Heterokonts)." *Molecular Biology and Evolution* 33 (11): 2890–98.

Derelle, Romain, Guifré Torruella, and Vladimír Klime. 2015. "Bacterial Proteins Pinpoint a Single Eukaryotic Root," Proceedigsof the National Academy of Sciences 112 (7): E693-E699. doi:10.1073/pnas.1420657112.

Dìez, Beatriz, Carlos Pedrós-alió, and Ramon Massana. 2001. "Study of Genetic Diversity of Eukaryotic Picoplankton in Different Oceanic Regions by Small-Subunit rRNA Gene Cloning and Sequencing Study of Genetic Diversity of Eukaryotic Picoplankton in Different Oceanic Regions by Small-Subunit rRNA Gene Cloning and." *Applied and Environmental Microbiology* 67 (7): 2932–41. doi:10.1128/AEM.67.7.2932.

Dirren, Sebastian, Gianna Pitsch, Marisa O.D. Silva, and Thomas Posch. 2017. "Grazing of Nuclearia Thermophila and Nuclearia Delicatula (Nucleariidae, Opisthokonta) on the Toxic Cyanobacterium Planktothrix Rubescens." *European Journal of Protistology* 60. Elsevier GmbH: 87–101. doi:10.1016/j.ejop.2017.05.009.

Dirren, Sebastian, and Thomas Posch. 2016. "Promiscuous and Specific Bacterial Symbiont Acquisition in the

Amoeboid Genus Nuclearia (Opisthokonta).” *FEMS Microbiology Ecology* 92 (8): 1–16. doi:10.1093/femsec/fiw105.

Dos Reis, Mario, Yuttapong Thawornwattana, Konstantinos Angelis, Maximilian J. Telford, Philip C J Donoghue, and Ziheng Yang. 2015. “Uncertainty in the Timing of Origin of Animals and the Limits of Precision in Molecular Timescales.” *Current Biology* 25 (22). The Authors: 2939–50. doi:10.1016/j.cub.2015.09.066.

Dunn, Casey W., Gonzalo Giribet, Gregory D. Edgecombe, and Andreas Hejnol. 2013. “Animal Phylogeny and Its Evolutionary Implications.” *Annual Review of Ecology, Evolution, and Systematics* 45 (1): 141007174654001. doi:10.1146/annurev-ecolsys-120213-091627.

Dunn, Casey W, Andreas Hejnol, David Q Matus, Kevin Pang, William E Browne, Stephen A Smith, Elaine Seaver, et al. 2008. “Broad Phylogenomic Sampling Improves Resolution of the Animal Tree of Life” 452 (April): 745–50. doi:10.1038/nature06614.

Eme, Laura, Susan C Sharpe, Matthew W Brown, and Andrew J Roger. 2014. “On the Age of Eukaryotes: Evaluating Evidence from Fossils and Molecular Clocks.” doi:10.1101/cshperspect.a016139.

Erwin, D. H., M. Laflamme, S. M. Tweedt, E. A. Sperling, D. Pisani, and K. J. Peterson. 2011. “The Cambrian Conundrum: Early Divergence and Later Ecological Success in the Early History of Animals.” *Science* 334 (6059): 1091–97. doi:10.1126/science.1206375.

Esteban, G. F., A. J. Gooday, and K. J. Clarke. 2007. “Siliceous Scales of Filose-Amoebae (Pompholyxophryidae, Rotosphaerida) from Deep Southern Ocean Sediments, Including First Records for the Southern Hemisphere.” *Polar Biology* 30 (7): 945–50. doi:10.1007/s00300-007-0280-4.

Esteban, J A, M Salas, and L Blanco. 1993. “Fidelity of Phi 29 DNA Polymerase. Comparison between Protein-Primed Initiation and DNA Polymerization.” *J Biol Chem* 268.

Fairclough, Stephen R, Zehua Chen, Eric Kramer, Qiandong Zeng, Sarah Young, Hugh M Robertson, Emina Begovic, et al. 2013. “Premetazoan Genome Evolution and the Regulation of Cell Differentiation in the Choanoflagellate Salpingoeca Rosetta.” *Genome Biology* 14 (2): R15. doi:10.1186/gb-2013-14-2-r15.

Finet, Cédric, Ruth E. Timme, Charles F. Delwiche, and Ferdinand Marlétaz. 2010. “Multigene Phylogeny of the Green Lineage Reveals the Origin and Diversification of Land Plants.” *Current Biology* 20 (24): 2217–22. doi:10.1016/j.cub.2010.11.035.

Fonseca, Vera G, Gary R Carvalho, Way Sung, Harriet F Johnson, Deborah M Power, Simon P Neill, Margaret Packer, et al. 2010. “Second-Generation Environmental Sequencing Unmasks Marine Metazoan Biodiversity.” *Nature Communications* 1 (7). Nature Publishing Group: 98. doi:10.1038/ncomms1095.

Frank, Nitsche, Thomsen Helge Abuldhauge, and J. Richter Daniel. 2017. “Bridging the Gap between Morphological Species and Molecular Barcodes ??? Exemplified by Loricate Choanoflagellates.” *European Journal of Protistology* 57. Elsevier GmbH: 26–37. doi:10.1016/j.ejop.2016.10.006.

Gasmi, Samah, Gabriel Nève, Nicolas Pech, Saïda Tekaya, André Gilles, and Yvan Perez. 2014. “Evolutionary History of Chaetognatha Inferred from Molecular and Morphological Data : A Case Study for Body Plan Simplification,” 1–25.

Gawad, Charles, Winston Koh, and Stephen R. Quake. 2016. “Single-Cell Genome Sequencing: Current State of the Science.” *Nature Reviews Genetics* 17 (3). Nature Publishing Group: 175–88. doi:10.1038/nrg.2015.16.

Giribet, Gonzalo. 2015. “New Animal Phylogeny: Future Challenges for Animal Phylogeny in the Age of Phylogenomics.” doi:10.1007/s13127-015-0236-4.

Glockling, Sally L., Wyth L. Marshall, and Frank H. Gleason. 2013. “Phylogenetic Interpretations and Ecological Potentials of the Mesomycetozoea (Ichthyosporea).” *Fungal Ecology* 6 (4). Elsevier Ltd: 237–47. doi:10.1016/j.funeco.2013.03.005.

Gómez, Fernando, David Moreira, Karim Benzerara, and Purificación López-García. 2011. “Solenicola Setigera Is the First Characterized Member of the Abundant and Cosmopolitan Uncultured Marine Stramenopile Group MAST-3.” *Environmental Microbiology* 13: 193–202. doi:10.1111/j.1462-2920.2010.02320.x.

Gorokhova, Svetlana, Stéphanie Bibert, Käthi Geering, and Nathaniel Heintz. 2007. “A Novel Family of Transmembrane Proteins Interacting with β Subunits of the Na,K-ATPase.” *Human Molecular Genetics* 16 (20): 2394–2410. doi:10.1093/hmg/ddm167.

Grau-Bové, Xavier. 2017. “L’origen de La Multicel·lularitat En Animals: Una Aproximació Genòmica.” Universitat de Barcelona.

Grau-Bové, Xavier, Guifré Torruella, Stuart Donachie, Hiroshi Suga, Guy Leonard, Thomas A Richards, and Iñaki Ruiz-Trillo. 2017. “Dynamics of Genomic Innovation in the Unicellular Ancestry of Animals.” *eLife* 6. doi:10.7554/eLife.26036.

Haeckel, Ernts. 1874. “Die Gastraea-Theorie, Die Phylogenetische Klassifikation Des Thierreichs Und Die Homologie Der Keimblatter.” *Jenaische Z. Naturwiss* 8: 1–55.

Hardy, D M, and D L Garbers. 1995. “A Sperm Membrane-Protein That Binds in a Species-Specific Manner to the Egg Extracellular-Matrix Is Homologous to Von-

Willebrand-Factor." *Journal of Biological Chemistry* 270 (44): 26025–28.

Hassett, Brandon T., J. Andrés López, and Rolf Gradinger. 2015. "Two New Species of Marine Saprotrophic Sphaeroformids in the Mesomycetozoea Isolated From the Sub-Arctic Bering Sea." *Protist* 166 (3): 310–22. doi:10.1016/j.protis.2015.04.004.

Hehenberger, Elisabeth, Denis V. Tikhonenkov, Martin Kolisko, Javier del Campo, Anton S. Esaulov, Alexander P. Mylnikov, and Patrick J. Keeling. 2017. "Novel Predators Reshape Holozoan Phylogeny and Reveal the Presence of a Two-Component Signaling System in the Ancestor of Animals." *Current Biology*. 27 (13): p2043-2050. doi:10.1016/j.cub.2017.06.006.

Herr, Roger A., Libero Ajello, John W. Taylor, Sarath N. Arseculeratne, and Leonel Mendoza. 1999. "Phylogenetic Analysis of Rhinosporidium Seeberi's 18S Small-Subunit Ribosomal DNA Groups This Pathogen among Members of the Protoctistan Mesomycetozoa Clade." *Journal of Clinical Microbiology* 37 (9): 2750–54.

Hertel, Lynn A., Eric S. Loker, and Christopher J. Bayne. 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." *International Journal for Parasitology* 32 (9): 1183–91. doi:10.1016/S0020-7519(02)00066-8.

Heywood, Jane L, Michael E Sieracki, Wendy Bellows, Nicole J Poulton, and Ramunas Stepanauskas. 2011. "Capturing Diversity of Marine Heterotrophic Protists: One Cell at a Time." *The ISME Journal* 5 (4). Nature Publishing Group: 674–84. doi:10.1038/ismej.2010.155.

Hirai, J., M. Kuriyama, T. Ichikawa, K. Hidaka, and A. Tsuda. 2015. "A Metagenetic Approach for Revealing Community Structure of Marine Planktonic Copepods."

Molecular Ecology Resources 15 (1): 68–80. doi:10.1111/1755-0998.12294.

Humes, AG. 1994. “How Many Copepods?” *Hydrobiologia* 292/293 (1): 1–7.

James-Clark, H. 1866a. “Conclusive Proofs on the Animality of the Ciliat Sponges, and Their Affinities with Infusioria Flagellata.” *Am J Sci Ser* 2 (42): 320–25.

James, Timothy Y., and Mary L. Berbee. 2012. “No Jacket Required - New Fungal Lineage Defies Dress Code: Recently Described Zoosporic Fungi Lack a Cell Wall during Trophic Phase.” *BioEssays* 34 (2): 94–102. doi:10.1002/bies.201100110.

Jiang, Zhengwen, Xingqi Zhang, Ranjan Deka, and Li Jin. 2005. “Genome Amplification of Single Sperm Using Multiple Displacement Amplification.” *Nucleic Acids Research* 33 (10): e91. doi:10.1093/nar/gni089.

Jones, Meredith D M, Irene Forn, Catarina Gadelha, Martin J Egan, David Bass, Ramon Massana, and Thomas a Richards. 2011. “Discovery of Novel Intermediate Forms Redefines the Fungal Tree of Life.” *Nature* 474 (7350). Nature Publishing Group: 200–203. doi:10.1038/nature09984.

Kang, Dongwan D, Rob Egan, and Zhong Wang. 2015. “MetaBAT, an Efficient Tool for Accurately Reconstructing Single Genomes from Complex Microbial Communities,” *PeerJ* 3: e1165. doi:10.7717/peerj.1165.

Karpov, Sergey A., Maria A. Mamkaeva, Karim Benzerara, David Moreira, and Purificación López-García. 2014. “Molecular Phylogeny and Ultrastructure of Aphelidium Aff. Melosirae (Aphelida, Opisthosporidia).” *Protist* 165 (4): 512–26. doi:10.1016/j.protis.2014.05.003.

Karpov, Sergey a, Kirill V Mikhailov, Gulnara S Mirzaeva, Iskandar M Mirabdullaev, Kira a Mamkaeva, Nina N

Titova, and Vladimir V Aleoshin. 2013. “Obligately Phagotrophic Aphelids Turned out to Branch with the Earliest-Diverging Fungi.” *Protist* 164(2): 195-205. doi:10.1016/j.protis.2012.08.001.

Keeling, Patrick. 2009. “Five Questions about Microsporidia.” *PLoS Pathogens* 5 (9): 1–3. doi:10.1371/journal.ppat.1000489.

King, N. 2003. “Evolution of Key Cell Signaling and Adhesion Protein Families Predates Animal Origins.” *Science* 301 (5631): 361–63. doi:10.1126/science.1083853.

King, Nicole. 2004. “The Unicellular Ancestry of Animal Development.” *Developmental Cell* 7 (3): 313–25. doi:10.1016/j.devcel.2004.08.010.

King, Nicole, M Jody Westbrook, Susan L Young, Alan Kuo, Monika Abedin, Jarrod Chapman, Stephen Fairclough, et al. 2008. “The Genome of the Choanoflagellate Monosiga Brevicollis and the Origin of Metazoans.” *Nature* 451 (7180): 783–88. doi:10.1038/nature06617.

Knoll, A. H., and E. A. Sperling. 2014. “Oxygen and Animals in Earth History.” *Proceedings of the National Academy of Sciences* 111 (11): 3907–8. doi:10.1073/pnas.1401745111.

Knoll, Andrew H. 2011. “The Multiple Origins of Complex Multicellularity.” *Annual Review of Earth and Planetary Sciences* 39 (1): 217–39. doi:10.1146/annurev.earth.031208.100209.

Lang, B. F., C. O’Kelly, T. Nerad, M. W. Gray, and G. Burger. 2002. “The Closest Unicellular Relatives of Animals.” *Current Biology* 12 (20): 1773–78. doi:10.1016/S0960-9822(02)01187-9.

Lara, Enrique, David Moreira, and Purificación López-García. 2010. “The Environmental Clade LKM11 and Rozella Form the Deepest Branching Clade of Fungi.” *Protist* 161

(1): 116–21. doi:10.1016/j.protis.2009.06.005.

Leadbeater, B.S.C. 2015. *The Choanoflagellates: Evolution, Biology, and Ecology*. Cambridge: Cambridge University Press.

Lechauve, Christophe, Muriel Jager, Laurent Laguerre, Laurent Kiger, Gaëlle Correc, Cédric Leroux, Serge Vinogradov, Mirjam Czjzek, Michael C. Marden, and Xavier Bailly. 2013. “Neuroglobins, Pivotal Proteins Associated with Emerging Neural Systems and Precursors of Metazoan Globin Diversity.” *Journal of Biological Chemistry* 288 (10): 6957–67. doi:10.1074/jbc.M112.407601.

Lee, S. C., M. Ni, W. Li, C. Shertz, and J. Heitman. 2010. “The Evolution of Sex: A Perspective from the Fungal Kingdom.” *Microbiology and Molecular Biology Reviews* 74 (2): 298–340. doi:10.1128/MMBR.00005-10.

Levin, Tera C., and Nicole King. 2013. “Evidence for Sex and Recombination in the Choanoflagellate Salpingoeca Rosetta.” *Current Biology* 23(21): 2176-80. doi:10.1016/j.cub.2013.08.061.

Levin, Tera C, Allison J Greaney, Laura Wetzel, and Nicole King. 2014. “The Rosetteless Gene Controls Development in the Choanoflagellate S. Rosetta.” *eLife* 3 (October): 1–23. doi:10.7554/eLife.04070.

Lichtwardt, RW, MJ Cafaro, and MM White. 2001. *The Trichomycetes: Fungal Associates of Arthropods*. Edited by Joseph Seckbach. Springer.

Liebeskind, Benjamin J. 2011. “Evolution of Sodium Channels and the New View of Early Nervous System Evolution.” *Evolution*, no. December: 679–83. doi:10.4161/cib.4.6.17069.

Lindeque, Penelope K, Helen E Parry, Rachel a Harmer, Paul J Somerfield, and Angus Atkinson. 2013. “Next

Generation Sequencing Reveals the Hidden Diversity of Zooplankton Assemblages." *PloS One* 8 (11): e81327. doi:10.1371/journal.pone.0081327.

Liu, Yu, Emma T Steenkamp, Henner Brinkmann, Lise Forget, Hervé Philippe, and B Franz Lang. 2009. "Phylogenomic Analyses Predict Sistergroup Relationship of Nucleariids and Fungi and Paraphyly of Zygomycetes with Significant Support." *BMC Evolutionary Biology* 9: 272. doi:10.1186/1471-2148-9-272.

Logares, Ramiro, Stéphane Audic, David Bass, Lucie Bittner, Christophe Boutte, Richard Christen, Jean-Michel Claverie, et al. 2014. "Patterns of Rare and Abundant Marine Microbial Eukaryotes." *Current Biology : CB* 24 (8): 813–21. doi:10.1016/j.cub.2014.02.050.

Logares, Ramiro, Stephane Audic, Sebastien Santini, Massimo C Pernice, Colomban de Vargas, and Ramon Massana. 2012. "Diversity Patterns and Activity of Uncultured Marine Heterotrophic Flagellates Unveiled with Pyrosequencing." *The ISME Journal* 6 (10). Nature Publishing Group: 1823–33. doi:10.1038/ismej.2012.36.

López-García, P, F Rodríguez-Valera, C Pedrós-Alió, and D Moreira. 2001. "Unexpected Diversity of Small Eukaryotes in Deep-Sea Antarctic Plankton." *Nature* 409 (February): 603–7. doi:10.1038/35054537.

Love, Gordon D, Emmanuelle Grosjean, Charlotte Stalvies, David A Fike, John P Grotzinger, Alexander S Bradley, Amy E Kelly, et al. 2008. "Fossil Steroids Record the Appearance of Demospongiae during the Cryogenian Period." *Nature* 457 (7230). Nature Publishing Group: 718–21. doi:10.1038/nature07673.

Lyons, Timothy W, Christopher T Reinhard, and Noah J Planavsky. 2014. "The Rise of Oxygen in Earth's Early Ocean and Atmosphere." *Nature* 506 (7488). Nature Publishing Group: 307–15. doi:10.1038/nature13068.

Magner, Lois N. 2002. “History of the Life Sciences.” In , 3rd ed. CRC Press.

Mangot, Jean-françois, Ramiro Logares, Pablo Sanchez, and Fran Latorre. 2017. “Accessing to the Genomic Information of Unculturable Oceanic Picoeukaryotes by Combining Multiple Single Cells.” *Scientific Reports* 7: 41498. doi:10.1038/srep41498.

Markmann, Melanie, and Diethard Tautz. 2005. “Reverse Taxonomy: An Approach towards Determining the Diversity of Meiobenthic Organisms Based on Ribosomal RNA Signature Sequences.” *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences* 360 (1462): 1917–24. doi:10.1098/rstb.2005.1723.

Marshall, Wyth L., and Mary L. Berbee. 2011. “Facing Unknowns: Living Cultures (Pirum Gemmata Gen. Nov., Sp. Nov., and Abeoforma Whisleri, Gen. Nov., Sp. Nov.) from Invertebrate Digestive Tracts Represent an Undescribed Clade within the Unicellular Opisthokont Lineage Ichthyosporea (Mesomycetozoea).” *Protist* 162 (1). Elsevier GmbH.: 33–57. doi:10.1016/j.protis.2010.06.002.

Marshall, Wyth L, and Mary L Berbee. 2013. “Comparative Morphology and Genealogical Delimitation of Cryptic Species of Sympatric Isolates of Sphaeroforma (Ichthyosporea, Opisthokonta).” *Protist* 164(2):287-311. doi:10.1016/j.protis.2012.12.002.

Massana, Ramon, Jose Castresana, Vanessa Balague, Laure Guillou, Khadidja Romari, Klaus Valentin, and Carlos Pedro. 2004. “Phylogenetic and Ecological Analysis of Novel Marine Stramenopiles.” *Applied and Environmental Microbiology* 70 (6): 3528–34. doi:10.1128/AEM.70.6.3528.

Massana, Ramon, Massimo Pernice, John a Bunge, and Javier del Campo. 2011. “Sequence Diversity and

Novelty of Natural Assemblages of Picoeukaryotes from the Indian Ocean." *The ISME Journal* 5 (2). Nature Publishing Group: 184–95. doi:10.1038/ismej.2010.104.

Massana, Ramon, Ramon Terrado, Irene Forn, Connie Lovejoy, and Carlos Pedrós-Alió 2006. "Distribution and Abundance of Uncultured Heterotrophic Flagellates in the World Oceans." *Environmental Microbiology* 8 (9): 1515–22. doi:10.1111/j.1462-2920.2006.01042.x.

Medina, Mnica, Allen G. Collins, John W. Taylor, James W. Valentine, Jere H. Lipps, Linda Amaral-Zettler, and Mitchell L. Sogin. 2003. "Phylogeny of Opisthokonta and the Evolution of Multicellularity and Complexity in Fungi and Metazoa." *International Journal of Astrobiology* 2 (3): 203–11. doi:10.1017/S1473550403001551.

Medlin, Linda, Hille J. Elwood, Shawn Stickel, and Mitchell L. Sogin. 1988. "The Characterization of Enzymatically Amplified Eukaryotic 16S-like rRNA-Coding Regions." *Gene* 71 (2): 491–99. doi:10.1016/0378-1119(88)90066-2.

Meixner, Martin J, Carsten Lu, Carsten Eckert, Valeria Itskovich, Dorte Janussen, Alexandra V Bohne, Johannes M Meixner, and Wolfgang R Hess. 2007. "Phylogenetic Analysis of Freshwater Sponges Provide Evidence for Endemism and Radiation in Ancient Lakes" 45: 875–86. doi:10.1016/j.ympev.2007.09.007.

Mende, Daniel R., Frank O. Aylward, John M. Eppley, Torben N. Nielsen, and Edward F. DeLong. 2016. "Improved Environmental Genomes via Integration of Metagenomic and Single-Cell Assemblies." *Frontiers in Microbiology* 7 (February): 1–9. doi:10.3389/fmicb.2016.00143.

Mendoza, Leonel, John W Taylor, and Libero Ajello. 2002. "The Class Mesomycetozoea: A Heterogeneous Group of Microorganisms at the Animal-Fungal Boundary." *Annual Review of Microbiology* 56: 315–44. doi:10.1146/annurev.micro.56.012302.160950.

Mikhailov, Kirill V, Anastasiya V Konstantinova, Mikhail A Nikitin, Peter V Troshin, Vassily A Lyubetsky, Yuri V Panchin, Alexander P Mylnikov, Leonid L Moroz, Sudhir Kumar, and Vladimir V Aleoshin. 2009. “The Origin of Metazoa: A Transition from Temporal to Spatial Cell Differentiation,” 758–68. doi:10.1002/bies.200800214.

Mikrjukov, Kirill A. 1999. “Taxonomic Revision of Scale-Bearing Heliozoon-like Amoebae (Pompholyxophryidae, Rotosphaerida).” *Acta Protozoologica* 38 (2): 119–31.

Mikrjukov, Kirill A., and Alexander P. Mylnikov. 2001. “A Study of the Fine Structure and the Mitosis of a Lamellicristate Amoeba, Micronuclearia Podoventralis Gen. et Sp. Nov. (Nucleariidae, Rotosphaerida).” *European Journal of Protistology* 37 (1): 15–24. doi:10.1078/0932-4739-00783.

Mills, Daniel B., and Donald E. Canfield. 2014. “Oxygen and Animal Evolution: Did a Rise of Atmospheric Oxygen ‘trigger’ the Origin of Animals?” *BioEssays* 36 (12): 1145–55. doi:10.1002/bies.201400101.

Mills, Daniel B, Lewis M Ward, Carriayne Jones, Brittany Sweeten, Michael Forth, Alexander H Treusch, and Donald E Canfield. 2014. “Oxygen Requirements of the Earliest Animals.” *Proceedings of the National Academy of Sciences of the United States of America* 111 (11): 4168–72. doi:10.1073/pnas.1400547111.

Moczydlowska, Malgorzata, Victoria Pease, Willman Sebastian, Linda Wickström, and Heda Agic. 2017. “A Tonian Age for the Visingsö Group in Sweden Constrained by Detrital Zircon Dating and Biochronology: Implications for Evolutionary Events.” *Geol. Mag.*, 1–15. doi:10.1017/S0016756817000085.

Moon-van der Staay, Seung Yeo, R de Wachter, and Daniel Vaulot. 2001. “Oceanic 18S rDNA Sequences from Picoplankton Reveal Unsuspected Eukaryotic Diversity.” *Nature* 409 (6820): 607–610. doi:10.1038/35054541.

Moreira, David, and Purificación López-García. 2002. “The Molecular Ecology of Microbial Eukaryotes Unveils a Hidden World” 10 (1): 31–38.

Moroz, Leonid L., Kevin M. Kocot, Mathew R. Citarella, Sohn Dosung, Tigran P. Norekian, Inna S. Povolotskaya, Anastasia P. Grigorenko, et al. 2014. “The Ctenophore Genome and the Evolutionary Origins of Neural Systems.” *Nature* 510 (7503). Nature Publishing Group: 109–14. doi:10.1038/nature13400.

Mutlu Bingel, F., E. 1999. “Distribution and Abundance of Ctenophores, and Their Zooplakton Food in the Black Sea. 1. Pleurobrachia Pileus.” *Marine Biology* 135 (1999): 589–601.

Newman, S A. 2012. “Physico-Genetic Determinants in the Evolution of Development,” no. October: 217–20.

Nguyen, Tu Anh, Ousmane H. Cissé, Jie Yun Wong, Peng Zheng, David Hewitt, Minou Nowrousian, Jason E. Stajich, and Gregory Jedd. 2017. “Innovation and Constraint Leading to Complex Multicellularity in the Ascomycota.” *Nature Communications* 8: 14444. doi:10.1038/ncomms14444.

Nielsen, Claus. 2008. “Six Major Steps in Animal Evolution: Are We Derived Sponge Larvae?” *Evolution and Development* 10 (2): 241–57. doi:10.1111/j.1525-142X.2008.00231.x.

Nielsen, Claus. 2012. *Animal Evolution: Interrelationships of the Living Phyla*. New York: Oxford University Press.

Nosenko, Tetyana, Fabian Schreiber, Maja Adamska, Marcin Adamski, Michael Eitel, Jörg Hammel, Manuel Maldonado, et al. 2013. “Deep Metazoan Phylogeny: When Different Genes Tell Different Stories.” *Molecular Phylogenetics and evolution*, 1–11. doi:10.1016/j.ympev.2013.01.010.

Paps, Jordi, Luis a Medina-Chacón, Wyth Marshall, Hiroshi Suga, and Iñaki Ruiz-Trillo. 2013. “Molecular Phylogeny of Unikonts: New Insights into the Position of Apusomonads and Ancyromonads and the Internal Relationships of Opisthokonts.” *Protist* 164 (1): 2–12. doi:10.1016/j.protis.2012.09.002.

Parfrey, Laura Wegener, Erika Barbero, Elyse Lasser, Micah Dunthorn, Debashish Bhattacharya, David J Patterson, and Laura a Katz. 2006. “Evaluating Support for the Current Classification of Eukaryotic Diversity.” *PLoS Genetics* 2 (12): e220. doi:10.1371/journal.pgen.0020220.

Patterson, David J. 1983. “On the Organization of the Naked Filose Amoeba, Nuclearia Moebiusi Frenzel, 1897 (Sarcodina, Filosea) and Its Implications.” *The Journal of Protozoology* 30 (2): 301–7. doi:10.1111/j.1550-7408.1983.tb02920.x.

Patterson, David J 1984. “The Genus Nuclearia (Sarcodina, Filosea): Species Composition and Characteristics of the Taxa.” *Archiv Für Protistenkunde*. doi:10.1016/S0003-9365(84)80034-2.

Pawlowski, Jan, Philippe Esling, Franck Lejzerowicz, Tomas Cedhagen, and Thomas A. Wilding. 2014. “Environmental Monitoring through Protist next-Generation Sequencing Metabarcoding: Assessing the Impact of Fish Farming on Benthic Foraminifera Communities.” *Molecular Ecology Resources* 14 (6): 1129–40. doi:10.1111/1755-0998.12261.

Pearman, John K., Mohsen M. El-Sherbiny, Anders Lanzen, Ali M. Al-Aidaroos, and Xabier Irigoien. 2014. “Zooplankton Diversity across Three Red Sea Reefs Using Pyrosequencing.” *Frontiers in Marine Science* 1 (July): 1–11. doi:10.3389/fmars.2014.00027.

Pearman, John K., and Xabier Irigoien. 2015. “Assessment of Zooplankton Community Composition along a Depth

Profile in the Central Red Sea." *PLoS ONE* 10 (7): 1–14. doi:10.1371/journal.pone.0133487.

Philippe, Hervé, Romain Derelle, Philippe Lopez, Kerstin Pick, Carole Borchiellini, Nicole Boury-Esnault, Jean Vacelet, et al. 2009. "Phylogenomics Revives Traditional Views on Deep Animal Relationships." *Current Biology : CB* 19 (8): 706–12. doi:10.1016/j.cub.2009.02.052.

Ponce-Toledo, Rafael, Philippe Deschamps, Purificación López-garcía, Yvan Zivanovic, Karim Benzerara, and David Moreira. 2016. "An Early-Branching Freshwater Cyanobacterium at the Origin of Plastids." *Current Biology* 27 (3): 386–91. doi:10.1016/j.cub.2016.11.056.

Porazinska, Dorota L, Way Sung, Robin M Giblin-Davis, and W Kelley Thomas. 2010. "Reproducibility of Read Numbers in High-Throughput Sequencing Analysis of Nematode Community Composition and Structure." *Molecular Ecology Resources* 10 (4): 666–76. doi:10.1111/j.1755-0998.2009.02819.x.

Putnam, N. H., M. Srivastava, U. Hellsten, B. Dirks, J. Chapman, A. Salamov, A. Terry, et al. 2007. "Sea Anemone Genome Reveals Ancestral Eumetazoan Gene Repertoire and Genomic Organization." *Science* 317 (5834): 86–94. doi:10.1126/science.1139158.

Raghu-Kama, S., D. Chandramohan, and N. Ramaiah. 1987. "Contribution of the Thraustochytrid Corallochytrium Limacisporum Raghu-Kumar to Microbial Biomass in Coral Reef Lagoons." *Indian J. Mar. Sci.* 16: 122–25.

Reports, Cell. 2012. "Report Premetazoan Origin of the Hippo Signaling Pathway." *Cell* 1 (1): 1–8. doi:10.1016/j.celrep.2011.11.004.

Richards, Thomas A., Meredith D.M. Jones, Guy Leonard, and David Bass. 2012. "Marine Fungi: Their Ecology and Molecular Diversity." *Annual Review of Marine Science* 4

(1): 495–522. doi:10.1146/annurev-marine-120710-100802.

Richards, Thomas A, Guy Leonard, F. Mahe, J. del Campo, Sarah Romac, Meredith D M Jones, Finlay Maguire, et al. 2015. “Molecular Diversity and Distribution of Marine Fungi across 130 European Environmental Samples.” *Proceedings of the Royal Society B: Biological Sciences* 282 (1819): 1–10. doi:10.1098/rspb.2015.2243

Richards, Thomas A, Guy Leonard, and Jeremy G Wideman. 2017. “What Defines the ‘ Kingdom ’ Fungi ?,” Microbiol Spectr. Jun; 5(3):1–21. doi:10.1128/microbiolspec.FUNK-0044-2017.

Richter, Daniel, and Nicole King. 2013. “The Genomic and Cellular Foundations of Animal Origins.” *Annual Review of Genetics*, 1–31. doi:10.1146/annurev-genet-111212-133456.

Rinke, C, J Lee, N Nath, D Goudeau, B Thompson, and N Poulton. 2014. “Obtaining Genomes from Uncultivated Environmental Microorganisms Using FACS-Based Single-Cell Genomics.” *Nat Protoc* 9. doi:10.1038/nprot.2014.067.

Rinke, Christian, Patrick Schwientek, Alexander Sczyrba, Natalia N Ivanova, Iain J Anderson, Jan-Fang Cheng, Aaron Darling, et al. 2013. “Insights into the Phylogeny and Coding Potential of Microbial Dark Matter.” *Nature* 499 (7459): 431–37. doi:10.1038/nature12352.

Roy, Rajat S, Dana C Price, Alexander Schliep, Guohong Cai, Anton Korobeynikov, Hwan Su Yoon, Eun Chan Yang, and Debashish Bhattacharya. 2014. “Single Cell Genome Analysis of an Uncultured Heterotrophic Stramenopile.” *Scientific Reports* 4: 4780. doi:10.1038/srep04780.

Ruiz-Trillo, Iñaki, Gertraud Burger, Peter W H Holland, Nicole King, B Franz Lang, Andrew J Roger, and Michael W Gray. 2007. “The Origins of Multicellularity: A Multi-Taxon Genome Initiative.” *Trends in Genetics : TIG* 23 (3): 113–18. doi:10.1016/j.tig.2007.01.005.

Ruiz-Trillo, Iñaki, Yuji Inagaki, Lesley A Davis, Bjarne Landfald, and Andrew J Roger. 2004. “Capsaspora Owczarzaki Is an Independent Opisthokont Lineage Known as the DRIPs or” 14 (22): 946–47.

Ruiz-Trillo, Iñaki, Christopher E. Lane, John M. Archibald, and Andrew J. Roger. 2006. “Insights into the Evolutionary Origin and Genome Architecture of the Unicellular Opisthokonts Capsaspora Owczarzaki and Sphaeroforma Arctica.” *Journal of Eukaryotic Microbiology* 53 (5): 379–84. doi:10.1111/j.1550-7408.2006.00118.x.

Ruiz-Trillo, Iñaki, Andrew J. Roger, Gertraud Burger, Michael W. Gray, and B. Franz Lang. 2008. “A Phylogenomic Investigation into the Origin of Metazoa.” *Molecular Biology and Evolution* 25 (4): 664–72. doi:10.1093/molbev/msn006.

Ryan, J. F., K. Pang, C. E. Schnitzler, A.-D. Nguyen, R. T. Moreland, D. K. Simmons, B. J. Koch, et al. 2013. “The Genome of the Ctenophore Mnemiopsis Leidyi and Its Implications for Cell Type Evolution.” *Science* 342 (6164): 1242592–1242592. doi:10.1126/science.1242592.

Sebé-Pedrós, Arnau, Ana Ariza-Cosano, Matthew Weirauch, Sven Lininger, Yang Ally, Guifré Torruella, Marcin Adamski, et al. 2013. “Early Evolution of the T-Box Transcription Factor Family.” *Proceedings of the National Academy of Sciences of the United States of America* 110 (40): 16050–55. doi:10.1073/pnas.1309748110/-/DCSupplemental.www.pnas.org/cgi/doi/10.1073/pnas.1309748110.

Sebé-Pedrós, Arnau, Andrew J Roger, Franz B Lang, Nicole King, and Iñaki Ruiz-Trillo. 2010. “Ancient Origin of the Integrin-Mediated Adhesion and Signaling Machinery.” *Proceedings of the National Academy of Sciences of the United States of America* 107 (22): 10142–47. doi:10.1073/pnas.1002257107.

Sebé-Pedrós, Arnau, Alex De Mendoza, B. Franz Lang, Bernard M. Degnan, and Iñaki Ruiz-Trillo. 2011. “Unexpected Repertoire of Metazoan Transcription Factors in the Unicellular Holozoan Capsaspora Owczarzaki.” *Molecular Biology and Evolution* 28 (3): 1241–54. doi:10.1093/molbev/msq309.

Sebé-Pedrós, Arnau, Yonggang Zheng, Iñaki Ruiz-Trillo, and Duojia Pan. 2012. “Premetazoan Origin of the Hippo Signaling Pathway.” *Cell Reports* 1 (1): 13–20. doi:10.1016/j.celrep.2011.11.004.

Sebé-Pedrós, Arnau, Manuel Irimia, Javier del Campo, Helena Parra-acero, Carsten Russ, Chad Nusbaum, Benjamin J Blencowe, and Iñaki Ruiz-Trillo. 2013. “Regulated Aggregative Multicellularity in a Close Unicellular Relative of Metazoa,” *eLife* 2: e01287. doi:10.7554/eLife.01287.

Sebé-Pedrós, Arnau, Pawel Burkhardt, Núria Sánchez-Pons, Stephen R. Fairclough, B. Franz Lang, Nicole King, and Iñaki Ruiz-Trillo. 2013. “Insights into the Origin of Metazoan Filopodia and Microvilli.” *Molecular Biology and Evolution* 30 (9): 2013–2023. doi:10.1093/molbev/mst110.

Sebé-Pedrós, Arnau, Cecilia Ballarí, Helena Parra-Acero, Cristina Chiva, Juan J. Tena, Eduard Sabidó, José Luis Gómez-Skarmeta, Luciano Di Croce, and Iñaki Ruiz-

Trillo. 2015. “The Dynamic Regulatory Genome of Capsaspora and the Origin of Animal Multicellularity.” *Cell* 165(5):1224–1237. doi:10.1016/j.cell.2016.03.034.

Sebé-Pedrós, Arnau, Marcia Ivonne Peña, Salvador Capella-Gutiérrez, Meritxell Antó, Toni Gabaldón, Iñaki Ruiz-Trillo, and Eduard Sabidó. 2016. “High-Throughput Proteomics Reveals the Unicellular Roots of Animal Phosphosignaling and Cell Differentiation,” *Developmental cell* 39(2): p186–197.

Sebé-Pedrós, Arnau, Bernard M Degnan, and Iñaki Ruiz-Trillo. 2017. “The Origin of Metazoa, a Unicellular Perspective.” *Nature Reviews Genetics* 18(8):498-512. doi:10.1038/nrg.2017.21.

Shalchian-Tabrizi, Kamran, Marianne a Minge, Mari Espelund, Russell Orr, Torgeir Ruden, Kjetill S Jakobsen, and Thomas Cavalier-Smith. 2008. “Multigene Phylogeny of Choanozoa and the Origin of Animals.” *PloS One* 3 (5): e2098. doi:10.1371/journal.pone.0002098.

Shen, Xin, Song Sun, Fang Qing Zhao, Guang Tao Zhang, Mei Tian, Ling Ming Tsang, Jin Feng Wang, and Ka Hou Chu. 2016. “Phylomitogenomic Analyses Strongly Support the Sister Relationship of the Chaetognatha and Protostomia.” *Zoologica Scripta* 45 (2): 187–99. doi:10.1111/zsc.12140.

Shen, Xing-Xing, Chris Todd Hittinger, and Antonis Rokas. 2017. “Contentious Relationships in Phylogenomic Studies Can Be Driven by a Handful of Genes.” *Nature Ecology & Evolution* 1 (5). Macmillan Publishers Limited, part of Springer Nature.: 126. doi:10.1038/s41559-017-0126.

Shiratori, Takashi, Rabindra Thakur, and Ken-ichiro Ishida. 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." *Annals of Anatomy*. Elsevier GmbH. doi:10.1016/j.protis.2017.06.004.

Simakov, Oleg, and Takeshi Kawashima. 2016. "Independent Evolution of Genomic Characters during Major Metazoan Transitions." *Developmental Biology* 427 (August 2016): 179–92. doi:10.1016/j.ydbio.2016.11.012.

Simion, Paul, Hervé Philippe, Denis Baurain, Muriel Jager, Daniel J. Richter, Arnaud Di Franco, Béatrice Roure, et al. 2017. "A Large and Consistent Phylogenomic Dataset Supports Sponges as the Sister Group to All Other Animals." *Current Biology* 27(7):958–967. doi:10.1016/j.cub.2017.02.031.

Snelgrove, Paul V. R. 1999. "Getting to the Bottom of Marine Biodiversity: Sedimentary Habitats." *BioScience* 49 (2): 129. doi:10.2307/1313538.

Spatafora, Joseph W, Ying Chang, Gerald L Benny, Katy Lazarus, Matthew E Smith, Mary L Berbee, Gregory Bonito, et al. 2016. "A Phylum-Level Phylogenetic Classification of Zygomycete Fungi Based on Genome-Scale Data." *Mycologia* 108 (5): 1028–46. doi:10.3852/16-042.

Sperling, Erik A., Christina A. Frieder, Akkur V. Raman, Peter R. Girguis, Lisa A. Levin, and Andrew H. Knoll. 2013. "Oxygen, Ecology, and the Cambrian Radiation of Animals." *Proceedings of the National Academy of Sciences of the United States of America* 110 (33): 13446–51. doi:10.1073/pnas.1312778110.

Srivastava, Mansi, Oleg Simakov, Jarrod Chapman, Bryony Fahey, Marie E A Gauthier, Therese Mitros, Gemma S Richards, et al. 2010. "The Amphimedon Queenslandica Genome and the Evolution of Animal Complexity." *Nature* 466 (7307). Nature Publishing Group: 720–26. doi:10.1038/nature09201.

Steenkamp, Emma T., Jane Wright, and Sandra L. Baldauf. 2006. “The Protistan Origins of Animals and Fungi.” *Molecular Biology and Evolution* 23 (1): 93–106. doi:10.1093/molbev/msj011.

Stepanauskas, Ramunas. 2012. “Single Cell Genomics: An Individual Look at Microbes.” *Current Opinion in Microbiology* 15 (5). Elsevier Ltd: 613–20. doi:10.1016/j.mib.2012.09.001.

Stepanauskas, Ramunas, Elizabeth A. Fergusson, Joseph Brown, Nicole J. Poulton, Ben Tupper, Jessica M. Labonté, Eric D. Becraft, et al. 2017. “Improved Genome Recovery and Integrated Cell-Size Analyses of Individual Uncultured Microbial Cells and Viral Particles.” *Nature Communications* 8 (1). Springer US: 84. doi:10.1038/s41467-017-00128-z.

Stepanauskas, Ramunas, and Michael E Sieracki. 2007. “Matching Phylogeny and Metabolism in the Uncultured Marine Bacteria, One Cell at a Time.” *Proceedings of the National Academy of Sciences of the United States of America* 104 (21): 9052–57. doi:10.1073/pnas.0700496104.

Stibbs, Henry H., Alfred Owczarzak, Christopher J. Bayne, and Peggy DeWan. 1979. “Schistosome Sporocyst-Killing Amoebae Isolated from Biomphalaria Glabrata.” *Journal of Invertebrate Pathology* 33 (2): 159–70. doi:10.1016/0022-2011(79)90149-6.

Stoeck, Thorsten, David Bass, Markus Nebel, Richard Christen, Meredith D M Jones, Hans-Werner Breiner, and Thomas A Richards. 2010. “Multiple Marker Parallel Tag Environmental DNA Sequencing Reveals a Highly Complex Eukaryotic Community in Marine Anoxic Water.” *Molecular Ecology* 19 (March): 21–31. doi:10.1111/j.1365-294X.2009.04480.x.

Suga, H., M. Dacre, A. de Mendoza, K. Shalchian-Tabrizi, G. Manning, and I. Ruiz-Trillo. 2012. “Genomic Survey of

Premetazoans Shows Deep Conservation of Cytoplasmic Tyrosine Kinases and Multiple Radiations of Receptor Tyrosine Kinases." *Science Signaling* 5 (222): ra35-ra35. doi:10.1126/scisignal.2002733.

Suga, Hiroshi, Zehua Chen, Alex de Mendoza, Arnau Sebé-Pedrós, Matthew W Brown, Eric Kramer, Martin Carr, et al. 2013. "The *Capsaspora* Genome Reveals a Complex Unicellular Prehistory of Animals." *Nature Communications* 4: 2325. doi:10.1038/ncomms3325.

Suga, Hiroshi, Guifré Torruella, Gertraud Burger, Matthew W. Brown, and Iñaki Ruiz-Trillo. 2014. "Earliest Holozoan Expansion of Phosphotyrosine Signaling." *Molecular Biology and Evolution* 31 (3): 517–28. doi:10.1093/molbev/mst241.

Surek, Barbara, and Michael Melkonian. 1980. "The Filose Amoeba Vampyrellidium Perforans Nov. Sp. (Vampyrellidae, Aconchulinida): Axenic Culture, Feeding Behaviour and Host Range Specifity." *Archiv Für Protistenkunde* 123 (2): 166–91. doi:10.1016/S0003-9365(80)80003-0.

Swan, Brandon K, Manuel Martinez-Garcia, Christina M Preston, Alexander Sczyrba, Tanja Woyke, Dominique Lamy, Thomas Reinthaler, et al. 2011. "Potential for Chemolithoautotrophy among Ubiquitous Bacteria Lineages in the Dark Ocean." *Science (New York, N.Y.)* 333 (6047): 1296–1300. doi:10.1126/science.1203690.

Tang, Cuong Q, Francesca Leasi, Ulrike Obertegger, Alexander Kieneke, Timothy G Barraclough, and Diego Fontaneto. 2012. "The Widely Used Small Subunit 18S rDNA Molecule Greatly Underestimates True Diversity in Biodiversity Surveys of the Meiofauna." *Proceedings of the National Academy of Sciences of the United States of America* 109 (40): 16208–12. doi:10.1073/pnas.1209160109.

Tautz, Diethard, and Tomislav Domazet-los. 2010. “A Phylogenetically Based Transcriptome Age Index Mirrors Ontogenetic Divergence Patterns,” *Nature* 468: 815–818. doi:10.1038/nature09632.

Telford, Maximilian J. 2013. “Evolution. The Animal Tree of Life.” *Science (New York, N.Y.)* 339 (6121): 764–66. doi:10.1126/science.1234378.

Torruella, Guifré, Romain Derelle, Jordi Paps, B Franz Lang, Andrew J Roger, Kamran Shalchian-Tabrizi, and Iñaki Ruiz-Trillo. 2012. “Phylogenetic Relationships within the Opisthokonta Based on Phylogenomic Analyses of Conserved Single-Copy Protein Domains.” *Molecular Biology and Evolution* 29 (2): 531–44. doi:10.1093/molbev/msr185.

Torruella, Guifré, Alex De Mendoza, Xavier Grau-Bové, Meritxell Antó, Mark A. Chaplin, Javier Del Campo, Laura Eme, et al. 2015. “Phylogenomics Reveals Convergent Evolution of Lifestyles in Close Relatives of Animals and Fungi.” *Current Biology* 25 (18): 2404–10. doi:10.1016/j.cub.2015.07.053.

Towe, Kenneth M. 1970. “Oxygen-Collagen Priority and the Early Metazoan Fossil Record.” *Proceedings of the National Academy of Sciences* 65 (4): 781–88. doi:10.1073/pnas.65.4.781.

Troell, Karin, Björn Hallström, Anna-maria Divne, Cecilia Alsmark, Romanico Arrighi, Mikael Huss, Jessica Beser, and Stefan Bertilsson. 2016. “Cryptosporidium as a Testbed for Single Cell Genome Characterization of Unicellular Eukaryotes.” *BMC Genomics*. BMC Genomics, 1–12. doi:10.1186/s12864-016-2815-y.

Tsagkogeorga, Georgia, Xavier Turon, Russell R Hopcroft, Ka Tilak, Tamar Feldstein, Noa Shenkar, Yossi Loya, Dorothée Huchon, Emmanuel J P Douzery, and Frédéric Delsuc. 2009. “An Updated 18S rRNA Phylogeny of Tnunicates Based on Mixture and Secondary Structure

Models.” *BMC Evolutionary Biology* 16 (9:187): 1–16. doi:10.1186/1471-2148-9-187.

Uriarte, Ibon, and Fernando Villate. 2004. “Effects of Pollution on Zooplankton Abundance and Distribution in Two Estuaries of the Basque Coast (Bay of Biscay)” 49: 220–28. doi:10.1016/j.marpolbul.2004.02.010.

Vávra, Jiří, and Julius Lukeš. 2013. *Microsporidia and “the Art of Living Together”. Advances in Parasitology*. Vol. 82. Elsevier. doi:10.1016/B978-0-12-407706-5.00004-6.

Wainright, PO, G Hinkle, ML Sogin, and SK Stickel. 1993. “Monophyletic Origins of Metazoa: An Evolutionary Link with Fungi.” *Science* 260 (5106): 340–42.

Weber, Felix, Javier del Campo, Claudia Wylezich, Ramon Massana, and Klaus Jürgens. 2012. “Unveiling Trophic Functions of Uncultured Protist Taxa by Incubation Experiments in the Brackish Baltic Sea.” *PLoS ONE* 7 (7). doi:10.1371/journal.pone.0041970.

Whelan, Nathan V, Kevin M Kocot, Leonid L Moroz, and Kenneth M Halanych. 2015. “Error, Signal, and the Placement of Ctenophora Sister to All Other Animals” 112 (18): 5773–78. doi:10.1073/pnas.1503453112.

Winberg, Margaret L., Jasprina N. Noordermeer, Luca Tamagnone, Paolo M. Comoglio, Melanie K. Spriggs, Marc Tessier-Lavigne, and Corey S. Goodman. 1998. “Plexin A Is a Neuronal Semaphorin Receptor That Controls Axon Guidance.” *Cell* 95 (7): 903–16. doi:10.1016/S0092-8674(00)81715-8.

Woese, C. R., and G. E. Fox. 1977. “Phylogenetic Structure of the Prokaryotic Domain: The Primary Kingdoms.” *Proceedings of the National Academy of Sciences* 74 (11): 5088–90. doi:10.1073/pnas.74.11.5088.

Woese, Carl R. 1996. “Phylogenetic Trees: Whither Microbiology?” *Current Biology* 6 (9): 1060–63. doi:10.1016/S0960-9822(02)70664-7.

Woese, Carl R, Otto Kandlert, and Mark L Wheelis. 1990. “Towards a Natural System of Organisms : Proposal for the Domains” 87 (June): 4576–79.

Woodland, Hugh R. 2016. *The Birth of Animal Development: Multicellularity and the Germline*. *Current Topics in Developmental Biology*. 1sted. Vol. 117. Elsevier Inc. doi:10.1016/bs.ctdb.2015.10.020.

Worley, Ann C, Kenneth B Raper, and Marianne Hohl. 1979. “Mycological Society of America Fonticula Alba : A New Cellular Slime Mold (Acrasiomycetes) Published by : Mycological Society of America.” *Mycologia* 71 (4): 746–60.

Woyke, Tanja, Gary Xie, Alex Copeland, José M. González, Cliff Han, Hajnalka Kiss, Jimmy H. Saw, et al. 2009. “Assembling the Marine Metagenome, One Cell at a Time.” *PLoS ONE* 4 (April). doi:10.1371/journal.pone.0005299.

Woznica, Arielle, Joseph P. Gerdt, Ryan E. Hulett, Jon Clardy, and Nicole King. 2017. “Mating in the Closest Living Relatives of Animals Is Induced by a Bacterial Chondroitinase.” *Cell* 170 (6). Elsevier Inc.: 1175–1183.e11. doi:10.1016/j.cell.2017.08.005.

Xiao, S., A. Knoll, J. Schiffbauer, C. Zhou, and Y. Xunlai. 2012. “Comment on ‘Fossilized Nuclei and GerminationStructures Identify Ediacaran “Animal Embryos” as Encysting Protists.’” *Science* 335 (March).

Yoon, Hwan Su, Dana C Price, Ramunas Stepanauskas, Veeran D Rajah, Michael E Sieracki, William H Wilson, Eun Chan Yang, Siobain Duffy, and Debashish Bhattacharya. 2011. “Single-Cell Genomics Reveals Organismal Interactions in Uncultivated Marine Protists.” *Science (New York, N.Y.)* 332 (6030): 714–17. doi:10.1126/science.1203163.

Yoshida, Masaki, Takeshi Nakayama, and Isao Inouye. 2009. “Nuclearia Thermophila Sp. Nov. (Nucleariidae), a New Nucleariid Species Isolated from Yunoko Lake in Nikko (Japan).” *European Journal of Protistology* 45 (2): 147–55. doi:10.1016/j.ejop.2008.09.004.

Zettler LAA, T a Nerad, C J O’Kelly, and M L Sogin. 2001. “The Nucleariid Amoebae: More Protists at the Animal-Fungal Boundary.” *The Journal of Eukaryotic Microbiology* 48 (3): 293–97. http://www.ncbi.nlm.nih.gov/pubmed/11411837.

Zhu, Fei, Ramon Massana, Fabrice Not, Dominique Marie, and Daniel Vaulot. 2005. “Mapping of Picoeucaryotes in Marine Ecosystems with Quantitative PCR of the 18S rRNA Gene.” *FEMS Microbiology Ecology* 52 (1): 79–92. doi:10.1016/j.femsec.2004.10.006.

Zong, Chenghang, Sijia Lu, Alec R Chapman, and X Sunney Xie. 2012. “Genome-Wide Detection of Single-Nucleotide and Copy-Number Variations of a Single Human Cell.” *Science (New York, N.Y.)* 338 (6114): 1622–26. doi:10.1126/science.1229164.