Diversity and Evolution of Nassellaria and Spumellaria (Radiolaria)

Miguel Méndez Sandín

Sorbonne Université

École doctorale 227 : Sciences de la Nature et de l'Homme : Écologie et Évolution

Station Biologique de Roscoff / Adaptation et diversité en milieu marin, UMR 7144

DIVERSITÉ ET ÉVOLUTION DES NASSELLAIRES ET SPUMELLAIRES (RADIOLAIRES)

Par Miguel Méndez Sandín

Thèse de doctorat d'Évolution et Diversité Moléculaire

Dirigée par Fabrice Not

Présentée et soutenue publiquement le 22 de Novembre de 2019

Devant un jury composé de :

Linda Amaral-Zettler Senior Scientist, Netherlands Institute for Sea Research	Rapportrice
Fabien Burki Assistant professor, Uppsala University	Rapporteur
John Dolan Research Director, Observatorie Océanologique de Villefranche-sur-Mer, Sorbonne Université, CNRS	Examinateur
Spela Gorican Senior Researcher, Slovenian Academy of Sciences and Arts	Examinatrice
Nathalie Simon Associate professor, Station Biologique de Roscoff, Sorbonne Université, CNRS	Examinatrice
Cédric Berney Associate Researcher, Station Biologique de Roscoff, Sorbonne Université, CNRS	Invité
Fabrice Not Research Director, Station Biologique de Roscoff, Sorbonne Université, CNRS	Directeur de Thèse

I would like to thank the ANR grant IMPEKAB (15-CE02-0011) and the Brittany Region ARED (C16 1520A01), to have funded my PhD thesis and the experiments on it.

To my grandfather:

Suso, who gave me everything having nothing

A mi abuelo:

Suso, quien me dio todo sin tener nada

À mon grand-père :

Suso, qui m'a tout donné sans rien avoir.

Summary

Nassellaria and Spumellaria (Polycystines, Radiolaria) are planktonic amoeboid protists belonging to the Rhizaria lineage. They are widely distributed and abundant in the global ocean. Their silicified skeleton preserves very well in sediments, displaying an excellent and continuous fossil record dating back to the early Cambrian. Radiolarian fossil record is extremely valuable for paleo-environmental reconstruction studies. Radiolaria are difficult to maintain in culture preventing an accurate perception of their extant diversity and ecology in contemporary oceans, and most of it is inferred from the fossil record and sediment samples. Despite recent effort, their taxonomy and evolutionary history remains poorly known and controversial.

Here I explored the diversity and evolutionary patterns of Nassellaria and Spumellaria, based on a single cell integrative classification obtained from taxonomic marker genes (18S and 28S ribosomal DNA) and morphological characteristics. Our phylogenetic analyses established a morpho-molecular framework partly agree with the latest classifications relying essentially on the overall symmetry of the skeleton at Superfamily and Family level. This comprehensive morpho-molecular framework was integrated with recent phylogenetic studies of Acantharia and Collodaria in order to reconstruct the most extensive rDNA phylogenetic analysis of Radiolaria to date. The integrative classification of Radiolaria established the Acantharia, with Strontium sulphate skeleton, a sister clade of the Taxopodida and the Polycystines, both with silicate skeletons. Radiolarian evolutionary patterns were explored using a fossil calibrated molecular clock dating an origin of Radiolaria in the Early Neoproterozoic. Two major events characterized their diversification, the development of the skeleton in the Early Paleozoic and the establishment of the symbiosis in the Middle to Upper Jurassic, when oligotrophy and anoxia governed the oceans. A large environmental diversity was found associated with basal nodes, that following the morpho-molecular framework and evolutionary patterns led to hypothesize a large skeleton-less diversity related to Taxopodida (Rad-B) at basal positions in the radiolarian phylogenetic tree.

The intracellular genetic variability of Nassellaria and Spumellaria was explored finding an important taxonomic bias in both the variability and number of sequences when dealing with short read High-Throughput Sequencing (HTS) output. Sequencing platforms, such as Oxford Nanopore Technologies, provided interesting results for the full rDNA sequences, despite their high error rate. Our analyses allowed a better understanding of the global biodiversity and biogeography of Radiolaria, that was later explored through a metabarcoding approach across samples collected globally during *Tara* Oceans and Malaspina expeditions and regionally compared with the MOOSE-GE cruises in the western Mediterranean-sea. Radiolaria contributed about 9% of the total eukaryotic reads in the studied datasets. The colonial Collodaria was the group more abundant in large size fractions, adapted to oligotrophic and surface waters. Acantharia took the leading role in smaller size fractions and more productive waters. Spumellaria dominated the mesopelagic and bathypelagic followed by a big importance of environmental diversity believed to be associated to the skeleton-less Radiolaria.

This work brings a new comprehensive perspective of the evolutionary relationships and diversity of extant Radiolaria, highlighting their important planktonic role in both contemporary and past oceans.

Résumé

Les Nassellaires et les Spumellaires (Polycystines, Radiolaires) sont des protistes planctoniques amiboïdes appartenant au super-groupe des Rhizaires. Ils sont abondants et largement distribués dans l'océan. Leur squelette silicifié se conserve très bien dans les sédiments et présente un excellent enregistrement fossile remontant jusqu'au début du Cambrien. Les fossils de Radiolaires sont extrêmement précieux pour les études de reconstruction paléo-environnementales. Les Radiolaires sont difficiles à maintenir en culture, ce qui limite nos connaissances sur leur diversité et leur écologie dans les océans actuels, qui reposent essentiellement sur les données fossiles et les échantillons de sédiments. Malgré des efforts récents, leur taxinomie et leur histoire évolutive demeurent peu connues et controversées.

Pendant mon doctorat, j'ai exploré la diversité et l'évolution des Nassellaires et des Spumellaires sur la base d'une classification intégrative obtenue à partir de gènes marqueurs taxinomiques (ADN ribosomal 18S et 28S) et de caractéristiques morphologiques issues de cellules uniques isolées depuis l'environnement. Nos analyses phylogénétiques ont permit de mettre en evidence un cadre morpho-moléculaire correspondant partiellement aux classifications les plus récentes reposant largement sur la symétrie globale du squelette pour les Superfamilles et Familles. Nos nouvelles données morpho-moléculaires ont été intégrée aux études précédentes sur les Acanthaires et des Collodaires afin de reconstruire la phylogénie des Radiolaires la plus complete à ce jour. Cette nouvelle classification intégrée des Radiolaires établit les Acanthaires, avec leur squelette en sulfate de strontium, comme le groupe frère des Taxopodida et des Polycystines, tous deux dotés d'un squelette en silicate. Associé à ces phylogenies, l'utilisation d'horloges moléculaires nous a permis de dater l'origine des Radiolaires au Néoprotérozoïque. Mais également de metre en evidence deux événements majeurs caractérisent leur diversification : le développement du squelette au Paléozoïque ancien et l'établissement de la symbiose dans le Jurassique moyen à supérieur, lorsque les eaux oligotrophes et anoxiques dominaient les océans. La grande diversité environnementale trouvée aux nœuds basaux de nos phylogenies nous conduit à émettre l'hypothèse de l'existence d'une grande diversité de Radiolaires sans squelette associée aux Taxopodida (Rad-B).

La variabilité génétique intracellulaire des Nassellaires et des Spumellaires a été étudiée et a révélé un biais taxinomique important dans la variabilité et le nombre de séquences obtenues par différentes méthodes de séquençage à haut débit (HTS). Cependant, le séquençage par *Oxford Nanopore Technologies*, a fourni des résultats intéressants pour le séquençage complet des ADNr, malgré son taux d'erreur élevé. L'ensemble de ces analyses a permis de mieux comprendre la biodiversité et la biogéographie des Radiolaires, qui ont été étudiées selon une approche de *metabarcoding* sur des échantillons collectés a travers le monde au cours des expéditions *Tara Oceans* et Malaspina puis comparés à l'échelle régionale avec les campagnes océanographiques MOOSE-GE en Méditerranée Occidentale. Les Radiolaires contribuent pour environ 9% du total des séquences eucaryotes dans les jeux de données étudiés. Les Collodaires coloniaux représentent le groupe le plus abondant dans les grandes fractions de taille, ils sont adaptés aux eaux oligotrophes et de surface. Les Acanthaires dominent les fractions de plus petites tailles et les eaux plus productives, tandis que les Spumellaires prédominent dans les eaux mésopélagiques et bathypélagiques. Enfin, en profondeur, il existe une importante diversité environnementale probablement représentée par des Radiolaires sans squelette.

Ce travail apporte une perspective globale des relations évolutives et de la diversité des Radiolaires, soulignant leur importance dans les communautés planctoniques des océans contemporains et passés.

Index

Glossary

Amplicon Sequence Variant (ASV): Individual nucleotide sequence inferred after correction of artefacts (process known as denoising) produced during amplification and/or sequencing obtained from NGS. In contrast to OTUs, ASVs are based in the expected error rate. Further reading: Callahan et al. (2017) and references therein.

Apomorphy: A character innovation that is different from that found in a previous ancestor.

Barcoding: In molecular biology, a tool for individual specimen identification, characterization and discovery by using a nucleotide sequence as unique identifier, such as the mitochondrial gene cytochrome *c* oxidase I (COI) for animals (Hebert et al., 2003) or the 18S rDNA gene for protists. I.e.: Chapter 1.

Basal: In phylogenetic analysis, refers to a clade (or node) holding a position close to the root or relative to a specified clade (or node).

Biological species: Groups of actually or potentially interbreeding natural populations, which are reproductively isolated from other such group (Mayr 1942).

Chrono-species: "Species along a single phyletic lineage that are distinguished morphologically from earlier and later forms" (quoted from Zachos, 2016).

Cryptic species: Two or more different "species" hidden under the same binomial name.

Distal: In phylogenetic analysis, refers to a clade (or node) far from the root or the specified clade, in contrast to basal.

Ecological species: A lineage (or a closely related set of lineages) which occupies an adaptive zone minimally different from that of any other lineage in its range and which evolves separately from all lineages outside its range (Van Valen, 1976).

Environmental DNA, or **eDNA**: DNA sampled from the environment, and typically, with no other associated identifier or characterization than the environment itself in which it was collected.

Evolutionary species: (Not to be confused with Evolutionary Significant Unit) ancestor-descendant (or population) lineages that evolve separately from other such lineages and have their own evolutionary tendencies and historical fate (Wiley, 1978).

Genetic species: A group of organisms so constituted and so situated in nature that a hereditary character of any one of these organisms may be (possibly, but not necessarily) transmitted to a descendant of any other (Simpson, 1943).

Holobiont: An evolutionary and ecological unit of different species that live together in symbiosis. Normally composed of a large entity (host) harbouring smaller partners (symbionts and/or microbiota). Further reading: Dittami et al. (2019) and references therein.

Lineage: A monophyletic subset of a phylogenetic tree.

Long Branch Attraction (LBA): In phylogenetic analysis, when the amount of accumulated change within a lineage is similar to another long branch lineage, producing an artefact in the topology and interpretation by relating distantly related long branches.

Metabarcoding: In molecular biology, is the large-scale taxonomic comparison of a marker gene extracted from the environment against homologous barcodes. I.e.: Chapter 3. Further reading: Taberlet and Coissac (2012) and references therein.

Mixotroph: An organism, generally protist, that can use both inorganic and organic carbon from different energetic sources, combining photoautotrophy and phagotrophy. Further reading: Mitra et al. (2016) and references therein.

Molecular clock: A technique that uses the mutation rate of a nucleotide or protein sequence to infer their relative divergence (Zuckerkandl and Pauling, 1962). The molecular clock can be "calibrated" resulting in absolute values. Further reading: Bromham and Penny (2003), Ho and Duchêne (2014) and references therein.

Monophyly, or **clade**: a group of organisms, genes or other evolving elements (taxa) that all of them share the same most recent common ancestor. I.e.: Acantharia, Collodaria in Chapter 2.

Morpho-species: The smallest groups that are consistently and persistently distinct, and distinguishable on the basis of morphological recognizable characters (Cronquist 1978).

Next Generation Sequencing (NGS): Also known as High-Throughput Sequencing (HTS), refers to different techniques able to sequence thousands to millions of nucleotide sequences inexpensively. Further reading: Goodwin et al. (2016), Levy and Myers (2016) and references therein.

Node: In phylogenetic analysis, refers to a point where a lineage splits into two or more lineages.

Operational Taxonomic Unit (OTU): Originally described as "a group of organisms currently being studied" (Sokal and Sneath, 1965). In molecular biology is used to cluster nucleotide sequences based in a similarity threshold, normally 97%.

Paraphyly: A group of organisms, genes or other evolving elements (taxa) that share the most recent common ancestor but not all its descendants. I.e.: Hexastyloidea in chapter 1.2, Nassellaria in Chapter 2.

Phyletic: related to or based on the evolutionary history of a single line of descent without including branching.

Phylogenetics: In molecular biology, is the study of the relationships and evolutionary history among taxa of observed heritable traits (normally nucleotide sequences) under a model of evolution.

Phytoplankton: Planktonic organisms that use light for the acquisition of energy and the reduction of inorganic carbon (photoautotrophy). This classification represents a functional group, occurring in many different lineages of eukaryotes and even bacteria, such as the cyanobacteria.

Plankton: Organisms present in the water column that cannot swim against oceanic currents.

Polyphyly: A group of organisms, genes or other evolving elements (taxa) that do not share the same most recent common ancestor, in contrast with monophyly. I.e.: "flat spumellarians" in chapter 1.2.

Reference sequence: Is a DNA sequence belonging to a specific organism/specimen that has been taxonomically annotated and described.

Ribosomal DNA (rDNA): Is a DNA sequence coding for the ribosomal RNA gene, constituting the main component of ribosomes (proteins in charge of translating RNA into proteins). Briefly, the rDNA structure in eukaryotes is, in most of the studied cases, repeated in several copies in the genome and it is subdivided in the 18S, 5.8S and 28S genes delimited by Internal Transcribed Spacers (ITS1 and ITS2) between them. The 18S gene forms the Small Sub-Unit (SSU) of the ribosomes and the 28S gene the Large Sub-Unit (LSU) along the 5.8S gene. The specific evolutionary rate, the repeated copies and the wide distribution in every eukaryotic organism make the rDNA operon the preferred marker for phylogenetic inference analysis among protists (Woese and Fox, 1977; Hillis and Dixon, 1991; Weisburg et al., 1991).

Ring species: A chain of neighbour populations, that can interbreed among closely related populations, in which at least two ends cannot interbreed despite potential gene flow across the populations that connect them.

Root (or outgroup): In phylogenetic analysis, the most recent common ancestor of the taxa being studied (ingroup).

Species: A concept, and therefore an idea, for delimiting biological unique entities. Zachos (2016) has listed more than 30 different species definitions, although there have been other definitions proposed, and different versions to accommodate asexual species.

Symbiosis: A close and lasting relationship between organisms living together. Including mutualism, commensalism, amensalism and parasitic relationships.

Synapomorphy: An apomorphy only shared by a monophyletic clade.

Zooplankton: Planktonic organisms that cannot produce their own organic carbon, relying on the intake from other sources (heterotrophy, in protists normally phagotrophy or osmotrophy). As for phytoplankton, this is a functional group occurring in many different lineages of eukaryotes.

INTRODUCTION

1. Unicellular plankton diversity in the marine environment

The marine environment represents the largest ecosystem on Earth. More than 70% of the Earth is covered by the vast open ocean, housing a wide variety of life from the largest to some of the smallest living forms on earth. The oceans are believed to be the cradle of life, being witness of major biological diversification events since more than 3,770 million years ago (Ma) (Dodd et al., 2017). Attempting to classify the broad variety of organisms in the ocean, Victor Hensen (1887) coined the functional term plankton to refer to those organisms that cannot swim against the current. The plankton gathers a broad range of life forms, from virus and bacteria to unicellular and small pluricellular eukaryotes, spanning more than six orders of magnitude in size. In contrast to the plankton, Ernst Haeckel (1890) proposed nekton for those that can actively swim.

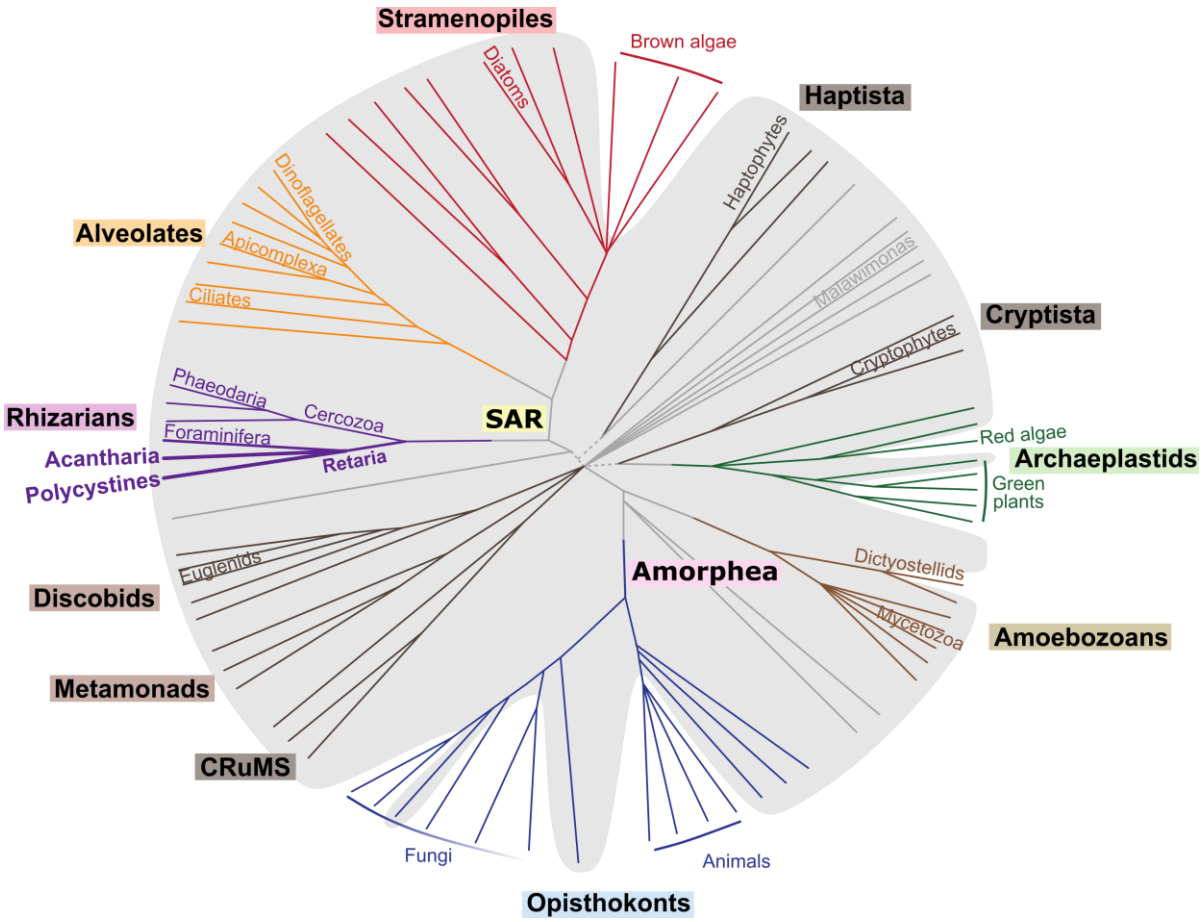


Figure 1. Schematic representation of the eukaryotic tree showing under the grey area unicellular eukaryotic lineages (protists). Considered as unicellular those lineages where most of the diversity is unicellular in the majority of its life cycle and/or there are not cellular organization. Adapted from Keeling and Burki (2019).

For many years Plankton diversity has been focused on animals, plants and fungi relegating to “other eukaryotes” the rest of the diversity. Ironically, multicellular eukaryotes fall within the minority of plankton diversity, most of it being unicellular (Fig. 1). These single-celled eukaryotes are the so-called Protists, and were firstly noted by Leeuwenhoek in 1674 as “many little animalcules” living in the waters of a nearby lake (Rothschild, 1989; O’Malley et al., 2013). From here on, I am going to be referring to protists as single celled eukaryotes, understanding the polyphyly of the group and avoiding any debate or ambiguity in the multi-cellularity organization level. Protists represent an astonishing diversity in the eukaryotic tree of life. Within protists, there are plankton able to photosynthesize, such as the phytoplankton, others are heterotrophic, such as the zooplankton, and others are able to do both constituting the mixotrophs (Flynn et al., 2013; Stoecker et al., 2016). Protists share cellular, genomic and evolutionary characteristics with their multicellular relatives but also ecological and evolutionary patterns with prokaryotes (Keeling and Burki, 2019). Their immense distribution across many different lineages and environments make them excellent models for answering fundamental evolutionary and biological processes.

Despite the large variety of protists, our vision of their molecular diversity is certainly biased, corresponding to less than 4% of the described eukaryotic diversity (Del Campo et al., 2014). Furthermore, Archaeplastids, Stramenopiles and Alveolates share more studied diversity than the other unicellular groups together; mainly due to the facility to maintain them in culture and to their impact on human economy and health (harmful algal blooms, parasites, etc.; Del Campo et al., 2014). In the last decades, molecular advances have allowed the exploration of unculturable organisms providing a better understanding of the uncharted eukaryotic diversity, its relationships and its role in the ecosystems. By comparing a DNA sequence extracted from the environment (environmental DNA) to that of an identified organism several studies have accessed the unexplored diversity and inferred its geographical and evolutionary distribution (e.g. López-García et al., 2001; Lovejoy et al., 2007; Not et al., 2007; Vaulot et al., 2008). With the recent advent of high throughput Next Generation Sequencing technologies it has been possible to access to the global environmental diversity through metabarcoding approaches, discovering that less than 1% of the global planktonic protists diversity in the oceans corresponded to described organisms (de Vargas et al., 2015; Pernice et al., 2016). The exploration of this unculturable diversity has shown the importance of previously overlooked groups, such as the Rhizaria. In the meantime, other studies using in situ imaging technologies have confirmed the contribution of Rhizaria to the ecosystems, representing up to 5.2% of the total standing carbon stock in the oceans (Biard et al., 2016; Guidi et al., 2016). However, most of the rhizarian environmental diversity remain undescribed (Grattepanche et al., 2018).

Rhizaria is a major lineage of eukaryotes comprising an immense diversity of clades. They were firstly described including Cercozoa and Retaria (Cavalier-Smith, 2002) and recent molecular phylogenetic analyses have also included Endomyxa (Cavalier-Smith et al., 2018; Adl et al., 2019). Their basal position in the SAR supergroup (Stramenopiles, Alveolates and Rhizaria, Fig 1) makes them a key element to understand the early evolution and diversification of Eukaryotes. Rhizarians are mainly free-living heterotrophic amoeboid

organisms, naked, with testate or skeleton and long prolongations of the cytoplasm, called pseudopods, that are used more for feeding than for locomotion. In general, they are difficult to maintain in culture. These descriptions contrast the biases mentioned above towards photosynthetic, culturable and/or parasitic preferences for molecular studies. This is basically why our understanding of Rhizaria stands far behind other eukaryotic groups (Burki and Keeling, 2014).

Within the Rhizaria, Radiolaria and Foraminifera (both grouped in the Retaria) are active predators, hunting preys with their pseudopodia as big as copepods or mollusc larvae (Anderson, 1983; Swanberg et al., 1986). In surface waters of the ocean, some Retaria are mixotrophs as they host photosynthetic symbionts, such as the symbiotic dinoflagellates Brandtodinium (Probert et al., 2014) and Gymnoxanthella (Yuasa et al., 2016) that have been rarely documented in free-living state. This mixotrophic behaviour allows them to reach important abundances in equatorial and tropical waters (Leles et al., 2017), contributing to the food web not only as predators but also to the primary productivity and carbon fixation (Michaels, 1988; Stoecker et al., 2009). Besides, their skeleton made of calcium carbonate in Foraminifera and opaline silica in Polycystines (Radiolaria) preserves very well in sediments, showing a continuous fossil record that dates back to the early Cambrian (~520 Ma), or earlier (Suzuki and Oba, 2015; Fig. 2). Their extensive fossil record represents an extremely valuable tool for both paleo-environmental reconstruction (e.g. Abelmann and Nimmergut, 2005; Kamikuri et al., 2009; Lariviere et al., 2012) and bio-stratigraphic studies (e.g.: O'Dogherty et al., 2011; Aitchison et al., 2017).

Despite the ecological and evolutionary importance of Retaria little attention has been paid to untangle phylogenetic relationships within and between them. The Retaria was firstly introduced by Cavalier-Smith (1999) mentioning a long branch attraction artefact that may alter phylogenetic topologies. Ever since, phylogenetic studies on Retaria have given contrasting and complementary results, challenging the monophyly of Radiolaria where Foraminifera sometimes hold a sister position to Radiolaria (Nikolaev et al., 2004; Ishitani et al., 2011; Cavalier-Smith et al., 2018) whereas in other analyses Foraminifera appears among Radiolaria (Burki et al., 2013; Sierra et al., 2013; Krabberød et al., 2017). First molecular barcoding and phylogenetic analysis within Retaria were carried out in Foraminifera (e.g.: Pawlowski and Holzmann, 2002; Pawlowski et al., 2003a, 2013; Shaked and De Vargas, 2006), probably due to a preference for the larger cell size regarding Radiolaria. Such studies have deeply contributed to the understanding of the diversity and evolution of Foraminifera, by establishing a comprehensive evolutionary history linked to their morpho-molecular diversity. Yet, radiolarian molecular diversity exploration has always lagged that of Foraminifera. And, our current knowledge on radiolarian diversity and evolution relies mainly on morphology-based approaches performed in the fossil record, sediments and plankton samples.

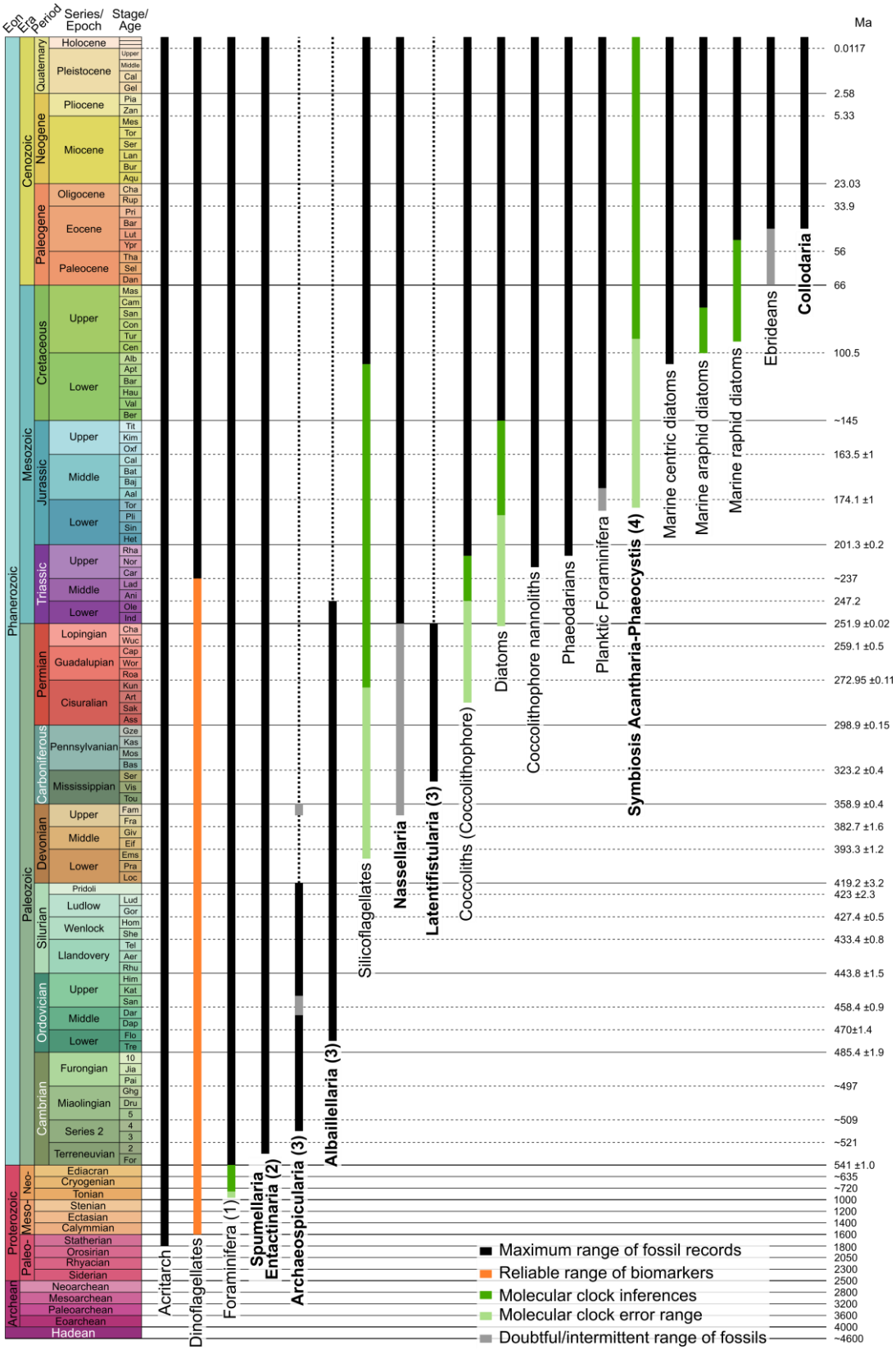


Figure 2 Schematic representation of main fossilizable marine protists range in the fossil record and major Radiolaria events. Adapted from Suzuki and Oba (2015). 1-corrected after Pawlowski et al. (2003a) and Groussin et al. (2011), 2- corrected after Zhang and Feng (2019), 3-Added after Aitchison et al. (2017), 4- Added after Decelle et al. (2012a).

2. Nassellaria and Spumellaria in the radiolarian context

First classifications of Radiolaria were carried out by Ernst Haeckel documenting their diversity in his extensively illustrated monographs (1862, 1887; Fig. 3). However, first taxonomic works correspond to Franz J. F. Meyen (1834) on colonial Collodaria, Christian G. Ehrenberg (1838) describing the Polycystines, and Johanes P. Müller (1858) proposing the name Radiolaria. Haeckel's Radiolaria classification was established based on morphological characters such as the double-layered cytoplasm and the pseudopodia. At that time, Radiolaria was encompassing the Acantharia, the Phaeodaria and the Polycystines (Spumellaria and Nassellaria, including the Collodaria as colonial Spumellaria). After Haeckel's classifications, many taxonomists continued the work, being mostly studied by micropaleontologists due to their worldwide distribution and abundance in sediments. Only few studies performed on live specimens linked the skeleton's morphology and cytological structures (Hollande and Enjumet, 1960). The latest classifications and evolutionary histories proposed were established based on the structure of the first elements developed in the skeletal growth, the so called initial spicular system (Petrushevskaya, 1971a; De Wever et al., 2001; Afanasieva et al., 2005).

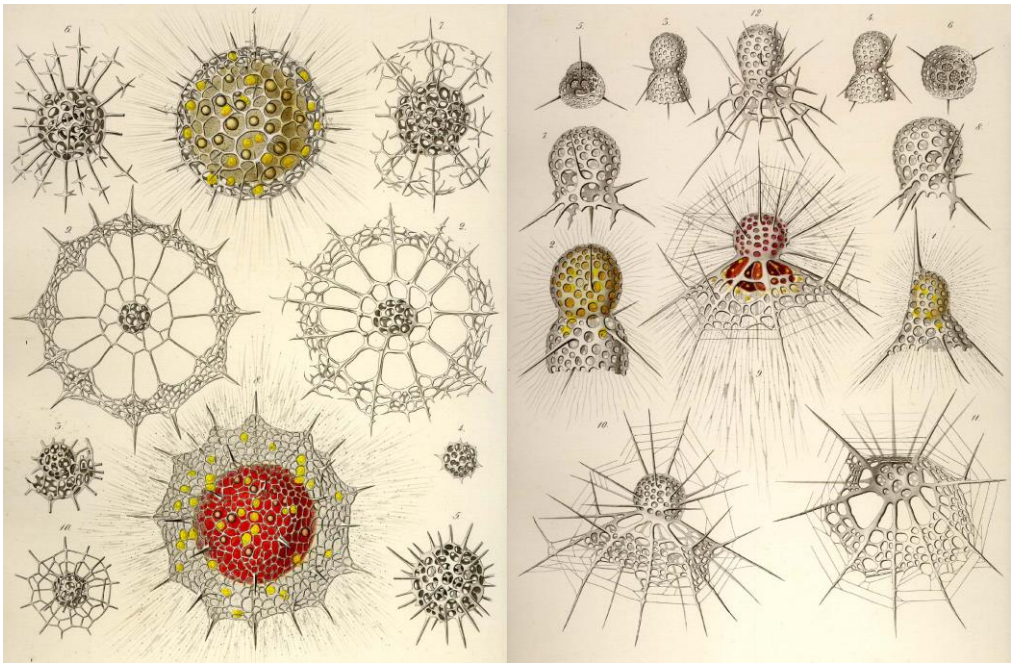


Figure 3. Drawings of Spumellaria (left) and Nassellaria (right) from Haeckel's original plates (1862).

2.1. The fossil record

From the first radiolarian representatives of the early Palaeozoic to contemporary specimens there is a well-documented range of appearance and extinction in the fossil record of many different forms (De Wever et al., 2001, 2003; Fig. 2). Their abundance in sediments can reach considerable importance, which led to name rocks (the radiolarite) after this group. First representatives of Polycystines in the Cambrian (541-485.4 Ma) were an aggregate of isolated spicules difficult to distinguish from those of siliceous sponges (Maletz, 2011). These primitive forms are classified in the extinct order Archaeospicularia, where a set of spicules

arrange around an internal cavity. This order is considered to be the oldest and the ancestor of all Polycystines (Dumitrica et al., 2000) although, their diversity was rather low and fragmented. Albaillellaria is the only extinct order gathering bilateral radiolarians. Families within this group are used to defined the Carboniferous (358.9-298.9 Ma) and the Permian (298.9-251.9 Ma) biozones (Aitchison et al., 2017). The latest extinct order corresponds to the Latentifistularia, created due to their different initial spicular system (Caridroit et al., 1999).

Recent studies have found the oldest forms of Radiolaria, taxonomically assigned to Spumellaria with a consolidated skeletal structure (Zhang and Feng, 2019) contrasting that of Archaeospicularia. However, the Spumellaria is very similar and sometimes confused with the Entactinaria. These two orders were classified on the basis of the presence (Entactinaria) or absence (Spumellaria) of the initial spicular system (De Wever et al., 2001). Due to this definition in their classification and their similar external morphology several families moved from one order to the other depending on the preservation state of similar fossil specimens. Despite the effort to unveil differences between the two orders, their classification remained always unsteady. Yet, their evolutionary history is marked by a higher presence of Entactinaria during the Palaeozoic (541-251.9 Ma), changing towards a dominance of Spumellaria after the Permian-Triassic boundary (251.9 Ma), where most of the diversity of Entactinaria got extinct (De Wever et al., 2003). Advances in imaging tools have shown homologous elements of the initial spicular system in Spumellaria, concluding their tight relationships and doubting the monophyly of the Entactinaria (Kachovich et al., 2019).

The evolution of the order Nassellaria is also subject of discussions, since some Radiolaria families in the Devonian (419.2-358.9 Ma) have a nassellarian-like initial spicular system, but the external morphology resemblance that of Entactinaria (De Wever et al., 2001). Yet, it is difficult to link the first nassellarian-like fossils from the Devonian with the first nassellarian representatives from the Triassic (251.9-201.4 Ma) since there are no evidences of intermediate morphologies along the Permian (298.9-251.9 Ma; Isakova and Nazarov, 1986; De Wever et al., 2003). Afanasieva et al., (2005) proposed a different scenario for the origin of Nassellaria, with an independent evolution from that of Spumellaria and/or Entactinaria orders since the Cambrian. Although, the morphological evolutionary relationships of Nassellaria remain obscure.

The extraordinary and detailed work carried out from sediments and fossil records have established a comprehensive view of the diversity and evolution of the Polycystines from a morphological point of view (Petrushevskaya, 1971b; De Wever et al., 2001; Afanasieva et al., 2005; O'Dogherty et al., 2011; Aitchison, 2017). However, gaps in stratigraphic ranges or lack of intermediate morphologies to link fossil representatives raised important questions that cannot be answered by a morphological approach. Furthermore, differences in the preservation state or similarities between different specimens may change the interpretation of the results, being strongly subjective to the expertise and background of the scientist performing the study. Besides these limitations, other questions such as evolutionary patterns in Acantharia or the origin and relationships of Collodaria are challenging, if not impossible,

to answer relying on morphological observations only. In the first example due to the absence of the direct observation in fossils because of the rapid dissolution of the strontium sulphate skeleton upon cell death and in the second due to the degradation of the organic matter that constitutes the colonies and their very contrasting skeletal morphology regarding other Radiolaria.

2.2. The advent of molecular phylogenetic analysis

Recent studies implementing molecular tools have the potential to answer radiolarian classification and evolution questions beyond the fossil record and the morphology. Based on rDNA and multigene phylogenies, Phaeodaria, that were always considered within Radiolaria, have been moved to the Cercozoa, and Taxopodida, previously related to Heliozoa, are now belonging to Radiolaria (Nikolaev et al., 2004; Burki et al., 2013). Besides, multi-gene and multi-protein phylogenies have demonstrated the monophyly of the Polycystines group with Acantharia as a sister clade (Krabberød et al., 2017; Cavalier-Smith et al., 2018). These recent advances in the phylogenetic relationships of Radiolaria are contributing to a better understanding of their extant diversity and evolution at high taxonomic levels. Decelle et al. (2012b) have explored the genetic diversity of Acantharia finding morphological synapomorphies and describing a comprehensive classification by merging molecular and morphological data. This study allowed inferring Acantharia evolutionary history and to reconstruct their morphological evolution, finding a consolidation of the strontium spicules in a robust central junction over time. Following a similar approach, Biard et al. (2015) have shown that the previously considered solitary Collodaria family (Thalassicollidae) is very likely a life cycle stage of the colonial Collodaria.

In contrast to the recent molecular advances performed in Acantharia and Collodaria, the Polycystines Nassellaria and Spumellaria have received little attention. Few studies have dug into their genetic diversity characterization (Kunitomo et al., 2006; Yuasa et al., 2009; Krabberød et al., 2011; Ishitani et al., 2012a). With no more than sixteen rDNA genes sequences corresponding to morphologically identified Nassellaria and 35 for Spumellaria their phylogenetic relationships remain still in its infancy. This poor number of sequences contrasts with the amount of morpho-species described, gathering Nassellaria (~430 extant species described) or Spumellaria (~380 sp.) more morpho-species described than Collodaria (~95 sp.) and Acantharia (~145 sp.) together. Matsuzaki et al. (2015) have attempted to merge the extensive morphological criteria with recent rDNA molecular studies. They recognized the “overwhelmingly artificial” systematics and reflected the outweighed morphological criteria in the classification of Nassellaria and Spumellaria. Therefore, the poor taxonomic coverage for which molecular data is available prevents a proper agreement between the molecular taxonomic markers and the morphology, challenging the understanding of fundamental questions among radiolarian classification, diversity and evolution. For instance, is the initial spicular system the best morphological character for their classification? When is Nassellaria appearing as a group? Are Spumellaria and Entactinaria two different orders or are they closely related?

2.3. Nassellaria, Spumellaria and the molecular clock

In a further attempt to complement the fossil record and morphological approaches, molecular tools can take advantage of the exhaustive work performed on the fossil record to calibrate a molecular clock. Time calibrated phylogenies allow accessing diversification times and therefore help contextualizing a molecular-based phylogeny with geological and environmental global changes. Such contextualization leads to a better understanding of the biotic and abiotic factors driving evolutionary and diversity patterns. Furthermore, molecular dating allows tracing the origin and evolutionary history beyond fossil limits, such as on the origin of eukaryotes (Berney and Pawlowski, 2006; Eme et al., 2014). In this context, Decelle et al. (2012a) estimated the date of the establishment of the photosymbiosis between Acantharia and the haptophyte Phaeocystis in the Middle Jurassic (~175 Ma), when the oceans reached their maximum oligotrophic state. Following the same approach, Ishitani et al. (2012b) estimated the first diversification of Collodaria in the Eocene (56-33.9 Ma), suggesting that the first collodarian representatives lacked silicified skeleton.

The above-mentioned studies were performed with external calibrations of the molecular clock, based on their closest relatives, Nassellaria and Spumellaria. The calibration of the molecular clock is a critical step that requires a comprehensive link between the morphology and the molecular clade, to ensure accurate estimates of divergence times between clades (Ho and Phillips, 2009). The excellent and continuous fossil record of Polycystines makes them a privileged group for molecular dating analyses, allowing the connection from the Cambrian to the present (Suzuki and Oba, 2015). The importance of Polycystines in the fossil record and their described morphological diversity sets Nassellaria and Spumellaria as key taxa in order to answer fundamental evolutionary questions within the Radiolaria. Yet, the low number of reference sequences of both Nassellaria and Spumellaria avoids an appropriate link between molecular barcodes and morphological groups challenging a correct calibration of the molecular clock.

The fossil record is the most widely used source for calibration of molecular clocks (Sauquet, 2013). Meaning that a molecular clade is going to be dated based on the first appearance in the fossil record of the most similar morphology to the one having a molecular signature available. In most of the cases, such approach actually estimates only the minimum date of appearance for a given morphology. Signal in fossil record, only represent species found in relatively large population and that have lived in a favourable environment over a considerable amount of time. This contributes to explain why molecular clock tends to provide older estimates than what it is reflected by observation in the fossil record (e.g.: Pawlowski et al., 2003a; Groussin et al., 2011; Foster et al., 2017; Betts et al., 2018; Lahr et al., 2019). These discrepancies have led to fruitful discussions about the limitations of both approaches (e.g.: Donoghue and Benton, 2007; Pulquério and Nichols, 2007). Yet, for many calibrations, the morphology that it is based on is extinct. Besides, we are not aware of how the early representatives of a clade look like, or even if they had structures that preserve in the fossil record. Therefore, that the molecular clock estimates earlier dates than the fossil record might

not indicate a discrepancy but rather a different morphology in the fossil record that have been ignored, if it ever had representatives in the fossil record. On the other side, when the molecular clock dates much younger ages can be (i) a failure in the clock model measuring the rate of variation, for instance due to large differences in the mutation rate between taxa (ii) or simply a misinterpretation of the clades, in which the fossil clade used for the calibration is not nested in the phylogenetic clade it is believed to belong.

2.4. Distribution and biogeography

Despite the importance of Polycystine skeleton in the fossil record, little is known about their contemporary distribution and biogeography in the oceans, and most of it is inferred from sediments and plankton samples. The use of Polycystines skeleton in paleo-environmental reconstruction analysis rely in the link between defined morpho-types and preferred environmental characteristics (e.g.; Abelmann and Nimmergut, 2005). This link is based on observations from modern analogs (e.g.; Shuman, 2013). And in order to perform an accurate paleoenvironmental reconstruction it is required a strong relationship of the contemporary biogeography to well physic-chemically characterized water masses.

First studies exploring the abundance and ecology of Radiolaria grouped together the Acantharia, planktonic Foraminifera and Polycystine under the name Sarcodines, due to their similar ecological niches. For this group, Michaels et al. (1995) estimated an average contribution of 15.5% of the total carbon flux in the upper 150 m of the Sargasso Sea. Sarcodines represent on average the 23% of the total macrozooplankton in the global mesopelagic ocean (Stemmann et al., 2008). And among Sarcodines, Radiolaria contribute up to the 90% of the total vertical export in specific areas such as the California Current Ecosystem (Gutierrez-Rodriguez et al., 2019). Recently, Boltovskoy and Correa (2016) have delimited six major biogeographic provinces based on different Polycystine assemblages (Fig. 4). These biogeographic provinces correlate with the temperature. The intertropical province shows the biggest Polycystine species richness followed by the subpolar region, and the northern polar areas shows different assemblages than in the southern areas. Finally, there is a small province of transition between the tropical and subpolar with a high variance in species composition and one region restricted to the eastern equatorial Pacific. Plankton samples showed that in general tropical surface waters gather the highest species richness, and then decreases towards depth and poles (Boltovskoy, 2017). In contrast, sediment traps deployed at various depth along the water column have recorded a higher number of radiolarians in deeper samples, and there seems to have an overrepresentation of Nassellaria in this type of samples compare to surface sediment traps or planktonic samples where presence of Nassellaria is more balanced with that of Spumellaria.

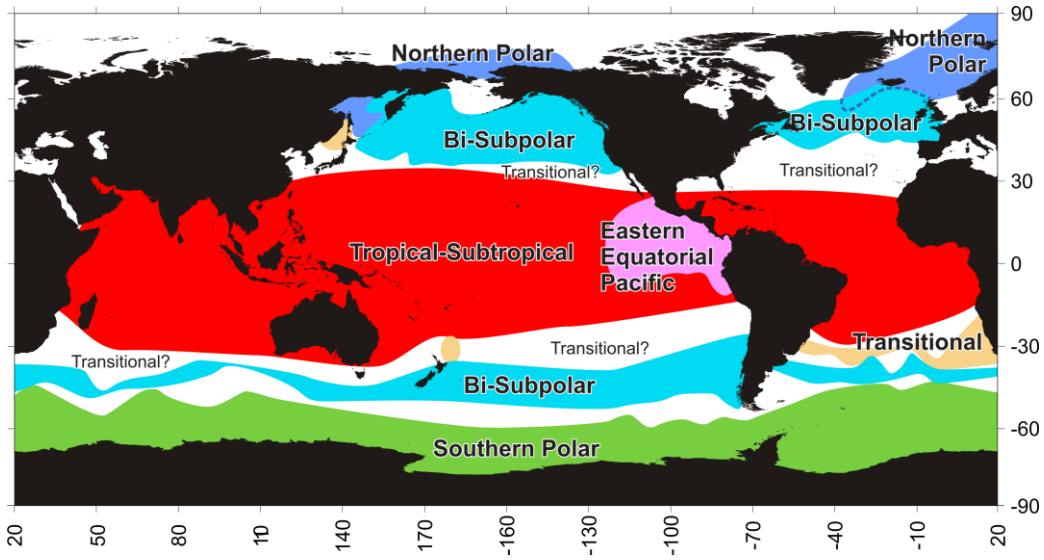


Figure 4. Major biogeographic provinces of Radiolaria assessed by differences in species assemblages obtained by plankton samples, sediment traps and surface sediment samples, from Boltovskoy and Correa (2016).

Morphology-based studies are time consuming and the level of expertise required prevents accurate identification for a large number of samples. In contrast, metabarcoding surveys are exploring the biodiversity of the water column by a rapid approach and at a rather fine taxonomic resolution. Global molecular environmental surveys have detected an important abundance of Radiolaria along the water column, especially towards deep environments (de Vargas et al., 2015; Pernice et al., 2016). Other studies have also showed their importance in the pico- and nano-size fraction (Not et al., 2007, 2009; Edgcomb et al., 2011), where paradoxically, it is not expected to find Radiolaria. Yet, metabarcoding surveys rely in reference morpho-molecular framework to accurately link genetic sequences (metabarcodes) to their morpholocially described taxonomic group. Following such approach, Biard et al. (2017) have explored environmental preferences in Collodaria, finding higher diversified communities towards more oceanic and oligotrophic biomes. Still many questions regarding Radiolaria biogeography and ecology remain. For instance, what about the diversity found in the small size fraction? Is it corresponding to already known morphological diversity? Spumellaria and Nassellaria show an outstanding morphological diversity compared to that of Acantharia and Collodaria, yet, in contrast, molecular surveys are showing Collodaria and Acantharia more important (de Vargas et al., 2015). Therefore, what is the meaning of such environmental molecular pattern? Can it be related to morphological observations?

3. Challenging the species concept: morphology, DNA and the strength of teamwork

Molecular tools have provided access to previously undetected diversity, allowing a better understanding of evolutionary and taxonomic relationships among marine plankton. Yet, in order to unravel evolutionary and ecological patterns among eukaryotic diversity, it is required a relationship between the DNA molecules and what they actually represent. In Eukaryotes, the most extensively used measure for describing diversity is the morphology. Yet

due to the incredible long evolutionary history and high rank lineage diversity within Protists, applying a morphologic "currency exchange" leads to more issues than solutions. By trying to match a molecular clade with a morphology, two different species concepts are combined and, in most of the cases, these two concepts fail to agree. This is part of the so-called "species problem", a long debate in biology when species delimitation comes to play.

Far from being a modern debate, the species problem has wander around even before evolutionary notions were driving biology. According to Wilkins (2009), John Ray was the first naturalist describing a species in a biological context in his *Historia plantarum* in 1686. Ray realised about the great variation among individuals from the same "species" and the ambiguous boundaries between them. Later in the 18th century, the father of modern taxonomy, Carl Linnaeus, mentioned that some species were the result of hybridization (1767-1770). Despite his believes in creationism he stated that species were not fixed and not clearly delimited, growing differently according to the environment. When Darwin proposed the famous theory of Evolution in his book *On the Origin of Species*, the concept of the species started becoming a problem since at that time it was not possible to find a definition for a species:

> "[..] *No one definition has satisfied all naturalists; yet every naturalist knows vaguely what he means when he speaks of a species* [..]."
>
> "[..] *I look at the term species, as one arbitrarily given for the sake of convenience to a set of individuals closely resembling each other, and that it does not essentially differ from the term variety, which is given to less distinct and more fluctuating forms. The term variety, again, in comparison with mere individual differences, is also applied arbitrarily, and for mere convenience sake.*"
>
> Darwin (1859) Chapter II, variation under nature.

Most of these inconsistencies are found when applying the morphological species concept only, since it is subjective to the "naturalist" decision, as Darwin mentioned. Nowadays, we know the existence of cryptic species (e.g.: Bickford et al., 2007) for which morphological delimitation is not clearly defined. When applying the biological species concept (i.e. interbreeding meta-populations) the definition of a species is neither clarified. Examples such as ring species (e.g.: Irwin et al., 2001; Ivalu and Baum, 2012) keep challenging the species concept of fix and clearly delimited boundaries between different biological entities. In a continuum meta-population where there are, at least, two populations that cannot directly interbreed (ring species), it is not possible to distinguish between one or two species. Despite these two populations cannot interbreed, they are connected by the same meta-population. The concept of ring species have been argued due to the faster Earth's climatic shifts compared to the time required for reproductive isolation, yet it has also been proposed as a gradual speciation event (Alcaide et al., 2014). The morphological and biological species concept may go far beyond the species level, since in some cases it is not possible to distinguish a single individual (i.e.: incrusting calcareous algae, colonial fungi). Furthermore, what about asexual organisms? Or those that we do not know how they reproduce?

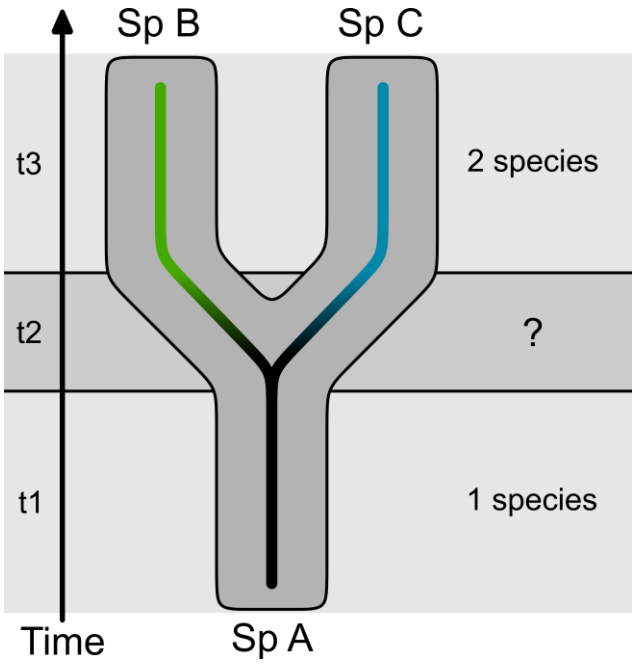


Figure 5. A simplified representation of the process of speciation over time. Modified after de Queiroz (1998).

Applying an evolutionary scenario to the previous definitions changes the static vision of the concept and becomes even harder to think about species as a group of individuals following mating rules. In figure 5 there is an exemplified process of speciation in which one species (*SpA*) at a given time (*t1*) gradually diversifies (*t2*), founding two different species (*SpB* and *SpC*) at the end of the process (*t3*). In this example there are new questions arising: When does *SpA* stop being *SpA*? Is *SpA* becoming *spB* or *spC*? At what moment can we consider *SpB* independent from *SpA*? Is this moment the same for *SpC*? Is *t2* as clearly defined as in the figure? This is part of the evolutionary species concept issue, yet as for any other question regarding the species concept the answer is open to interpretation. In a constant changing environment, individuals are constantly adapting and therefore diversifying. So, if species are constantly diversifying, we are constantly living in *t2*. In a geographical scale, there are going to be different environments or selection pressures creating different allopatric or sympatric barriers, challenging also the ecological species concept. Therefore, at what time should we use one definition or another? In most of the cases, species definitions are approaching a debate towards species delimitation leaving aside the conceptualization of the entities itself (De Queiroz, 2007). Since *t2* is not easily defined and during the speciation process there are more than one responsible event, different species criteria (or concepts) could apply to delimitate realities, and therefore cannot be wrong (de Queiroz, 1998; Zachos, 2016). The concept used to define a species will depend on the question to be answered.

The problem of a concept moves to delimitation of the concept itself. Most of the species definitions mention lineages, meta-populations, populations, groups of organisms or individuals, among others, as elementary units. However, it is not clear yet what is the unit of selection. Following the genetic species concept (and related concepts) the selection occurs at the genetic level. Yet, once again, other aspects are being neglected, for example the role of interactions with other species such as symbionts (e.g.: Rosenberg et al., 2007; Gilbert et

al., 2010). This idea brings into the game that species alone (and all their different levels) might not be the only unit of selection, introducing the holobiont concept and the role of co-evolution. But again, comes a problem of delimitation, since the boundaries of the holobiont are not always properly defined, from the local influence of the host to a whole new concept called “nested ecosystems” (McFall-Ngai et al., 2013; Dittami et al., 2019).

After having scratched the surface of how complicated can get the concept of species, processes such as evolution of a single lineage (including many distinct species) may become extremely complex. This question gets further complicated when talking about Prokaryotes, where even the concept of phylogenetic tree loses his meaning due to the gene flow between distantly related individuals (Corel et al., 2016). Therefore, given the complexity of eukaryotic genomes and their evolution (Snel et al., 2005), is it accurate to represent such complex processes into a tree? Maybe not all of them, but as Doolittle and Bapteste (2007) have discussed, it might be possible to reconstruct some complex patterns of similarity and differences among living things, without “seeking some elusive unifying meta-narrative”. Since there is no tool or approach carrying the only-one-answer (Fig. 6), it is important to integrate all variety of tools in order to improve the understanding of the processes that drive the diversity and evolution of any given taxa, without searching for one only conclusive result.

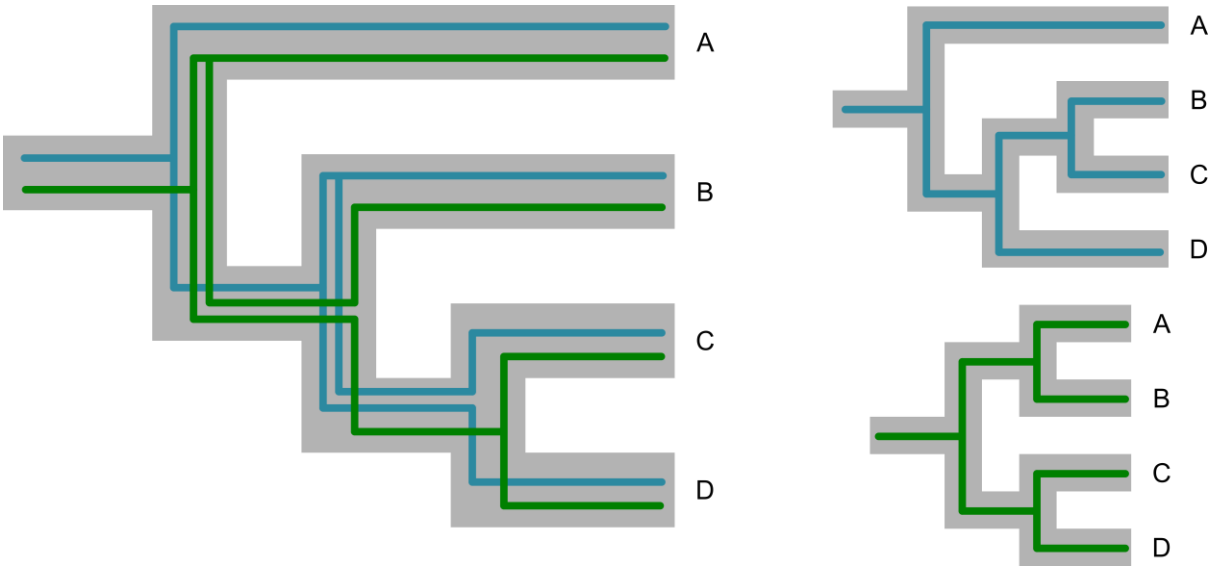


Figure 6. A simplified schematic representation of a “true” historic diversification event (left) compared to the vision obtained by two different tools, approaches or species concepts (right: blue and green). I.e.: the evolutionary hypothesis of a species following morphology and that of the genetic marker (e.g.: rDNA), or two different evolutionary hypotheses obtained by two different marker genes (e.g.: 18S rDNA gene vs 28S rDNA gene). Modified after Slowinski et al. (1999).

PhD Objectives

Both Nassellaria and Spumellaria represent a valuable tool for paleo-environmental reconstruction studies thanks to their extensive and continuous fossil record dating back to the Cambrian. Their phylogenetic relationships with other important radiolarian lineages such as Acantharia and Collodaria, makes them key components for the understanding of fundamental radiolarian evolutionary and ecological questions.

1- The first objective of this PhD thesis is to explore the extant diversity of Nassellaria and Spumellaria. To do so, a phylogeny will be built based on single cell specimens to link their phylogenetic marker and their morphology. This will lead to the establishment of a reference morpho-molecular framework for Nassellaria and Spumellaria classification and evolution.

2- My second objective is to integrate the current knowledge on molecular diversity of all Radiolaria together and study their relationships. Their excellent fossil record allows calibrating the molecular clock and explore in detail their evolutionary patterns related to global geological and environmental changes.

3- Finally, I will use the exhaustive morpho-molecular framework produced to investigate radiolarian biogeography and distribution in the ocean through a metabarcoding approach in order to infer their main ecological patterns in the global ocean.

Chapter 1

Molecular diversity of Nassellaria and Spumellaria

Context of the work

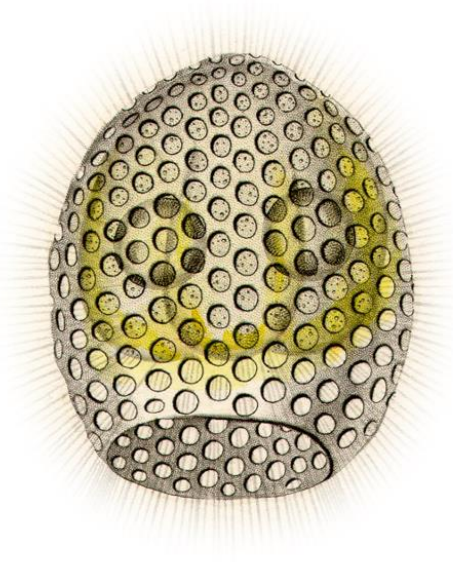
The first chapter of this thesis is dedicated to the exploration of the molecular diversity of both Nassellaria and Spumellaria. I essentially used a single cell approach to characterize the diversity of these two groups by integrating phylogenetic relationships with morphological characteristics. Besides, their excellent fossil record allowed us contextualizing their evolutionary history in relation to global environmental and geological changes, leading to the reconsideration of certain fossil groups and a better understanding of their extant diversity.

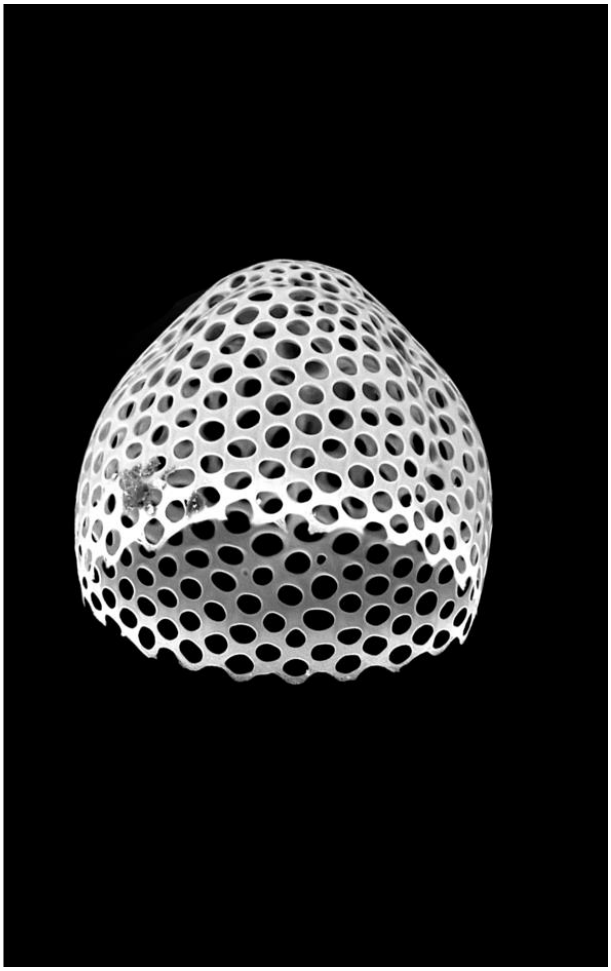
During the sampling of the specimens used in this study I had the opportunity to visit the *Observatoire Oceanologique de* Villefranche-sur-Mer where I collaborated with Emilie Villar in the response of the thermal stress of collodarian symbionts (See collaborations 1 in annexes, page 197). In this collaboration I set up and performed respiration experiments to measure the respiration-photosynthesis ratio over a thermal stress gradient.

To carry out the work presented in this chapter, I also had the opportunity to participate in the 2017 Mediterranean Ocean Observing System for the Environment (MOOSE-GE) oceanographical cruise (Annexes, Oceanographic Expeditions, page 194). On board I contributed to start a biomonitoring strategy through a metabarcoding approach (data that I have used later in chapter 3.2). During this cruise I sampled Radiolaria specimens that were used in the following phylogenetic analysis. Besides, I collaborated with Natalia Llopis in the estimation of the silica production of Rhizaria in the global ocean (see collaborations 2 in annexes, page 198) by teaching and helping in the isolation and identification of Radiolaria specimens found during the cruise.

Chapter 1.1

Time calibrated morpho-molecular classification of Nassellaria (Radiolaria)





Time calibrated morpho-molecular classification of Nassellaria (Radiolaria)

Miguel M. Sandin[1], Loic Pillet[1], Tristan Biard[1], Camille Poirier[1], Estelle Bigeard[1], Sarah Romac[1], Noritoshi Suzuki[2], Fabrice Not[1]

1-Sorbonne Université, CNRS - UMR7144 - Ecology of Marine Plankton Group - Station Biologique de Roscoff, 29680 Roscoff, France.
2-Department of Earth Science, Graduate School of Science, Tohoku University, Sendai 980–8578, Japan.

Published in: *Protist,* 2019
doi: 10.1016/j.protis.2019.02.002

Abstract

Nassellaria are marine protists belonging to the Radiolaria lineage (Rhizaria). Their skeleton, made of opaline silica, exhibit an excellent fossil record, extremely valuable in micro-paleontological studies for paleo-environmental reconstruction. Yet, to date very little is known about the extant diversity and ecology of Nassellaria in contemporary oceans, and most of it is inferred from their fossil record. Here we present an integrative classification of Nassellaria based on taxonomical marker genes (18S and 28S ribosomal DNA) and morphological characteristics obtained by optical and scanning electron microscopy imaging. Our phylogenetic analyses distinguished 11 main morpho-molecular clades relying essentially on the overall morphology of the skeleton and not on internal structures as previously considered. Using fossil calibrated molecular clock we estimated the origin of Nassellaria among radiolarians primitive forms in the Devonian (ca. 420 Ma), that gave rise to living nassellarian groups in the Triassic (ca. 250 Ma), during the biggest diversification event over their evolutionary history. This morpho-molecular framework provides both a new morphological classification easier to identify under light microscopy and the basis for future molecular ecology surveys. Altogether, it brings a new standpoint to improve our scarce understanding of the ecology and worldwide distribution of extant nassellarians.

Introduction

Along with Foraminifera, Radiolaria constitute the Phylum Retaria, within the supergroup Rhizaria, one of the 8 major branches of eukaryotic life (Burki and Keeling, 2014). Radiolarians are marine heterotrophic protists, currently classified in 5 taxonomic orders based on morphological features and chemical composition of their biomineralized skeleton. Acantharia possess a skeleton made out of strontium sulfate ($SrSO_4$), while opaline silica (SiO_2 nH_2O) is found in skeletons of Taxopodia and the polycystines Collodaria, Nassellaria and Spumellaria (Suzuki and Not, 2015). The robust silica skeleton of polycystines preserves well

in sediments and hard sedimentary rocks, providing an extensive fossil record throughout the Phanerozoic (De Wever et al., 2001). Essentially studied by micro-palaeontologists, classification and evolutionary history of Radiolaria are largely based on morphological criteria (Suzuki and Oba, 2015) and very little is known about the ecology and diversity of contemporary species.

Among polycystines, Nassellaria actively feed on a large variety of preys, from bacteria to mollusc larvae (Anderson, 1993; Sugiyama et al., 2008), contributing significantly to trophic webs dynamic of oceanic ecosystems. Some nassellarian species host photosynthetic algal symbionts, up to 50 symbionts per host cells, mainly identified as dinoflagellates (Suzuki and Not, 2015; Zhang et al., 2018, Decelle et al., 2015; Probert et al., 2014). This mixotrophic behaviour may influence their distribution patterns, being in surface tropical waters the greatest diversity and abundance values, and decreasing towards the poles and at depth (Boltovskoy and Correa, 2016a; Boltovskoy, 2017). Only a few taxa are restricted to deep waters (1000-3000m) in which no photosymbionts have ever been described (Suzuki and Not, 2015).

Unlike other radiolarians, nassellarian skeleton is heteropolar, aligned along an axis and not a centre. This skeleton is divided in three main different segments: the cephalis always present (or 1st segment), the thorax (2nd segment) and sometimes an abdomen (3rd segment) and post-abdominal segments (Campbell, 1954). The cephalis contains the initial spicular system, which structure was widely used for the taxonomic classification at family or higher levels due to its early development in the ontogenetic growth (De Wever et al., 2001; Petrushevskaya, 1971b). Its basic architecture is the component of *A*-rod (apical rod), *D*- (dorsal rod), *V*- (ventral rod), *MB* (median bar), occasionally Ax (axobate node), *l*-rod (lateral rod from *MB* at the *A*-rod side) and *L*-rod (lateral rod from *MB* at the *V*-rod side). All these initial spicules except for *V*- and *l*-rods are always present in Nassellaria (Supplementary Material Fig. S1). The architecture of the initial spicular system has been used not only in nassellarian classification but also for Collodaria, Spumellaria and other Polycystines such as Entactinaria, a group considered to be an early lineage in the Paleozoic. Morphology based taxonomic classifications have divided extant Nassellaria in nearly 25 families, 140 genera and 430 recognized species (Suzuki and Not, 2015). At higher level, they are currently divided in 7 super-families: Acanthodesmoidea (Hertwig, 1879; sensu Dumitrica in De Wever et al., 2001), Acropyramioidea (Haeckel, 1882; sensu emend. Petrushevskaya, 1981), Artostrobioidea (Riedel, 1967; sensu O'Dogherty, 1994), Cannobotryoidea (Haeckel, 1882; sensu Petrushevskaya, 1971a), Eucyrtidioidea (Ehrenberg, 1846; sensu Dumitrica in De Wever et al., 2001), Plagiacanthoidea (Hertwig, 1879; sensu Petrushevskaya, 1971a), Pterocorythoidea (Haeckel, 1882; sensu Matsuzaki et al., 2015) and some undetermined families (i.e., Theopiliidae, Bekomidae, Carpocaniidae) (Matsuzaki et al., 2015).

The expertise required to collect, sort and identify living nassellarian specimens along with their short maintenance time in cultures (Anderson et al., 1989; Suzuki and Not, 2015) make the study of their taxonomy and ecology arduous. In addition, the low DNA concentration per individual cell challenges the molecular approach to address such questions. Phylogenetic studies have demonstrated the effectiveness of combining single cell DNA sequencing and imaging data in assessing classification and evolutionary issues beyond

morphological characteristics (Bachy et al., 2012; Decelle et al., 2012b; Biard et al., 2015). Acquisition of reference DNA barcode based on single cell sequencing of isolated specimens have also established the basis for further molecular ecology surveys inferring the actual diversity and ecology in the nowadays oceans (Decelle et al., 2013; Nitsche et al., 2016; Biard et al., 2017). In addition, the use of fossil-calibrated molecular clock has become a popular tool in molecular evolution, addressing diversification rates (e.g., in diatoms; Lewitus et al., 2018), exploring co-evolution processes (e.g., Acantharia and *Phaeocystis*; Decelle et al., 2012a; or bees and eudicots; Cardinal and Danforth, 2013), resolving evolutionary patterns along with the fossil record (e.g., on Ray finned fishes; Giles et al., 2017; or in holothuroids; Miller et al., 2017) or tracing the origin and the evolutionary history outside fossil limits (e.g., on the origin of eukaryotes; Douzery et al., 2004; Berney and Pawlowski, 2006; Eme et al., 2014). To date, few phylogenetic studies have explored the extant genetic diversity of Nassellaria and the relationships among families and their evolutionary patterns remains still elusive (Kunitomo et al., 2006; Yuasa et al., 2009; Krabberød et al., 2011). So far, with a total of 16 sequences from morphologically described specimens, covering 5 of the 7 super-families identified, Eucyrtidioidea was considered as the most basal and the rest of the represented groups have uncertain phylogenetic positions (Krabberød et al., 2011).

The latest classification of Nassellaria (Matsuzaki et al., 2015) attempted to integrate the extensive morphological knowledge (Hertwig, 1879; Haeckel, 1882; Petrushevskaya, 1971a; De Wever et al., 2001) with the few molecular analyses performed for Nassellaria (Krabberød et al., 2011). Here we introduce an integrative morpho-molecular classification of Nassellaria obtained by ribosomal DNA taxonomic marker genes (18S and partial 28S) and imaging techniques (Light microscopy, Scanning Electron Microscopy and/or Confocal Microscopy), compared with the current morphological classification. In addition, the extensive fossil record available for Nassellaria allowed us to time-calibrate our phylogenetic analysis based on molecular dating and infer their evolutionary history contextualized with geological environmental changes at a global scale. Finally, the inclusion of environmental sequences gave insights in the extant genetic diversity of nassellarians in contemporary oceans.

Material and Methods

Sampling and single cell isolation: Plankton samples were collected off Sendai (38° 0’ 28.8’’ N, 142° 0’ 28.8’’ E), Sesoko (26° 48’ 43.2’’ N, 73° 58’ 58.8’’ E), the Southwest Islands, South of Japan (28° 14’ 49.2’’ N, 129° 5’ 27.6’’ E), in the bay of Villefranche-sur-Mer (43° 40’ 51.6’’ N, 7° 19’ 40.8’’ E) and in the west Mediterranean (MOOSE-GE cruise) by net tows, Vertical Multiple-opening Plankton Sampler (VMPS) or Bongo net (20-300 µm). More information on sampling methodology can be found in the RENKAN database (http://abims.sb-roscoff.fr/renkan). Targeted specimens were individually handpicked with Pasteur pipettes from the samples and transferred 3-4 times into 0.2 µm filtered seawater to allow self-cleaning from debris, particles attached to the cell or preys digestion. Images of live specimens were taken under an inverted microscope and thereafter transfer into 1.5 ml Eppendorf tubes containing 50 µl of molecular grade absolute ethanol and stored at -20 ºC until DNA extraction.

Single cell morphological identification: Nassellaria specimens were identified at the species level, referring to pictures of holotypes and reliable specimens, through observation of live images and posterior analysis of the skeleton by scanning electron microscopy and/or confocal microscopy. At the species, genus and family levels, we respected the Articles 66 to 70 of "types in the genus group" in the International Code of Zoological Nomenclature 2000 (hereinafter ICZN 2000; http://www.iczn.org/iczn/index.jsp). The family level was decided by the similarity of the type genus of the family based on ICZN 2000. This procedure has the advantage to directly use the fossil taxonomy (Matsuzaki et al., 2015) and the taxonomic concept of superfamily (Matsuzaki et al., 2015; Suzuki and Not, 2015) without ambiguity in the name used. Based on this rule, we updated the species name of specimens illustrated in previous studies and that are included in our phylogeny (see Supplementary Material Table 1 for taxonomic authority of specimens included in our study).

DNA extraction, amplification and sequencing: DNA was extracted using the MasterPure Complete DNA and RNA Purification Kit (Epicentre) following manufacturer's instructions. Once DNA was extracted and recovered, waste (i.e., pellet debris) from the extraction procedure were diluted in milliQ water to preserve skeletons and stored at -20°C. Both 18S rDNA and partial 28S rDNA genes (D1 and D2 regions) were amplified by Polymerase Chain Reaction (PCR) using Radiolaria and Nassellaria specific and general primers (Table 1). For further details about rDNA amplification see: dx.doi.org/10.17504/protocols.io.t5req56. PCR amplicons were visualized on 1% agarose gel stained with ethidium bromide. Positive reactions were purified using the Nucleospin Gel and PCR Clean up kit (Macherey Nagel), following manufacturer's instructions and sent to Macrogen Europe for sequencing.

Table 1. Primer sequences used for DNA amplification and sequencing.

Targeted gene	Primer	Specificity	Sequence 5'-3'	Direction	Tm °C	Reference
18S (1st part)	SA	Eukaryotic	AACCTGGTTGATCCTGCCAGT	Forward	56-60	Medlin et al., 1988
	18S NasIbR	Nassellaria	GAGACTACGACGGTATCTGATC	Reverse	60	This study
	S879	Radiolaria	CCAACTGTCCCTATCAATCAT	Reverse	56	Decelle et al., 2012b
18S (2nd part)	S32_TASN	Radiolaria	CCAGCTCCAATAGCGTATRC	Forward	52	Ishitani et al. 2012
	V9R	Eukaryotic	CCTTCYGCAGGTTCACCTAC	Reverse	52	Romac (Unpub.)
	18S NassIIF	Nassellaria	AGCATGGAATAATAACTGATGA	Forward	57	This study
	18S NassIIR	Nassellaria	CACCARTTCATCCAATCGGTAG	Reverse	57	This study
28S (D1+D2)	28S NasF	Nassellaria	AGTAACGGCGAGTGAAGC	Forward	56	This study
	28S NasR	Nassellaria	CCAACATACDTGCTCTTGT	Reverse	56	This study
	28S Rad2	Radiolaria	TAAGCGGAGGAAAAGAAA	Forward	50	Ando et al., 2009
	ITSa3	Radiolaria	TCACCATCTTTCGGGTCCCAACA	Reverse	50	Ando et al., 2009

Phylogenetic analyses: After sequencing, forward and reverse sequences were checked and assembled using ChromasPro software version 2.1.4 (2017). Sequences were compared to reference database (GenBank) using the *BLAST search* tool integrated in ChromasPro to discriminate radiolarian sequences from possible contamination. Similar sequences identified in GenBank were retrieved and integrated in our databases. Two different datasets for each genetic marker (18S rDNA gene and partial 28S rDNA gene) were obtained and align separately using Muscle (Edgar, 2004) implemented in SeaView version 4.6.1 (Gouy et al., 2010) and manually checked. For both genes, the 18S rDNA (61 taxa, 1890 positions) and the 28S rDNA (57 taxa, 700 positions), phylogenetic analyses were performed independently. The

best nucleotide substitution model was chosen following the corrected Akaike Information Criterion (AICc) using the *modelTest* function implemented in the R (R Core Team, 2014) package *phangorn* (Schliep, 2011). The obtained model (General Time Reversible with Gama distribution and proportion of Invariable sites, GTR+G+I) was applied to each data set in R upon the packages *APE* (Paradis et al., 2004) and *phangorn* (Schliep, 2011). A Maximun Likelihood (ML) method (Felsenstein, 1981) with 1000 replicates of bootstrap (Felsenstein, 1985) was performed to infer phylogenies. Since the topology and bootstrap support within main clades of both markers agree, the 18S rDNA and the 28S rDNA were concatenated in order to improve phylogenetic resolution. A final data set containing 90 taxa and 2590 positions was used to infer phylogenies following the previous methodology. The best model obtained was GTR+G+I, with 4 intervals of the discrete gamma distribution, and a ML method with 100 000 bootstraps were performed. In parallel, a Bayesian analysis was performed using BEAST version 1.8.4 (Drummond et al., 2012) with the same model parameters over 10 million generations sampled every 1000 states, to check the consistency of the topology and to calculate posterior probabilities. Final tree was visualized and edited with FigTree version 1.4.3 (Rambaut 2016). All sequences obtained in this study and used for the phylogeny were submitted in GenBank under the accession numbers: MK396913 - MK397005.

Molecular clock analyses: The resulting concatenated dataset of the 18S rDNA and the partial 28S rDNA used to infer phylogenies, was used for the molecular clock analyses following established protocols implemented in BEAST version 1.8.4 (Drummond et al., 2012). In order to avoid the assumption of substitution rates correlation between neighbour branches, a Bayesian uncorrelated relaxed clock method (Drummond et al., 2006) was performed to calculate the divergence times of Nassellaria with a relaxed lognormal distribution. GTR+G+I with estimated base frequencies and 4 gamma categories was chosen for the substitution model, according to previous analyses (see section above). Both Spumellaria (outgroup) and Nassellaria nodes were forced to be monophyletic according to bootstrap results in the phylogenetic analysis. According to software options, a speciation Yule process (Yule, 1925; Gernhard, 2008) with random starting tree was chosen as tree prior, since our analyses are beyond population level. Markov chains were run in three parallel replicates for 100 million generations sampled every 1000 states and operators were left as default. Different replicates were combined in LogCombiner version 1.8.4 (Drummond et al., 2012) after removing the first 25% states. Twelve nodes were chosen to carry out this calibration (explained below from the oldest to the newest calibration age, and given the name of the node for the taxa they cover):

· 1. Root: The calibration for the root of the tree corresponds to the hypothesized last common ancestor between Nassellaria and Spumellaria (De Wever et al., 2001). A uniform distribution with a minimum bound of 200 million years ago (Ma) and a maximum of 600 Ma (U [200, 600]) was set to allow uncertainty in the diversification of both groups and to establish a threshold restricting the range of solutions for the entire tree.

· 2. Spumellaria: The second node corresponds to the origin of Spumellaria, the outgroup of our phylogenetic analyses. The first spumellarians appear in the fossil record in the Lower

Ordovician (ca. 477.7 - 470 Ma) with the genera *Antygopora* (J. C. Aitchison et al., 2017). Therefore, the node was normally distributed with a mean of 474 and a standard deviation of 50 to avoid uncertainty due to the low genetic diversity of the outgroup: N (474, 50).

· 3. Nassellaria: The first nassellarian-like fossil appear in the Devonian (ca. 419.2 – 358.9 Ma) with internal spicules typical from Nassellaria (Cheng, 1986; Kiessling and Tragelehn, 1994; Schwartzapfel and Holdsworth, 1996). But it is not until the Early Triassic (ca. 250 – 240 Ma) that the first multi-segmented nassellarians appeared in the fossil record (Sugiyama, 1992, 1997; De Wever et al., 2003; Hori et al., 2011; Dumitrica, 2017a). Because it is unclear whether the first nassellarian-like fossils are true nassellarians, a uniform distribution with a maximum boundary of 500 and a minimum of 180 (U [180, 500]) was chosen to allow uncertainty in the origin of living Nassellaria (Suzuki and Oba, 2015).

· 4. Eucyrtidioidea N (246, 10): The first appearance of this family in the fossil record is with the genus *Triassocampe* (Sugiyama, 1997) in the early Middle Triassic (Anisian: ca. 246.8 - 241.5). And there are continuous fossil record since the Carnian (ca. 237 – 228.5 Ma; as well Late Triassic) to present of the genus *Eucyrtidium* (Petrushevskaya, 1971a; De Wever et al., 2001, 2003). Hence, the node was normally distributed with an average mean of 246 and a standard deviation of 10 (N (246, 10)).

· 5. Lophophaenidae N (191, 10): This family belongs to the Superfamily Plagiacanthoidea, and the genus *Thetisolata* in the Early Jurassic (Pliensbachian: ca. 191.4 – 183.7 Ma) is the oldest fossil record associated to this family (De Wever, 1982).

· 6. Artostrobioidea N (182, 10): The genus *Artostrobium,* questionably assigned by Matsuoka (2004), is the oldest fossil for this family, and its first appearance in the fossil record is dated in the Early Jurassic (Toarcian: ca. 183.7 - 174.2 Ma) (Matsuoka, 2004).

· 7. Sethoperidae N (180, 30): Like Lophophaenidae, this family also belongs to the Plagiacanthoidea, but its first appearance in the fossil record dates in the Middle Jurassic (Bajocian: ca. 170.3 – 168.3 Ma) with the genus *Turriseiffelus* (Dumitrica and Zügel, 2003). However, the fossil record of this family seems to be patched, since the very first member associated to this family is restricted to a single stage in the Early Jurassic (genus *Carterwhalenia*, Pliensbachian: ca. 191.4 – 183.7) (O'Dogherty et al., 2009a). Due to this uncertainty in the appearance of this family, we decided to increase the node prior standard deviation to 30 Ma.

· 8. Cannobotryoidea N (122, 20): Petrushevskaya (1971b) dated the first appearance of this family in the Cretaceous (ca. 145 – 66 Ma), and later the time frame was restricted to the Early Cretaceous (ca. 145 – 100 Ma) (De Wever et al., 2001, 2003). However, there is no consensus about whether the first morphologically similar genera (*Ectonocorys* or *Solenotryma*) to this group that appear in the fossil record really belong to it or not. Therefore, since no adequate evidence has been shown we slightly increase the standard deviation (sd: 20 Ma) for this node prior to allow uncertainty (O'Dogherty et al., 2009a; Matsuzaki et al., 2015).

· 9. Acanthodesmoidea N (66, 10): The three families belonging to this superfamily (Acanthodesmiidae, Stephaniidae, Triospyrididae) have their first appearance in the fossil

record in the Paleocene (ca. 66 - 56) (Petrushevskaya, 1971a; De Wever et al., 2001, 2003), and the first genus described associated to this families corresponds to *Tholospyris* in the Early Paleocene (Kozlova, 1983, 1999).

· 10. Bekomidae N (62, 10): The genus *Bekoma* (as representative of the family Bekomidae) has its first appearance in the Late Paleocene (ca. 66 - 56) (Nishimura, 1992; De Wever et al., 2001).

· 11. Pterocorythidae N (57, 10): *Cryptocarpium* is the first, but doubted, representative of this family in the fossil record and it is dated in the Late Paleocene (ca. 66 - 56 Ma), followed by *Podocyrtis*, the first true representative, in the transition of the Eocene (ca. 56 – 33 Ma) (Sanfilippo and Riedel, 1992; De Wever et al., 2001; Hollis, 2006).

· 12. Carpocaniidae N (38, 10): The first representative of this family appears in the late Eocene (ca. 38 – 33.9 Ma) and correspond to the genus *Carpocanium* (De Wever et al., 2001; Kamikuri et al., 2012).

Post hoc analyses: In order to provide statistical support to our conclusions two different analyses were developed: a diversification of taxa over time (Lineages Through Time: LTT) and an ancestral state reconstruction. The former analysis was carried out with the *ltt.plot* function implemented in the package *APE* (Paradis et al., 2004) upon the tree obtained by the molecular dating analyses after removing the outgroup. The second analysis uses the resulting phylogenetic tree to infer the evolution of morphological characters. A numerical value was assigned to each state of a character trait, being 0 for the outgroup or the considered ancestral state, and 1 to 4 for Nassellaria or the presumed divergence state. In total 5 traits were consider (Table 2): the number of cyrtids/segments (1-monocyrtid, 2-dicyrtid, 3-tricyrtid, 4-multicyrtid), the complexity of the cephalis (1-simple / spherical, 2-hemispherical / elongated / reduced / fussed with thorax, 3- Lobes, 4-complex), the apical (A) ray, the ventral (V) ray (1-not projecting the skeleton, 2-small, 3-medium/big, 4-united with A/V) and the dorsal (D) and lateral rays (Lr and Ll) treated as one single character (D+Lr+Ll: 1-not projecting the skeleton, 2-small, 3-medium, 4-big/foot/wings). Once the character matrix was established a parsimony ancestral state reconstruction was performed to every character independently in Mesquite version 3.2 (Madison & Madison, 2017).

Environmental sequences: Each nassellarian sequence of the 18S rDNA and partial 28S rDNA was compared with publicly available environmental sequences in GenBank (NCBI) using BLAST (as of December 2017); in order to estimate the environmental genetic diversity of Nassellaria and eventually to check the genetic coverage of our phylogenetic tree. Environmental sequences were placed in our reference phylogenetic tree using the pplacer software (Matsen et al., 2010). A RAxML (GTR+G+I) tree was built for the placement of the sequences with a rapid bootstrap analysis and search for best-scoring ML tree and 1000 bootstraps.

Table 2. List of morphological characters (traits) and their states (1 to 4) used for the ancestral state reconstruction analysis.

		Traits				
		Number of segments	Complexity of cephalis	Apical ray	Ventral ray	Dorsal and lateral rays
State	1	Monocyrtid	Simple, Spherical	Not projecting the skeleton	Not projecting the skeleton	Not projecting the skeleton
	2	Dicyrtid	Hemispherical, elongated, Reduced, fussed with thorax	Small	Small	Small
	3	Tricyrtid	Lobes	Medium, big	Medium, big	Medium
	4	Multicyrtid	Complex	United with V	United with A	Big, foot, wings

Confocal (CM) and Scanning Electron Microscopy (SEM): After DNA extraction, nassellarian skeletons were recovered from the eluted pellet and handpicked under binoculars or inverted microscope. After cleaning and preparing the skeletons (detailed protocol in dx.doi.org/10.17504/protocols.io.ug9etz6 for SEM and dx.doi.org/10.17504/protocols.io.t5qeq5w for CM; protocol for CM imaging extracted from dx.doi.org/10.17504/protocols.io.vuze6x6) images were taken with an inverted Confocal Microscopy (Leica TCS SP5 AOBS) and/or FEI Phenom table-top Scanning Electron Microscope (FEI technologies).

Results

Comparative molecular phylogeny and morphological taxonomy

Final molecular phylogeny is composed of 90 distinct nassellarian specimens generated by 61 sequences of the 18S rDNA gene and of 57 sequences of the partial 28S (D1 & D2 regions) rDNA gene (Supplementary Material Table S2). From the 61 final sequences of the 18S rDNA gene, 38 were obtained in this study and 23 were previously available, of which 16 have been morphologically identified and 7 are environmental. Regarding the partial 28S rDNA gene, 55 new sequences were obtained in this study and 2 were previously available and morphologically described. The final alignment matrix has 32.4% of invariant sites, and in total includes 67 new sequences. All of these specimens cover 7 superfamilies (Acanthodesmoidea, Acropyramioidea, Artostrobioidea, Cannobotryoidea, Eucyrtidioidea, Plagiacanthoidea, Pterocorythoidea, and three undefined families), based on morphological observations performed with light microscopy (LM; Supplementary Material Fig. S2), scanning electron microscopy (SEM) and/or confocal microscopy (CM) on the exact same specimens for which we obtain a sequence. Overall, molecular phylogeny is consistent with morphological classification at the superfamily level, although there are some specific discrepancies. The phylogenetic analysis shows 11 clades (Fig. 1) clearly differentiated with high values of ML bootstrap (BS > 99) and posterior probabilities (PP > 0.86).

Clade A holds the most basal position with 16 sequences of which 10 are novel sequences, 5 were previously available and one is environmental. All specimens clustering within this clade (Fig. 2.A, Fig. 3.A) have a simple and round cephalis, an apical horn, a small

ventral rod and multisegmented (cephalis, thorax and several abdomen) skeleton with distinctive inner rings, that correspond to the Superfamily Eucyrtidioidea. All multisegmented nassellarians with spherical cephalis encountered in our study belong to this superfamily, both morphologically and phylogenetically. The rest of the clades group together with a high BS (100) and PP (1) values. Thereafter clades B, C, D and E cluster together in lineage II. Clades X, F and G constitute the lineage III, highly supported (100 BS and 1 PP) as sister group of the lineage IV. This last lineage it is composed by the clades H, as the basal group and clades I and J highly related phylogenetically (100 BS and 1 PP).

Within lineage II, clade B is represented by only one sequence from this study (Osh128), and its morphology matches with the superfamily Acropyramioidea (Fig. 2.B, Fig. 3.B) exhibiting a pyramidal skeleton constituted of a reduced cephalis and thorax. Clade C is composed by two novel sequences (Fig. 2.C, Fig. 3.C). Their overall round morphology and a characteristic small and flat cephalis not well distinguished from the thorax agrees with the undetermined family Carpocaniidae (Petrushevskaya, 1971a; De Wever et al., 2001). Its sister clade, the Clade D is constituted by three new (Ses58, Mge17-70 and Mge17-9) and one environmental sequences, and the morphology of these specimens (Fig. 2.D, Fig. 3.D) agrees with the definition of the Superfamily Artostrobioidea. They share a multisegmented skeleton without significant inner rings, a hemispherical cephalis and an important ventral rod. The last clade of the lineage II, clade E, gathers four new sequences, one and two environmental sequences. This clade includes all the monocyrtid (cephalis) nassellarians where the spines A and V are merged forming the so-called D-ring (Acanthodesmoidea). There are representatives for two out of three families (Stephaniidae is the missing family), yet no phylogenetic differences were found for the included families, Acanthodesmiidae (Fig. 2.E1, Fig. 3.E1) and Triospyrididae (Fig. 2.E2-E4, Fig. 3.E2).

In lineage III, clades X and F are highly supported as sister clades (100 BS and 1 PP). Clade X is established by two novel sequences morphologically identified as *Archipilium johannismonicae* (Fig. 2.X1, Fig. 3.X1) and *Enneaphormis enneastrum* (Fig. 2.X2, Fig. 3.X2). These two specimens have a very short or missing median bar (MB) allowing the development of three large feet with a three-pointed star (dorsal and lateral, left and right, rays) shape forming a significant circular frame where they build the thorax when present. In Clade F there are two novel sequences (Ses59 and Mge17-94) and a previously available sequence. These specimens are morphologically identified in the genus *Eucecryphalus* (Fig. 2.F, Fig. 3.F) and a third specimen as *Tricolocamptra* (AB246685). The former genus belongs to the undetermined family Theopiliidae (De Wever et al., 2001; Matsuzaki et al., 2015) with a hat shaped morphology and pronounced segmentation. The genus *Eucecryphalus*, the senior synonym of *Theopilium*, is the type genus of the family Theopiliidae (Matsuzaki et al., 2015), and thus Clade F is automatically specified as the “true” Theopiliidae. The taxonomic position of the second genus, *Tricolocamptra*, varies through references and the available image for the specimen lacks taxonomical resolution, therefore it is not possible to establish the link with any described morphological group.

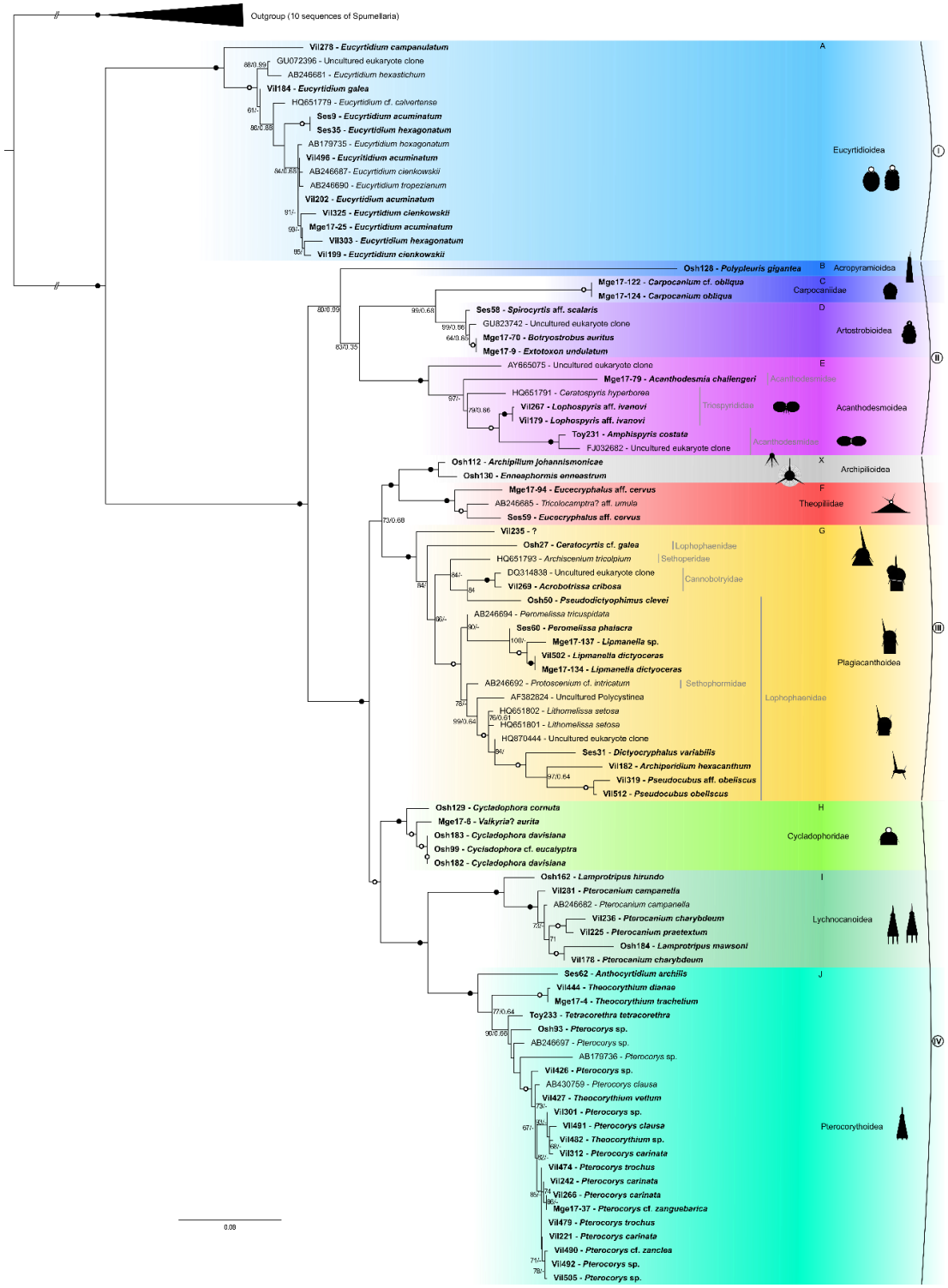


Figure 1. Molecular phylogeny of Nassellaria inferred from the concatenated complete 18S and partial 28S (D1-D2 regions) rDNA genes (90 taxa and 2590 aligned positions). The tree was obtained by using a phylogenetic Maximum likelihood method implemented using the GTR + γ + I model of sequence evolution. PhyML bootstrap values (100 000 replicates, BS) and posterior probabilities (PP) are shown at the nodes (BS/PP). Black circles indicate BS of 100% and PP > 0.99. Hollow circles indicates BS > 90% and PP > 0.90. Sequences obtained in this study are shown in bold. Eleven main clades are defined based on statistical support and morphological criteria (A, B, C, D, E, X, F, G, H, I, J). Figures besides clade names represent main features in the overall morphology of specimens included in the phylogeny. Ten Spumellaria sequences were assembled as out-group. Branches with a double barred symbol are fourfold reduced for clarity.

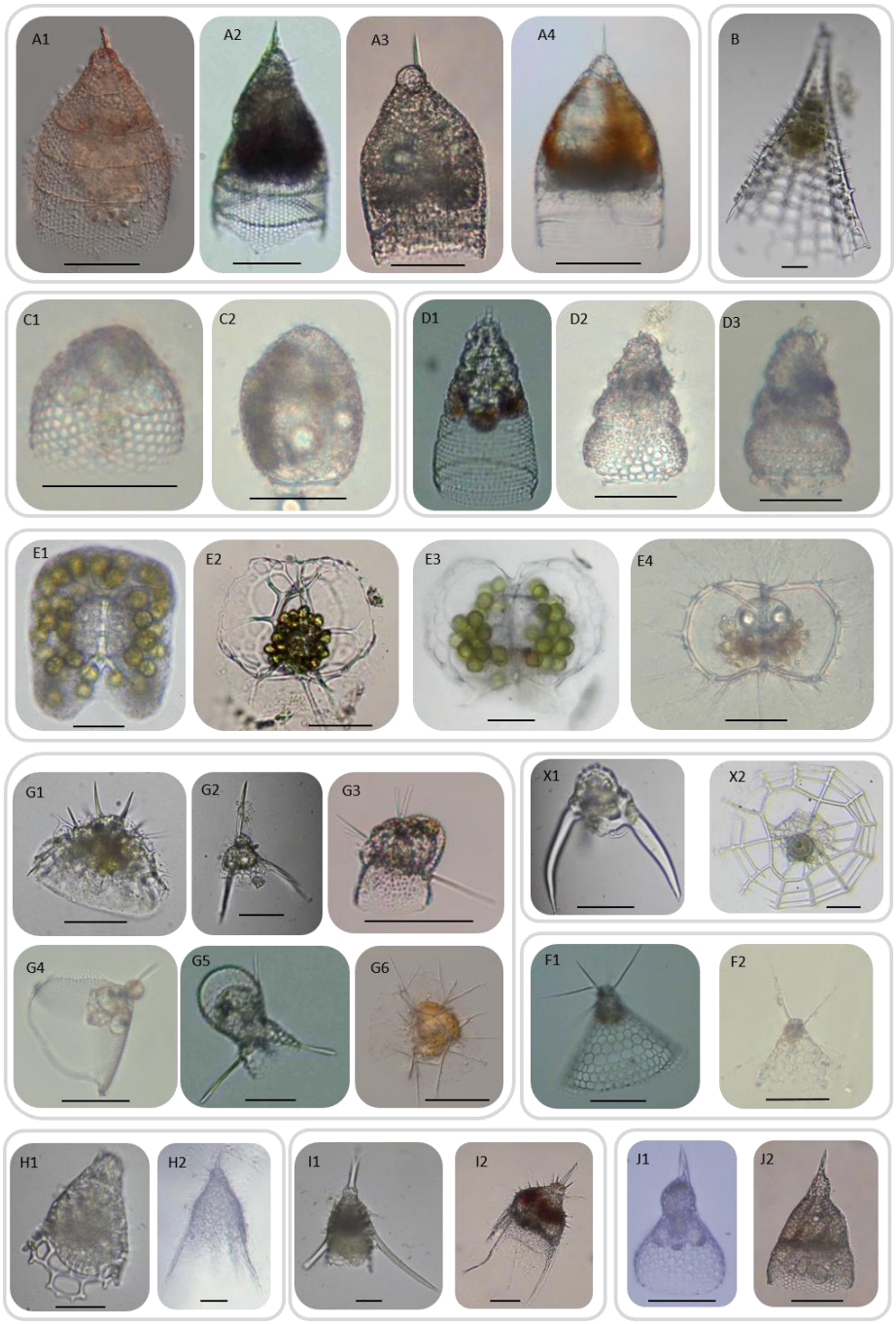


Figure 2. Light Microscopy (LM) images of live Nassellarian specimens used in this study for phylogenetic analysis. Letters correspond to its phylogenetic clade in Fig. 1. Scale bars (when available) = 50µm. (A1) Vil496: *Eucyrtidium acuminatum*. (A2) Ses35: *E. hexagonatum*. (A3) Vil278: *E. campanulatum*. (A4) Vil184: *E. galea*. (B) Osh128: *Polypleuris gigantea*. (C1) Mge17-122: *Carpocanium* cf. *obliqua*. (C2) Mge17-124: *C. obliqua*. (D1) Ses58: *Spirocyrtis* aff. *scalaris*. (D2) Mge17-9: *Extotoxon undulatum*. (D3) Mge17-70: *Botryostrobus auritus*. (E1) Toy231: *Amphispyris costata*. (E2) Vil267: *Lophospyris* aff. *ivanovi*. (E3) Vil179: *Lophospyris* aff. *ivanovi*. (E4) Mge17-79: *Acanthodesmia challengeri*. (X1) Osh112: *Archipilium johannismonicae*. (X2) Osh130: *Enneaphormis enneastrum*. (F1) Ses59: *Eucecryphalus* aff. *cervus*. (F2) Mge17-94: Eucecryphalus aff. cervus. (G1) Osh27: *Ceratocyrtis* cf. *galea*. (G2) Osh50: *Pseudodictyophimus clevei*. (G3) Vil269: *Acrobotrissa cribosa*. (G4) Mge17-134: *Lipmanella dictyoceras*. (G5) Ses60: *Peromelissa phalacra*. (G6) Vil512: *Pseudocubus obeliscus*. (H1) Osh129: *Cycladophora cornuta*. (H2) Mge17-6: *Valkyria*? *aurita*. (I1) Osh162: *Lamprotripus hirundo*. (I2) Vil225: *Pterocanium praetextum*. (J1) Vil444: *Theocorythium dianae*. (J2) Vil312: *Pterocorys carinata*.

With 14 different morpho-species and 11 different genus, Clade G is the most morphologically diverse clade. While all specimens have consistently one or two segments, the complexity of the cephalis and the sizes of the rays vary widely (Fig. 2.G1-G6, Fig. 3.G2-G6). Specimen HQ651793 (*Archiscenium tricolpium*) corresponds to the Family Sethophormidae (Sethophormidae Haeckel, 1882, *sensu emend*. Petrushevskaya, 1971a) within the Superfamily Plagiacanthoidea. The large cephalis and the umbrella shape of the thorax are characteristics of this family. The specimen Vil269 (Fig. 2.G3, Fig. 3.G3, *Acrobotrissa cribosa*) agrees with the description of the Superfamily Cannobotryoidea and its complex cephalic structure subdivided in different lobes is the main characteristic. The rest of the specimens of clade G can be grouped within the Family Lophophaenidae (Lophophaenidae Haeckel, 1882, *sensu* Petrushevskaya, 1971a), still within Plagiacanthoidea. Within this family there are two different subclades, those specimens with a prominent thorax (e.g. Mge17-134, *Lipmanella dictyoceras*, Fig. 2.G4; Mge17-137, *Lipmanella* sp. Fig. 3.G4) and those with a barely defined thorax (e.g., Ses60, *Peromelissa phalacra,* Fig. 2.G5, Fig. 3.G5) or not-defined thorax (e.g. Vil512, *Pseudocubus obeliscus*, Fig. 2.G6, Fig. 3.G6). Two last specimens, Osh27 (*Ceratocyrtis* cf. *galea*, Fig. 2.G1) and Osh50 (*Pseudodictyophimus clevei*; Fig. 2.G2, Fig. 3.G2), are morphologically identified as Lophophaenidae (HQ651793: De Wever et al., 2001; & Osh50: Matsuzaki et al., 2015; Petrushevskaya, 1971b) despite their phylogenetic position closely related to the Superfamily Cannobotryoidea.

Lineage IV is the most distal regarding the root of the tree. Within it, clade H is constituted by six new sequences very closely related to each other (BS>96 and PP>0.99). These specimens are morphologically assigned to the genus *Cycladophora* (Fig. 2.H1, Fig. 3.H) and *Valkyria* (Fig. 2.H2). Specimens from clade H show a conical and campanulate shaped morphology with a distinguishable cephalis and an apical horn compared to the wide thorax and not distinguished cephalis with two (V and A) rods of *Eucecryphalus*. Clade I is composed of six novel sequences and one previously available, while clade J gathers twenty new sequences and 3 previously available. These clades comprise nassellarians with an apical stout horn, a spherical (clade I) or elongated cephalis (clade J) and a truncated conical thorax; the abdomen, if present, is also truncated and sometimes not well defined, agreeing with the definition of the Superfamily Pterocorythoidea. Clade J corresponds to those with elongated cephalis (Fig. 2.J, Fig. 3.J) or Pterocorythidae (Pterocorythidae Haeckel, 1882, *sensu* De Wever et al., 2001). Clade I includes nassellarians with spherical head, ventral short ray and three feet (Fig. 2.I2, Fig. 3.I2), all characteristics of the Family Lychnocanidae (Lychnocaniidae Haeckel, 1882, *sensu* emend. Suzuki in Matsuzaki et al. 2015). Two specimens (Fig. 2.I1, Fig. 3.I1) of the undetermined family Bekomidae (Bekomidae Dumitrica in De Wever et al., 2001), are scattered among clade I showing no phylogenetic differences between members of both families. The BS (93) and the PP (0.99) values establish a strong phylogenetic relationship within this lineage, and so does the overall skeleton shape.

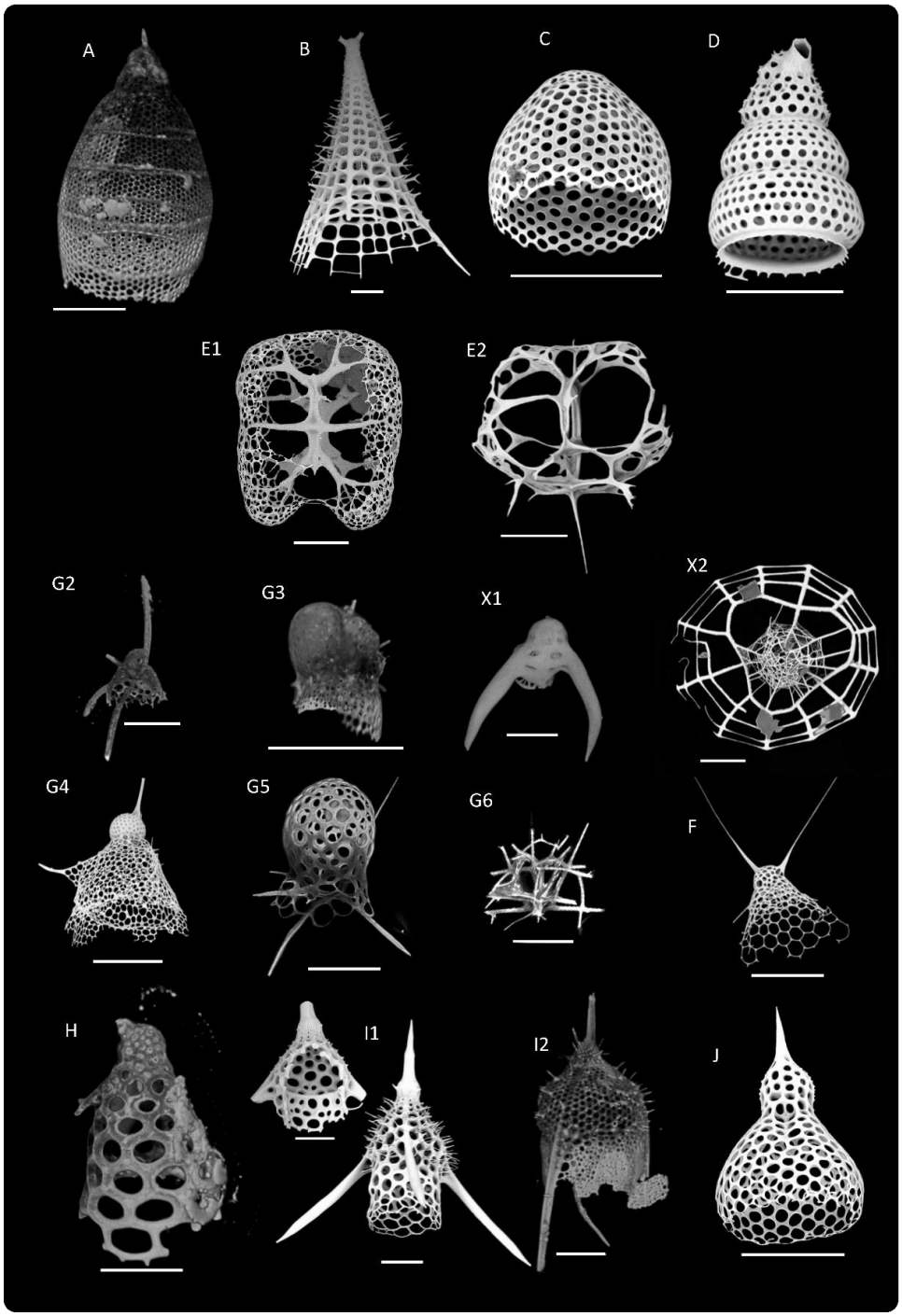


Figure 3. Scanning Electron Microscopy (SEM) and/or Confocal Microscopy (CM) images of Nassellarian specimens used in this study for phylogenetic analysis or morphologically related to one of the morpho-molecular clades of Fig. 1. Letters correspond to its phylogenetic clade in Fig. 1. Scale bars = 50µm. (A) Vil496: *Eucyrtidium acuminatum* (CM). (B) Osh128: *Polypleuris gigantea* (SEM). (C) Mge17-122: *Carpocanium* cf. *obliqua* (SEM). (D) Mge17-70: *Botryostrobus auritus* (SEM). (E1) Toy231: *Amphispyris costata* (SEM). (E2) Vil267: *Lophospyris* aff. *ivanovi* (SEM). (X1) Osh112: *Archipilium johannismonicae* (CM). (X2) Osh130: *Enneaphormis enneastrum* (SEM). (F) Mge17-94: *Eucecryphalus* aff. *cervus* (SEM). (G2) Osh50: *Pseudodictyophimus clevei* (CM). (G3) Vil269: *Acrobotrissa cribosa* (CM). (G4) Mge17-137: *Lipmanella* sp. (SEM). (G5) Ses60: *Peromelissa phalacra* (CM). (G6) Vil512: *Pseudocubus obeliscus* (CM). (H) Osh129: *Cycladophora cornuta* (CM). (I1 left) Osh184: *Lamprotripus hirundo* (SEM). (I1 right) Osh127: *Lamprotripus hirundo* (SEM). (I2) Vil225: *Pterocanium praetextum* (CM). (J1) Vil444: *Theocorythium dianae* (SEM).

Molecular dating

The molecular clock dated the diversification between Spumellaria and Nassellaria (the root of the tree) with a median value of 512 Ma (95% Highest Posterior Density -HPD-: between 600 and 426 Ma) (Fig. 4). From now on, all dates are expressed as median values followed by the 95% HPD interval. Despite the large uniform distribution given to the ingroup (U [180, 500]), the first diversification of Nassellaria was settled at 423 (500-342) Ma. Clade A had its first radiation dated in 245 (264-225) Ma. The common ancestor to all the other clades diversified at around 340 (419-271) Ma into two main groups, the so-called lineage II and the lineages III-IV splitting into two other lineages soon afterwards. Within lineage II, the first diversification occurs at 276 (354-209) Ma, and clades C and D diversified 197 (250-160) Ma. Thereafter in this lineage the phylogenetic relationships are dubious, however clade D diversified from any other clade 248 (315-191) Ma. Diversification within these clades was 28 (49-4) Ma for clade C, 175 (194-155) Ma for clade D and 77 (95-60) Ma for clade E. The lineages III and IV diversified from each other 274 (344-215) Ma. Lineage III rapidly diversified 243 (304-197) Ma when clades F and X split from clade G, followed by the fast diversification of clade G 196 (241-170) Ma. It was 168 (243-76) Ma when clade F separate of clade X, and 86 (157-32) Ma and 40 (106- 4) Ma when they respectively diversified. The last lineage (IV) diversified between clades H, I and J 206 (287-129) Ma, and clades I and J 138 (206-88) Ma. Despite this early divergence between clades, radiation within clades was more recent, being 87 (171-26) Ma, 70 (87-53) Ma and 73 (90-58) Ma for clades H, I and J, respectively.

Post-hoc analyses

The lineages through time analysis (Fig. 5) shows a classic exponential diversification slope, with a 0.0109 rate of speciation (Ln(lineages)·million years^{-1}). The first diversification of extant Nassellaria (~423 Ma) corresponds to the divergence between clade A and the rest of the clades. However, the first increase in the slope (up to: 0.017, t-test: $p<0.001$) occurs at ~275 Ma, when the main lineages start expanding; lineage II splits between clade B and clades C, D and E, clade A diversifies and the evolutionary lineages III and IV diverged from each other followed by the rapid diversification of lineage III. After this sudden increase of the main lineages, a second diversification event happened (ca. 198 Ma) where both the evolutionary lineage IV and the clade G diversify and clade C splits from clade D. Thereafter, the diversification seems to be stepped and separate in different periods of time, were only lineage III keep diversifying. The last and relatively uninterrupted diversification occurred ~82 Ma corresponding to the speciation within the already present clades and the first diversification of the clades H, F, E, J, I, X and C.

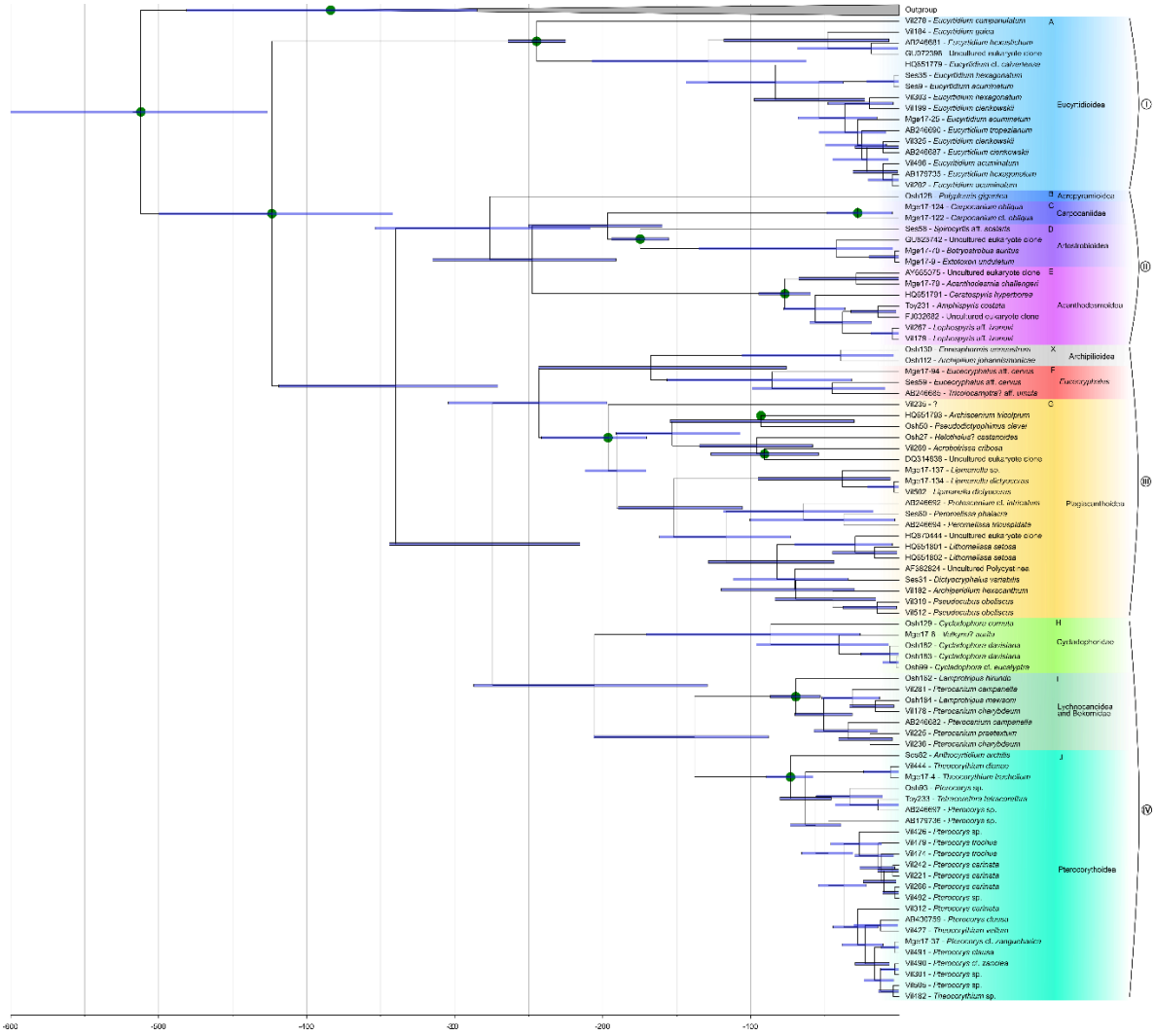


Figure 4. Time-calibrated tree (Molecular clock) of Nassellaria, based on alignment matrix used for phylogenetic analyses. Node divergences were estimated with a Bayesian relaxed clock model and the GTR + γ + I evolutionary model, implemented in the software package BEAST. Twelve different nodes were selected for the calibration (green dots). Blue bars indicate the 95% highest posterior density (HPD) intervals of the posterior probability distribution of node ages.

The ancestral state reconstruction analyses (Fig. 6) establish that the common ancestor to all living Nassellaria should be either multicyrtid or dicyrtid, with 1st or 2nd cephalic state (Table 2), a small apical horn, with a small or not projecting ventral ray and not projecting dorsal and lateral rays out of the skeleton. Soon after the diversification between clade A and the rest of the clades, the analysis establishes for the common ancestor of lineages II, III and IV, a dicyrtid state, with a 2nd cephalic state (Table 2), a small apical horn, and not projecting ventral, dorsal and lateral rays. The same morphology was found for the common ancestor of the lineage II, for that of lineages III and IV and that of lineage III. The dicyrtid state remains the most parsimonious state for the common ancestor of all families except for clades A and D and E. The cephalis has a 2nd state (Table 2) for every common ancestor of each clade except for clade A (1st state; Table 2), F (1st or 2nd; Table 2) and I-J (2nd or 3rd; Table 2). Regarding states of the apical ray, it remained small in every node with the exception of lineage IV, where it is medium/big. The ventral ray remained inside the skeleton with the exception of clade A, D, G

and I, where it is small. Finally, the dorsal and lateral rays remained inside the skeleton for the common ancestor of clades A, B, C, D and F, but quite different states for the common ancestor of the rest of the clades: clade E has either a not projecting or small dorsal & lateral rays, clade X is characterized by big/feet/wings, clade G has a small dorsal and lateral rays, clades H and J either not projecting out of the skeleton, small or medium dorsal and lateral rays and clade I has a big/feet/wings dorsal and lateral rays.

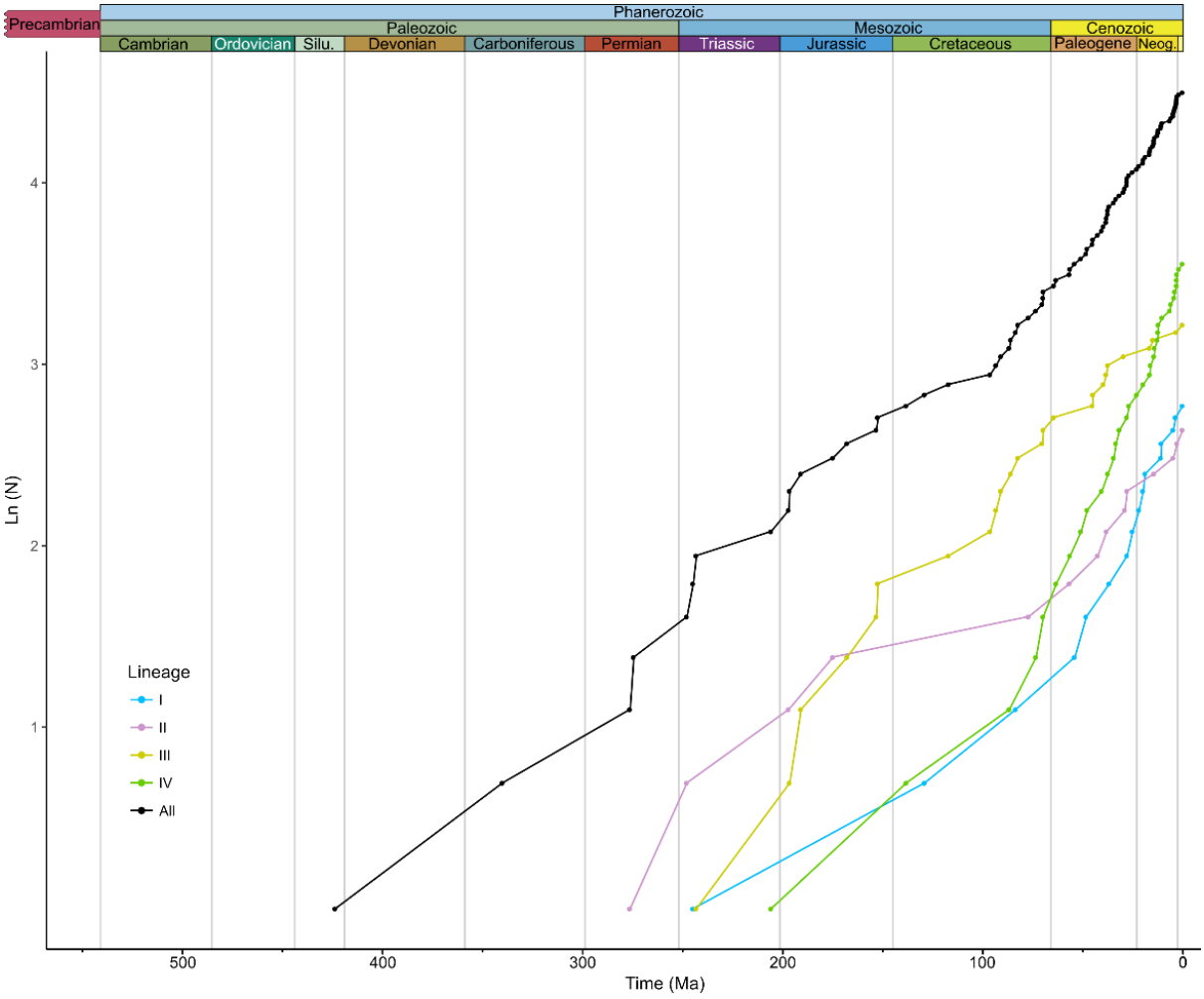


Figure 5. Lineages Through Time (LTT) analysis based on the molecular clock results for Nassellaria (removing the outgroup; in black), and of each lineage independently: lineage I (Blue), lineage II (purple), lineage III (yellow) and lineage IV (green). The y-axis represents the number of lineages (N) expressed in logarithmic (base e) scale (Ln(N)) and in the x-axis it is represented the time in million of years ago (Ma).

Environmental genetic diversity of Nassellaria

A total of 229 18S rDNA and two 28S rDNA environmental sequences affiliated to Nassellaria were retrieved from public database and placed in our reference phylogenetic tree (Fig. 7; Supplementary Material Table S3). Most of the environmental sequences (151) belonged to clade G, while 41 other sequences were scattered between clades A, D, E, F, I and J (2, 6, 26, 1, 2 and 4 sequences respectively). The rest of the sequences could not be placed within any existing clade. From those, 28 were closely related to clade E, and the others at basal nodes mainly over lineage II. These 37 sequences were then included in a phylogenetic

tree (with a GTR+G+I model, 4 invariant sites and 1000 bootstraps) and 3 new and highly supported clades appeared (Supplementary Material Fig. S3). The first clade, constituted by 31 sequences, was assigned to Collodaria in Biard et al. (2015). A second clade formed by 2 sequences (BS: 94) was highly related to clades B and C (BS: 96). And the last clade, constituted as well of 2 sequences (BS: 100) was basal to the node formed by clades C, D and the previous mentioned environmental clade (BS: 58). The 2 remaining sequences were highly related to clades I (BS: 84) and to the node comprising clades I and J (BS: 100). Finally, 2 sequences were aligned within clade D and the last sequence was closely related to clade B, C and D. Regarding the sequences blasted for the partial 28S matrix, 2 sequences were extracted and mapped within clades E and G respectively.

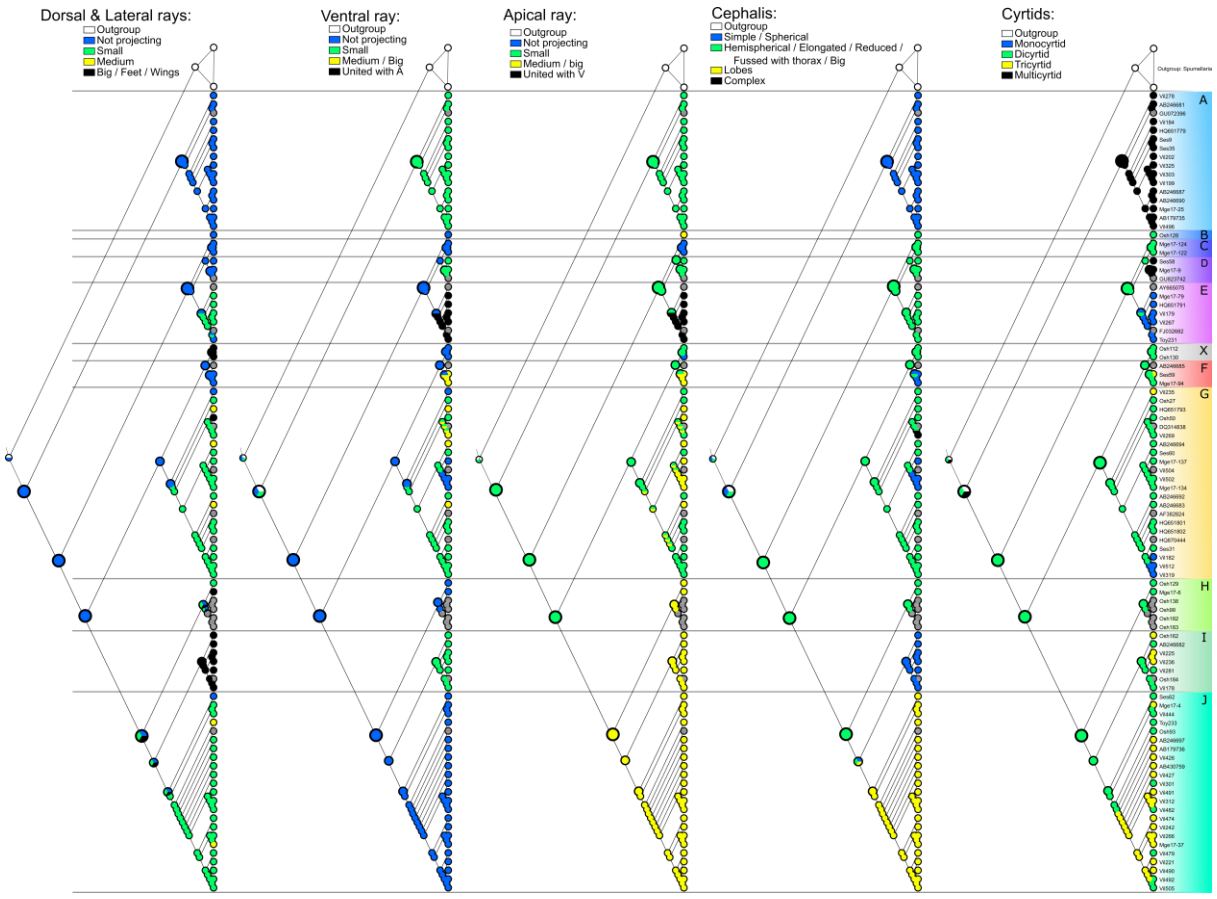


Figure 6. Parsimony Ancestral State Reconstruction analysis based in the resulting phylogenetic tree for the 5 characters chosen. Relevant nodes are increased for clarity.

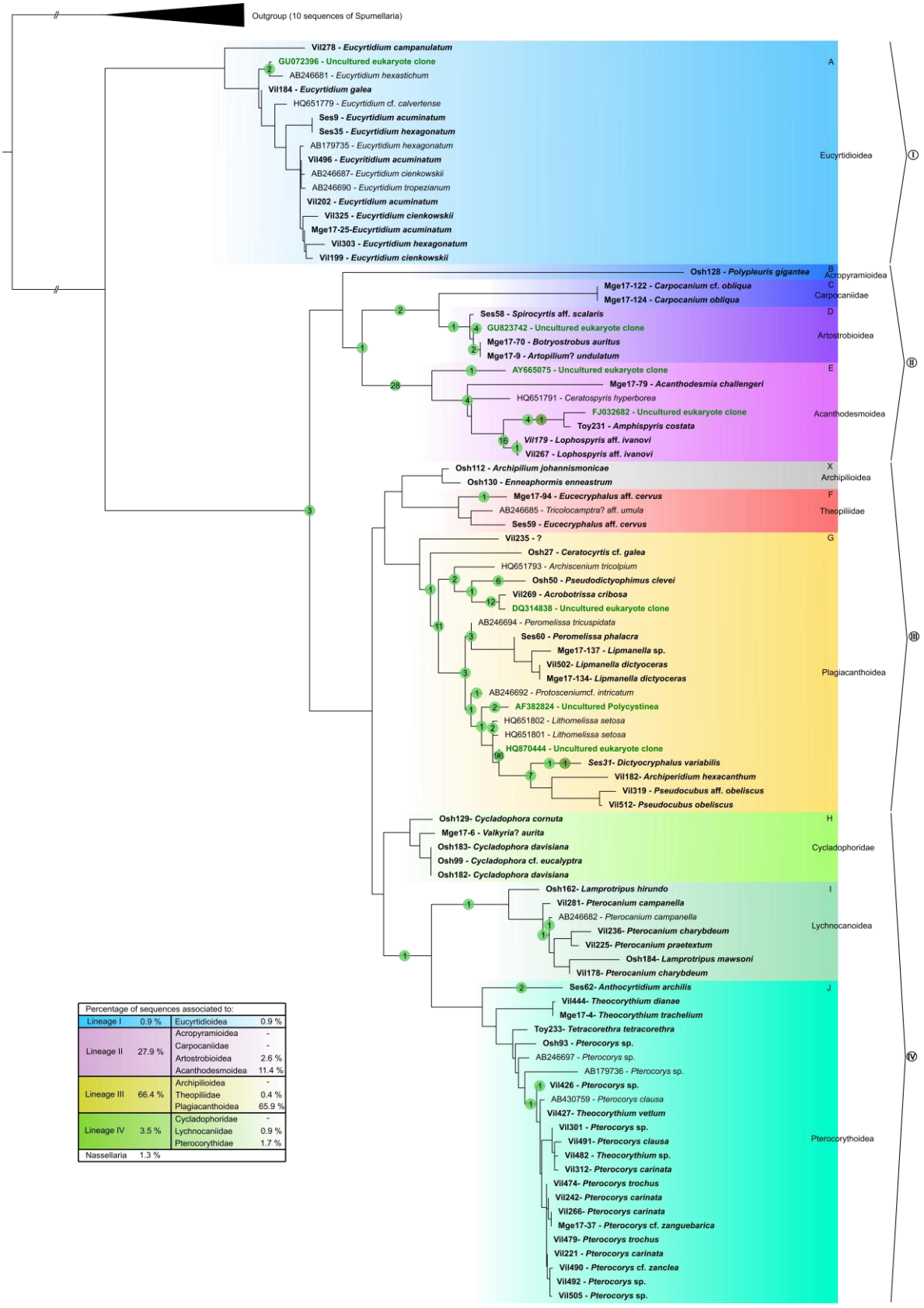


Figure 7. Pplacer phylogenetic placement of 229 environmental sequences for the 18S rDNA gene and 2 environmental sequences for the 28S rDNA gene into a concatenated phylogenetic tree of Nassellaria (complete 18S + partial 28S rDNA genes). Numbers at nodes represent the amount of environmental sequences assigned to a branch or a node; in light green for the 18S rDNA gene and in dark green for the 28S rDNA gene.

Discussion

Morpho-molecular classification of living Nassellaria

In our phylogenetic classification the overall morphology of Nassellaria, rather than the initial spicular system, is the most accurate feature to differentiate clades at the Superfamily level (Supplementary Material Fig. S1). We could not find any pattern in the initial spicular system that enables to separate clades, suggesting that the complexity of the initial spicular system is not related to Nassellaria phylogeny (e.g., Acanthodesmoidea, Plagiacanthoidea). Yet, this morphological character has been used to discriminate families and higher taxonomical levels (Petrushevskaya, 1971b). As well, the arches connecting these spicules were used in previous classifications and, in some cases, to discern at genus, and partly family levels (Petrushevskaya, 1971b; De Wever et al., 2001). In addition to our results, it has never been hypothesized the evolution of the initial spicular system complexity through the radiolarian fossil record. The use of the overall morphology in Nassellaria classification makes recognition of live cells easier under a light microscope facilitating morphology based ecological studies.

The extent Nassellaria included in our study can be divided in 11 morpho-molecular clades, grouped in four main evolutionary lineages based on phylogenetic clustering support, common morphological features and molecular dating: Eucyrtidioidea in lineage I; Acropyramioidea, Carpocaniidae, Artostrobioidea and Acanthodesmoidea in lineage II; Archipilioidea, Theopiliidae and Plagiacanthoidea in lineage III, and *Cycladophora*, Lychnocanoidea and Pterocorythoidea in lineage IV. The unified morpho-molecular framework here proposed reveals a partial agreement between the traditional taxonomy and the molecular classification. Such discrepancies have already been reported not only in other Radiolaria groups (Decelle et al., 2012b; Biard et al., 2015) but as well in other SAR taxa such as Foraminifera (Pawlowski and Holzmann, 2002), Phaeodaria (Nakamura et al., 2015) or Tintinnida (Bachy et al., 2012), being a common issue in protists classification (Schlegel and Meisterfeld, 2003; Caron, 2013).

Our revised morpho-molecular classification confirms the monophyly of the ancient Superfamilies Eucyrtidioidea and Acropyramioidea, represented by the clades A and B respectively, as well as the undetermined Family Carpocaniidae, the Superfamilies Artostrobioidea and Acanthodesmoidea, (clades C, D and E, respectively). The Superfamily Plagiacanthoidea is however paraphyletic, appearing in two different clades (X and G). Similarly, previously proposed families are scattered within clade G or even including the Superfamily Cannobotryoidea. Clade F and H both display an identical architecture of the initial spicular system, yet they are clustered based in the overall morphology in two different clades, the family here proposed as Theopiliidae (clade F) and *Cycladophora*-like specimens (clade H). As the genus *Eucecryphalus* (clade F) is the genus of the family Theopiliidae (see Matsuzaki et al., 2015), clade F holds the name Theopiliidae. Regarding clade H, a new family Cycladophoridae is established herein as defined in the taxonomic note. Regarding clades I and J, Matsuzaki et al. (2015) include them within the same Superfamily (Pterocorythoidea) due to the strong phylogenetic relationship reported by Krabberød et al. (2011). Yet due to the evolutionary patterns and the phylogenetic distance, these two clades should be

considered as two different superfamilies, Pterocorythoidea and Lychnocanoidea (Lychnocanoidea Haeckel, 1882, sensu Kozur and Mostler, 1984). In our study the undetermined family Bekomidae was scattered within clade I, showing no phylogenetic differences. Re-examination of the cephalic structure in *Lamprotripus* concludes the assignation of the genus to the Superfamily Lychnocanoidea, yet intergeneric morphological differences remain still elusive. Further molecular analyses must reveal these discrepancies between morphological and molecular classification.

Evolutionary history of Nassellaria

The morphological evolution of Nassellaria is marked by their dubious appearance in the fossil record, whether it happened with the primitive nassellarian forms (in the Upper Devonian) or with the first multi-segmented nassellarians (in the Early Triassic) (De Wever et al., 2001; Suzuki and Oba, 2015). Our results showed that the first diversification of Nassellaria agrees in both time (~423 Ma; 95% HPD: 500-342 Ma; Figs. 4, 5) and reconstructed morphology (two- or multi-segmented last common ancestor) with the primitive nassellarians. Thus, the most likely scenario, proposed by Petrushevskaya (1971a) and continued by Cheng (1986), is where Nassellaria originated during the Devonian from primitive radiolarian forms and not during the Triassic. This hypothesis can explain the sudden appearance of forms in the Middle Triassic already pointed out by De Wever et al. (2003) and confirmed by our results.

The oldest known nassellarian like fossils are Popofskyellidae and Archocyrtiidae from the Devonian (Cheng, 1986). The former family shares with lineage I a multisegmented nature of the skeleton whereas the Archocyrtiidae has a unique and large cephalic segment and three long feet, derived characteristics that can be found over lineages II, III and IV. Such similarities in the morphology and the accordance of the molecular clock in the branching times with the fossil record allow us to hypothesis these two families as possible ancestors of living Nassellarians; as already suggested by Cheng (1986). Thereafter, the evolutionary connection of Popofskyellidae and lineage I is debatable, due to a fragmented fossil record along the Permian (Isakova and Nazarov, 1986; De Wever et al., 2003). Regarding lineage II, it is likely that diverged from Archocyrtiidae by reduction of the feet, though phylogenetic relationships within this lineage remain still unclear. A second morphological modification might have happened in the separation of the cephalis and thorax and the appearance of lineage IV (probable candidate is the Ultranaporidae). Therefore, Archocyrtiidae could be a direct ancestor of lineage III due the strong similarities within the members of this lineage and the Palaeozoic Archocyrtiidae.

Despite the early appearance of the Order Nassellaria, living nassellarian groups diversified ca. 250 Ma after the third, and biggest of all, mass extinction (Sepkoski, 1981; Bambach et al., 2004). This bottle neck led to a rapid increase in the global marine diversity (Twitchett et al., 2004), where all the surviving and isolated populations followed different pathways; an expected process in the aftermath of an evolutionary crisis (Twitchett and Barras, 2004; Hull, 2015). The lineage of Acropyramioidea/Acanthodesmoidea (lineage II) splits apart into many different forms that will led to new families, lineages III and IV diverge from each other followed by a fast branching of the plagiacanthoids, while eucyrtidioids

(lineage I) slowly diversifies. By the end of the Late Triassic-Early Jurassic, radiolarian fossil record reaches its highest diversity measures (De Wever et al., 2006). Along with Nassellaria, other groups of marine protists also diversified such as dinoflagellates (Fensome et al., 1996; Janouškovec et al., 2017) or even appeared like diatoms (Kooistra and Medlin, 1996; Sims et al., 2006).

The second part of the Mesozoic is characterized by a global oceanic anoxic event happening at the end of the early Jurassic (Jenkyns, 1998) and a widespread series of Oceanic Anoxic Events (known as OAEs) during the Cretaceous (Jenkyns, 2010; Schalanger and Jenkyns, 1976). These events have been proposed as the major responsible force for the appearance of planktic Foraminifera during the Jurassic (Hart et al., 2003), and for extinction-speciation processes in planktonic evolution during the Cretaceous, especially for Foraminifera and Radiolaria (Leckie et al., 2002). During the same period, we can find the rise of the Artostrobioidea, the divergence of the superfamilies Pterocorythoidea and Lychnocanoidea or the continuous diversification of the lineage III. After the OAEs and during the Early Cenozoic populations isolated from each other and probably thrived by the favourable conditions, start diversifying into the most recent families (Acanthodesmoidea, Theopiliidae, Cycladophoridae fam. nov., Lychnocanoidea, Pterocorythoidea and probably Archipilioidea). During the Cenozoic, other groups were also recovered from small populations linked to climate oscillations such as Foraminifera (Hallock et al., 1991) or calcareous Coccolithophores (Bown et al., 2004). Finally, during the late Cenozoic, the most recent family (Carpocaniidae) and groups (i.e., *Lipmanella* clade within Plagiacanthoidea) appear or diversify probably just before the opening of the Tasmanian gateway between Antarctica and Australia to form Antarctic Circumpolar Current.

Environmental genetic diversity of Nassellaria

Little is known about the diversity and ecology of contemporary Nassellaria. Despite recent studies using plankton net data publicly available (Boltovskoy et al., 2010; Boltovskoy and Correa, 2016a; Boltovskoy, 2017), most of the knowledge on this taxa is inferred from their extensive fossil record worldwide (Petrushevskaya, 1971a; De Wever et al., 2001). The present morpho-molecular framework proposes a morphological classification based on the overall structure, a feature easier to access compared to internal spicular structures, and therefore facilitating morphology-based ecological studies. Our framework also establishes a reference for sequences-based environmental diversity surveys and molecular ecology studies, allowing the accurate and fast taxonomic assignment of metabarcoding data.

Here we considered publicly available environmental sequences closely related to Nassellaria, from various environment (e.g. Edgcomb et al., 2011; Orsi et al., 2012; Lie et al., 2014) into our morpho-molecular data frame (Fig. 7). Most of environmental sequences clustered within the Superfamily Plagiacanthoidea, revealing an underrepresentation of this taxon in our study despite the number of sequences included (23%). This clade can be found at high relative abundances all year round (Motoyama et al., 2005; Ikenoue et al., 2015), in every latitude (Boltovskoy et al., 2010) and at every depth (Boltovskoy, 2017). Most of the known plagiacantoid species are very small related to the average nassellarian size, so they could go through the plankton nets or be overlooked during isolation. Also,

Acanthodesmoidea and Artostrobioidea are two clades highly represented in environmental sequences compared to the number of morphologically described sequences included in our study. The former superfamily is represented by sequences coming mainly from the subtropical and tropical South China Sea where Radiolaria are very abundant (Wu et al., 2014). The Artostrobioidea is mainly characterised by deep environmental sequences, an environment poorly represented in our phylogenies mainly due to the difficulties in the DNA amplification of nassellarian specimens habiting the deep ocean. Likewise, the Superfamily Acropyramioidea, restricted to the deep ocean, is represented by only one sequence of the 28S rDNA gene. Therefore, due to the lack of 18S rDNA marker for this clade, it is likely that some of the sequences clustered within this lineage actually belong to the Acropyramioidea, since most of them come from deep environments (e.g., Kim et al., 2012; Lie et al., 2014; Xu et al., 2017). Increasing molecular coverage of the reference sequence database, through exploration of a variety of ecosystems, will be critical to properly decipher placement of environmental sequences. Our integrated approach enables providing a robust evolutionary history and classification of Nassellaria, and allows future ecological and diversity studies based in both a morphology or through a metabarcoding approach.

Taxonomic Notes

Following the molecular phylogenetic results definition of superfamilies is revised and a new family is established here following ICZN2000.

Superfamily **Eucyrtidioidea** Ehrenberg, 1846 sensu emend here by Suzuki.

Type genus. *Eucyrtidium* Ehrenberg, 1846 (type species: *Lithocampe acuminata* Ehrenberg).

Emended definition. A spherical cephalis with a sharply constricted basal aperture and many segmentations which are generally divided by inner rings. No feet.

Remarks. This Superfamily corresponds to clade A. Although many suggestions about the definition have been made so long (see Matsuzaki et al., 2015), these two points are only different points from other nassellarians.

Superfamily **Archipilioidea** Haeckel, 1882 stat. nov. by Sandín, Not and Suzuki.

Type genus. *Archipilium* Haeckel, 1882 (type species: *Archipilium orthopterum* Haeckel).

Definition. Practically single segment, although the upper and lower parts can be recognized by the position of *MB*. Initial spicular system is characterized by a very short *MB* (or missing) to form a three-pointed star rod system and a significant circular frame. Three feet or relevant structure is present.

Remarks. This corresponds with our Clade X. The figured specimens of the type species of *Archipilium* apparently lack a three-pointed star rod system and “a significant circular frame” (PR in Figure S1). The intermediate form between *Archipilium* and *Enneaphormis* was described as “*Enneaphormis trippula*” by Renaudie and Lazarus (2015).

Family **Archipilidiae** Haeckel, 1882.

Type genus. *Archipilium* Haeckel, 1882 (type species: *Archipilium orthopterum* Haeckel).

Remarks. Here we include genera *Archibursa* Haeckel, 1882 (type species: *Archibursa tripodiscus* Haeckel), *Chitascenium* Sugiyama, 1994 (type species: *Chitascenium cranites* Sugiyama), and *Enneaphormis* Haeckel, 1882 (type species: *Sethophormis* (*Enneaphormis*) *rotula* Haeckel), because they have a three-pointed star rod system and a significant circular frame (PR).

Superfamily **Plagiacanthoidea** Hertwig, 1879 emended here by Sandín and Suzuki.

Type genus. *Plagiacantha* Claparede and Lachmann, 1859 (type species: *Acanthometra arachnoides* Claparède).

Emended definition. Test can be divided into upper and lower parts by the level of *MB* or the neck of test. Test is small, and the initial spicular system contains variable types of arches. By these arches, antecephalic lobe, eucephalic lobe and postcephalic lobes are formed. This Superfamily also includes nassellarians made of bony frame only.

Remarks. This definition includes both Plagiacanthoidea and Cannobotryoidea. The name “Plagiacanthoidea” has a priority to Cannobotryoidea.

Family **Cycladophoridae** fam. nov. Suzuki

Type genus. *Cycladophora* Ehrenberg, 1846 (type species: *Cycladophora davisiana* Ehrenberg)

Definition. Initial spicule with *A*-, *V*-, *D*- and two *L*-rods, and an *MB*. No *PR* and No tubular cephalis horn. Test robust, helmet-conical, consisting of two segments with or without frill-like fringe. Cephalis small, spherical, and pore-less or relict pores; distinctive from thorax; three wing-like rod or rims on upper thoracic wall.

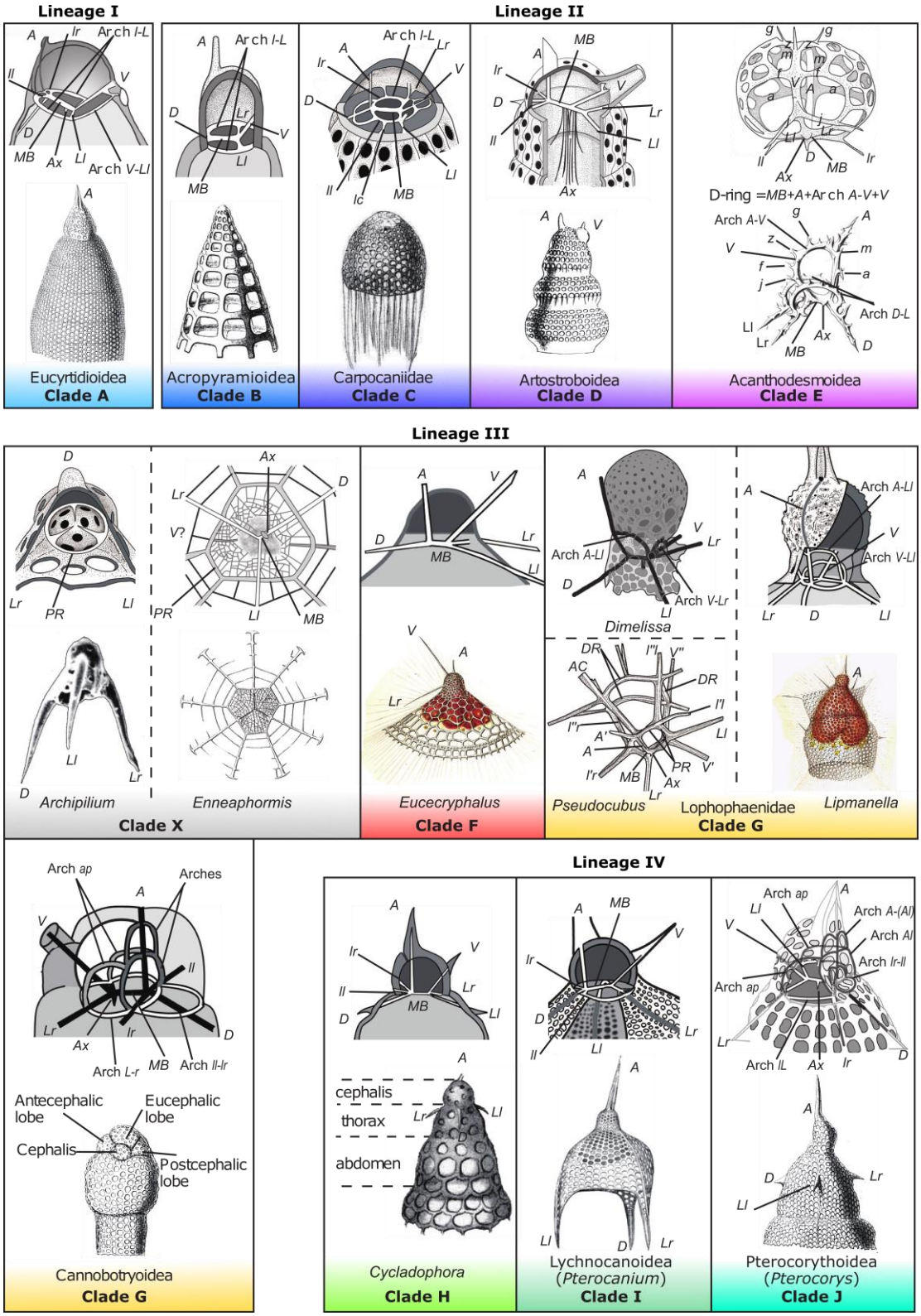
Remarks. This new family is newly proposed to separate genus *Eucecryphalus*, the type genus of the family Theopiliidae, from *Cycladophora* which was used to belong to the Theopiliidae. More study needs to determine differences between Cycladophoridae fam. nov. and Theopiliidae. The former’s test is robust whereas the latter’s one is fragile. Superfamily is not determined to Cycladophoridae because nothing is known about the phylogenetic relationship with the Mesozoic family Neosciadiocapsidae. Cycladophoridae fam. nov. differs from the Superfamily Lychnocanoidea in that the latter has three distinctive feet which are connected through thoracic rims from *A*-, *V*-, and two *L*-rods. Cycladophoraidae fam nov. is easily distinguished from the Superfamily Pterocorythoidea in having a pterocorythid cephalic structure with special lobes.

Acknowledgments

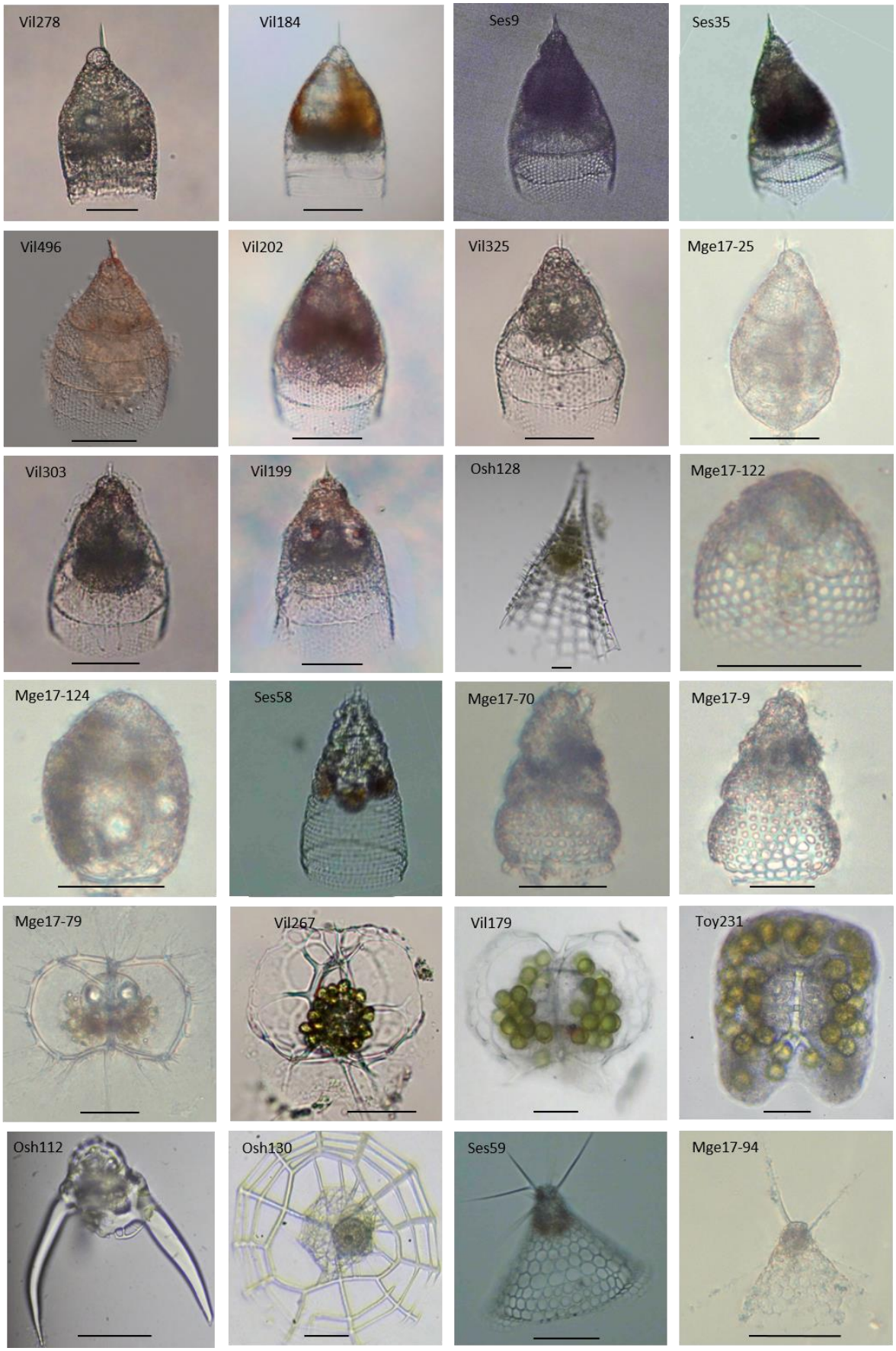
This work was supported by the IMPEKAB ANR 15-CE02-0011 grant and the Brittany Region ARED C16 1520A01, the Japan Society for Promotion of Science KAKENHI Grand No. K16K0-74750 for N. Suzuki and “the Cooperative Research Project with the Japan Science and Technology Agency (JST) and Centre National de la Recherche Scientifique (CNRS, France) “Morpho-molecular Diversity Assessment of Ecologically, Evolutionary, and Geologically Relevant Marine Plankton (Radiolaria)”. We are grateful to T/S Oshoro-maru (Hokkaido Univ.), T/S Toyoshio-maru (Hiroshima Univ.) and their directors, Susumu Ohtsuka (Hiroshima Univ.) and Atsushi Yamaguchi (Hokkaido Univ.), as well as Akihiro Tuji, Yasuhide Nakamura and

Yoshiaki Aita for sampling in Oshoro-Maru. We would like to thank the MOOSE cruise and program for the opportunity of sampling and the facilities given onboard, as well as John Dolan for hosting us multiple times at the Laboratoire d'Océanographie of Villefranche sur Mer. We thank Sebastien Colin for the Confocal Microscope imaging. We are also very grateful to Peter Baumgartner for his valuable comments in geological events, and specially to Luis O'dogherty, Spela Gorican and Taniel Danelian for sharing their knowledge in the fossil record and improving the calibration of the molecular clock.

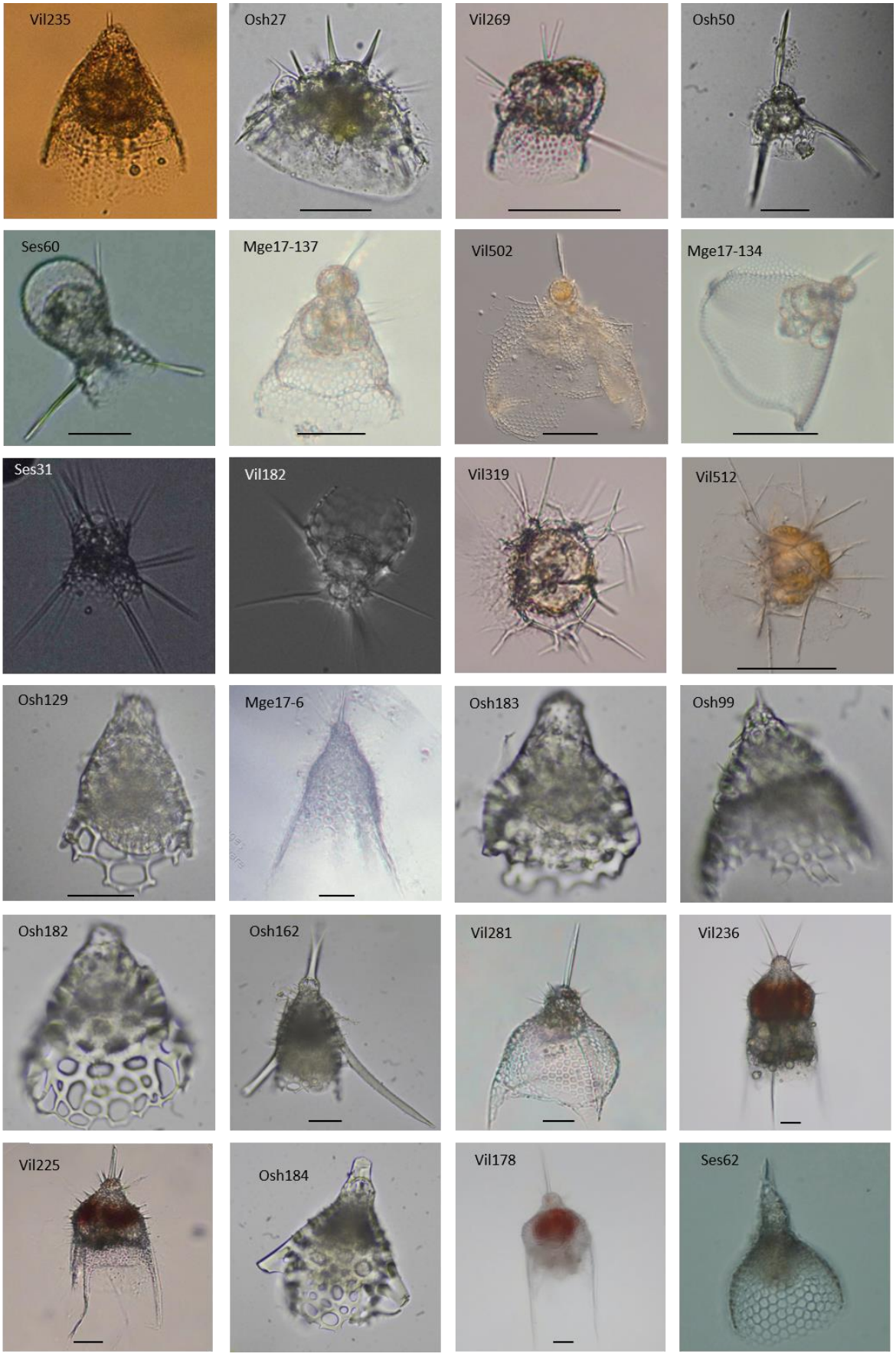
Supplementary Material



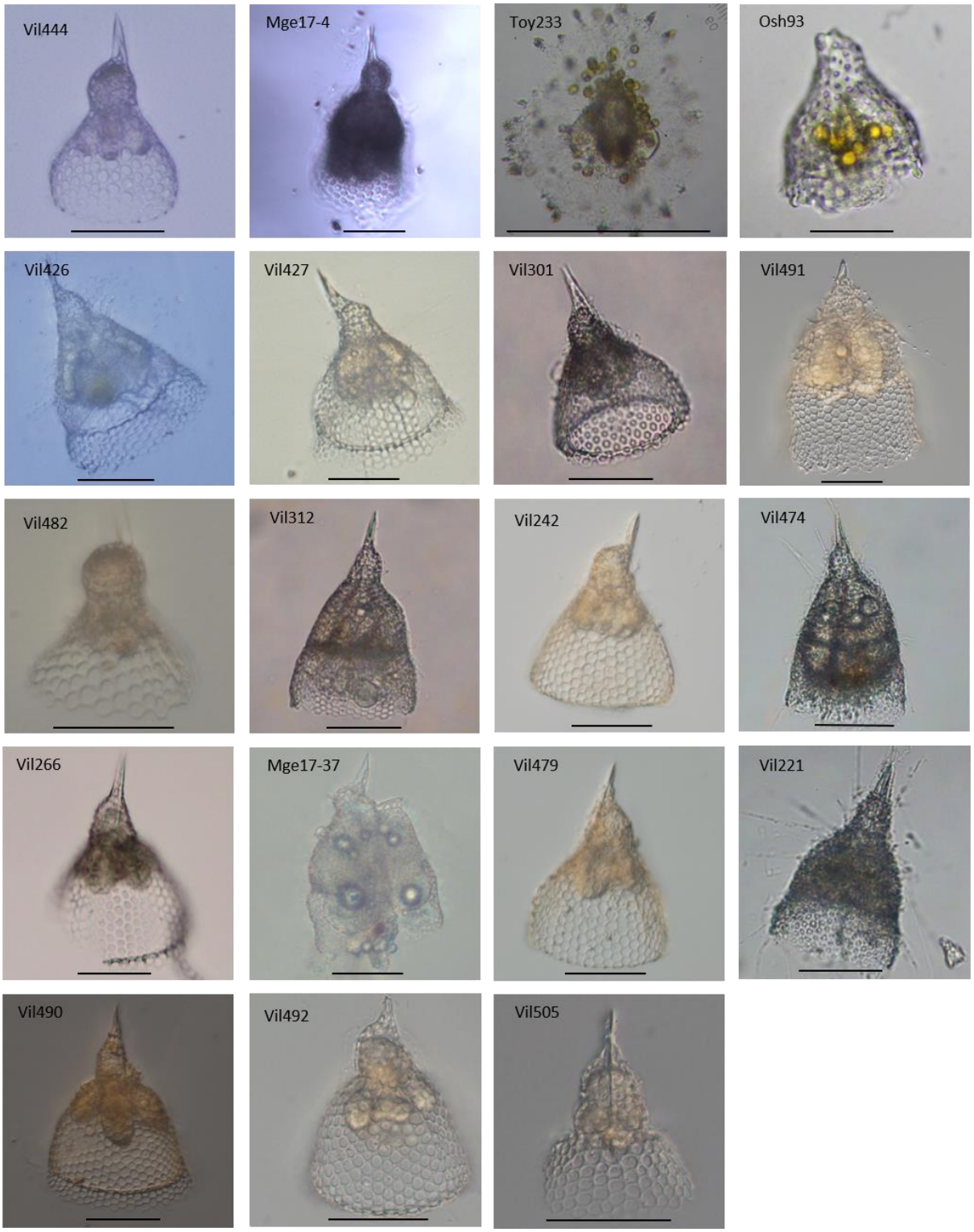
Supplementary Figure S1. A comparison between the detailed internal spicular system (upper row of each box) and the general skeleton morphology (lower row of each box) for a representative genus of each phylogenetic clade (different boxes) from Fig. 1. Each component of the internal spicular systems is coded by alphabetical letters which are not sourced from any abbreviated words. Overall images came from variable sources archived in Biodiversity Heritage Library (https://www.biodiversitylibrary.org/).



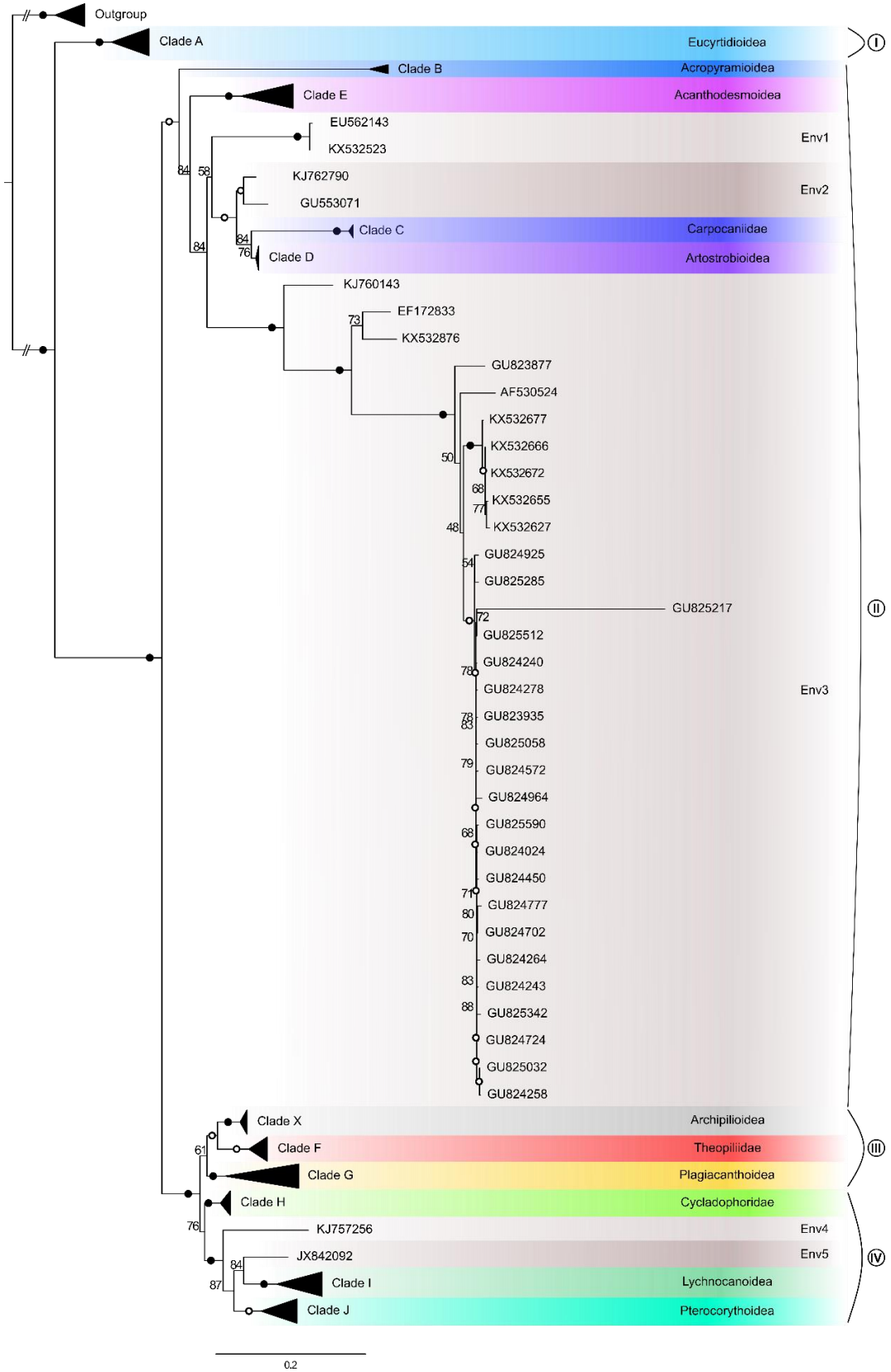
Supplementary Figure S2. Light Microscopy (LM) images of live Nassellarian specimens used in this study for phylogenetic analysis in Fig. 1. Scale bars (when available) = 50µm. Part 1/3



Supplementary Figure S2. Light Microscopy (LM) images of live Nassellarian specimens used in this study for phylogenetic analysis in Fig. 1. Scale bars (when available) = 50µm. Part 2/3



Supplementary Figure S2. Light Microscopy (LM) images of live Nassellarian specimens used in this study for phylogenetic analysis in Fig. 1. Scale bars (when available) = 50µm. Part 3/3



Supplementary Figure S2. Molecular phylogeny of environmental sequences (18S rDNA) associated to Nassellaria. The tree was obtained by using a phylogenetic Maximum likelihood method implemented using the GTR + γ + I model of sequence evolution. PhyML bootstrap values (1000 replicates, BS) are shown at the nodes. Black circles indicate BS of 100% and hollow circles indicates BS > 90%. Branches with a double barred symbol are fourfold reduced for clarity.

Supplementary Table S1. Taxonomic authority of specimens used to obtain nassellarian phylogeny.

Type	This paper	Superfamily (mainly from Matsuzaki et al., 2015 and Suzuki and Not, 2015)	Family (mainly from Matsuzaki et al. 2015, and subsequently De Wever et al. 2001)	Species name	Note
I	Eucyrtidioidea	Eucyrtidioidea	Eucyrtidiidae	*Eucyrtidium campanulatum* (Ehrenberg)	
				Eucyrtidium hexastichum (Haeckel)	
				Eucyrtidium galea (Haeckel)	
				Eucyrtidium af. *calvertense* (Martin)	
				Eucyrtidium acuminatum (Ehrenberg)	Type of the genus
				Eucyrtidium hexagonatum (Haeckel)	
				Eucyrtidium cienkowskii (Haeckel)	
				Eucyrtidium tropezianum (Müller)	
II	Acropyramoidea	Acropyramoidea	Acropyramiidae	*Polypleuris gigantea* (Haecker)	
	Carpocanioidea	Incertae superfamilia	Carpocaniidae	*Carpocanium obliqua* (Haeckel)	
	Artostrobooidea	Artostroboioidea	Artostroboiidae	*Spirocyrtis* aff. *scalaris* (Haeckel)	Type of the genus
				Botryostrobus auritus (Ehrenberg)	
		Incertae superfamilia	Incertae familia	*Extotoxon undulatum* (Popofsky)	
	Acanthodesmoidea	Acanthodesmoidea	Acanthodesmiidae	*Acanthodesmia challengeri* (Haeckel)	
			Triospyrididae	*Ceratospyris hyperborea* (Jörgensen)	
				Lophospyris aff. *ivanovi* (Petrushevskaya)	
			Acanthodesmiidae	*Amphispyris costata* (Haeckel)	
III	Archipillioidea stat. nov.	---	Archipillidae	*Archipilium johannismonicae* (Deflandre)	
		---	Sethophormidae?	*Enneaphormis enneastrum* (Haeckel)	
	Theopillioidea	Incertae superfamilia	Theopilliidae	*Eucecryphalus* aff. *cervus* (Ehrenberg)	
		Eucyrtidioidea	Eucyrtidiidae	*Tricolocamptra* aff. *umula* (Haeckel)	
	Plagiacanthoidea	Plagiacanthoidea	Lophophaenidae	*Ceratocyrtis* cf. *galea* (Cleve)	
			Sethoperidae	*Archiscenium tricolpium* (Haeckel)	
		Cannobotrioidea	Cannobotryidae	*Acrobotrissa cribosa* Popofsky	Type of the genus
		Plagiacanthoidea	Lophophaenidae	*Pseudodictyophimus clevei* (Jörgensen)	
				Peromelissa tricuspidata (Popofsky)	
				Peromelissa phalacra Haeckel	Type of the genus
				Lipmanella dictyoceras (Haeckel)	Type of the genus
			Sethophormidae	*Protoscenium* cf. *intricatum* (Cleve)	
			Lophophaenidae	*Lithomelissa setosa* Jörgensen	
				Dictyocryphalus variabilis (Popofsky)	
				Archiperidium hexacanthum (Popofsky)	
				Pseudocubus obeliscus Haeckel	
IV	Cycladophorioidea	Incertae superfamilia	Theopilliidae	*Cycladophora cornuta* (Bailey)	
		---	---	*Valkyria*? *aurita* (Nigrini and Caulet)	No taxanomic reliability at the genus level and higher
		Incertae superfamilia	Theopilliidae	*Cycladophora davisiana* Ehrenberg	Type of the genus
				Cycladophora cf. *eucalyptra* (Haeckel)	
	Lychnocanoidea	Incertae superfamilia	Bekomidae	*Lamprotripus hirundo* (Haeckel)	
		Lychnocanoidea	Lychnocaniidae	*Pterocanium campanella* (Ehrenberg)	
				Pterocanium charybedum (Müller)	
				Pterocanium pratextum (Ehrenberg)	
		Incertae superfamilia	Bekomidae	*Lamprotripus mawsoni* (Riedel)	
	Pterocorythoidea	Pterocorythoidea	Pterocorythidae	*Anthocyrtidium archilis* (Haeckel)	
				Theocorythium dianae (Haeckel)	Type of the genus
				Tetracorethra tetracorethra (Haeckel)	Type of the genus
				Pterocorys clausa (Popofsky)	
				Theocorythium vetulum Nigrini	
				Pterocorys carinata (Haeckel)	
				Pterocorys trochus (Ehrenberg)	
				Pterocorys cf. *zanguebarica* (Ehrenberg)	
				Pterocorys cf. *zanclea* (Müller)	

Supplementary Table S2. List of specimens used to obtain nassellarian phylogeny. Abbreviations: WMS, Westearn Mediterranean Sea; VsM, Villefranche-sur-Mer (Mediterranean Sea); Ses, Sesoko, Okinawa (Japan); DCM, Deep Chlorophyl Maxium; OMZ, Oxygen Minimum Zone; AWC, Anoxic Water column

ID	Clade	Species	Location	Depth	Reference	Accession number 18S	Accession number 28S
AB179735	A	Eucyrtidium hexagonatum	NA	NA	Yuasa et al. (2005)	AB179735	NA
AB179736	J	Pterocorys sp.	NA	NA	Yuasa et al. (2005)	AB179736	NA
AB246681	A	Eucyrtidium hexastichum	NA	NA	Kunitomo et al. (2006)	AB246681	NA
AB246682	I	Pterocanium campanella	NA	NA	Kunitomo et al. (2006)	AB246682	NA
AB246685	F	Tricolocamptra? aff. umula	NA	NA	Kunitomo et al. (2006)	AB246685	NA
AB246687	A	Eucyrtidium cienkowskii	NA	NA	Kunitomo et al. (2006)	AB246687	NA
AB246690	A	Eucyrtidium tropezianum	NA	NA	Kunitomo et al. (2006)	AB246690	NA
AB246692	G	Protoscenium cf. intricatum	NA	NA	Kunitomo et al. (2006)	AB246692	NA
AB246694	G	Peromelissa tricuspidata	NA	NA	Kunitomo et al. (2006)	AB246694	NA
AB246697	J	Pterocorys sp.	NA	NA	Kunitomo et al. (2006)	AB246697	NA
AB430759	J	Pterocorys clausa	Japan	NA	Ando et al. (2009)	AB430759	AB430759
AF382824	G	Uncultured Polycystinea	NA	NA	Lopez-Garcia et al. (2002)	AF382824	NA
AY665075	E	Uncultured eukaryote clone	Sargasso Sea	DCM	Armbrust et al. (Unpub.)	AY665075	NA
DQ314838	G	Uncultured eukaryote clone	Arctic waters	NA	Lovejoy et al. (2006)	DQ314838	NA
FJ032682	E	Uncultured eukaryote clone	Marmara Sea	25-500m	Marande et al. (2009)	FJ032682	NA
GU072396	A	Uncultured eukaryote clone	Arabian Sea	OMZ	Jebaraj et al. (2009)	GU072396	NA
GU823742	D	Uncultured eukaryote clone	Cariaco Bassin	AWC	Edgcomb et al. (2011)	GU823742	NA
HQ651779	A	Eucyrtidium cf. calvertense	Norway	NA	Krabberod et al. (2011)	HQ651779	HQ651779
HQ651791	E	Ceratospyris hyperborea	Norway	NA	Krabberod et al. (2011)	HQ651791	NA
HQ651793	G	Archiscenium tricolpium	Norway	NA	Krabberod et al. (2011)	HQ651793	NA
HQ651801	G	Lithomelissa setosa	Norway	NA	Krabberod et al. (2011)	HQ651801	NA
HQ651802	G	Lithomelissa setosa	Norway	NA	Krabberod et al. (2011)	HQ651802	NA
HQ870444	G	Uncultured eukaryote clone	Vancouver	120m	Orsi et al. (2012)	HQ870444	NA
Mge17-122	C	Carpocanium cf. obliqua	WMS	0-500m	This study	MK396913	NA
Mge17-124	C	Carpocanium obliqua	WMS	0-500m	This study	MK396914	NA
Mge17-134	G	Lipmanella dictyoceras	WMS	0-500m	This study	NA	MK396951
Mge17-137	G	Lipmanella sp.	WMS	0-500m	This study	NA	MK396952
Mge17-25	A	Eucyrtidium acuminatum	WMS	0-500m	This study	MK396915	NA
Mge17-37	J	Pterocorys cf. zanguebarica	WMS	0-500m	This study	NA	MK396953
Mge17-4	J	Theocorythium trachelium	WMS	0-500m	This study	NA	MK396954
Mge17-6	H	Valkyria? aurita	WMS	0-500m	This study	NA	MK396955
Mge17-70	D	Botryostrobus auritus	WMS	0-500m	This study	MK396916	NA
Mge17-79	E	Acanthodesmia challengeri	WMS	0-500m	This study	NA	MK396956
Mge17-9	D	Extotoxon undulatum	WMS	0-500m	This study	MK396918	NA
Mge17-94	F	Eucecryphalus aff. cervus	WMS	0-500m	This study	MK396917	MK396957
Osh112	X	Archipilium johannismonicae	Japan	2000-3000m	This study	NA	MK396958
Osh128	B	Polypleuris gigantea	Japan	2000-3000m	This study	NA	MK396959
Osh129	H	Cycladophora cornuta	Japan	2000-3000m	This study	MK396919	MK396960
Osh130	X	Enneaphormis enneastrum	Japan	2000-3000m	This study	MK396920	MK396961
Osh162	I	Lamprotripus hirundo	Japan	2000-3000m	This study	MK396921	MK396962
Osh182	H	Cycladophora davisiana	Japan	2000-3000m	This study	NA	MK396963
Osh183	H	Cycladophora davisiana	Japan	2000-3000m	This study	NA	MK396964
Osh184	I	Lamprotripus mawsoni	Japan	2000-3000m	This study	NA	MK396965
Osh27	G	Ceratocyrtis cf. galea	Japan	NA	This study	NA	MK396966
Osh50	G	Pseudodictyophimus clevei	Japan	NA	This study	MK396922	MK396967
Osh93	J	Pterocorys sp.	Japan	750-1000m	This study	NA	MK396968
Osh99	H	Cycladophora cf. eucalyptra	Japan	2000-3000m	This study	NA	MK396969
Ses31	G	Dictyocryphalus variabilis	Ses	NA	This study	MK396923	MK396970
Ses35	A	Eucyrtidium hexagonatum	Ses	NA	This study	MK396924	NA
Ses58	D	Spirocyrtis aff. scalaris	Ses	NA	This study	MK396925	MK396971
Ses59	F	Eucecryphalus aff. cervus	Ses	NA	This study	MK396926	MK396972
Ses60	G	Peromelissa phalacra	Ses	NA	This study	MK396927	NA
Ses62	J	Anthocyrtidium archilis	Ses	NA	This study	MK396928	MK396973
Ses9	A	Eucyrtidium acuminatum	Ses	NA	This study	MK396929	NA
Toy231	E	Amphispyris costata	Okinawa	0-150m	This study	MK396930	MK396974
Toy233	J	Tetracorethra tetracorethra	Okinawa	0-150m	This study	NA	MK396975
Vil178	I	Pterocanium charybdeum	VsM	NA	This study	NA	MK396976
Vil179	E	Lophospyris aff. ivanovi	VsM	NA	This study	MK396931	MK396977
Vil182	G	Archiperidium hexacanthum	VsM	NA	This study	NA	MK396978
Vil184	A	Eucyrtidium galea	VsM	NA	This study	MK396932	NA
Vil199	A	Eucyrtidium cienkowskii	VsM	NA	This study	NA	MK396979
Vil202	A	Eucyrtidium acuminatum	VsM	NA	This study	NA	MK396980

Vil221	J	Pterocorys carinata	VsM	NA	This study	MK396933	MK396981
Vil225	I	Pterocanium praetextum	VsM	NA	This study	MK396934	MK396982
Vil235	G	?	VsM	NA	This study	NA	MK396983
Vil236	I	Pterocanium charybdeum	VsM	NA	This study	MK396935	MK396984
Vil242	J	Pterocorys carinata	VsM	NA	This study	NA	MK396985
Vil266	J	Pterocorys carinata	VsM	NA	This study	MK396936	MK396986
Vil267	E	Lophospyris aff. ivanovi	VsM	NA	This study	MK396937	MK396987
Vil269	G	Acrobotrissa cribosa	VsM	NA	This study	MK396938	NA
Vil278	A	Eucyrtidium campanulatum	VsM	NA	This study	MK396939	MK396988
Vil281	I	Pterocanium campanella	VsM	NA	This study	NA	MK396989
Vil301	J	Pterocorys sp.	VsM	NA	This study	MK396940	MK396990
Vil303	A	Eucyrtidium hexagonatum	VsM	NA	This study	NA	MK396991
Vil312	J	Pterocorys carinata	VsM	NA	This study	MK396941	MK396992
Vil319	G	Pseudocubus obeliscus	VsM	NA	This study	NA	MK396993
Vil325	A	Eucyrtidium cienkowskii	VsM	NA	This study	NA	MK396994
Vil426	J	Pterocorys sp.	VsM	Surface	This study	MK396942	MK396995
Vil427	J	Theocorythium vetlum	VsM	Surface	This study	MK396943	MK396996
Vil444	J	Theocorythium dianae	VsM	0-450m	This study	NA	MK396997
Vil474	J	Pterocorys trochus	VsM	Surface	This study	MK396944	NA
Vil479	J	Pterocorys trochus	VsM	Surface	This study	NA	MK396998
Vil482	J	Theocorythium sp.	VsM	Surface	This study	MK396945	NA
Vil490	J	Pterocorys cf. zanclea	VsM	0-30m	This study	MK396946	MK396999
Vil491	J	Pterocorys clausa	VsM	0-30m	This study	MK396947	MK397000
Vil492	J	Pterocorys sp.	VsM	0-30m	This study	NA	MK397001
Vil496	A	Eucyrtidium acuminatum	VsM	0-30m	This study	MK396948	MK397002
Vil502	G	Lipmanella dictyoceras	VsM	0-30m	This study	MK396949	MK397003
Vil505	J	Pterocorys sp.	VsM	0-450m	This study	MK396950	MK397004
Vil512	G	Pseudocubus obeliscus	VsM	0-450m	This study	NA	MK397005

Supplementary Table S3. List of environmental sequences related to Nassellaria. Abbreviations: OWC, oxygenated water column; mOWC, micro-oxic marine water column; AWC, Anoxic water column sample; OWC, Oxygenated water column sample; SCS, South China Sea; CS, Caribbean Sea

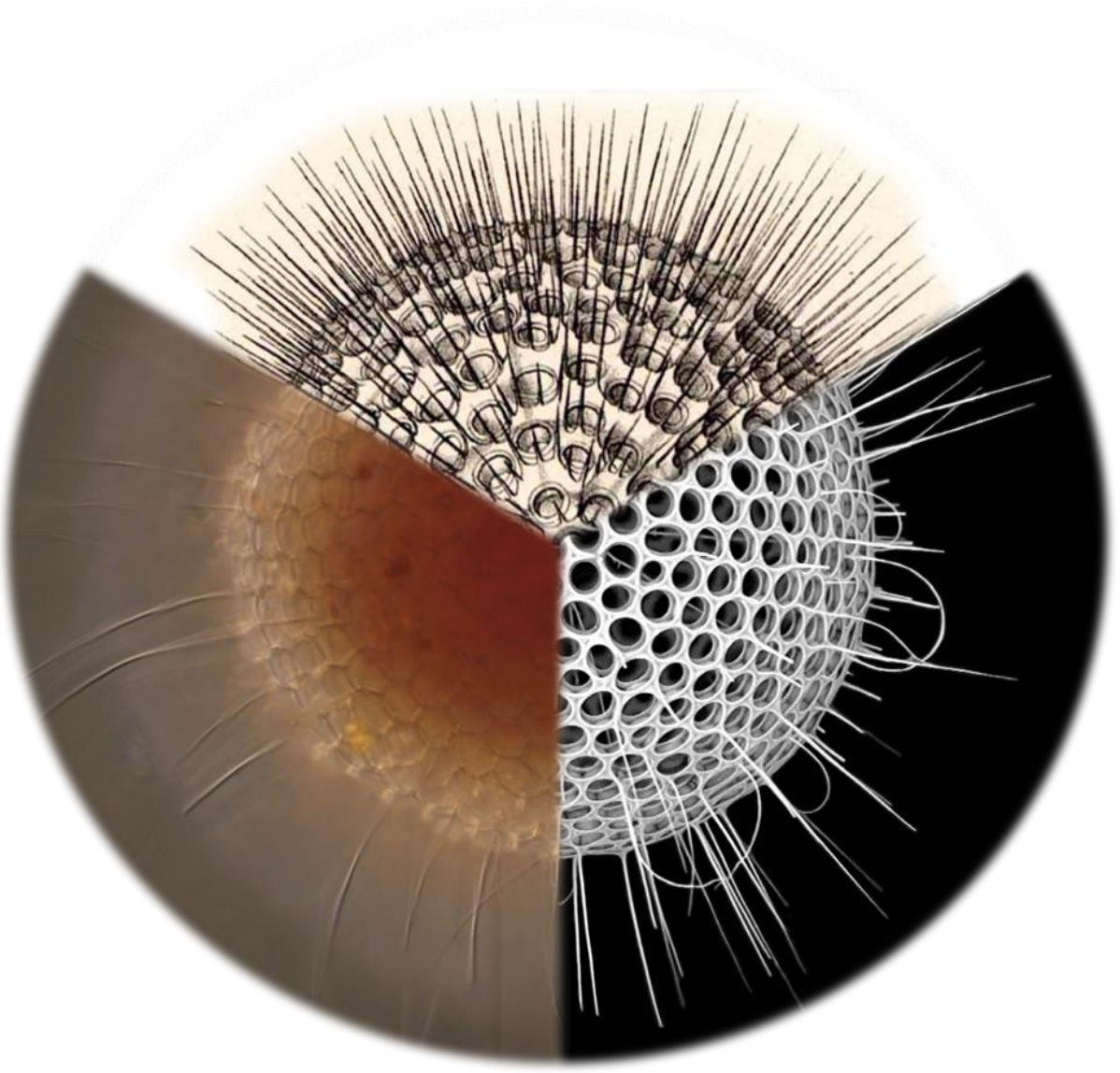
Accesion number	rDNA	Associated Linneage	Associated Clade	Location	Depth	Reference	Observations
AF530524	18S	II	E (basal)	Hydrothermal Mid Atlantic Ridge	-	Lopez-Garcia et al. (2003)	
AJ829823	18S	III	G	SCS, Nansha Islands	-	Yuan et al. (2004)	
DQ001518	18S	III	G	Mediterraneas sea	80m	Marie et al. (2006)	Picoplankton
DQ001519	18S	III	G	Mediterraneas sea	80m	Marie et al. (2006)	Picoplankton
DQ314834	18S	III	G	Arctic waters	-	Lovejoy et al. (2006)	
DQ314835	18S	III	G	Arctic waters	-	Lovejoy et al. (2006)	
DQ314837	18S	III	G	Arctic waters	-	Lovejoy et al. (2006)	
DQ386169	28S	III	G	Antarctic water	500 m	Moreira et al. (2007)	-
EF172833	18S	II	E (basal)	Sargasso Sea	500m	Not et al. (2007)	<2µm
EU333037	18S	III	G	SCS	-	Li et al. (2011)	
EU562100	18S	III	G	Indian Ocean	75m	Not et al. (2008)	
EU562143	18S	II	C-D (basal)	Indian Ocean	75m	Not et al. (2008)	
EU682659	18S	III	G	Arctic Sea: Beaufort Sea	-	Potvin & Lovejoy (2009)	
EU682660	18S	III	G	Arctic Sea: Beaufort Sea	-	Potvin & Lovejoy (2009)	
EU682661	18S	III	G	Arctic Sea: Beaufort Sea	-	Potvin & Lovejoy (2009)	
EU682662	18S	III	G	Arctic Sea: Beaufort Sea	-	Potvin & Lovejoy (2009)	
FJ032683	28S	II	E	Marmara sea	25 or 500m	Marande et al. (2009)	0.2-30µm
FN598293	18S	IV	J	South Pacific Ocean	1000m	Sauvadet et al. (2010)	0.2-3µm
GQ383024	18S	II	D	North west Pacific Ocean	-	Caron et al. (2009); Schnetzer et al. (2011)	
GQ383040	18S	II	E	North west Pacific Ocean	-	Caron et al. (2009); Schnetzer et al. (2011)	
GQ383100	18S	III	G	North west Pacific Ocean	-	Caron et al. (2009); Schnetzer et al. (2011)	
GU072381	18S	I	A	Arabian Sea water, OMZ	-	Jebaraj et al. (2009)	GPW2C8
GU072394	18S	I	A	Arabian Sea water, OMZ	-	Jebaraj et al. (2009)	GPW2E7
GU218927	18S	II	E	Eastern Guinea	5060-5066m	Scheckenbach et al. (2010)	sediment overlaying water
GU553067	18S	II	E	Nansha sea area, SCS	100m	Li et al. (2010)	NS51A256
GU553069	18S	II	E	Nansha sea area, SCS	100m	Li et al. (2010)	NS51C236
GU553070	18S	II	E	Nansha sea area, SCS	100m	Li et al. (2010)	NS51C240
GU553071	18S	II	C-D (basal)	Nansha sea area, SCS	100m	Li et al. (2010)	NS371B07
GU553072	18S	III	G	Nansha sea area, SCS	100m	Li et al. (2010)	NS371B41
GU553073	18S	II	E	Nansha sea area, SCS	100m	Li et al. (2010)	NS371C94
GU820848	18S	III	G	Cariaco Basin, CS	-	Edgcomb et al. (2011)	OWC
GU821228	18S	III	G	Cariaco Basin, CS	-	Edgcomb et al. (2011)	OWC
GU821503	18S	III	G	Cariaco Basin, CS	-	Edgcomb et al. (2011)	OWC
GU821560	18S	III	G	Cariaco Basin, CS	-	Edgcomb et al. (2011)	OWC
GU821868	18S	III	G	Cariaco Basin, CS	-	Edgcomb et al. (2011)	OWC
GU822007	18S	III	G	Cariaco Basin, CS	-	Edgcomb et al. (2011)	OWC
GU822146	18S	III	G	Cariaco Basin, CS	-	Edgcomb et al. (2011)	OWC
GU822482	18S	III	G	Cariaco Basin, CS	-	Edgcomb et al. (2011)	AWC
GU822745	18S	III	G	Cariaco Basin, CS	-	Edgcomb et al. (2011)	AWC
GU822871	18S	III	G	Cariaco Basin, CS	-	Edgcomb et al. (2011)	OWC
GU823751	18S	II	D	Cariaco Basin, CS	-	Edgcomb et al. (2011)	AWC
GU823877	18S	II	E (basal)	Cariaco Basin, CS	-	Edgcomb et al. (2011)	AWC
GU823935	18S	II	E (basal)	Cariaco Basin, CS	-	Edgcomb et al. (2011)	AWC
GU823990	18S	III	G	Cariaco Basin, CS	-	Edgcomb et al. (2011)	OWC

GU824024	18S	II	E (basal)	Cariaco Basin, CS	-	Edgcomb et al. (2011)	AWC
GU824240	18S	II	E (basal)	Cariaco Basin, CS	-	Edgcomb et al. (2011)	AWC
GU824243	18S	II	E (basal)	Cariaco Basin, CS	-	Edgcomb et al. (2011)	AWC
GU824258	18S	II	E (basal)	Cariaco Basin, CS	-	Edgcomb et al. (2011)	AWC
GU824264	18S	II	E (basal)	Cariaco Basin, CS	-	Edgcomb et al. (2011)	AWC
GU824278	18S	II	E (basal)	Cariaco Basin, CS	-	Edgcomb et al. (2011)	AWC
GU824450	18S	II	E (basal)	Cariaco Basin, CS	-	Edgcomb et al. (2011)	AWC
GU824572	18S	II	E (basal)	Cariaco Basin, CS	-	Edgcomb et al. (2011)	AWC
GU824702	18S	II	E (basal)	Cariaco Basin, CS	-	Edgcomb et al. (2011)	AWC
GU824724	18S	II	E (basal)	Cariaco Basin, CS	-	Edgcomb et al. (2011)	AWC
GU824777	18S	II	E (basal)	Cariaco Basin, CS	-	Edgcomb et al. (2011)	AWC
GU824925	18S	II	E (basal)	Cariaco Basin, CS	-	Edgcomb et al. (2011)	AWC
GU824964	18S	II	E (basal)	Cariaco Basin, CS	-	Edgcomb et al. (2011)	AWC
GU825032	18S	II	E (basal)	Cariaco Basin, CS	-	Edgcomb et al. (2011)	AWC
GU825058	18S	II	E (basal)	Cariaco Basin, CS	-	Edgcomb et al. (2011)	AWC
GU825217	18S	II-III-IV (basal)	-	Cariaco Basin, CS	-	Edgcomb et al. (2011)	AWC
GU825285	18S	II	E (basal)	Cariaco Basin, CS	-	Edgcomb et al. (2011)	AWC
GU825342	18S	II	E (basal)	Cariaco Basin, CS	-	Edgcomb et al. (2011)	AWC
GU825397	18S	II	D	Cariaco Basin, CS	-	Edgcomb et al. (2011)	AWC
GU825512	18S	II	E (basal)	Cariaco Basin, CS	-	Edgcomb et al. (2011)	AWC
GU825575	18S	III	G	Cariaco Basin, CS	-	Edgcomb et al. (2011)	mOWC
GU825590	18S	II	D	Cariaco Basin, CS	-	Edgcomb et al. (2011)	AWC
HM859020	18S	III	F		500m	Countway et al. (2010)	33.55 N 118.40 W
HQ156870	18S	III	G	Isfjorden Area, Svalbard		Sorensen et al. (2012)	0.22-3 µm
HQ865271	18S	III	G	Canada: Vancouver	100m	Orsi et al. (2012)	Saanich Inlet, OWC
HQ865359	18S	III	G	Canada: Vancouver	100m	Orsi et al. (2012)	Saanich Inlet, OWC
HQ865388	18S	III	G	Canada: Vancouver	100m	Orsi et al. (2012)	Saanich Inlet, OWC
HQ865403	18S	III	G	Canada: Vancouver	100m	Orsi et al. (2012)	Saanich Inlet, OWC
HQ865428	18S	III	G	Canada: Vancouver	100m	Orsi et al. (2012)	Saanich Inlet, OWC
HQ865480	18S	III	G	Canada: Vancouver	120m	Orsi et al. (2012)	Saanich Inlet,mOWC
HQ865532	18S	III	G	Canada: Vancouver	120m	Orsi et al. (2012)	Saanich Inlet,mOWC
HQ865536	18S	III	G	Canada: Vancouver	120m	Orsi et al. (2012)	Saanich Inlet,mOWC
HQ865549	18S	III	G	Canada: Vancouver	120m	Orsi et al. (2012)	Saanich Inlet,mOWC
HQ865637	18S	III	G	Canada: Vancouver	120m	Orsi et al. (2012)	Saanich Inlet,mOWC
HQ865670	18S	III	G	Canada: Vancouver	120m	Orsi et al. (2012)	Saanich Inlet,mOWC
HQ865676	18S	III	G	Canada: Vancouver	120m	Orsi et al. (2012)	Saanich Inlet,mOWC
HQ865688	18S	III	G	Canada: Vancouver	120m	Orsi et al. (2012)	Saanich Inlet,mOWC
HQ865691	18S	III	G	Canada: Vancouver	120m	Orsi et al. (2012)	Saanich Inlet,mOWC
HQ865707	18S	III	G	Canada: Vancouver	120m	Orsi et al. (2012)	Saanich Inlet,mOWC
HQ866202	18S	III	G	Canada: Vancouver	10m	Orsi et al. (2012)	Saanich Inlet, OWC
HQ866269	18S	III	G	Canada: Vancouver	100m	Orsi et al. (2012)	Saanich Inlet, OWC
HQ866271	18S	III	G	Canada: Vancouver	100m	Orsi et al. (2012)	Saanich Inlet, OWC
HQ866281	18S	III	G	Canada: Vancouver	100m	Orsi et al. (2012)	Saanich Inlet, OWC
HQ866283	18S	III	G	Canada: Vancouver	100m	Orsi et al. (2012)	Saanich Inlet, OWC
HQ866289	18S	III	G	Canada: Vancouver	100m	Orsi et al. (2012)	Saanich Inlet, OWC
HQ866294	18S	III	G	Canada: Vancouver	100m	Orsi et al. (2012)	Saanich Inlet, OWC
HQ866301	18S	III	G	Canada: Vancouver	100m	Orsi et al. (2012)	Saanich Inlet, OWC
HQ866311	18S	III	G	Canada: Vancouver	100m	Orsi et al. (2012)	Saanich Inlet, OWC
HQ866319	18S	III	G	Canada: Vancouver	100m	Orsi et al. (2012)	Saanich Inlet, OWC
HQ866321	18S	III	G	Canada: Vancouver	100m	Orsi et al. (2012)	Saanich Inlet, OWC
HQ866323	18S	III	G	Canada: Vancouver	100m	Orsi et al. (2012)	Saanich Inlet, OWC
HQ866326	18S	III	G	Canada: Vancouver	100m	Orsi et al. (2012)	Saanich Inlet, OWC
HQ866328	18S	III	G	Canada: Vancouver	100m	Orsi et al. (2012)	Saanich Inlet, OWC
HQ866329	18S	III	G	Canada: Vancouver	100m	Orsi et al. (2012)	Saanich Inlet, OWC
HQ866330	18S	III	G	Canada: Vancouver	100m	Orsi et al. (2012)	Saanich Inlet, OWC
HQ866333	18S	III	G	Canada: Vancouver	100m	Orsi et al. (2012)	Saanich Inlet, OWC
HQ866337	18S	III	G	Canada: Vancouver	100m	Orsi et al. (2012)	Saanich Inlet, OWC
HQ866339	18S	III	G	Canada: Vancouver	100m	Orsi et al. (2012)	Saanich Inlet, OWC
HQ866349	18S	III	G	Canada: Vancouver	100m	Orsi et al. (2012)	Saanich Inlet, OWC
HQ866359	18S	III	G	Canada: Vancouver	100m	Orsi et al. (2012)	Saanich Inlet, OWC
HQ866361	18S	III	G	Canada: Vancouver	100m	Orsi et al. (2012)	Saanich Inlet, OWC
HQ866388	18S	III	G	Canada: Vancouver	100m	Orsi et al. (2012)	Saanich Inlet, OWC
HQ866396	18S	III	G	Canada: Vancouver	100m	Orsi et al. (2012)	Saanich Inlet, OWC
HQ866400	18S	III	G	Canada: Vancouver	100m	Orsi et al. (2012)	Saanich Inlet, OWC
HQ866415	18S	III	G	Canada: Vancouver	100m	Orsi et al. (2012)	Saanich Inlet, OWC
HQ866419	18S	III	G	Canada: Vancouver	100m	Orsi et al. (2012)	Saanich Inlet, OWC
HQ866434	18S	III	G	Canada: Vancouver	100m	Orsi et al. (2012)	Saanich Inlet, OWC
HQ866438	18S	III	G	Canada: Vancouver	100m	Orsi et al. (2012)	Saanich Inlet, OWC
HQ866448	18S	III	G	Canada: Vancouver	100m	Orsi et al. (2012)	Saanich Inlet, OWC
HQ866451	18S	III	G	Canada: Vancouver	100m	Orsi et al. (2012)	Saanich Inlet, OWC
HQ866452	18S	III	G	Canada: Vancouver	100m	Orsi et al. (2012)	Saanich Inlet, OWC
HQ866457	18S	III	G	Canada: Vancouver	100m	Orsi et al. (2012)	Saanich Inlet, OWC
HQ866458	18S	III	G	Canada: Vancouver	100m	Orsi et al. (2012)	Saanich Inlet, OWC
HQ866460	18S	III	G	Canada: Vancouver	100m	Orsi et al. (2012)	Saanich Inlet, OWC
HQ866474	18S	III	G	Canada: Vancouver	100m	Orsi et al. (2012)	Saanich Inlet, OWC
HQ866482	18S	III	G	Canada: Vancouver	100m	Orsi et al. (2012)	Saanich Inlet, OWC
HQ866483	18S	III	G	Canada: Vancouver	100m	Orsi et al. (2012)	Saanich Inlet, OWC
HQ866491	18S	III	G	Canada: Vancouver	100m	Orsi et al. (2012)	Saanich Inlet, OWC
HQ866518	18S	III	G	Canada: Vancouver	100m	Orsi et al. (2012)	Saanich Inlet, OWC
HQ866524	18S	III	G	Canada: Vancouver	100m	Orsi et al. (2012)	Saanich Inlet, OWC
HQ866528	18S	III	G	Canada: Vancouver	100m	Orsi et al. (2012)	Saanich Inlet, OWC
HQ866533	18S	III	G	Canada: Vancouver	100m	Orsi et al. (2012)	Saanich Inlet, OWC
HQ866549	18S	III	G	Canada: Vancouver	120m	Orsi et al. (2012)	Saanich Inlet,mOWC
HQ866550	18S	III	G	Canada: Vancouver	120m	Orsi et al. (2012)	Saanich Inlet,mOWC
HQ866607	18S	III	G	Canada: Vancouver	120m	Orsi et al. (2012)	Saanich Inlet,mOWC
HQ866608	18S	III	G	Canada: Vancouver	120m	Orsi et al. (2012)	Saanich Inlet,mOWC
HQ866757	18S	III	G	Canada: Vancouver	120m	Orsi et al. (2012)	Saanich Inlet,mOWC
HQ866836	18S	III	G	Canada: Vancouver	120m	Orsi et al. (2012)	Saanich Inlet,mOWC
HQ867255	18S	III	G	Canada: Vancouver	10m	Orsi et al. (2012)	Saanich Inlet, OWC
HQ867403	18S	III	G	Canada: Vancouver	10m	Orsi et al. (2012)	Saanich Inlet, OWC
HQ867502	18S	III	G	Canada: Vancouver	100m	Orsi et al. (2012)	Saanich Inlet, OWC
HQ867509	18S	III	G	Canada: Vancouver	100m	Orsi et al. (2012)	Saanich Inlet, OWC
HQ867528	18S	III	G	Canada: Vancouver	100m	Orsi et al. (2012)	Saanich Inlet, OWC
HQ867544	18S	III	G	Canada: Vancouver	100m	Orsi et al. (2012)	Saanich Inlet, OWC
HQ867552	18S	III	G	Canada: Vancouver	100m	Orsi et al. (2012)	Saanich Inlet, OWC

HQ867556	18S	III	G	Canada: Vancouver	100m	Orsi et al. (2012)	Saanich Inlet, OWC
HQ867564	18S	III	G	Canada: Vancouver	100m	Orsi et al. (2012)	Saanich Inlet, OWC
HQ867566	18S	III	G	Canada: Vancouver	100m	Orsi et al. (2012)	Saanich Inlet, OWC
HQ867573	18S	III	G	Canada: Vancouver	100m	Orsi et al. (2012)	Saanich Inlet, OWC
HQ867606	18S	III	G	Canada: Vancouver	100m	Orsi et al. (2012)	Saanich Inlet, OWC
HQ867622	18S	III	G	Canada: Vancouver	100m	Orsi et al. (2012)	Saanich Inlet, OWC
HQ867630	18S	III	G	Canada: Vancouver	100m	Orsi et al. (2012)	Saanich Inlet, OWC
HQ867639	18S	III	G	Canada: Vancouver	100m	Orsi et al. (2012)	Saanich Inlet, OWC
HQ867646	18S	III	G	Canada: Vancouver	100m	Orsi et al. (2012)	Saanich Inlet, OWC
HQ867648	18S	III	G	Canada: Vancouver	100m	Orsi et al. (2012)	Saanich Inlet, OWC
HQ867652	18S	III	G	Canada: Vancouver	100m	Orsi et al. (2012)	Saanich Inlet, OWC
HQ867684	18S	III	G	Canada: Vancouver	100m	Orsi et al. (2012)	Saanich Inlet, OWC
HQ867704	18S	III	G	Canada: Vancouver	100m	Orsi et al. (2012)	Saanich Inlet, OWC
HQ867783	18S	III	G	Canada: Vancouver	100m	Orsi et al. (2012)	Saanich Inlet, OWC
HQ867784	18S	III	G	Canada: Vancouver	100m	Orsi et al. (2012)	Saanich Inlet, OWC
HQ867796	18S	III	G	Canada: Vancouver	100m	Orsi et al. (2012)	Saanich Inlet, OWC
HQ867807	18S	III	G	Canada: Vancouver	100m	Orsi et al. (2012)	Saanich Inlet, OWC
HQ867810	18S	III	G	Canada: Vancouver	100m	Orsi et al. (2012)	Saanich Inlet, OWC
HQ867828	18S	III	G	Canada: Vancouver	100m	Orsi et al. (2012)	Saanich Inlet, OWC
HQ867831	18S	III	G	Canada: Vancouver	100m	Orsi et al. (2012)	Saanich Inlet, OWC
HQ868165	18S	III	G	Canada: Vancouver	200m	Orsi et al. (2012)	Saanich Inlet, OWC
HQ868223	18S	III	G	Canada: Vancouver	200m	Orsi et al. (2012)	Saanich Inlet, OWC
HQ868229	18S	III	G	Canada: Vancouver	200m	Orsi et al. (2012)	Saanich Inlet, OWC
HQ868355	18S	III	G	Canada: Vancouver	200m	Orsi et al. (2012)	Saanich Inlet, OWC
HQ869288	18S	III	G	Canada: Vancouver	100m	Orsi et al. (2012)	Saanich Inlet, OWC
HQ869413	18S	III	G	Canada: Vancouver	100m	Orsi et al. (2012)	Saanich Inlet, OWC
HQ869580	18S	III	G	Canada: Vancouver	120m	Orsi et al. (2012)	Saanich Inlet,mOWC
HQ870133	18S	III	G	Canada: Vancouver	100m	Orsi et al. (2012)	Saanich Inlet, OWC
HQ870144	18S	III	G	Canada: Vancouver	100m	Orsi et al. (2012)	Saanich Inlet, OWC
HQ870179	18S	III	G	Canada: Vancouver	100m	Orsi et al. (2012)	Saanich Inlet, OWC
HQ870318	18S	III	G	Canada: Vancouver	100m	Orsi et al. (2012)	Saanich Inlet, OWC
HQ870364	18S	III	G	Canada: Vancouver	100m	Orsi et al. (2012)	Saanich Inlet, OWC
HQ870413	18S	III	G	Canada: Vancouver	100m	Orsi et al. (2012)	Saanich Inlet, OWC
HQ870539	18S	III	G	Canada: Vancouver	120m	Orsi et al. (2012)	Saanich Inlet,mOWC
HQ870540	18S	III	G	Canada: Vancouver	120m	Orsi et al. (2012)	Saanich Inlet,mOWC
HQ870606	18S	III	G	Canada: Vancouver	120m	Orsi et al. (2012)	Saanich Inlet,mOWC
HQ870633	18S	III	G	Canada: Vancouver	120m	Orsi et al. (2012)	Saanich Inlet,mOWC
HQ870770	18S	III	G	Canada: Vancouver	120m	Orsi et al. (2012)	Saanich Inlet,mOWC
JQ226204	18S	III	G	Canada: Vancouver	100m	Wright & Hallam (Unpub)	
JQ226414	18S	III	G	Canada: Vancouver	10m	Wright & Hallam (Unpub)	
JQ956235	18S	III	G	Canada: Arctic	33m	Terrado et al. (2012)	Baffin Bay
JQ956296	18S	III	G	Canada: Arctic	70m	Terrado et al. (2012)	Baffin Bay
JX842092	18S	IV	I (basal)	Eastern North Pacific	500m	Kim et al. (2012)	33.55 N 118.40 W
JX842410	18S	IV	J	Eastern North Pacific	500m	Kim et al. (2012)	33.55 N 118.40 W
JX842634	18S	II	D	Eastern North Pacific	500m	Kim et al. (2012)	33.55 N 118.40 W
JX842654	18S	III	G	Eastern North Pacific	500m	Kim et al. (2012)	33.55 N 118.40 W
KC455077	18S	III	G	Juturnaiba freshwater lake	-	Froes et al. (Unpub)	
KC488587	18S	III	G	Canada (42.93N 61.82W)	30m	Dasilva et al. (Unpub)	
KC583143	18S	III	G	Red Sea	10m	Acosta et al. (2013)	
KF129697	18S	III	G	SCS	Surface	Wu et al. (2014)	
KF129924	18S	III	G	SCS	60m	Wu et al. (2014)	
KF130072	18S	II	E	SCS	60m	Wu et al. (2014)	
KF130075	18S	II	E	SCS	60m	Wu et al. (2014)	
KF130078	18S	II	E	SCS	60m	Wu et al. (2014)	
KF130087	18S	II	E	SCS	60m	Wu et al. (2014)	
KF130088	18S	II	E	SCS	60m	Wu et al. (2014)	
KF130100	18S	II	E	SCS	60m	Wu et al. (2014)	
KF130104	18S	II	E	SCS	60m	Wu et al. (2014)	
KF130119	18S	II	E	SCS	60m	Wu et al. (2014)	
KF130124	18S	II	E	SCS	60m	Wu et al. (2014)	
KF130127	18S	II	E	SCS	60m	Wu et al. (2014)	
KF130141	18S	II	E	SCS	60m	Wu et al. (2014)	
KF130142	18S	II	E	SCS	60m	Wu et al. (2014)	
KF130144	18S	II	E	SCS	60m	Wu et al. (2014)	
KF130147	18S	II	E	SCS	60m	Wu et al. (2014)	
KF130151	18S	II	E	SCS	60m	Wu et al. (2014)	
KF130430	18S	III	G	SCS	60m	Wu et al. (2014)	
KF130452	18S	II	E	SCS	60m	Wu et al. (2014)	
KF130460	18S	II	E	SCS	60m	Wu et al. (2014)	
KF130468	18S	II	E	SCS	60m	Wu et al. (2014)	
KJ757061	18S	IV	I	East Pacific Rise	2500m	Lie et al. (2014)	
KJ757256	18S	IV	I-J (basal)	East Pacific Rise	2500m	Lie et al. (2014)	
KJ759920	18S	IV	I	Gulf Stream	2500m	Lie et al. (2014)	
KJ759983	18S	III	G	Gulf Stream	2500m	Lie et al. (2014)	
KJ760143	18S	II-III-IV (basal)	-	Gulf Stream	2500m	Lie et al. (2014)	
KJ760258	18S	II	E	Gulf Stream	2500m	Lie et al. (2014)	
KJ762302	18S	III	G	Arctic Ocean	35m	Lie et al. (2014)	
KJ762790	18S	II	D (basal)	SPOT station	500m	Lie et al. (2014)	
KJ762806	18S	III	G	SPOT station	500m	Lie et al. (2014)	
KJ763458	18S	IV	J	Gulf Stream	105m	Lie et al. (2014)	
KJ763826	18S	IV	J	Gulf Stream	105m	Lie et al. (2014)	
KJ763874	18S	III	G	Gulf Stream	105m	Lie et al. (2014)	
KX532517	18S	II	E	SCS	-	Xu et al. (2017)	Bathypelagic
KX532522	18S	II	D	SCS	-	Xu et al. (2017)	Bathypelagic
KX532523	18S	II	C-D (basal)	SCS	-	Xu et al. (2017)	Bathypelagic
KX532534	18S	II	D	SCS	-	Xu et al. (2017)	Bathypelagic
KX532627	18S	II	E (basal)	SCS	-	Xu et al. (2017)	Bathypelagic
KX532655	18S	II	E (basal)	SCS	-	Xu et al. (2017)	Bathypelagic
KX532666	18S	II	E (basal)	SCS	-	Xu et al. (2017)	Bathypelagic
KX532672	18S	II	E (basal)	SCS	-	Xu et al. (2017)	Bathypelagic
KX532677	18S	II	E (basal)	SCS	-	Xu et al. (2017)	Bathypelagic
KX532876	18S	II-III-IV (basal)	-	SCS	-	Xu et al. (2017)	Bathypelagic

Chapter 1.2

A morpho-molecular classification of Spumellaria dated with the fossil record



Fossil dated morpho-molecular classification reveals possible new lifestyles within Spumellaria

Miguel M. Sandin[1], Tristan Biard[1], Sarah Romac[1], Luis O'Dogherty[2], Noritoshi Suzuki[3], Fabrice Not[1]

1-Sorbonne Université, CNRS -UMR7144- Station Biologique de Roscoff, 29680 Roscoff, France.
2-Facultad de Ciencias del Mar y Ambientales, Cadiz University, E-11510 Puerto Real, Spain.
3-Department of Earth Science, Graduate School of Science, Tohoku University, Sendai 980–8578, Japan.

In preparation

Abstract

Spumellaria are holoplanktonic ameboid protists (Radiolaria, Rhizaria) widely and abundantly distributed in the global oceans. Their silicified skeleton preserves very well in sediments displaying an excellent and continuous fossil record from the early Cambrian, extremely valuable for paleo-environmental reconstruction studies. The short period of time they survive in lab conditions prevents an accurate perception of their extant diversity and ecology in contemporary oceans, and most of it is inferred from sediments. Several attempts have explored their taxonomy, yet no convincing results have been reported. Here we present an integrative classification of Spumellaria based on taxonomic marker genes (18S and 28S ribosomal DNA) and morphological characteristics obtained by optical and scanning electron microscopy imaging. Our phylogenetic analysis distinguished 13 main morpho-molecular clades partly agreeing with the latest morphological classifications, yet overall symmetry of the skeleton takes more importance than internal structures at higher rank classification in contrast to previous work. Using fossil calibrated molecular clock we estimated the origin of Spumellaria among the first Polycystine representatives in the lower Cambrian (ca. 515 Ma). A great molecular environmental diversity appeared at early diverging positions by the phylogenetic placement of environmental sequences. The results obtained in this study led us to hypothesis the existence of a new mode of symbiotic Spumellaria, in which a non-bearing skeleton organism lives within another shell-Spumellaria. This study brings the first comprehensive classification of Spumellaria and contributes to the understanding of both their diversity and evolutionary history.

Introduction

Radiolaria are holoplanktonic ameboid protists belonging to the Rhizaria lineage (SAR), one of the main branches of the eukaryotic tree of life (Burki and Keeling, 2014). Along with minifera they have classically composed the phylum Retaria, and recently named the hylum Ectoreta for Radiolaria and Foraminifera with Endomyxa as a sister clade within the Retaria (Cavalier-Smith et al., 2018). Spumellaria constitute an important order within the

Radiolaria (Suzuki and Not, 2015) extensively studied across the fossil record thanks to its opaline silica skeleton (De Wever et al., 2001). Their detailed fossil record, along with that of other radiolarians, date back to the Cambrian (Suzuki and Oba, 2015; Aitchison et al., 2017) constituting an important tool for paleoenvironmental reconstructions analysis (e.g., Abelmann and Nimmergut, 2005). Molecular-based metabarcoding surveys performed at global scale have shown Radiolaria contribute significantly to plankton communities (de Vargas et al., 2015; Pernice et al., 2016). Although very little is known about their contemporary ecology and diversity.

Living Spumellaria, as other Radiolaria, are characterized by a complex protoplasmic meshwork of pseudopodia extending radially from their skeleton (Suzuki and Aita, 2011). With this complex, they actively capture prey by adhesion such as copepod nauplii or tintinnids (Sugiyama and Anderson, 1997; Sugiyama et al., 2008). Besides, this trophic behaviour, many spumellarian species harbour photosynthetic algal symbionts mainly identified as dinoflagellates (Probert et al., 2014; Yuasa et al., 2016; Zhang et al., 2018). Although several studies have reported symbiotic associations of Spumellaria with cyanobacteria (Foster et al., 2006a, 2006b; Yuasa et al., 2012), prasinophytes (Gast and Caron, 2001), and haptophytes (Anderson, 1983; Yuasa et al., 2019) in the sunlit ocean. This mixotrophic behaviour may influences their distribution patterns, being in surface tropical waters the greatest diversity and abundance values, and decreasing towards the poles and at depth as that of Nassellaria (Boltovskoy and Correa, 2016b; Boltovskoy, 2017).

Spumellaria classifications are historically based on morphological criteria. Radiolarians with a concentric structure have been classified as Spumellaria relying on the absence of the initial spicular system, an early develop skeletal structure considered the foundation of the systematics at family and higher levels. Traditionally, the absence or presence of this morphological character was concluding the Spumellaria or Entactinaria nature, respectively, despite their similar morphology. The most recent spumellarian classification studies (Matsuzaki et al., 2015) have attempted to merge the extensive morphological criteria (De Wever et al., 2001; Afanasieva et al., 2005; O'Dogherty et al., 2009b; Noble et al., 2017) with recent rDNA molecular studies. The current scheme describes 9 extant superfamilies: Actinommoidea (Haeckel, 1862; O'Dogherty, 1994), Hexastyloidea (Haeckel, 1882), Liosphaeroidea (Haeckel, 1882), Lithelioidea (Haeckel, 1862; Petrushevskaya, 1975), Pylonioidea (Haeckel, 1882; Dumitrica, 1989), Saturnalioidea (Deflandre, 1953), Spongodiscoidea (Haeckel, 1862; De Wever et al., 2001), Sponguroidea (Haeckel, 1862; De Wever et al., 2001), Stylodictyoidea (Haeckel, 1882; Matsuzaki et al., 2015); and two undetermined families (Heliodiscidae, Haeckel, 1887; De Wever et al., 2001; and Spongosphaeroidae Haeckel, 1862). However, despite the attempt in combining different approaches they have stated the "overwhelmingly artificial" systematics of Spumellaria and reflected the outweighed morphological criteria in the classification.

To date, few studies have explored the genetic diversity of Spumellaria unveiling relationships among higher rank taxonomical groups (Kunitomo et al., 2006; Yuasa et al., 2009; Krabberød et al., 2011; Ishitani et al., 2012a). With a total of 35 sequences from morphologically described specimens, covering 5 of the 9 superfamilies described, two main groups of Spumellaria are found with different innermost shell structure; Hexalonchidae is

included within Spumellaria in group I and Pylonioidea in group II; also the family Spongodiscidae shows to be polyphyletic (Yuasa et al., 2009; Ishitani et al., 2012a). Despite such important insights, the understanding of these relationships and the rest of the families remain still elusive. As previously seen (e.g., Decelle et al., 2012a; Lewitus et al., 2018; Sandin et al., 2019, Chapter 1.1), the time-calibration of such phylogenies thanks to the fossil record allows a better understanding of relationships among extinct groups and a contextualized evolutionary history. In addition, the acquisition of single cell reference DNA barcodes morphologically described establish the basis for further molecular ecology surveys, inferring the actual diversity and ecology in the nowadays oceans (Decelle et al., 2013; Nitsche et al., 2016; Biard et al., 2017).

Here we present an integrative morpho-molecular classification of Spumellaria obtained by combining ribosomal DNA taxonomical marker genes (18S and 28S partial rDNA) and imaging techniques (light and scanning electron microscopy). The extensive fossil record of Spumellaria allowed to calibrate in time our phylogenetic analysis and infer their evolutionary history contextualized with global scale geological and environmental changes. Finally, phylogenetic placement of environmental sequences provided insights in the extant genetic diversity of Spumellaria in the contemporary oceans.

Material and Methods

Sampling and single cell isolation: Plankton samples were collected off Sendai (11 samples: 38° 0’ 28.8’’ N, 142° 0’ 28.8’’ E), Sesoko (8 samples: 26° 48’ 43.2’’ N, 73° 58’ 58.8’’ E), the Southwest Islands, South of Japan (2 samples: 28° 14’ 49.2’’ N, 129° 5’ 27.6’’ E), in the bay of Villefranche-sur-Mer (38 samples: 43° 40’ 51.6’’ N, 7° 19’ 40.8’’ E) and in the western Mediterranean Sea (9 samples: MOOSE-GE cruise) by net tows (20, 64 or 200 µm), Vertical Multiple-opening Plankton Sampler (VMPS) or Bongo net (24-200 µm). Samples related metadata can be found in the RENKAN database (http://abims.sb-roscoff.fr/renkan). Spumellarian specimens were individually handpicked with Pasteur pipettes from the samples and transferred 3 to 4 times into 0.2 µm filtered seawater to allow self-cleaning from debris, particles attached to the cell or prey digestion. Images of live specimens were taken under an inverted microscope and thereafter transfer into 1.5 ml Eppendorf tubes containing 50 µl of molecular grade absolute ethanol and stored at -20 ºC until DNA extraction.

Single cell morphological identification: Spumellaria specimens were identified at the species level, referring to pictures of holotypes, through observation of live images and posterior analysis of the skeleton by scanning electron microscopy when available (see Supplementary material Table 1 for taxonomic authority of specimens included in our study). Further details in taxonomic assignment of the specimens can be found in the Material and Methods of Sandin et al. (2019), Chapter 1.1.

DNA extraction, amplification and sequencing: DNA was extracted using the MasterPure Complete DNA and RNA Purification Kit (Epicentre) following manufacturer’s instructions. Both 18S rDNA and partial 28S rDNA genes (D1 and D2 regions) were amplified by Polymerase Chain Reaction (PCR) using Radiolaria and Spumellaria specific and general primers (Table 1). For further details about rDNA amplification see: dx.doi.org/10.17504/protocols.io.xwvfpe6. PCR amplicons were visualized on 1% agarose gel stained with ethidium bromide. Positive

reactions were purified using the Nucleospin Gel and PCR Clean up kit (Macherey Nagel), following manufacturer's instructions and sent to Macrogen Europe for sequencing.

Table 1. Primer sequences and temperatures used for DNA amplification and sequencing.

Targeted gene	Primer	Specificity	Sequence 5'-3'	Direction	Tm °C	Reference
18S (1st part)	SA	Eukaryotes	AACCTGGTTGATCCTGCCAGT	Forward	56	Medlin et al., 1988
	S879	Radiolaria	CCAACTGTCCCTATCAATCAT	Reverse		Decelle et al., 2012b
18S (2nd part)	S32_TASN	Radiolaria	CCAGCTCCAATAGCGTATRC	Forward	57	Ishitani et al. 2012
	V9R	Eukaryotes	CCTTCYGCAGGTTCACCTAC	Reverse		Romac (unpub.)
28S (D1+D2)	28S Rad2	Radiolaria	TAAGCGGAGGAAAAGAAA	Forward	52	Ando et al., 2009
	ITSa4R	Radiolaria	TCACCATCTTTCGGGTCCCGGCAT	Reverse		This study

Phylogenetic analyses: After sequencing, forward and reverse sequences were checked and assembled using ChromasPro software version 2.1.4 (2017). Sequences were compared to the GenBank reference database (GenBank) using the *BLAST search* tool integrated in ChromasPro to discriminate radiolarian sequences from possible contamination. Presence of chimeras was detected by mothur v.1.39.3 (Schloss et al., 2009) against previously available reference sequences of Spumellaria. Sequences not detected as chimeras, were in turn included in our reference database for later chimeric analysis.

Two different datasets for each genetic marker (18S rDNA gene and partial 28S rDNA gene) were aligned separately using MAFFT v7.395 (Katoh and Standley, 2013) with a L-INS-i algorithm ('--localpair') and 1000 iterative refinement cycles for high accuracy. Each alignment was manually checked in SeaView version 4.6.1 (Gouy et al., 2010) and trimmed automatically using trimal (Capella-Gutiérrez et al., 2009) with a 30% gap threshold. For both genes, the 18S rDNA (133 taxa, 1789 positions) and the 28S rDNA (55 taxa, 746 positions), phylogenetic analyses were performed independently. The best nucleotide substitution model was chosen following the corrected Akaike Information Criterion (AICc) using the *modelTest* function implemented in the R version 3.6.0 (R Core Team, 2014) package *phangorn* version 2.5.5 (Schliep, 2011). The obtained model (General Time Reversible with Gama distribution and proportion of Invariable sites, GTR+G+I) was applied to each data set in R upon the packages *APE* version 5.3 (Paradis et al., 2004) and *phangorn* version 2.5.5 (Schliep, 2011). A Maximun Likelihood (ML) method (Felsenstein, 1981) with 10 000 replicates of bootstrap (Felsenstein, 1985) was performed to infer phylogenies.

Despite specific discrepancies in the topology of the two analysis (Supplementary material Fig. S1), the two genes were concatenated in order to increase taxonomic coverage and improve phylogenetic resolution. A final data set containing 133 taxa and 2535 positions was used to infer phylogenies following the previous methodology. Sixteen sequences of Nassellaria were assembled to form the outgroup as seen in previous classifications to be the sister clade (eg; Krabberød et al., 2011; Cavalier-Smith et al., 2018). The best model obtained was GTR+G+I, with 4 intervals of the discrete gamma distribution, and a ML method with 100 000 bootstraps were performed. In parallel, a Bayesian analysis was performed using BEAST version 1.8.4 (Drummond et al., 2012) with the same model parameters over 100 million generations sampled every 1000 states, to check the consistency of the topology and to

calculate posterior probabilities. Final tree was visualized and edited with FigTree version 1.4.3 (Rambaut 2016).

Molecular clock analyses: Molecular clock estimates were performed according to Sandin et al. (2019), Chapter 1.1. Nine nodes were chosen to carry out the calibration. The selection of these nodes is explained below from the oldest to the newest calibration age, and given the name of the node for the taxa they cover:

· 1. Root: The calibration for the root of the tree corresponds to the hypothesized last common ancestor between Nassellaria and Spumellaria (De Wever et al., 2001). A uniform distribution (U) with a minimum bound of 300 million years ago (Ma) and a maximum of 800 Ma (U[300, 800]) was set to allow uncertainty in the diversification of both groups and to establish a threshold restricting the range of solutions for the entire tree.

· 2. Nassellaria N(410, 20): The outgroup of the phylogeny is calibrated based in a consensus between the first appearance of nassellarian-like fossils in the Upper Devonian (ca. 372.2 Ma) in the fossil record (Cheng, 1986) and the first diversification of Nassellaria dated with previous analysis of the molecular clock (ca. 423 Ma; 95% HPD: 500-342 Ma; Sandin et al., 2019) , Chapter 1.1. Therefore, the node was normally distributed (N) with a mean of 410 and a standard deviation of 20: N(410, 20).

· 3. Spumellaria U[700, 200]: Recent studies have found the oldest spumellarian representatives in the Early Cambrian (Zhang and Feng, 2019). Yet, De Wever et al. (2001) have argued that the initial spicular system may be subjected to preservation bias, a character defining by its absence the Spumellarian authority. Since many spumellarians and entactinarians (presence of the initial spicular system) are superficially similar and it is not possible to distinguish from one another (Suzuki and Oba, 2015), a uniform distribution was set to allow uncertainty in the diversification of Spumellaria.

· 4. Hexastyloidea N(242, 10): The family Hexastylidae is the first representative of the superfamily Hexastyloidea and has its first appearance in the fossil record in the Middle Triassic (Late Anisian: ca. 246.8 - 241.5 Ma; O'Dogherty et al., 2011).

· 5. Liosphaeroidea N(233, 20): The Liosphaeroidea seems to have appeared in the Triassic (Pessagno and Blome, 1980; De Wever et al., 2001). Although it is not sure whether there is a continuity between members from the Mesozoic and from the Cenozoic (Matsuzaki et al., 2015).

· 6. Actinommidae N(170, 20): The family Actinommidae appears for the first time in the fossil record in the Middle Jurassic (Aalenian: ca. 174.2-170.3 Ma; O'Dogherty et al., 2011) yet some morphologies from the Triassic resemblance this Superfamily (De Wever et al., 2001).

· 7. Rhizosphaeroidea N(148, 10): The Rhizosphaeridae appeared during the late Jurassic (Tithonian: ca. 145-152.1 Ma) in the fossil record (Petrushevskaya, 1975; De Wever et al., 2001; Afanasieva and Amon, 2006; Dumitrica, 2017b).

· 8. Pylonioidea N(97, 10): The family Larnacillidae are the first representatives of the superfamily Pylonioidea appearing at the beginning of the Late Cretaceous (Cenomanian: ca. 100.5-93.9 Ma; De Wever et al., 2001, 2003; Afanasieva and Amon, 2006).

· 9. Coccodiscoidea N(45, 10): The family Coccodiscoidea appears for the first time in the fossil record in the Early Eocene (De Wever et al., 2001; Afanasieva and Amon, 2006).

Post hoc analyses: Two different analyses were performed *a posteriori*: a diversification of taxa over time (Lineages Through Time: LTT) and an ancestral state reconstruction. The former analysis was carried out with the *ltt.plot* function implemented in the package *APE* (Paradis et al., 2004) upon the tree obtained by the molecular dating analyses after removing the outgroup. The second analysis uses the resulting phylogenetic tree to infer the evolution of morphological characters. A numerical value was assigned to each state of a character trait, being 0 for the considered ancestral state, and 1 to 3 or 4 for the presumed divergence state. In total 5 traits were considered (Table 2): the skeleton shape, the central structure, the number of spines arisen from the central structure, the number of distinctive cortical shells and the internal structure outside the central part. Once the character matrix was established a parsimony ancestral state reconstruction was performed to every character independently in Mesquite version 3.2 (Madison & Madison, 2017).

Table 1. List of morphological characters (traits) and their states (1 to 5) used for the ancestral state reconstruction analysis.

		Traits				
		Skeleton shape	Central structure	Number of spines arisen from central structure	Number of distinctive cortical shell	Internal structure outside the central part
State	0	Spherical	Empty	0	1	Empty
	1	Spherical modified	Filament, ring, spicular system	6	2	Concentric structures
	2	Cubic-, box-, dice-shaped	Double medullary shell, rhizosphaerid centre	More or less than 6	3	Spongy
	3	Flat	Concentric structure, highly dense centre	Significantly more than 6	More than 3	Complicated
	4	-	Six faces along three vertical axes	-	-	-

Environmental sequences: Each of the reference 18S rDNA and partial 28S rDNA spumellarian sequence available in our study was compared with publicly available environmental sequences in GenBank (NCBI) using BLAST (as of May 2019). It allowed estimating the environmental genetic diversity of Spumellaria and to assess the genetic coverage of our phylogenetic tree. Environmental sequences were placed in our reference phylogenetic tree using the pplacer software (Matsen et al., 2010). A RAxML (GTR+G+I) tree was built for the placement of the sequences with a rapid bootstrap analysis and search for best-scoring ML tree and 1000 bootstraps.

Scanning Electron Microscopy (SEM): After DNA extraction, spumellarian skeletons were recovered from the eluted pellet and handpicked under binoculars or inverted microscope. After cleaning and preparing the skeletons (detailed protocol in dx.doi.org/10.17504/protocols.io.ug9etz6) images were taken with a FEI Phenom table-top Scanning Electron Microscope (FEI technologies).

Results

Comparative molecular phylogeny and morphological taxonomy:

Our final molecular phylogeny is composed of 133 distinct spumellarian sequences of the full 18S rDNA gene, completed with 55 sequences of the partial 28S (D1 & D2 regions) rDNA gene (Supplementary material Table S2). From the 133 final sequences of the 18S rDNA, 67 were obtained in this study, 58 were previously available sequences morphologically described and 8 were environmental sequences not morphologically described added to increase phylogenetic support in poorly represented clades. Regarding the partial 28S rDNA, a total of 37 sequences were obtained in this study and 18 were previously available. The final alignment matrix has 25.45% of invariant sites. Morphological observations performed with light and scanning electron microscopy (Supplementary material Fig. S2) assign all these sequences to 7 Superfamilies (Actinommoidea, Hexastyloidea scattered in 3 different clades, Liosphaeroidea, Pylonioidea, Spongodiscoidea scattered in 3 different clades -Lithocyclioidea, Spongodiscoidea and Spongopyloidea-, Spongosphaeroidea, Stylodictyoidea) and 2 superfamilies considered to belong to Entactinaria (Rhizosphaeridae and Centrocuboidea). Phylogenetic analysis shows 13 different clades (Fig. 1) clearly differentiated by BS and PP values (Clades A, B, C, D, E, F, G, H, I J, K, L and M). In general, morphological classification agrees with molecular phylogeny at the clade level, although several discrepancies are found.

All morpho-molecular clades were highly supported in both the 18S rDNA and the 28S rDNA gene phylogenies, yet the general topology and relationships between clades slightly disagree in between the 18S rDNA and the 28S rDNA gene phylogenies (Supplementary material Figure S1). Such discrepancies are basically due to the variant position of Clade F that in the 18S rDNA gene phylogeny appears basal to all clades and in the 28S rDNA gene appears basal to clades J, K, L and M, as a sister group of clades H and G (Supplementary material Figure S1). Another example is, clades B, C, D and E, which appear in the same highly supported group in the 18S rDNA gene phylogeny, whereas in the 28S rDNA gene phylogeny these clades appear scattered at basal positions regarding the rest of the tree (Supplementary material Figure S1). In the concatenation of the two genes clade F appears highly supported as a group with the clades G to M (Fig. 1). Yet low BS and PP values in the node G-M and a short phylogenetic distance reflect a still variant position of Clade F. Due to this conflicting position and the high BS (100) and PP (1) values this clade constitutes the so-called lineage II by itself. The other lineages show high BS (>97) and PP (1) and their positions are constant across the different phylogenetic analysis. Lineage I includes clades A, B, C, D and E and it is the sister group of lineages III (including clades G, H and I) and IV (including clades J, K, L and M). It worth mentioning the high ratio of chimeric sequences all along the sequences acquisition process, especially among lineage I specimens.

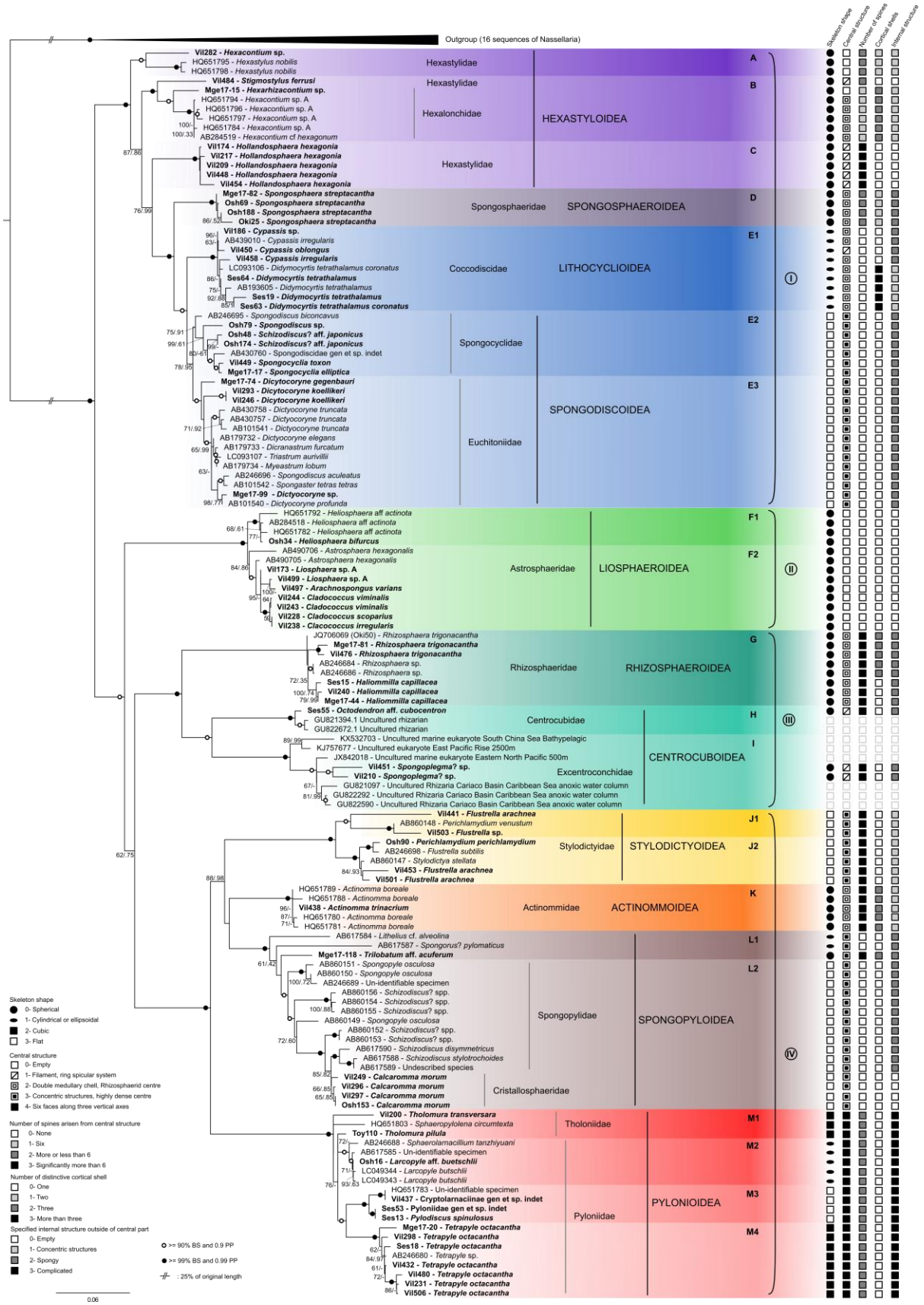


Figure 1. Legend on the following page.

Lineage I is characterized by the presence of one or two concentric shells or full of spongy test, sometimes non-distinguishable, where the main spines grow from the innermost shell and go through the outermost shell, when present (Fig. 2. A-E). Only clade A and B shows two concentric shells with six primary spines from the centre (Fig. 2. A-B). The former clade is composed of 1 novel and 2 previously available sequences and a common characteristic is the fragile and hexagonal mesh constituting the outermost shell (Fig. 2.A). Clade B is composed of 2 novel and 5 previously available sequences clustered with high BS and PP values despite their phylogenetic distance (Fig. 1). All specimens of this clade show a more irregular and thicker mesh of the shell compare to that of clade A, and, when present, very thin spines (generally called as by-spines) coming out from all the surface of the shell (Fig. 2. Ba and Bb). Clade C is composed of 5 novel sequences showing a big single double layered shell with no main spines but a large abundance of long and thin spines (Fig. 2. C). These three clades agree with the definition of the Superfamily Hexastyloidea (Haeckel, 1882; Matsuzaki et al., 2015). The innermost shell if present always shows pyriform with six radial beams. Yet families within Hexastyloidea appear scattered among the three clades with moderate to high BS and PP values. Clades D and E are the more distal clades in this lineage and highly supported together (100 BS and 1 PP). The presence of spongy structures differentiates these two clades from the others on the lineage (Fig. 2. D-E). Clade D is composed of 4 novel sequences obtained in this study. All specimens of this clade show a single, very small spherical shell where several long and three bladed spines grow interconnected by a spongy mesh (Fig. 2. D). This definition agrees with the genus *Spongosphaera* belonging to the Superfamily Spongosphaeroidea (Haeckel, 1862). The last clade of Lineage I (Clade E) is composed of 30 sequences of which 15 are novel. Members of this clade tend to lose the concentric symmetry towards a cylindrical to ellipsoidal (subclade E1; Fig. 2. E1a and E1b) or flat symmetry (subclade E2, Fig. 2. E2a and Eb; and Subclade E3, Fig. 2. E3), and so does the spines and the shell. Spongy structures take more importance complicating the inner structure. These morphologies agree with the definition of the Superfamily Spongodiscoidea (Haeckel, 1862; De Wever et al., 2001). Within this clade there are representatives of three families belonging to the Lithocyclioidea: Coccodiscidae (Haeckel, 1862) highly supported in subclade E1, and Spongodiscoidea: Spongocyclidae (Haeckel, 1862) with moderate BS and PP values in subclade E2 and Eucthitoniidae (Haeckel, 1882) highly supported in subclade E3.

Figure 1. Molecular phylogeny of Spumellaria inferred from the concatenated complete 18S and partial 28S (D1-D2 regions) rDNA genes (135 taxa and 2459 aligned positions). The tree was obtained by using a phylogenetic Maximum likelihood method implemented using the GTR + γ + I model of sequence evolution. PhyML bootstrap values (100 000 replicates, BS) and posterior probabilities (PP) are shown at the nodes (BS/PP). Black circles indicate BS ≥ 99% and PP ≥ 0.99. Hollow circles indicates BS ≥ 90% and PP ≥ 0.90. Sequences obtained in this study are shown in bold. Thirteen main clades are defined based on statistical support and morphological criteria (A, B, C, D, E, F, G, H, I, J, K, L, M). Main skeletal traits are identified on the right for each spumellarian specimen. Sixteen Nassellaria sequences were assembled as outgroup. Branches with a double barred symbol are fourfold reduced for clarity.

Lineage II is composed of 13 sequences of which 5 were previously available and 8 are novel. All specimens within this clade are characterized by one large and hollow shell where long spines grow from its surface (Fig. 2. F1 and F2), that agree with the definition of the Superfamily Liosphaeroidea (Haeckel, 1882). Members of the subclade F1 (Fig. 2. F1) share a robust shell compared to that of subclade F2 and short spicules are coming out of the spines. Those specimens agree with the definition of the family Astrosphaeridae (Haeckel, 1887; Matsuzaki et al., 2015) in the large hollow and fragile shell with the presence of long or short spicules coming out at the far end of the spines.

Within Lineage III, Clade G is represented by 5 novel and 3 previously available sequences. Members of this clade have a very unique innermost shell, named "rhizosphaerid-type microsphere" with a large and fragile spherical skeleton (Fig. 2. Ga) or several cortical shells with a robust skeleton (Fig. 2. Gb). Either morphology agree with the definition of the family Rhizosphaeridae (Hollande and Enjumet, 1960; Dumitrica, 2017b). Clade H is composed by 1 novel sequence (Ses55, Supplementary material figure S2) and 2 environmental sequences not morphologically described. The picture available for this specimen lacks taxonomic resolution, although based in the overall shape, the characteristic interconnected spines and the highly supported position between clade G and I we believed it agrees with the definition of the family Centrocubidae (Hollande and Enjumet, 1960; emend. Dumitrica, 1994). The last clade of this lineage (Clade I) is represented by 2 novel sequences obtained in this study and 6 environmental sequences. Specimens of this clade show only spongy structures with radiated fibres and there is no evidence of either shell or main spines (Fig. 2. I) not finding any described family with these morphological characteristics.

Figure 2. Scanning Electron Microscopy (SEM) images of Spumellaria specimens used in this study for phylogenetic analysis or morphologically related to one of the morpho-molecular clades of Fig. 1. Letters correspond to its phylogenetic clade in Fig. 1. Scale bars = 50µm. (A) Osh194: *Hexacontium* sp. (specimen not in phylogeny). (Ba) Vil484: *Stigmostylus ferrusi*. (Bb) Mge17-15: *Hexarhizacontium* sp. (C) Vil174: *Hollandosphaera hexagonia*. (D) Mge17-82: *Spongosphaera streptacantha*. (E1a) Vil186: *Cypassis* sp. (E1b) Vil458: *Cypassis irregularis*. (E2a) Osh48: *Schizodiscus*? aff. *japonicus*. (E2b) Mge17-17: *Spongocyclia elliptlica*. (E3) Vil246: *Dictyocoryne koellikeri*. (F1) Osh34: *Heliosphaera bifurcus*. (F2) Vil499: *Liosphaera* sp. A. (Ga) Vil240: *Haliommilla capillacea*. (Gb) Mge17-81: *Rhizosphaera trigonacantha*. (I) Vil210: *Spongoplegma*? Sp. (J1) Vil441: *Flustrella arachnea*. (J1') detail of Vil441. (J2) Osh90: *Perichlamydium perichlamydium*. (K) Vil438: *Actinomma trinacrium*. (L1) Mge17-118: *Trilobatum* aff. *acuferum*. (L1') detail of Mge17-118. (L2a) Osh191: *Schizodiscus* sp. (specimen not in phylogeny). (L2b) Vil296: *Calcaromma morum*. (M1) Vil200: *Tholomura transversara*. (M2) Osh16: *Larcopyle buetschli buetschli*. (M4a) Mge17-20: *Tetrapyle octacantha*. (M4b) Vil452: *Tetrapyle octacantha*.

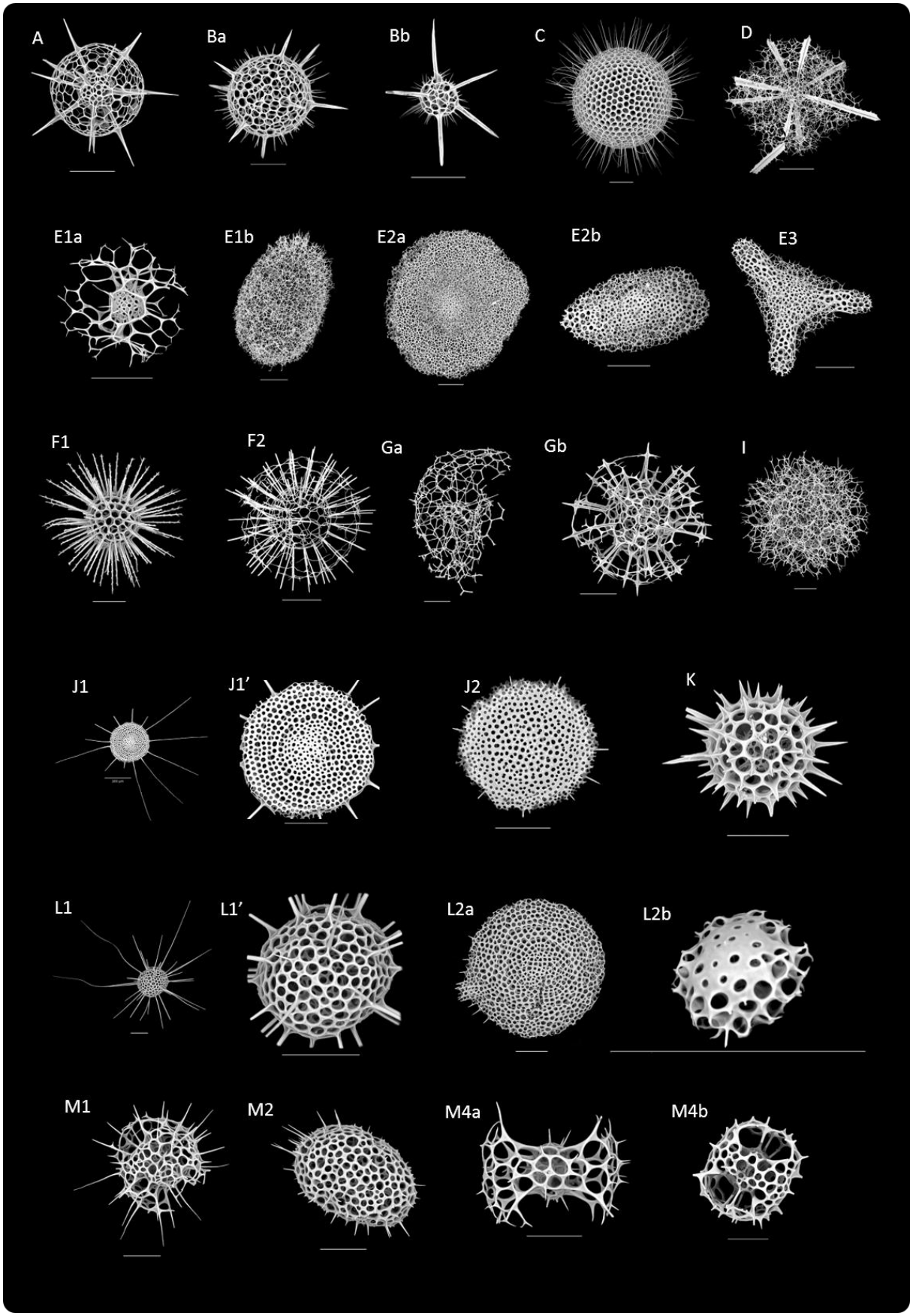


Figure 2. Legend on previous page.

In Lineage IV, Clade J, K and L cluster together with relatively high BS (88) and PP (0.98) values as a sister group of clade M. Although relationships between clades J, K and L remain still elusive due to the short phylogenetic distance and the different topology when using a Maximum Likelihood (Fig. 1) or a Bayesian (Fig. 3) approach (BS: 49 between clade K and L and PP: 0.38 between clade J and K). All members of this Lineage share a very small spherical inner most shell with four or more radial beams connecting to the second shell, where 12 or more radial beams come out (Fig. 2. J-M). Differs from Lineage I in the absence of discrete radial beams from the second or later shells. Clade J is characterised by 5 sequences obtained in this study and 3 previously available sequences. Members of this clade share a flattened skeleton with 2 or more concentric shells and radial spines from the innermost or second innermost shell (Fig. 2. J1, J1' and J2), agreeing with the definition of the Superfamily Stylodictyoidea (Haeckel, 1882; Matsuzaki et al., 2015). Morphological differences between subclade J1 (Fig. 2. J1) and J2 (Fig. 2. J2) were not possible to be determined, yet they have been separated into two different sub-clades due to high BS (96 in J1 and 100 in J2) and PP (0.98 in J1 and 1 in J2) values. Clade K contains 1 novel and 4 previously available sequences. All specimens of this clade have three spherical shells with more than eight main spines that extend from the inner shell (Fig. 2. K), representing the Superfamily Actinommoidea (Haeckel, 1862). Clade L is represented by 5 novel and 14 previously available sequences. Members of this clade are morphologically distant. Subclade L1 has a tightly concentric or coiled centre with one thick protoplasmic pseudopodium (axopodium) (AB61787) or a rhomboidally inflated centre (Fig. 2. L1 and L1'; Mge17-118) which is not observed in any other Spumellaria. In the first case, this morphology is attributed to Sponguroidea, and in the second to Lithelioidea. Subclade L2 is either marked by a flatten circular test with a tunnel-like pylome and a test comprised by a very densely concentric structure (Fig. 2. L2a) attributed to Spongodiscoidea or by a spherical translucent protoplasm with an encrypted flat consolidated skeletal shell in distal position and star-like solid soluble materials (Fig. 2. L2b and Supplementary material Fig. S2) assigned to Collodaria. In Clade M there are 13 novel sequences and 7 previously available sequences. A morphological characteristic of this clade is the broken, or fenestrated, outermost shell with the presence of pyloniid central structure (Fig. 2. M), corresponding to the Superfamily Pylonioidea (Haeckel, 1882), thereafter, the symmetry and the different opening of the outermost shell distinguish the subclades. Yet, phylogenetic relationships within this clade remain elusive. The first subclade (M1) consists of three basal sequences poorly supported as a clade yet their cubic morphology and the two opposite and closed gates (Fig. 2. M1) agrees with the family Tholoniidae (Haeckel, 1887). The three other subclades have different morphologies (Fig. 2. M2; Ses13, Ses53 and Vil437, Supplementary material Fig. S2) agreeing with the definition of different families within Pylonioidea. Subclade M2 (Fig. 2. M2) has a spherical to ellipsoidal skeleton without distinctive openings. On the other side, subclades M3 and M4 are highly supported as a group with a flatten box-shaped skeleton for the former subclade (Ses13, Ses53 and Vil437, Supplementary material Fig. S2) and a cubic skeleton characterized by big openings of the second shell for the later subclade (Fig. 2. M4a and 4b).

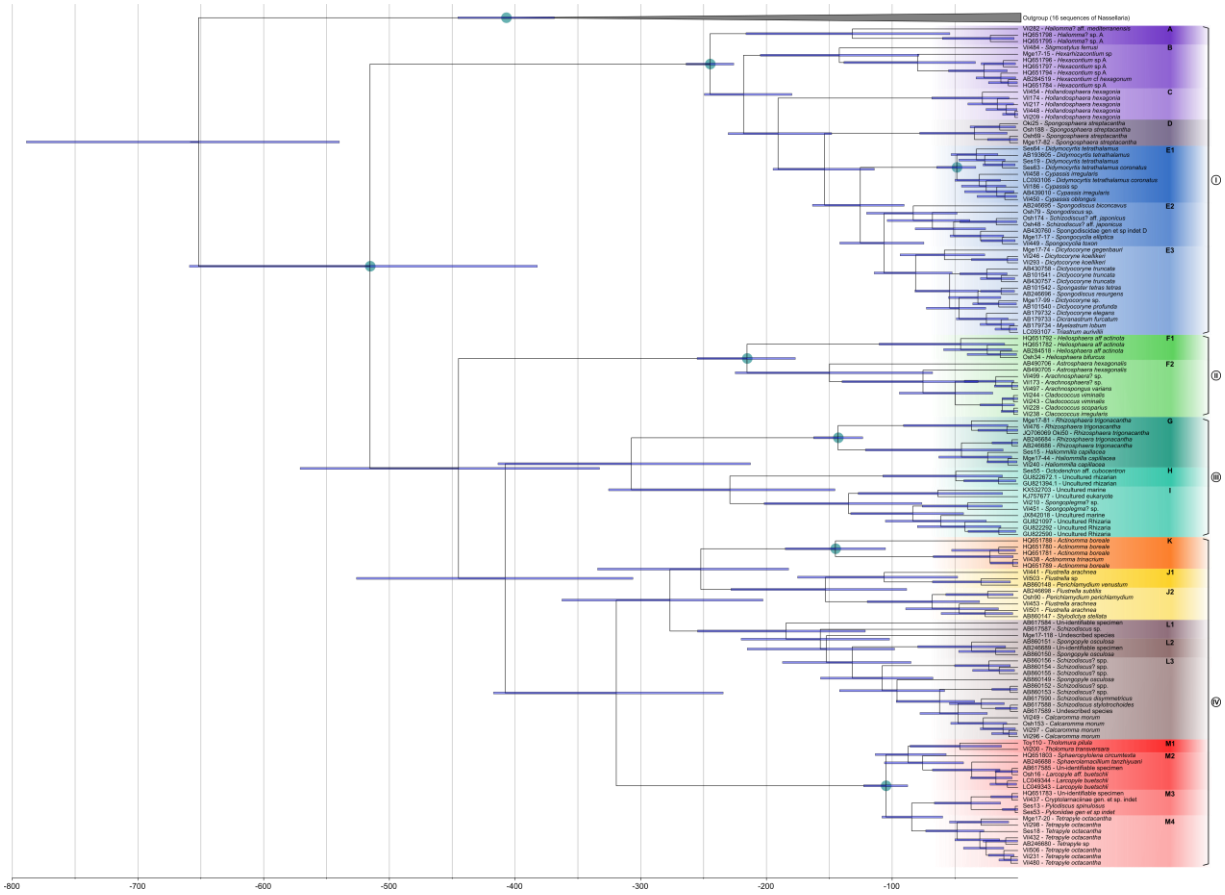


Figure 3. Time-calibrated tree (Molecular clock) of Spumellaria, based on alignment matrix used for phylogenetic analyses. Node divergences were estimated with a Bayesian relaxed clock model and the GTR + γ + I evolutionary model, implemented in the software package BEAST. Nine different nodes were selected for the calibration (blue dots). Blue bars indicate the 95% highest posterior density (HPD) intervals of the posterior probability distribution of node ages.

Molecular dating

The molecular clock dated the diversification between Nassellaria and Spumellaria (the root of the tree) at a median value of 651 Ma (95% Highest Posterior Density -HPD-: between 789 and 540 Ma) (Fig. 3). From here on, all dates are expressed as median values followed by the 95% HPD interval. The first diversification of Spumellaria happened at around 515 (659-382) Ma, followed by a rapid branching of the lineages. Despite their dubious phylogenetic relationships, Lineage II splits apart 445 (571-332) Ma and Lineage III and IV diversified from each other 407 (526-306) Ma. The next diversification events correspond to the first radiation of Lineage IV 319 (417-234) Ma and that of Lineage III 307 (414-213) Ma. Within Lineage IV, the phylogenetic relationships are doubtful, yet two radiation events at 277 (363-203) Ma and 252 (334-182) Ma separate Clades J, K and L. Clade L is the first of these clades diversifying at 184 (255-121) Ma, followed by clade J 153 (228-88) Ma, clade K at 145 (185-105) Ma and Clade M at 105 (123-88) being the youngest clade of Lineage IV. Lineage III diverged soon after Lineage IV, yet the next ramification was between Clade G and Clade I 229 (326-145) Ma. It was not until 143 (162-123) and 135 (202-76) Ma when Clade G and I respectively radiate, and the latest diversification within this Lineage corresponds to Clade H 49 (107-12) Ma. Lineage I appeared 245 (264-226) Ma, and thereafter is characterized by a series of relatively

continuous diversification events until the radiation of Clades B, A, E, D and C at 142 (205-79), 131 (216-54), 153 (163-90), 34 (78-8) and 28 (68-7) Ma respectively. Regarding Lineage II, is the youngest of all the lineages appearing at 215 (255-177) Ma despite its early branching from any other lineage.

Post-hoc analyses

The lineages through time analysis (Fig. 4) shows a classic exponential diversification slope, with a 0.0102 rate of speciation (Ln(lineages)·million years^{-1}). The first diversification of extant Spumellaria corresponds to an early and fast divergence of the different lineages from ~515 to ~407 Ma. After that, the diversification of living groups remained standstill until ~313 Ma when Lineage IV and III radiated. From ~276 to ~215 Ma all different groups diversified: Clades K, J and L (within Lineage IV) split apart; Lineage I appears and starts a continued diversification; within Lineage III the three clades split apart; and Lineage II emerge. Another important diversification event happens from ~157 to ~125 Ma when Lineage IV greatly diversifies, followed by Lineage I and II. During the following years there is a tiered diversification, appearing Clade M in isolated events as well as the ramification of Clade L or the branching between subclades E2 and E3. Finally, from ~54 Ma on wards there is a continuous diversification where the rest of the clades appear or keep diversifying. During this time lineage III diversifies notably and so does lineage II.

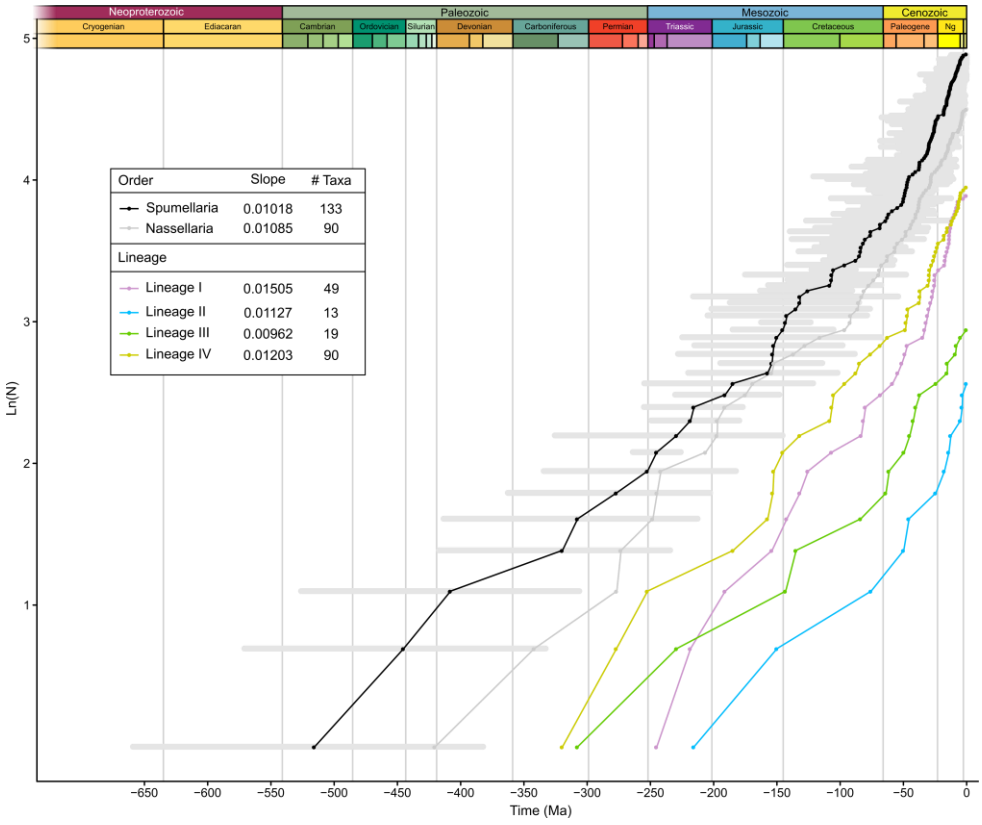


Figure 4. Lineages Through Time (LTT) analysis based on the molecular clock results for Spumellaria (removing the outgroup; in black), of each lineage independently: lineage I (purple), lineage II (blue), lineage III (green) and lineage IV (yellow) and of Nassellaria (data taken from Sandín et al., 2019, Chapter 1.1). The y-axis represents the number of lineages (N) expressed in logarithmic (base e) scale (Ln(N)) and in the x-axis it is represented the time in million of years ago (Ma). Horizontal grey bars in black slope represent the 95% Highest Posterior Density (HPD) of molecular clock estimates.

The ancestral state reconstruction analysis (Supplementary Material Fig. S3) establish a spherical skeleton shape for the common ancestor to all Spumellaria, with a central structure that can be either empty (state 0), with filaments, rings or a spicular system (state 1) or with a double medullary shell (state 2), but without spines arising from it, one distinctive cortical shell and concentric internal structures outside the central part. The common ancestor of Lineage I and that of Lineages II, III and IV share a spherical skeleton shape, a central structure similar to the common ancestor to all Spumellaria, with one cortical shell and concentric structures outside the central part. The difference between these two ancestors is that the ancestor of lineage I is characterized by more or less 6 radial spines and beams arising from the central structure, whereas that of lineages II, III, IV are variable, from no radial beams in Lineage II to more than eight radial beams (Lineage IV) arising from the central structure in Lineage IV and six or more fibrous radial beams from the central structure in Lineage III. The common ancestor of lineage III and that of lineage IV have also a similar morphology with the exception of the skeleton shape, where for that of lineage III is spherical but for that of lineage IV is either spherical, cylindrical or ellipsoidal or cubic. The central structure in these ancestors is undefined since it has equal probabilities every character state.

Environmental genetic diversity of Spumellaria

A total of 1171 18S rDNA and 6 28S rDNA environmental sequences affiliated to Spumellaria were retrieved from NCBI public database and placed in our reference phylogenetic tree (Fig. 5). Of the 6 28S rDNA sequences, 4 were also found in the 18S rDNA survey, and therefore removed to avoid duplicates, as for the 8 sequences already included in the phylogeny (2 in clade H and 8 in clade I). From the 1165 final environmental sequences, 617 were closely related to Clade F. Two other clades with a high number of environmental sequences related to are Clade I and Clade M, with 145 and 47 sequences respectively. While up to 40 sequences were scattered between clades E, L, B, G, D and J (17, 14, 3, 3, 1 and 1 respectively). Whereas no environmental sequences were clustered within clades A, C, H and K. The rest of the sequences (317) where distributed across the tree, mainly at basal nodes, with no possible assignation: 206 basal to clade F in lineage II, 70 in Lineage I (of which 41 basal to the lineage, 11 basal to clades C, D and E, 12 basal to clades D and E, 6 basal to clade B and 1 basal to clade C), 26 in lineage IV (of which 16 basal to clade J, 5 basal to clade M and 5 basal to the lineage), 4 in lineage III basal to clade G, 2 basal to lineages II, III and IV and 8 final sequences basal to all Spumellaria.

To go beyond rapid phylogenetic placement using the pplacer tool, these 317 environmental sequences were later included in a phylogenetic tree of the 18S rDNA gene (RAxML v8.2.10, GTR+G+I, with 1000 rapid bootstraps, Stamatakis, 2014) after removing chimeras (52 sequences were detected as chimeric). Up to 6 different clades were found mainly at basal positions (Fig. 6), 14 different sequences were scattered over the phylogenetic tree and 2 last sequences were clustered with high support basal to Clade M. (further details of sequence assignations can be found in Supplementary material Table S3). From these clades, Env5 gathers 152 sequences related to clade F in lineage II, yet low boostrap values (51) avoid a formal assignation. Two clades (Env1 with 63 sequences and Env2 with 7) are found at basal positions of lineage I. Another environmental clade appears in lineage IV (Env4

with 16 sequences) at basal positions. And two final clades are found basal to lineages I, III and IV (Env6 with 8 sequences) and basal to lineage IV (Env3 with 3 sequences).

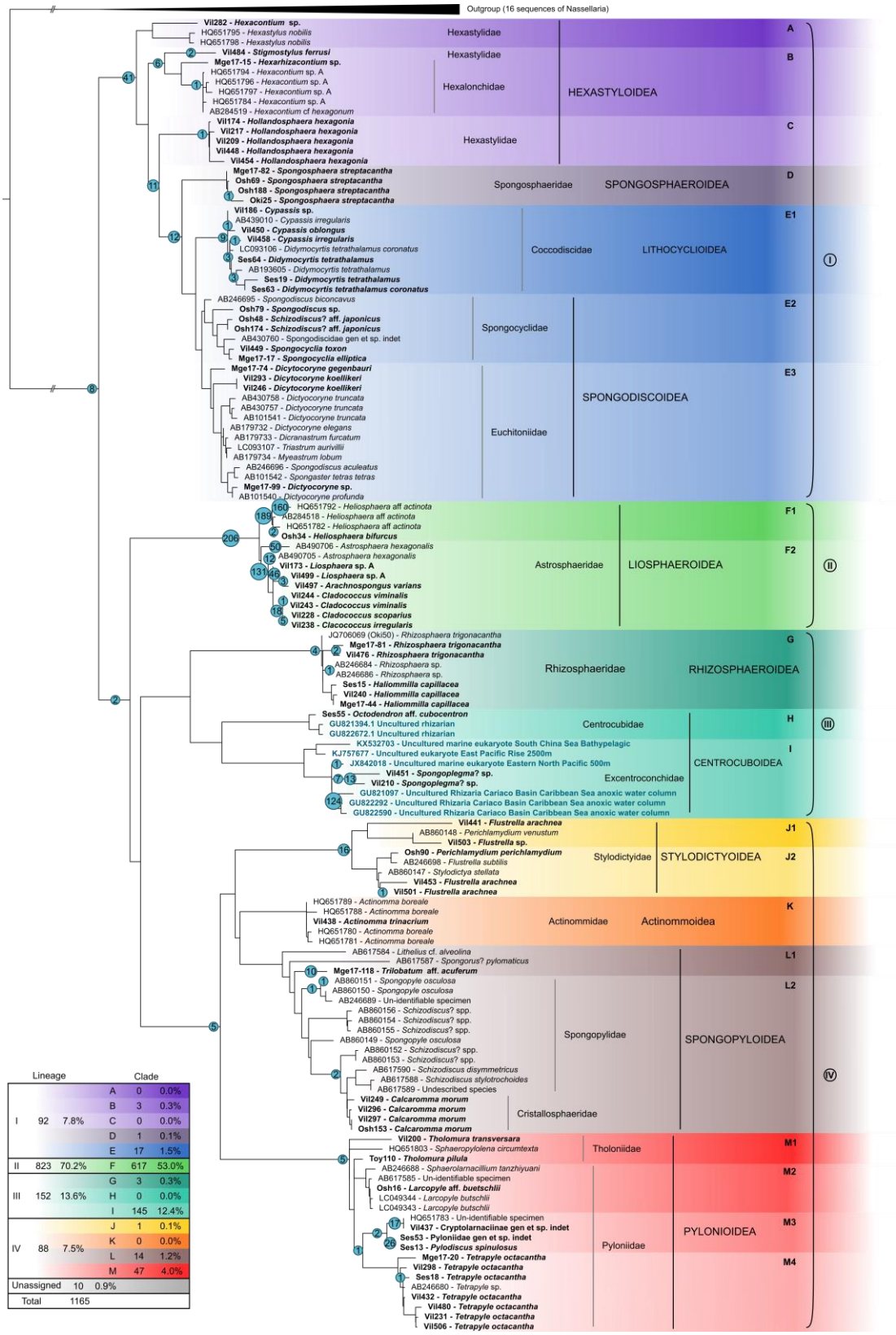


Figure 5. Pplacer phylogenetic placement of 1171 environmental sequences into a concatenated phylogenetic tree of Spumellaria (complete 18S + partial 28S rDNA genes). Numbers at nodes represent the amount of environmental sequences assigned to a branch or a node.

In order to improve the understanding of the high chimeric ratio and the diversity of the environmental clades, we performed a phylogenetic analysis (RAxML v8.2.10, GTR+G+I, with 1000 rapid bootstraps, Stamatakis, 2014) including those sequences that were removed from principal analysis believed to be contaminations. From these sequences, 2 were removed finally considered as contaminations, and up to 9 sequences clustered with moderate to high support among one of the environmental clades (Env1; Supplementary Material Fig. S4; Supplementary material Table S2). They were previously considered as contaminations due to the contrasting morphology in the clustered basal clade. In addition, one of these specimens (Osh174) is also found within clade E. Yet in clade E it is composed by the concatenation of the first part of the 18S rDNA and the 28S rDNA gene (1489 bp in total) and that of clade Env1 by the second part of the 18S rDNA gene (871 bp).

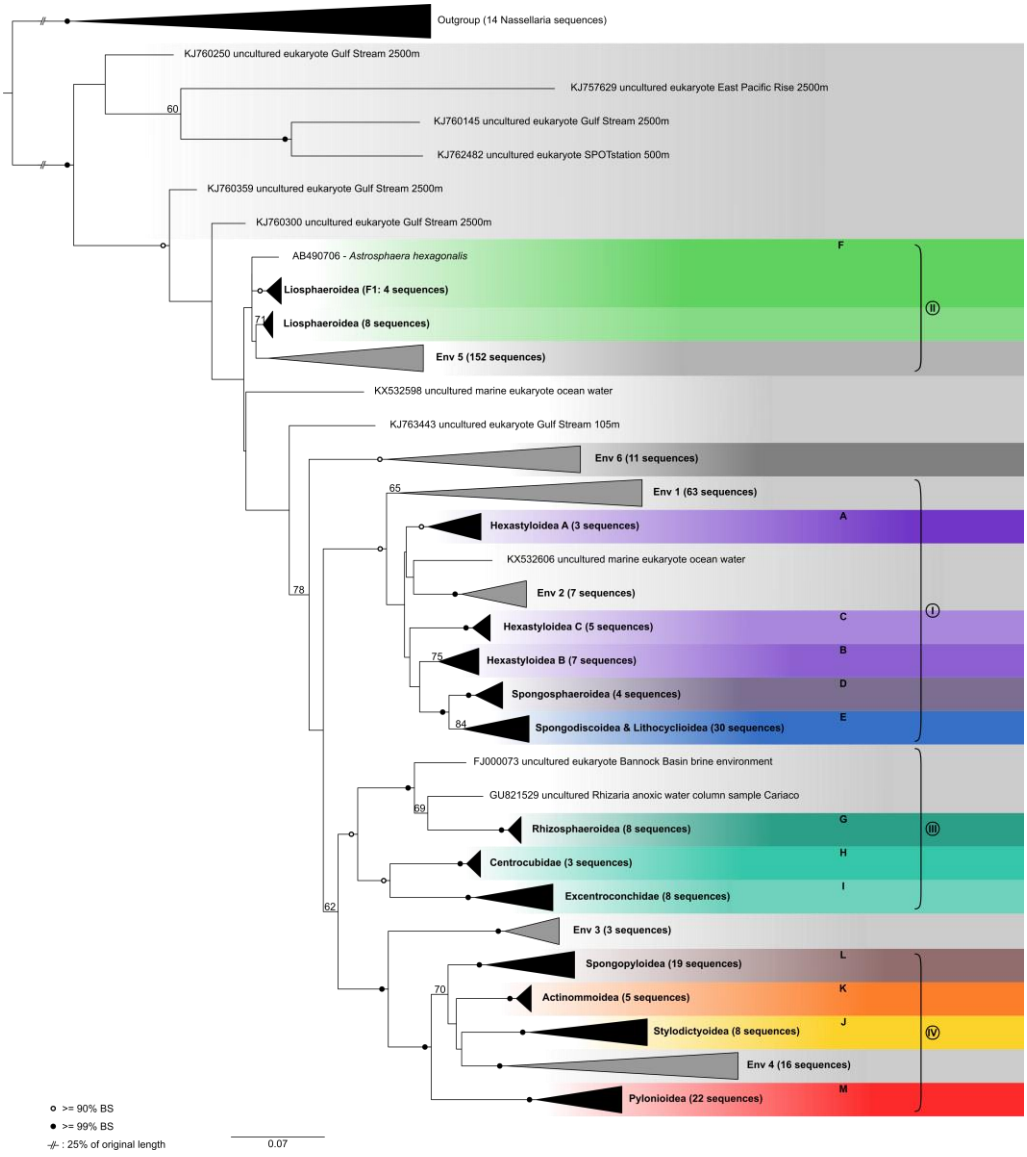


Figure 6. Molecular phylogeny of environmental sequences associated to Spumellaria. The tree was obtained by using 1000 rapid bootstrap RAxML GTR + γ + I model of sequence evolution. Black circles at nodes indicate BS ≥ 99% and hollow circles indicates BS ≥ 90%. BS lower than 60 were removed for clarity. Branches with a double barred symbol are fourfold reduced for clarity. Black triangles represent clades which morphology is known (Fig. 1, Fig. 2). Grey triangles represent environmental clades. Sequence composition of each environmental clade are described in Table S3.

Discussion

Morpho-molecular classification of living Spumellaria

Despite the recognized importance of internal structures in previous radiolarian classifications the overall shape have always played a significant role in higher rank classifications (De Wever et al., 2001; Afanasieva et al., 2005). Our study shows that the symmetry of the skeleton is a key feature describing spumellarian phylogeny, as for other radiolarian groups such as Acantharia (Decelle et al., 2012b) or Nassellaria (Sandin et al., 2019, Chapter 1.1). Although central structure patterns are well defined between clades helping to discern similar symmetric trends in distantly related clades (e.g. flat symmetry and the superfamilies Spongodiscoidea, Spongopyloidea and Stylodictyoidea).

Extant Spumellaria included in our study can be divided in 13 morpho-molecular clades, grouped in four main evolutionary lineages based on phylogenetic clustering support, common morphological features and molecular dating: Hexastyloidea, Spongosphaeroidea, Lythocyclioidea and Spongodiscoidea in lineage I; Liosphaeroidea in lineage II; Rhizosphaeroidea and Centrocuboidea in Lineage III and Actinommoidea, Stylodictyoidea, Spongopyloidea and Pylonioidea in lineage IV. Our results confirm the monophyly of certain families (e.g., Liosphaeroidea, Rhizosphaeridae, Actinommoidea or Pylonioidea) yet other groups have shown to be polyphyletic such as Hexastyloidea in clades A, B and C or clades E and L previously classified under Spongodiscoidea. The former superfamily have been classified independently in previous morphology based taxonomic studies (De Wever et al., 2001; Afanasieva et al., 2005). On the contrary, flat spumellarians have been classified all together on the basis of its flat symmetry, yet it has been stated the very likely polyphyletic nature of this group (De Wever et al., 2001). Lineage III gathers very different skeleton shapes, previously considered as Entactinaria. The two superfamilies constituting Lineage III, Rhizosphaeridae and Centrocuboidea have been grouped together based in common cytological structures in previous studies along with Actinommoidea and Hexastyloidea, two groups placed at basal positions in their respective lineages (Hollande and Enjumet, 1960). The general agreement between morphology-based taxonomy and molecular classification along with the amount of exceptions at a higher taxonomic level, leads to a better understanding of spumellarian phylogeny. Although further analysis must reveal the position of missing families (i.e. Heliodiscidae, Suttoniidae, Ethmosphaeridae, Sphaerostylidae and Saturnalidae) for a proper understanding of spumellarian and radiolarian diversity and evolution. However, these families are represented by very few species, probably emended within a clade presented herein (i.e.; clade L gathers 3 different families) or representing an environmental clade (i.e.; Env 2, Env 4, Env 5).

The morpho-molecular framework established herein understands the Spumellaria as Polycystines radiolarians with concentric structure and a spherical or radial symmetry. This concept includes some living groups classified under the Order Entactinaria in De Wever et al. (2001) already pointed out by Yuasa et al. (2009) and Ishitani et al. (2012) and corrected in Matsuzaki et al. (2015) such as Hexastyloidea. Our results also suggest the inclusion of the Entactinaria families Centrocuboidea and Rhizosphaeridae within Spumellaria. The Radiolaria orders Spumellaria and Entactinaria were traditionally separated based on the absence or

presence of internal structures respectively (De Wever et al., 2001). Our results show the Entactinaria order, in the sense of De Wever et al. (2001), is scattered into Clades A, B, C and lineage III. Recent advances in imaging techniques have suggested the Spumellaria as a suborder within the Entactinaria based on findings of homologous structures with the initial spicular system in Spumellaria among primitive radiolarian forms (Kachovich et al., 2019). Although our results agree with Kachovich et al. (2019), we consider Entactinaria as a polyphyletic group within the Rhizaria (Nakamura et al. submitted, annexes) and Spumellaria as the name of these shapes due to the use of internal structures at a lower taxonomic rank and the prevalence of the overall shape at higher taxonomic rank. In addition to our conclusions, there are no genetic evidences of an entactinarian clade so far. Therefore, it is possible that other groups sharing these symmetric patterns might also be emended within the Spumellaria, as we have seen with the family Centrocubidae. Yet we doubt the existence of Entactinaria as a monophyletic group, we do not argue the suitability on the use of this group from a biostratigraphic point of view due to the differences in appearance and extinction times along the fossil record.

Evolutionary history of Spumellaria

Our molecular clock results dated the first diversification of Spumellaria at ~515 (HPD: 659-382) Ma, in agreement with recent findings of the oldest radiolarian fossils with spherical forms (Early Cambrian, ca. 520 Ma), taxonomically classified in the extinct spumellarian genus *Paraantygopora* (Zhang and Feng, 2019). This early diversification probably separated the two contrasting lineages I and IV, followed soon afterwards by the branching of linages II and III, during the so-called Great Ordovician (485.4 – 443.8 Ma) diversification event (Noble and Danelian, 2004; Servais et al., 2016b). At this moment many different symmetrically spherical radiolarians appeared in the fossil record (Aitchison et al., 2017; Noble et al., 2017). Yet, poor radiolarian-baring rocks (Aitchison et al., 2017), different topologies of the different marker genes and the short phylogenetic distance at these nodes in the concatenated phylogenetic analysis make relationships between lineages obscure. From this period until the end of the Paleozoic the diversification of living lineages is low and basically restricted to the end of the Carboniferous (358.9 – 298.9 Ma) when lineages III and IV start diversifying.

All over the Palaeozoic, extinct specimens previously attributed to Entactinaria show a great diversity that most of it become extinct towards the end of this Era and during the beginning of the Mesozoic (De Wever et al., 2003, 2006). Some representatives of this group have been already tentatively considered as Nassellaria in latest classifications (Noble et al., 2017) and later suggested as ancestors of this order due to overall morphology and molecular clock results (Sandin et al., 2019, Chapter 1.1). Probably long branches of clades belonging to lineages II, III and IV of the phylogenetic tree followed by a sudden diversification may reflect a bottleneck effect, in which many groups got extinct before radiation. Understanding some of the previously considered entactinarian families as possible candidates for such extinct groups.

Thereafter at the beginning of the Mesozoic (251.9 – 66 Ma) the first living representatives appear with the diversification of Lineage II (Liosphaeroidea) and Hexastyloidea. Although, the polyphyletic nature of Hexastyloidea shows a later

diversification within the clades (A, B and C), explaining why the Triassic (251.9 – 201.4 Ma) genera does not resemble that of the Cenozoic (66 – 0 Ma) (O'Dogherty et al., 2011). During this period, Lineage III also diversifies and all clades from lineage IV are already separated from each other, yet their diversifications are not happening until the end of the Mesozoic. Similar diversification patterns at the beginning of the Mesozoic have been reported in living Nassellaria inferred from the combination of the molecular clock and the direct observation of the fossil record (De Wever et al., 2006; Sandin et al., 2019, Chapter 1.1); in which ancient forms, presumably extinct during the end-Permian extinction (Sepkoski, 1981; De Wever et al., 2006; Takahashi et al., 2009; O'Dogherty et al., 2011; Aitchison et al., 2017) led to the diversification of living groups.

The second part of the Mesozoic is characterized by a stepped diversification of Spumellaria, as already described for Nassellaria (Sandin et al., 2019, Chapter 1.1) and Foraminifera (Leckie et al., 2002; Hart et al., 2003). Probably due to the onset of a global oceanic anoxia during the Jurassic (Jenkyns, 1998) and a series of Oceanic Anoxic Events during the Cretaceous (Schalanger and Jenkyns, 1976; Erbacher et al., 1996; Jenkyns, 2010; Yilmaz et al., 2012; Kemp and Izumi, 2014). In contrast to Nassellaria diversification, towards the end of the Jurassic (201.4 – 145 Ma) and beginning of Cretaceous (145 – 66 Ma), there is a big increase in the diversification where most of the clades suddenly diversify, also reported in the fossil record (Kiessling, 2002; De Wever et al., 2003). Such differences between diversification patterns of Spumellaria regarding other radiolarian groups is a question already arisen in recent studies (Kachovich et al., 2019). Both extant Nassellaria and Spumellaria share similar environmental preferences (Suzuki and Not, 2015), yet Spumellaria tend to possess a larger protoplasmic volume (Takahashi, 1982; Des Combes and Abelmann, 2009) probably providing the advantage to thrive during low nutrient availability periods, as known to be the Jurassic (Cárdenas and Harries, 2010). Once the conditions become more favourable for the heterotrophic Spumellaria, they started a sustained diversification that followed all over the Cenozoic (66 – 0 Ma). Similar diversification patterns were found in Nassellaria (Sandin et al., 2019, Chapter 1.1), Foraminifera (Hallock et al., 1991) and Coccolithophores (Bown et al., 2004) linked to climatic oscillations.

Environmental genetic diversity of Spumellaria

The present morpho-molecular framework of Spumellaria allows an accurate phylogenetic placement of environmental sequences for which it would be not possible to infer their taxonomic position otherwise.

Here we placed publicly available environmental sequences from diverse environments (e.g.: Orsi et al., 2011; Kim et al., 2012; Lie et al., 2014; Wu et al., 2014; Xu et al., 2017) closely related to Spumellaria into our phylogenetic database. Most of the environmental sequences (not morphologically described) placed in our reference morpho-molecular framework are highly related to Liosphaeroidea, coming primarily from deep environments (1500-2500m). Another group displaying a large environmental diversity is the family Excentroconchidae, mainly represented by sequences coming from anoxic environments. These findings contrast the data obtained by morphological observations from the fossil record (De Wever et al., 2001) or from plankton and sediments traps and surface sediment materials (Boltovskoy et

al., 2010; Boltovskoy and Correa, 2016a), in which both the Liosphaeroidea and the Excentroconchidae were never among the most abundant groups. Members belonging to these groups have a thin and fragile skeleton, and probably it dissolves fast or break apart in the sediments, avoiding a proper preservation of the skeleton and therefore identification. Besides, living specimens have a compact and dense protoplasm making sometimes hard to recognize the skeleton. Probably this feature led to the misidentification or overlook of such specimens in previous morphological-based surveys of Spumellaria.

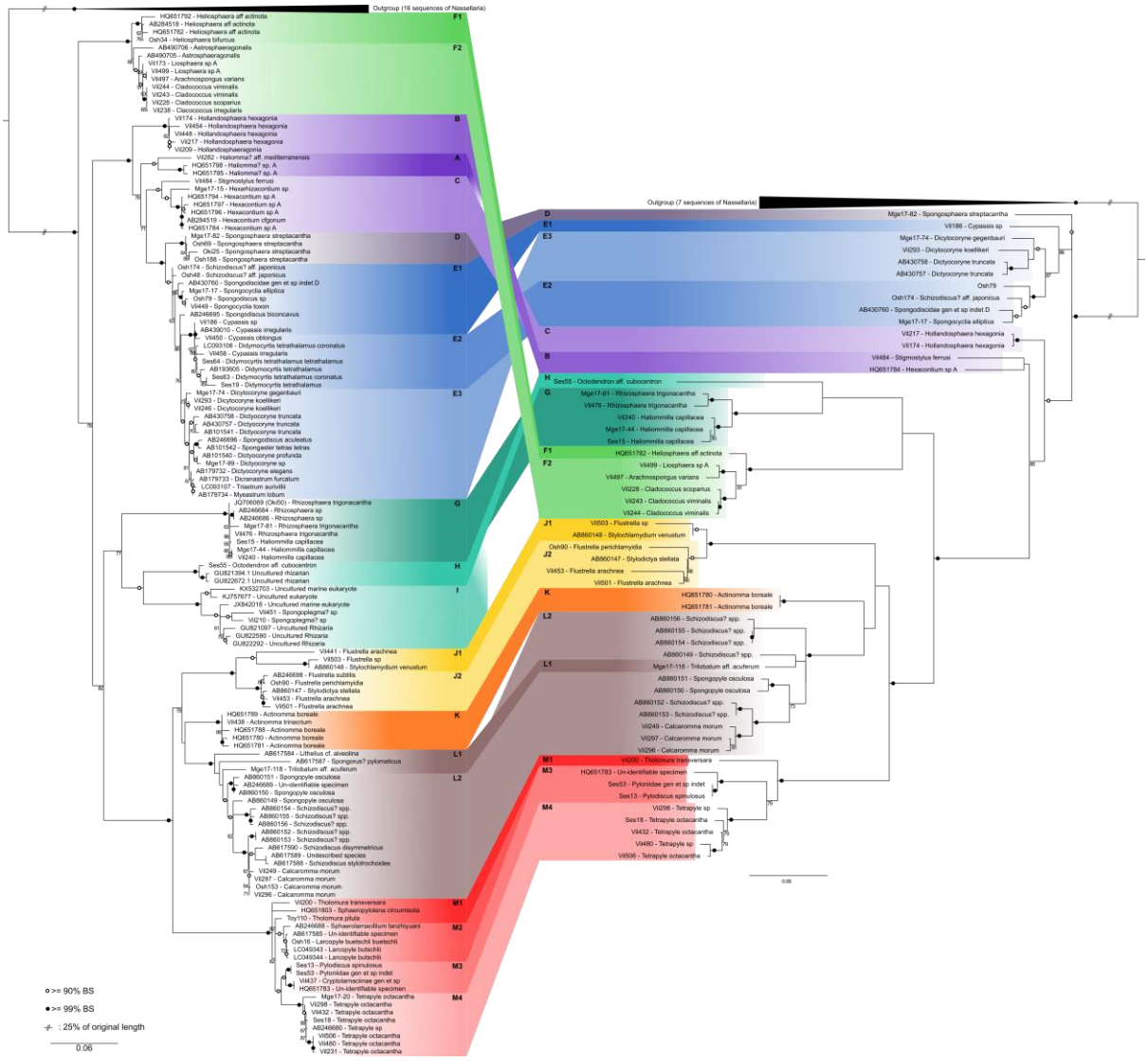
The great environmental diversity at early diverging positions (i.e.; Env 1, Env 3, Env 6) and the few possible candidates in the fossil record prevent a deep understanding of the extant spumellarian morpho-molecular diversity. Furthermore, it questions the reliability of such molecular clades and its diversity in the environment. In our analysis, various lines of evidences led us to hypothesise the existence of two different and phylogenetically distant 18S rDNA gene copies within the same specimen: First, the notably high ratio of chimeric sequences obtained (in particular for specimens related to lineage I), the unexpected phylogenetic position of different morpho-species within the same clade, the presence of the same individual in 2 different and highly supported clades and the common relatively long branches within the supposed clade (Env1: Supplementary Material Fig. S4). Given the extend of symbiotic relationships within the SAR supergroup (Stoecker et al., 2009, 2016; Bjorbækmo et al., 2019), the most likely explanation left for Env 1 is the existence of a naked symbiotic group of Spumellaria developing within other, skeleton-bearing, Spumellaria. The symbiotic nature (parasitic, mutualism, …) of such group remains beyond the scope of this study, although further molecular or imaging analysis (e.g. fluorescence *in situ* hybridization) must reveal this issue. Similar associations have been already identified between benthic bathyal foraminifera and dead tests of Xenophyophorea (Foraminifera) (Pawlowski et al., 2003b; Hughes and Gooday, 2004). Besides, to our knowledge the few molecular data available of marine Heliozoa have appeared to be polyphyletic (Cavalier-Smith et al., 2015; Burki et al., 2016), as for the fresh water groups (Nikolaev et al., 2004). Indeed, some groups of Heliozoa, such as Gymnosphaerida, have been moved to the Retaria lineage (precisely within Radiozoa) in based of ultrastructural features of the axopodial complex (Yabuki et al., 2012). Therefore, it is possible that some of these basal environmental clades are represented by heliozoan-like organisms, and some of them (clade Env1) developed symbiosis with other Spumellaria.

Acknowledgments

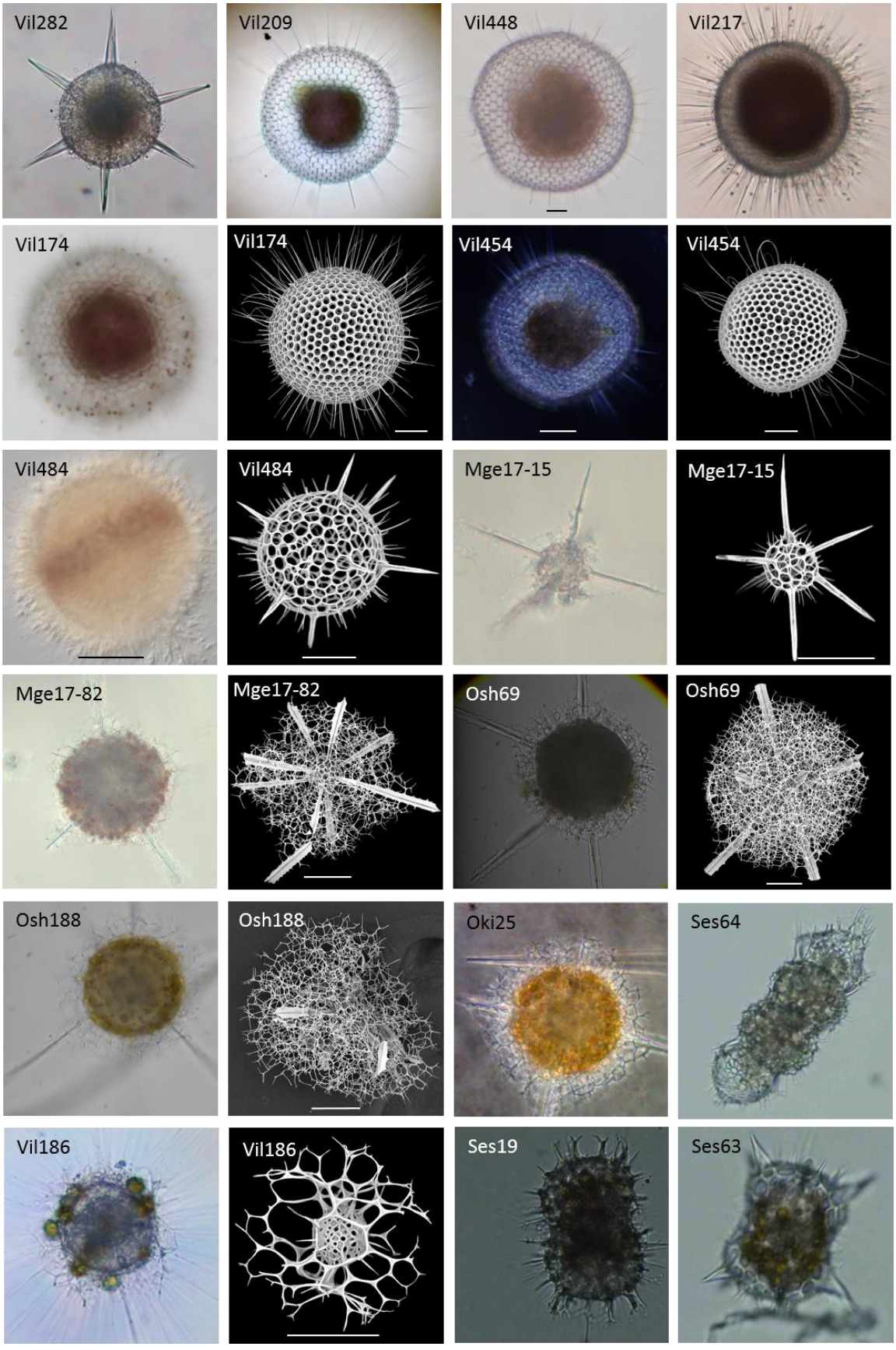
This work was supported by the IMPEKAB ANR 15-CE02-0011 grant and the Brittany Region ARED C16 1520A01, the Japan Society for Promotion of Science KAKENHI Grand No. K16K0-74750 for N. Suzuki and “the Cooperative Research Project with the Japan Science and Technology Agency (JST) and Centre National de la Recherche Scientifique (CNRS, France) “Morpho-molecular Diversity Assessment of Ecologically, Evolutionary, and Geologically Relevant Marine Plankton (Radiolaria)”. We are grateful to the CNRS-Sorbonne University ABiMS bioinformatics platform (http://abims.sb-roscoff.fr) for providing computational resources. We would like to thank the MOOSE cruise and program for the opportunity of sampling and the facilities given on board, as well as John Dolan for hosting us multiple times at the Laboratoire d’Océanographie of Villefranche sur Mer. We are greatly thankful to Cedric

Berney for the phylogenetic advises and the valuable help on the interpretation of the “symbiotic” clade, as well as Vasily Zlatogursky for his contributions and feedback on the heliozoan-like organism.

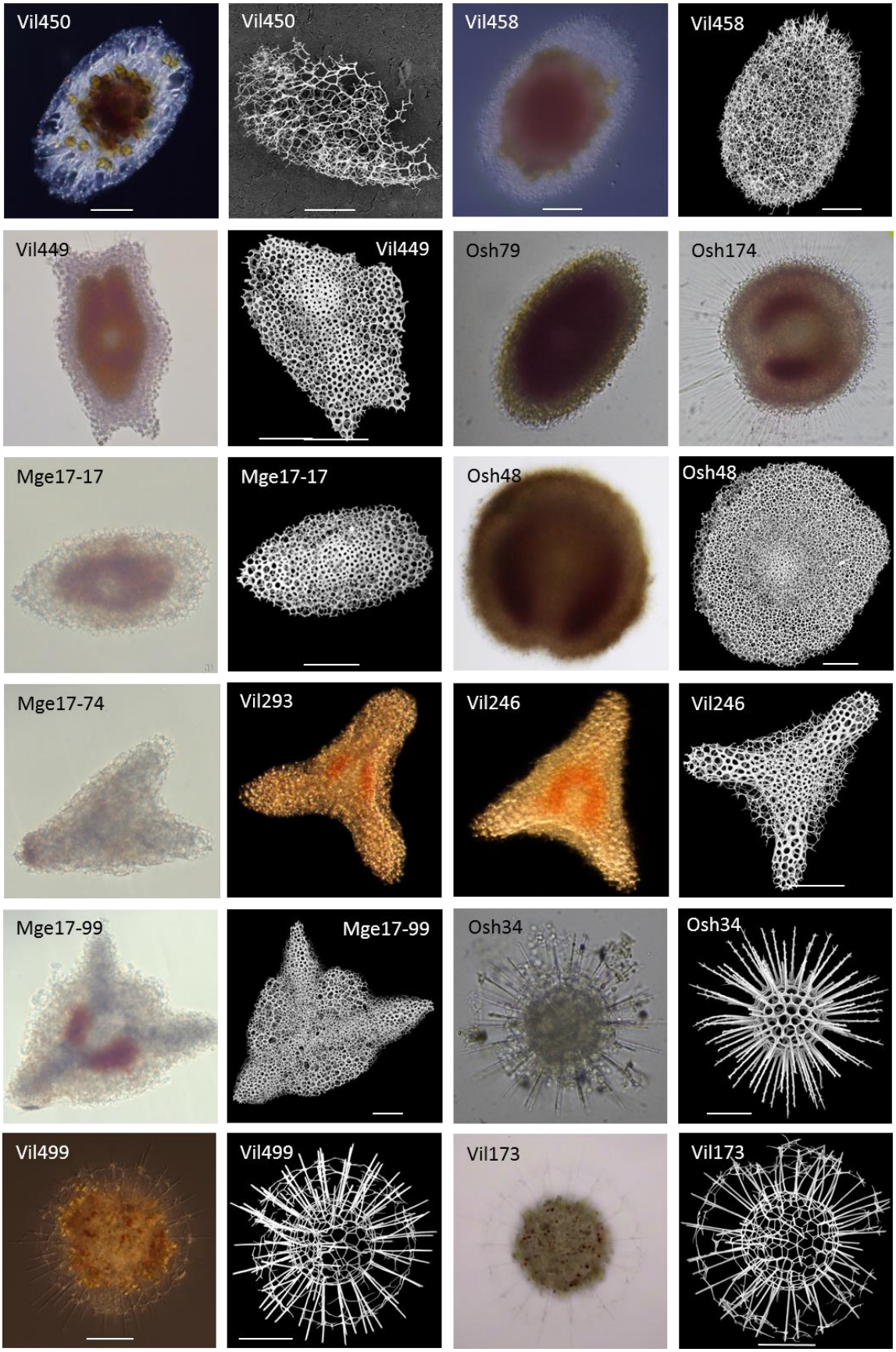
Supplementary Material



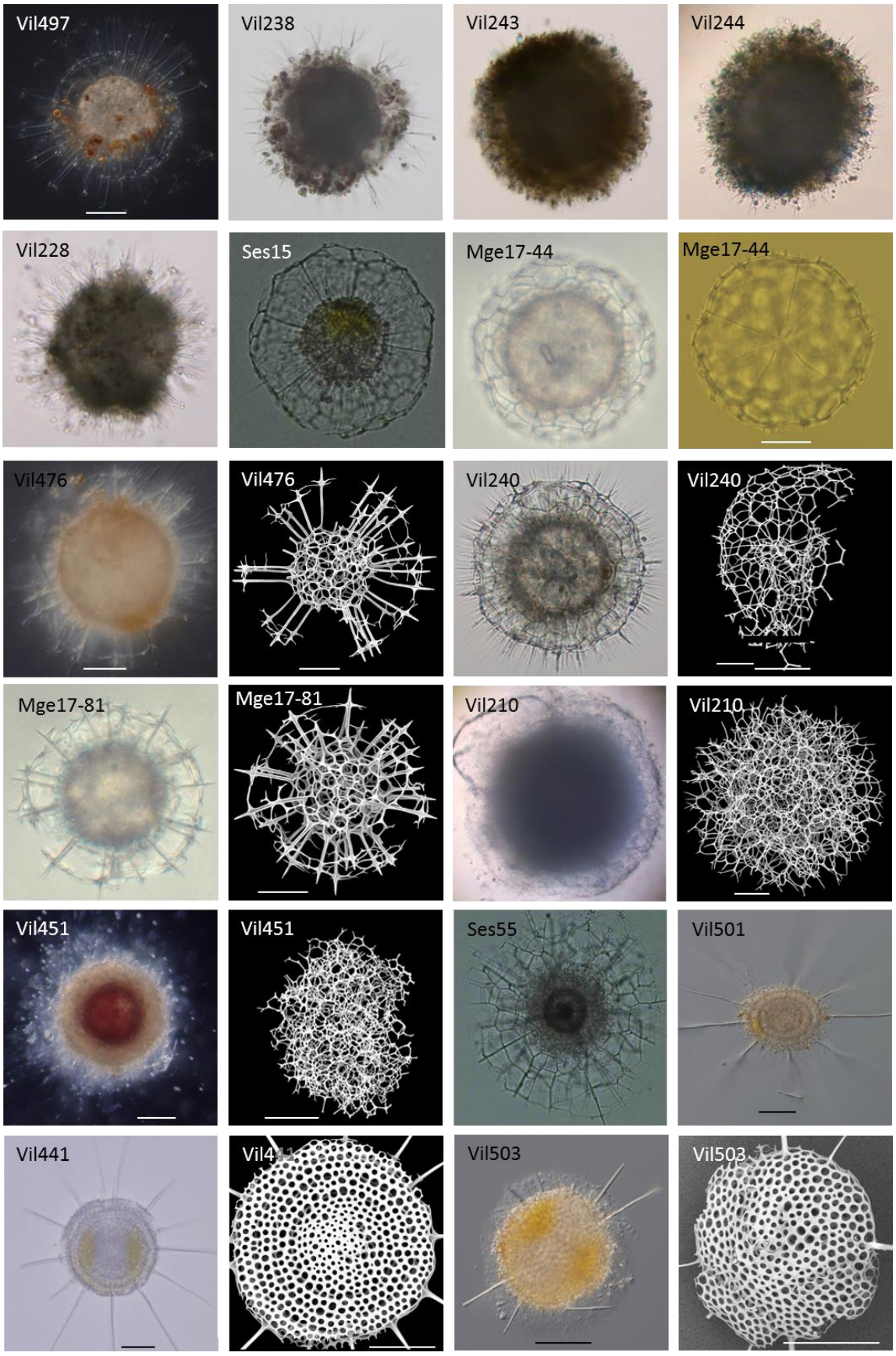
Supplementary Figure S1. A compared molecular phylogeny of Spumellaria inferred from the 18S (on the left) and partial 28S (D1-D2 regions, on the right) rDNA genes. Trees were obtained by using a phylogenetic Maximum likelihood method implemented using the GTR + γ + I model of sequence evolution. PhyML bootstrap values (100 000 replicates, BS) are shown at the nodes (BS). Black circles indicate BS ≥ 99%. Hollow circles indicate BS ≥ 90%. Branches with a double barred symbol are fourfold reduced for clarity.



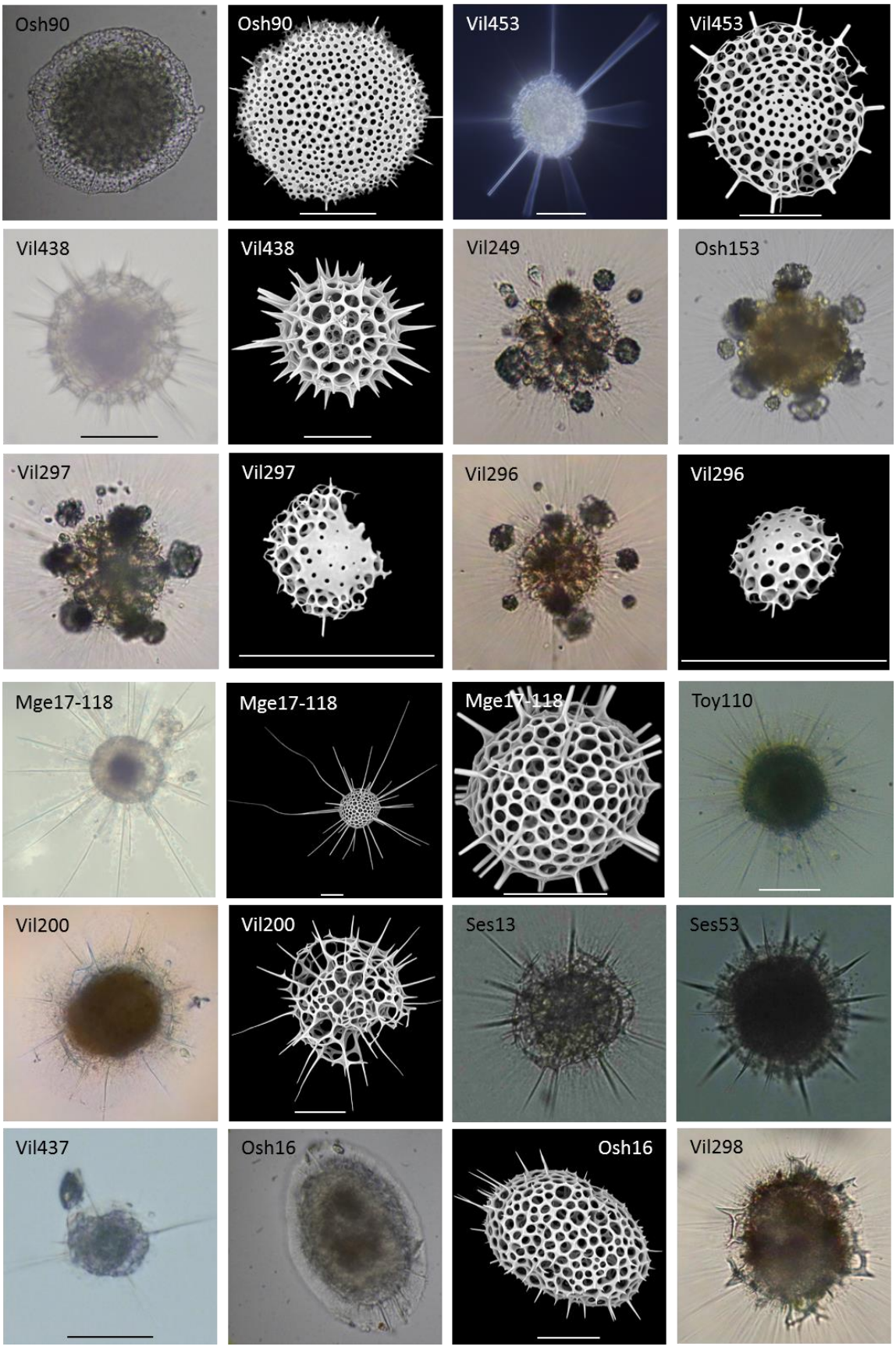
Supplementary Figure S2. Light Microscopy (LM) and Scanning Electron Microscopy (SEM, when available) images of live Spumellaria specimens used in this study for phylogenetic analysis in Fig. 1. Scale bars (when available) = 50µm. Part 1/5



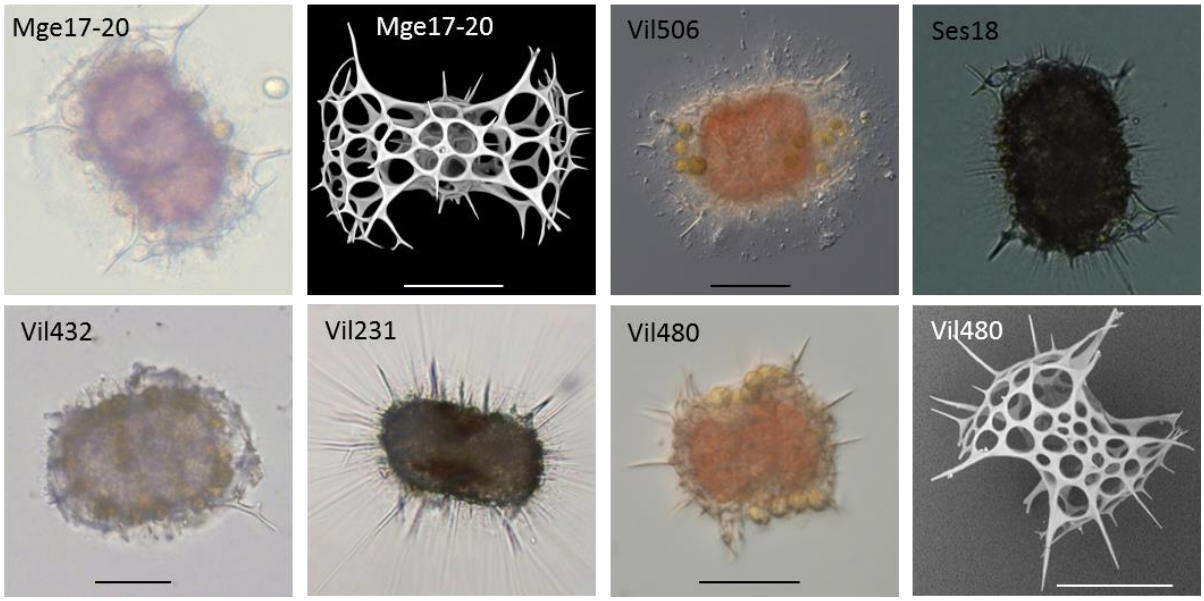
Supplementary Figure S2. Light Microscopy (LM) and Scanning Electron Microscopy (SEM, when available) images of live Spumellaria specimens used in this study for phylogenetic analysis in Fig. 1. Scale bars (when available) = 50µm. Part 2/5



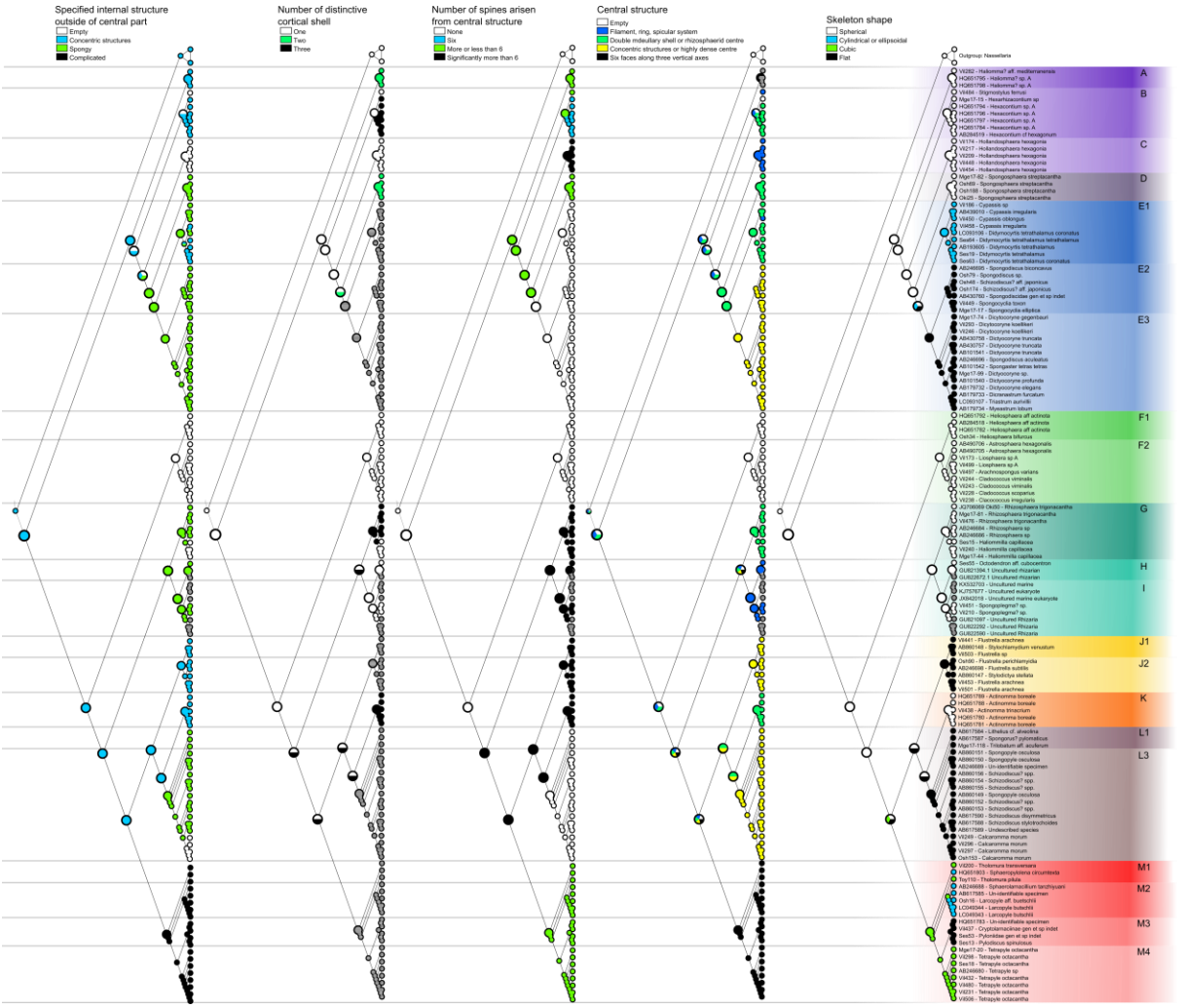
Supplementary Figure S2. Light Microscopy (LM) and Scanning Electron Microscopy (SEM, when available) images of live Spumellaria specimens used in this study for phylogenetic analysis in Fig. 1. Scale bars (when available) = 50µm. Part 3/5



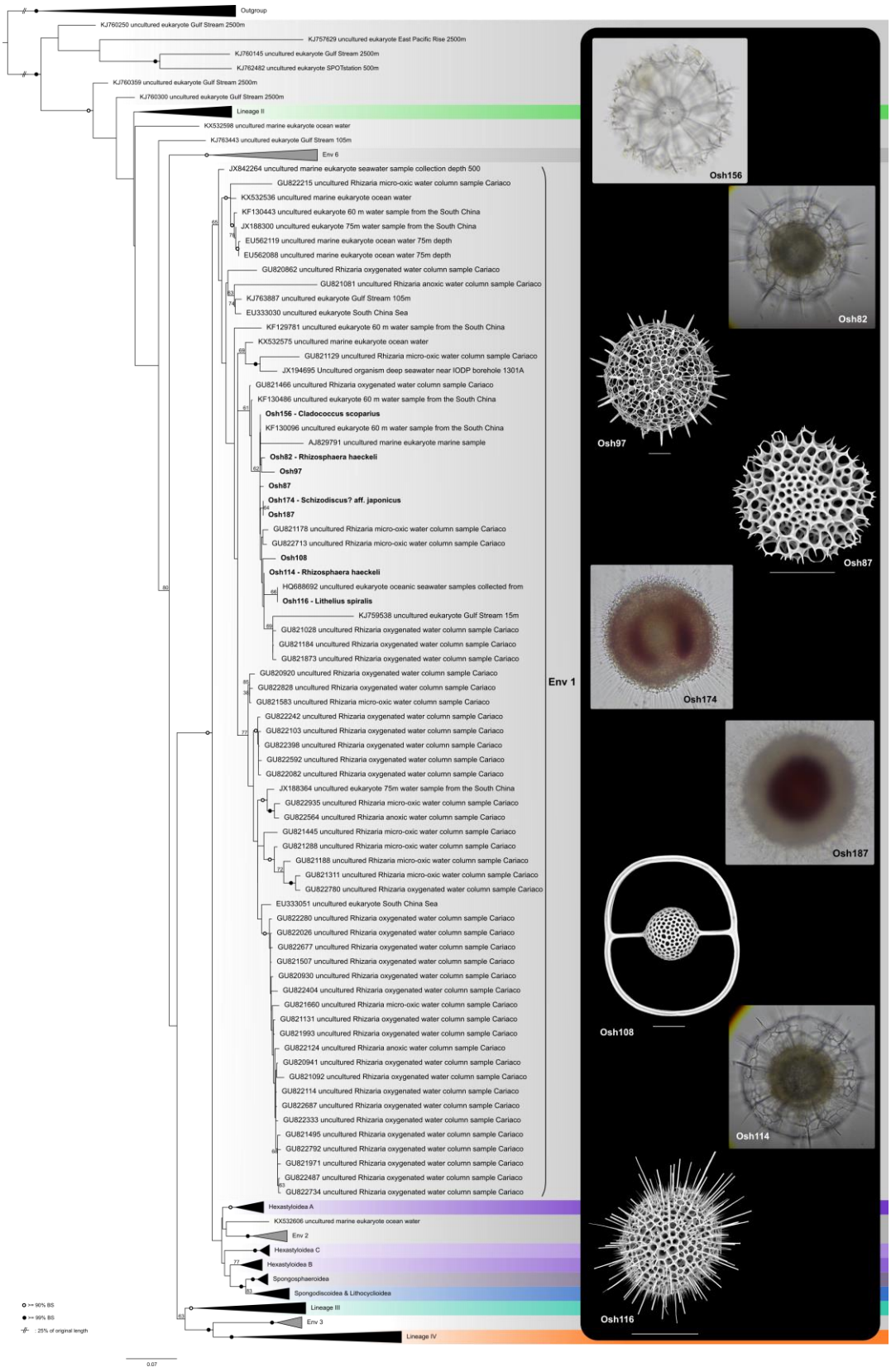
Supplementary Figure S2. Light Microscopy (LM) and Scanning Electron Microscopy (SEM, when available) images of live Spumellaria specimens used in this study for phylogenetic analysis in Fig. 1. Scale bars (when available) = 50µm. Part 4/5



Supplementary Figure S2. Light Microscopy (LM) and Scanning Electron Microscopy (SEM, when available) images of live Spumellaria specimens used in this study for phylogenetic analysis in Fig. 1. Scale bars (when available) = 50µm. Part 5/5



Supplementary Figure S3. Parsimony Ancestral State Reconstruction analysis based in the resulting phylogenetic tree for the 5 characters chosen. Relevant nodes are increased for clarity.



Supplementary Figure S4. Molecular phylogeny of probably symbiotic clade of Spumellaria, showing the different morphotypes within the same phylogenetic clade. RAxML bootstrap values (1000 replicates, BS) are shown at the nodes. Black circles indicate BS ≥ 99% and hollow circles indicates BS ≥ 90%. BS lower than 60 were removed for clarity. Branches with a double barred symbol are fourfold reduced for clarity.

Supplementary Table S1. Taxonomic authority of specimens used to obtain spumellarian phylogeny.

Lineage	Clade	Superfamily	Family	Taxon name
I	A	Hexastyloidea	Hexalonchidae	Hexacontium axotrias (Haeckel)
I	A	Hexastyloidea	Hexastylidae	Hexastylus nobilis (Cleve)
I	B	Hexastyloidea	Hexastylidae	Stigmostylus ferrusi (Hollande and Enjumet)
I	B	Hexastyloidea	Hexalonchidae	Hexarhizacontium sp.
I	B	Hexastyloidea	Hexalonchidae	Hexacontium sp. A
I	B	Hexastyloidea	Hexalonchidae	Hexacontium cf. hexagonum (Ehrenberg)
I	C	Hexastyloidea	Hexastylidae	Hollandosphaera hexagonia (Hollande and Enjumet)
I	D	Spongosphaeroidea	Spongosphaeridae	Spongosphaera streptacantha (Haeckel)
I	E1	Lithocyclioidea	Coccodiscidae	Didymocyrtis tetrathalamaus (Haeckel)
I	E1	Lithocyclioidea	Coccodiscidae	Didymocyrtis tetrathalamaus coronatus (Haeckel)
I	E1	Lithocyclioidea	Coccodiscidae	Cypassis irregularis (Nigrini)
I	E1	Lithocyclioidea	Coccodiscidae	Cypassis oblongus (Takahashi)
I	E2	Spongodiscoidea	Spongocyclidae	Schizodiscus? aff. japonicus (Matsuzaki and Suzuki)
I	E2	Spongodiscoidea	Spongocyclidae	Spongocyclia elliptlica (Haeckel)
I	E2	Spongodiscoidea	Spongocyclidae	Spongocyclia toxon (Tan and Su)
I	E3	Spongodiscoidea	Eucthitonidae	Dictyocoryne gegenbauri (Haeckel)
I	E3	Spongodiscoidea	Eucthitonidae	Dictyocoryne koellikeri (Haeckel)
I	E3	Spongodiscoidea	Eucthitonidae	Dictyocoryne elegans (Ehrenberg)
I	E3	Spongodiscoidea	Eucthitonidae	Dicranastrum furcatum (Haeckel)
I	E3	Spongodiscoidea	Eucthitonidae	Myelastrum lobum (Swanberg, Anderson and Bennett)
I	E3	Spongodiscoidea	Eucthitonidae	Triastrum aurivillii (Cleve)
I	E3	Spongodiscoidea	Eucthitonidae	Dictyocoryne truncata (Ehrenberg)
I	E3	Spongodiscoidea	Eucthitonidae	Dictyocoryne profunda Ehrenberg
I	E3	Spongodiscoidea	Eucthitonidae	Spongodiscus? biconcavus (Haeckel)
I	E3	Spongodiscoidea	Eucthitonidae	Spongodiscus aculeatus
I	E3	Spongodiscoidea	Eucthitonidae	Spongaster tetras tetras (Ehrenberg)
II	F1	Liosphaeroidea	Astrosphaeridae	Heliosphaera aff. actinosa (Haeckel)
II	F1	Liosphaeroidea	Astrosphaeridae	Heliosphaera bifurcus (Haeckel)
II	F2	Liosphaeroidea	Astrosphaeridae	Astrosphaera hexagonalis (Haeckel)
II	F2	Liosphaeroidea	Astrosphaeridae	Arachnospongus varians
II	F2	Liosphaeroidea	Astrosphaeridae	Cladococcus scoparius (Haeckel)
II	F2	Liosphaeroidea	Astrosphaeridae	Cladococcus viminalis (Haeckel)
II	F2	Liosphaeroidea	Astrosphaeridae	Cladococcus irregularis (Popofsky)
III	G	Rhizosphaeridae	Rhizosphaeridae	Haliommilla capillacea (Haeckel)
III	G	Rhizosphaeridae	Rhizosphaeridae	Rhizosphaera trigonacantha (Haeckel)
III	H	Centrocuboidae	Centrodubidae	Octodendron aff. cubocentron (Haeckel)
III	I	Centrocuboidae	Excentroconchidae	Spongoplegma? sp.
III	I	Centrocuboidae	Excentroconchidae	Spongoplegma? sp.
IV	J1, J2	Stylodictytoidea	Stylodictyidae	Flustrella arachnea (Müller)
IV	J1	Stylodictytoidea	Stylodictyidae	Perichlamydium venustum (Bailey)
IV	J2	Stylodictytoidea	Stylodictyidae	Perichlamydium perichlamydium (Stöhr)
IV	J2	Stylodictytoidea	Stylodictyidae	Flustrella subtilis (Ehrenberg)
IV	J2	Stylodictytoidea	Stylodictyidae	Stylodictya stellata (Bailey)
IV	K	Actinommoidea	Actinommidae	Actinomma boreale (Cleve)
IV	K	Actinommoidea	Actinommidae	Actinomma trinacrium (Haeckel)
IV	L1	Spongopyloidea	Lithelioidea	Trilobatum aff. acuferum
IV	L1	Spongopyloidea	Lithelioidea	Spongorus? pylomaticus
IV	L1	Spongopyloidea	Sponguroidea	New genus and new species A
IV	L2	Spongopyloidea	Spongopylidae	Spongopyle osculosa
IV	L2	Spongopyloidea	Spongopylidae	Schizodiscus? spp.
IV	L2	Spongopyloidea	Spongopylidae	Schizodiscus disymmetricus (Dogiel in Dogiel and Reshetnyak)
IV	L2	Spongopyloidea	Spongopylidae	Schizodiscus stylotrochoides (Dogiel in Dogiel and Reshetnyak)
IV	L2	Spongopyloidea	Cristallosphaeridae	Calcaromma morum (Müller)
IV	M1	Pylonioidea	Tholoniidae	Tholomura pilula (Zhang and Suzuki)
IV	M1	Pylonioidea	Tholoniidae	Tholomura transversara (Chen)
IV	M1	Pylonioidea	Pyloniidae	Sphaeropylolena circumtexta (Zhang and Suzuki)
IV	M2	Pylonioidea	Pyloniidae	Sphaerolarnacilium tanzhiyuani (Zhang and Suzuki)
IV	M2	Pylonioidea	Pyloniidae	Larcopyle aff. buetschli (Dreyer)
IV	M3	Pylonioidea	Pyloniidae	Pylodiscus spinulosus (Chen and Tan)
IV	M3	Pylonioidea	Pyloniidae	Cryptolarnaciinae gen. et sp. indet
IV	M4	Pylonioidea	Pyloniidae	Tetrapyle octacantha (Müller)

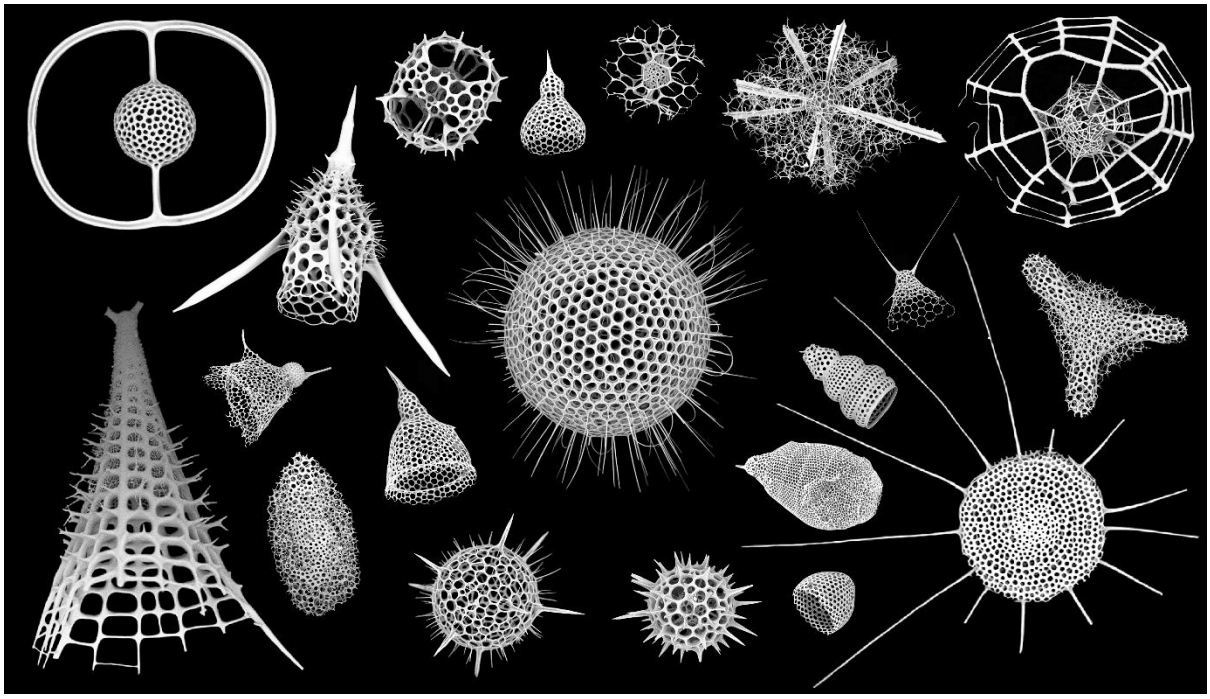
Supplementary Table S2. List of specimens used to obtain spumellarian phylogeny. Abbreviations: WMS, Westearn Mediterranean Sea; VsM, Villefranche-sur-Mer (France); Ses, Sesoko (Japan); Oki, Okinawa, Japan; EJ, East off Japan; Shi, Shimoda, Japan; Sag, Sagami Bay, Japan; Aki, Akajima Island, Japan; Sog, Sogndalsfjorden (Norway); OWC, oxygenated water column, Cariaco Basin; AWC, Anoxic water column sample, Cariaco Basin; NPO, North Pacific; EPR, East Pacific Rise; CP, Central Pacific

ID	Clade	Species	Location	Depth	Reference	Accession number 18S	Accession number 28S
HQ651795	A	*Hexastylus nobilis*	Sog		Krabberod et al. (2011)	HQ651795	
HQ651798	A	*Hexastylus nobilis*	Sog		Krabberod et al. (2011)	HQ651798	
Vil282	A	*Hexacontium sp A*	VsM		This study	in prep.	
AB284519	B	*Hexacontium cf hexagonum*	Sog	250-25	Yuasa et al. (2009)	AB284519	
HQ651784	B	*Hexacontium sp A*	Sog		Krabberod et al. (2011)	HQ651784	HQ651784
HQ651794	B	*Hexacontium sp A*	Sog		Krabberod et al. (2011)	HQ651794	
HQ651796	B	*Hexacontium sp A*	Sog		Krabberod et al. (2011)	HQ651796	
HQ651797	B	*Hexacontium sp A*	Sog		Krabberod et al. (2011)	HQ651797	
Mge17-15	B	*Hexarhizacontium sp*	WMS	500-0	This study	in prep.	
Vil484	B	*Stigmostylus ferrusi*	VsM	30	This study	in prep.	in prep.
Vil174	C	*Hollandosphaera hexagonia*	VsM	Euphotic	This study	in prep.	in prep.
Vil209	C	*Hollandosphaera hexagonia*	VsM	Euphotic	This study	in prep.	
Vil217	C	*Hollandosphaera hexagonia*	VsM	Euphotic	This study	in prep.	in prep.
Vil448	C	*Hollandosphaera hexagonia*	VsM	450-0	This study	in prep.	
Vil454	C	*Hollandosphaera hexagonia*	VsM	450-0	This study	in prep.	
Mge17-82	D	*Spongosphaera streptacantha*	WMS	500-0	This study	in prep.	in prep.
Oki25	D	*Spongosphaera streptacantha*	Oki		This study	in prep.	
Osh188	D	*Spongosphaera streptacantha*	EJ	25-0	This study	in prep.	
Osh69	D	*Spongosphaera streptacantha*	EJ		This study	in prep.	
AB193605	E1	*Didymocyrtis tetrathalamus*	Oki		Yuasa et al. (2004)	AB193605	
AB439010	E1	*Cypassis irregularis*	Sag	Surface	Kimoto et al. (2011)	AB439010	
LC093106	E1	*Didymocyrtis tetrathalamus c.*	Oki	Surface	Yuasa et al. (2016)	LC093106	
Ses19	E1	*Didymocyrtis tetrathalamus*	Ses		This study	in prep.	
Ses63	E1	*Didymocyrtis tetrathalamus c.*	Ses		This study	in prep.	
Ses64	E1	*Didymocyrtis tetrathalamus*	Ses		This study	in prep.	
Vil186	E1	*Cypassis sp*	VsM		This study	in prep.	in prep.
Vil450	E1	*Cypassis oblongus*	VsM	450-0	This study	in prep.	
Vil458	E1	*Cypassis irregularis*	VsM	450-0	This study	in prep.	
AB246695	E2	*Spongodiscus biconcavus*	Shi	Surface	Kunitomo et al. (2006)	AB246695	
AB430760	E2	*Spongocyclia toxon*	Shi		Ando et al. (2009)	AB430760	AB430760
Mge17-17	E2	*Spongocyclia elliptica*	WMS	500-0	This study	in prep.	in prep.
Osh174	E2	*Schizodiscus? aff. japonicus*	EJ	25-0	This study	in prep.	in prep.
Osh48	E2	*Schizodiscus? aff. japonicus*	EJ		This study	in prep.	
Osh79	E2	*Spongodiscus sp*	EJ	25-0	This study	in prep.	in prep.
Vil449	E2	*Spongocyclia toxon*	VsM	450-0	This study	in prep.	
AB101540	E3	*Dictyocoryne profunda*	Oki	Surface	Takahashi et al. (2004)	AB101540	
AB101541	E3	*Dictyocoryne truncata*	Oki	Surface	Takahashi et al. (2004)	AB101541	
AB101542	E3	*Spongaster tetras tetras*	Oki	Surface	Takahashi et al. (2004)	AB101542	
AB179732	E3	*Dictyocoryne elegans*	Oki	Surface	Yuasa et al. (2005)	AB179732	
AB179733	E3	*Dicra strum furcatum*	Oki	Surface	Yuasa et al. (2005)	AB179733	
AB179734	E3	*Myeastrum lobum*	Oki	Surface	Yuasa et al. (2005)	AB179734	
AB246696	E3	*Spongodiscus aculeatus*	Shi	Surface	Kunitomo et al. (2006)	AB246696	
AB430757	E3	*Dictyocoryne truncata*	Shi		Ando et al. (2009)	AB430757	AB430757
AB430758	E3	*Dictyocoryne truncata*	Shi		Ando et al. (2009)	AB430758	AB430758
LC093107	E3	*Triastrum aurivillii*	Oki	Surface	Yuasa et al. (2016)	LC093107	
Mge17-74	E3	*Dicytocoryne gegenbauri*	WMS	500-0	This study	in prep.	in prep.
Mge17-99	E3	*Dicytocoryne sp*	WMS	500-0	This study	in prep.	
Vil246	E3	*Dicytocoryne koellikeri*	VsM	Euphotic	This study	in prep.	
Vil293	E3	*Dicytocoryne koellikeri*	VsM		This study	in prep.	in prep.
AB284518	F1	*Heliosphaera aff actinota*	Oki		Yuasa et al. (2009)	AB284518	
HQ651782	F1	*Heliosphaera aff actinota*	Sog		Krabberod et al. (2011)	HQ651782	HQ651782
HQ651792	F1	*Heliosphaera aff actinota*	Sog		Krabberod et al. (2011)	HQ651792	
Osh34	F1	*Heliosphaera bifurcus*	EJ		This study	in prep.	
AB490705	F2	*Astrosphaera hexago lis*	Oki		Yuasa et al. (2009)	AB490705	
AB490706	F2	*Astrosphaera hexago lis*	Oki		Yuasa et al. (2009)	AB490706	
Vil173	F2	*Liosphaera sp A*	VsM		This study	in prep.	
Vil228	F2	*Cladococcus scoparius*	VsM		This study	in prep.	in prep.
Vil238	F2	*Clacococcus irregularis*	VsM		This study	in prep.	
Vil243	F2	*Cladococcus vimi lis*	VsM		This study	in prep.	in prep.
Vil244	F2	*Cladococcus vimi lis*	VsM		This study	in prep.	in prep.
Vil497	F2	*Arachnospongus varians*	VsM	30	This study	in prep.	in prep.
Vil499	F2	*Liosphaera sp A*	VsM	30	This study	in prep.	in prep.

AB246684	G	*Rhizosphaera sp*	Shi	Surface	Kunitomo et al. (2006)	AB246684	
AB246686	G	*Rhizosphaera sp*	Shi	Surface	Kunitomo et al. (2006)	AB246686	
JQ706069	G	*Rhizosphaera trigo cantha*	Aka		Decelle et al. Unpublished	JQ706069	
Mge17-44	G	*Haliommilla capillacea*	WMS	500-0	This study	in prep.	in prep.
Mge17-81	G	*Rhizosphaera trigo cantha*	WMS	500-0	This study	in prep.	in prep.
Ses15	G	*Haliommilla capillacea*	Ses		This study	in prep.	in prep.
Vil240	G	*Haliommilla capillacea*	VsM		This study	in prep.	in prep.
Vil476	G	*Rhizosphaera trigo cantha*	VsM	Surface	This study	in prep.	in prep.
GU821394	H	*Uncultured rhirarian clone*	OWC		Edgcomb et al. (2011)	GU821394	
GU822672	H	*Uncultured rhirarian clone*	OWC		Edgcomb et al. (2011)	GU822672	
Ses55	H	*Octodendron aff cubocentron*	Ses		This study	in prep.	in prep.
GU821097	I	*Uncultured Rhizaria*	AWC		Edgcomb et al. (2011)	GU821097	
GU822292	I	*Uncultured Rhizaria*	AWC		Edgcomb et al. (2011)	GU822292	
GU822590	I	*Uncultured Rhizaria*	AWC		Edgcomb et al. (2011)	GU822590	
JX842018	I	*Uncultured marine eukaryote*	-	500	Kim et al. (2012)	JX842018	
KJ757677	I	*Uncultured eukaryote*	EPR	2500	Lie et al. (2014)	KJ757677	
KX532703	I	*Uncultured marine eukaryote*	-		Xu et al. press	KX532703	
Vil210	I	*Spongoplegma sp*	VsM		This study	in prep.	
Vil451	I	*Spongoplegma sp*	VsM	450-0	This study	in prep.	
AB860148	J1	*Perichlamydium venustum*	CP		Ishitani et al. (2014)	AB860148	AB860148
Vil441	J1	*Flustrella arachnea*	VsM	450-0	This study	in prep.	
Vil503	J1	*Flustrella sp*	VsM	450-0	This study	in prep.	in prep.
AB246698	J2	*Flustrella subtilis*	Shi	Surface	Kunitomo et al. (2006)	AB246698	
AB860147	J2	*Stylodictya stellata*	CP		Ishitani et al. (2014)	AB860147	AB860147
Osh90	J2	*Perichlamydium perichlamydium*	EJ	250-150	This study	in prep.	in prep.
Vil453	J2	*Flustrella arachnea*	VsM	450-0	This study	in prep.	in prep.
Vil501	J2	*Flustrella arachnea*	VsM	30	This study	in prep.	in prep.
HQ651780	K	*Actinomma boreale*	Sog		Krabberod et al. (2011)	HQ651780	HQ651780
HQ651781	K	*Actinomma boreale*	Sog		Krabberod et al. (2011)	HQ651781	HQ651781
HQ651788	K	*Actinomma boreale*	Sog		Krabberod et al. (2011)	HQ651788	
HQ651789	K	*Actinomma boreale*	Sog		Krabberod et al. (2011)	HQ651789	
Vil438	K	*Actinomma tri crium*	VsM	450-0	This study	in prep.	
AB617584	L1	*Lithelius cf alveoli*	NPO	1000-750	Ishitani et al. (2012)	AB617584	
AB617587	L1	*Schizodiscus sp*	NPO	500-250	Ishitani et al. (2012)	AB617587	
Mge17-118	L1	*Trilobatum aff acuferum*	WMS	500-0	This study	in prep.	in prep.
AB246689	L2	*Un-identifiable specimen*	Shi	Surface	Kunitomo et al. (2006)	AB246689	
AB617588	L2	*Schizodiscus stylotrochoides*	NPO	500-350	Ishitani et al. (2012)	AB617588	
AB617589	L2	*Undescribed species*	NPO	150-0	Ishitani et al. (2012)	AB617589	
AB617590	L2	*Schizodiscus disymmetricus*	NPO	150-0	Ishitani et al. (2012)	AB617590	
AB860149	L2	*Spongopyle osculosa*	CP		Ishitani et al. (2014)	AB860149	AB860149
AB860150	L2	*Spongopyle osculosa*	CP		Ishitani et al. (2014)	AB860150	AB860150
AB860151	L2	*Spongopyle osculosa*	CP		Ishitani et al. (2014)	AB860151	AB860151
AB860152	L2	*Schizodiscus spp*	CP		Ishitani et al. (2014)	AB860152	AB860152
AB860153	L2	*Schizodiscus spp*	CP	180	Ishitani et al. (2014)	AB860153	AB860153
AB860154	L2	*Schizodiscus spp*	CP		Ishitani et al. (2014)	AB860154	AB860154
AB860155	L2	*Schizodiscus spp*	CP		Ishitani et al. (2014)	AB860155	AB860155
AB860156	L2	*Schizodiscus spp*	CP		Ishitani et al. (2014)	AB860156	AB860156
Osh153	L2	*Calcaromma morum*	EJ	25-0	This study	in prep.	
Vil249	L2	*Calcaromma morum*	VsM	Euphotic	This study	in prep.	in prep.
Vil296	L2	*Calcaromma morum*	VsM	Euphotic	This study	in prep.	in prep.
Vil297	L2	*Calcaromma morum*	VsM	Euphotic	This study	in prep.	in prep.
HQ651803	M1	*Sphaeropylole circumtexta*	Sog		Krabberod et al. (2011)	HQ651803	
Toy110	M1	*Tholomura pilula*	Oki		This study	in prep.	
Vil200	M1	*Tholomura transversara*	VsM		This study	in prep.	in prep.
AB246680	M2	*Tetrapyle sp*	Shi	Surface	Kunitomo et al. (2006)	AB246680	
AB246688	M2	*Sphaerolar cillium tanzhiyuani*	Shi	Surface	Kunitomo et al. (2006)	AB246688	
AB617585	M2	*Un-identifiable specimen*	NPO	150-0	Ishitani et al. (2012)	AB617585	
LC049343	M2	*Larcopyle butschlii*	Japan sea		Ishitani et al. (2015)	LC049343	
LC049344	M2	*Larcopyle butschlii*	Japan sea		Ishitani et al. (2015)	LC049344	
Mge17-20	M2	*Tetrapyle octacantha*	WMS	500-0	This study	in prep.	
Osh16	M2	*Larcopyle aff buetschlii*	EJ		This study	in prep.	
Ses18	M2	*Tetrapyle octacantha*	Ses		This study	in prep.	in prep.
Vil231	M2	*Tetrapyle octacantha*	VsM	Euphotic	This study	in prep.	
Vil298	M2	*Tetrapyle octacantha*	VsM		This study	in prep.	in prep.
Vil432	M2	*Tetrapyle octacantha*	VsM	450-0	This study	in prep.	in prep.
Vil480	M2	*Tetrapyle octacantha*	VsM	Surface	This study	in prep.	in prep.
Vil506	M2	*Tetrapyle octacantha*	VsM	450-0	This study	in prep.	in prep.
HQ651783	M3	*Un-identifiable specimen*	Sog		Krabberod et al. (2011)	HQ651783	HQ651783
Ses13	M3	*Pylodiscus spinulosus*	Ses		This study	in prep.	in prep.
Ses53	M3	*Pyloniidae gen et sp indet*	Ses		This study	in prep.	in prep.
Vil437	M3	*Cryptolar cii e gen et sp indet*	VsM	450-0	This study	in prep.	

Chapter 2

Radiolaria classification and evolution, an integrative approach



Context of the work

The second chapter presents a revised classification and evolutionary history of Radiolaria. Here I integrated previous morpho-molecular classifications performed on Acantharia (Decelle et al., 2012b) and Collodaria (Biard et al., 2015) and those performed in the first chapter of this thesis. I achieved the most extensive rDNA phylogenetic framework for Radiolaria available so far. Then I compared my results with recent advances in multi-gene and multi-protein phylogenies in order to propose a comprehensive integrative classification for Radiolaria. Finally, I calibrated the molecular clock for rDNA genes of Radiolaria leading to a better understanding of their early evolutionary history within the Retaria and Rhizaria context.

I presented my previous phylogenetic work (from Chapter 1) at the 15th InterRad congress held in Niigata (Japan; see Oral and poster presentations in annexes, page 194), where I had the opportunity to meet with Yasuhide Nakamura. Together we started a collaboration on the phylogenetic relationships of the controversial order Entactinaria (see collaborations 3 in annexes, page 199).

While conducting my general phylogenetic analyses on Radiolaria I realised that every symbiotic clade were holding distal positions and/or had long branches. To better understand the origin of the symbiotic relationships, I supervised Elsa Gadoin during her Master Thesis on the exploration of the molecular diversity of symbiotic microalgae in Nassellaria and Spumellaria (see collaborations 4 in annexes, page 200).

Radiolaria classification and evolution, an integrative approach

Miguel M. Sandin[1], Luis O'Dogherty[2], Spella Gorican[3], Noritoshi Suzuki[4], Fabrice Not[1]

1-Sorbonne Université, CNRS -UMR7144- Station Biologique de Roscoff, 29680 Roscoff, France.
2-Universidad de Cadiz, Facultad de Ciencias del Mar, E-11510, Puerto Real, Cadiz, Spain.
3-Paleontoloski Institut Ivana Rakovca ZRC SAZU, Novi trg 2, SL-1000 Ljubljana, Slovenia.
4-Department of Earth Science, Graduate School of Science, Tohoku University, Sendai 980–8578, Japan.

In preparation

Introduction

Radiolaria are ameboid planktonic protists, ubiquitous and abundant in the world ocean. Along with Foraminifera they constitute the Retaria supergroup, a main branch in the eukaryotic tree of life within the Rhizaria lineage (SAR; Adl et al., 2019). Radiolaria is divided in five groups, Acantharia with a strontium sulphate skeleton on one hand and those with a siliceous skeleton Taxopodida and the Polycystines Collodaria, Nassellaria and Spumellaria on the other hand (Suzuki and Not, 2015). The robust skeleton of Polycystines preserves very well in sediments, showing a continuous fossil record back to the early Cambrian (Suzuki and Oba, 2015; Aitchison et al., 2017). This fossil record represents a valuable tool for paleoenvironmental reconstruction studies (e.g., Abelmann and Nimmergut, 2005), and has been extensively studied by micro-paleontologists, establishing radiolarian taxonomy and evolutionary history from a morphology-based approach (Petrushevskaya, 1971a; De Wever et al., 2001; Afanasieva et al., 2005). Radiolaria ecology and bio-geography have been assessed from sediments and planktonic sediment traps essentially (Boltovskoy et al., 2010; Boltovskoy and Correa, 2016a; Boltovskoy, 2017). Although, an important part of radiolarian diversity does not preserve in fossils and sediments, such as Acantharia and Taxopodida due to the fast dissolution of their skeleton upon cell death.

Molecular based diversity surveys performed at a global scale have highlighted the significant contribution of Radiolaria to the protistan planktonic community, demonstrating the complementarity of DNA to that of morphology-based studies providing access to the unexplored morphological diversity. These metabarcoding surveys have shown Acantharia and Collodaria to be very abundant and diverse in the sunlit ocean (de Vargas et al., 2015), spumellarian greatest diversity and abundances being towards deep environments (Pernice et al., 2016). These distribution patterns may be influenced by the fact that some radiolarian groups harbor photosynthetic algae as symbionts. To our knowledge, Collodaria was never reported without symbionts, suggesting a strong dependency to the dinoflagellate *Brandtodinium nutricula* (Probert et al., 2014). Although this dinoflagellate was also reported in other Polycystines and even other planktonic hosts such as the jellyfish *Velella velella* (Gast

and Caron, 1996). The symbiotic association of Acantharia with the haptophyte *Phaeocystis*, occurs in the great majority for clades E and F being less flexible and with more intimate interactions such as algal symbiont plastids remodeling (Decelle et al., 2012a, 2015, 2019). On the contrary, Spumellaria can harbour a wide variety of photosynthetic symbionts (Zhang et al., 2018) from the dinoflagellates *Brandtodinium nutricula* and *Gymnoxanthella radiolariae* (Probert et al., 2014; Yuasa et al., 2016) to cyanobacteria, prasinophytes, and haptophytes (Gast and Caron, 2001; Yuasa et al., 2012, 2019)

Molecular approaches on radiolarian single cells combined with morphological identification have recently contributed to explore co-evolutionary patterns between Acantharia and their symbionts (Decelle et al., 2012a, Decelle et al., 2012b) but also allowed an accurate description of collodarian molecular biogeography and diversity at a fine taxonomic level (Biard et al., 2017). Recent studies have used the extensive knowledge in the fossil record of polycystines to time-calibrated morpho-molecular classifications in Nassellaria (Sandin et al., 2019, Chapter 1.1) and Spumellaria (Sandin et al. *in prep.*, Chapter 1.2), leading to the reinterpretation of ancient fossil groups. Further analyses have also demonstrated that previously considered distinct orders are actually scattered among Rhizaria, such is the case of the polyphyletic group Entactinaria (Sandin et al. *in prep*, Chapter 1.2; Nakamura et al. *in prep*.).

A number of studies have used molecular approaches to unveil the relationships of Radiolaria at high taxonomic level (Ishitani et al., 2011; Krabberød et al., 2017; Cavalier-Smith et al., 2018), but little attention has been payed within the different orders of Radiolaria. The fast-evolving nature of the rDNA of this group challenges phylogenetic reconstructions giving contrasting interpretations (Yuasa et al., 2005; Kunitomo et al., 2006; Krabberød et al., 2011; Yuasa and Takahashi, 2016). For instance, Nassellaria and Collodaria have always shown to be a monophyletic group and, along with Foraminifera, they have the longest branches in phylogenetic inferences producing unresolved relationships among the other radiolarian groups (Kunitomo et al., 2006; Krabberød et al., 2011). Also, Taxopodida, with only one species described, have been related to either Acantharia (Ishitani et al., 2011), Spumellaria (Kunitomo et al., 2006; Yuasa and Takahashi, 2016), Nassellaria (Yuasa et al., 2005) or basal to all radiolarians (Krabberød et al., 2017). Despite all these different topologies, Spumellaria and Acantharia have always shown to be phylogenetically distant, Nassellaria and Collodaria to be highly related and Taxopodida to hold always basal positions regardless its phylogenetic relationships.

Integrating previous phylogenetic morpho-molecular studies (Decelle et al., 2012b; Biard et al., 2015; Sandin et al., 2019, Chapter 1.1; Sandin et al. *in prep.* Chapter 1.2), here we present a broad radiolarian rDNA gene (18S and 28S partial rDNA) phylogeny, amended with the extant knowledge about their classification. The extensive fossil record of Radiolaria allows to calibrate in time their phylogeny and reconstruct their evolutionary history contextualized with biotic and abiotic drivers. Altogether, it brings an up-to-date comprehensive framework for the study of the molecular diversity and evolution of Radiolaria, improving the understanding of Rhizaria early diversification.

Material and Methods

Sampling, single cell isolation, morphological identification and DNA extraction, amplification and sequencing

In total 2 Nassellaria specimens were collected during MOOSE-cruise 2018 (western Mediterranean Sea) in order to complete missing morpho-genera (Fig S1). Protocols regarding single cell isolation, morphological identification and DNA extraction, amplification and sequencing were followed according to Chapter 1.1 (Sandin et al., 2019, Chapter 1.1).

Phylogenetic analysis

Taxa was selected in order to cover the main genetic diversity described by Decelle et al. (2012b) for Acantharia, Biard et al. (2015) for Collodaria, Sandin et al. (2019, Chapter 1.1) for Nassellaria, Nakamura et al. (2015) for Phaeodaria and Sandin et al. (*in prep.,* Chapter 1.2) for Spumellaria. Regarding environmental genetic diversity affiliated to Radiolaria, the PR2 database (Guillou et al., 2013; https://github.com/pr2database/pr2database) was examined in order to extract representative clades of main environmental groups according to diverse environmental studies (López-García et al., 2002; Not et al., 2007; Pernice et al., 2013; Wu et al., 2014).

Two different datasets for the 18S rDNA and the partial 28S rDNA gene (D1+D2 regions) were aligned using MAFFT v7.395 (Katoh and Standley, 2013) with a L-INS-i algorithm ('--localpair') and 1000 iterative refinement cycles for high accuracy. Due to the fast-evolving nature of these taxa (especially Nassellaria and Collodaria lineage), an exhaustive manual checking was carried out in SeaView version 4.6.1 (Gouy et al., 2010) in order to decrease gaps and correct similar but unmatching regions. Final alignments were trimmed automatically using trimal (Capella-Gutiérrez et al., 2009) with a 30% gap threshold.

As previously seen in Radiolaria the concatenation of 18S rDNA gene and 28S rDNA gene improves phylogenetic resolution (Decelle et al., 2012; Biard et al., 2015; Sandin et al., 2019, Chapter 1.1; Sandin et al. *in prep.*, Chapter 1.2). The final data set of the concatenation of the two genes contains 217 taxa and 2530 positions (211 taxa and 1794 positions regarding the 18S rDNA gene and 98 taxa and 736 positions the partial 28S rDNA gene; D1+D2 regions). Foraminifera is considered to be the closest relative of Radiolaria (Cavalier-Smith et al., 2018) yet its fast evolving rDNA may produce a long branch attraction artefact interfering in the topology (Krabberød et al., 2011). Therefore, 16 sequences of Phaeodaria were chosen as the closest undoubtable relative to form the outgroup. In order to strengthen the output, three different evolutionary methods were implemented in a maximum likelihood approach, a PhyML GTR+G+I, a RAxML GTR+G and a RAxML GTR+G+CAT with 10 000 bootstraps. In parallel a Bayesian analysis was performed using BEAST version 1.10.1 (Drummond et al., 2012) with a GTR+G+I model over 100 million generations sampled every 1000 states. Final trees were visualized and edited with FigTree version 1.4.3 (Rambaut 2016).

Resulting phylogenetic analysis were used as a reference to update Radiolaria sequences (precisely those of Acantharia, Nassellaria and Spumellaria) in PR2 database following the curation pipeline stated in http://eukref.org/curation-pipeline-overview/.

Molecular clock

Molecular clock estimates were performed according to Sandin et al. (2019), Chapter 1.1. In total 17 nodes were chosen to carry out the calibration, explained below from the oldest to the newest calibration age and given the name of the node for the taxa they cover:

- Spumellaria: Both the fossil record (De Wever et al., 2001; Pouille et al., 2011; Aitchison, 2017) and previous molecular clock analysis (Chapter 1.2) agree that the first spherical polycystines appeared in the early Cambrian. Therefore the Spumellaria clade was calibrated with a normal distribution (N) of average 515 Ma and a standard deviation of 40 Ma to allow some uncertainty in the estimation due to the broad context of the analysis: N(515, 40)
- Nassellaria N(415, 30): The first radiolarian fossils attributed to Nassellaria appear in the Upper Devonian (ca. 372.2 Ma; Cheng, 1986), although molecular clock analysis estimate their diversification earlier at ~423 Ma (Sandin et al., 2019, Chapter 1.1). Therefore, this node was calibrated with a normal distribution of average 415 Ma and a standard deviation of 30 Ma.
- Acantharia N(355, 100): Previous molecular clock analysis of Acantharia has settled its first diversification at 355 Ma (Decelle et al., 2012a). Although due to the absence of a direct observation in the fossil record of Acantharia this estimation was calibrated through closely related groups (Spumellaria and Nassellaria). Therefore, a big standard deviation (100 Ma) was chosen to allow uncertainty in the estimation.
- Acropyramioidea N(245, 10): The first appearance of this superfamily happen in the Early Anisian (ca. 241.5-246.8 Ma) with the genus *Celluronta* (Sugiyama, 1997; O'Dogherty et al., 2009b).

The rest of the nodes were taken according to previous calibrations of the molecular clock: Seven nodes within Nassellaria (Eucyrtidioidea N(246, 10), Lophophaenidae N(191, 10), Artostrobioidea N(182, 10), Acanthodesmoidea N(66, 10), Bekomidae N(62, 10), Pterocorythoidea N(57, 10) and Carpocaniidae N(38, 10)) included in Sandin et al. (2019), Chapter 1.1, and six within Spumellaria (Hexastyloidea N(242, 10), Liosphaeroidea N(233, 20), Actinommoidea N(170, 20), Rhizosphaeroidea N(148, 10), Pylonioidea N(97, 10) and Coccodiscoidea N(45, 10)) included in Sandin et al. *in prep.,* Chapter 1.2, In order to restrict the possible solutions during the Bayesian inference, both the Phaeodaria clade (considered the outgroup) and the Radiolaria clade (considered the ingroup) were forced to be monophyletic. Also were forced to be monophyletic nodes appearing in different positions along the different phylogenetic analysis (such as Acantharian clades B and C, closely related to clades D, E, F and the environmental clade Acan4; Decelle et al., 2012b), clades appearing in different position as originally described (Carpocaniidae appears closely related to Artostrobioidea; Sandin et al., 2019, Chapter 1.1) or with a low bootstrap support but confirmed to be a clade in other phylogenetic studies (such es the Polycystines group; Cavalier-Smith et al., 2018).

Results

Phylogenetic analysis

Our final molecular phylogeny is composed of 217 distinct radiolarian specimens, covering 211 sequences for the 18S rDNA gene and 98 sequences of the partial (D1 & D2 regions) 28S rDNA gene. From the 217 sequences, 135 correspond to morphologically described specimens and 82 correspond to environmental sequences with no morphological data associated (Supplementary Material Table S1). The final alignment matrix has a 13.84% of invariant sites and 2090 distinct alignment patterns. Based on phylogenetic clustering support, phylogenetic distance and monophyletic delimitation, sequences included in our study cluster in 7 main groups (Fig. 1): Acantharia, Nassellaria, Rad-A, Rad-B, Rad-C, Rad-X and Spumellaria. Every main group were highly supported (>93 BS and > .99 PP) and consistent along the different phylogenetic analysis: PhyML GTR+G+I, RAxML GTR+G, RAxML GTR+CAT, Bayesian GTR+G+I. On the contrary, relationships between these groups were inconsistent among the different phylogenetic analysis, especially the position Nassellaria and Rad-A. Final phylogeny (Fig. 1) is represented by RAxML GTR+G analysis due to the preferred topology and relationships between main groups.

Composed of 4 environmental sequences, Rad-X clade is highly supported (BS=100 and PP=1) as a sister clade to Acantharia in all different phylogenetic analysis, despite the clear phylogenetic distance between these groups. Also, in every phylogenetic analysis, Acantharia and Rad-X are regarded as a sister clade to all the other groups. Although Nassellaria, Rad-A, Rad-B, Rad-C and Spumellaria are grouped together but weakly supported (BS=50 in PhyML, 53 in RAxML GTR+G, 69 in RAxML GTR+CAT and PP= 0.54 in Bayesian GTR+G+I). These differences are mainly due to the variant position of Nassellaria and Rad-A. In RAxML GTR+CAT and Bayesian GTR+G+I, Nassellaria appears related to Spumellaria, Rad-A, Rad-B and Rad-C. These four groups cluster together with moderate to high support (BS= 63 & PP=0.93), being Rad-A, Rad-B and Rad-C (weakly supported as a group, BS=34 & PP=0.57) basal to Spumellaria. In RAxML GTR+G, Rad-A, Rad-B and Rad-C forms a weakly supported clade (BS=31) basal to the Spumellaria and Nassellaria, also weakly clustered together (BS=45). And in PhyML GTR+G+I, Nassellaria appears related to two groups clustered together (BS=88), Rad-A and Spumellaria moderately supported as a clade (BS=64) and Rad-B and Rad-C strongly supported together (BS=97). Despite these differences, Rad-C and Rad-B were always clustered together (BS>87 and PP=1).

Acantharia is composed of a total of 11 clades based on phylogenetic support and consistency among replicates, of which 7 are morphologically described (Clades A, B1, B2, C, D, E and F) and 4 correspond to environmental sequences (clades Acan1, Acan2, Acan3 and Acan4). Clade A along with Acan1 form a highly supported clade basal to the rest of the clades. Clades Acan2 and Acan3 are also holding a basal position with a shorter phylogenetic distance to the common node of all Acantharia. Clades B and C appear weakly supported as sister lineage to clades D, E, F and Acan4. Within the former of these groups, Clade B appears scattered in two clades (B1 and B2) of which B1 is weakly (BS<50 & PP=.55) related to Clade C holding a long branch relative to Clade B2 and C. And clades D, E, F and Acan4 are clustered together with high bootstrap (>95) and posterior probabilities (1) together. From these, Clade D branches first, and clades E and F are sister clades.

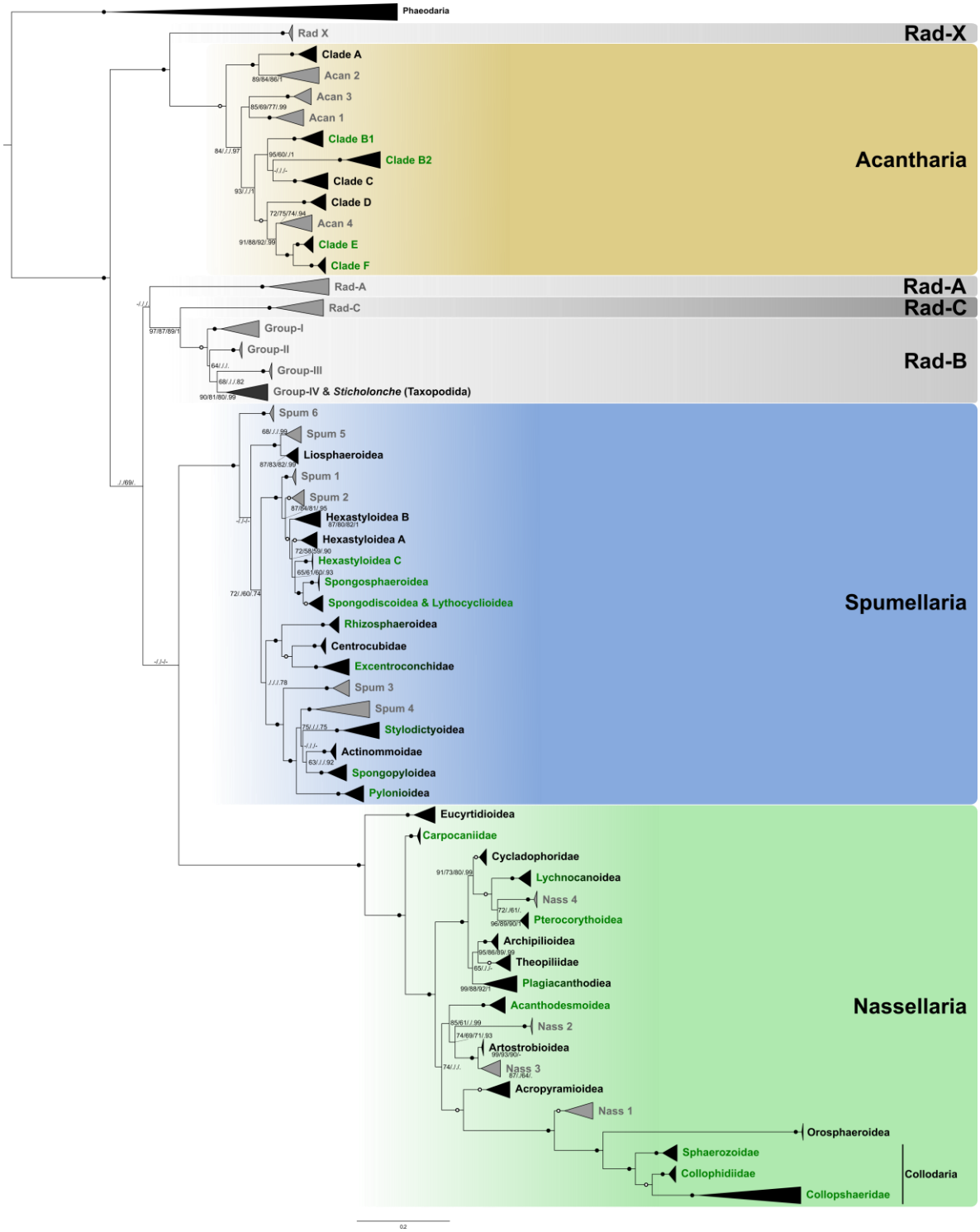


Figure 1. Molecular phylogeny of Radiolaria inferred from the concatenated complete 18S and partial 28S (D1-D2 regions) rDNA genes (217 taxa and 2533 aligned positions). The tree represents the best scoring tree obtained by RAxML using the GTR + GAMMA model of sequence evolution. Values at nodes represents PhyML (GTR+G+I), RAxML (GTR+GAMMA), RAxML (GTR+CAT) bootstrap values (10 000 replicates) and posterior probabilities obtained by Bayesian inference (GTR+G+I). Black circles indicate BS ≥ 99% and PP ≥ 0.99. Hollow circles indicate BS ≥ 90% and PP ≥ 0.90. Values below 60 BS or 0.6 PP are represented as a dot (“.”). Not existing node in a replicate phylogenetic analysis is represented as “-“. Names in black represent morphologically described clades, in grey environmental clades, in green symbiotic clades and black degraded to green indicates presence of symbiotic subclades. Sixteen phaeodarian sequences were assembled as outgroup.

Each of Rad-A, Rad-B and Rad-C are highly supported clades (BS>95 & PP=1) composed essentially of environmental sequences coming from different studies and different environments (Supplementary Material Table S1). Both Rad-A and Rad-C are represented by only one clade each. Rad-B is the most genetically diverse with up to 4 different subclades including the only described species among these environmental clades, *Sticholonche zanclea* (Taxopodida). This is represented by 2 different sequences clustering at basal position within group IV (Supplementary Material Fig. S1). One of these sequences (HQ651784) covers the only representation for the 28S rDNA gene within the environmental clades. In general, both Rad-A and the group Rad-C and Rad-B, always hold basal positions regarding the root of the tree.

Spumellaria is constituted by 19 groups of which 13 are morphologically described and 6 are environmental clades. Liosphaeroidea and the environmental clades Spum5 and Spum6 hold the most basal position regarding the other Spumellaria clades. In PhyML and bayesian analysis they cluster together (BS=54 & PP=0.94). Yet, in RAxML (both gamma and CAT) Spum6 holds the most basal position regarding all Spumellaria. Although, in all different phylogenetic analysis, Spum5 and Liosphaeroidea constitute a highly supported clade (BS=100 & PP=.99) with relatively short phylogenetic distance between them. The environmental clades Spum1 and Spum2 are clustered in the basal position with the clades Hexastyloidea B, A and C, Spongodiscoidea and Lythocyclioidea and Spongosphaeroidea (from the more basal to the more distal) with high bootstrap (BS=100) and posterior probabilities (1). Within this group all clades are clearly supported (BS>80 & PP>.96) despite the short phylogenetic distance between them. Rhizosphaeroidea, Centrocubidae and Excentroconchidae cluster together with high support (BS>90 & PP=1) and a great phylogenetic distance constituting the sister clade of the environmental clades Spum3, Spum4, Actinommoidea, Spongopyloidea, Stylodictyoidea and Pylonioidea (from the more basal to the more distal), also clustering together with high support (BS=100 & PP=1). Despite Spum4, Stylodictyoidea, Actinommoidea and Spongopyloidea clustered together in all phylogenetic analysis, their relationships are low supported (BS<67 & PP<.92).

The last of the groups belongs to Nassellaria, and they are monophyletic in all phylogenetic analyses, with a large phylogenetic distance and high support (BS=100 & PP=1). There are up to 19 different clades, of which 15 are morphologically described and 4 environmental clades. From the 15 morphologically described clades, 3 correspond to Collodaria (Sphaerozoidae, Collophidiidae and Collosphaeridae) and 1 to Orosphaeridae. Eucyrtidioidea holds the most basal position, and Carpocaniidae the second most basal position, both with high support (BS>99 & PP=1) regarding the rest of Nassellaria clades. Archipilioidea, Theopiliidae and Plagiacanthoidea forms a highly supported clade BS>99 & PP=1) with Cycladophoridae, Lychnocanoidea, the environmental clade Nass4 and Pterocorythoidea. The rest of the clades cluster together with relatively low support (BS=74/55/55 & PP=.60). Of them, Acanthodesmoidea with moderate support (BS=85/61/59 & PP=.99) constitutes the sister clade of the environmental clade Nass2 and the highly supported clade (BS=100 & PP=1) constituted by Artostrobioidea and the environmental clade Nass3. Acropyramioidea is sister to the environmental clade Nass1, Orosphaeridae and the

three families of Collodaria. Within this last group the phylogenetic distance between clades is relatively big and the support is high at all nodes (BS<93 & PP=.97).

Molecular clock

The molecular clock dated the root of the tree with a median value of 1035 Ma (95% Highest Posterior Density -HPD-: between 1317 and 793 Ma) (Fig. 2). From here on, all dates are expressed as median values followed by the 95% HPD interval. The Radiolaria diversify short afterwards at 941 (1156-750) Ma. From this point, until the following ~250 million years, there is only one diversification event happening at 804 (975-646) Ma that corresponds to the diversification within silicate radiolarians, where Rad (A, B and C) diverge from the Polycystines. The first significant diversification event happened at 692 (897-500) Ma, 685 (879-510) Ma and 677 (828-558) Ma, when Rad-A branch apart from Rad-B and C, Acantharia diversifies from Rad-X, and Nassellaria and Spumellaria separate, respectively. Again at 519 (584-453) Ma, 507 (618-406) Ma and 504 (699-340) Ma there is a second rapid diversification, when Spumellaria and Acantharia diversify for the first time and Rad-C diverge from Rad-B, respectively. Nassellaria diversifies soon afterwards at 466 (514-418) Ma. On the contrary, environmental clades start the speciation later at 328 (448-221) Ma for Rad-B, 288 (520-120) Ma for Rad-A, 178 (297-85) Ma for Rad-C and 36 (83-10) Ma respectively for Rad-X. The first living representatives appear at 243 Ma with the nassellarian groups Eucyrtidioidea (263-223) and Acropyramioidea (263-224) followed by the spumellarian polyphyletic group Hexastyloidea at 240 (259-221) Ma. Other relevant nodes are the monophyletic group of symbiotic Spumellaria (Hexastyloidea-C, Spongosphaeroidea, Spongodiscoidea and Lythocyclioidea) diversifying at 187 (228-140) Ma, Collodaria diversifying at 186 (244-132) Ma, *Sticholonche* (emended in group-IV within Rad-B), diversifying at 184 (277-101) Ma and the monophyletic group of symbiotic Acantharia (clades E and F) diversifying at 159 (233-90) Ma.

Discussion

About the environmental diversity

Despite the increased effort in recent years unveiling radiolarian morpho-molecular reference sequences and the great amount of described families covered (Decelle et al., 2012b; Biard et al., 2015; Sandin et al., 2019, Chapter 1.1; Sandin et al. *n prep.*, Chapter 1.2), up to 20 environmental clades (out of 56) remain still without morphological representatives. Some of them have been hypothesized to be regarded to described families not included in the phylogenetic analysis. Especially those holding distal positions in the tree, such as Spum4 or Acan1, mainly due to a low sampling effort towards deep or extreme environments (Decelle et al., 2012b; Sandin et al. *in prep*, Chapter 1.2). Although, it is important to stress the relatively high coverage of morphologically described groups. For example, in Acantharia there are represented the four orders described, covering 14 out of ~18 families described or for Nassellaria there are representatives of the seven superfamilies, covering 16 out of ~25 families described based on morphology. This leaves few candidates to speculate possible morphologies for any of the many environmental clades.

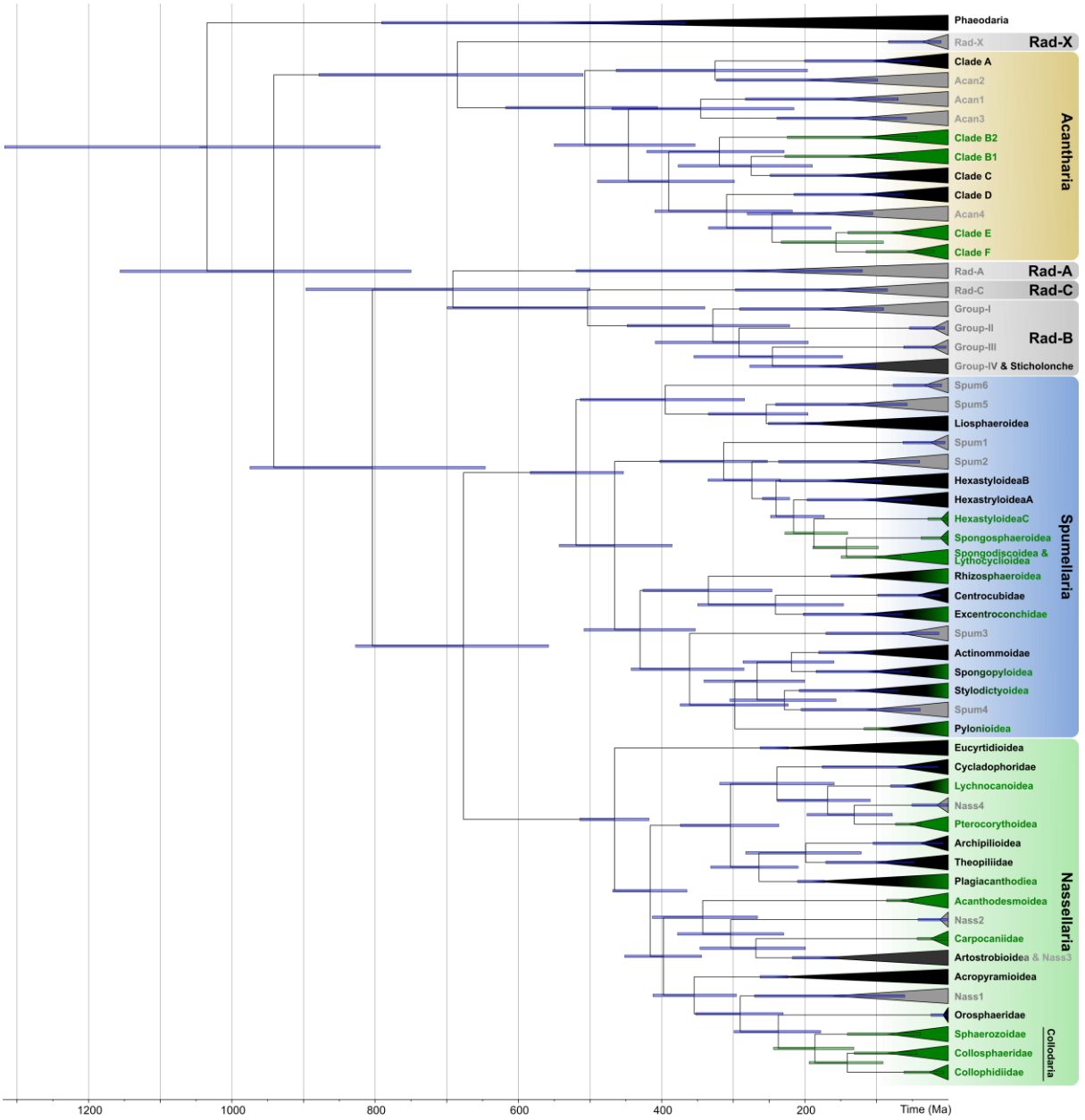


Figure 2. Time-calibrated tree (Molecular clock) of Radiolaria, based on alignment matrix used for phylogenetic analyses. Node divergences were estimated with a Bayesian relaxed clock model and the GTR + γ + I evolutionary model, implemented in the software package BEAST. See material and methods for more explanations on model selection. Blue bars indicate the 95% highest posterior density (HPD) intervals of the posterior probability distribution of node ages. Black names represent morphologically described clades, grey environmental clades, green symbiotic clades and black degraded to green indicates presence of symbiotic subclades.

A considerable part of the environmental diversity holds basal positions, such as Rad-A, Rad-B, Rad-C and Rad-X. Most of the sequences affiliated within and among these clades come from diverse environments sequenced in different studies: as the Cariaco Basin (Edgcomb et al., 2011; Orsi et al., 2011), the East Pacific Rise and the North Atlantic (Lie et al., 2014) or the South China Sea (Wu et al., 2014) from surface down to 2500m deep. The wide geographical distribution along with the few morphological candidates hinders the possibility of characterizing the great environmental diversity. Although, recent molecular environmental

surveys (Giner et al., 2019) have shown different water masses preferences among these environmental groups; being Rad-A more abundant at a deep chlorophyll maximum depth, Rad-C in mesopelagic and Rad-B increasing towards the bathypelagic and becoming very abundant and diversified (Pernice et al., 2016; see Chapter 3.2). These results highlight the need of both a higher sampling effort towards deep environments and a more detailed molecular environmental survey in order to explore specific patterns in their distribution and diversity.

The basal position of *Sticholonche* within its clade (Group-IV in Rad-B; Supplementary Material Figure S1) and the considerable amount of sequences related to it suggest that *Sticholonche* may not be the only representative of its clade (group-IV) or may belong to a species complex, as previously proposed by Nikolaev et al. (2004). Also, in Sandin et al. *in prep.*, Chapter 1.2, it has been hypothesized that some of these basal groups probably are represented by skeleton-less, or heliozoan-like, organism living in tight biotic interactions with other radiolarians (precisely Spum1 living within other spumellarians). Therefore, it is likely that part of the molecular diversity related to Rad-B may be related to naked amoeboid shapes or has a similar morphology to Taxopodida. Similar naked species have been hypothesis for basal groups in Foraminifera, related to Monothalamids (Pawlowski et al., 2011, 2013). Indeed, some amoeboid Heliozoans have been proposed as members of Radiolaria based on ultrastructural features of the axopodial complex, such as Gymnosphaerida (Yabuki et al., 2012). For these reasons, in order to unveil the great and basal environmental diversity, it may be worth exploring different morphologies rather than looking for the classical star-radiolarian shape or target remote environments, with a preference towards the deep ocean.

A revised morpho-molecular classification of Radiolaria

As seen in previous classifications of Retaria, the fast-evolving nature of the rDNA of Foraminifera challenges a consistent molecular classification because of long branch attraction artefact (Moreira et al., 2007). This effect placed Foraminifera as a sister clade of the Polycystines (Irwin et al., 2019) or to Acantharia (Sierra et al., 2013) in some analysis, challenging the monophyly of Radiolaria. The long branch of Nassellaria (+Collodaria) produces a similar result, that is reflected in the different position of this group in different rDNA phylogenetic studies (Ishitani et al., 2011; Krabberød et al., 2011; Yuasa and Takahashi, 2016, this study). Recently, and in agreement with our results (RAxML GTR+G), multigene and multiprotein phylogenies have confirmed the monophyly of the polycystines group (including Spumellaria and Collodaria) with Acantharia as a sister clade (Krabberød et al., 2017; Cavalier-Smith et al., 2018). Although, further analysis are required to help clarify long branches of multigene and multiprotein phylogenies in both Radiolaria and Foraminifera, since contaminations from symbiotic entities may be frequent in genomes and transcriptomes extraction altering relationships (Burki et al., 2010). In addition, in multi-genes phylogenies, Radiolaria specimens are often represented by fewer proportion of genes (Burki et al., 2013), and some taxa such as Taxopodida and the rest of environmental clades are neglected in most of the multigene and multiprotein phylogenies.

Despite several incongruencies in molecular classifications of Radiolaria, Acantharea and Spumellaria always form two separate clades, Collodaria appears nested within Nassellaria and Taxopodida holding basal positions. Therefore, here we understand the Radiolaria with three different classes, the Acantharia, with strontium sulphate skeleton, and the Taxopodida (=Sticholonchia) and Polycystines, with opaline silica skeleton, as established in Adl et al. (2019) and in agreement with Yuasa and Takahashi (2016). This separation of the Radiolaria in three different lineages contrasts with recent studies (Cavalier-Smith et al., 2018) regarding their taxonomic level, where *Sticholonche* (Taxopodida, Rad-B) holds an infraphylum position along with Radiozoa (=Radiolaria). Then Polycystina, is divided in two orders and not three: Nassellaria (including the three families of Collodaria) and Spumellaria. Following this hypothesis, it is possible to identify a complexification of the skeleton for both the Acantharia (as already reported by Decelle et al., 2012b) and the silicate radiolarians: Taxopodida and Polycystines. The most basal position in the phylogenetic tree correspond to Taxopodida, emended in Rad-B, lacking any symmetry with silicate spicules coming out of the cell. In contrast the Polycystines, holding a more distal position, gained a complex and robust skeleton. Similar patterns seem to be shared in Foraminifera, where monothalamids are either naked or have an agglutinated wall and Tubolothalamea and Globothalamea have tubular and globular calcareous chambers respectively (Pawlowski et al., 2013). Within the Polycystines, they have differentiated in the symmetry, being radial or spherical in Spumellaria and heteropolar in Nassellaria. Finally, Collodaria as seen by its very divergent rDNA, have suffered a drastic change by simplification of the skeleton.

The importance of the skeleton symmetry in the classification of the different lineages of Radiolaria has taken relevance in recent years thanks to developments in single cell DNA barcoding and its combination with morphological data (Decelle et al., 2012b; Sandin et al., 2019, Chapter 1.1; Sandin et al. *in prep*., Chapter 1.2). That led to the reinterpretation of the fossil record, considering ancient radiolarian morphologies as possible ancestors of Nassellaria (Sandin et al., 2019, Chapter 1.1) and Spumellaria (Sandin et al. *in prep.*, Chapter 1.2). Yet, the most likely ancestor of Spumellaria is considered among primitive polycystines forms (Archaeospicularia: Pouille et al., 2011; Aitchison et al., 2017; Sandin et al. *in prep*., Chapter 1.2). And if these fossils are the ancestors of all Polycystines, as proposed in De Wever et al. (2001) and Noble et al. (2017), Nassellaria should be emended within Spumellaria. Although this scenario is not supported by any molecular phylogeny, and neither does the fossil record. Therefore, and based on the symmetric pattern of the skeleton and our phylogenetic results, nassellarian skeleton probably evolved independently from that of Spumellaria, as Afanasieva et al. (2005) have suggested. Simple forms like the extinct genus *Palaeospiculum* (Archaeospicularia) could be possible early ancestors of Nassellaria, diversifying later in the first heteropolar shapes (Proventocitidae). This could also explain the presence of the extinct order Albaillelaria with bilateral symmetry.

Evolutionary history of Radiolaria

The early Neoproterozoic ocean is characterized by a gradual eukaryotic diversification mostly driven by predation (Cohen and Macdonald, 2015; Loron et al., 2018). Probably, as suggested by Cavalier-Smith et al. (2018), an amoeboid Retaria diversified in two different ecosystems, the planktonic and the benthic, in order to expand their niches and avoid

competition (Fig. 3). This differentiation would have given later rise to the Radiolaria and the Foraminifera lineages respectively. During the mid-Neoproterozoic the drifting of the supercontinent Rodinia has been proposed as a trigger for an increased weathering and high biomineralization rates, being mainly abiotic factors shaping eukaryotic speciation (Halverson et al., 2010; Cohen and Macdonald, 2015). Biomineralization was already documented in protist (Cohen et al., 2011) at this time and both Foraminifera (Pawlowski et al., 2003a; Groussin et al., 2011) and, based on the molecular clock, silicate radiolarians diversified for the first time. During the late Neoproterozoic the first representatives of fossilized Foraminifera appeared in the fossil record (Bosak et al., 2012). At the same time, our molecular clock estimates the diversification between Acantharea and Rad-X, and that of Polycystines and those silicate environmental lineages. Shortly afterwards, planktic algae radiate for the first time driven by the melting of the long Sturtian glaciation (Brocks et al., 2017). Also at this moment, appears for the first time, and in relatively big concentrations, sterane biomarkers related to Rhizaria (Nettersheim et al., 2019). Most likely, the end of the long glaciation favored the development of the different lineages of Radiolaria and allowed their proliferation and establishment in the oceans as important predators.

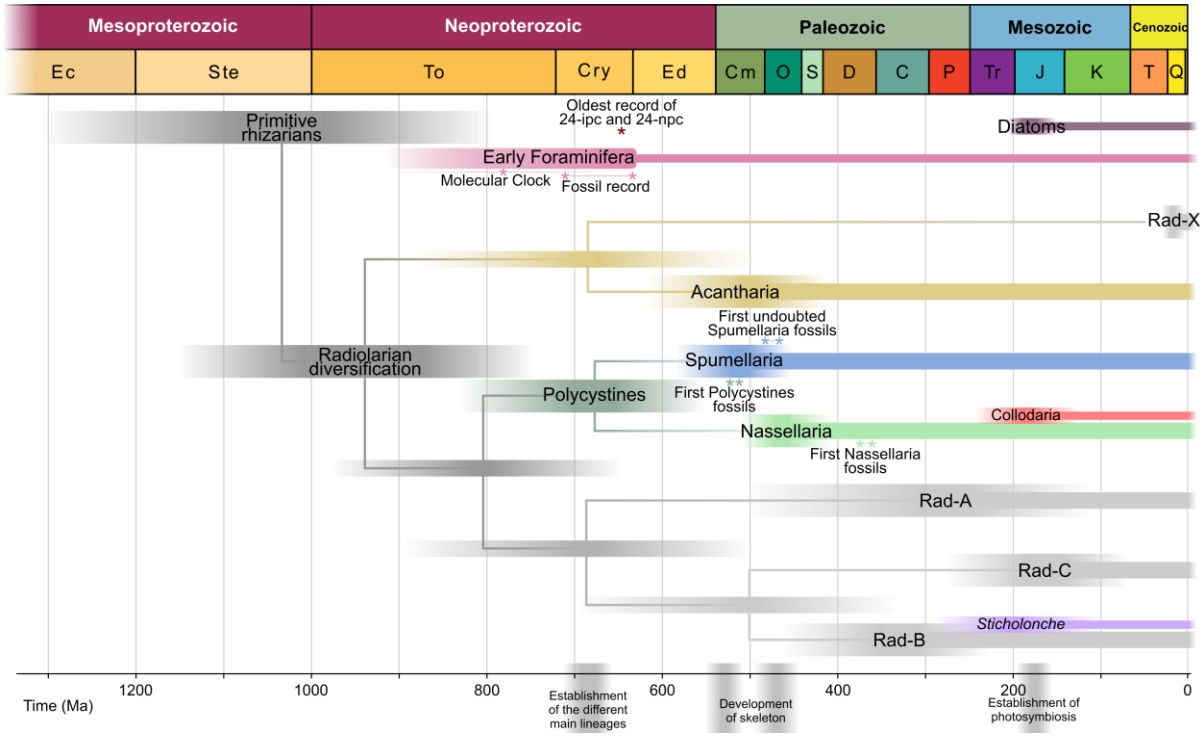


Figure 3. Schematic representation of radiolarian molecular evolution, along with possible biotic and abiotic factors contributing to their diversification. See text (Evolutionary history of Radiolaria) for more detailed explanation and references.

In the early phanerozoic multicellular animal lineages were established (Erwin et al., 2011) and potentially bigger, macrophagous metazoans have been proposed to be responsible of the development of skeleton among protists to escape predation in the early Cambrian (Butterfield, 1997; Porter, 2011). This hypothesis might explain the sudden and independent diversification of the different radiolarian groups (Acantharia, Spumellaria, Nassellaria) dated in our results and found in the fossil record (De Wever et al., 2001;

Afanasieva et al., 2005; Aitchison et al., 2017). On the contrary, Foraminifera probably developed its skeleton earlier due to the benthic adaptation and the higher mineralization rate, as proposed for animals (Cohen, 2005; Wood, 2011). After this morphological innovation, Radiolaria become one of the most important planktonic groups of the early Paleozoic, contributing significantly to the food web and the silica cycle in the onset of the Ordovician Plankton Revolution (Servais et al., 2008, 2016a; Kidder and Tomescu, 2016). On the other side, Rad groups (A, B, C and X) never developed a skeleton (with the exception of the more recent spicules in *Sticholonche*), probably due to a specialization towards deep environments where the predation pressure is lower and where Rad-B has been detected at high relative abundances (Pernice et al., 2016; Giner et al., 2019).

At the beginning of the Mesozoic contemporary groups of Radiolaria diversified. This timing is based on both the fossil record (De Wever et al., 2001; O'Dogherty et al., 2011), results from this study, as well as in previous molecular clock studies (Sandin et al., 2019, Chapter 1.1; Sandin et al. *in prep*., Chapter 1.2). The onset of the photosymbiosis between Acantharia and the haptophyte *Phaeocystis* has been estimated to be related to the extreme oligotrophy of oceanic waters during the Jurassic (Decelle et al., 2012a). Not only our results supported this hypothesis but extend it to other photosymbiotic taxa belonging to Nassellaria and Spumellaria since our estimates of the first symbiotic interactions with the dinoflagellates *Brandtodinium* (Probert et al., 2014) and *Gymnoxanthella* (Yuasa et al., 2016) happens to match at around ~180 Ma independently. Very likely acting in synergy along with oligotrophy to foster the establishment of photosymbiosis, at that period, diatoms appeared (~186 Ma; Lewitus et al., 2018) becoming an important competitor of Polycystines for Silica in the plankton realm (Conley et al., 2017; Hendry et al., 2018) and a long period of oceanic anoxia characterized the oceans during the Toarcian (183.7-174.2 Ma) (Jenkyns, 1998, 2010).

The establishment of the photosymbiosis brought a high ecological success in Radiolaria, allowing the heterotrophic host to thrive in the Jurassic ocean. Yet, the vast pelagic environment limits the encounter of host and symbiont threatening the fitness of the host in geological time scales. Probably extinction-diversification events favored a low specificity host-symbiont (due to the risk assessed to the host in finding the symbiont in the pelagic environment) along with previously abiotic factors proposed for their diversification (such as oceanic anoxic events: Leckie et al., 2002; Chapter 1.1, Sandin et al., 2019, Chapter 1.1; Sandin et al. *in prep.*, Chapter 1.2), contributed to the speciation of symbiotic groups and in a lesser term of non-symbiotic groups. This could explain the large variety of symbionts and the earlier appearance of extant Collodaria compared to the fossil record (dated to be in the Eocene (56-33.9 Ma), De Wever et al., 2001). Most likely, Collodaria diversified from ancient forms (Lineage II in Sandin et al. 2019, Chapter 1.1), and with the establishment of the photosymbiosis along with the ability to form colonies their fitness increased considerably diversifying over the following periods. Finally, with the opening of the drake passage (~41 Ma; Scher and Martin, 2006) Collodaria diversified into today's morphologies.

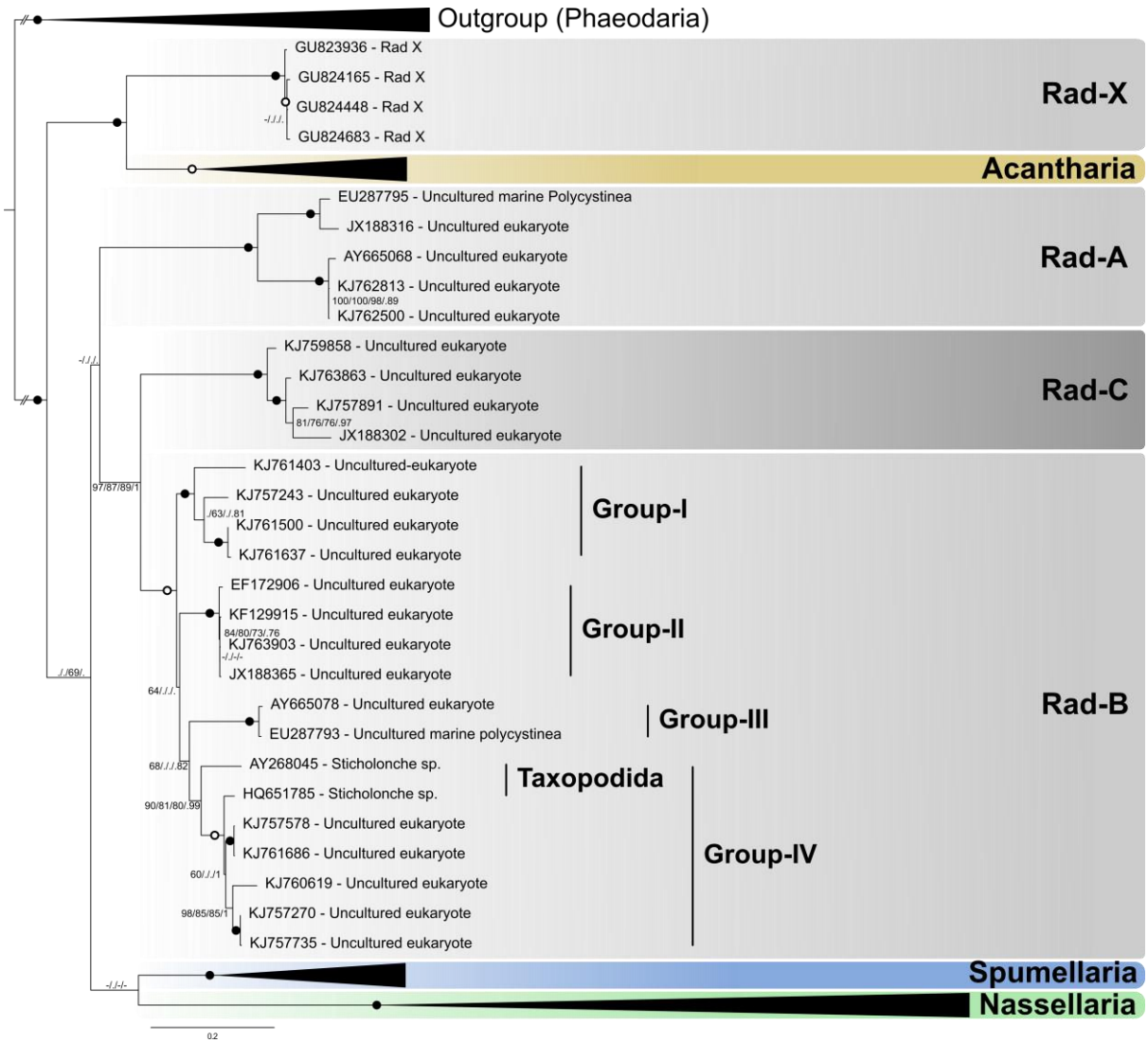
Conclusion

The main change introduced in the revised Radiolaria classification presented here is the inclusion of Collodaria within the order Nassellaria. Although it can still refer as a

shortening of colonial polycystines. Combining this morpho-molecular framework with previous hypothesis on amoeboid protists or clades (Nikolaev et al., 2004; Chapter 1.2) and environmental surveys (Pernice et al., 2016; Giner et al., 2019), we argued the possibility of different and new morphologies within Radiolaria (skeleton-less) as possible candidates for the great environmental diversity present at basal nodes, probably with a certain preference towards deep environments. Future research combining novel sequencing techniques able to sequence long and low concentrated molecules (such as Oxford Nanopore or PacBio) with single-cell morphological data may improve the current morpho-molecular framework.

Using fossil-calibrated phylogenies we inferred the radiation of Radiolaria in the early Neoproterozoic (~940 Ma). Other studies have also reported an early and slow diversification in testate Amoebae along with the gradual oxygenation of the oceans (Lahr et al., 2019). The early evolution resulted by molecular dating contrast the late diversification proposed by the fossil record and hypothesized by Cavalier-Smith et al. (2018). Although, the fossil record is showing minimum ages, and during the mid-Neoproterozoic (Cryogenian: ~720-~635 Ma) has been noticed a bias in the assemblages diversity mainly due to long periods of glaciations. Therefore, the early radiation of Radiolaria, along with that of other lineages, such as Foraminifera (Pawlowski et al., 2003a; Groussin et al., 2011) or Arcellinids (Lahr et al., 2019), support the idea of an early diversification of eukaryotes long before the oxygenation of the oceans (Butterfield, 2015). Yet, the exploration of early diverging processes through the direct observation of the fossil record is arduous due to the relative recent development of hard body structures. Besides, phylogenetic patterns at broad taxonomic level (such as that of Retaria and the position of Foraminifera) remain still elusive obscuring the interpretation and understanding of the drivers shaping the diversification.

Supplementary Material



Supplementary Figure 1. Molecular phylogeny of Radiolaria inferred from the concatenated complete 18S and partial 28S (D1-D2 regions) rDNA genes (217 taxa and 2533 aligned positions). Branches with a double barred symbol are fourfold reduced for clarity. The tree corresponds to that of Fig. 1 with emphasis in the environmental clades. See Fig. 1 for further details.

Supplementary Table 1. List of taxa used to obtain radiolarian phylogenetic study.

Clade	Reference	environment	Accession number 18S	Accession number 28S
Rad-X	Edgcomb et al. (2011)	micro-oxic water column sample, Cariaco Basin, Caribbean Sea	GU823936	
Rad-X	Edgcomb et al. (2011)	oxygenated water column sample, Cariaco Basin, Caribbean Sea	GU824165	
Rad-X	Edgcomb et al. (2011)	oxygenated water column sample, Cariaco Basin, Caribbean Sea	GU824448	
Rad-X	Edgcomb et al. (2011)	anoxic water column sample, Cariaco Basin, Caribbean Sea	GU824683	
A	Lopez-Garcia et al. (2006)	Antarctic	AF290072	
A	Decelle et al. (2012b)	Eilat, Israel	JN811161	JN811246
A	Gilg et al. (2010)	north western Pacific, San Pedro Channel	GU246591	
Acan 2	Wu et al. (2014)	60 m water sample from the South China Sea	KF129879	
Acan 2	Wu et al. (2014)	60 m water sample from the South China Sea	KF130152	
Acan 2	Wu et al. (2014)	60 m water sample from the South China Sea	KF129749	
Acan 2	Not et al. (2007)	Sargasso Sea; 500; picoplankton	EF172802	
Acan 3	Not et al. (2007)	Sargasso Sea; 3000; picoplankton	EF172929	
Acan 3	Not et al. (2007)	Sargasso Sea; 3000; picoplankton	EF172909	
Acan 3	Edgcomb et al. (2011)	anoxic water column sample, Cariaco Basin, Caribbean Sea	GU822868	

Acan 3	Gilg et al. (2010)	north western Pacific, San Pedro Channel	GU246582	
Acan 1	Wu et al. (2014)	75m water sample from the South China Sea	JX188361	
Acan 1	Wu et al. (2014)	60 m water sample from the South China Sea	KF129913	
Acan 1	Edgcomb et al. (2011)	micro-oxic water column sample, Cariaco Basin, Caribbean Sea	GU821070	
Acan 1	Edgcomb et al. (2011)	micro-oxic water column sample, Cariaco Basin, Caribbean Sea	GU822163	
B1	Lie et al. (2014)	East Pacific Rise 1500m	KJ761157	
B1	Edgcomb et al. (2011)	anoxic water column sample, Cariaco Basin, Caribbean Sea	GU825535	
B1	Decelle et al. (2012b)	Eilat, Israel	JN811162	JN811247
B1	Lie et al. (2014)	Gulf Stream 2500m	KJ760026	
B2	Lie et al. (2014)	Arctic Ocean 500m	KJ758551	
B2	Decelle et al. (2012b)	Villefranche-sur-Mer, France	JN811211	JN811318
B2	Decelle et al. (2012b)	Villefranche-sur-Mer, France	JN811206	JN811310
B2	Decelle et al. (2012b)	Villefranche-sur-Mer, France	JN811207	JN811314
C	Decelle et al. (2012b)	Akajima Island, Japan	JN811176	JN811266
C	Decelle et al. (2012b)	Akajima Island, Japan	JN811197	JN811291
C	Decelle et al. (2012b)	Eilat, Israel	JN811163	JX660724
C	Decelle et al. (2012b)	Eilat, Israel	JN811157	JN811240
C	Decelle et al. (2012b)	Akajima Island, Japan	JN811171	JN811261
D	Decelle et al. (2012b)	Roscoff, France	JN811202	JN811298
D	Decelle et al. (2012b)	Akajima Island, Japan	JN811182	JN811272
D	Decelle et al. (2012b)	Akajima Island, Japan	JN811184	JN811275
D	Decelle et al. (2012b)	Akajima Island, Japan	JN811174	JN811264
Acan4	Lie et al. (2014)	SPOTstation 5m	KJ763045	
Acan4	Lie et al. (2014)	Gulf Stream 105m	KJ763536	
Acan4	Wu et al. (2014)	60 m water sample from the South China Sea	KF129893	
Acan4	Lie et al. (2014)	East Pacific Rise 1500m	KJ760964	
E	Decelle et al. (2012b)	Villefranche-sur-Mer, France	JN811204	JN811307
E	Decelle et al. (2012b)	Akajima Island, Japan	JN811166	JN811254
E	Decelle et al. (2012b)	Akajima Island, Japan	JN811185	JN811276
E	Decelle et al. (2012b)	Akajima Island, Japan	JN811180	JN811270
E	Decelle et al. (2012b)	Akajima Island, Japan	JN811193	JN811287
F	Decelle et al. (2012b)	Villefranche-sur-Mer, France	JN811218	JN811326
F	Decelle et al. (2012b)	Akajima Island, Japan	JN811192	JN811285
F	Decelle et al. (2012b)	Eilat, Israel	JN811150	JN811232
F	Decelle et al. (2012b)	Villefranche-sur-Mer, France	JN811217	JN811325
F	Decelle et al. (2012b)	Eilat, Israel	JN811159	JN811242
F	Decelle et al. (2012b)	Eilat, Israel	JN811160	JN811243
Rad-A	Viprey et al. (2008)	cruise OLIPAC, Stn 11, depth 75m, picoplancton	EU287795	
Rad-A	Wu et al. (2014)	75m water sample from the South China Sea	JX188316	
Rad-A	Armbrust et al. (2004)	Sargasso Sea, DCM	AY665068	
Rad-A	Lie et al. (2014)	SPOTstation 500m	KJ762813	
Rad-A	Lie et al. (2014)	SPOTstation 500m	KJ762500	
Rad-C	Lie et al. (2014)	Gulf Stream 2500m	KJ759858	
Rad-C	Lie et al. (2014)	Gulf Stream 105m	KJ763863	
Rad-C	Lie et al. (2014)	Ross Sea 20m	KJ757891	
Rad-C	Wu et al. (2014)	75m water sample from the South China Sea	JX188302	
Group-I	Lie et al. (2014)	East Pacific Rise 1500m	KJ761403	
Group-I	Lie et al. (2014)	East Pacific Rise 2500m	KJ757243	
Group-I	Lie et al. (2014)	East Pacific Rise 1500m	KJ761500	
Group-I	Lie et al. (2014)	East Pacific Rise 1500m	KJ761637	
Group-II	Not et al. (2007)	Sargasso Sea; 3000; picoplankton	EF172906	
Group-II	Wu et al. (2014)	60 m water sample from the South China Sea	KF129915	
Group-II	Lie et al. (2014)	Gulf Stream 105m	KJ763903	
Group-II	Wu et al. (2014)	75m water sample from the South China Sea	JX188365	
Group-III	Armbrust et al. (2004)	Sargasso Sea, DCM	AY665078	
Group-III	Viprey et al. (2008)	cruise OLIPAC, Stn 11, depth 75m, picoplancton	EU287793	
Group-IV	Nikolaev et al. (2004)	Villefranche-sur-Mer, France	AY268045	
Group-IV	Krabberod et al. (2011)	Norway	HQ651785	HQ651785
Group-IV	Lie et al. (2014)	East Pacific Rise 2500m	KJ757578	
Group-IV	Lie et al. (2014)	East Pacific Rise 1500m	KJ761686	
Group-IV	Lie et al. (2014)	East Pacific Rise 20m	KJ760619	
Group-IV	Lie et al. (2014)	East Pacific Rise 2500m	KJ757270	
Group-IV	Lie et al. (2014)	East Pacific Rise 2500m	KJ757735	
Eucyrtidioidea	Sandin et al. (2019)	Japan, Okinawa, Sesoko	MK396929	
Eucyrtidioidea	Krabberod et al. (2011)	Norway	HQ651779	HQ651779
Eucyrtidioidea	Kunitomo et al. (2006)		AB246681	
Eucyrtidioidea	Sandin et al. (2019)	Mediterranean Sea, Villefranche-sur-Mer, 0-30m	MK396948	MK397002
Eucyrtidioidea	Sandin et al. (2019)	Western Mediterranean Sea, 0-500m	MK396915	
Carpocaniidae	Sandin et al. (2019)	Western Mediterranean Sea, 0-500m	MK396913	
Carpocaniidae	Sandin et al. (2019)	Western Mediterranean Sea, 0-500m	MK396914	
Cycladophoridae	Sandin et al. (2019)	Japan (40.499, 143.849), 2000-3000m		MK396963

Cycladophoridae	Sandin et al. (2019)	Japan (39.021, 143.691), 2000-3000m	MK396919	MK396960
Lychnocanoidea	Sandin et al. (2019)	Japan (39.737, 143.894), 2000-3000m	MK396921	MK396962
Lychnocanoidea	Kunitomo et al. (2006)		AB246682	
Lychnocanoidea	Sandin et al. (2019)	Mediterranean Sea, Villefranche-sur-Mer	MK396934	MK396982
Nass4	Kim et al. (2012)	Pacific Ocean: Eastern North Pacific	JX842092	
Nass4	Kim et al. (2012)	Pacific Ocean: Eastern North Pacific	JX842410	
Pterocorythoidea	Kunitomo et al. (2006)		AB246697	
Pterocorythoidea	Ando et al. (2009)	Japan (34.80N, 138.54E)	AB430759	AB430759
Pterocorythoidea	Sandin et al. (2019)	Mediterranean Sea, Villefranche-sur-Mer, 0-30m	MK396946	MK396999
Pterocorythoidea	Sandin et al. (2019)	Mediterranean Sea, Villefranche-sur-Mer, surface	MK396942	MK396995
Pterocorythoidea	Sandin et al. (2019)	Mediterranean Sea, Villefranche-sur-Mer, surface	MK396943	MK396996
Archipilioidea	Sandin et al. (2019)	Japan (39.021, 143.691), 2000-3000m		MK396958
Archipilioidea	Sandin et al. (2019)	Japan (39.021, 143.691), 2000-3000m	MK396920	MK396961
Theopiliidae	Sandin et al. (2019)	Western Mediterranean Sea, 0-500m	MK396917	MK396957
Theopiliidae	Kunitomo et al. (2006)		AB246685	
Theopiliidae	Sandin et al. (2019)	Japan, Okinawa, Sesoko	MK396926	MK396972
Plagiacanthoidea	This study	Western Mediterranean Sea, 0-500m		in prep.
Plagiacanthoidea	Krabberod et al. (2011)	Norway	HQ651801	
Plagiacanthoidea	Krabberod et al. (2011)	Norway	HQ651802	
Plagiacanthoidea	Kunitomo et al. (2006)		AB246694	
Plagiacanthoidea	Sandin et al. (2019)	Japan, Okinawa, Sesoko	MK396927	
Plagiacanthoidea	Sandin et al. (2019)	Western Mediterranean Sea, 0-500m		MK396952
Plagiacanthoidea	Sandin et al. (2019)	Mediterranean Sea, Villefranche-sur-Mer, 0-30m	MK396949	MK397003
Plagiacanthoidea	Sandin et al. (2019)	Japan (39.001, 141.002)	MK396922	MK396967
Plagiacanthoidea	Sandin et al. (2019)	Japan (38.008, 142.008)		MK396966
Acanthodesmoidea	Krabberod et al. (2011)	Norway	HQ651791	
Acanthodesmoidea	Sandin et al. (2019)	Japan, Okinawa, 0-150m	MK396930	MK396974
Acanthodesmoidea	Sandin et al. (2019)	Mediterranean Sea, Villefranche-sur-Mer	MK396937	MK396987
Acanthodesmoidea	Sandin et al. (2019)	Mediterranean Sea, Villefranche-sur-Mer	MK396931	MK396977
Nass2	Xu et al. (2016)	South China Sea, Bathypelagic	KX532523	
Nass2	Not et al. (2007)	Indian Ocean; 75m depth	EU562143	
Artostrobioidea	Sandin et al. (2019)	Japan, Okinawa, Sesoko	MK396925	MK396971
Artostrobioidea	Sandin et al. (2019)	Western Mediterranean Sea, 0-500m	MK396918	
Artostrobioidea	Sandin et al. (2019)	Western Mediterranean Sea, 0-500m	MK396916	
Nass3	Lie et al. (2010)	Nansha sea area, South China Sea	GU553071	
Nass3	Lie et al. (2014)	SPOTstation 500m	KJ762790	
Nass3	Kim et al. (2012)	Pacific Ocean: Eastern North Pacific	JX842634	
Acropyramioidea	Sandin et al. (2019)	Japan (39.021, 143.691), 2000-3000m		MK396959
Acropyramioidea	This study	Western Mediterranean Sea, 0-500m		in prep.
Nass1	Not et al. (2007)	Sargasso Sea; 500; picoplankton	EF172833	
Nass1	Xu et al. (2016)		KX532876	
Orosphaeroidea	Nakamura et al. (in prep.)	Eastern North Pacific; 0-1000m	in prep.	in prep.
Orosphaeroidea	Nakamura et al. (in prep.)	Eastern North Pacific; 0-1000m	in prep.	in prep.
Sphaerozoidae	Biard et al. (2015)	Villefranche-sur-Mer, France	KR058249	KR058307
Sphaerozoidae	Biard et al. (2015)	North Atlantic Ocean	KR058224	KR058280
Sphaerozoidae	Biard et al. (2015)	South Pacific Ocean	KR058225	KR058282
Sphaerozoidae	Biard et al. (2015)	Sesoko, Japan	KR058229	KR058286
Collophidioidea	Biard et al. (2015)	South Atlantic Ocean	KR058212	KR058269
Collophidioidea	Biard et al. (2015)	Indian Ocean	KR058213	KR058270
Collophidioidea	Biard et al. (2015)	North Pacific Ocean	KR058219	KR058275
Collophidioidea	Biard et al. (2015)	North Pacific Ocean	KR058218	KR058274
Collosphaeroidea	Biard et al. (2015)	South Pacific Ocean	KR058196	KR058259
Collosphaeroidea	Biard et al. (2015)	South Pacific Ocean	KR058197	KR058260
Collosphaeroidea	Biard et al. (2015)	South Atlantic Ocean	KR058201	KR058264
Collosphaeroidea	Biard et al. (2015)	South Atlantic Ocean	KR058202	KR058265
Spum6	Edgcomb et al. (2011)	micro-oxic water column Cariaco Basin	GU820872	
Spum6	Edgcomb et al. (2011)	oxygenated water column Cariaco Basin	GU821243	
Spum6	Lie et al. (2014)	SPOTstation; 500	KJ762797	
Spum6	Lie et al. (2014)	SPOTstation; 500	KJ762533	
Spum5	Li et al. Unpublished	central Pacific Ocean	KP175032	
Spum5	Edgcomb et al. (2011)	oxygenated water column Cariaco Basin	GU822576	
Spum5	Lie et al. (2014)	East Pacific Rise; 1500	KJ760904	
Spum5	Lie et al. (2014)	East Pacific Rise; 1500	KJ761666	
Spum5	Lie et al. (2014)	SPOTstation; 500	KJ762723	
Spum5	Lie et al. (2014)	East Pacific Rise; 1500	KJ761107	
Liosphaeroidea	Krabberod et al. (2011)	Sogndalsfjorden (Norway)	HQ651782	HQ651782
Liosphaeroidea	Yuasa et al. (2009)	Okinawa, Japan	AB490706	
Liosphaeroidea	Yuasa et al. (2009)	Okinawa, Japan	AB490705	
Liosphaeroidea	Sandin et al. (in prep.)	Villefranche-sur-mer	in prep.	in prep.
Liosphaeroidea	Sandin et al. (in prep.)	Villefranche-sur-mer; 30	in prep.	in prep.
Liosphaeroidea	Sandin et al. (in prep.)	Villefranche-sur-mer; 30	in prep.	in prep.

Spum1	Wu et al. (2014b)	South China; 60	KF130443	
Spum1	Wu et al. (2014a)	South China; 75	JX188300	
Spum1	Not et al. (2008)	ocean water; 75	EU562119	
Spum2	Edgcomb et al. (2011)	oxygenated water column Cariaco Basin	GU821466	
Spum2	Edgcomb et al. (2011)	oxygenated water column Cariaco Basin	GU820920	
Spum2	Edgcomb et al. (2011)	micro-oxic water column Cariaco Basin	GU821583	
Hexastyloidea	Sandin et al. (in prep.)	Villefranche-sur-mer; 30	in prep.	in prep.
Hexastyloidea	Edgcomb et al. (2011)	micro-oxic water column Cariaco Basin	GU822737	
Hexastyloidea	Krabberod et al. (2011)	Sogndalsfjorden (Norway)	HQ651796	
Hexastyloidea	Yuasa et al. (2009)	Sogndalsfjord, Norway; 250-25	AB284519	
Hexastyloidea	Krabberod et al. (2011)	Sogndalsfjorden (Norway)	HQ651784	HQ651784
Hexastyloidea	Sandin et al. (in prep.)	Villefranche-sur-mer	in prep.	
Hexastyloidea	Krabberod et al. (2011)	Sogndalsfjorden (Norway)	HQ651798	
Hexastyloidea	Krabberod et al. (2011)	Sogndalsfjorden (Norway)	HQ651795	
Hexastyloidea	Sandin et al. (in prep.)	Villefranche-sur-mer; Euphotic	in prep.	in prep.
Hexastyloidea	Sandin et al. (in prep.)	Villefranche-sur-mer; Euphotic	in prep.	in prep.
Hexastyloidea	Sandin et al. (in prep.)	Villefranche-sur-mer; 450-0	in prep.	
Spongosphaeroidea	Sandin et al. (in prep.)	Westearn mediterranean sea; 500-0	in prep.	in prep.
Spongosphaeroidea	Sandin et al. (in prep.)	East off Japan; 25-0	in prep.	
Lithocyclioidea	Sandin et al. (in prep.)	Villefranche-sur-mer	in prep.	in prep.
Lithocyclioidea	Yuasa et al. (2016)	Okinawa, Japan; Surface	LC093106	
Lithocyclioidea	Yuasa et al. (2004)	Okinawa, Japan	AB193605	
Spongodiscoidea	Ando et al. (2009)	Shimoda, Izu peninsula, Japan	AB430760	AB430760
Spongodiscoidea	Sandin et al. (in prep.)	Westearn mediterranean sea; 500-0	in prep.	in prep.
Spongodiscoidea	Sandin et al. (in prep.)	Westearn mediterranean sea; 500-0	in prep.	in prep.
Spongodiscoidea	Ando et al. (2009)	Shimoda, Izu peninsula, Japan	AB430757	AB430757
Spongodiscoidea	Ando et al. (2009)	Shimoda, Izu peninsula, Japan	AB430758	AB430758
Rhizosphaeroidea	Sandin et al. (in prep.)	Villefranche-sur-mer	in prep.	in prep.
Rhizosphaeroidea	Sandin et al. (in prep.)	Westearn mediterranean sea; 500-0	in prep.	in prep.
Rhizosphaeroidea	Sandin et al. (in prep.)	Villefranche-sur-mer; Surface	in prep.	in prep.
Rhizosphaeroidea	Sandin et al. (in prep.)	Westearn mediterranean sea; 500-0	in prep.	in prep.
Centrocubidae	Sandin et al. (in prep.)	Sesoko Okinawa Japan	in prep.	in prep.
Centrocubidae	Edgcomb et al. (2011)	oxygenated water column Cariaco Basin	GU821394	
Centrocubidae	Edgcomb et al. (2011)	oxygenated water column Cariaco Basin	GU822672	
Excentroconchidae	Lie et al. (2014)	East Pacific Rise; 2500	KJ757677	
Excentroconchidae	Edgcomb et al. (2011)	anoxic water column Cariaco Basin	GU822292	
Excentroconchidae	Sandin et al. (in prep.)	Villefranche-sur-mer; 450-0	in prep.	
Excentroconchidae	Sandin et al. (in prep.)	Villefranche-sur-mer	in prep.	
Spum3	Xu et al. (2016)	ocean water	KX532622	
Spum3	Lie et al. (2014)	Gulf Stream; 2500	KJ760039	
Pylonioidea	Ishitani et al. (2012)	North Pacific Ocean; 150-0	AB617585	
Pylonioidea	Sandin et al. (in prep.)	Sesoko Okinawa Japan	in prep.	in prep.
Pylonioidea	Sandin et al. (in prep.)	Villefranche-sur-mer	in prep.	in prep.
Pylonioidea	Sandin et al. (in prep.)	Villefranche-sur-mer; 450-0	in prep.	in prep.
Pylonioidea	Sandin et al. (in prep.)	Villefranche-sur-mer; 450-0	in prep.	in prep.
Spum4	Xu et al. (2016)	ocean water	KX533313	
Spum4	Lie et al. (2014)	East Pacific Rise; 2500	KJ757620	
Spum4	Lie et al. (2014)	Gulf Stream; 2500	KJ759906	
Spum4	Lie et al. (2014)	Gulf Stream; 2500	KJ760144	
Stylodictyoidea	Sandin et al. (in prep.)	Villefranche-sur-mer; 450-0	in prep.	
Stylodictyoidea	Sandin et al. (in prep.)	Villefranche-sur-mer; 450-0	in prep.	in prep.
Stylodictyoidea	Ishitani et al. (2014)	Central Pacific	AB860147	AB860147
Stylodictyoidea	Sandin et al. (in prep.)	Villefranche-sur-mer; 30	in prep.	in prep.
Actinommoidea	Krabberod et al. (2011)	Sogndalsfjorden (Norway)	HQ651781	HQ651781
Actinommoidea	Krabberod et al. (2011)	Sogndalsfjorden (Norway)	HQ651780	HQ651780
Actinommoidea	Krabberod et al. (2011)	Sogndalsfjorden (Norway)	HQ651788	
Actinommoidea	Krabberod et al. (2011)	Sogndalsfjorden (Norway)	HQ651789	
Spongopyloidea	Ishitani et al. (2014)	Central Pacific	AB860150	AB860150
Spongopyloidea	Ishitani et al. (2014)	Central Pacific	AB860156	AB860156
Spongopyloidea	Sandin et al. (in prep.)	Villefranche-sur-mer; Euphotic	in prep.	in prep.
Spongopyloidea	Sandin et al. (in prep.)	Villefranche-sur-mer; Euphotic	in prep.	in prep.

Chapter 3

From intracellular variability to community ecology

Context of the work

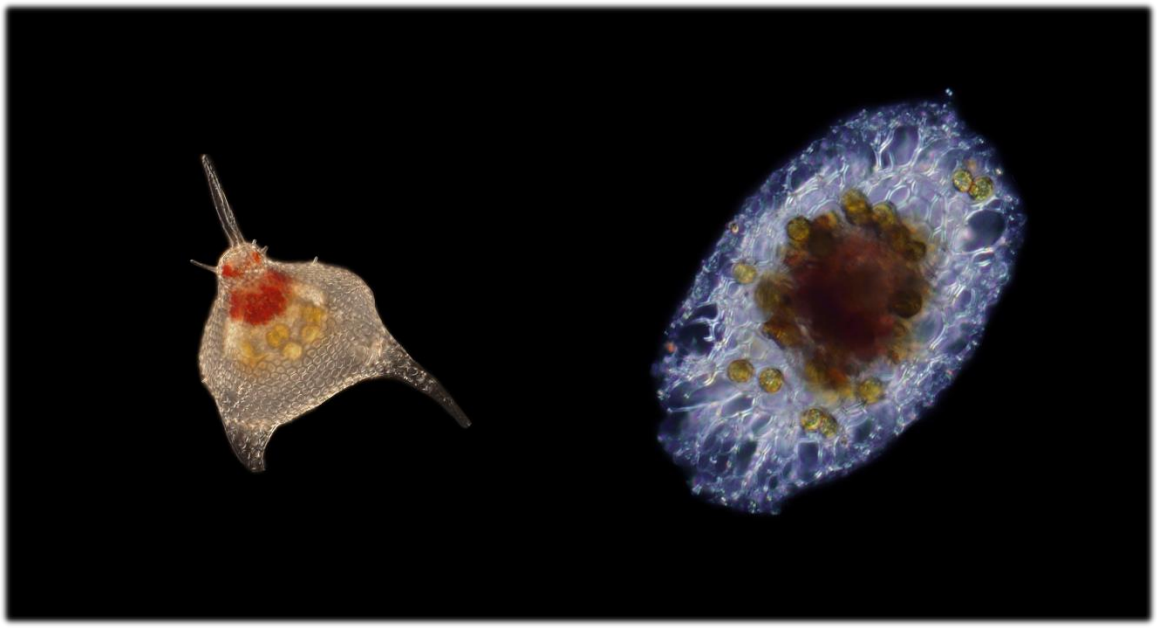
In this chapter I investigated the intracellular molecular variability of ribosomal taxonomic marker genes using various sequencing technologies. Considering this later information along with the reference morpho-molecular framework built in previous chapters I explored the molecular diversity and ecology of Nassellaria and Spumellaria by a metabarcoding approach in the global ocean.

The emerging sequencing platform of Oxford Nanopore Technologies (ONT) provides a promising High-Through Put approach sequencing long reads and using direct DNA amplification (PCR-free). We wanted to compare this technology with the current mostly used High-Through Put sequencing platform, Illumina, to better understand the potential of ONT technology for future environmental molecular diversity surveys and evolutionary purposes.

The second part of this chapter explored the biodiversity and biogeography of Radiolaria through a metabarcoding approach. By combining the output obtained in different global environmental surveys, such as Tara Oceans and Malaspina expeditions, with regional surveys, such as MOOSE-GE, I aimed at obtaining a picture of the extant molecular diversity and biogeography of Radiolaria in the world oceans both at surface and along the water column depth.

Chapter 3.1

Intracellular gene variability in Nassellaria and Spumellaria



Intra-cellular 18S rDNA gene variability of Nassellaria and Spumellaria (Rhizaria, Radiolaria) assessed by Sanger, MinION and Illumina sequencing

Miguel M. Sandin, Sarah Romac, Fabrice Not

Sorbonne Université, CNRS - UMR7144 - Ecology of Marine Plankton Group - Station Biologique de Roscoff, 29680 Roscoff, France.

In preparation

Introduction

Ribosomal DNA (rDNA) genes are known to be valuable markers for the barcoding of eukaryotic life and its phylogenetic classification at different taxonomic levels. This is mainly due to its intra-genomic tandem repeated structure, the presence of conserved and variable regions and its appearance in all eukaryotes (Pawlowski et al., 2012; del Campo et al., 2018). The 18S rDNA gene, coding for the small sub-unit of the ribosome, has been widely used in molecular environmental surveys, especially the short hyper-variable regions V4 and V9, thanks to the extensive occurrence in public databases and the presence of generalist primers flanking their sides (Amaral-zettler et al., 2009; Stoeck et al., 2010). The advent of High-Throughput Sequencing (HTS) techniques has allowed the massive sequencing of molecular environmental diversity supporting its exploration through a metabarcoding approach (de Vargas et al., 2015; Massana et al., 2015; Forster et al., 2016; Pernice et al., 2016). The large amount of reads (or amplicons) generated by HTS is normally classified in Operational Taxonomic Units (OTUs) based on arbitrary similarity thresholds. OTUs are used not only to identify taxonomic entities, but also to describe community structure (Blaxter et al., 2005). The increasing use of the HTS, have led to the development of different clustering methods resulting in finer-scale OTUs that focus in single nucleotide differences (Mahé et al., 2015) or in the correction of sequencing errors based in the error rate entropy (so-called Amplicon Sequence Variants or ASVs; Callahan et al., 2016).

HTS produce a vast amount of reads carrying errors that are difficult to distinguish from real biological variation, that is considered as a main factor inflating diversity (Kunin et al., 2010). Intragenomic rDNA Polymorphism and its different copy numbers among taxa can also affect diversity assessments (Gong et al., 2013; Gong and Marchetti, 2019). Other less common causes, yet important, have also been reported as factors inflating diversity estimates, such as lateral gene transfer (Yabuki et al., 2014) or presence of pseudogenes (Thornhill et al., 2007). Regarding the considerable sequencing depth of the HTS technologies it is possible that most of this intra-cellular diversity could be sequenced, leading to a large

overestimation of the environmental diversity. Several studies have argued that the number of molecular clusters largely exceed that of morphological counts (Medinger et al., 2010; Bachy et al., 2013; Santoferrara et al., 2016), leading to hypothesis that scientists are actually measuring intracellular variability (Caron and Hu, 2018). Current HTS technologies used for environmental surveys can sequence short fragments of DNA of about 400 base pairs (bp) only, such as the hyper-variable regions (V4 and V9 most commonly used in protist) of the 18S rDNA gene. Comparing such short hyper-variable regions to, far from exhaustive, reference sequences databases (Pitsch et al., 2019) may also contribute to inflating environmental diversity by misidentification of environmental clusters or lack of intragenomic rDNA variability representation, among other causes.

New sequencing technologies are arriving with the ability of high throughput sequencing longer nucleotide fragments in real time, such as Oxford Nanopore Technologies (ONT) or Pacific Bioscience (PacBio). These sequencing methods have already showed their useful capabilities, for example PacBio have developed a circular consensus sequencing resulting in near-zero error long reads, improving aspects such as genome assembly (Wenger et al., 2019) or even phylogenetic analysis of environmental diversity (Jamy et al., 2019). However, its limited high throughput sequencing ability and its relatively high cost (Goodwin et al., 2016), may affect the sequencing depth at a metabarcoding community level. On the other side, ONT provides a large amount of reads, inexpensively and highly portable with the MinION device (Levy and Myers, 2016). However, despite that the error rate of the ONT reads are improving since it was firstly released (for MinION, from ~40% in 2014 to <15% error rate; Rang et al., 2018) it is still a major concern, reaching up to 3-6% of errors in the best case scenarios (Tyler et al., 2018).

In this study, we assess the intracellular gene variability within protists by comparing three different sequencing methods, Sanger, ONT and Illumina. We focused our efforts in two groups of Radiolaria, the Nassellaria and Spumellaria (Polycystines). The Radiolaria is an important group of protists in the eukaryotic plankton community contributing for a major fraction of the total reads in environmental molecular surveys (de Vargas et al., 2015; Pernice et al., 2016). However, the number of morpho-species described among Radiolaria taxa (Suzuki and Not, 2015) does not match with the molecular barcodes found in environmental surveys. Nassellaria and Spumellaria environmental diversity lags far behind other radiolarian, despite of possessing the largest morphological diversity described. Recent studies have dwelled on exploring their extant morpho-molecular diversity (Sandin et al., 2019, Chapter 1.1; Sandin et al. in prep, Chapter 1.2), showing their uncharted diversity among Radiolaria (Sandin et al. in prep, Chapter 2). Previous studies have shown the low intracellular variability among Nassellaria, yet it can be important in other multinuclear radiolarian groups such as Acantharia (Decelle et al., 2014). The ecological importance of these groups and their observed low molecular diversity in environmental surveys stresses the need for understanding such differences among.

Material and Methods

Single-cell Sampling, isolation and DNA extraction

Plankton samples were collected in the Bay of Villefranche-sur-Mer (France) and in the West Mediterranean Sea (MOOSE-GE 2017 expedition) by plankton nets tows (from 20 to 64 µm mesh size). More information on sampling methodology for specific samples can be found in the RENKAN database (http://abims.sb-roscoff.fr/renkan/). Specimens were individually handpicked with Pasteur pipettes from the plankton community and maintained in 0.2 µm filtered seawater for several hours. They were transferred 3-4 times into new 0.2µm filtered seawater to allow self-cleaning from debris, particles attached to the cell or prey(s) digestion. By doing so, it is expected to keep only essential entities from the holobiont (the radiolarian host + associated symbionts and bacteria). Live specimens were imaged using an inverted microscope and thereafter transfer into 1.5 ml Eppendorf tubes containing 50 µl of molecular grade absolute ethanol and finally stored at -20 °C until DNA extraction. DNA was extracted using the MasterPure Complete DNA and RNA purification Kit (Epicentre) following manufacturer's instructions.

Amplification and sequencing

A schematic representation of the study design and the different amplification and sequencing steps can be found in Supplementary material Fig S1.

Four holobionts were selected to amplify the full length of the rDNA gene (Supplementary material Fig S2), 2 of them belonging to Spumellaria: Mge17-81 (*Rhizosphaera trigonacantha*, Rhizosphaeroidea) and Mge17-82 (*Spongosphaera streptacantha*, Spongosphaeroidea); and 2 to Nassellaria: Vil325 (*Eucyrtidium cienkowskii*, Eucyrtidioidea) and Vil496 (*Eucyrtidium acuminatum*, Eucyrtidioidea). Each holobiont was PCR-amplified in three different technical replicates with the eukaryotic primers SA (AACCTGGTTGATCCTGCCAGT) / D2C-R (CCTTGGTCCGTGTTTCAAGA) (Medlin et al., 1988; Scholin et al., 1994). Amplifications were performed with the enzyme Phusion® High-Fidelity DNA Polymerase (Finnzymes). The PCR mixture contained 1µL (~0.5ng) of DNA template with 0.35µM of each primer (final concentration), 3% of DMSO and 2X of GC buffer Phusion MasterMix (Finnzymes), in a final volum of 25µL. Amplifications were done in a SIMPLIAMP (Applied BiosystemsTM) with following the PCR program: initial denaturation step at 98°C for 30 s, 37 amplification cycles of 10 s at 98°C, 30 s at 55°C, 30 s at 72°C and final elongation step at 72°C for 10 minutes. For each of the 12 amplified samples, a 23-base pairs (bp) multiplex identifier or BC sequence was designed and included in one of the primers in order to identify the origin of every single read from the pooled population generated on a single run. Structures of the "BC-primers" were as follows: Primer 1: (5'- Tag + Forward primer -3'); Primer 2: (5'- Reverse primer -3'). To avoid single cell DNA sample contamination, PCR mixtures were carried out under a DNA–free vertical laminar flow hood. Negative and positive controls were added to check if there were no sample contaminations (DNA from operator, reagents or material contamination). PCR products were then purified using NucleoSpin®Gel and PCR Clean up kit (Macherey-NagelTM). Final elution product was divided in two for sequencing by Sanger (after cloning) and Oxford Nanopore technologies using the MinION device (Laver et al., 2015; Jain et al., 2016).

· *Sanger sequencing*: After purification, PCR amplicons were then cloned using TOPO TA Cloning for sequencing kit (Invitrogen) following manufacturer's recommendations. After a first step of 6 min ligation into pCRTM4-TOPO® vector, ligated PCR products were then transformed into TOP10 Chemically Competent Escherichia coli (ampicillin resistant) by heat-shock treatment at 42°C for 30 s. Cells were then incubated 2h at 37°C with rotation tray at 200 rpm with an enrichment medium. Afterwards, 80 µL of each transformant was spread on selective solid LB-agar medium (Sigma) with 50 µg/mL of ampicillin and incubated overnight at 37°C. On the following day, presence of the insert was verified in obtained colonies by PCR using the primers M13 forward and M13 reverse complementary to cloning site flanking sequences. A total of 24 clones from each amplicon were amplified directly with GoTaq polymerase (Promega, Lyon France) in a 25 µL reaction volume using the following PCR parameters: 10 min at 95 °C, 35 amplification cycles of 30 s of denaturation at 95 °C, 30s annealing at 56°C and 1 min extension at 72 °C, with a final elongation 10 min at 72 °C. PCR products were then purified using ExoStar (IllustraTM ExostarTM 1_Step, GE Healthcare Bio-Sciences Corp.). Final PCR product were sent to Macrogen Europe for sequencing using the primers SA (AACCTGGTTGATCCTGCCAGT; Medlin et al., 1988), S69f (AAHCTYAAAGGAAHTGACGG), D1Rf (ACCCGCTGAATTTAAGCATA; Scholin et al., 1994).

· *MinION library preparation and sequencing*: Each PCR product was pooled together in equimolar conditions to get 200 ng of final amplicon concentration in 45µL. Quality, size, and concentration of PCR products were assessed on a Bioanalyzer 2100 (Agilent) using a DNA1000 Labchip and on a Qbit fluorometer using a Qbit HS DNA quantification kit (Invitrogen). For this study, we optimized a protocol using Oxford Nanopore Technology (ONT) 1D^2 sequencing chemistry combined with Flowcell R 9.5.
Library preparation: DNA products (45 µL amplicon) were treated by an end-repair/dA tailing using NEBNext® Ultra™ II End Repair/dA-Tailing Module (New England Biolabs) and NEB Blunt/TA Ligase Master Mix (New England Biolabs), and D2 Adapter. Samples were washed using AMPure beads (Agencourt) and 70% EtOH and resuspended in 45 µL water. A first step of adapter ligation was conducted with 25 µL NEB Blunt/TA Ligase Master Mix (New England Biolabs), and 2.5µL 1D2 Adapter, and then carried out with 5µL BAM and 50µL of Blunt/TA ligase. A second washing step was done using AMPure beads (Agencourt) and 70% EtOH and resuspended in 15 µL water of final elution.
Sequencing: Pre sequence-library was prepared following ONT protocols by mixing 12 µL amplicon, 2.5 µL water, 25.5 µL LBB and 35 µL RBF. Flow cell R 9.5 was inserted in the MinION frame and priming port was loaded with 800 µL of priming mix (576 µL RBF and 624 µL water) with SpotON cover closed. We additionally loaded 200 µL of priming mix into the flowcell via the priming port, avoiding the introduction of air bubbles. 75µL of sample library was then added to SpotOn port via dropwise fashion. Finally, we covered SpotOn and priming ports, close the MinION lid and open the MinKNOW GUI software to proceed sequencing using a half of a total flow cell R 9.5.

· *Illumina sequencing*: In total 8 holobionts were selected to amplify the V4 region (~380pb) of the 18S rDNA gene (Supplementary material Fig S2), 4 belonging to Spumellaria: Mge17-81 (common to previous section), Mge17-82 (common to previous section), Vil480 (Tetrapyle octacantha, Pylonioidea), Vil497 (Arachnospongus varians, Liosphaeroidea); and 4 to

Nassellaria: Mge17-9 (Extotoxon undulatum, Artostrobioidea), Mge17-124 (Carpocanium obliqua, Carpocaniidae), Vil490 (Pterocorys cf. zanclea, Pterocorythoidea), Vil496 (common to previous section). Each holobiont was PCR-amplified with the eukaryotic primers TAR-EukF1 (5'-CCAGCA(G⁄C)C(C⁄T)GCGG-TAATTCC-3') / TAR-EukR3 (ACTTTCGTTCTTGAT(C⁄T)(A⁄G)A) (Stoeck et al., 2010). The forward primer had a barcode adapted for Illumina sequencing and each holobiont was amplified with 3 different barcode-adapted-primers to get 3 technical sequencing replicates. PCR reactions contained 1x MasterMix GC Phusion High-Fidelity DNA Polymerase (Finnzymes), 0.35µM of each primer, 3% dimethylsulphoxide and 1µL of DNA in a final volume of 25µl. The PCR program had an initial denaturation step at 98°C during 30 s, followed by 15 cycles of 10 s at 98°C, 30 s at 53°C and 30 s at 72°C, then 22 similar cycles but with 48°C annealing temperature, and a final elongation step at 72°C for 10 min. Vil325 gave no positive reactions and therefore was no possible to be included in this step. Polymerase chain reaction, in triplicates, were purified with NucleoSpin Gel and PCR Clean-Up kit (Macherey-Nagel), and quantified with the Quant-It PicoGreen double stranded DNA Assay kit (Invitrogen). About 500 ng of pooled amplicons were sent to Fasteris (https://www.fasteris.com, Switzerland) for Illumina sequencing on a Miseq nano V2 2x250.

Quality checking and similarity of amplicons

In order to compare the output of the three different sequencing results, similar and broadly known methods were used to quickly check the quality of the sequences/amplicons obtained. Firstly, raw sequences/amplicons were measured in base pairs (bp) length. Thereafter, sequences/amplicons were compared against reference sequences by local alignment (BLAST) for Sanger and MinION results and global alignment (vsearch) for Illumina results. This difference in the alignment algorithm used is due to the big differences in the sequence length and the method mostly used for each of the sequencing results. Finally, in order to obtain sequences per holobiont, Sanger reads belonging to the same replicate were concatenated manually, MinION reads were demultiplexing with cutadapt (Martin, 2011) and Illumina reads were clustered as described in the next section.

Clustering of V4 hyper-variable region

Two different pipelines were used to cluster the V4 hyper-variable region reads: a distance-based method and an error-based method. The former was carried out using swarm (Mahé et al., 2015) following the steps specified in https://github.com/frederic-mahe/swarm/wiki/Fred%27s-metabarcoding-pipeline for demultiplexing, filtering and clustering of reads with a difference threshold of 1 nucleotide. The second method was carried out using dada2 and the pipeline described in Callahan et al. (2016). Final amplicons were compared against PR2 v4.11.0 updated with Radiolaria sequences from Sandín et al. (in prep). Regarding sequences coming from Sanger and MinION sequencing, the V4 region was extracted with cutadapt (Martin, 2011) using the primer set used to amplify the V4 region (see previous section: TAR-EukF1 / TAR-EukR3). Sequences were later clustered using swarm with 1, 2 and 3 differences, and for MinION results dada2 was also used. Note dada2 was not used for Sanger sequences since the error-correction algorithm relies on number of reads.

Phylogenetic analysis

Minion sequences taxonomically assigned to Polycystines were used to carry out phylogenetic analysis. These sequences were aligned against a reference alignment using MAFFT v7.395 (Katoh and Standley, 2013) with a L-INS-i algorithm ('--localpair') and 1000 iterative refinement cycles. The reference alignment is composed of reference sequences (from Sandín et al., in prep, Chapter 2) and sequences obtained by Sanger sequencing in this study. A consensus sequence of the different cells used in sanger sequencing was obtained with a 60% threshold, and sanger sequences were finally removed for phylogenetic analysis. Final alignments were trimmed automatically using trimal (Capella-Gutiérrez et al., 2009) with a 30% gap threshold. A first phylogenetic analysis was done by a RAxML and GTR+G evolutionary model with 100 rapid bootstraps, in order to filter long branches and poorly resolved sequences. The final data set contains 90 taxa and 2670 positions for Nassellaria (+ 18 sequences of Spumellaria as outgroup) and 105 taxa and 2508 positions for Spumellaria (+ 14 sequences of Nassellaria as outgroup). Same method as previously described was used to infer phylogenetic position of the sequences. Final trees were visualized and edited with FigTree version 1.4.3 (Rambaut 2016).

Entropy analysis

Sanger and MinION sequences were aligned independently using MAFFT v7.395 (Katoh and Standley, 2013) with a L-INS-i algorithm and 1000 iterative refinement cycles for high accuracy against a reference alignment (extracted from Sandín et al., in prep, Chapter 2). All Sanger sequences were considered for this analysis, and only those sequences used for phylogenetic analysis were used regarding MinION. In total 4 different alignments (Reference + Sanger and reference + MinION for both Nassellaria and Spumellaria) were manually checked in SeaView version 4.6.1 (Gouy et al., 2010). For every position of each alignment, Shannon entropy was calculated without considering insertions or deletions ("-") due to the incompleteness of Nassellaria reference sequence and the large number of gaps produced in MinION alignment. The entropy of the full V4 region was also measured to compare the three different sequencing technologies. The full V4 region was extracted as previously described (cutadapt) from the reference alignment, Sanger and MinION results (considering only these sequences used for the phylogenetic analysis regarding MinION results). Regarding Illumina sequencing results, were considered only these sequences taxonomically assigned under Nassellaria or Spumellaria. Final datasets included sequences from 4 different origins: reference sequences and sequences obtained in this study by Sanger, MinION and Illumina sequencing. Two dataset (for Nassellaria and Spumellaria) were aligned independently using MAFFT v7.395 (Katoh and Standley, 2013) with a L-INS-i algorithm and 1000 iterative refinement cycles for high accuracy and the Shannon entropy was calculated for every position.

Results

Quality of sequencing results

In total 832 reads were successfully retrieved from Sanger sequencing for the 4 holobionts (208 ±3.16 reads per holobiont; 3 reads (primers) per sequence, 24 sequences per replicate and 3 replicates per holobiont; see Supplementary material Fig S1 for a schematic representation). These amplicons had an average length of 996.86 (± 108.59) bp, with a

median of 1028 bp and an average BLAST similarity identity of 97.62 (± 3.09) % with a reference sequence (Fig. 1, Sanger). MinION sequencing resulted in 864 total sequences. These sequences were compared against NCBI database by BLAST tool and 81 had no possible matching, 185 were assigned to bacteria, 593 to eukaryotes and 5 matching different domains (e.g. the same sequence was matching virus, uncultured prokaryote and uncultured organism). The remaining 783 sequences had an average sequence length of 2043.48 bp (± 1143.42) bp, with a median of 2746 bp and an average identity of 86.03 (± 2.91) % with a reference sequence. Demultiplexing MinION sequences resulted in 55 sequences among the different replicates. When demultiplexing sequences the average sequence length increase to 2641.82 bp (± 681.8) bp and the median length to 2921 bp although the average identity remain stable at 86.48 (± 2.33) % similarity on average with a reference sequence (Fig. 1, MinION). Regarding Illumina sequencing, on average 12,135 (± 6,470) reads were obtained for each cell. These reads were merged resulting on 13,500 (± 5,718) amplicons on average per cell and 80,627 unique amplicons. They had an average length of 383.04 (± 5.62) bp, a median of 383 bp and an average identity of 96.37 (± 4.8) % with a reference sequence (Fig. 1, Illumina).

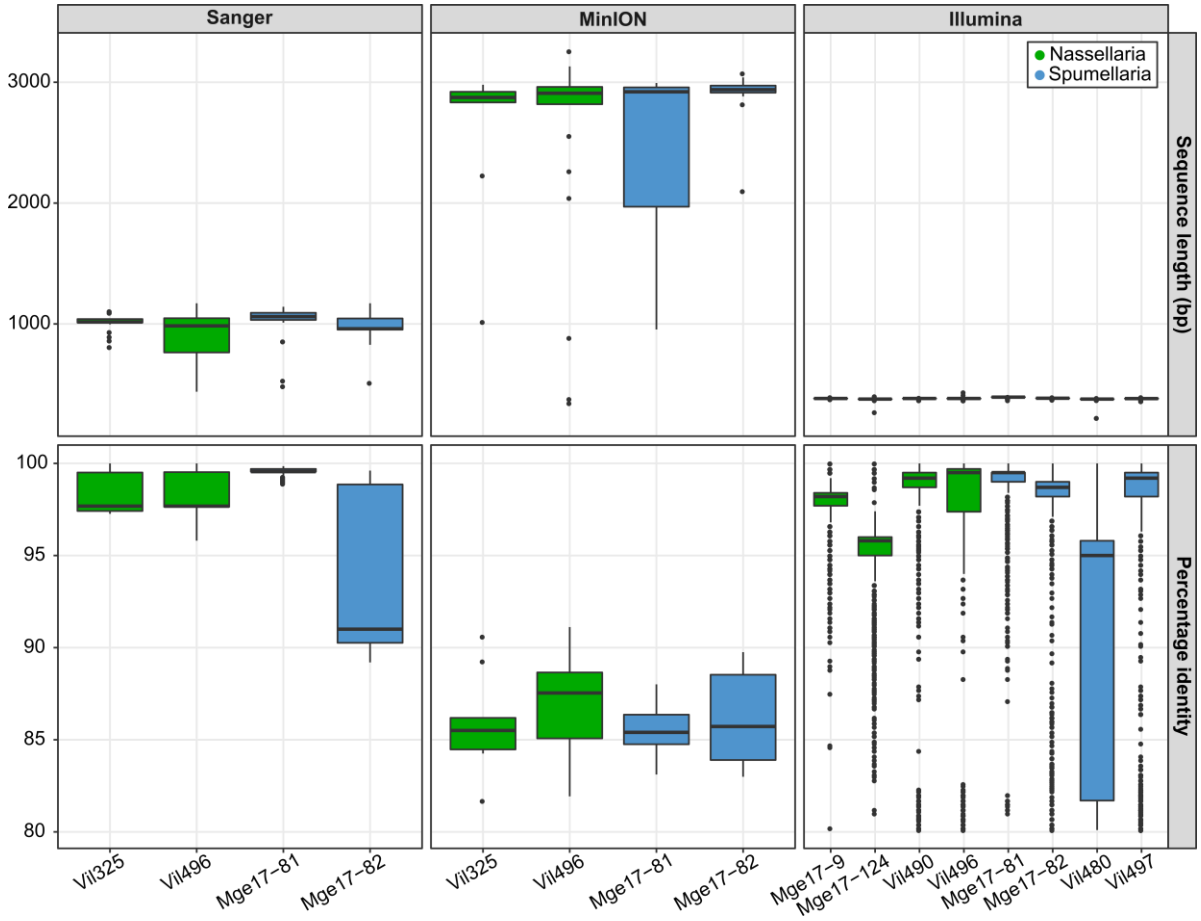


Figure 1. Boxplot summarizing sequencing results of Sanger, MinION and Illumina sequencing of the sequence length (top row) and the percentage identity of the first match against a reference sequence (bottom row) per cell. Note that identity for Sanger and MinION was performed by BLAST and for Illumina by global alignment (usearch: due to the short length of the reads).

Taxonomic assignation and diversity of reads

Sanger sequenced amplicons of the different parts of the ribosomal genes belonging to the same replicate were concatenated resulting in 288 sequences with an average length of 2879.82 (± 498.88) bp. This covers about the first ~1000 bp of the 18S rDNA gene (primer s30), the last ~600-800 bp of the 18S and the ITS1 rDNA gene (primer s69f) and the regions D1 and D2 from the 28S rDNA gene (primer d1r). In Vil325 sequences were assigned to Nassellaria (56), Ciliates (11) and Tunicates (4); in Vil496 to Nassellaria (53), golden algae (15) and diatoms (5); in Mge17-81 to Spumellaria only (72) and in Mge17-82 to Spumellaria (63), ciliates (8) and diatoms (1) (Fig. 2, Sanger).

From the demultiplexed sequences extracted from MinION sequencing, in Vil325 was found only Nassellaria (9); in Vil496 was found Nassellaria (16), diatoms (6), golden algae (3) and fungi (1); in Mge17-81, only Spumellaria (8), as in Sanger sequencing, and in Mge17-82, Spumellaria (7), dinoflagellates (3), alveolates (1) and fungi (1) (Fig. 2, MinION). The low amount of demultiplexed sequences in MinION led us to search for more Polycystine sequences from the raw sequences. Two phylogenetic analysis were done independently for Nassellaria and Spumellaria. The first phylogenetic analysis considered all sequences taxonomically assigned with BLAST to Radiolaria. In the first analysis, 19 sequences within Nassellaria and 15 within Spumellaria were removed due to long branches and clustering between the outgroup and the ingroup. Final phylogenetic analysis included 39 sequences for Nassellaria (Supplementary material Fig S3) and 26 for Spumellaria (Supplementary material Fig S4) that were not previously considered. For both Spumellaria and Nassellaria, these sequences clustered in their respective clades, yet in long branches. Bootstrap values in general tend to be low within the clades bearing MinION sequences, being most of the times below 60, yet sporadically it can get higher. The consensus sequences obtained from Sanger sequencing branched basal to the MinION sequences and phylogenetically close to the sequences used as reference.

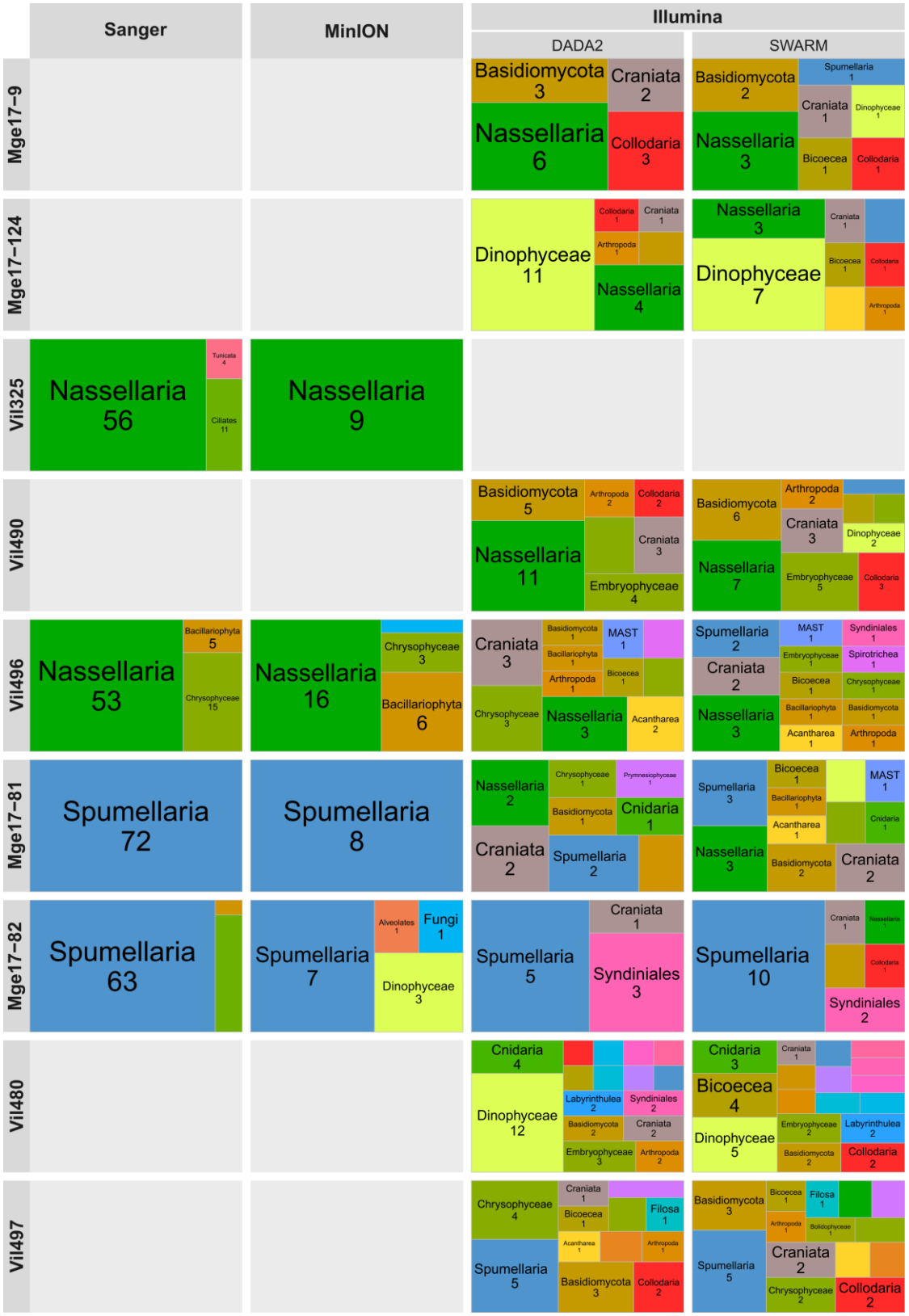


Figure 2. Taxonomic affiliation of sequences/amplicons obtained by Sanger (after concatenation of the primers), MinION and Illumina (reads processed by dada2 and swarm) sequencing for each cell, the area represents the proportion of total number of unique sequences/amplicons affiliated to the specific taxonomic entity (tree map). Numbers below the taxonomic group represent the number of unique sequences/amplicons.

Amplicons of Illumina were clustered using an error-based method (dada2) and a distance-based method (swarm). For the former method a total of 127 ASVs (with a total of 325,656 reads) were found and for the second 93 ASVs (with 495,717 reads). These ASVs had an average length of 372.8 (± 26.27) bp and a median length of 380 bp for dada2 and 376.82 (± 15.66) bp and a median length of 380 bp for swarm. Both methods gave similar results after taxonomic assignation, with an average similarity identity of 99.01 (± 2.34) % for dada2 and 99.04 (± 2.23) % for swarm against reference sequences. Yet, there are slight differences in the number of ASVs within taxonomic groups. The most diverse groups were Nassellaria (23 ASVs for dada2 and 10 for swarm), Dinophyceae (23 ASVs in dada2 and 11 in swarm), Spumellaria (13 ASVs and 16 in swarm) and Basidiomycota (12 ASVs in dada2 and 10 in swarm; Fig. 2, MinION). In average there were 4.66 (± 3.07) ASVs associated to the host in dada2 and 4.38 (± 2.89) in swarm. For dada2 in Mge17-9 there were mainly found Nassellaria (6), fungi (Basidiomycota: 3), Collodaria (3) and Craniata (2); in Mge17-124, dinoflagellates (11) and Nassellaria (4); in Vil490, there were Nassellaria (11), fungi (Basidiomycota: 5), Embryophycea (4), Golden algae (3), Craniata (3), Arthropoda (2) and Collodaria (2); in Vil496, Nassellaria (3), Golden algae, (3), Craniata (3) and Acantharea (2); in Mge17-81, Spumellaria (2), Nassellaria (2) and Craniata (2); in Mge17-82, Spumellaria (5) and Syndiniales (3); in Vil480, Dinophyceae (12), Cnidaria (4), Embryophyceae (3), Arthropoda (2), Basidiomycota (2), Craniata (2), Labyrinthulea (2) and Syndiniales (2) and in Vil497, Spumellaria (5), Chrysophyceae (4), Basidiomycota (3) and Collodaria (2). Regarding swarm in Mge17-124 there were mainly Dinophyceae (7) and Nassellaria (3); in Mge17-9, Nassellaria (3) and fungi (Basidiomycota: 2); in Vil490, Nassellaria (7), fungi (Basidiomycota 6), Embryophyceae (5), Collodaria (3), Craniata (3), Arthropoda (2) and dinoflagellates (2); in Vil496, Nassellaria (3), Craniata (2) and Spumellaria (2); in Mge17-81, Nassellaria (3), Spumellaria (3), fungi (Basidiomycota: 2) and Craniata (2); in Mge17-82, Spumellaria (10) and Syndiniales (2); in Vil480, dinoflagellates (5), Bicoecea (4), Cnidaria (3), fungi (Basidiomycota: 2), Collodaria (2), Embryophyceae (2) and Labyrinthulea (2) and in Vil497, Spumellaria (5), fungi (Basidiomycota: 3), Golden algae (2), Collodaria (2) and Craniata (2). Although the most abundant ASVs were Spumellaria, Dinophyceae, Nassellaria, Biocoecea, Chrysophyceae and Syndiniales (Supplementary material Fig S5). For two holobionts (Mge17-124 and Vil480) the relative abundance of Nassellaria and Spumellaria (respectively) is very low compared to that of other taxonomic groups. It is important to note the presence of Spumellaria sequences among Nassellaria holobionts and the other way around, besides the presence of other Radiolaria sequences such as Collodaria (e.g. in Mge17-9, Mge17-124, Vil490, Mge17-82, Vil480 and Vil497) or Acantharia (e.g. in Mge17-124, Vil496, Mge17-81 and Vil497).

Due to the great taxonomic diversity found within holobionts by Illumina sequencing, only ASVs present in at least 3 samples (triplicates in the PCR) and with a total abundance equal or higher than the median value of the abundance for both methods independently (that is 69 reads for dada2 or 42 for swarm) were considered. We have chosen stringent thresholds in an attempt to remove artefacts and/or contaminations. After processing, the number of ASVs changed considerably (31 ASVs in dada2 and 35 in swarm) but the total reads did not change drastically with 311,880 for dada2 and 483,628 for swarm, representing up to 95.8% and 97.6% of the total reads respectively. Main taxonomic affiliations were similar to

previously described (before filtering the ASVs). Although, the average number of ASVs associated to the host decreased to 1.63 (± 0.92) in dada2 and 2.38 (±1.30) in swarm. Furthermore, the relative proportion of unexpected ASVs took more importance; such as an ASV affiliated to Collodaria present in 4 holobionts (Mge17-24, Mge17-9, Vil480 and Vil497) in both clustering methods, an ASV affiliated to Acantharea present in 4 holobionts (Mge17-24, Vil496, Mge17-81 and Vil497) only clustered by swarm and 1-2 ASVs affiliated to Craniata in every sample (Supplementary material Fig S6). When exploring in deeper detail the ASVs affiliated to Polycystines and their abundance and distribution among samples after applying the stringent thresholds, it is possible to find up to 3 different and highly abundant ASVs within the same cell (e.g. Mge17-82 for both methods or Vil497 in dada2; Fig. 3 and Supplementary material Fig S7 showing the abundance for each ASV). In Mge17-9 there were two ASV, of which one ("c") was present in the three replicates of the holobiont and in a fourth replicate from Mge17-81, the second ASV ("m") was only present in one replicate for dada2 or 2 in swarm. Mge17-124 did not have any ASV present in the three replicates, only ASV "m" present in one replicate for both clustering methods. Vil490 had two ASVs in dada2 ("b" and "g") but only one in swarm ("b"). In Vil496 there was only one ASV ("e") present in the three replicates of the holobiont, and in a fourth replicate from Mge17-81, yet in a very low relative abundance. Mge17-81 had also one ASV present in the three replicates ("a") and two other ASVs ("c" and "e") present each of them in one different replicate. Vil480 had one ASV ("k") present in the three replicates and a second ASV ("m") present in one replicate in dada2 and in the three replicates in swarm but with low relative abundances. The ASVs "c" and "e" are taxonomically assigned to Nassellaria and appear in a fourth sample from a different organism of Spumellaria, although in very low relative abundance. Finally, the ASV named "m", that appears in several samples (in dada2 never within the same three samples (or replicates) of a holobiont), its taxonomic affiliation is to Collodaria, with a 100% identity score.

Intracellular gene variability

In order to compare the intracellular gene variability between different sequencing method and discriminate between sequencing errors, we calculated the entropy of the alignment between sequences obtained in this study and reference sequences (from Sandin et al. *in prep*., Chapter 1.2) at every position. Since it is expected that errors are random, a single error in the alignment would have low evenness and appear with a low entropy (yet not 0). On the other side, intra-genomic variability is expected to be sequenced in several replicates increasing the evenness and with it the entropy, and also with a tendency to appear towards hyper-variable regions. In general, concatenated sequences from Sanger sequencing belonging to Nassellaria showed low Shannon entropy values among them (Supplementary material Fig S8). Towards the end of the primers the entropy slightly increased, meaning that most probably are suspected to be errors. Specially at the end of the region D2 of the 28S rDNA gene the entropy reaches its highest values along a region of ~100 bp length, probably meaning variability among the different copies of the rDNA. Regarding Spumellaria, Mge17-81 has a similar trend than Nassellaria (Supplementary material Fig S8). In contrast, Mge17-82 shows regions of high entropic values at around the position 750 (V4 region from the 18S rDNA gene), along the beginning of the 28S RDNA and especially over a region of ~250 bp on the ITS1, showing a most probable big intra-genomic variability of the rDNA.

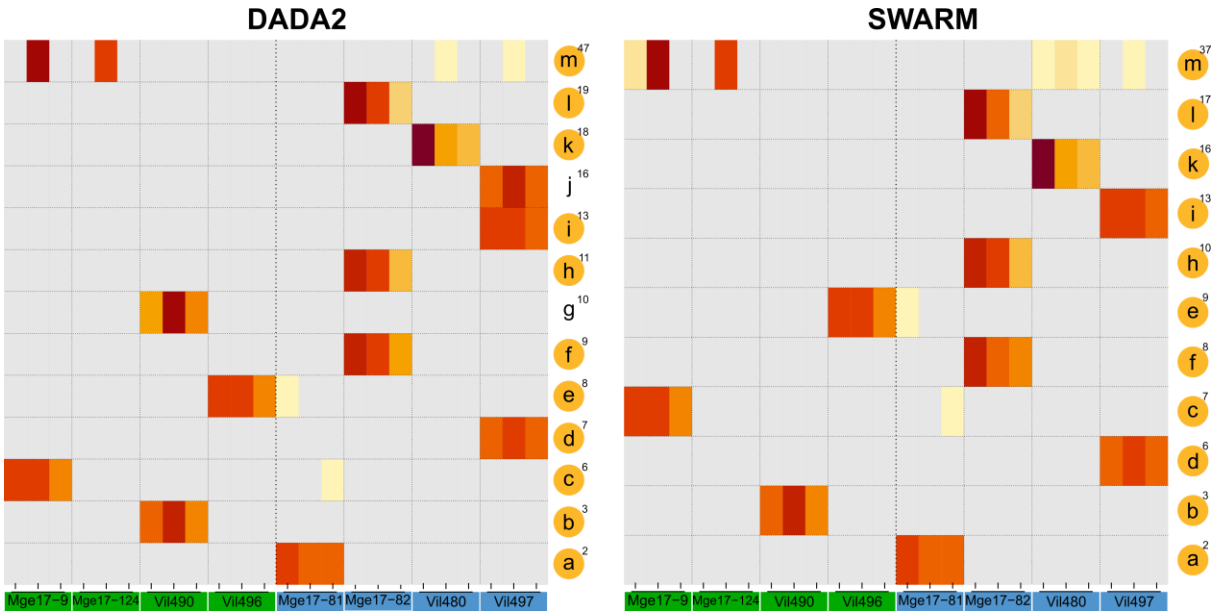


Figure 3. Heatmap of amplicons obtained from dada2 (left) and swarm (right). Color represent the square root of the abundance normalized by samples (Hellinger transformation). Only amplicons taxonomically assigned to Polycystines, present in 3 or more samples and with a total abundance equal or higher than 69 reads for dada2 or 42 for swarm (median) were considered. Letters on the right represent an arbitrary label for unique amplicons with no meaning outside this figure. Orange circles in the ASV names represent presence in both dada2 and swarm. Numbers besides the orange circles represent the abundance rank of the ASV in its clustering method.

In general MinION sequences aligned against the reference sequence with many gaps (Supplementary material Fig S8). Nassellaria 18S rDNA gene alignment moved from 1863 bp length to 2593 bp when aligned with the 39 sequences from MinION, and from 698 bp to 968 bp for the 28S rDNA gene. Aligning the 27 Spumellaria sequences from MinION resulted in a smaller change for the alignment compared to Nassellaria, moving from 1931 bp of the reference alignment for 18S rDNA gene and 859 bp for the 28S rDNA gene to 2513 and 958 respectively. These sequences had high entropy values all along the rDNA alignment. With the exception of the 18S rDNA gene of Nassellaria, where there is a constant entropy trend, there are variations of the entropy depending on the region, but these variations do not match exactly those of the reference alignment.

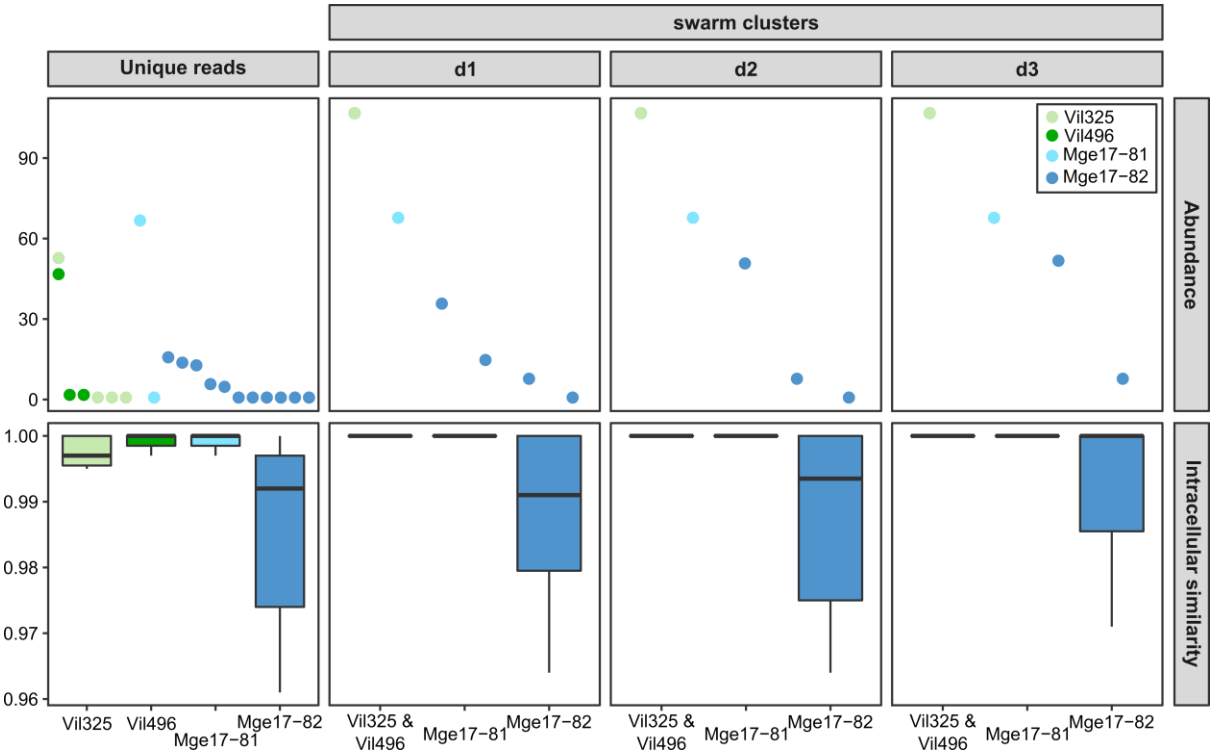


Figure 4. Abundance and intracellular similarities of the V4 region extracted from Sanger sequencing results of the total unique reads and clustered with swarm with 1 (d1), 2 (d2) and 3 (d3) differences.

The hyper-variable region V4 of the 18S rDNA gene was extracted from the sequences affiliated to Radiolaria from both Sanger and MinION sequencing, in order to compare with Illumina sequencing. In total 51, 56, 68 and 60 sequences from Sanger sequencing were extracted belonging to Vil496, Vil325, Mge17-81 and Mge17-82 respectively (Fig. 4). The same unique sequence was found to be the most abundant in both Vil325 and Vil496 (53 and 47 reads respectively) being exactly identical despite of being morphologically identified like different species, and up to 5 other unique sequences were found with a very low abundance (2, 2, 1, 1 and 1 reads, green dots on Fig. 4 under unique reads). These sequences were very similar, showing similarities within them close to 100% with a maximum of one base of difference. Similar patterns are seen in the Spumellaria Mge17-81, with one sequence found 67 times and only one sequence different with one read. In contrast, Mge17-82 shows up to 3 different reads relatively abundant (16, 14 and 13 reads) and 8 other sequences with a lower abundance, having a similarity among them of 96.1%. It is important to note, that these three sequences were sequenced in the three different replicates (PCR reactions). When these sequences were clustered with Swarm and 1 difference of threshold, Vil325 and Vil496 share the same amplicons. In Mge17-81 the single sequence was grouped in the unique amplicon found for this holobiont, and up to 4 different amplicons were found in Mge17-82 (still with a similarity among these amplicons of 96.4%). Increasing the difference threshold to 2 and 3, the number of amplicons in Mge17-82 decreases to 3 and 2, with low changes in the similarities among them (96.4% and 97.1% respectively). Yet there is one amplicon a lot more abundant than the other (52 against 8 reads). Same protocol was implemented in MinION sequences in order to extract the V4 hyper-variable region of the 18S rDNA (data not shown). In total 17 sequences were extracted from Mge17-81, 20 from Mge17-82, 34 from Vil325 and

25 from Vil496. It was not possible to cluster the sequences with swarm due to their large dissimilarities. Up to a difference threshold of 10, there were still no clusters, keeping an average intracellular similarity of 86.92% (± 3.56%). Besides, the wrong assignation of samples was very severe, finding Spumellaria sequences in Nassellaria samples and the other way around.

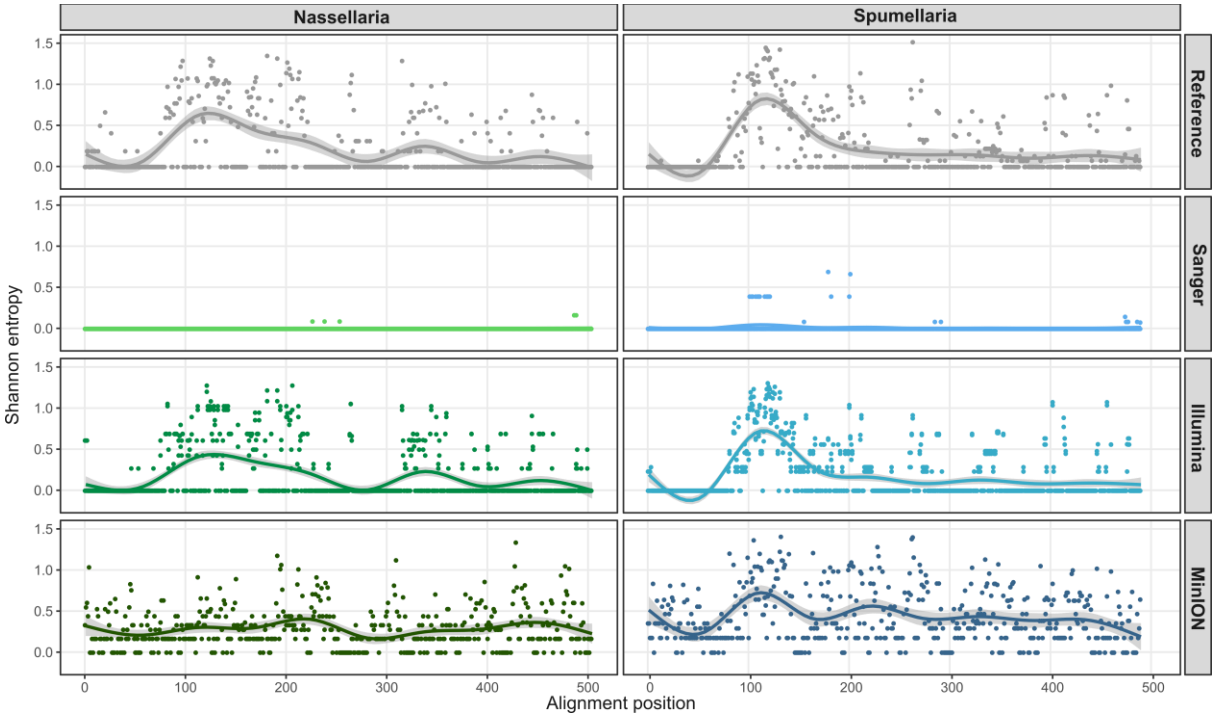


Figure 5. Shannon entropy analysis for every position (on *x* axis) of the V4 hypervariable region of the 18S rDNA gene for the reference sequences and for Sanger, MinION and Illumina sequencing results aligned all together. For further results on the near-full rDNA alignment entropy obtained by Sanger and MinION sequencing see supplementary figure S8.

The V4 hyper-variable region of the 18S rDNA gene used in Fig. 4 along with that extracted from the sequences obtained used for phylogenetic analysis in MinION sequencing were pooled together in one alignment for the V4 region comparison. In total, 107 sequences for Nassellaria and 128 sequences for Spumellaria were extracted from Sanger sequencing and 25 nassellarian sequences and 23 spumellarian sequences were extracted from MinION sequencing. The V4 region dataset was completed with those amplicons extracted from Illumina taxonomically assigned to Nassellaria (13 in dada2 and 10 in swarm) and Spumellaria (12 in dada2 and 16 in swarm) and re-aligned with all sequences. Final aligned dataset had a length of 504 bp for Nassellaria and 490 bp for Spumellaria (Fig. 5). Reference sequences showed a hotspot of higher Shannon diversity within the hyper-variable region in both Nassellaria and Spumellaria spanning from the position ~80 until ~220 with maximum values around the position ~110. Nassellaria reference sequences showed two other hotspots regions peaking at around the positions ~330 and ~450, yet average entropy values are half of the first hotspot region. Regarding Spumellaria, these two last hotspot regions are less marked than for Nassellaria due to overall higher entropy values. Sanger sequences maintain a near-0 entropy values, with the exception of a region between ~100-200 bp in Spumellaria showing higher entropy values (corresponding to Mge17-82). Illumina ASVs follow same

patterns as the reference alignment, with the highest entropy values at around the position ~110. Despite the poor trends found for the full length of the rDNA for MinION sequences (Supplementary material Fig S8), when focusing on the V4 region, shows similar patterns that those found in the reference alignment and illumina. Although these peaks are smoother due to the relative higher entropy values all along the alignment.

Discussion

Our results showed that intracellular gene variability of Nassellaria and Spumellaria is generally limited, showing in some cases the same V4 rDNA hyper-variable region between different morpho-species (yet belonging to the same genus; Vil325, *Eucyrtidium cienkowskii* and Vil496, *E. acuminatum*), as previously found in tintinnids (Bachy et al., 2013). Although in some groups it can be very important, finding two highly distinct V4 rDNA hyper-variable regions within the same cell (Mge17-82 ~97% similarity, representing 14 different nucleotide positions). Such taxonomic differences among closely related groups have also been found in Oligotrich and Peritrich Ciliates (Gong et al., 2013). Most of this intracellular gene variability is however overlooked due to the presence of a highly repeated copy that predominates over the low abundant copies, as previously found in Nassellaria or among other orders of Radiolaria such as Acantharia (Decelle et al., 2014). However, in Acantharia the intracellular variability could also become important, finding up to 3 different OTUs (V9: clustered at 97% and present in 2 replicates; Decelle et al., 2014). Similar studies have shown a relationship between the intragenomic variability and the number of macronuclei (Zhao et al., 2019) and between the rDNA copy number in ciliates (Gong and Marchetti, 2019) and in alveolates (Medinger et al., 2010). That could explain the taxonomic differences in the intragenomic variability of Radiolaira, since both Nassellaria and Spumellaria have only one nucleus that tend to be small whereas Acantharia and Collodaria have several nuclei (Suzuki et al., 2009). In the former case, they tend to show low intracellular variability, whereas in the second case can be relatively high (Decelle et al., 2014).

Despite the low intracellular gene variability found within the hosts, Illumina sequencing has identified *a priori* a very diverse host gene variability; finding in some hosts up to 10-11 different ASVs (Fig. 2). And when comparing the holobiont community obtained by the three different sequencing methods, Illumina shows an incredibly diverse holobiont community. The unexpected presence of taxonomic groups such as Acantharia, Collodaria or Craniata within holobionts constituted by Nassellaria and Spumellaria as hosts, questions the reliability of the so-called “rare” ASVs in environmental studies. Part of this rare biosphere has been proposed as artefacts inflating diversity estimates (Kunin et al., 2010; Bachy et al., 2013). Other ASVs present in a fourth sample (i.e.: “c” and “e” in Fig. 3) also question technical issues such as cross-contamination during sequencing (Kircher et al., 2012) or tag-jumping during library preparation (Schnell et al., 2015). Other ASVs have had full similarities against a reference sequence and have passed the stringent abundance filters (e.g. Craniata in fig. 2 or ASV “m” in fig. 3) suggesting the presence of environmental DNA (eDNA). Number of reads have been found to be correlated with the number of nuclei and the cell size (Biard et al.,

2017; Pitsch et al., 2019). Our results indicate that eDNA contamination is showing a bias towards organisms with a higher copy number, such as Collodaria (Biard et al., 2017) or the metazoans Craniata. A problem that might be accentuated when the targeted DNA is relatively low abundant (*i.e.* that of Nassellaria) and have to "compete" for the available space in the flow-cell during sequencing, resulting in large differences of relative abundances, as seen in Mge17-124 and Vil480 (Fig. 2 and supplementary figure S5). DNA amplification success in Foraminifera has been shown to be taxa-specific (Weiner et al., 2016). Probably, DNA from Acantharia and Collodaria are more prone to be successfully amplified regarding that of Nassellaria and Spumellaria. In this study we have used general eukaryotic primers, that are equally binding these radiolarian groups. Therefore, further analysis must discriminate such differences in shell or cell architecture regarding enzymes for DNA extraction and amplification.

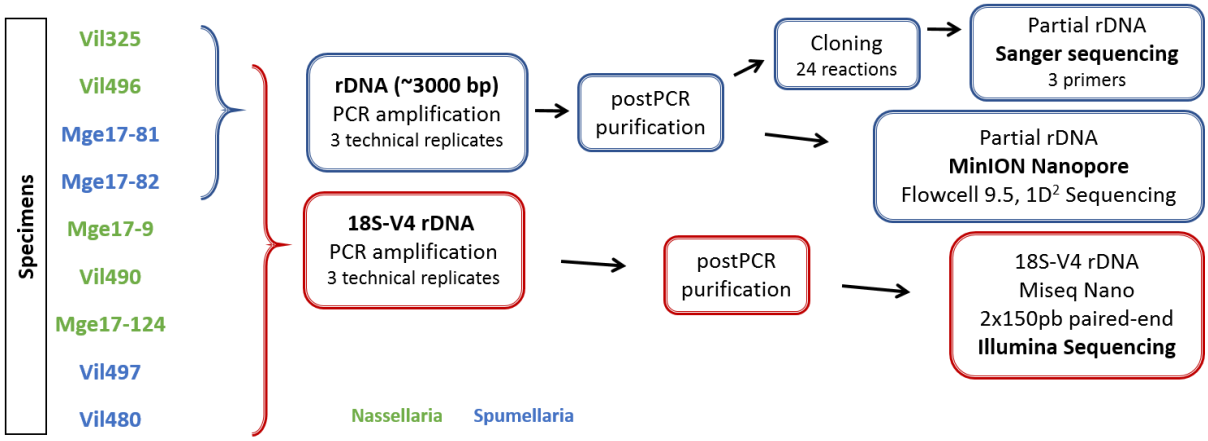
These results highlight the need for a careful interpretation of metabarcoding surveys due to the taxonomic bias found among different taxa (e.g. Weiner et al., 2016; Gong and Marchetti, 2019). In some cases, illumina sequencing tends to lead to an underrepresentation of the environmental diversity (e.g.: Mge17-124 and Vil480 in Fig. 2) and for some other taxa an over estimation from eDNA (e.g.: Collodaria and Craniata). Most of these biases and technical problems cannot be identified and sorted out in environmental metabarcoding surveys leading to poor estimations of the environmental molecular diversity. For example, the presence of the ASVs "c" or "e" in a fourth sample would have been considered as low abundant ASV in that specific sample due to the lack of replicates in most of the global environmental metabarcoding surveys. This emphasize the need for considering replication in metabarcoding surveys to ensure an accurate estimation of the environmental microbial diversity (Prosser, 2010). Besides, when working with specific taxa, may help the estimation of the rDNA copy number (Biard et al., 2017; Gong and Marchetti, 2019), the exploration of differences in relative abundance (Morton et al., 2019) or even exploring previous steps such as DNA extraction bias, amplification primers or intracellular architecture. However, a big part of the overestimated diversity might also be due to sequencing errors, as previously proposed (Bachy et al., 2013; Decelle et al., 2014), since the 50% less abundant ASVs accounted for less than 5% of the total reads meaning a high presence of singleton and low abundant clusters.

Illumina sequencing showed to be the method with higher amount of likely contaminations from eDNA, yet after clustering into ASVs, the error rate showed to be the lower from the three methods and randomly distributed. Besides, most of the suspected artefacts produced during Illumina sequencing were removed with the stringent filters. Sanger sequencing also showed a limited error rate, yet singletons were found in the V4 rDNA region of every sample sequenced, that were later corrected when clustering (Fig. 4). However, towards the end of the primers the error rate increased substantially. For this reason, we suggest that filtering of Illumina ASVs should be done more importantly based in abundance than identity thresholds, especially when comparing against the V9 region that is located near the end of general and specific primers. On the other side, MinION is the

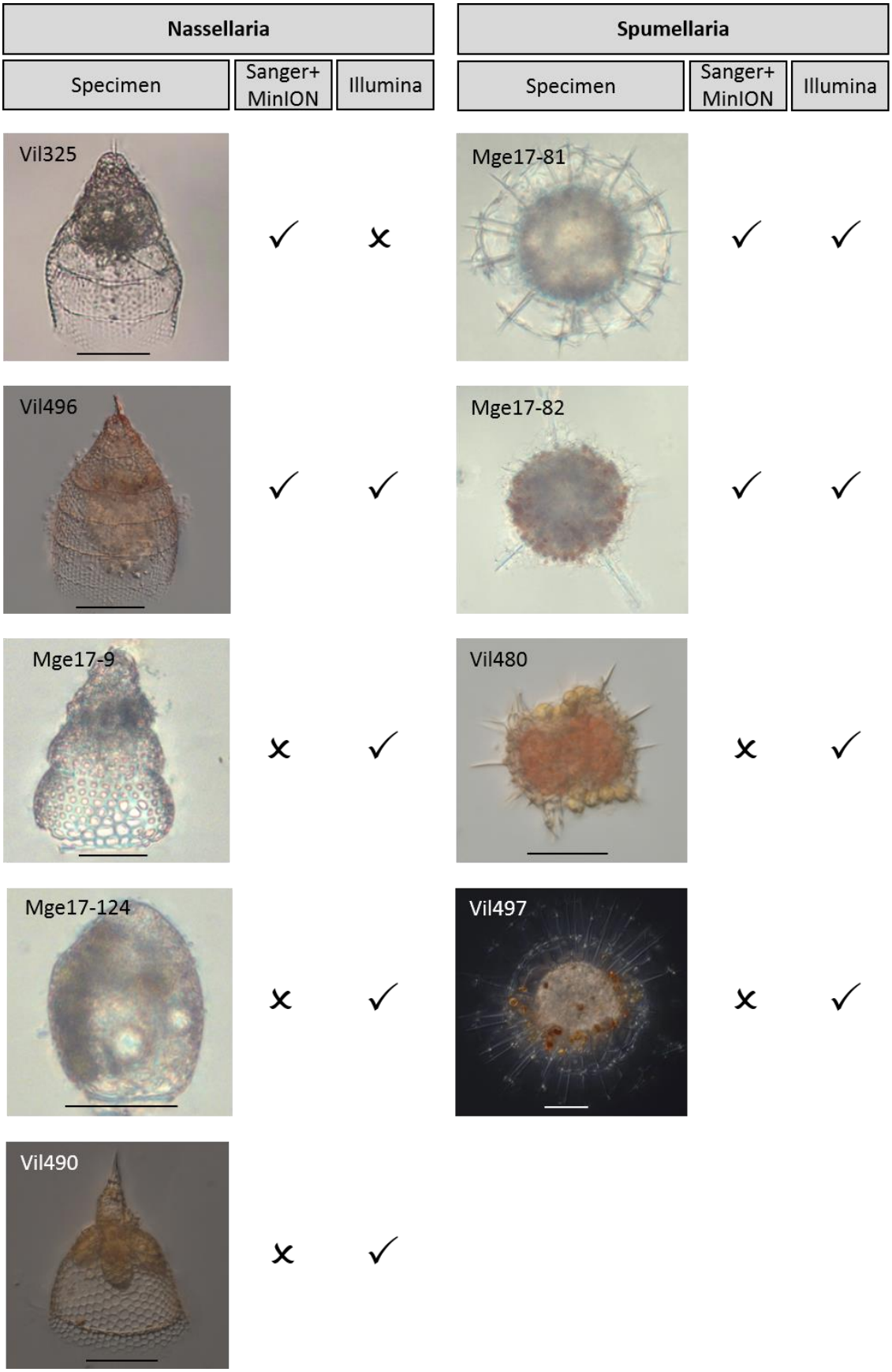
technique with the highest error rate, finding alignment similarities of 94-97% in bacteria (Tyler et al., 2018) better than our results with an average of 86% for the best scoring sequences. Although, recent studies have generated consensus sequences decreasing considerably the error rate and reaching results comparable or better than those obtained by Sanger sequencing (Pomerantz et al., 2018; Wurzbacher et al., 2019). Despite the high error rate of MinION found in our analysis without correction, and a poor alignment against reference sequences, phylogenetic analysis showed an expected clustering of the targeted sequences thanks to their long reads. Besides, when looking in detail in the V4 region, MinION sequencing had a similar pattern to that obtained by Illumina. In addition, the contamination rate was lower, showing comparable holobiont communities to those obtained by Sanger (Fig. 2).

ONT have the advantage of directly sequence the DNA strand with no need of PCR amplification, and therefore removing those biases associated in these steps that most likely produced the artificial environmental diversity in Illumina. This also brings the possibility to work with absolute reads, and not relative values providing a more quantitative picture of the community. Yet, the error rate of ONT is still very high for a fine scale resolution. However, its accuracy is improving since it was firstly released (Rang et al., 2018), and new methods for correcting and analysing ONT results are coming along (e.g.: demultiplexing: Wick et al., 2018; base-calling: Wick et al., 2019; consensus sequence building: Pomerantz et al., 2018; Wurzbacher et al., 2019). Therefore, if the error rate is overcome, MinION could generate fruitful results in the near future of metabarcoding surveys by taking advantage of the extensive work of the 18S reference sequences already done, the high variability and taxonomic resolution of the ITS and the 28S, the portability and the sequencing depth and length. Future perspectives are the implementation of the correction and analysing tools mentioned above in order to assess the reliability taxonomic level of ONT sequencing with the current error rate.

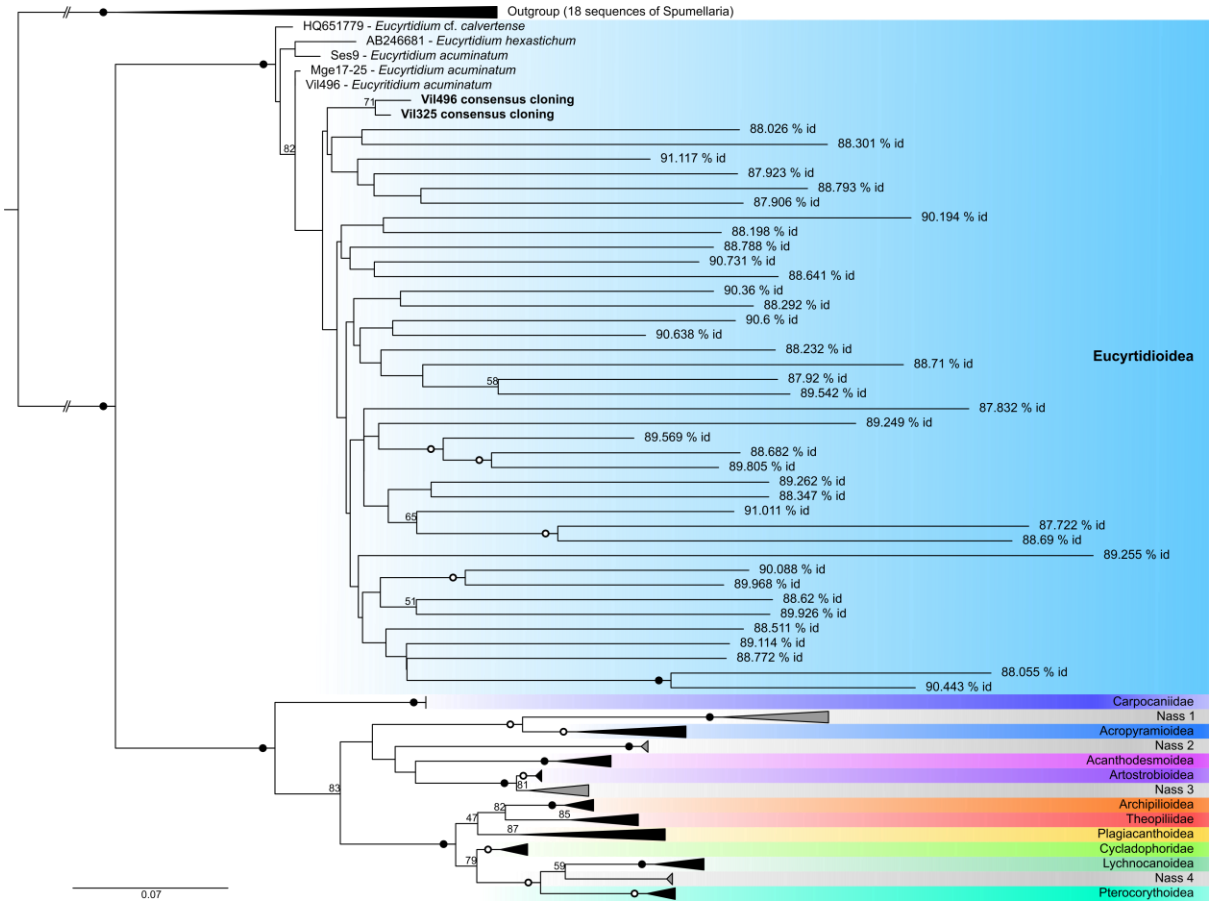
Supplementary Material



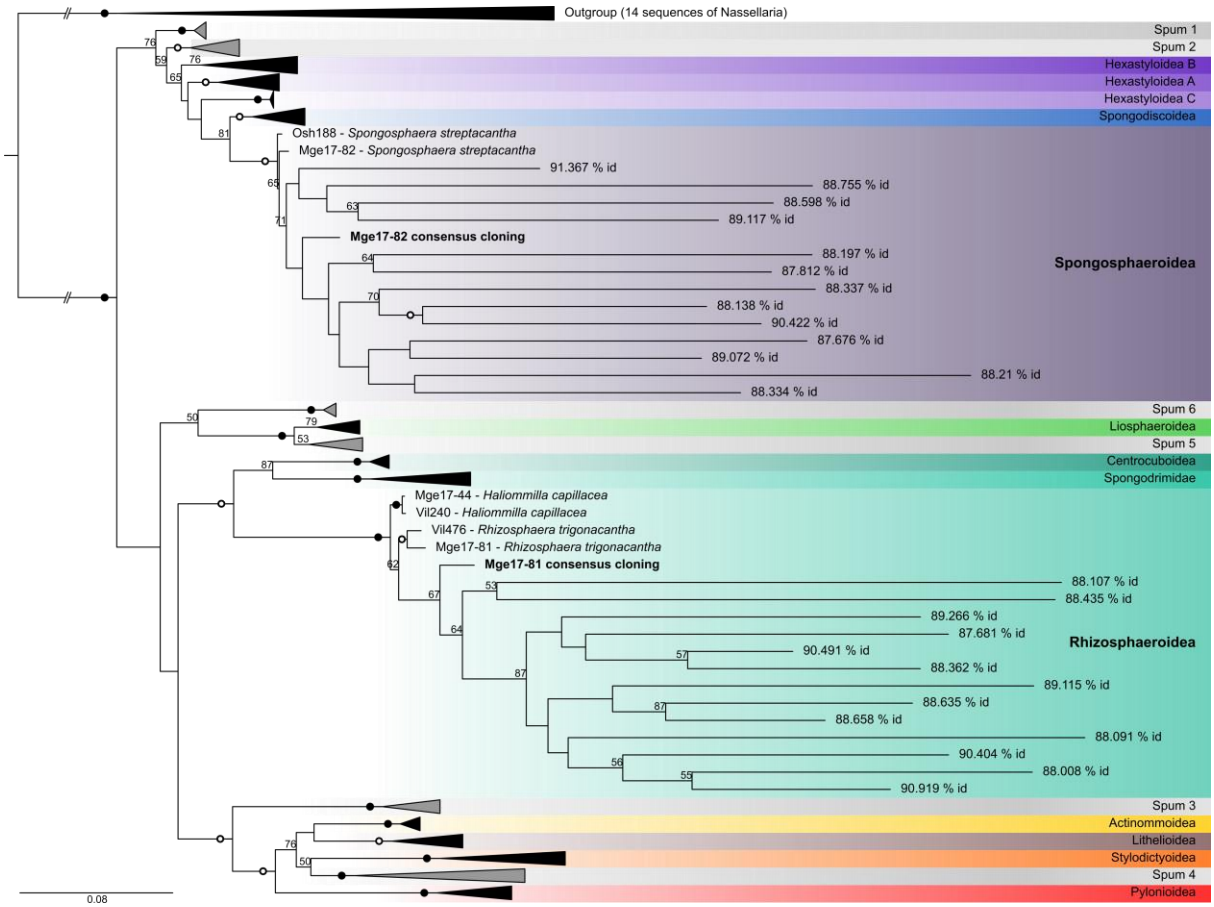
Supplementary Figure 1. Schematic representation of the study design.



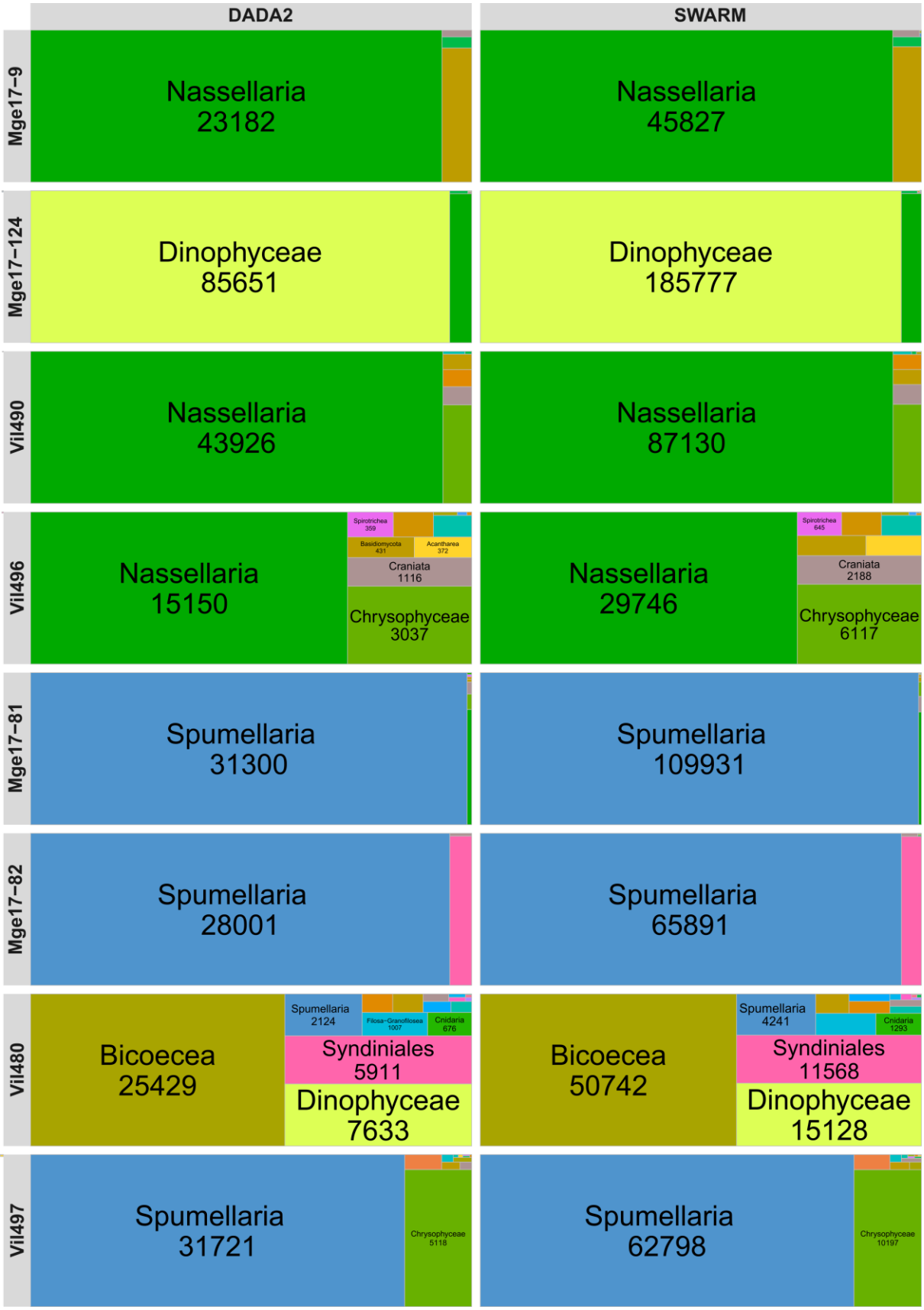
Supplementary Figure 2. Light microscopy images of specimens used in this study. On the right of each specimen it is indicated whether it was used (tick mark) or not (cross mark) for Sanger + MinION sequencing and/or Illumina. Scale bar (when available; black) represents 50 µm.



Supplementary Figure 3. Molecular phylogeny of MinION sequences, a consensus of Sanger results (in bold) and reference sequences (extracted from Sandin et. al. *in prep*., Chapter 1.2) for Nassellaria inferred from the concatenated complete 18S and partial 28S (D1-D2 regions) rDNA genes and automatically trimmed at 30% threshold (trimal). Sequences obtained by MinION are shown with the best scoring identity to a reference sequence in NCBI. The tree was obtained by using a phylogenetic Maximum likelihood method implemented in RAxML using the GTR + G model of sequence evolution and 100 rapid bootstraps (BS, shown at the nodes). Bootstrap values below 50 are not shown. Black circles indicate BS of 100%. Hollow circles indicate BS > 90%. Branches with a double barred symbol are fourfold reduced for clarity.

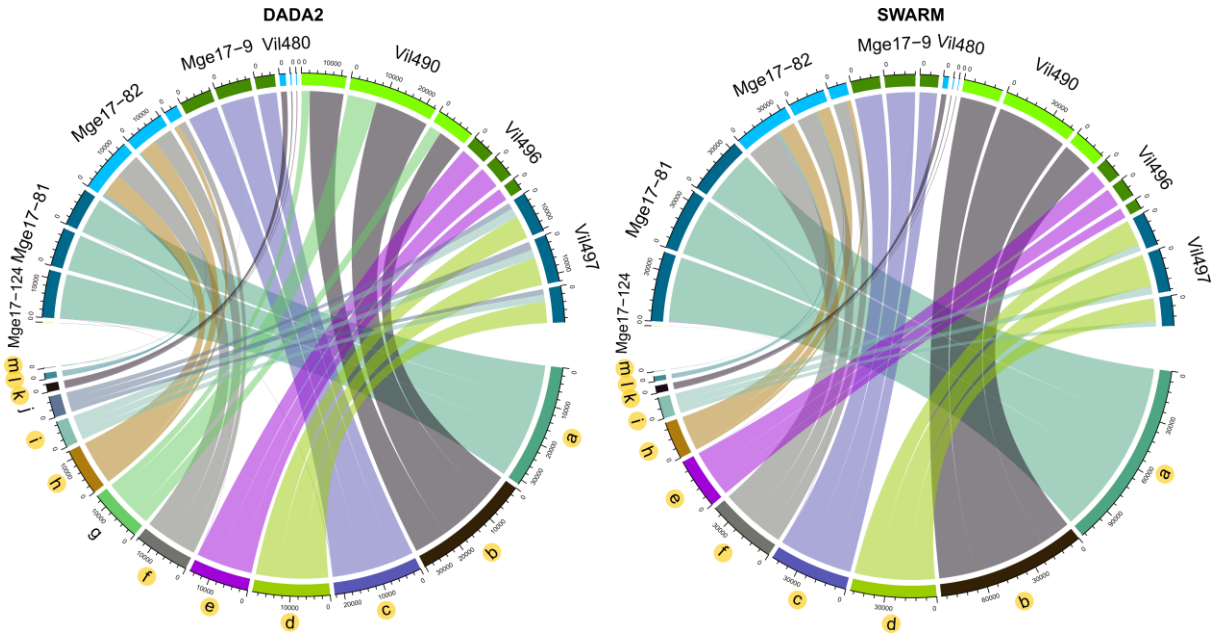


Supplementary Figure 4. Molecular phylogeny of MinION sequences, a consensus of Sanger results (in bold) and reference sequences (extracted from Sandin et. al. *in prep.*, Chapter 1.2) for Spumellaria (bottom) inferred from the concatenated complete 18S and partial 28S (D1-D2 regions) rDNA genes and automatically trimmed at 30% threshold (trimal). Sequences obtained by MinION are shown with the best scoring identity to a reference sequence in NCBI. The tree was obtained by using a phylogenetic Maximum likelihood method implemented in RAxML using the GTR + G model of sequence evolution and 100 rapid bootstraps (BS, shown at the nodes). Bootstrap values below 50 are not shown. Black circles indicate BS of 100%. Hollow circles indicate BS > 90%. Branches with a double barred symbol are fourfold reduced for clarity.

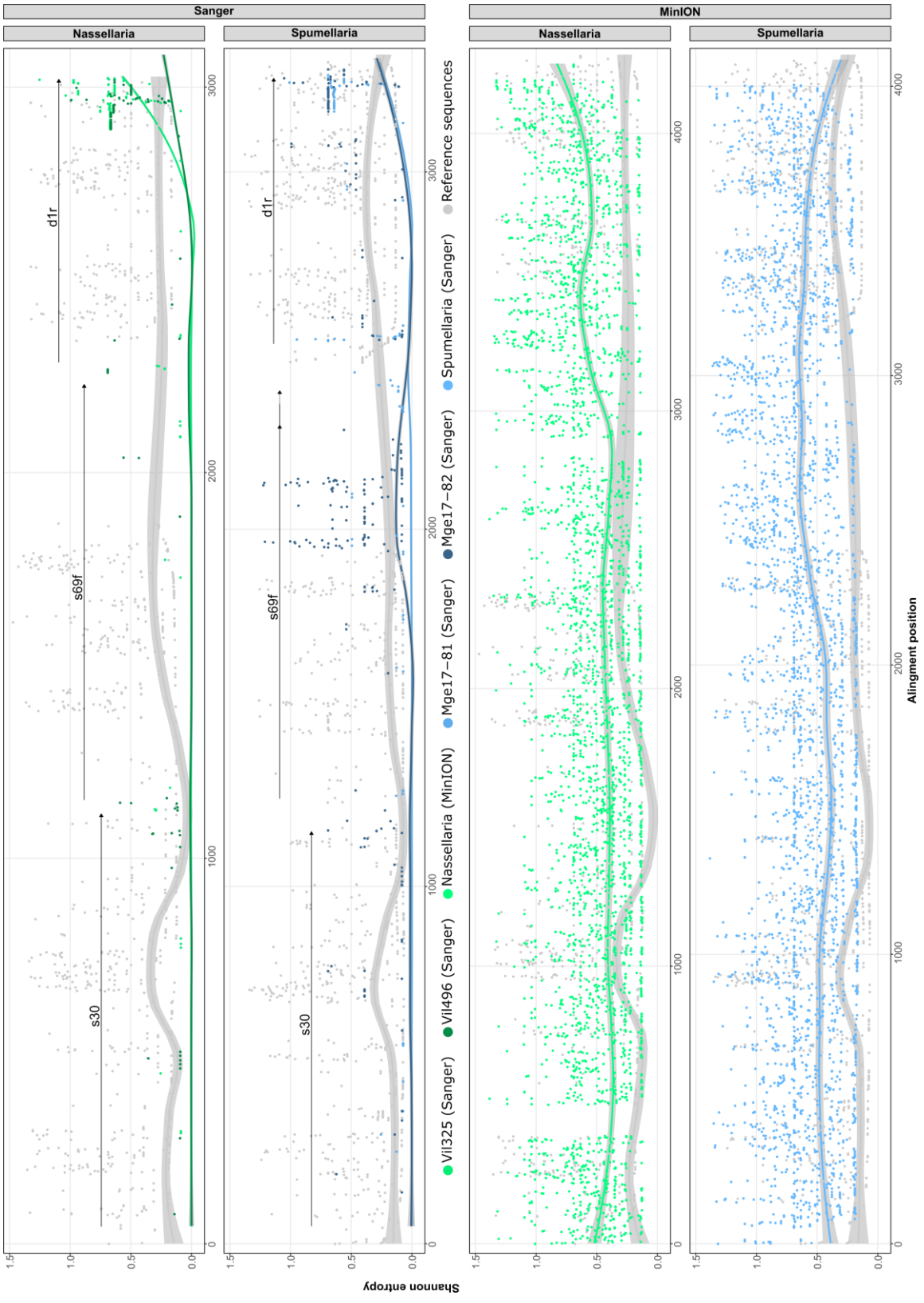


Supplementary Figure 5. Tree map of the total number of reads affiliated to a taxonomic group of raw reads obtained by Illumina sequencing for each cell. Reads were processed by dada2 and swarm.

Supplementary Figure 6. Tree map of the total number of amplicons affiliated to a taxonomic group of amplicons obtained by Illumina sequencing for each cell. Only amplicons present in 3 or more samples and with a total abundance equal or higher than 69 reads for dada2 or 42 for swarm (median) were considered.



Supplementary Figure 7. Circular plot representing the abundance of polycystines amplicons (lower half of the circles) clustered with dada2 (left) and swarm (right) and sample affiliation (upper half of the circles). Only amplicons affiliated to polycystines, present in 3 or more samples and with a total abundance equal or higher than 69 reads for dada2 or 42 for swarm (median) were considered. Letters below represent an arbitrary name for unique amplicons. Orange circles in the ASV names represent presence in both dada2 and swarm.



Supplementary Figure 8. Shannon entropy analysis for every position (on x axis) of Sanger and Minion results aligned independently against a reference alignment (in grey, extracted from Sandin et. al. *in prep.*, Chapter 1.2) for Nassellaria (green) and for Spumellaria (blue). Lines represent the tendency of the entropy for each alignment. Black arrows in Sanger boxes represent direction and approximate position of the primers used for Sanger sequencing. Insertions and deletions ("-") were not considered due to the missing regions of Nassellaria reference sequences.

Chapter 3.2

Biodiversity and biogeography of Radiolaria in the world oceans through metabarcoding

CC by Solène Rigaut

Biodiversity and biogeography of Radiolaria in the world oceans through metabarcoding

Miguel M. Sandin[1], Nicolas Henry[1], Sarah Romac[1], Aleix Obiol[2], Ramon Massana[2], Fabrice Not[1]

1-Sorbonne Université, CNRS - UMR7144 - Ecology of Marine Plankton Group - Station Biologique de Roscoff, 29680 Roscoff, France.
2-Department of Marine Biology and Oceanography, Institut de Ciències del Mar (CSIC), Passeig Marítim de la Barceloneta 37-49, 08003 Barcelona, Catalonia, Spain.

In preparation

Introduction

Radiolaria, along with Foraminifera, belong to the Retaria lineage within the Rhizaria eukaryotic supergroup (Cavalier-Smith et al., 2018; Adl et al., 2019). Radiolaria are marine heterotrophic amoeboid protist, currently divided in 3 orders based on molecular phylogenetic analysis, morphological features and chemical composition of the skeleton; Acantharia, with strontium sulphate ($SrSO_4$) skeleton and Taxopodida and the Polycystines with opaline silica (SiO_2 nH_2O) skeleton (Adl et al., 2019). The Polycystines is the most morphologically diverse group among Radiolaria (Suzuki and Not, 2015) and is traditionally divided in Spumellaria, displaying spherical and radial symmetry, Nassellaria, with an heteropolar symmetry, and the colonial Collodaria. Yet, Collodaria was recently emended within Nassellaria due to their strong phylogenetic relationships (Sandin et al. *in prep*., Chapter 2). The robust skeleton of Polycystines preserves well in sediments and hard rocks, showing an extensive and continuous fossil record dating back to the early Cambrian (De Wever et al., 2001; Afanasieva et al., 2005). Based on the specific taxonomic composition and environmental preferences, this fossil record represents a valuable tool for both bio-stratigraphic and paleoenvironmental reconstruction studies (Abelmann and Nimmergut, 2005; Kamikuri et al., 2009; O'Dogherty et al., 2011; Lariviere et al., 2012; Aitchison et al., 2017). Yet, due to their short maintenance time in culture very little is known about their living ecology and environmental preferences.

The distribution and diversity of extant Radiolaria was globally explored through sediment traps and plankton samples based on their morphological features. The Pacific Ocean gathers the highest species richness, also found in equatorial surface waters and decreasing towards poles and depth, with a second peak of species richness in the mesopelagic environment (Boltovskoy and Correa, 2016a; Boltovskoy, 2017). Differences in species assemblages related to temperature gradients, led previous authors to delimit six major biogeographic provinces. Five of them are circumglobal, and one restricted to the Eastern Equatorial Pacific. The subtropical province is the largest, gathering the biggest

species richness, then a bi-subpolar is similar in northern and southern hemisphere regarding species assemblages, but the southern and northern pole show different specific species. Finally, the last province is a transitional region in between the subtropical and bi-subpolar, showing the biggest variability in species assemblages among all provinces. Radiolaria also play an important role in the trophic food web by hosting symbiotic algae, mainly identified as haptophytes in Acantharia (Decelle et al., 2012a) and dinoflagellates in Polycystines (Probert et al., 2014; Yuasa et al., 2016). This mixotrophic behaviour allows Radiolaria to dominate the community of intertropical surface waters (Leles et al., 2017). Recently, *in situ* imaging techniques, emphasized the quantitative significance of large Rhizaria by estimating their contribution up to 5.2% of the total standing carbon stock in the euphotic layer in the ocean (Biard et al., 2016).

Morphology-based approaches are tedious and tend to underrepresent groups such as Acantharia. Because of the fast dissolution of their strontium sulphate skeleton upon cell death. Recent molecular environmental surveys have demonstrated their effectiveness in assessing diversity and biogeography beyond morphology and encompassing uncultured organisms (de Vargas et al., 2015; Pernice et al., 2016; Brisbin et al., 2019; Gutierrez-Rodriguez et al., 2019). These metabarcoding approaches are showing the great contribution of Radiolaria in the eukaryotic plankton community and are helping to better understand the important environmental diversity within Radiolaria (Not et al., 2007; Giner et al., 2019). It has been recently shown the contrasting biogeographies between Collodaria and Acantharia (Faure et al., 2019), where the colonial Collodaria tend to prefer oligotrophic areas whereas Acantharia more nutrient-rich waters based in their trophic mode. Furthermore, fine scale molecular environmental studies helped to identify specific families of Collodaria with certain preferences to coastal or oceanic biomes (Biard et al., 2017), or contribute to the understanding of the vertical life cycle of Acantharia (Decelle et al., 2013).However, no integrative fine scale metabarcoding analysis have been yet carried out to unveil Radiolaria biogeography.

Within this context, here we explore biodiversity and biogeographic patterns among the different groups of Radiolaria through a metabarcoding approach folowing the recently proposed morpho-molecular framework of Radiolaria (Sandin et al. *in prep*, Chapter 2), gathering a detailed taxonomic description for each of the main groups (Decelle et al., 2012b; Biard et al., 2015; Sandin et al., 2019, Chapter 1.1; Sandin et al. *in prep*, , Chapter 1.2). The three different datasets we selected for this study (TARA Oceans, Malaspina and MOOSE-GE) allow assessing a detailed picture of the extant diversity and biogeography of Radiolaria in the world oceans. The TARA Oceans expedition (Pesant et al., 2015) represents a global sampling of the euphotic ocean associated with environmental parameters contextualizing the biological data, in order to identify major environmental features driving their biogeography. The Malaspina expedition (Duarte, 2015) present detailed vertical profiles sampled in different oceanic regions, allowing further characterization of radiolarian communities across the water column. Finally, the MOOSE-GE expedition (https://doi.org/10.17600/17001500) combines an exhaustive environmental characterization along with a vertical profile of the western Mediterranean Sea, providing a more regional perspective on radiolarian

distribution. Altogether it brings a detailed picture of radiolarian biodiversity and biogeography in contemporary oceans, related with well-characterized water masses.

Material and Methods

Defining datasets

Three different datasets used were:

· TARA Oceans dataset was downloaded from the publicly available repository European Nucleotide Archive (accession number PRJEB6610). Detailed description on data collection can be found in Pesant et al. (2015). Briefly, small size fractions (<20µm) were collected by Niskin bottles mounted on a rosette or by a peristaltic pump. Bigger size fractions (>20µm and frequently 5-20 µm) were collected by plankton net tows. Environmental variables were downloaded from the publicly available repository PANGAEA (https://doi.org/10.1594/PANGAEA.875582). Including metadata associated to the stations such as latitude, longitude and depth and a set of environmental parameters such as temperature, inorganic carbon, alkalinity, calcite, aragonite, nitrates, phosphates, silica, salinity, density, oxygen, chlorophyll A, angular scattering coefficient, Backscattering coefficient (these two are related with the dispersion and reflection of the light, or in other words, the clearness of the water column), sunshine duration, iron, sea surface gradient temperature or distance to coast among others. This dataset represents an exhaustive sampling in the surface waters of the global ocean, with a detailed list of environmental parameters contextualizing biological data.

· Malaspina expedition samples were collected during the Malaspina-2010 expedition (REF). In total, 115 stations were sampled at different depths with Niskin bottles mounted on a rosette equipped with a CTD (detailed protocol can be found in Giner et al. n.d.). A vertical profile was performed in 13 stations sampled at 7 different depth: surface (3m deep), Deep Chlorophyll Maximum (DCM), 2-3 depths in Mesopelagic (200-1000m deep) and 2-3 depth in bathypelagic (1000-4000m deep) waters from the Indian Ocean, the Pacific Ocean and the Atlantic Ocean. This dataset represents a detailed vertical profile of main oceanic regions across the world oceans.

· MOOSE-GE dataset was collected along 15 stations in 2017 and 14 stations in 2018 in the western Mediterranean Sea. Three different depths were sampled: surface, DCM and 2000m (when possible or 10m from the sea floor otherwise). Plankton sampling was carried out in two different approaches in function of the size fraction. For the small size fractions about 24 litters of seawater from CTD Niskin bottles were prefiltered through a 200 µm nylon mesh and then sequentially filtered through a 3µm and 0.2 µm pore-sized polycarbonate filters (47 mm) by a peristaltic pump. The largest size fractions were collected by net tows deployed at 500 m deep (when possible) of 64, 200 and 500 µm mesh size. Cod-ends were split in 2 L of water and 1 L was filtered through 10 µm polycarbonate-filter. Filters from the different size fractions were flash frozen in liquid nitrogen and stored at -80°C until DNA extraction. This dataset represents a detailed sampling in a localized region in order to compare global patterns to local scale patterns.

Data acquisition

· TARA Oceans expedition: Detailed protocol can be found at http://taraoceans.sb-roscoff.fr/EukDiv/ and (Alberti et al., 2017). Briefly, DNA was extracted by cryogenic grinding of cryopreserved membrane filters followed by nucleic acid extraction with NucleoSpin RNA kits (Macherey-Nagel, Düren, Germany) combined with DNA Elution buffer kit (Macherey-Nagel). PCR amplification were performed with the general eukaryotic primer pair 1389F (5'-TTGTACACACCGCCC -3') and 1510R (5'- CCTTCYGCAGGTTCACCTAC-3') targeting the V9 hypervariable region of the 18S rDNA (Amaral-zettler et al., 2009). Sequencing was performed using the Illumina MiSeq platform.

· Malaspina expedition: Detailed protocol can be found in Giner et al. (2019). Briefly, DNA was extracted using the Nucleospin RNA kit (Macherey-Nagel) plus the NucleoSpin RNA/DNA Buffer Set (Macherey-Nagel) procedures. PCR amplification were performed with the general eukaryotic primer pair TAReuk454FWD1 (5'-CCAGCASCYGCGGTAATTCC-3') and TAReukREV3 (5'-ACTTTCGTTCTTGATYRA-3') targeting the V4 hypervariable region of the 18S rDNA (Stoeck et al., 2010). Sequencing was performed using the Illumina MiSeq platform at the Research and Testing Laboratory (Lubbock, USA; http://www.researchandtesting.com).

· Moose-GE: DNA was extracted using the MasterPure Complete DNA and RNA Purification Kit (Epicentre) following manufacturer's instructions. PCR amplification were performed with the general eukaryotic primer pair TAReuk454FWD1 (5'-CCAGCASCYGCGGTAATTCC-3') and TAReukREV3 (5'-ACTTTCGTTCTTGATYRA-3') targeting the V4 hypervariable region of the 18S rDNA (Stoeck et al., 2010). Sequencing was performed using the Illumina MiSeq platform (2x250 bp).

Data curation

Raw reads from the different datasets were independently processed and clustered following DADA2 (Callahan et al., 2016). Resulted amplicons sequence variants (ASVs) were taxonomically assigned using global search implemented in Vsearch (Rognes et al., 2016) against PR2 v4.11.0 database updated with Radiolaria sequences following Sandín et al. (*in prep*, Chapter 2). ASVs were considered for further analyses if (i) the similarity against a reference sequence was above the 90%, (ii) present in at least 2 samples and (iii) with more than 10 reads. These criteria were chosen based on previous results in the exploration of the molecular intracellular diversity of radiolarian specimens (Decelle et al., 2014; Sandin et al. *in prep.*, Chapter 3.1). These filters led to the removal of 2322 ASVs in Tara Oceans expedition (54.84% of the total number of ASVs), 1406 ASVs in Malaspina expedition (52.04%) and 1159 for MOOSE-GE expedition (55.19 %). In general, these removed ASVs accounted for about 4% of the total reads. Final datasets were composed of 1912 ASVs for Tara Oceans expedition, 1296 ASVs for Malaspina expedition and 941 ASVs for MOOSE-GE expedition.

Statistical analyses

In order to integrate the information of the different datasets from the different expeditions, ASVs belonging to the same taxonomical assignation (down to species level) were merged into lineages independently for each expedition and normalized by samples to relative abundance. Final datasets had 113 lineages (and 601 samples) for TARA Oceans, 121

lineages (and 181 samples) for Malaspina and 83 lineages (and 273 samples) for MOOSE-GE. In order to homogenise the differences in taxon reads, a squared root transformation of the relative abundance (Hellinger transformation; Legendre and Gallagher, 2001) was used to carry out statistical analysis: redundancy analysis (RDA), Shannon diversity indexes and indicator value estimates:

The detailed contextual environmental data of TARA Oceans expedition allows to perform RDA to analyse what environmental parameters best explain the distribution and biogeography of Radiolaria. Due to differences in the sampling protocol and in the taxonomic composition among the different size fractions (Fig. 1) two RDA analyses were performed: one in the smallest size fraction and a second in the entire sample pool. The RDA were done over the TARA Oceans environmental dataset and a selection of the lineages dataset that explain 90% of the variability by Escoufier's vectors (Escoufier, 1973). Resulting dataset contained 46 lineages for the smallest size fraction dataset and 54 lineages for all samples. Environmental parameters were selected through a two directional stepwise model selection (10000 steps from an empty model and a maximal model for scope) based on the Akaike information criterion. Final models contained 7 and 22 environmental parameters for the smallest size fraction and all samples respectively. Same protocol was implemented in MOOSE-GE for those samples collected by Niskin bottle, representing the smallest size fractions (due to the same reasons as before, Fig. 1). A selection of 41 lineages and 9 environmental parameters were selected.

Shannon index was calculated to estimate the diversity indexes and an indicator value was calculated (function *indval* from the R package *labdsv*) for all taxonomic lineages within each separate dataset and several environmental clusters, such as different water layers, anoxic environments or oligotrophic areas (where $Chla_{sat} < 0.1$ mg·m^{-3}; Field et al., 1998). All analysis were carried out in R (R Core Team, 2014) under the packages *vegan* (Oksanen et al., 2015), unless specified, and visualized with *ggplot2* (Wickham, 2016).

Results

Global distribution and diversity of Radiolaria in the sunlit ocean

Radiolaria sequences represented a total of 9.03% of total reads in TARA Oceans expedition, 7.85% in Malaspina expedition and 13.06% in MOOSE-GE expedition. Radiolarian sequences were evenly distributed, being present in almost every sample among the different stations and size fractions (Supplementary Figures S1, S2). Within Radiolaria, Collodaria was the most abundant group in the size fractions above 5µm across the different oceanic regions of TARA Oceans (contributing on average 77.74% in the 5-20, 20-180 and 180-2000 µm size fractions) and in MOOSE-GE above 64µm (contributing on average 87.39% in the >64, >200 and > 5000 µm size fractions Fig. 1). In the smallest size fractions Collodaria only dominated in TARA Oceans expedition, being Acantharia the most dominant group in both Malaspina and MOOSE-GE expeditions, contributing to the 59.37 and 38.47% respectively to the radiolarian community. Rad-B is the second most abundant group in the smallest size fractions of the three datasets, followed by Spumellaria.

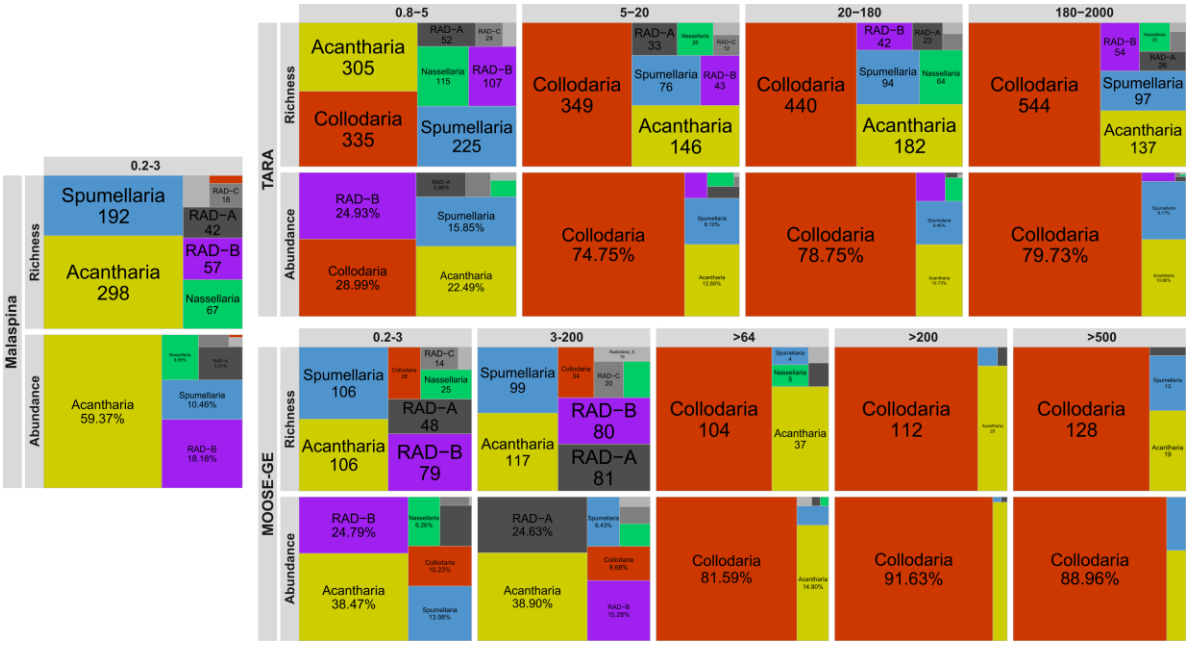


Figure 1. Tree map of the taxonomic affiliation of amplicons for each expedition and size fraction, the area represents the proportion of total number of unique amplicons (Richness) and relative abundance (Abundance).

Considering the biogeography of smallest size fraction only, differences were observed between datasets (Fig. 2). In TARA Oceans expedition, Collodaria is the most abundant order (with an average relative abundance per sample of 30.82%), followed by Rad-B (23.75%), Acantharia (23.06%) and Spumellaria (13.90%). Whereas in Malaspina the most abundant group is Acantharea (59.15%), followed by Rad-B (20.20%) and Spumellaria (11.94%) and then Rad-A (10.41%) and Nassellaria (10.32 %). And in MOOSE Acantharia (37.94%) is the most abundant group per sample, followed by Rad-B (25.37%) and Spumellaria (13.90%). In general, the small size fraction of Radiolaria contributed to the total eukaryotic community in greater relative proportion in the euphotic Pacific Ocean (4.09% on average in TARA Oceans and 2.99% on average in Malaspina) or the Atlantic ocean (3.03% on average in TARA Oceans and 3.03% on average in Malaspina) than in the Southern Ocean (0.12% on average in TARA Oceans) or the South Australian Basin (1.53% on average in Malaspina). Worth mentioning that in both Malaspina and MOOSE-GE datasets, Nassellaria takes more importance than in TARA oceans, where only in one station related to constant upwelling region (067) stands out.

Regarding the bigger size fractions of TARA Oceans expedition, in the South Pacific Ocean and in the Indian Ocean Radiolaria contributed on average to the 9.79% and 9.65% respectively to the total eukaryotic community in the different size fractions. Whereas in the Red Sea and Southern Ocean were found the lowest values of radiolarian contribution (1.59% and 0.30% respectively; Supplementary Figure S1). In the Southern Ocean, Acantharia accounted for 89.67% of the reads within Radiolaria, being the most abundant order. Also in the Red Sea and Western Mediterranean sea, Collodaria relative abundance decrease, taking the dominant role Acantharia and Spumellaria. These differences are not seen in MOOSE-GE, where Collodaria dominates in the bigger size fractions (Supplementary Figure S2). Only the size fraction 3-200 µm follow similar trends of those found in the small (0.2-3 µm) size fraction, with the exception of Rad-A that its relative abundant increase.

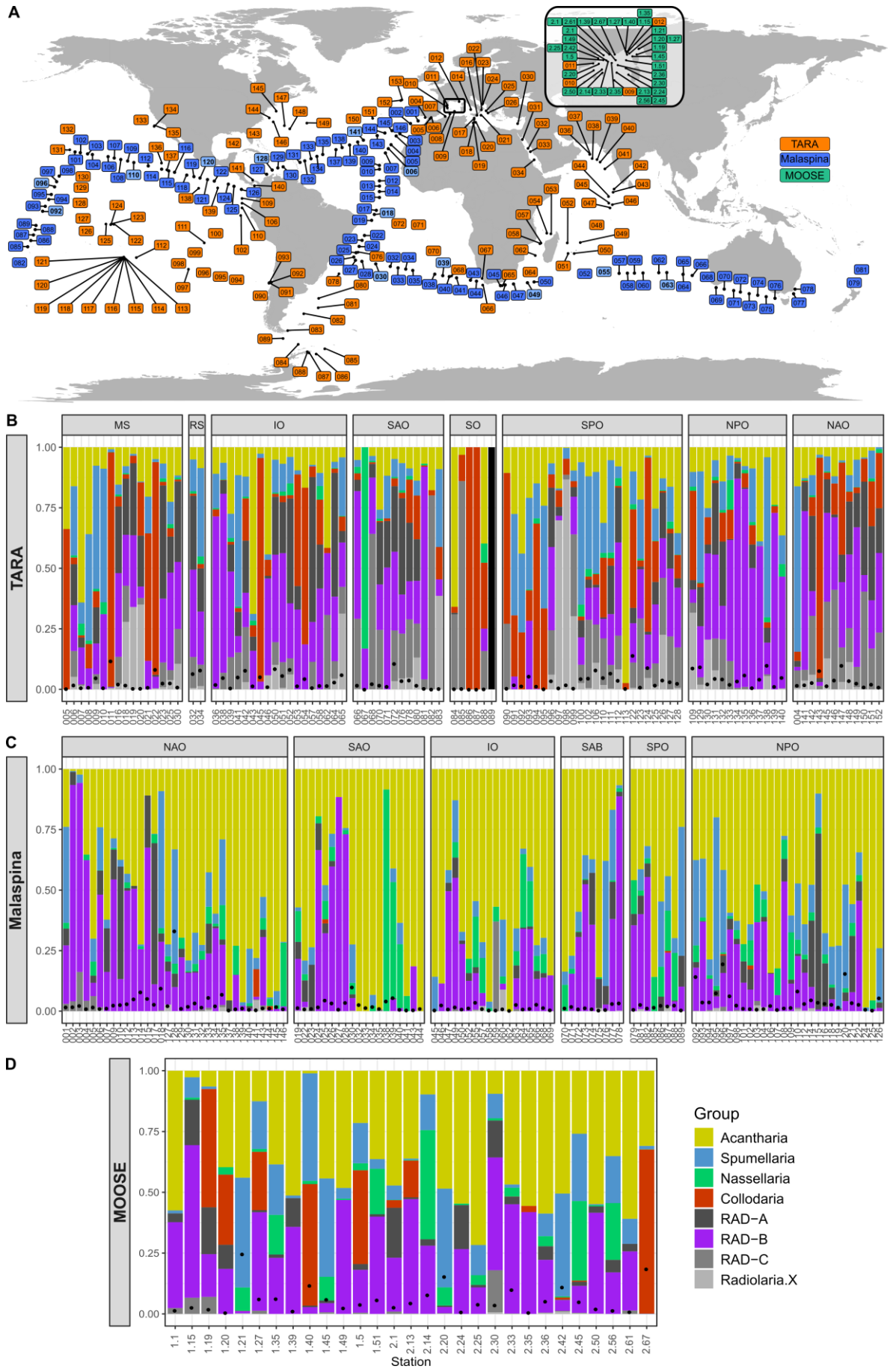


Figure 2. Legend on following page.

Differences in the size fractions are reflected in the Shannon diversity indexes, finding more diverse communities in the smallest size fraction than in larger size fractions (Supplementary Figure S3). For further analysis, in order to have a global picture of radiolarian diversity and biogeography by integrating the three datasets studied, we chose to focus on the most diverse size fraction.

Environmental parameters driving the biogeography in surface waters

The 46 selected lineages for the redundancy analysis encompassed 3 lineages within Rad-B, 14 in Spumellaria, 1 in RAD-A, 11 in Acantharia, 1 in RAD-C, 5 in Collodaria, 10 in Nassellaria and 1 in Radiolaria-X. The adjusted R-squared of the analysis was of 28.28% (44.31% unadjusted), and the two first axis explain a 19.68% of the variance (Fig. 3). The first axis separates stations related to clear waters and with a higher day light duration of those with more particles and deeper. This axis is positively correlated to the day light duration (min), the backscattering coefficient (1/m), the angular scattering coefficient (m/sr) and the sea surface temperature gradient (°C/100 Km). It is negatively correlated to depth (m), carbonate (µmol/l) and temperature (°C). The second RDA axis separates warm and cold stations. This axis is positively correlated to temperature and negatively correlated to the sea surface temperature gradient, angular scattering coefficient, the day length, the particulate backscattering coefficient and in a lesser proportion to the carbonates and the depth.

From the selected lineages, up to 18 had a significant response to environmental parameters in their distribution. Among the 3 lineages in Rad-B, two stands out having preferences towards deep environments and so does the environmental lineages Rad-A and Rad-C. From the 14 lineages within Spumellaria, 4 have certain preference towards cold and deep environments *Liosphaera* sp. being the most discriminant lineage of this group and *Spongosphaera streptacantha* preferred surface and warm samples. Acantharia had 6 out of 11 lineages relatively highly related to environmental parameters and with contrasting preferences like that of Spumellaria. In one side, *Lychnaspis giltschii*, *Xiphacantha quadridentate* and *Haliommatidium* sp. (belonging to clades E1, F3 and F3 respectively) preferred warm and shallower waters, and *Litholopus* sp. (from clade C3) and the environmental clade Acantharia 2 had certain preference towards deeper environments. Within Collodaria, only one (i.e. Raphidozoum acuferum) showed a strong response, directly related to the clearness of the water column. Finally, 2 lineages within Nassellaria showed a direct relationship towards deep and cold environments (*Lophospyris* aff *ivanovi* and *Protoscenium* cf *intricatum*).

Figure 2. (A) World map showing the different sampling stations of the Tara Oceans (orange), Malaspina (blue) and MOOSE-GE (green) expeditions from which the metabarcoding datasets were analyzed. Malaspina stations in bold and light blue indicate samples used for the vertical profile. (B, C & D) Bar charts showing the diversity and relative abundance of main taxonomic groups of Radiolaria in the euphotic zone for the smallest size fraction (pico-nano) of each dataset (B: TARA, 0.8-5 µm; C: Malaspina, 0.2-3 µm, D: MOOSE-GE, 0.2-3 µm) across pelagic stations of different oceanic basins (MS: Mediterranean Sea; RS: Red Sea; IO: Indian Ocean; SAO: South Atlantic Ocean; SO: Southern Ocean; SAB: South Australian Basin; SPO: South Pacific Ocean; NPO: North Pacific Ocean; NAO: North Atlantic Ocean). Black dots represent the average relative abundance of radiolarian reads in the sample. Empty columns represent no radiolarian reads for given sample.

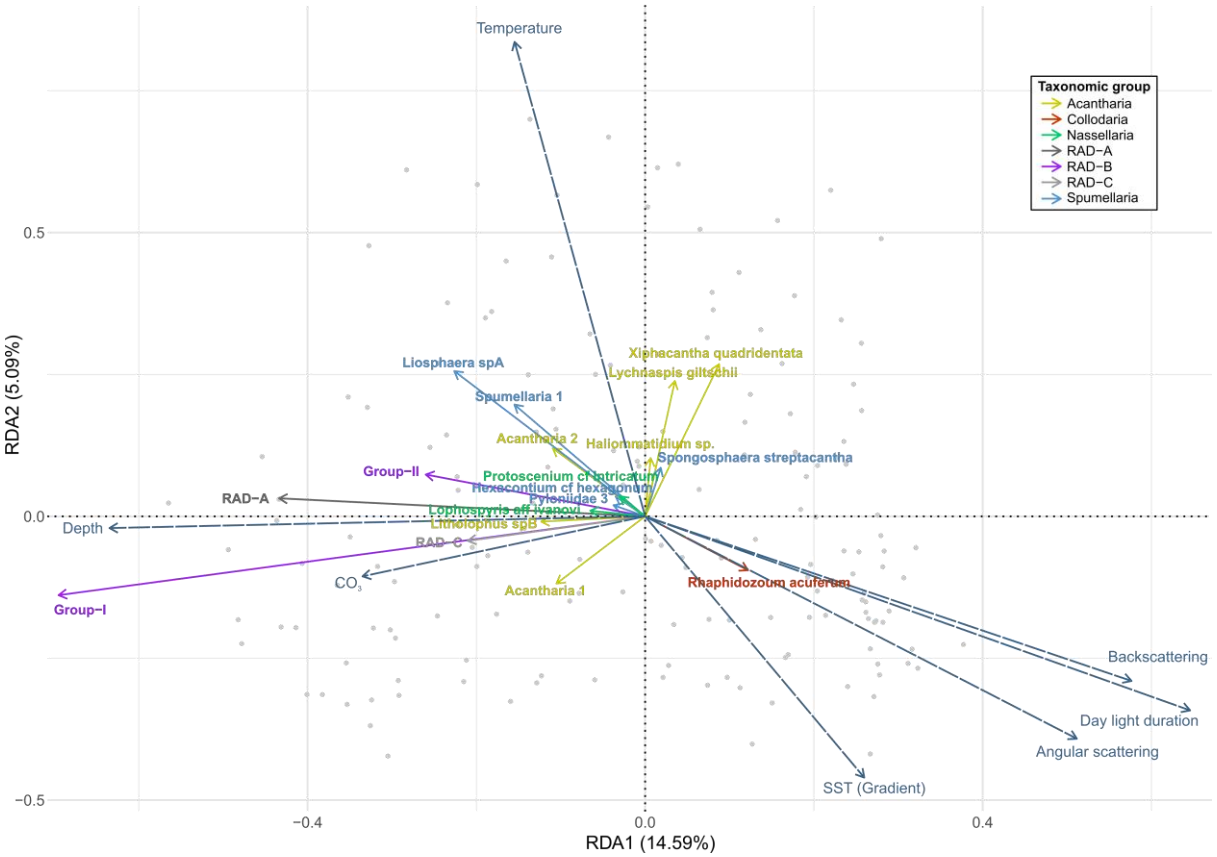


Figure 3. Redundancy analysis (RDA) triplot showing the impact of environmental variables on the distribution of the 46 Escoufier-selected taxa (only most relevant shown) for the smallest size fraction (0.3-5µm). The adjusted R-squared of the analysis is of 28.28% (44.31% unadjusted). Abbreviations of environmental parameters and units: Angular Scattering: Angular Scattering coefficient at 470nm (1/m); Backscattering: Backscattering coefficient of particles at 470 nm (1/m); CO3: Carbonate (µmol/l); Day light duration (min); Depth (m); SST (Gradient): Sea Surface Temperature gradient horizontal (°C/100Km); Temperature (°C).

When analysing all different size fractions there are clearly two contrasting group of samples (Supplementary Figure S4). In one group there are those corresponding to the smallest size fraction (0.8-5µm), mainly driven by temperature and bathymetry with lineages belonging to Acantharia, Rad-B, Rad-A, Spumellaria and Nassellaria. The other group corresponds to those samples from the size fractions bigger than 5µm, mainly driven by nutrients, day light duration, Chlorophyll A and clearness of the water with lineages belonging to Collodaria.

Vertical community structure in the global ocean

In contrast to surface waters where Acantharia dominated or was among the most dominant groups, Mesopelagic samples are clearly dominated by Spumellaria (Fig. 4). The Mesopelagic water layer gathered the biggest richness for all taxonomic groups except for Rad-A, showing their biggest richness at DCM and surface. The biggest values of Shannon diversity where found in the mesopelagic for most of the groups, except for Acantharia being at the DCM. Also, in the mesopelagic was found the greater contribution of Radiolaria to the plankton community. Concerning the bathypelagic water layer Spumellaria was still the most abundant group and Rad-B reached its maximum relative abundances (32%).

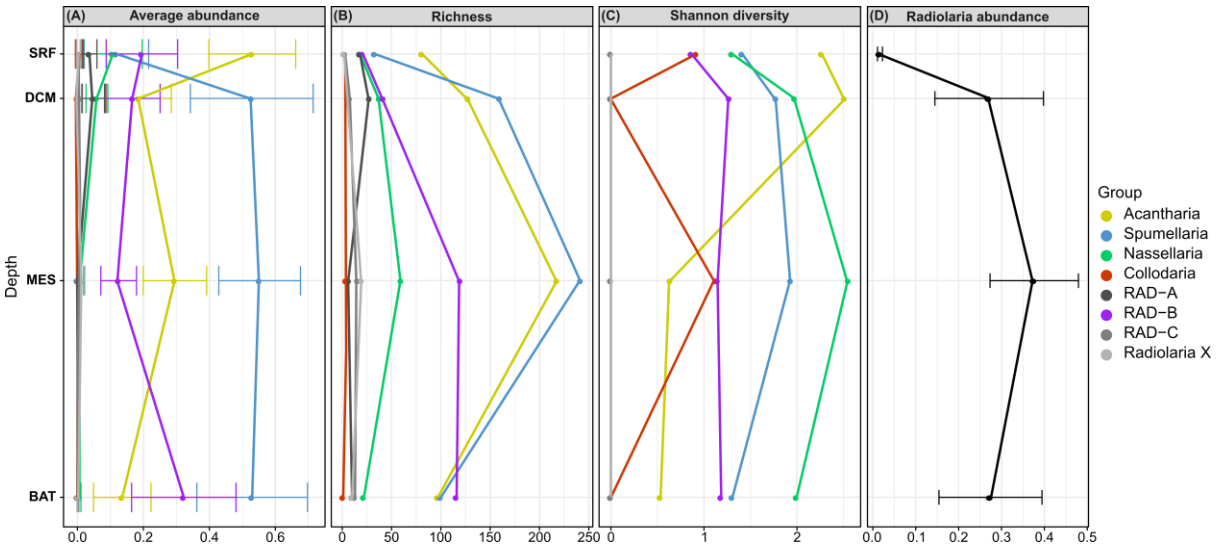


Figure 4. (A) Average relative abundance, (B) total richness and (C) Shannon diversity for the main taxonomic groups and (D) average relative contribution of radiolarian reads to the total eukaryotic community at the four water depths derived from the V4 rDNA of Malaspina expedition. Horizontal bars represent the standard deviation for the average values.

Indicator values indicated *Lychnaspis giltschii* (Acantharia, Clade E) as highly related to surface waters (Table 1). At the DCM Spumellaria 1, *Didymocyrtis tetrathalamus* (Spumellaria, Spongodiscoidea), Rad-A were highly indicative. *Litholopus* sp., *Arachnospongus varians*, Acantharia 4, Spumellaria X, *Gigartacon muelleri*, Acantharia 2b and *Cladococcus irregularis* still were indicative of the DCM yet in a lesser extent. The mesopelagic water layer had the greatest number of indicative lineages. Spumellaria 5, Acantharia 1, Radiolaria X, *Acanthoplegma* sp., *Liosphaeroidea* sp. and *Heliosphaera bifurcus* were highly indicative of the mesopelagic water layer, followed by Spongodrymidae (Spumellaria, Clade I), *Eucecryphalus cervus* (Nassellaria, Theopiliidae), *Octodendron cubocentron (*Spumellaria, Clade H), *Haliommilla capillacera* (Spumellaria, Rhizosphaeroidea) and *Gigartacon fragilis* (Acantharia, Clade C). The bathypelagic had relatively high indicative lineage the group II from Rad-B, followed by group I (Rad-B), *Heliosphaera* aff. *actinote* and Spumellaria 4.

Mediterranean scale

A redundancy analysis carried out on the MOOSE dataset may be used to improve the understanding of local compared to global scale patterns in biogeography and environmental diversity (Fig. 5). The analysis had an adjusted r squared of 42.15% (46.94% unadjusted) and the two first axis explained a 29.56% of the variance. Yet, the depth of the sample had a great impact in the clustering of the RDA showing three different groups; the surface samples, the DCM samples and the mesopelagic and Bathypelagic samples. The RD1 axis explains the 18.06% of the variance and separates those samples coming from mesopelagic and bathypelagic environments low in oxygen (RDA1>0) of those from the surface with a higher concentration of oxygen (RDA1<0). This axis is positively correlated to silica concentration, depth, the bathypelagic environment, salinity, the mesopelagic and DCM environment and nitrates, and negatively correlated with the temperature, the concentration of oxygen and the surface environment. The second axis (RDA2) explains 11.5% of the variance and separates cold samples (RDA2>0) of warm samples (RDA2<0). This axis is positively correlated

with the DCM, the concentration of Oxygen and nitrates, and negatively correlated with the concentration of silica, the depth, the bathypelagic and mesopelagic environment, the salinity and the surface environment.

Table 1. Indicator species of each water depth for Malaspina and MOOSE-GE expeditions. ASVs assigned to the same species were pooled together to obtain the given species. Only species with a significant (p-value < 0.05) indicative value and above 0.4 are shown.

Depth	Malaspina			MOOSE		
	indval	Group	Species	indval	Order	Species
Surface	0.620	Acantharia	Lychnaspis giltschii			
	0.469	Nassellaria	Ceratocyrtis cf galea			
	0.427	Acantharia	Amphibelone anomala			
	0.401	Acantharia	Amphiastrus tetrapterus			
	0.400	Acantharia	Clade F sp.			
DCM	0.803	Spumellaria	Spumellaria 1 sp.	0.525	RAD-A	RAD-A sp.
	0.667	Spumellaria	Didymocyrtis tetrathalamus	0.504	Acantharia	Acanthocolla cruciata
	0.638	RAD-A	RAD-A sp.			
	0.554	Acantharia	Litholophus spB			
	0.549	Spumellaria	Arachnospongus varians			
	0.535	Acantharia	Acantharia 4 sp.			
	0.526	Spumellaria	Spumellaria sp.			
	0.526	Acantharia	Gigartacon muelleri			
	0.503	Acantharia	Acantharia 2b X sp.			
	0.467	Spumellaria	Cladococcus irregularis			
	0.458	Acantharia	Acanthocolla cruciata			
	0.411	Acantharia	Clade E sp.			
	0.410	Acantharia	Acanthochiasma sp.			
Mesopelagic	0.753	Spumellaria	Spumellaria 5 sp.	0.814	Spumellaria	Heliosphaera aff actinota
	0.750	Acantharia	Acantharia 1 sp.	0.768	Spumellaria	Liosphaera spA
	0.740	RadiolariaX	Radiolaria sp.	0.703	Acantharia	Acantharia 1 sp.
	0.730	Acantharia	Acanthoplegma spA	0.683	Spumellaria	Spumellaria sp.
	0.718	Spumellaria	Liosphaeroidea XX sp.	0.672	Spumellaria	Spumellaria 5 sp.
	0.711	Spumellaria	Heliosphaera bifurcus	0.594	Acantharia	Clade C sp.
	0.549	Spumellaria	Uncultured Rhizaria	0.568	RadiolariaX	Radiolaria sp.
	0.496	Nassellaria	Eucecryphalus aff cervus	0.452	Acantharia	Acanthoplegma spA
	0.473	Spumellaria	Octodendron aff cubocentron	0.427	Spumellaria	Uncultured eukaryote
	0.463	Spumellaria	Uncultured marine			
	0.462	Spumellaria	Haliommilla capillacea			
	0.434	Acantharia	Gigartacon fragilis			
	0.413	RAD-B	RAD-B-Group-IV sp.			
Bathypelagic	0.506	RAD-B	RAD-B-Group-II sp.	0.749	RAD-B	RAD-B-Group-II sp.
	0.498	RAD-B	RAD-B-Group-I sp.	0.746	RAD-B	RAD-B sp.
	0.473	Spumellaria	Heliosphaera aff actinota	0.691	RAD-C	RAD-C sp.
	0.438	Spumellaria	Spumellaria 4 sp.	0.669	RAD-B	RAD-B-Group-I sp.
				0.666	RAD-B	RAD-B-Group-IV sp.
				0.658	Nassellaria	Lithomelissa setosa
				0.581	Spumellaria	Rhizosphaera sp.
				0.545	Spumellaria	Spongopyle osculosa
				0.530	Acantharia	Acanthochiasma sp.
				0.500	Acantharia	Acantharia 2a sp.
				0.484	RAD-B	Sticholonche sp.
				0.467	Acantharia	Phyllostaurus echinoides
				0.411	Acantharia	Acantharia 3 sp.

From the 41 selected species, 17 had a substantial response to environmental parameters (values above 0.05 in both RDA1 and RDA2). The four lineages within Rad-B showed strong response against environmental parameters, group I and Rad-B X had clear

preferences towards deep and cold samples (as well as Radiolaria X), group III towards the DCM (as well as Rad-A) and enriched with nitrates, and group-IV showed preferences in between (as well as Rad-C). Five out of 13 acantharian lineages showed a strong response to the environmental parameters and having contrasted preferences; Clade F sp., *Amphiastrus tetrapterus* (from clade F) and *Lychnaspis giltschii* (from clade E) had certain preferences towards shallow and warm samples; *Acanthocolla cruciata* (from clade D) and *Gigartacon muelleri* (from clade C) prefers DCM environments with high concentrations of oxygen and nitrates; and Acantharia 2 towards deep and cold samples. From the 14 lineages in Spumellaria, 5 showed strong response and were scattered in two groups; *Heliosphaera* aff *actinota* (Liosphaeroidea), *Heliosphaera bifurcus* (Liosphaeroidea) and *Rhizosphaera* sp. (Rhizosphaeroidea) had clear preferences towards deep and cold environments and *Hollandosphaera hexagonia* (Hexastyloidea, Clade C) and *Cypassis irregularis* (Spongodiscoidea) towards enriched DCM waters. Regarding Nassellaria, 2 lineages (out of 4) showed a contrasting response; *Protoscenium* cf *intricatum* (Plagiacanthoidea) preferred surface and warm waters and *Lithomelissa setosa* (Plagiacanthoidea) deep and cold environments. The 3 lineages belonging to Collodaria had a very weak response (RDA axes < 0.024).

At the local scale of the Western Mediterranean Sea the vertical profile structure is similar to that shown for Malaspina dataset. Regarding specificities, only two lineages showed to be indicative of the surface layer, yet with very low indicative values (>0.3). In the DCM Rad-A and *Acanthocolla cruciata* (Acantharia, Clade D) showed to be the most indicative lineages (Table 1). *Heliosphaera actinota* (Spumellaria, Liosphaeroidea) and *Liosphaera* sp. (Spumellaria, Liosphaeroidea) had both a very high indicative value for the mesopelagic (>0.75, p-value<0.001). Also important are Acantharia 1, Spumellaria X, Spumellaria 5, Acantharia C and Radiolaria X, followed by *Acanthoplegma* sp. (Acantharia, Clade A) and Spongodrymidae (Spumellaria, Clade I). In the bathypelagic layer indicative lineages are leaded by Rad-B (groups II, X, I and IV), followed by Rad-C, *Lithomelissa setosa* (Nassellaria, Plagiacanthoidea), *Rhizosphaera* sp. (Spumellaria, Rhizosphaeroidea), *Spongopyle osculosa* (Spumellaria, Spongodiscoidea), *Acanthochiasma* sp. (Acantharia, Clade B) and Acantharia 2. Also important are *Sticholonche* sp. (Rad-B), *Phyllostaurus echinoides* (Acantharia, Clade B) and Acantharia 3.

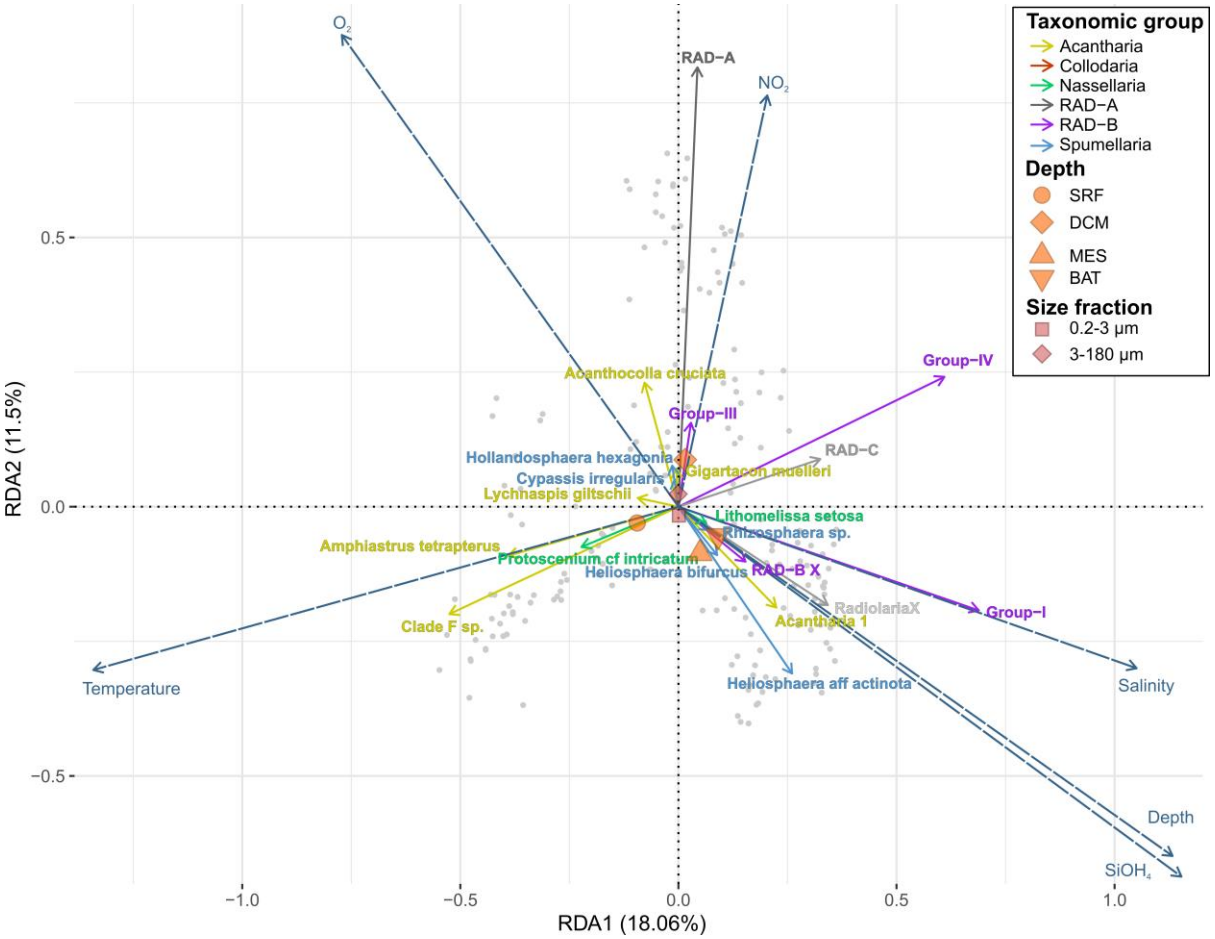


Figure 5. Redundancy analysis (RDA) triplot showing the impact of environmental variables on the distribution of the 41 Escoufier-selected taxa (only most relevant shown) for MOOSE-GE expedition on the samples collected by Niskin Bottle. The adjusted R-squared of the analysis is of 42.15% (46.94% unadjusted).

Discussion

Global distribution and diversity of Radiolaria in the sunlit ocean

Our results showed that Radiolaria represent an important component of the eukaryotic plankton community in every dataset studied, accounting for about 9% of the total reads, as previously reported in global environmental molecular surveys (de Vargas et al., 2015; Pernice et al., 2016; Giner et al., 2019) and several regional studies (e.g.: Not et al., 2007; Brisbin et al., 2019; Gutierrez-Rodriguez et al., 2019). In general, Radiolaria tend to prefer the open ocean than enclosed seas, finding higher relative abundances in the Pacific, Indian and Atlantic Oceans than in the Mediterranean and Red sea. Yet, there have been morphology-based reports on Nassellaria blooms in brackish waters (Boltovskoy et al., 2003), therefore this should be considered as a global trend. Temperature is one of the most important factors shaping radiolarian community in the photic ocean, finding the lowest relatives abundances in the Southern Ocean, as previously reported from the sediments (Boltovskoy and Correa, 2016a). Within Radiolaria, there are two different communities depending on the size fraction studied. In one side the colonial Collodaria dominate the bigger size fractions, in agreement with studies performing *in situ* imaging techniques on large protists (Biard et al., 2016).

Whereas the smaller size fractions follow similar trends than those seen for the entire eukaryotic community (de Vargas et al., 2015) finding higher diversified assemblages.

The separation in two different groups regarding their sizes is reflected in their contrasted living modes, as already pointed out by Faure et al. (2019). Colonial Collodaria tend to inhabit large oligotrophic and shallow waters. The big gelatinous organic matter constituting the colonies, allows them to capture nutrients extracellularly and feed on prey as big as copepods or mollusc larvae (Swanberg et al., 1986). They also possess a large amount of dinoflagellates symbionts (Probert et al., 2014) which were suggested to contribute to the primary production of the holobiont (Caron et al., 1995). In contrast, solitary radiolarians are more abundant in eutrophic waters, where Acantharia tend to be the dominant group followed by Spumellaria. Both Acantharia and Spumellaria harbour a great number of algal symbionts. Yet, Acantharia live with the cosmopolitan haptophyte *Phaeocystis* (Decelle et al., 2012a, 2018), and it has been proven the tight relationships and the ability from the host to exploit the photosynthetic efficiency of the algae (Decelle et al., 2019). On the contrary, little is known about the dinoflagellates algal symbionts of Spumellaria (Probert et al., 2014; Yuasa et al., 2016), although it seems that the relationships is not as tight as for Acantharia, being in some cases facultative (Zhang et al., 2018). Based in this distribution pattern, most probably the mixotrophic behaviour of Radiolaria shaped their distribution in euphotic water layers, allowing their great ecological success.

Vertical community structure in the global ocean

When the dominance of the mixotrophic groups disappear in the mesopelagic water layers, it is possible to find more diverse assemblages, as recently reported for the global eukaryotic community (Giner et al., 2019), where Spumellaria takes dominance over Acantharia compare to the photic layer. Probably, the silicate skeleton of Spumellaria plays an important role in their higher dominance in mesopelagic waters over the fast dissolving strontium skeleton of Acantharia. Stands out the large importance of environmental clades at depth, already mention by the latest morpho-molecular classifications performed in Radiolaria (Decelle et al., 2012b; Biard et al., 2015; Sandin et al., 2019, Chapter 1.1; Sandin et al. *in prep*, Chapter 1.2), where they have noticed a significantly lower sampling representation towards deep and extreme environments. Also, it worth mentioning the big relative abundance of Rad-B all over the water column but especially towards deep environments. In Sandin et al. (*in prep*, Chapter 2), they have shown the highly diverse environmental clade of Rad-B, with up to 4 different subclades. *Sticholonche* belongs to the subclade called group-IV, yet there are only 2 sequences morphologically described, they cluster basal to this clade. Therefore, as already mentioned by Nikolaev et al. (2004) and Sandin et al. (*in prep*, Chapter 2), due to its evolutionary pattern, its ecological importance and its evasiveness from different studies, they might represent a species complex or have a different morphology compare to the characteristic radiolarian star-shaped.

Most of this environmental diversity associated to Rad-B was indicative of the bathypelagic ocean. In contrast, the mesopelagic environment is highly represented by Spumellaria groups, precisely from the lineage II and III from Sandin et al. (*in prep*, Chapter 1.2). The later of these spumellarian groups was traditionally classified under Entactinaria, a

controversial order recently believed to be polyphyletic (Nakamura et al. *in prep.*). Probably previous classifications based on the fossil record (De Wever et al., 2001) found differences between the Spumellaria and Entactinaria due to their differences in water mass preferences that might be reflected in a contrasting stratigraphic range. Although further analysis must discriminate if these differences are due to ecological preferences or to a lack of biogeography in radiolarian assemblages from the dark ocean. Since similar vertical patterns are found in both Malaspina and MOOSE-GE, a multivariate analysis could be implemented in order to test the lack of biogeography in Radiolaria assemblages of the deep ocean.

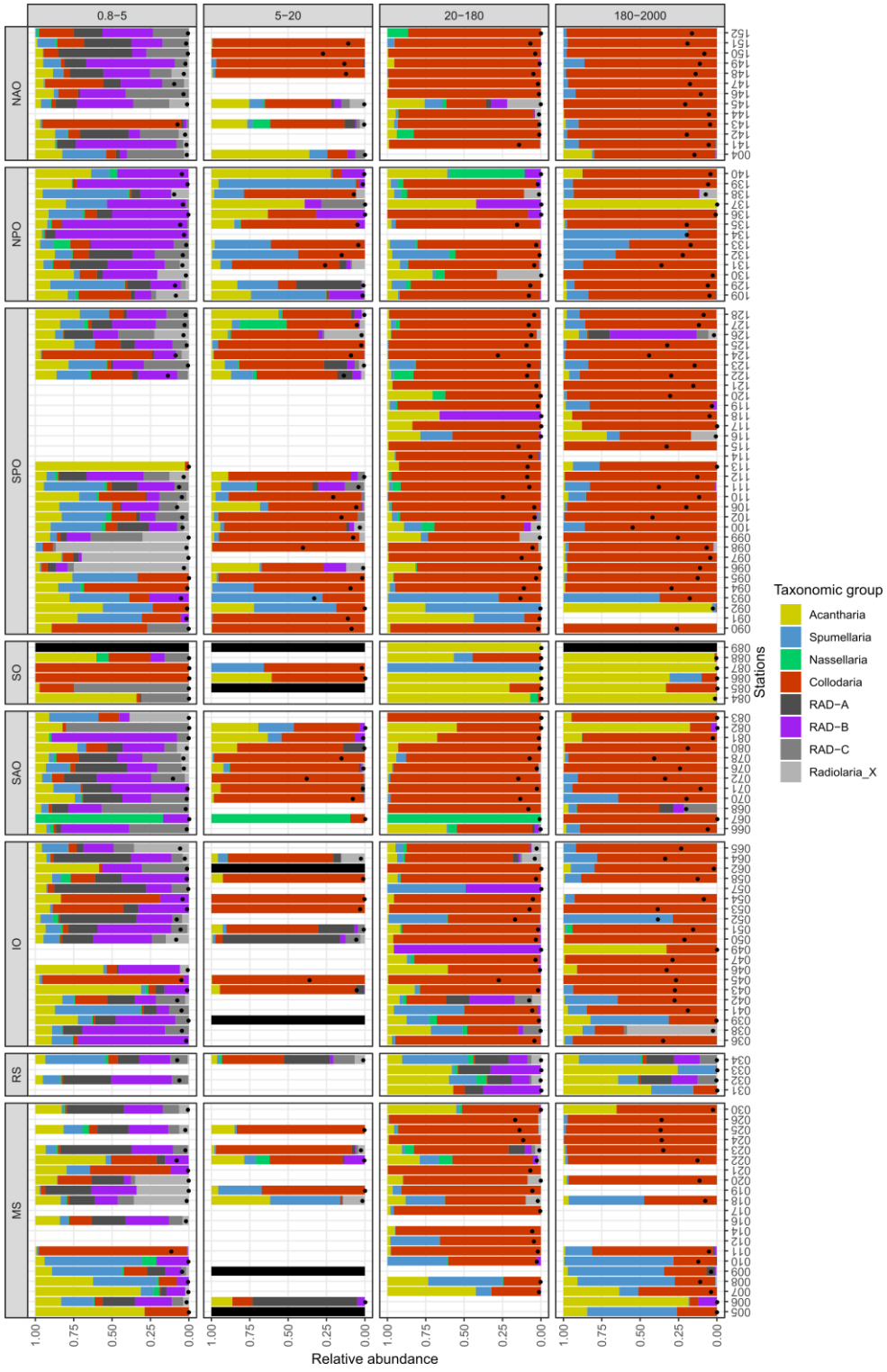
However, no lineages, or very few, were found specifically associated to the euphotic layer, meaning that they are also present in aphotic water masses. Since we were focusing our efforts in the lower size fraction we could be sequencing swarmers (Not et al., 2007), that they are believed to reproduce in the deep ocean. The few lineages indicative of the photic layer were mostly Acantharia belonging to the mixotrophic clades E and F (Decelle et al., 2012b). And as Decelle et al. (2013) have proposed, probably symbiotic acantharian clades reproduce in the photic layer never reaching the dark ocean, also suggested by our results. Therefore, sinking particles could be another reason of the lower indicative species found in the sunlit ocean. However, it could also be due to methodological or technical issues, such as the taxonomic resolution, in which we pooled together several ASVs clustered under the same taxonomic identity. Besides, as already mention by Sandin et al. (*in prep*, Chapter 3.1), some very abundant ASVs are also present in additional samples due, most probable, to cross-contaminations or tag-jumping. Therefore, we cannot exclude that by applying different detection threshold in our analytical pipelines, other patterns would appear.

General remarks on interpreting metabarcoding analysis from different datasets

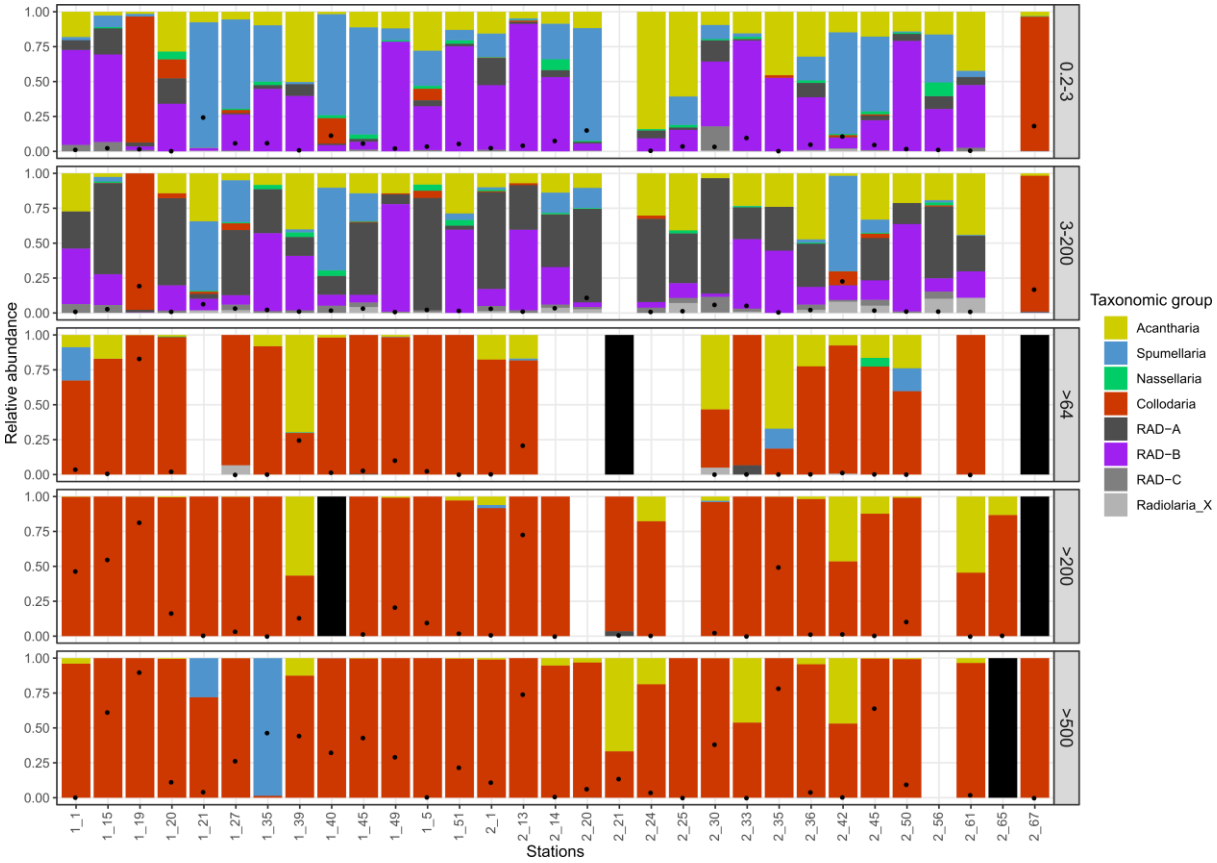
Our results showed similar global patterns in radiolarian distribution among the different datasets studied, despite that they were designed to answer different questions and implemented different protocols for samples collection and data acquisition. Furthermore, studying different datasets allow better understanding the taxonomic biases (e.g. the overrepresentation of Collodaria regarding the underestimation of Nassellaria) already mentioned by Sandin et al. (*in prep*, Chapter 3.1), and showed by Weiner et al. (2016) for Foraminifera. The clearer taxonomic difference is the bias in the sampling method towards overestimating Collodaria when using net tows regarding Niskin bottles or pump, since the bigger size fractions were sampled by net tows, whereas the smaller size fractions with Niskin bottles or pump. Another difference might be due to the DNA extraction procedure, since TARA Oceans followed freeze-thaw cycles, whereas in Malaspina and MOOSE-GE was performed a chemical extraction. And as already pointed out by Not et al. (2007), cell breakage may play a crucial role in molecular diversity studies, where both Spumellaria and Nassellaria tend to have a robust skeleton compared to their relatives. This bias is also reflected in the higher amount of environmental sequences regarded to weaker-skeleton specimens in these two groups (Sandin et al., 2019, Chapter 1.1; *in prep*., Chapter 1.2). When the DNA was extracted chemically, Nassellaria shows a higher relative contribution to the total reads. Yet, their low representation compared to that of Spumellaria might indicate other biases such as lower intracellular variability (Sandin et al. *in prep*, Chapter 3.1) or lower rDNA copy number. Since despite of being one of the groups with lower representation within

Radiolaria accounts for the higher morphological diversity described (Suzuki and Not, 2015) and showed the highest diversity across the water column. Lastly, it is also important to stress the high dependency of metabarcoding analysis to the reference database. Since many more ASVs will be pooled together within the same environmental name (i.e.; Rad-B Group II or Rad-A), whereas those described will tend to be more scattered, probably showing a lower ecological response to environmental variables. Altogether, it brings a better understanding of the extant biodiversity and biogeography of Radiolaria and their interpretation in molecular environmental surveys.

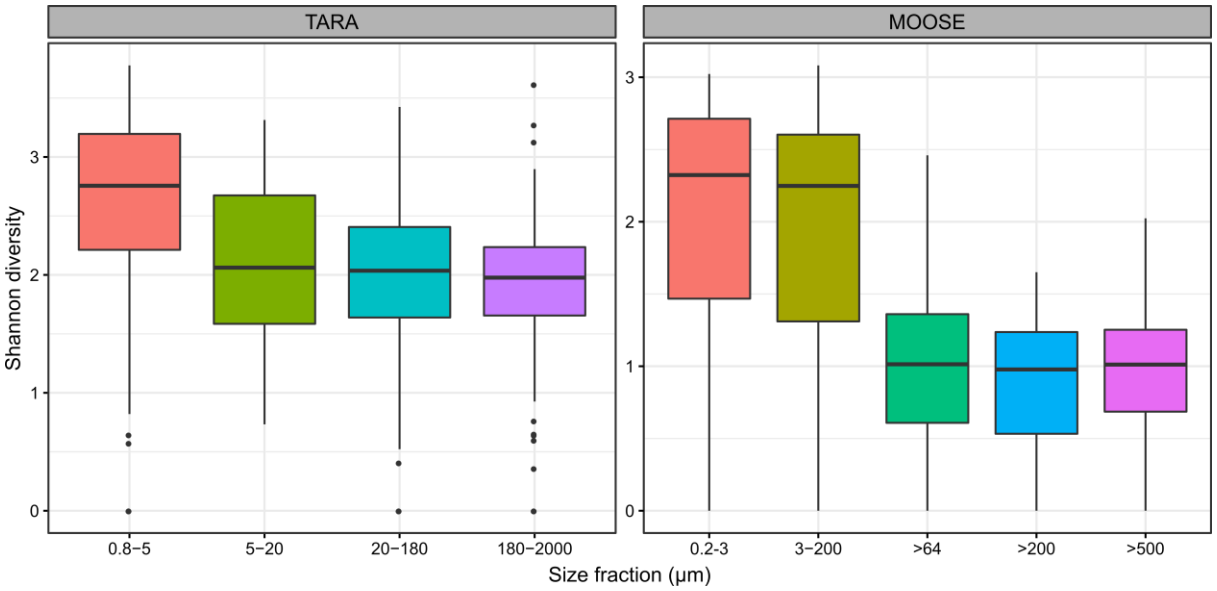
Supplementary Material



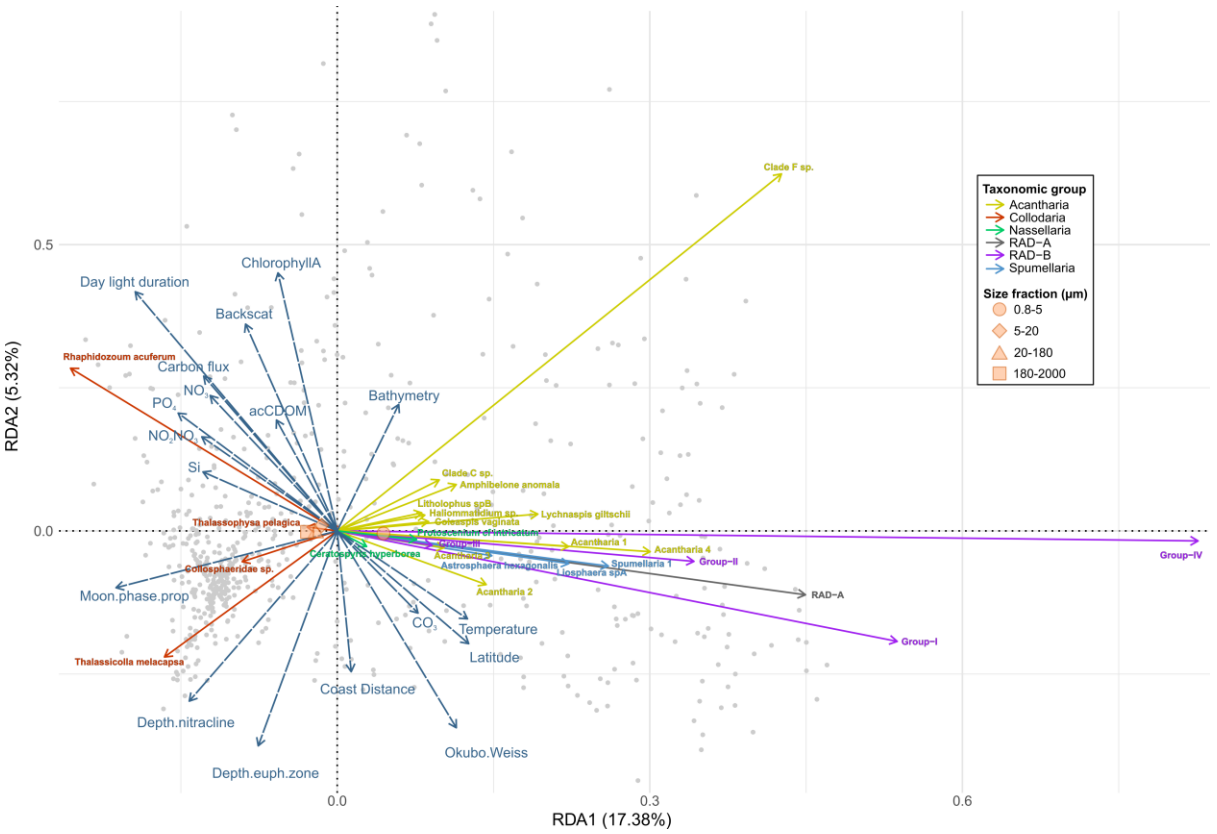
Supplementary Figure S1. Bar chart showing the diversity and relative abundance of main taxonomic groups of Radiolaria in the different size fractions of the V9 rDNA from Tara Oceans expedition dataset from surface and DCM across pelagic stations of different oceanic basins (MS: Mediterranean Sea; RS: Red Sea; IO: Indian Ocean; SAO: South Atlantic Ocean; SO: Southern Ocean; SPO: South Pacific Ocean; NPO: North Pacific Ocean; NAO: North Atlantic Ocean). Black dots represent the average relative abundance of radiolarian reads in the sample. Empty columns represent no data for given sample. Black bars represent no radiolarian reads for given sample.



Supplementary Figure S2. Bar chart showing the diversity and relative abundance of main taxonomic groups of Radiolaria in the different size fractions (0.2-3, 3-200, >64, >200 and >500) of the V4 rDNA from MOOSE-GE expedition dataset across Mediterranean stations from vertical tow Nets. Black dots represent the relative abundance of radiolarian reads in the sample. Empty columns represent no data for given sample. Black bars represent no radiolarian reads for given sample.



Supplementary Figure S3. Box plot of the shannon index diversity for the different size fractions of TARA Oceans expedition (V9 rDNA) and MOOSE-GE expedition (V4 rDNA) of the euphotic water layer.



Supplementary Figure S4. Redundancy analysis (RDA) triplot showing the impact of environmental variables on the distribution of the 52 Escoufier-selected taxa (only most relevant shown) for the pool of all samples from TARA Oceans expedition. The adjusted R-squared of the analysis is of 32.28% (40.07% unadjusted).

Discussion and perspectives

1. Towards an integrative classification: limitations and advantages

Morphological features still represent the gold standard for delineating species. Yet phenotypic variability, growth stages, sex differences or ecotypes are never well established boundaries. When integrating the time dimension with morphology, it becomes even more difficult to establish limits between two similar morphologies that are believed to be sister species. How to settle different binomial names to a constantly (on geological time frame) evolving morphology? On top of that, morphology may be biased by preservation of specimens (i.e. fossils), the observation techniques used (i.e. most of the cases ultrafine scale resolution require dehydration, changing the morphology) or the interpretation and background knowledge of the observer (for further details see introduction part 3, page 10, and references therein).

Molecular tools have become in the last century an exciting alternative to morphological observation, by providing access to the vast “unculturable” world of microbes. Back in 1977, Carl Woese showed the advantages of using rRNA comparisons to discriminate between morphologically similar bacteria, delimiting the three domains of cellular life (Woese and Fox, 1977). DNA comparison allowed exploring previously unknown diversity, and to define and discover new lineages of life. Nevertheless, molecular tools, as for morphological approaches, are also subjected to a number of drawbacks. These include the choice of the molecular marker best suited to the phylogenetic question asked and to the specific taxonomic level required. As seen in chapter 1 and in previous classifications of Acantharia (Decelle et al., 2012b) and Collodaria (Biard et al., 2015), rDNA (18S gene + partial 28S gene; D1 & D2 regions) defines rather good phylogenetic relationships at the superfamily and family level in Radiolaria. Yet long branch artefacts may lead to a wrong interpretation at the order or higher rank classifications (as in chapter 2). The picture gets even more complicated when adding Foraminifera in the game, since its fast-evolving rDNA gene produces poorly resolved phylogenetic relationships (Moreira et al., 2007). Besides, in Chapter 3.1 we have seen intracellular diversity and sequencing errors can lead towards an overestimation of actual diversity in most of the cases. But rDNA sequencing approaches may lead to an underrepresentation in some other cases, finding highly similar molecular markers for different morphologically described species among Nassellaria. Such taxa specific differences can become important issues when interpreting other studies, for example molecular environmental surveys. In chapter 3.2 Nassellaria was certainly underrepresented compared to other radiolarian groups in metabarcoding analysis. Therefore, as any model in nature, the more complex it gets the more vulnerable points it may have. And molecular tools have many different steps susceptible to biases.

In this manuscript we have revisited from a molecular point of view, previous questions addressed by morphological approaches. An impressive work in the fossil record has resulted (and continues to result) in an exhaustive classification and description of many different morphologies from the Cambrian to the present, inferring evolutionary patterns and relationships between groups (De Wever et al., 2001; Afanasieva et al., 2005; Caridorit et al., 2017; O'Dogherty et al. *in prep.*). Also, from sediments their ecology has been inferred, providing a detailed distribution in the world oceans for these morphologies (Boltovskoy and Correa, 2016a; Boltovskoy, 2017). Describing molecular clades and combining it with the extensive morphological knowledge brings a new perspective to the understanding of radiolarian evolution and diversity. Awareness of the limitations of each technique is a must, yet integrating knowledge helps building robust hypotheses and therefore improve overall understanding. Every study should be considered as a hypothesis based on the data considered therein. Especially within the vast diversity of protists, where species concepts remain an open question (e.g.: Caron, 2013; Caron and Hu, 2018).

In recent years, molecular tools have brought a new perspective to radiolarian classification and evolution, by probing congruencies and inconsistencies between traditional taxonomy and molecular-based phylogenies. Applying an integrative morpho-molecular classification has also revolutionized other groups of protists such as Foraminifera (Pawlowski et al., 2013), Phaeodaria (Nakamura et al., 2015) or Tintinnida (Bachy et al., 2012) among others, showing the strength of combining different approaches. The first molecular phylogenies of Radiolaria proved the monophyly of the Polycystines and that of the Acantharia with 18S rDNA, yet relationships between these groups were not proven (Amaral Zettler et al., 1997). Soon afterwards these relationships between Acantharia and Polycystines were resolved with a more exhaustive radiolarian representation (López-García et al., 2002). Ever since many different works have contributed to Radiolarian phylogeny and classification, providing a continued basis for further research.

At a finer taxonomic scale, Decelle et al. (2012) showed the importance of the central junction in Acantharia classification. This allowed to explore their evolution beyond morphological features, since this group lacks a fossil record and the direct observation of their past is thus not possible. Regarding Collodaria, Biard et al. (2015) inferred their life cycle by detecting solitary specimens among colonial clades. In this doctoral work I have provided a new scheme in Nassellaria and Spumellaria classification partially agreeing with the traditional taxonomy. The reduction of the initial spicular system to a lower taxonomic rank (Chapter 1), allowed the placement of ambiguous fossils as ancestors or connections between and within such Polycystines groups. By integrating these results with previous molecular and morphological results, we have finally proposed a comprehensive framework for Radiolaria classification (Chapter 2).

2. A new framework for the evolution and classification of Radiolaria

Ribosomal DNA phylogenies have shown that these genes are not the best phylogenetic markers to study relationships at the Radiolaria-Retaria level (Fig.1). Due to the fast-evolving nature of the rDNA of Foraminifera (Moreira et al., 2007) and that of Nassellaria (Chapter 2) previous classifications and phylogenies have given contrasted topologies (Ishitani et al., 2011; Krabberød et al., 2011; Yuasa and Takahashi, 2016). Multigene and multiprotein phylogenies have contributed to a better understanding of their relationships, yet the few specimens included and the biased number of genes for each specimen, avoid a deep interpretation of their relationships (Burki et al., 2013; Krabberød et al., 2017; Cavalier-Smith et al., 2018). Besides, their long branches in phylogenetic analysis of the rDNA are also seen in multigene and multiprotein phylogenies producing an unsteady perception of their classification. Therefore, the monophyletic or polyphyletic nature of Radiolaria and its relationships with Foraminifera remains a mystery difficult to be solved. Despite discrepancies and poorly resolved phylogenies, most of the classifications consider Radiolaria and Foraminifera at the same taxonomic level (Cavalier-Smith et al., 2018; Adl et al., 2019). Such distinction looks more likely based in the early evolution of Radiolaria proposed in this manuscript, where these two groups diversified independently in the plankton and in the benthos respectively. Also, the differences in the endoskeleton of Radiolaria and the test or agglutinated wall of Foraminifera may support their independent evolution. Yet, attempting to trace their hypothetical early evolution and their classification with the data available so far is closer to a tale than a plausible hypothesis.

The radiolarian integrative morpho-molecular framework reviewed and proposed in this manuscript (Chapter 2, Fig. 1, page 98; in agreement with recent multigene and multiprotein studies: Cavalier-Smith et al., 2018; Adl et al., 2019) allows a better understanding of the molecular diversity and their relationships. Acantharia, with strontium sulphate skeleton, as sister clade of the Polycystines, with opaline silica skeleton, and an unexplored group (yet environmentally important; RAD-B in Chapter 3.2) basal to the polycystines with so far only one species described (Taxopodida: *Sticholonche zanclea*). The three families of Collodaria are considered within Nassellaria due to their strong phylogenetic support, as seen in Chapter 2 and supported by, hitherto, all the other studies that include them (e.g.: Ishitani et al., 2011; Krabberød et al., 2011, 2017; Yuasa and Takahashi, 2016). This consideration suggests a drastic morphological innovation, in which Collodaria loses the heteropolar symmetry towards a decrease in silicified structures, as suggested by Lazarus et al. (2009) from the Cenozoic. Other groups have also seen a dramatic morphological innovation, such as Spongodiscoidea within Hexastyloidea (Chapter 1.2), or clades E and F within Acantharia (Decelle et al., 2012b), corresponding to the most important symbiotic clades within Radiolaria. Another remarkable change in the taxonomy of Radiolaria is the polyphyly of the Entactinaria order, as seen in Chapter 2 and in Nakamura et al. (*in prep*; Annexes). Most of the previously considered entactinarian families fall now in the description of Spumellaria, as Polycystines with spherical or radial symmetry and concentric structures. However, some of

these families previously considered within Entactinaria, have shown in chapter 3.2 specific environmental preferences towards mesopelagic water masses. Therefore, due to their different distribution patterns, I understand the advantages on the use of the Entactinaria group from a paleoenvironmental or bioestratigraphic point of view. For the same reason, and for the sake of convenience, Collodaria may refer to colonial Radiolaria.

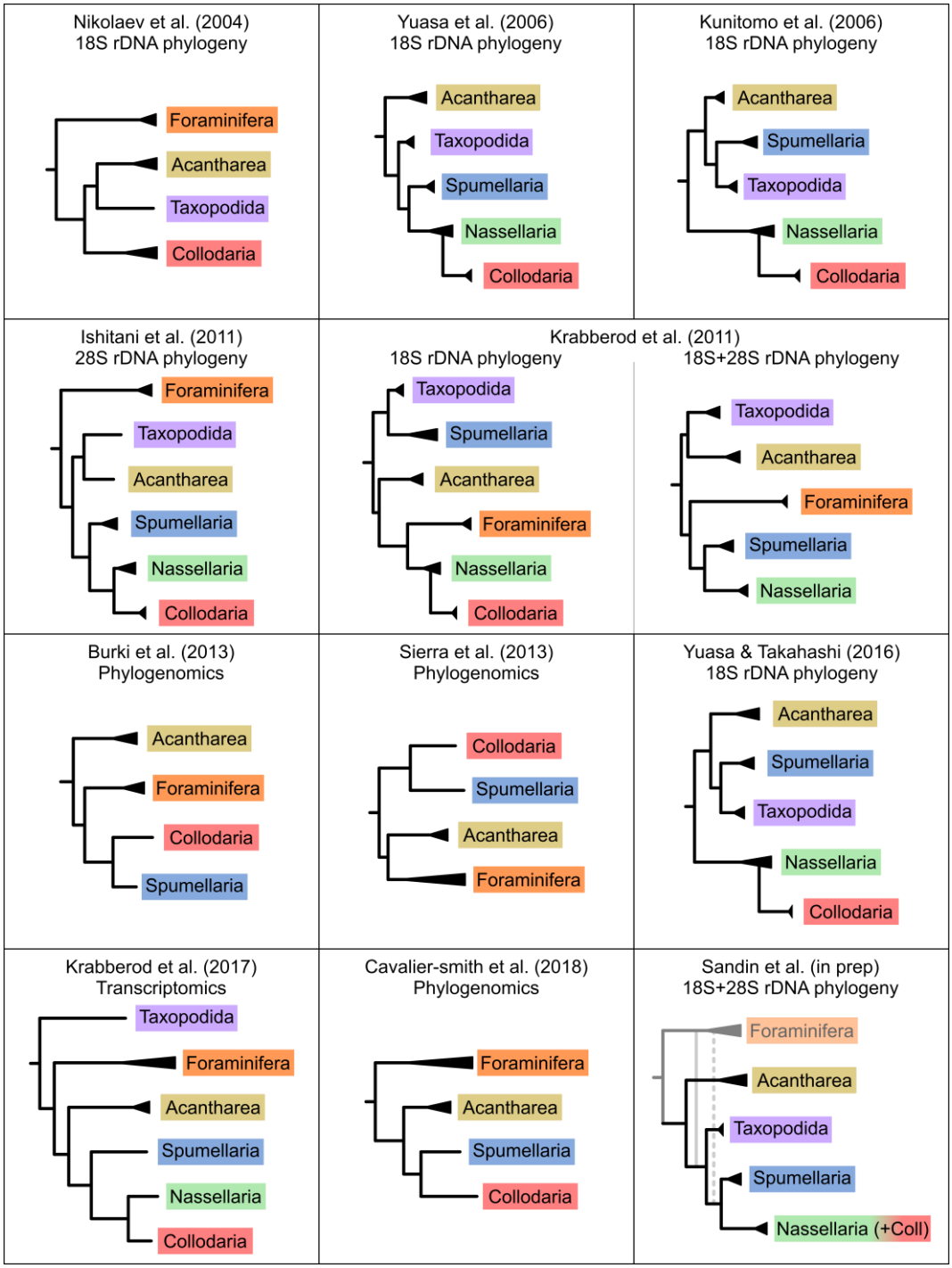


Figure 1. Schematic representation of phylogenetic trees of Radiolaria/Retaria inferred among different studies. Below the citation it is mentioned the main marker/approach used to obtain the phylogenetic tree.

The great environmental diversity described in Chapter 2 holding an early branching positions showed in Chapter 3.2 a big representation in the pico- and micro-plankton of the mesopelagic and bathypelagic water masses. Similar molecular diversity have been speculated in Foraminifera to correspond to naked amoeboid shapes (see Monothalamids; Pawlowski et al., 2011, 2013). Besides, Nikolaev et al. (2004) have already suggested that *Sticholonche zanclea* might not be the only member of this clade or might represent a species complex. Along these lines, we have hypothesized that most of this environmental diversity might be represented by skeleton-less ameboid protists. Indeed, some amoeboid Heliozoa have been proposed as part of Radiolaria, based on ultrastructural features of the axopodial complex, such as Gymnosphaerida (Yabuki et al., 2012). In this regard, the diversity surrounding Rad-B most probably correspond to naked amoeboid protists in which only few groups develop a simple skeleton composed of spicules, such as group-IV where *Sticholonche zanclea* belongs to. Since the arrangements of these spicule is relatively simple, it is possible that a large cryptic diversity is hidden within Taxopodida. Further research implementing both morphological and molecular data must unveil this hidden diversity present at early-branching lineages.

The lack of skeleton for these environmental clades within Retaria, is not surprising if we understand their early evolution and the lack of possible candidate morphologies in the fossil record. Phylogenetic placement of morphological data in DNA based phylogenies (as detailed in Berger and Stamatakis, 2010) could potentially resolve the position of missing families and groups. Or even going beyond and carry out the placement of fossil families (thanks to the extensive morphological knowledge from the fossil record) to properly understand their evolution. So far, the biggest described diversity of Retaria possess a skeleton made of either calcium carbonate (Foraminifera), strontium sulphate (Acantharia) or opaline silica (polycystines and Taxopodida). Therefore, the preferred distribution of environmental clades towards meso- and bathypelagic environments (Chapter 3.2) agrees with the hypothesis established in Chapter 2, and supported by Chapter 3.2. Probably, Radiolaria developed the skeleton to avoid predation (Porter, 2011) in highly active regions such as surface waters, outcompeting those lacking the ability to biomineralize. Then they found shelter in the vast deep and open ocean, where encountering a prey is less probable. Building of the skeleton supposed a big success for Radiolaria in the early Palaeozoic, that allowed them to diversify fast and in a big variety of forms (Chapter 2, Fig. 2, page 101), both Acantharia and Polycystines.

The last evolutionary innovation in Radiolaria was the establishment of the symbiosis back in the Mid Mesozoic with the dinoflagellates that we know today as *Brandtodinium* (Probert et al., 2014) and *Gymnoxanthellae* (Yuasa et al., 2016). Diatoms appeared at ~186 Ma and became an important competitor for Silica (Conley et al., 2017; Hendry et al., 2018; Lewitus et al., 2018). Besides, at that time, the oceans reached their most oligotrophic status (Cárdenas and Harries, 2010) and underwent a period of anoxia (Jenkyns, 2010). Probably the establishment of the symbiosis allowed the threaten heterotrophic Polycystines to thrive

despite the unfavourable conditions. Since it happened independently in Nassellaria at ~186 Ma (with the first diversification of Collodaria) and Spumellaria at ~187 Ma. Soon afterwards, the Acantharia established symbiosis with the Haptophyte *Phaeocystis* at ~157 Ma, as also suggested by Decelle et al. (2012a).

The evolutionary success of Radiolaria in the establishment of symbiosis is undoubted. Despite the low abundance of mixotrophic clades in the phylogenetic tree (Chapter 2, Fig. 1, page 98), their diversity in the oceans is comparable to that of their non-symbiotic relatives (Fig. 2). As all known Collodaria are photosymbiotic, such important evolutionary innovation could explain their great diversification, from Nassellaria, and their very divergent rDNA and completely different morphologies. That is also reflected in the higher morphological diversity of symbiotic Acantharia regarding non-symbiotic groups, or the high diversity of symbionts found in Spumellaria (dinoflagellates: Probert et al., 2014; Yuasa et al., 2016; Zhang et al., 2018; Cyanobacteria: Foster et al., 2006a, 2006b; Yuasa et al., 2012, Prasinophytes: Gast and Caron, 2001; Haptophytes: Anderson, 1983; Yuasa et al., 2019). Looking for specific spatial correlations between described symbionts and hosts of Radiolaria, could help elucidate specific relationships and improve the understanding of the planktonic symbiosis. Although the free-living state of the dinoflagellate *Symbiodinium*, symbiont of corals, have been shown to be abundant (Decelle et al., 2018). And due to the characteristic association in pelagic ecosystems, where encountering is a matter of abundance, it is expected to find Polycystines symbionts in high relative abundances. However, very little is known about the free-living state of the relatives of *Symbiodinium* hosted by the Polycystines or Foraminifera in the plankton, suggesting interesting future perspectives in the understanding of the symbiotic relationships.

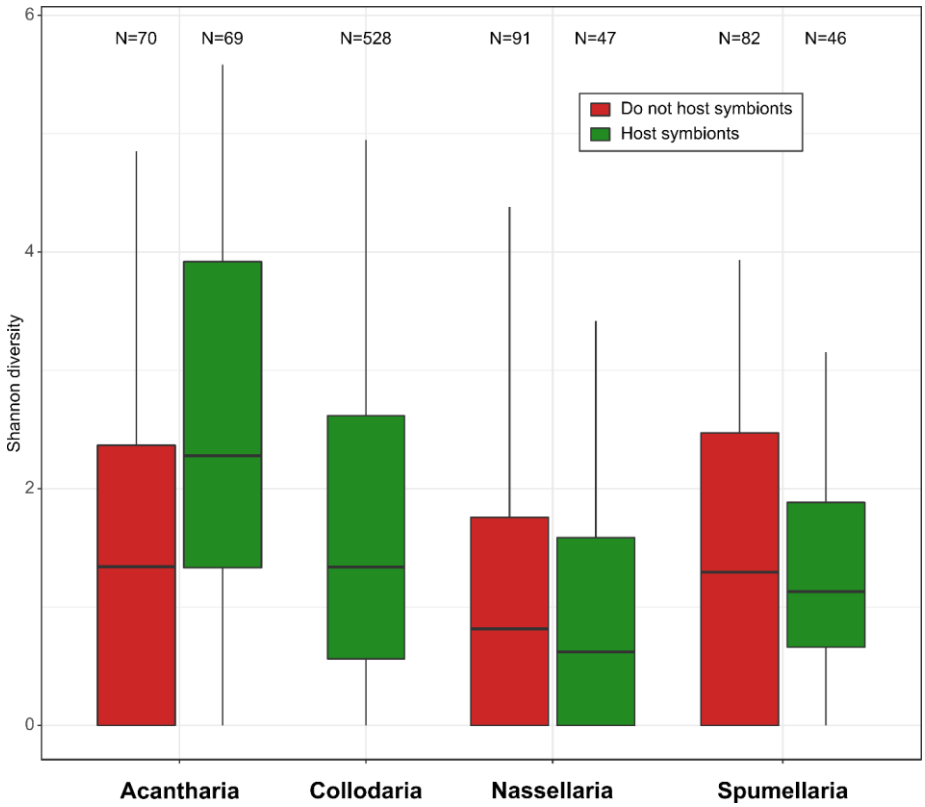


Figure 2. Shannon diversity of Hellinger transformed abundances for the main described groups of Radiolaria between symbiotic and non-symbiotic groups from TARA oceans V9 rDNA dataset. All depth and size fractions are considered. Only ASVs associated to morphologically described diversity and clear patterns of hosting/not-hosting symbionts were considered to avoid ambiguity. Note Collodaria is separated from Nassellaria for historical and practical reasons.

3. Approaching a better understanding of the molecular environmental diversity

The symbiotic association showed both an evolutionary success and an ecological advantage (as seen in Chapter 3.2). Surface waters are dominated by Acantharia in the small size fraction and Collodaria in the big size fraction, contributing about the 60% and 80% respectively to the radiolarian community based on sequence abundance. In the first case probably due to their ability to modify the symbionts metabolism, whereby Acantharia exploit their symbionts in its profit (Decelle et al., 2019). Whereas in the second case probably due to the ability to form colonies, in which the large organic matter that constitutes the colony allows them a better use and capture of the nutrients (Swanberg et al., 1986). This life mode of Collodaria allowed them to thrive in oligotrophic environments where they have been shown to be very abundant (Biard et al., 2017; Faure et al., 2019). In contrast, in the dark ocean the environmental clade Rad-B starts taking more importance as well as other non-symbiotic spumellarian clades. It is also important to highlight that in the biogeography described in Chapter 3.2, Nassellaria is always showing little importance, despite the great molecular diversity described in Chapter 2 (15 clades, +3 Collodaria and 1 Orosphaeridae) and being the taxonomic entity gathering the most extant morpho-species described (Fig. 3). Besides Collodaria is always showing a remarkable richness of ASVs, what is unexpected based on the number of morphologically described species.

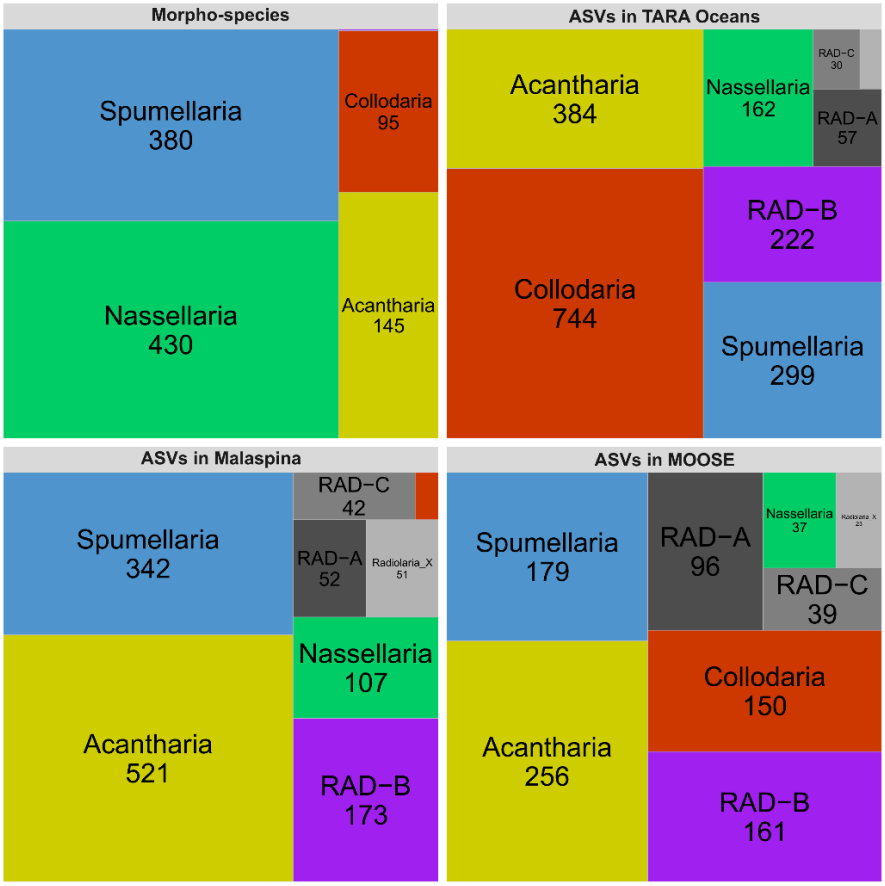


Figure 3. Relative proportion of the approximate number of described morpho-species (from Suzuki and Not. 2015) and total richness of ASVs in TARA Oceans (V9 rDNA, global expedition, mostly euphotic layer), Malaspina (V4 rDNA, global expedition, vertical profile) and MOOSE-GE (V4 rDNA, regional expedition, vertical profile) for each different taxonomic group described in Chapter 2. All size fractions and depth are included. Note only one morphologically described species stands for Rad-B (*Sticholonche zanclea*).

The coupling between high-throughput sequencing results and morphospecies count is a non-conclusive debate since it seems to be taxa related as it shows some correlation in freshwater ciliates (Santoferrara et al., 2016) but not with not in tintinnids ciliates, (Bachy et al., 2013) or Haptophytes (Egge et al., 2013). On the contrary, ASVs reads have been shown to be correlated with biomass (Egge et al., 2013; Giner et al., 2016; Pitsch et al., 2019), and the rDNA copy number correlates with cell size (Zhu et al., 2005; Godhe et al., 2008; Biard et al., 2017). Besides, intracellular diversity and artefacts produced by sequencing errors may also inflate ASVs numbers (Bachy et al., 2013; Decelle et al., 2014; Chapter 3.1). As discussed in Chapter 3.1 and in Decelle et al. (2014) the number and size of the nucleus in Radiolaria have been hypothesized to be related to intracellular variability, and Suzuki et al. (2009) have shown big differences in the nucleus of Radiolaria among taxa. Therefore, it is not surprising to find more ASVs in Collodaria and Acantharia (Fig. 3), that have many nuclei, compared to other groups, such as Nassellaria and Spumellaria, that have only one and small nucleus.

Environmental studies implement many different analytical filters and thresholds in order to get rid of errors and low abundant ASVs and avoid analysing false positives. Also,

different algorithms help the clustering into ASVs of most-probable errors in sequencing, like the so-called *fastidious* option in swarm (Mahé et al., 2015), reducing over-representation of low abundant OTUs. Errors are not only happening in high-throughput sequencing, Sanger sequencing has many errors towards the end of the sequence (and in a lesser extent at the beginning). This may affect reference sequences to whom the metabarcodes are compared against. The hypervariable region V9 is close to the end of universal primers, and many reference sequences may have such errors (and one error in such small sequence length -~130 bp- has a strong impact) leading to misidentification of metabarcodes when filtering by similarity. Although reference sequences normally are a consensus of two primers, decreasing the error probability. On the other side, intracellular (or intraspecific) genetic diversity in theory should be easy to detect in metaB surveys. Since it is expected to have very similar identity clusters or because it should correlate among the samples. This is why biological samples should be sequenced in duplicates (or triplicates better) for every metabarcoding survey (Prosser, 2010: Replicate or lie). Although errors in sequencing and intracellular diversity is simply one of the many biases of metabarcoding surveys (Fig. 4).

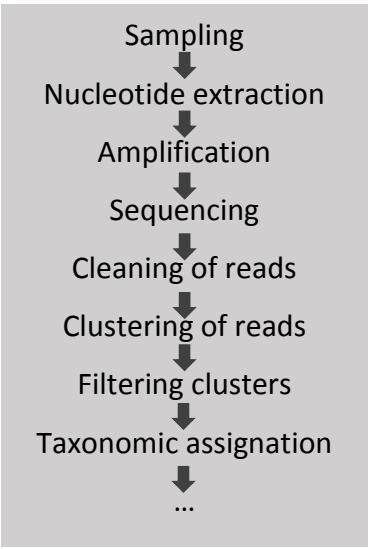


Fig 4. Schematic representation of main steps in the metabarcoding analysis.

Biases in ASV counts can happen at many different steps from the sampling to the final diversity analysis of the ASVs (Fig. 4). And especially when comparing two closely related organisms (Collodaria and Nassellaria) with such different lifestyles, where one tends to form colonies and the other correspond to small cells, failing in the first step of the sampling where it is required different approaches for their collection. As seen in Chapter 3.2, the relative contribution of Nassellaria to the total radiolarian community was higher when the DNA was extracted chemically than when it was extracted mechanically. Another example, DNA amplification success in Foraminifera have shown to be different with the taxa (Weiner et al., 2016). During the process of barcoding Nassellaria in Chapter 1.1, I found a very low success ratio in the amplification of the DNA compared to that obtained for Spumellaria in the following chapter (1.2). These differences raise further questions on the rDNA structure, the copy number, the physiology or the anatomy of nassellarian cells. Therefore, the contrasted results obtained by metabarcoding versus morphological studies should not be considered as misleading or wrong (as for any other study) but should be interpreted as complementary, setting the bases for further research. ASVs were never intended to be related to the species concept (by their origin as Operational Taxonomic Units, clustering sequences by similarity), and far less to the morpho-species definition. As here discussed, differences in ASV number and relative abundance might be a result of the high evolutionary and ecological success of Collodaria and Acantharia, reflected in their higher biomass regarding other radiolarians.

New sequencing technologies are arriving with the potential for high throughput sequencing of long nucleotide fragments, from low concentrated samples, nearly 0 errors and real time. The combined robustness of circular consensus sequencing of PacBIO along with the long and portability high throughput sequencing of MinION (Goodwin et al., 2016; Levy and Myers, 2016) could set a new standpoint for barcoding and metabarcoding studies of planktonic communities. Problems encountered during the building of the morpho-molecular reference framework, in particular for Nassellaria, could be reduce (or in theory overcome) by performing the single molecule real time sequencing (SMRT) of PacBIO. Obtaining several rDNA sequences from the different copies, reducing technical effort and creating a reference database with intracellular diversity resolution. Which could also improve the understanding of the rDNA structure in the different groups and therefore the metabarcoding output. Other issues such as dealing with relative abundances or number of ASVs (either very low or outstanding) might also be solved since these methods do not require amplification, reading single molecules and therefore absolute values (as discussed in chapter 3.1). Shortening the flowchart of figure 4 may lead to complementary results by removing important biases. Although, biases will always be there such as differences in rDNA copy number, and these techniques are still in its infancy, yet promising infancy.

Despite all the limitations and a-priori misleading information obtained by metabarcoding surveys, the final output can be easily link to morphological data. Here we have expanded the knowledge on radiolarian diversity, by having access to the unknown part. All along this manuscript we have provided a new perspective in the evolution and diversity of Radiolaria by combining different techniques and applying novel approaches. Yet, “understanding” is a long word, and longer is the way to it.

Acknowledgments

First of all, I would like to show my deepest gratitude to Fabrice, to have shown me this wonderful little big world of Radiolaria. Thanks for your time, your advices, your comprehension, your honesty and pragmatism, your motivation and your interest and will to know more. Thanks to have trusted in a motivated guy and supported me. Thanks to have introduced me in the molecular biology and to have guided me all along these three years. It was really an honour and a pleasure to share this time all over the world. I hope it won't be the last time!

A very big part of this work was undoubtedly thanks also to Noritoshi Suzuki. Thanks for all the detailed feedback and contribution about any radiolarian aspect. Thanks for the time in Japan while waiting for the typhoon to move away and the endless conversations, I learnt a lot! Thanks a lot for all the explanatory figures and especially for your interest and patient in all the collaborations.

I'm very grateful for the excellent committee I had, Lucie Bitner, Javier del Campo, Daniel Ritcher and Frederique Viard. It was a real pleasure to have all your experienced feedback and advises. And specially to have showed me the strongest and weakest points of my work from a very critical point of view. I'll definitely keep you as future references :)

I would like also to thank the Jury Linda Amaral-Zettler, Fabien Burki, John Dolan, Spela Gorican, Nathalie Simon and Cedric Berney. Thank you to have accepted to review and examine this work. It is an honour to count with you.

Of course, deserve especial attention the non-official committee Charles Bachy, Cedric Berney and Nicolas Henry. Thank you for all the patience and advises all along these three years. You have saved me many times with your knowledge, experience and humble opinion. Although most of the times I had more questions than answers after our conversations! I'm seriously saving every sentence and beer shared as wisdom.

Que dire de mes mentors de laboratoire ? Sarah Romac et Estelle Bigeard, merci simplement pour toute votre patience, vos heures et vos conseils pour faire au moins une séquence qui sorte! Surtout grâce à Sarah pour faire que toute la magie devienne réalité.

Also, I'm very thankful to Elsa Gadoin, it was a real pleasure to work with you. We have shared the joys and troubles of these little protists, the laughs and tears and the sweetness of the summit from a single sequence! I wish you all the best in your future steps.

Luis O'Dogherty, Špela Goričan, Peter Baumgartner, Claudia Baumgartner-Mora and Taniel Danelian thanks a lot for all the valuable feedback on the fossil record, and all your contribution to this manuscript. It was a real pleasure to have met you, such a good palaeontologists and better persons.

It was also a pleasure to work with Yasuhide Nakamura, Emilie Villar and Natalia Llopis. I have learnt and enjoyed a lot the time and work shared. Let's hope is not the last!

Bien sûr je ne peux pas oublier Florence Le Gall, Fabienne Rigaut-Jalabert, Priscillia Gourvil et toutes les personnes formidables d'ECOMAP. Vous mettez un sourire tous les jours qui rend le travail dans ces couloirs inoubliable. N'oubliez pas de venir me rendre visite où que je sois !

Also many thanks to Natalia Llopis and Magali Lescot to have made of MOOSE a wonderful experience! To Andres Gutierrez, Moira Decima and Adriana Lopes to have showed me all the good (and bad) sides of oceanography in New Zealand; and to Ramon Massana and Aleix Obiol to have hosted me in Barcelona. I can't neither forget Alexia, Florian, Michelle, Ola, Alicia, Parra, Nuria, Meri, … Thanks for all the good moments besides the boat or the lab!

Many thanks to all the people that I had the pleasure to meet before this thesis, Consolacion Fernandez, Jose Luis Acuña, Rick Officer, Jean-Luc Jung, Angel Manteca, Eva Garcia, Jaisel Borrel, Sara Fernandez. In one way or another you have helped me and contributed to my personal development. Muchas gracias también a Sonia, Carlinos, Axa, Richy y todos los demas.

Johan, Nico, Julie, Solenne, Yann, Pierre-Yves, Tanguy, … Merci beaucoup pour toute la peau laissée dans les rochers de Bretagne ! Les falaises des Asturies vous attendent.

Merci aussi aux Roscovittes temporaires ! Laure, Alexis, Thibaut, Marine, Eloïse, Damien, Louise, Mariana, Jeremy, Mariarita, Axelle, Charles, Lydia, Laura, Ulysse, Laura, Ewen, Ewan, David, Martin, Florian, Theophile, Hugo, … Et tout les eutres Brestois aussi ! Natalia, Marc et Anaïs. Merci pour tout le temps partagé à l'intérieur et à l'extérieur de la station. Surtout dehors !! Merci Julie pour les soirées sushi, les bières et les grimpes ! Merci Laetitia pour toutes les soirées et tous les jours à Berrien ! Merci également à toutes les personnes de la station et à Rosko pour avoir fait de ces trois années une expérience merveilleuse. Va certainement être difficile de sortir de Rosko. Ou peut-être on peut faire un autre endroit aussi génial !

Y ya que estamos por agradecer, que sería sin los amigos de Gijón, de biología, de Faro, de la montaña, de en general… Dani, Carlos, Jandro, Belén, Corti, Guillo, Croke y Katia (¡Felicidades por la boda por cierto!), Carlos, Nere, Irene, Aaron, Jana, Luna, Pata, Llata, Tuki, Jess, Ana, Afri, Belén, Lucía, Moritz, Isa, Pela, Sergi, Suco, Nico, Pepe, Gerar, Aitor, Jorge, Ale, … Y voy a poner puntos suspensivos porque ye imposible acordarse de todos los buenos momentos. Mil gracias por las infinitas rayadas mentales, los bancos, de todas las cerves y ciegos astrales, los viajes, las cordadas o sin más por estar ahí en ese preciso momento.

Deserve especial thanks Daniel and Nico, to have shared all those wonderful times in Roscoff, Fontainebleau, Barcelona, Rennes, …, everywhere you are just make it great! Thanks a lot Daniel for the last feedback in many parts of the manuscript. And specially thanks to Nico, who have stood me, suffered me and supported me quite patiently! Such cool climbing buddies!

A great part of this work was certainly thanks to all the support from Laure. I'm seriously very grateful (and lucky) I have met you. Thanks also for all the translations, paperwork and everything you have done selflessly. You neither need all the best, you just fight for it and you earn it! From the beginning until the end, thanks a lot for everything, once again.

Y sobretodo quiero agradecer y dedicar esta tesis a mis padres, a mi hermano y a mi abuelo. Quienes siempre me apoyaron en cada paso que dí y decisión que tomé. Me disteis todo al alcance de vuestras manos para poder llegar hasta aquí. Es un orgullo pertenecer a esta familia y seguir vuestros pasos. Desde luego, un ejemplo a seguir.

References

A

Abelmann, A., Nimmergut, A., 2005. Radiolarians in the Sea of Okhotsk and their ecological implication for paleoenvironmental reconstructions. Deep. Res. Part II Top. Stud. Oceanogr. 52, 2302–2331.

Adl, S.M., Bass, D., Lane, C.E., Lukeš, J., Schoch, C.L., Smirnov, A., Agatha, S., Berney, C., Brown, M.W., Burki, F., Cárdenas, P., Čepička, I., Chistyakova, L., del Campo, J., Dunthorn, M., Edvardsen, B., Eglit, Y., Guillou, L., Hampl, V., Heiss, A.A., Hoppenrath, M., James, T.Y., Karnkowska, A., Karpov, S., Kim, E., Kolisko, M., Kudryavtsev, A., Lahr, D.J.G., Lara, E., Le Gall, L., Lynn, D.H., Mann, D.G., Massana, R., Mitchell, E.A.D., Morrow, C., Park, J.S., Pawlowski, J.W., Powell, M.J., Richter, D.J., Rueckert, S., Shadwick, L., Shimano, S., Spiegel, F.W., Torruella, G., Youssef, N., Zlatogursky, V., Zhang, Q., 2019. Revisions to the Classification, Nomenclature, and Diversity of Eukaryotes. J. Eukaryot. Microbiol. 66, 4–119.

Afanasieva, M.S., Amon, E.O., 2006. Biotic crises and stages of radiolarian evolution in the Phanerozoic. Paleontol. J. 40, S453–S467.

Afanasieva, M.S., Amon, E.O., Agarkov, Y.V., Boltovskoy, D.S., 2005. Radiolarians in the Geological Record. Paleontol. J. 39, 135–392.

Aitchison, J., Suzuki, N., Caridroit, M., Danelian, T., Noble, P., 2017. Paleozoic radiolarian biostratigraphy, in: Danelian, Taniel, Caridroit, Martial, Noble, Paula, Aitchison, J.C. (Eds.), Catalogue of Paleozoic Radiolarian Genera. Geodiversitas, pp. 503–531.

Aitchison, J.C., Suzuki, N., O'Dogherty, L., 2017. Inventory of Paleozoic radiolarian species (1880-2016). Geodiversitas 39, 533–637.

Alberti, A., Poulain, J., Engelen, S., Labadie, K., Romac, S., Ferrera, I., Albini, G., Aury, J.M., Belser, C., Bertrand, A., Cruaud, C., Da Silva, C., Dossat, C., Gavory, F., Gas, S., Guy, J., Haquelle, M., Jacoby, E., Jaillon, O., Lemainque, A., Pelletier, E., Samson, G., Wessner, M., Acinas, S.G., Royo-Llonch, M., Cornejo-Castillo, F.M., Logares, R., Fernández-Gómez, B., Bowler, C., Cochrane, G., Amid, C., Hoopen, P. Ten, De Vargas, C., Grimsley, N., Desgranges, E., Kandels-Lewis, S., Ogata, H., Poulton, N., Sieracki, M.E., Stepanauskas, R., Sullivan, M.B., Brum, J.R., Duhaime, M.B., Poulos, B.T., Hurwitz, B.L., Pesant, S., Karsenti, E., Wincker, P., Bazire, P., Beluche, O., Bertrand, L., Besnard-Gonnet, M., Bordelais, I., Boutard, M., Dubois, M., Dumont, C., Ettedgui, E., Fernandez, P., Garcia, E., Aiach, N.G., Guerin, T., Hamon, C., Brun, E., Lebled, S., Lenoble, P., Louesse, C., Mahieu, E., Mairey, B., Martins, N., Megret, C., Milani, C., Muanga, J., Orvain, C., Payen, E., Perroud, P., Petit, E., Robert, D., Ronsin, M., Vacherie, B., Bork, P., Boss, E., Follows, M., Gorsky, G., Hingamp, P., Iudicone, D., Karp-Boss, L., Not, F., Raes, J., Sardet, C., Speich, S., Stemmann, L., Sunagawa, S., 2017. Viral to metazoan marine plankton nucleotide sequences from the Tara Oceans expedition. Sci. Data 4, 1–20.

Alcaide, M., Scordato, E.S.C., Price, T.D., Irwin, D.E., 2014. Genomic divergence in a ring species complex. Nature 511, 83–85.

Amaral-zettler, L.A., Mccliment, E.A., Ducklow, H.W., Huse, S.M., 2009. A Method for Studying Protistan Diversity Using Massively Parallel Sequencing of V9 Hypervariable Regions of Small-Subunit Ribosomal RNA Genes. PLoS One 4, e6372.

Amaral Zettler, L., Sogin, M.L., Caron, D.A., 1997. Phylogenetic relationships between the Acantharea and the Polycystinea: A molecular perspective on Haeckel's Radiolaria. Proc. Natl. Acad. Sci. 94, 11411–11416.

Anderson, O.R., 1993. The trophic role of planktonic foraminifera and radiolaria. Mar. Microb. Food Webs 7, 31–51.

Anderson, O.R., 1983. Radiolaria.

Anderson, O.R., Bennett, P., Bryan, M., 1989. Experimental and observational studies of radiolarian physiological ecology: 3. Effects of temperature, salinity and light intensity on the growth and survival of Spongaster tetras tetras maintained in laboratory culture. Mar. Micropaleontol. 14, 275–282.

Ando, H., Kunitomo, Y., Sarashina, I., Iijima, M., Endo, K., Sashida, K., 2009. Intraspecific variations in the ITS region of Recent radiolarians. Earth Evol. Sci. 3, 37–44.

Armbrust,E.V., Wieneke,S.G., McGillicuddy,D.J. and Olson,R.J. 2004. Phylogenetic diversity of nanoplankton in Sargasso Sea Eddies. Unpublished.

B

Bachy, C., Dolan, J.R., López-García, P., Deschamps, P., Moreira, D., 2013. Accuracy of protist diversity assessments: Morphology compared with cloning and direct pyrosequencing of 18S rRNA genes and ITS regions using the conspicuous tintinnid ciliates as a case study. ISME J. 7, 244–255.

Bachy, C., Gómez, F., López-García, P., Dolan, J.R., Moreira, D., 2012. Molecular Phylogeny of Tintinnid Ciliates (Tintinnida, Ciliophora). Protist 163, 873–887.

Bambach, R.K., Knoll, A.H., Wang, S.C., 2004. Origination, extinction, and mass depletions of marine diversity. Paleobiology 30, 522–542.

Berger SA, Stamatakis A., 2010. Accuracy of morphology-based phylogenetic fossil placementunder maximum likelihood. In:ACS/IEEE International Conference on Computer Systems andApplications, AICCSA 2010.Hammamet: IEEE, 1–8.

Berney, C., Pawlowski, J., 2006. A molecular time-scale for eukaryote evolution recalibrated with the continuous microfossil record. Proc. R. Soc. B Biol. Sci. 273, 1867–1872.

Betts, H.C., Puttick, M.N., Clark, J.W., Williams, T.A., Donoghue, P.C.J., Pisani, D., 2018. Integrated genomic and fossil evidence illuminates life's early evolution and eukaryote origin. Nat. Ecol. Evol. 2, 1556–1562.

Biard, T., Bigeard, E., Audic, S., Poulain, J., Stemmann, L., Not, F., 2017. Biogeography and diversity of Collodaria (Radiolaria) in the global ocean. Nat. Publ. Gr. 1–42.

Biard, T., Pillet, L., Decelle, J., Poirier, C., Suzuki, N., Not, F., 2015. Towards an Integrative Morpho-molecular Classification of The Collodaria (Polycystinea, Radiolaria). Protist 166, 374–388.

Biard, T., Stemmann, L., Picheral, M., Mayot, N., Vandromme, P., Hauss, H., Gorsky, G., Guidi, L., Kiko, R., Not, F., 2016. In situ imaging reveals the biomass of giant protists in the global ocean. Nature 000, 1–16.

Bickford, D., Lohman, D.J., Sodhi, N.S., Ng, P.K.L., Meier, R., Winker, K., Ingram, K.K., Das, I., 2007. Cryptic species as a window on diversity and conservation. Trends Ecol. Evol. 22, 148–155.

Bjorbækmo, M.F.M., Evenstad, A., Røsæg, L.L., Krabberød, A.K., Logares, R., n.d. The planktonic protist interactome: emerging trends after a century of research. bioRxiv.

Blaxter, M., Mann, J., Chapman, T., Thomas, F., Whitton, C., Floyd, R., Abebe, E., 2005. Defining operational taxonomic units using DNA barcode data. Philos. Trans. R. Soc. B Biol. Sci. 360, 1935–1943.

Boltovskoy, D., 2017. Vertical distribution patterns of Radiolaria Polycystina (Protista) in the World Ocean: living ranges, isothermal submersion and settling shells. J. Plankton Res. 39, 330–349.

Boltovskoy, D., Correa, N., 2016a. Biogeography of Radiolaria Polycystina (Protista) in the World Ocean. Prog. Oceanogr. 149, 82–105.

Boltovskoy, D., Correa, N., 2016b. Planktonic equatorial diversity troughs: fact or artifact? Latitudinal diversity gradients in Radiolaria. Ecology.

Boltovskoy, D., Kling, S.A., Takahashi, K., Bjørklund, K., 2010. World atlas of distribution of recent Polycystina (Radiolaria). Palaeontol. Electron. 13, 230.

Boltovskoy, D., Kogan, M., Alder, V.A., Mianzan, H., 2003. First record of a brackish radiolarian (Polycystina): Lophophaena rioplatensis n. sp. in the Río de la Plata estuary. J. Plankton Res. 25, 1551–1559.

Bosak, T., Lahr, D.J.G., Pruss, S.B., Macdonald, F.A., Gooday, A.J., Dalton, L., Matys, E.D., 2012. Possible early foraminiferans in post-sturtian (716-635 Ma) cap carbonates. Geology 40, 67–70.

Bown, P.R., Lees, J.A., Young, J.R., 2004. Calcareous nannoplankton evolution and diversity through time, in: Thierstein, H.R., Young, J.R. (Eds.), Coccolithophores. Springer Berlin Heidelberg, pp. 81–508.

Brisbin, M.M., Conover, A.E., Mitarai, S., 2019. Hydrothermal activity and water mass composition shape microbial eukaryote diversity and biogeography in the Okinawa Trough. bioRxiv 1–26.

Brocks, J.J., Jarrett, A.J.M., Sirantoine, E., Hallmann, C., Hoshino, Y., Liyanage, T., 2017. The rise of algae in Cryogenian oceans and the emergence of animals. Nature 548, 578–581.

Bromham, L., Penny, D., 2003. The Modern Molecular Clock. Nat. Rev. Genet. 4, 216–224.

Burki, F., Corradi, N., Sierra, R., Pawlowski, J., Meyer, G.R., Abbott, C.L., Keeling, P.J., 2013. Phylogenomics of the intracellular parasite mikrocytos mackini reveals evidence for a mitosome in rhizaria. Curr. Biol. 23, 1541–1547.

Burki, F., Kaplan, M., Tikhonenkov, D. V., Zlatogursky, V., Minh, B.Q., Radaykina, L. V., Smirnov, A., Mylnikov, A.P., Keeling, P.J., 2016. Untangling the early diversification of eukaryotes: a phylogenomic study of the evolutionary origins of Centrohelida, Haptophyta and Cryptista. Proc. R. Soc. B Biol. Sci. 283, 20152802.

Burki, F., Keeling, P.J., 2014. Rhizaria. Curr. Biol. 24, 103–107.

Burki, F., Keeling, P.J., Kudryavtsev, A., Matz, M. V., Aglyamova, G. V., Bulman, S., Fiers, M., Pawlowski, J., 2010. Evolution of Rhizaria: New insights from phylogenomic analysis of uncultivated protists. BMC Evol. Biol. 10.

Butterfield, N.J., 2015. Early evolution of the Eukaryota. Palaeontology 58, 5–17.

Butterfield, N.J., 1997. Plankton Ecology and the Proterozoic-Phanerozoic Transition. Paleobiology 23, 247–262.

C

Callahan, B.J., Mcmurdie, P.J., Holmes, S.P., 2017. Exact sequence variants should replace operational taxonomic units in marker-gene data analysis. ISME J. 1–5.

Callahan, B.J., Mcmurdie, P.J., Rosen, M.J., Han, A.W., Johnson, A.J., Holmes, S.P., 2016. DADA2: High resolution sample inference from amplicon data. Nat. Methods 13, 581–583.

Campbell, A.S., 1954. Part D, Protista 3. Protozoa (Chiefly Radiolaria and Tintinnina), in: Treatise on Invertebrate Paleontology.

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

Cárdenas, A.L., Harries, P.J., 2010. Effect of nutrient availability on marine origination rates throughout the Phanerozoic eon. Nat. Geosci. 3, 430–434.

Cardinal, S., Danforth, B.N., 2013. Bees diversified in the age of eudicots. Proc. R. Soc. B Biol. Sci. 280.

Caridorit, M., Danelian, T., O'Dogherty, L., Cuvelier, J., Aitchison, J.C., Pouille, L., Noble, P., Dumitrica, P., Suzuki, N., Kuwahara, K., Maletz, J., Feng, Q., 2017. An illustrated catalogue and revised classification of Paleozoic radiolarian genera, in: Danelian, T., Caridorit, M., Noble, P., Aitchison, J.C. (Eds.), Catalogue of Paleozoic Radiolarian Genera. Geodiversitas, pp. 363–417.

Caridroit, M., De Wever, P., Dumitrica, P., 1999. Un nouvel ordre, une nouvelle famille et un nouveau genre de Radiolaires du Paleozoique: Latentifistularia, Cauletellidae et Cauletella. Comptes Rendus l'Academie Sci. - Ser. IIa Sci. la Terre des Planetes 329, 603–608.

Caron, D.A., 2013. Towards a molecular taxonomy for protists: Benefits, risks, and applications in plankton ecology. J. Eukaryot. Microbiol. 60, 407–413.

Caron, D.A., Hu, S.K., 2018. Are We Overestimating Protistan Diversity in Nature? Trends Microbiol. xx, 1–9.

Caron, D.A., Michaels, A.F., Swanberg, N.R., Howse, F.A., 1995. Primary productivity by symbiont-bearing planktonic sarcodines (Acantharia, Radiolaria, Foraminifera) in surface waters near Bermuda. J. Plankton Res. 17, 103–129.

Cavalier-Smith, T., 2002. The phagotrophic origin of eukaryotes and phylogenetic classification on protozoa. Int. J. Syst. Evol. Microbiol. 52, 297–354.

Cavalier-Smith, T., 1999. Principles of protein and lipid targeting in secondary symbiogenesis: euglenoid, dinoflagellate, and sporozoan plastid origins and the eukaryote family tree. J. Eukaryot. Microbiol. 46, 347–366.

Cavalier-Smith, T., Chao, E.E., Lewis, R., 2018. Multigene phylogeny and cell evolution of chromist infrakingdom Rhizaria: contrasting cell organisation of sister phyla Cercozoa and Retaria. Protoplasma 255, 1517–1574.

Cavalier-Smith, T., Chao, E.E., Lewis, R., 2015. Multiple origins of Heliozoa from flagellate ancestors: New cryptist subphylum Corbihelia, superclass Corbistoma, and monophyly of Haptista, Cryptista, Hacrobia and Chromista. Mol. Phylogenet. Evol. 93, 331–362.

Cheng, Y.-N., 1986. Taxonomic studies on Upper Paleozoic Radiolaria. Spec. Publ. National Museum of Natural Science, Taiwan.

Claparede, E., Lachmann, J., 1859. Les infusoires et les Rhizopodes. Mem. l'institut Natl. Genev. 261–482.

Cohen, B.L., 2005. Not armour, but biomechanics, ecological opportunity and increased fecundity as keys to the origin and expansion of the mineralized benthic metazoan fauna. Biol. J. Linn. Soc. 85, 483–490.

Cohen, P.A., Macdonald, F.A., 2015. The Proterozoic Record of Eukaryotes. Paleobiology 41, 610–632.

Cohen, P.A., Schopf, J.W., Butterfield, N.J., Kudryavtsev, A.B., Macdonald, F.A., 2011. Phosphate biomineralization in mid-neoproterozoic protists. Geology 39, 539–542.

Conley, D.J., Frings, P.J., Fontorbe, G., Clymans, W., Stadmark, J., Hendry, K.R., Marron, A.O., De La Rocha, C.L., 2017. Biosilicification drives a decline of dissolved Si in the oceans through geologic time. Front. Mar. Sci. 4.

Corel, E., Lopez, P., Méheust, R., Bapteste, E., 2016. Network-Thinking: Graphs to Analyze Microbial Complexity and Evolution. Trends Microbiol. 24, 224–237.

Cronquist, A., 1978. Once again, what is a species? In: Knutson LV (ed) Biosystematics in agriculture. Alleheld Osmun, Montclair, NJ, pp 3–20

D

Darwin, C. 1859. On the origin of species by means of natural selection, or the preservation of favoured races in the struggle for life. London: Murray.

De Queiroz, K., 1998. The general lineage concept of species, and the process of speciation, in: Howard, D., Berlocher, S. (Eds.), Endless Forms: Species and Speciation. Oxford University Press, Oxford, pp. 57–75.

De Queiroz, K., 2007. Species concepts and species delimitation. Syst. Biol. 56, 879–886.

De Vargas, C., Audic, S., Henry, N., Decelle, J., Mahé, F., et al., 2015. Eukaryotic plankton diversity in the sunlit ocean. Science (80-.). 348, 1261605-1–12.

De Wever, P., 1982. Nassellaria (Radiolaires Polycystines) du Lias de Turquie. Rev. Micropaléontologie 244, 189–232.

De Wever, P., Dumitrica, P., Caulet, J.P., Nigrini, C., Caridroit, M., 2001. Radiolarians in the sedimentary record. Gordon & Breach Science Publishers.

De Wever, P., O'Dogherty, L., Caridroit, M., Dumitrica, P., Guex, J., Nigrini, C., Caulet, J.P., 2003. Diversity of radiolarian families through time. Bull. la Soc. Geol. Fr. 174, 453–469.

De Wever, P., O'Dogherty, L., Goričan, Š., 2006. The plankton turnover at the Permo-Triassic boundary, emphasis on radiolarians. Eclogae Geol. Helv. 99, S49–S62.

Decelle, J., Carradec, Q., Pochon, X., Henry, N., Romac, S., Mahe, F., Dunthorn, M., Kourlaiev, A., Voolstra, C.R., Wincker, P., de Vargas, C., 2018. Worldwide Occurrence and Activity of the Reef-Building Coral Symbiont Symbiodinium in the Open. Curr. Biol. 28, 1–9.

Decelle, J., Colin, S., Foster, R.A., 2015. Chapter 19. Photosymbiosis in Marine Planktonic Protists, in: Marine Protists. pp. 465–500.

Decelle, J., Martin, P., Paborstava, K., Pond, D.W., Tarling, G., 2013. Diversity, Ecology and Biogeochemistry of Cyst-Forming Acantharia (Radiolaria) in the Oceans. PLoS One 8.

Decelle, J., Probert, I., Bittner, L., Desdevises, Y., Colin, S., Vargas, C. de, Galí, M., Simó, R., Not, F., 2012a. An original mode of symbiosis in open ocean plankton. Proc. Natl. Acad. Sci. U. S. A. 109, 1–6.

Decelle, J., Romac, S., Sasaki, E., Not, F., Mahé, F., 2014. Intracellular Diversity of the V4 and V9 Regions of the 18S rRNA in Marine Protists (Radiolarians) Assessed by High-Throughput Sequencing. PLoS One 9

Decelle, J., Stryhanyuk, H., Gallet, B., Veronesi, G., Schmidt, M., Balzano, S., Marro, S., Uwizeye, C., Jouneau, P.H., Lupette, J., Jouhet, J., Maréchal, E., Schwab, Y., Schieber, N.L., Tucoulou, R., Richnow, H., Finazzi, G., Musat, N., 2019. Algal Remodeling in a Ubiquitous Planktonic Photosymbiosis. Curr. Biol. 29, 968-978.e4.

Decelle, J., Suzuki, N., Mahé, F., Vargas, C. De, Not, F., 2012b. Molecular Phylogeny and Morphological Evolution of the Acantharia (Radiolaria). Protist 163, 435–450.

Deflandre, G., 1953. Radiolaires fossiles, in: Grassé, P.P. (Ed.), Traité de Zoologie, Anatomie, Systématique, Biologie. p. vol. 1, part 2, pp. 389–436.

Del Campo, J., Kolisko, M., Boscaro, V., Santoferrara, L.F., Nenarokov, S., Massana, R., Guillou, L., Simpson, A., Berney, C., de Vargas, C., Brown, M.W., Keeling, P.J., Parfrey, L.W., 2018. EukRef: Phylogenetic curation of ribosomal RNA to enhance understanding of eukaryotic diversity and distribution. PLoS Biol. 16, 1–14.

Del Campo, J., Sieracki, M.E., Molestina, R., Keeling, P., Massana, R., Ruiz-Trillo, I., 2014. The others: Our biased perspective of eukaryotic genomes. Trends Ecol. Evol. 29, 252–259.

Des Combes, H.J., Abelmann, A., 2009. From species abundance to opal input: Simple geometrical models of radiolarian skeletons from the Atlantic sector of the Southern Ocean. Deep. Res. Part I Oceanogr. Res. Pap. 56, 757–771.

Dittami, S.M., Arboleda, E., Bigalke, A., Briand, E., Cárdenas, P., Cardini, U., Decelle, J., Eveillard, D., Gachon, C.M., Griffiths, S., Harder, T., Kayal, E., Kazamia, E., Lallier, F.H., Medina, M., Marzinelli, E.M., Morganti, T., Núñez Pons, L., Pardo, S., Pintado Valverde, J., Saha, M., Selosse, M.-A., Skillings, D., Stock, W., Sunagawa, S., Toulza, E., Vorobev, A., Not, F., 2019. A community perspective on the concept of marine holobionts: state-of-the-art, challenges, and future directions. PeerJ 1–24.

Dodd, M.S., Papineau, D., Grenne, T., Slack, J.F., Rittner, M., Pirajno, F., O'Neil, J., Little, C.T.S., 2017. Evidence for early life in Earth's oldest hydrothermal vent precipitates. Nature 543, 60–64.

Donoghue, P.C.J., Benton, M.J., 2007. Rocks and clocks: calibrating the Tree of Life using fossils and molecules. Trends Ecol. Evol. 22, 424–431.

Doolittle, W.F., Bapteste, E., 2007. Pattern pluralism and the Tree of Life hypothesis. Proc Natl Acad Sci U S A 104, 2043–2049.

Douzery, E.J.P., Snell, E.A., Bapteste, E., Delsuc, F., Philippe, H., 2004. The timing of eukaryotic evolution: Does a relaxed molecular clock reconcile proteins and fossils? PNAS 101, 15386–15391.

Drummond, A.J., Ho, S.Y.W., Phillips, M.J., Rambaut, A., 2006. Relaxed phylogenetics and dating with confidence. PLoS Biol. 4, 699–710.

Drummond, A.J., Suchard, M.A., Xie, D., Rambaut, A., 2012. Bayesian phylogenetics with BEAUti and the BEAST 1.7. Mol. Biol. Evol. 29, 1969–1973.

Duarte, C.M., 2015. Seafaring in the 21st century: The Malaspina 2010 circumnavigation expedition. Limnol. Oceanogr. Bull. 24, 11–14.

Dumitrica, P., 2017a. On the status of the Triassic radiolarian family Hexapylomellidae Kozur and Mostler: Taxonomic consequences. Rev. Micropaléontologie 60, 33–85.

Dumitrica, P., 2017b. Contribution to the knowledge of the Entactinaria radiolarian family Rhizosphaeridae Haeckel and description of some new genera and species. Rev. Micropaleontol. 60, 469–491.

Dumitrica, P., 1994. Pyloctostylys n. gen., a Cretaceous spumellarian radiolarian genus with initial spicule. Rev. Micropaleontol. 37, 235–244.

Dumitrica, P., 1989. Internal skeletal structures of the superfamily Pyloniacea (Radiolaria), a basis of a new systematics. Rev. Española Micropaleontol. XXI, 207–264.

Dumitrica, P., Caridroit, M., Wever, P. De, 2000. Archaeospicularia , ordre nouveau de radiolaires : une nouvelle étape pour la classification des radiolaires du Paléozoïque inférieur. C. R. Acad. Sci. 330, 563–569.

Dumitrica, P., Zügel, P., 2003. Lower Tithonian mono- and dicyrtid Nassellaria (Radiolaria) from the Solnhofen area (southern Germany). Geodiversitas 25, 72.

E

Edgar, R.C., 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. 32, 1792–1797.

Edgcomb, V., Orsi, W., Bunge, J., Jeon, S., Christen, R., Leslin, C., Holder, M., Taylor, G.T., Suarez, P., Varela, R., Epstein, S., 2011. Protistan microbial observatory in the Cariaco Basin, Caribbean. I. Pyrosequencing vs Sanger insights into species richness. ISME J. 5, 1344–1356.

Egge, E., Bittner, L., Andersen, T., Audic, S., Vargas, C. de, Edvardsen, B., 2013. 454 Pyrosequencing to Describe Microbial Eukaryotic Community Composition , Diversity and Relative Abundance : A Test for Marine Haptophytes. PLoS One 8.

Ehrenberg, C.G., 1846. Über eine halibiolithische, von Herrn R. Schomburgk entdeckte, vorherrschend aus mikroskopischen Polycystinen gebildete, Gebirgsmasse von Barbados. Bericht über die zur Bekanntmachung geeigneten Verhandlungen der Königlich Preußischen Akad. der Wissenschaften zu Berlin 1847, 382–385.

Ehrenberg, C.G., 1838. Über die Bildung der Kreidefelsen und des Kreidemergels durch unsichtbare Organismen, Abh. Preus. ed. Berlin.

Eme, L., Sharpe, S.C., Brown, M.W., Roger, A.J., 2014. On the Age of Eukaryotes: Evaluating Evidence from Fossils and Molecular Clocks. Cold Spring Harb. Perspect. Biol. 6.

Erbacher, J., Thurow, J., Littke, R., 1996. Evolution patterns of radiolaria and organic matter variations: A new approach to identify sea-level changes in mid-Cretaceous pelagic environments. Geology 24, 499–502.

Erwin, D.H., Laflamme, M., Tweedt, S.M., Sperling, E.A., Pisani, D., Peterson, K.J., 2011. The Cambrian Conundrum: Early Divergence and Later Ecological Success in the Early History of Animals. Science (80-.). 334, 1091–1097.

Escoufier, Y., 1973. Le Traitement des Variables Vectorielles. Biometrics 29, 751–760.

F

Faure, E., Not, F., Benoiston, A.-S., Labadie, K., Bittner, L., Ayata, S.-D., 2019. Mixotrophic protists display contrasted biogeographies in the global ocean. ISME J.

Felsenstein, J., 1985. Phylogenies and the comparative method. Am. Nat. 125, 1–15.

Felsenstein, J., 1981. Evolutionary trees from DNA sequences: A maximum likelihood approach. J. Mol. Evol. 17, 368–376.

Fensome, R.A., MacRae, R.A., Moldowan, J.M., Taylor, F.J.R., Williams, G.L., 1996. The Early Mesozoic Radiation of Dinoflagellates. Paleontol. Soc. 22, 329–338.

Field, C.B., Behrenfeld, M.J., Randerson, J.T., Falkowski, P., 1998. Primary Production of the Biosphere: Integrating Terrestrial and Oceanic Components. Science (80-.). 281, 237–240.

Flynn, K.J., Stoecker, D.K., Mitra, A., Raven, J.A., Glibert, P.M., Hansen, P.J., Granéli, E., Burkholder, J.M., 2013. Misuse of the phytoplankton-zooplankton dichotomy: The need to assign organisms as mixotrophs within plankton functional types. J. Plankton Res. 35, 3–11.

Forster, D., Dunthorn, M., Mahé, F., Dolan, J.R., Audic, S., Bass, D., Bittner, L., Boutte, C., Christen, R., Claverie, J.-M., Decelle, J., Edvardsen, B., Egge, E., Eikrem, W., Gobet, A., Kooistra, W.H.C.F., Logares, R., Massana, R., Montresor, M., Not, F., Ogata, H., Pawlowski, J., Pernice, M.C., Romac, S., Shalchian-Tabrizi, K., Simon, N., Richards, T.A., Santini, S., Sarno, D., Siano, R., Vaulot, D., Wincker, P., Zingone, A., Vargas, C. de, Stoeck, T., 2016. Benthic protists: the under-charted majority. FEMS Microbiol. Ecol.

Foster, C.S.P., Sauquet, H., Van Der Merwe, M., McPherson, H., Rossetto, M., Ho, S.Y.W., 2017. Evaluating the impact of genomic data and priors on Bayesian estimates of the angiosperm evolutionary timescale. Syst. Biol. 66, 338–351.

Foster, R.A., Carpenter, E.J., Bergman, B., 2006a. Unicellular cyanobionts in open ocean dinoflagellates, radiolarians, and tintinnids: Ultrastructural characterization and immuno-localization of phycoerythrin and nitrogenase. J. Phycol. 42, 453–463.

Foster, R.A., Collier, J.L., Carpenter, E.J., 2006b. Reverse transcription PCR amplification of cyanobacterial symbiont 16S rRNA sequences from single non-photosynthetic eukaryotic marine planktonic host cells. J. Phycol. 42, 243–250.

G

Gast, R.J., Caron, D.A., 2001. Photosymbiotic associations in planktonic foraminifera and radiolaria. Hydrobiologia 461, 1–7.

Gast, R.J., Caron, D.A., 1996. Molecular phylogeny of symbiotic dinoflagellates from planktonic foraminifera and radiolaria. Mol. Biol. Evol. 13, 1192–1197.

Gernhard, T., 2008. The conditioned reconstructed process. J. Theor. Biol. 253, 769–778.

Gilbert, S.F., McDonald, E., Boyle, N., Buttino, N., Gyi, L., Mai, M., Prakash, N., Robinson, J., 2010. Symbiosis as a source of selectable epigenetic variation: Taking the heat for the big guy. Philos. Trans. R. Soc. B Biol. Sci. 365, 671–678.

Giles, S., Xu, G., Near, T.J., Friedman, M., 2017. Early members of 'living fossil' lineage imply later origin of modern ray-finned fishes. Nature 549, 265–268.

Gilg, I.C., Amaral-Zettler, A., Countway, P.D., Moorthi, S., Schnetzer, A., Caron, D.A., 2010. Phylogenetic affiliations of mesopelagic Acantharia and acantharian-like environmental 18S rRNA genes off the southern California coast. Protist 161:197–211

Giner, C.R., Balagué, V., Pernice, M.C., Duarte, C.M., Gasol, J.M., Logares, R., Massana, R., 2019. Marked changes in diversity and relative activity of picoeukaryotes with depth in the global ocean. bioRxiv 552604.

Giner, C.R., Forn, I., Romac, S., Logares, R., de Vargas, C., Massana, R., 2016. Environmental sequencing provides reasonable estimates of the relative abundance of specific picoeukaryotes. Appl. Environ. Microbiol. AEM.00560-16.

Godhe, A., Asplund, M.E., Härnström, K., Saravanan, V., Tyagi, A., Karunasagar, I., 2008. Quantification of diatom and dinoflagellate biomasses in coastal marine seawater samples by real-time PCR. Appl. Environ. Microbiol. 74, 7174–7182.

Gong, J., Dong, J., Liu, X., Massana, R., 2013. Extremely High Copy Numbers and Polymorphisms of the rDNA Operon Estimated from Single Cell Analysis of Oligotrich and Peritrich Ciliates. Protist 164, 369–379.

Gong, W., Marchetti, A., 2019. Estimation of 18S Gene Copy Number in Marine Eukaryotic Plankton Using a Next-Generation Sequencing Approach. Front. Mar. Sci. 6, 1–5.

Goodwin, S., McPherson, J.D., McCombie, W.R., 2016. Coming of age: Ten years of next-generation sequencing technologies. Nat. Rev. Genet. 17, 333–351.

Gouy, M., Guindon, S., Gascuel, O., 2010. SeaView Version 4: A Multiplatform Graphical User Interface for Sequence Alignment and Phylogenetic Tree Building. Mol. Biol. Evol. 27, 221–224.

Grattepanche, J.-D.D., Walker, L.M., Ott, B.M., Paim Pinto, D.L., Delwiche, C.F., Lane, C.E., Katz, L.A., 2018. Microbial Diversity in the Eukaryotic SAR Clade: Illuminating the Darkness Between Morphology and Molecular Data. BioEssays 40, 1–12.

Groussin, M., Pawlowski, J., Yang, Z., 2011. Bayesian relaxed clock estimation of divergence times in foraminifera. Mol. Phylogenet. Evol. 61, 157–166.

Guidi, L., Chaffron, S., Bittner, L., Eveillard, D., et al., 2016. Plankton networks driving carbon export in the oligotrophic ocean. Nature 532, 465–470.

Guillou, L., Bachar, D., Audic, S., Bass, D., Berney, C., Bittner, L., Boutte, C., Burgaud, G., De Vargas, C., Decelle, J., Del Campo, J., Dolan, J.R., Dunthorn, M., Edvardsen, B., Holzmann, M., Kooistra, W.H.C.F., Lara, E., Le Bescot, N., Logares, R., Mahé, F., Massana, R., Montresor, M., Morard, R., Not, F., Pawlowski, J., Probert, I., Sauvadet, A.L., Siano, R., Stoeck, T., Vaulot, D., Zimmermann, P., Christen, R., 2013. The Protist Ribosomal Reference database (PR2): A catalog of unicellular eukaryote Small Sub-Unit rRNA sequences with curated taxonomy. Nucleic Acids Res. 41, 597–604.

Gutierrez-Rodriguez, A., Stukel, M.R., Lopes dos Santos, A., Biard, T., Scharek, R., Vaulot, D., Landry, M.R., Not, F., 2019. High contribution of Rhizaria (Radiolaria) to vertical export in the California Current Ecosystem revealed by DNA metabarcoding. ISME J. 13, 964–976.

H

Haeckel, E., 1890. Plankton-Studien, Jenaische Zeitschrift für Naturwissenschaft.

Haeckel, E., 1887. Report on the Radiolaria collected by H. M. S. Challenger during the years 1873–1876. Zoology 18, 1–1803.

Haeckel, E., 1882. Entwurf eines Radiolarien-Systems auf Grund von Studien der Challenger-Radiolarien. Jenaische Zeitschrift für Naturwiss. Hrsg. von der Medizinisch-naturwissenschaftlichen Gesellschaft zu Jena 15, 418–472.

Haeckel, E., 1862. Die Radiolarien (Rhizopoda Radiaria). Eine Monographie. Berlin: G. Reimer.

Hallock, P., Premoli Silva, I., Boersma, A., 1991. Similarities between planktonic and larger foraminiferal evolutionary trends through Paleogene paleoceanographic changes. Palaeogeogr. Palaeoclimatol. Palaeoecol. 83, 49–64.

Halverson, G.P., Wade, B.P., Hurtgen, M.T., Barovich, K.M., 2010. Neoproterozoic chemostratigraphy. Precambrian Res. 182, 337–350.

Hart, M.B., Hylton, M.D., Oxford, M.J., Price, G.D., Hudson, W., Smart, C.W., 2003. The search for the origin of the planktic Foraminifera. J. Geol. Soc. London. 160, 341–343.

Hebert, P.D.N., Cywinska, A., Ball, S.L., DeWaard, J.R., 2003. Biological identifications through DNA barcodes. Proc. R. Soc. B Biol. Sci. 270, 313–321.

Hendry, K.R., Marron, A.O., Vincent, F., Conley, D.J., Gehlen, M., Ibarbalz, F.M., Quéguiner, B., Bowler, C., 2018. Competition between Silicifiers and Non-silicifiers in the Past and Present Ocean and Its Evolutionary Impacts. Front. Mar. Sci. 5, 1–21.

Hensen, V., 1887. Uber die Bestimmung des Planktons oder des im Meere treibenden Materials an Pflanzen und Thieren, V. Bericht der Commission zur Wissenschaftlichen Untersuchung der Deutschen Meere, Jahrgang.

Hertwig, R., 1879. Der Organismus der Radiolarien. Gustav Fischer, Jena 149.

Hillis, D.M., Dixon, M.T., 1991. Ribosomal DNA: Molecular Evolution and Phylogenetic Inference. Q. Rev. Biol. 66, 411–453.

Ho, S.Y.W., Duchêne, S.S., 2014. Molecular-clock methods for estimating evolutionary rates and timescales. Mol. Ecol. 23, 5947–5965.

Ho, S.Y.W., Phillips, M.J., 2009. Accounting for calibration uncertainty in phylogenetic estimation of evolutionary divergence times. Syst. Biol. 58, 367–380.

Hollande, A., Enjumet, M., 1960. Cytologie, évolution et systématique des Sphaeroidés (Radioiaires). Arch. Mus. Nain. Hist. Nat. 7, 1–134.

Hollis, C.J., 2006. Radiolarian faunal turnover through the Paleocene-Eocene transition, Mead Stream, New Zealand. Eclogae Geol. Helv. 99, 79–100.

Hori, R.S., Yamakita, S., Ikehara, M., Kodama, K., Aita, Y., Sakai, T., Takemura, A., Kamata, Y., Suzuki, N., Takahashi, S., Sp??rli, K.B., Grant-Mackie, J.A., 2011. Early Triassic (Induan) Radiolaria and carbon-isotope ratios of a deep-sea sequence from Waiheke Island, North Island, New Zealand. Palaeoworld 20, 166–178.

Hughes, J.A., Gooday, A.J., 2004. Associations between living benthic foraminifera and dead tests of Syringammina fragilissima (Xenophyophorea) in the Darwin Mounds region (NE Atlantic). Deep - Sea Res. Part I - Oceanogr. Res. Pap. 51, 1741–1758.

Hull, P., 2015. Life in the aftermath of mass extinctions. Curr. Biol. 25, R941–R952.

I

Ikenoue, T., Bjørklund, K.R., Kruglikova, S.B., Onodera, J., Kimoto, K., Harada, N., 2015. Flux variations and vertical distributions of siliceous Rhizaria (Radiolaria and Phaeodaria) in the western Arctic Ocean: Indices of environmental changes. Biogeosciences 12, 2019–2046.

Irwin, D.E., Irwin, J.H., Price, T.D., 2001. Ring species as bridges between microevolution and speciation. Genetica 112–113, 223–243.

Irwin, N.A.T., Tikhonenkov, D. V., Hehenberger, E., Mylnikov, A.P., Burki, F., Keeling, P.J., 2019. Phylogenomics supports the monophyly of the Cercozoa. Mol. Phylogenet. Evol. 130, 416–423.

Isakova, T.N., Nazarov, B.B., 1986. Late Carboniferous – Early Permian stratigraphy and microfauna of Southern Urals. Tr. GIN 402, 1–183.

Ishitani, Y., Ishikawa, S.A., Inagaki, Y., Tsuchiya, M., Takahashi, K., Takishita, K., 2011. Multigene phylogenetic analyses including diverse radiolarian species support the " Retaria " hypothesis — The sister relationship of Radiolaria and Foraminifera. Mar. Micropaleontol. 81, 32–42.

Ishitani, Y., Ujiié, Y., de Vargas, C., Not, F., Takahashi, K., 2012a. Two distinct lineages in the radiolarian Order Spumellaria having different ecological preferences. Deep. Res. Part II Top. Stud. Oceanogr. 61–64, 172–178.

Ishitani, Y., Ujiié, Y., Vargas, C. de, Not, F., Takahashi, K., 2012b. Phylogenetic Relationships and Evolutionary Patterns of the Order Collodaria (Radiolaria). PLoS One 7, 1–8.

Ivalu, N.C., Baum, D.A., 2012. The Caribbean slipper spurge Euphorbia tithymaloides: The first example of a ring species in plants. Proc. R. Soc. B Biol. Sci. 279, 3377–3383.

J

Jain, M., Olsen, H.E., Paten, B., Akeson, M., 2016. Improved data analysis for the MinION nanopore sequencer. Nat. Methods 12, 351–356

Jamy, M., Foster, R., Barbera, P., Czech, L., Kozlov, A., Stamatakis, A., Bass, D., Burki, F., 2019. Long metabarcoding of the eukaryotic rDNA operon to phylogenetically and taxonomically resolve environmental diversity. bioRxiv 1–27.

Janouškovec, J., Gavelis, G.S., Burki, F., Dinh, D., Bachvaroff, T.R., Gornik, S.G., Bright, K.J., Imanian, B., Strom, S.L., Delwiche, C.F., Waller, R.F., Fensome, R.A., Leander, B.S., Rohwer, F.L., Saldarriaga, J.F., 2017. Major transitions in dinoflagellate evolution unveiled by phylotranscriptomics. Proc. Natl. Acad. Sci. 114, E171–E180.

Jenkyns, H.C., 2010. Geochemistry of oceanic anoxic events. Geochemistry Geophys. Geosystems 11, 1–30.

Jenkyns, H.C., 1998. The Early Toarcian Anoxic event: Stratigraphic, sedimentary, and geochrmical evidence. Am. J. Sci.

K

Kachovich, S., Sheng, J., Aitchison, J.C., 2019. Adding a new dimension to investigations of early radiolarian evolution. Sci. Rep. 9, 1–10.

Kamikuri, S.-I., Moore, T.C., Ogane, K., Suzuki, N., Pälike, H., Nishi, H., 2012. Early Eocene to early Miocene radiolarian biostratigraphy for the low-latitude Pacific Ocean. Stratigraphy 9, 77–108.

Kamikuri, S. ichi, Motoyama, I., Nishi, H., Iwai, M., 2009. Evolution of Eastern Pacific Warm Pool and upwelling processes since the middle Miocene based on analysis of radiolarian assemblages: Response to Indonesian and Central American Seaways. Palaeogeogr. Palaeoclimatol. Palaeoecol. 280, 469–479.

Katoh, K., Standley, D.M., 2013. MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. Mol. Biol. Evol. 30, 772–780.

Keeling, P.J., Burki, F., 2019. Progress towards the Tree of Eukaryotes. Curr. Biol. 29, R808–R817.

Kemp, D.B., Izumi, K., 2014. Multiproxy geochemical analysis of a Panthalassic margin record of the early Toarcian oceanic anoxic event (Toyora area, Japan). Palaeogeogr. Palaeoclimatol. Palaeoecol. 414, 332–341.

Kidder, D.L., Tomescu, I., 2016. Biogenic chert and the Ordovician silica cycle. Palaeogeogr. Palaeoclimatol. Palaeoecol. 458, 29–38.

Kiessling, W., 2002. Radiolarian diversity patterns in the latest Jurassic-earliest Cretaceous. Palaeogeogr. Palaeoclimatol. Palaeoecol. 187, 179–206.

Kiessling, W., Tragelehn, H., 1994. Devonian radiolarian faunas of conodont-dated localities in the Frankenwald (northern Bavaria, Germany). Abh. Geol. Bundesanst. Win 50, 219–225.

Kim, D.Y., Countway, P.D., Yamashita, W., Caron, D.A., 2012. A combined sequence-based and fragment-based characterization of microbial eukaryote assemblages provides taxonomic context for the Terminal Restriction Fragment Length Polymorphism (T-RFLP) method. J. Microbiol. Methods 91, 527–536.

Kircher, M., Sawyer, S., Meyer, M., 2012. Double indexing overcomes inaccuracies in multiplex sequencing on the Illumina platform. Nucleic Acids Res. 40, 1–8.

Kooistra, W.H.C.F., Medlin, K.L., 1996. Evolution of the Diatoms (Bacillariophyta). Mol. Phylogenet. Evol. 6, 391-407 ST-Evolution of the Diatoms (Bacillario.

Kozlova, G.E., 1999. Radiolarii paleogene boreal'noi oblasti Rossii. Practicheskoe rykovodstvo po microfayne Rossii. 9. VNIGRI Publishers, Sankt-Peterburg (In Russian).

Krabberød, A.K., Brate, J., Dolven, J.K., Ose, R.F., Klaveness, D., Kristensen, T., 2011. Radiolaria Divided into Polycystina and Spasmaria in Combined 18S and 28S rDNA Phylogeny. PLoS One 6.

Krabberød, A.K., Orr, R.J.S., Bråte, J., Kristensen, T., Bjørklund, K.R., Shalchian-Tabrizi, K., 2017. Single Cell Transcriptomics, Mega-Phylogeny, and the Genetic Basis of Morphological Innovations in Rhizaria. Mol. Biol. Evol. 34, 1557–1573.

Kunin, V., Engelbrektson, A., Ochman, H., Hugenholtz, P., 2010. Wrinkles in the rare biosphere: Pyrosequencing errors can lead to artificial inflation of diversity estimates. Environ. Microbiol. 12, 118–123.

Kunitomo, Y., Sarashina, I., Iijima, M., Endo, K., Sashida, K., 2006. Molecular phylogeny of acantharian and polycystine radiolarians based on ribosomal DNA sequences, and some comparisons with data from the fossil record. Eur. J. Protistol. 42, 143–153.

L

Lahr, D.J.G., Kosakyan, A., Lara, E., Mitchell, E.A.D., Morais, L., Porfirio-Sousa, A.L., Ribeiro, G.M., Tice, A.K., Pánek, T., Kang, S., Brown, M.W., 2019. Phylogenomics and Morphological Reconstruction of Arcellinida Testate Amoebae Highlight Diversity of Microbial Eukaryotes in the Neoproterozoic. Curr. Biol. 29, 991-1001.e3.

Lariviere, J.P., Ravelo, A.C., Crimmins, A., Dekens, P.S., Ford, H.L., Lyle, M., Wara, M.W., 2012. Late Miocene decoupling of oceanic warmth and atmospheric carbon dioxide forcing. Nature 486, 97–100.

Lazarus, D.B., Kotrc, B., Wulf, G., Schmidt, D.N., 2009. Radiolarians decreased silicification as an evolutionary response to reduced Cenozoic ocean silica availability. Proc. Natl. Acad. Sci. 106, 9333–9338.

Laver, T., Harrison, J., O'Neill, P.A., Moore, K., Farbos, A., Paszkiewicz, K., Studholme, D.J., 2015 Assessing the performance of the Oxford Nanopore technologies MinION. Biomol Detect Quantif. 3:1–8.

Leckie, R.M., Bralower, T.J., Cashman, R., 2002. Oceanic anoxic events and plankton evolution : Biotic response to tectonic forcing during the mid-Cretaceous. Paleoceanography 17.

Legendre, P., Gallagher, E.D., 2001. Ecologically meaningful transformations for ordination of species data. Oecologia 129, 271–280.

Leles, S.G., Mitra, A., Flynn, K.J., Stoecker, D.K., Hansen, P.J., Calbet, A., Mcmanus, G.B., Sanders, R.W., Caron, D.A., Not, F., Hallegraeff, G.M., Marı, C.P., 2017. Oceanic protists with different forms of acquired phototrophy display contrasting biogeographies and abundance. Proc. R. Soc. B 284, 1–6.

Levy, S.E., Myers, R.M., 2016. Advancements in Next-Generation Sequencing. Annu. Rev. Genomics Hum. Genet. 17, 95–115.

Lewitus, E., Bittner, L., Malviya, S., Bowler, C., Morlon, H., 2018. Clade-specific diversification dynamics of marine diatoms since the Jurassic. Nat. Ecol. Evol. 2, 1715–1723.

Lie, A.A.Y., Liu, Z., Hu, S.K., Jones, A.C., Kim, D.Y., Countway, P.D., Amaral-Zettler, L.A., Cary, S.C., Sherr, E.B., Sherr, B.F., Gast, R.J., Caron, D.A., 2014. Investigating microbial eukaryotic diversity from a global census: Insights from a comparison of pyrotag and full-length sequences of 18S rRNA genes. Appl. Environ. Microbiol. 80, 4363–4373.

Linné, C.v., 1767-1770. Systema naturae, per regna tria naturae : secundum classes, ordines, genera, species cum characteribus, differentiis, synonymis, locis. Vindobonae. Vienna.

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

López-García, P., Rodrıguez-Valera, F., Moreira, D., 2002. Toward the Monophyly of Haeckel's Radiolaria: 18S rRNA Environmental Data Support the Sisterhood of Polycystinea and Acantharea. Lett. to Ed. Mol. Biol. Evol. 2, 118–121.

Loron, C.C., Rainbird, R.H., Turner, E.C., Greenman, J.W., Javaux, E.J., 2018. Implications of selective predation on the macroevolution of eukaryotes: evidence from Arctic Canada. Emerg. Top. Life Sci. 2, 247–255.

Lovejoy, C., Massana, R., Pedros-Alio, C., 2006. Diversity and Distribution of Marine Microbial Eukaryotes in the Arctic Ocean and Adjacent Seas. Appl. Environ. Microbiol. 72, 3085–3095.

M

Mahé, F., Rognes, T., Quince, C., de Vargas, C., Dunthorn, M., 2015. Swarmv2: Highly-scalable and high-resolution amplicon clustering. PeerJ 2015, 1–12.

Maletz, J., 2011. Radiolarian skeletal structures and biostratigraphy in the early Palaeozoic (Cambrian-Ordovician). Palaeoworld 20, 116–133.

Martin, M., 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet.journal 17, 10–12.

Massana, R., Gobet, A., Audic, S., Bass, D., Bittner, L., Boutte, C., Chambouvet, A., Christen, R., Claverie, J.-M., Decelle, J., Dolan, J.R., Dunthorn, M., Edvardsen, B., Forn, I., Forster, D., Guillou, L., Jaillon, O., Kooistra, W.H.C.F., Logares, R., Mahé, F., Not, F., Ogata, H., Pawlowski, J., Pernice, M.C., Probert, I., Romac, S., Richards, T., Santini, S., Shalchian-Tabrizi, K., Siano, R., Simon, N., Stoeck, T., Vaulot, D., Zingone, A., Vargas, C. de, 2015. Marine protist diversity in European coastal waters and sediments as revealed by high- throughput sequencing. Environ. Microbiol. 17, 4035–4049.

Matsen, F.A., Kodner, R.B., Armbrust, E.V., 2010. pplacer: linear time maximum-likelihood and Bayesian phylogenetic placement of sequences onto a fixed reference tree. BMC Bioinformatics 11, 538.

Matsuoka, A., 2004. Toarcian (Early Jurassic) radiolarian in the Mino Terrane , fauna from the Nanjo Massif central Japan served radiolarian fauna of Toarcian age . Only nine nassellarian species have been illustrated in the sample but.

Matsuzaki, K.M., Suzuki, N., Nishi, H., 2015. Middle to Upper Pleistocene polycystine radiolarians from Hole 902-C9001C, northwestern Pacific. Paleontol. Res. 19(supp.), 1–77.

Mayr, E., 1942. Systematics and the origin of species. Columbia University Press, New York

McFall-Ngai, M., Hadfield, M.G., Bosch, T.C.G., Carey, H. V., Domazet-Lošo, T., Douglas, A.E., Dubilier, N., Eberl, G., Fukami, T., Gilbert, S.F., Hentschel, U., King, N., Kjelleberg, S., Knoll, A.H., Kremer, N., Mazmanian, S.K., Metcalf, J.L., Nealson, K., Pierce, N.E., Rawls, J.F., Reid, A., Ruby, E.G., Rumpho, M., Sanders, J.G., Tautz, D., Wernegreen, J.J., 2013. Animals in a bacterial world, a new imperative for the life sciences. Proc. Natl. Acad. Sci. U. S. A. 110, 3229–3236.

Medinger, R., Nolte, V., Pandey, R.V., Jost, S., Ottenwälder, B., Schlötterer, C., Boenigk, J., 2010. Diversity in a hidden world: Potential and limitation of next-generation sequencing for surveys of molecular diversity of eukaryotic microorganisms. Mol. Ecol. 19, 32–40.

Medlin, L., Elwood, H.J., Stickel, S., Sogin, M.L., 1988. The Characterization of enzymatically amplified eukaryotic 16S-like rRNA-coding regions. Genetics 71, 491–499.

Meyen, F. 1834. *Über das Leuchten des Meeres und Bes- chreibeng einiger Polypen und anderer niederer Thiere,* Verh. Kaiserl. Leopoldin.-Carolin. Akad. Naturwissensch. 16 (Suppl. 1), 125–216.

Michaels, A.F., 1988. Vertical distribution and abundance of Acantharia and their symbionts. Mar. Biol. 97, 559–569.

Michaels, A.F., Caron, D.A., Swanberg, N.R., Howse, F.A., Michaels, C.M., 1995. Planktonic sarcodines (Acantaria, Radiolaria, Foraminifera) in surface waters near Bermuda: abundance, biomass and vertical flux. J. Plankton Res. 17, 131–163.

Miller, A.K., Kerr, A.M., Paulay, G., Reich, M., Wilson, N.G., Carvajal, J.I., Rouse, G.W., 2017. Molecular phylogeny of extant Holothuroidea (Echinodermata). Mol. Phylogenet. Evol. 111, 110–131.

Mitra, A., Flynn, K.J., Tillmann, U., Raven, J.A., Caron, D., Stoecker, D.K., Not, F., Hansen, P.J., Hallegraeff, G., Sanders, R., Wilken, S., Mcmanus, G., Johnson, M., Pitta, P., Våge, S., Berge, T.,

Calbet, A., Thingstad, F., Jin, H., Burkholder, J., Glibert, P.M., Granéli, E., Lundgren, V., 2016. Defining Planktonic Protist Functional Groups on Mechanisms for Energy and Nutrient Acquisition: Incorporation of Diverse Mixotrophic Strategies. Protist 167, 106–120.

Moreira, D., von der Heyden, S., Bass, D., López-García, P., Chao, E., Cavalier-Smith, T., 2007. Global eukaryote phylogeny: Combined small- and large-subunit ribosomal DNA trees support monophyly of Rhizaria, Retaria and Excavata. Mol. Phylogenet. Evol. 44, 255–266.

Morton, J.T., Marotz, C., Washburne, A., Silverman, J., Zaramela, L.S., Edlund, A., Zengler, K., Knight, R., 2019. Establishing microbial composition measurement standards with reference frames. Nat. Commun. 10.

Motoyama, I., Ota, M., Kokushou, T., Tanaka, Y., 2005. Seasonal changes in fluxes and assemblages of radiolarians collected by sediment trap experiments in the northwestern Pacific: a family-level analysis. J. Geol. Soc. Japan 111, 404–416.

Müller, J., 1858. Über die Thalassicollen Polycystinen und Acanthometren des Mittelmeeres.

N

Nakamura, Y., Imai, I., Yamaguchi, A., Tuji, A., Not, F., Suzuki, N., 2015. Molecular Phylogeny of the Widely Distributed Marine Protists, Phaeodaria (Rhizaria, Cercozoa). Protist 166, 363–373.

Nettersheim, B.J., Brocks, J.J., Schwelm, A., Hope, J.M., Not, F., Lomas, M., Schmidt, C., Schiebel, R., Nowack, E.C.M., De Deckker, P., Pawlowski, J., Bowser, S.S., Bobrovskiy, I., Zonneveld, K., Kucera, M., Stuhr, M., Hallmann, C., 2019. Putative sponge biomarkers in unicellular Rhizaria question an early rise of animals. Nat. Ecol. Evol. 3, 577–581.

Nikolaev, S.I., Berney, C., Fahrn, J.F., Bolivar, I., Polet, S., Mylnikov, A.P., Aleshin, V. V., Petrov, N.B., Pawlowski, J., 2004. The twilight of Heliozoa and rise of Rhizaria, an emerging supergroup of amoeboid eukaryotes. Proc. Natl. Acad. Sci. U. S. A. 101, 8066–8071.

Nishimura, A., 1992. Paleocene Radiolarian Biostratigraphy in the Northwest Atlantic at Site 384, Leg 43, of the Deep Sea Drilling Project. Micropaleontology 38, 317.

Nitsche, F., Thomsen, H.A., Richter, D.J., 2016. Bridging the gap between morphological species and molecular barcodes − exemplified by loricate choanoflagellates. Eur. J. Protistol. 57, 26–37.

Noble, P., Aitchison, J.C., Danelian, T., Dumitrica, P., Maletz, J., Suzuki, N., Cuvelier, J., Caridroit, M., O'Dogherty, L., 2017. Taxonomy of Paleozoic radiolarian genera, in: Danelian, T., Caridroit, Martial, Noble, P., Aitchison, J.C. (Eds.), Catalogue of Paleozoic Radiolarian Genera. Geodiversitas, pp. 419–502.

Noble, P., Danelian, T., 2004. Radiolarians, in: Webby, B.D., Droser, M.L., Paris, F., Percival, I.G. (Eds.), The Great Ordovician Biodiversification Event. Columbia University Press, New York, pp. 97–101.

Not, F., del Campo, J., Balague, V., Vargas, C. De, Massana, R., 2009. New Insights into the Diversity of Marine Picoeukaryotes. PLoS One 4.

Not, F., Gausling, R., Azam, F., Heidelberg, J.F., Worden, A.Z., 2007. Vertical distribution of picoeukaryotic diversity in the Sargasso Sea. Environ. Microbiol. 9, 1233–1252.

O

O'Dogherty, L., 1994. Biochronology and Paleontology of Mid-Cretaceous Radiolarians from Northern Apennines (Italy) and Betic Cordillera (Spain). Mémoires de Géologie (Lausanne) 21, 1–415.

O'Dogherty, L., Carter, E.S., Dumitrica, P., Gori, Š., Wever, P. De, Bandini, A.N., Baumgartner, P.O., Matsuoka, A., 2009a. Catalogue of Mesozoic Radiolarian Genera . Part 2 : Jurassic- Cretaceous 31, 271–356.

O'Dogherty, L., Carter, E.S., Dumitrica, P., Gori, Š., Wever, P. De, Hungerbühler, A., Bandini, A.N., Takemura, A., 2009b. Catalogue of Mesozoic radiolarian genera. Part 1: Triassic. Geodiversitas 31, 213–270.

O'Dogherty, L., De Wever, P., Goričan, Š., Carter, E.S., Dumitrica, P., 2011. Stratigraphic ranges of Mesozoic radiolarian families. Palaeoworld 20, 102–115.

O'Malley, M.A., Simpson, A.G.B., Roger, A.J., 2013. The other eukaryotes in light of evolutionary protistology. Biol. Philos. 28, 299–330.

Ogane, K., Tuji, A., Suzuki, N., Matsuoka, A., Kurihara, T., Hori, R.S., 2010. Direct observation of the skeletal growth patterns of polycystine radiolarians using a fluorescent marker. Mar. Micropaleontol. 77, 137–144.

Oksanen, J., Blanchet, F.G., Kindt, R., Legendre, P., Minchin, P.R., Hara, R.B.O., Simpson, G.L., Solymos, P., Stevens, M.H.H., Wagner, H., 2015. vegan: Community Ecology Package.

Orsi, W., Edgcomb, V., Jeon, S., Leslin, C., Bunge, J., Taylor, G.T., Varela, R., Epstein, S., 2011. Protistan microbial observatory in the Cariaco Basin, Caribbean. II. Habitat specialization. ISME J. 5, 1357–1373.

Orsi, W., Song, Y.C., Hallam, S., Edgcomb, V., 2012. Effect of oxygen minimum zone formation on communities of marine protists. ISME J. 6, 1586–1601.

P

Paradis, E., Claude, J., Strimmer, K., 2004. APE: Analyses of phylogenetics and evolution in R language. Bioinformatics 20, 289–290.

Pawlowski, J., Audic, S., Adl, S., Bass, D., Belbahri, L., Berney, C. dric, Bowser, S.S., Cepicka, I., Decelle, J., Dunthorn, M., Fiore-Donno, A.M., Gile, G.H., Holzmann, M., Jahn, R., Jirku, M., Keeling, P.J., Kostka, M., Kudryavtsev, A., Lara, E., Lukes, J., Mann, D.G., Mitchell, E.A.D., Nitsche, F., Romeralo, M., Saunders, G.W., Simpson, A.G.B., Smirnov, A. V., Spouge, J.L., Stern, R.F., Stoeck, T., Zimmermann, J., Schindel, D., Vargas, C. de, 2012. CBOL Protist Working Group: Barcoding Eukaryotic Richness beyond the Animal, Plant, and Fungal Kingdoms. PLoS Biol. 10, e1001419.

Pawlowski, J., Fontaine, D., da Silva, A.A., Guiard, J., 2011. Novel lineages of Southern Ocean deep-sea foraminifera revealed by environmental DNA sequencing. Deep. Res. Part II Top. Stud. Oceanogr. 58, 1996–2003.

Pawlowski, J., Holzmann, M., 2002. Molecular phylogeny of Foraminifera - A review. Eur. J. Protistol. 38, 1–10.

Pawlowski, J., Holzmann, M., Berney, C., Fahrni, J., Gooday, A.J., Cedhagen, T., Habura, A., Bowser, S.S., 2003a. The evolution of early Foraminifera. Proc. Natl. Acad. Sci. U. S. A. 100, 11494–8.

Pawlowski, J., Holzmann, M., Tyszka, J., 2013. New supraordinal classification of Foraminifera: Molecules meet morphology. Mar. Micropaleontol. 100, 1–10.

Pawlowski, J., Holzmann, M., Fahrni, J., Richardson, S.L., 2003b. Small Subunit Ribosomal DNA Suggests that the Xenophyophorean Syringammina corbicula is a Foraminiferan. Eukaryot. Microbiol. 50, 483–487.

Pernice, M.C., Giner, C.R., Logares, R., Perera-bel, J., Acinas, S.G., Duarte, C.M., Gasol, J.M., Massana, R., 2016. Large variability of bathypelagic microbial eukaryotic communities across the world ' s oceans. ISME J. 10, 945–958.

Pernice, M.C., Logares, R., Guillou, L., Massana, R., 2013. General Patterns of Diversity in Major Marine Microeukaryote Lineages. PLoS One 8, e57170.

Pesant, S., Not, F., Picheral, M., Kandels-Lewis, S., Le Bescot, N., Gorsky, G., Iudicone, D., Karsenti, E., Speich, S., Troublé, R., Dimier, C., Searson, S., Acinas, S.G., Bork, P., Boss, E., Bowler, C., De Vargas, C., Follows, M., Grimsley, N., Hingamp, P., Jaillon, O., Karp-Boss, L., Krzic, U., Ogata, H., Raes, J., Reynaud, E.G., Sardet, C., Sieracki, M., Stemmann, L., Sullivan, M.B., Sunagawa, S., Velayoudon, D., Weissenbach, J., Wincker, P., 2015. Open science resources for the discovery and analysis of Tara Oceans data. Sci. Data 2, 150023.

Pessagno, E.A., 1976. Radiolarian zonation and stratigraphy of the upper Cretaceous portion of the Great Valley sequence, California Coast Ranges. Micropaleontol. Spec. Publ. No. 2, 100.

Pessagno, E.A., Blome, C.D., 1980. Upper Triassic and Jurassic Pantanelliinae from California, Oregon and British Columbia 26, 225–273.

Petrushevskaya, M.G., 1981. [Nassellarian Radiolarians from the World Oceans]. Zool. Institute, Acad. Sci. USSR. Nauk. Leningr. Otd. Leningrad, 405 p (in Russ.

Petrushevskaya, M.G., 1975. Cenozoic Radiolarians of the Antarctic, Leg 29, Dsdp, in: Kennett, J.P., Houtz, R.E., Al., E. (Eds.), Initial Reports of the Deep Sea Drilling Project. U. S. Government Printing Office, Washington, DC., pp. 541–675.

Petrushevskaya, M.G., 1971a. Radiolarians of the World Ocean. Acad. Sci. USSR. Inst. Zool. 393.
Petrushevskaya, M.G., 1971b. On the Natural System of Polycystine Radiolaria, in: Proceedings of the II Planktonic Conference. pp. 981–992.
Pitsch, G., Bruni, E.P., Forster, D., Qu, Z., Sonntag, B., Stoeck, T., Posch, T., 2019. Seasonality of planktonic freshwater ciliates: Are analyses based on V9 regions of the 18S rRNA gene correlated with morphospecies counts? Front. Microbiol. 10, 1–15.
Pomerantz, A., Peñafiel, N., Arteaga, A., Bustamante, L., Pichardo, F., Coloma, L.A., Barrio-Amorós, C.L., Salazar-Valenzuela, D., Prost, S., 2018. Real-time DNA barcoding in a rainforest using nanopore sequencing: Opportunities for rapid biodiversity assessments and local capacity building. Gigascience 7, 1–14.
Porter, S., 2011. The rise of predators. Geol. Soc. Am. 39, 607–608.
Pouille, L., Obut, O., Danelian, T., Sennikov, N., 2011. Comptes Rendus Palevol Lower Cambrian (Botomian) polycystine Radiolaria from the Altai Mountains (southern Siberia, Russia). Comptes rendus - Palevol 10, 627–633.
Probert, I., Siano, R., Poirier, C., Decelle, J., Biard, T., Tuji, A., Suzuki, N., Not, F., 2014. Brandtodinium Gen. Nov. and B. Nutricula Comb. Nov. (Dinophyceae), A Dinoflagellate Commonly Found In Symbiosis With Polycystine Radiolarians. Phycol. Soc. Am. 50, 388–399.
Prosser, J.I., 2010. Replicate or lie. Environ. Microbiol. 12, 1806–1810.
Pulquério, M.J.F., Nichols, R.A., 2007. Dates from the molecular clock: how wrong can we be? Trends Ecol. Evol. 22, 180–184.

R

R Core Team, 2014. R: A Language and Environment for Statistical Computing.
Rang, F.J., Kloosterman, W.P., de Ridder, J., 2018. From squiggle to basepair: Computational approaches for improving nanopore sequencing read accuracy. Genome Biol. 19, 1–11.
Ray, J., 1686. *Historia plantarum*. London: Clark.
Renaudie, J., Lazarus, D.B., 2015. New species of Neogene radiolarians from the Southern Ocean – part IV. J. Micropalaeontology 2014–026.
Riedel, W.R.E.X., 1967. Some new families of Radiolaria. Proc. Geol. Soc. London 1640, 148–149.
Rognes, T., Flouri, T., Nichols, B., Quince, C., Mahé, F., 2016. VSEARCH: a versatile open source tool for metagenomics. PeerJ 4, e2584.
Rosenberg, E., Koren, O., Reshef, L., Efrony, R., Zilber-Rosenberg, I., 2007. The role of microorganisms in coral health, disease and evolution. Nat. Rev. Microbiol. 5, 355–362.
Rothschild, L.J., 1989. Protozoa, Protista, Protoctista: What's in a Name? J. Hist. Biol. 22, 277–305.

S

Sandin, M.M., Pillet, L., Biard, T., Poirier, C., Bigeard, E., Romac, S., Suzuki, N., Not, F., 2019. Time Calibrated Morpho-molecular Classification of Nassellaria (Radiolaria). Protist 170, 187–208.
Sanfilippo, A., Riedel, W.R., 1992. the Origin and Evolution of Pterocorythidae (Radiolaria) - a Cenozoic Phylogenetic Study. Micropaleontology 38, 1–36.
Santoferrara, L.F., Grattepanche, J.D., Katz, L.A., Mcmanus, G.B., 2016. Patterns and processes in microbial biogeography: do molecules and morphologies give the same answers? ISME J. 10, 1779–1790.
Sauquet, H., 2013. A practical guide to molecular dating. Comptes Rendus - Palevol 12, 355–367.
Schalanger, S.O., Jenkyns, H.C., 1976. Cretaceous oceanic anoxic events: causes and consequences. Geol. Mijnb. 55, 179–184.
Scher, H.D., Martin, E.E., 2006. Timing and climatic consequences of the opening of drake passage. Science (80-.). 312, 428–430.
Schlegel, M., Meisterfeld, R., 2003. The species problem in protozoa revisited.pdf 355, 349–355.
Schliep, K.P., 2011. phangorn: Phylogenetic analysis in R. Bioinformatics 27, 592–593.
Schloss, P.D., Westcott, S.L., Ryabin, T., Hall, J.R., Hartmann, M., Hollister, E.B., Lesniewski, R.A., Oakley, B.B., Parks, D.H., Robinson, C.J., Sahl, J.W., Stres, B., Thallinger, G.G., Van Horn, D.J., Weber,

C.F., 2009. Introducing mothur: Open-source, platform-independent, community-supported software for describing and comparing microbial communities. Appl. Environ. Microbiol. 75, 7537–7541.

Schnell, I.B., Bohmann, K., Gilbert, M.T.P., 2015. Tag jumps illuminated - reducing sequence-to-sample misidentifications in metabarcoding studies. Mol. Ecol. Resour. 15, 1289–1303.

Scholin, C. a, Herzog, M., Sogin, M., Anderson, D.M., 1994. Identification of group-specific and strain-specific genetic-markers for globally distributed Alexandrium (Dinophyceae). II. Sequence analysis of a fragment of the LSU rRNA gene. J. Phycol.

Schwartzapfel, J.A., Holdsworth, B.., 1996. Upper Devonian and Mississippian radiolarian zonation and biostratigraphy of the Woodford, Sycamore, Caney and Goddard Formation, Oklahoma. Cushman Found. Foraminifer. Res. Spec. Publ. 33, 1–275.

Sepkoski, J.J., 1981. A Factor Analytic Description of the Phanerozoic Marine Fossil Record. Paleobiology 7, 36–53.

Servais, T., Lehnert, O., Li, J., Mullins, G.L., Munnecke, A., Nützel, A., Vecoli, M., 2008. The Ordovician Biodiversification: Revolution in the oceanic trophic chain. Lethaia 41, 99–109.

Servais, T., Perrier, V., Danelian, T., Klug, C., Martin, R., Munnecke, A., Nowak, H., Nützel, A., Vandenbroucke, T.R.A., Williams, M., Rasmussen, C.M.Ø., 2016a. The onset of the 'Ordovician Plankton Revolution' in the late Cambrian. Palaeogeogr. Palaeoclimatol. Palaeoecol. 458, 12–28.

Servais, T., Vincent Perrier, Danelian, T., Klug, C., Martin, R., Munnecke, A., Nowak, H., Nützel, A., Vandenbroucke, T.R.A., MarkWilliams, Rasmussen, C.M.Ø., 2016b. The onset of the 'Ordovician Plankton Revolution' in the late Cambrian. Palaeogeogr. Palaeoclimatol. Palaeoecol. 458, 12–28.

Shaked, Y., De Vargas, C., 2006. Pelagic photosymbiosis: rDNA assessment of diversity and evolution of dinoflagellate symbionts and planktonic foraminiferal hosts. Mar. Ecol. Prog. Ser. 325, 59–71.

Shuman, B.N., 2013. Approaches. Encycl. Quat. Sci. 179–184.

Sierra, R., Matz, M. V, Aglyamova, G., Pillet, L., Decelle, J., Not, F., Vargas, C. De, Pawlowski, J., 2013. Deep relationships of Rhizaria revealed by phylogenomics: A farewell to Haeckel' s Radiolaria. Mol. Phylogenet. Evol. 67, 53–59.

Simpson, G.G., 1943. Criteria for Genera, Species, and Subspecies in Zoology and Paleozoology. Ann. N. Y. Acad. Sci. 44, 145–178.

Sims, P.A., Mann, D.G., Medlin, L.K., 2006. Evolution of the diatoms: insights from fossil, biological and molecular data. Phycologia 45, 361–402.

Slowinski, J.B., Page, R.D.M., Biology, S., Dec, N., 1999. How Should Species Phylogenies Be Inferred from Sequence Data? Syst. Biol. 48, 814–825.

Snel, B., Huynen, M.A., Dutilh, B.E., 2005. Genome Trees and the Nature of Genome Evolution. Annu. Rev. Microbiol. 59, 191–209.

Sokal, R.R., Sneath, P.H.A., 1965. Principles of numerical taxonomy, San Francisco and London.

Stamatakis, A., 2014. RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics 30, 1312–1313.

Stemmann, L., Youngbluth, M., Robert, K., Hosia, A., Picheral, M., Paterson, H., Ibanez, F., Guidi, L., Lombard, F., Gorsky, G., 2008. Global zoogeography of fragile macrozooplankton in the upper 100-1000 m inferred from the underwater video profiler. ICES J. Mar. Sci. 65, 433–442.

Stoeck, T., Bass, D., Nebel, M., Christen, R., Meredith, D., 2010. Multiple marker parallel tag environmental DNA sequencing reveals a highly complex eukaryotic community in marine anoxic water 19, 21–31.

Stoecker, D.K., Hansen, P.J., Caron, D.A., Mitra, A., 2016. Mixotrophy in the Marine Plankton. Ann. Rev. Mar. Sci. 9, annurev-marine-010816-060617.

Stoecker, D.K., Johnson, M.D., Vargas, C. De, Not, F., 2009. Acquired phototrophy in aquatic protists. Aquat. Microb. Ecol. 57, 279–310.

Sugiyama, K., 1997. Triassic and Lower Jurassic radiolarian biostratigraphy in the siliceous claystone and bedded chert units of the southeasern Mino Terrane, Central Japan. Bull. Mizunami Foss. Mus 24, 79–193.

Sugiyama, K., 1994. Lower Miocene new nassellarians (Radiolaria) from the Toyohama Formation, Morozaki Group, central Japan . Bull. Mizunami Foss. Museum (Mizunami-shi Kadeki Hakubutsukan Kenkyu Hokoku) 21, 1–11.

Sugiyama, K., 1992. Lower and Middle Triassic Radiolarians from Mt. Kinkazan, Gifu Prefecture, central Japan. Trans. Proc. Palaeont. Soc. Japan 167, 1180–1223.

Sugiyama, K., Anderson, O.R., 1997. Experimental and observational studies of radiolarian physiological ecology, 6. Effects of silicate-supplemented seawater on the longevity and weight gain of spongiose radiolarians Spongaster tetras and Dictyocoryne truncatum. Mar. Micropaleontol. 29, 159–172.

Sugiyama, K., Hori, R.S., Kusunoki, Y., Matsuoka, A., 2008. Pseudopodial features and feeding behavior of living nassellarians Eucyrtidium hexagonatum Haeckel, Pterocorys zancleus (Müler) and Dictyocodon prometheus Haeckel. Paleontol. Res. 12, 209–222.

Suzuki, N., Aita, Y., 2011. Radiolaria: achievements and unresolved issues: taxonomy and cytology. Plankt. Benthos Res. 6, 69–91.

Suzuki, N., Not, F., 2015. Biology and Ecology of Radiolaria, in: Ohtsuka, S., Suzaki, T., Horiguchi, T., Suzuki, N., Not, F. (Eds.), Marine Protists. Springer Japan, pp. 179–222.

Suzuki, N., Oba, M., 2015. Chapter 15. Oldest Fossil Records of Marine Protists and the Geologic History Toward the Establishment of the Modern-Type Marine Protist World, in: Marine Protists. pp. 359–394.

Suzuki, N., Ogane, K., Aita, Y., Kato, M., Sakai, S., Kurihara, T., Matsuoka, A., Ohtsuka, S., Go, A., Nakaguchi, K., Yamaguchi, S., Takahashi, T., Tuji, A., 2009. Distribution Patterns of the Radiolarian Nuclei and Symbionts Using. Bull. Natl. Mus. Nat. Sci., Ser. B 35, 169–182.

Swanberg, N.R., Anderson, O.R., Lindsey, J.L., Bennet, P., 1986. The biology of Physematium muelleri : trophic activity. Deep - Sea Res. 33, 913–922.

T

Taberlet, P., Coissac, E., 2012. Towards next-generation biodiversity assessment using DNA metabarcoding - Buscar con Google 33, 2045–2050.

Takahashi, K., 1982. Dimensions and sinking speeds of tropical radiolarian skeletons from the PARFLUX sediment traps.

Takahashi, S., Yamakita, S., Suzuki, N., Kaiho, K., Ehiro, M., 2009. High organic carbon content and a decrease in radiolarians at the end of the Permian in a newly discovered continuous pelagic section: A coincidence? Palaeogeogr. Palaeoclimatol. Palaeoecol. 271, 1–12.

Thornhill, D.J., Lajeunesse, T.C., Santos, S.R., 2007. Measuring rDNA diversity in eukaryotic microbial systems: How intragenomic variation, pseudogenes, and PCR artifacts confound biodiversity estimates. Mol. Ecol. 16, 5326–5340.

Twitchett, R.J., Barras, C.G., 2004. Trace fossils in the aftermath of mass extinction events 1 School of Earth , Ocean and Environmental Sciences , University of Plymouth , 2Department of Palaeontology , The Natural History Museum , Cromwell Road , London S W 7 5BD , UK 3Department of Earth. Geol. Soc. London Special Pu, 397–418.

Twitchett, R.J., Krystyn, L., Baud, A., Wheeley, J.R., Richoz, S., 2004. Rapid marine recovery after the end-Permian mass-extinction event in the absence of marine anoxia. Geology 32, 805–808.

Tyler, A.D., Mataseje, L., Urfano, C.J., Schmidt, L., Antonation, K.S., Mulvey, M.R., Corbett, C.R., 2018. Evaluation of Oxford Nanopore's MinION Sequencing Device for Microbial Whole Genome Sequencing Applications. Sci. Rep. 8, 1–12.

V

Vaulot, D., Eikrem, W., Viprey, M., Moreau, H., 2008. The diversity of small eukaryotic phytoplankton (≤3 µm) in marine ecosystems. FEMS Microbiol. Rev. 32, 795–820.

Van Valen, L., 1976 Ecological species, multispecies, and oaks. Taxon 25:233–239

Viprey, M., Guillou, L., Ferréol, M., Vaulot, D., 2008. Wide genetic diversity of picoplanktonic green algae (Chloroplastida) in the Mediterranean Sea uncovered by a phylum-biased PCR approach. Environ. Microbiol. 10, 1804–1822.

W

Weiner, A.K.M., Morard, R., Weinkauf, M.F., Darling, K.F., André, A., Quillévéré, F., Ujiie, Y., Douady, C.J., de Vargas, C., Kucera, M., 2016. Methodology for single-cell genetic analysis of planktonic foraminifera for studies of protist diversity and evolution. Fontiers Mar. Sci. 3, 1–15.

Weisburg, W.G., Barns, S.M., Pelletier, D.A., Lane, D.J., 1991. 16S Ribosomal DNA Amplification for Phylogenetic Study. J. Bacteriol. 173, 697–703.

Wenger, A.M., Peluso, P., Rowell, W.J., Chang, P.-C., Hall, R.J., Concepcion, G.T., Ebler, J., Fungtammasan, A., Kolesnikov, A., Olson, N.D., Toepfer, A., Chin, C.-S., Alonge, M., Mahmoud, M., Qian, Y., Phillippy, A.M., Schatz, M.C., Myers, G., DePristo, M.A., Ruan, J., Marschall, T., Sedlazeck, F.J., Zook, J.M., Li, H., Koren, S., Carroll, A., Rank, D.R., Hunkapiller, M.W., 2019. Highly-accurate long-read sequencing improves variant detection and assembly of a human genome. bioRxiv 519025.

Wick, R.R., Judd, L.M., Holt, K.E., 2019. Performance of neural network basecalling tools for Oxford Nanopore sequencing. Genome Biol. 20, 1–10.

Wick, R.R., Judd, L.M., Holt, K.E., 2018. Deepbinner: Demultiplexing barcoded Oxford Nanopore reads with deep convolutional neural networks. PLOS Comput. Biol. 1–11.

Wickham, H., 2016. Elegant Graphics for Data Analysis Second Edition. Springer-Verlag, New York.

Wiley, E.O., 1978. The evolutionary species concept reconsidered. Syst Zool 27:17–26

Wilkins, J.S., 2009. Species. A history of the idea, University of California Press. Berkeley.

Woese, C.R., Fox, G.E., 1977. Phylogenetic structure of the prokaryotic domain: The primary kingdoms. Proc. Natl. Acad. Sci. 74, 5088–5090.

Wood, R.A., 2011. Paleoecology of the earliest skeletal metazoan communities: Implications for early biomineralization. Earth-Science Rev. 106, 184–190.

Wu, W., Huang, B., Liao, Y., Sun, P., 2014. Picoeukaryotic diversity and distribution in the subtropical-tropical South China Sea. FEMS Microbiol. Ecol. 89, 563–579.

Wurzbacher, C., Larsson, E., Bengtsson-Palme, J., Van den Wyngaert, S., Svantesson, S., Kristiansson, E., Kagami, M., Nilsson, R.H., 2019. Introducing ribosomal tandem repeat barcoding for fungi. Mol. Ecol. Resour. 19, 118–127.

X

Xu, D., Jiao, N., Ren, R., Warren, A., 2017. Distribution and Diversity of Microbial Eukaryotes in Bathypelagic Waters of the South China Sea. J. Eukaryot. Microbiol. 64, 370–382.

Y

Yabuki, A., Chao, E.E., Ishida, K.I., Cavalier-Smith, T., 2012. Microheliella maris (Microhelida ord. n.), an Ultrastructurally Highly Distinctive New Axopodial Protist Species and Genus, and the Unity of Phylum Heliozoa. Protist 163, 356–388.

Yabuki, A., Toyofuku, T., Takishita, K., 2014. Lateral transfer of eukaryotic ribosomal RNA genes: An emerging concern for molecular ecology of microbial eukaryotes. ISME J. 8, 1544–1547.

Yilmaz, I.O., Altiner, D., Tekin, U.K., Ocakoglu, F., 2012. The first record of the " Mid-Barremian" Oceanic Anoxic Event and the Late Hauterivian platform drowning of the Bilecik platform, Sakarya Zone, western Turkey. Cretac. Res. 38, 16–39.

Yuasa, T., Dolven, J.K., Bjørklund, K.R., Mayama, S., Takahashi, O., 2009. Molecular phylogenetic position of Hexacontium pachydermum Jørgensen (Radiolaria). Mar. Micropaleontol. 73, 129–134.

Yuasa, T., Horiguchi, T., Mayama, S., Matsuoka, A., Takahashi, O., 2012. Ultrastructural and molecular characterization of cyanobacterial symbionts in Dictyocoryne profunda (polycystine radiolaria). Symbiosis 57, 51–55.

Yuasa, T., Horiguchi, T., Mayama, S., Takahashi, O., 2016. Gymnoxanthella radiolariae gen. et sp. nov. (Dinophyceae), a dinoflagellate symbiont from solitary polycystine radiolarians. J. Phycol. 52, 89–104.

Yuasa, T., Kawachi, M., Horiguchi, T., Takahashi, O., 2019. Chrysochromulina andersonii sp. nov. (Prymnesiophyceae), a new flagellate haptophyte symbiotic with radiolarians. Phycologia 58, 211–224.

Yuasa, T., Takahashi, O., 2016. Light and electron microscopic observations of the reproductive swarmer cells of nassellarian and spumellarian polycystines (Radiolaria). Eur. J. Protistol. 54, 19–32.

Yuasa, T., Takahashi, O., Honda, D., Mayama, S., 2005. Phylogenetic analyses of the polycystine Radiolaria based on the 18s rDNA sequences of the Spumellarida and the Nassellarida. Eur. J. Protistol. 41, 287–298.

Yule, G.U., 1925. A mathematical theory of evolution based on the conclusions of Dr. J. C. Willis, F.R.S. Philos. Trans. R. Soc. Lond. B. Biol. Sci. 213, 21–87.

Z

Zachos, F.E., 2016. An Annotated List of Species Concepts, Species Concepts in Biology. Springer.

Zhang, K., Feng, Q.L., 2019. Early Cambrian radiolarians and sponge spicules from the Niujiaohe Formation in South China. Palaeoworld 28, 234–242.

Zhang, L., Suzuki, N., Yasuhide, N., Tuji, A., 2018. Modern shallow water radiolarians with photosynthetic microbiota in the western North Pacific. Mar. Micropaleontol. 139, 1–27.

Zhao, F., Filker, S., Xu, K., Li, J., Zhou, T., Huang, P., 2019. Effects of intragenomic polymorphism in the SSU rRNA gene on estimating marine microeukaryotic diversity : A test for ciliates using single-cell high-throughput DNA sequencing. Limnol. Oceanogr.

Zhu, F., Massana, R., Not, F., Marie, D., Vaulot, D., 2005. Mapping of picoeucaryotes in marine ecosystems with quantitative PCR of the 18S rRNA gene. FEMS Microbiol. Ecol. 52, 79–92.

Zuckerkandl, E., Pauling, L., 1965. Molecules as documents of evolutionary history. J. Theor. Biol. 8, 357–366.

Annexes

Oral and poster presentations

2019 VIII European Congress of Protistology – International Society of Protistology Joint meeting, Rome (Italy):

Talk: Molecular diversity and evolution of Radiolaria.

2018 Young Researchers Day, Station Biologique de Roscoff (France):

Talk & poster: Time calibrated morpho-molecular classification of Nassellaria (Radiolaria).

2017 15th InterRad congress, Niigata (Japan):

Talk: Time calibrated morpho-molecular classification of Nassellaria (Radiolaria).

2016 Young Researchers Day, Station Biologique de Roscoff (France):

Flash-talk: My thesis in 180 seconds.

Oceanographic Expeditions

2018 TAN1810 SalPOOP: Tangaroa Research Vessel.

The Ocean Vacuum cleaner: Salp bloom effects on the carbon cycle and marine food web.

2017 MOOSE-GE: L'Atalante Research Vessel.

Mediterranean Ocean Observing System for the Environment.

Awards, grants and rewards

2018 Third best oral presentation award in Young Researchers Day (JJC) Station Biologique de Roscoff.

2017 Best oral presentation award for Young Scientists in InterRad XV, Japan.

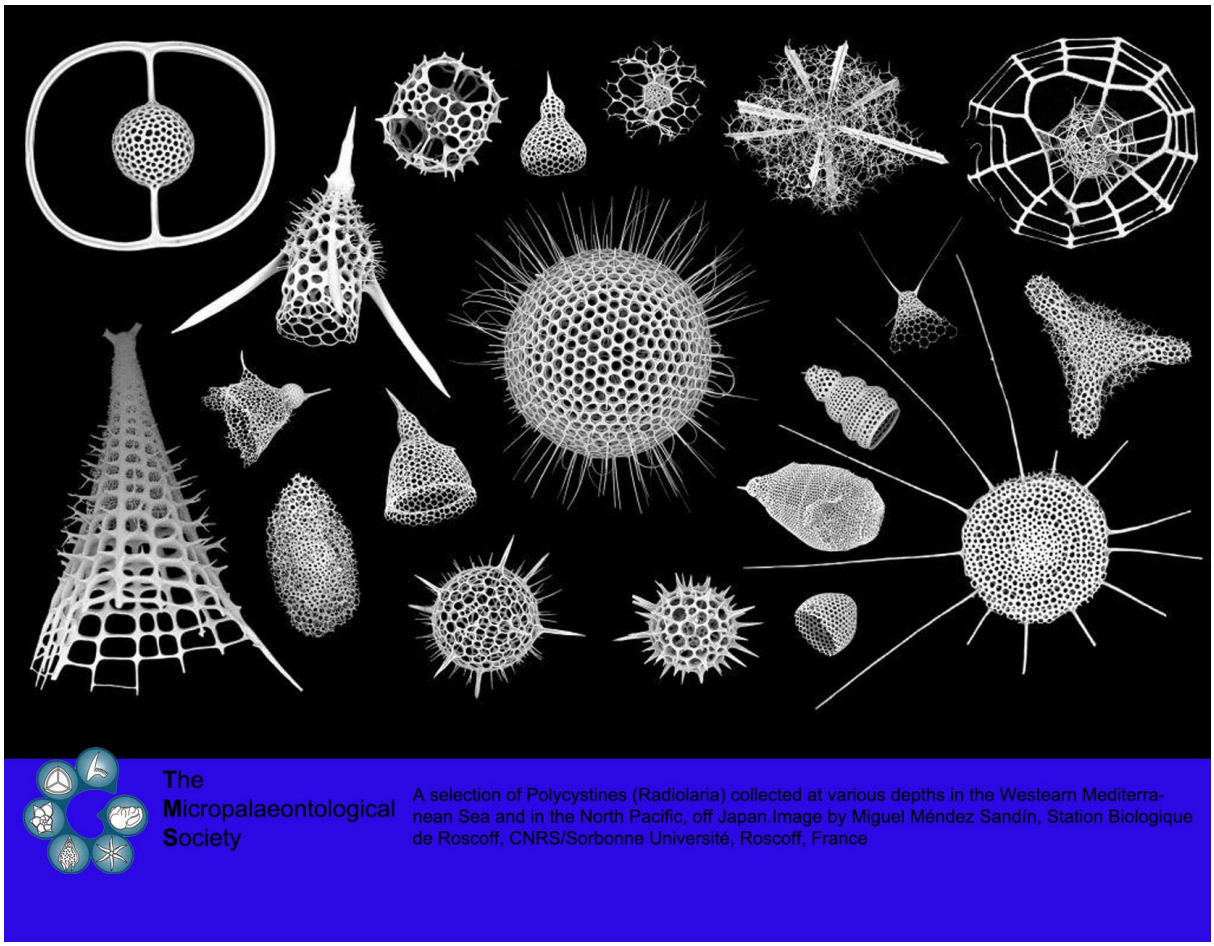
2017 Young Research Association (AJC, Station Biologique de Roscoff): Grant to attend InterRad XV, Japan.

(Continued on next page)

2019 Cover image of *Protist* journal, Volume 170, number 2, April 2019: Time calibrated morpho-molecular classification of Nassellaria (Cover Chapter 1.1).

This is a copy of the original cover

2019 The Micropaleontological Society: Microfossil Image Competition & Calendar 2019 (Cover Chapter 2)



Collaborations

Symbiont Chloroplasts Remain Active During Bleaching-Like Response Induced by Thermal Stress in Collozoum pelagicum (Collodaria, Retaria)

Emilie Villar[1], Vincent Dani[2], Estelle Bigeard[1], Tatiana Linhart[1], **Miguel Mendez-Sandin**[1], Charles Bachy[1], Christophe Six[1], Fabien Lombard[3], Cécile Sabourault[2] and Fabrice Not[1]

1-Sorbonne Université, CNRS – UMR7144 – Ecology of Marine Plankton Group – Station Biologique de Roscoff, Roscoff, France.
2-Université Côte d'Azur, Institut de Biologie Valrose UMR7277, Nice, France.
3- Sorbonne Université, CNRS – UMR 7093, Laboratoire d'Océanographie de Villefranche (LOV), Observatoire Océanologique, Villefranche-sur-Mer, France

Published in: Frontiers in Marine Science, 2018
doi.org/10.3389/fmars.2018.00387

Abstract

Collodaria (Retaria) are important contributors to planktonic communities and biogeochemical processes (e.g., the biologic pump) in oligotrophic oceans. Similarly to corals, Collodaria live in symbiosis with dinoflagellate algae, a relationship that is thought to explain partly their ecological success. In the context of global change, the robustness of the symbiotic interaction, and potential subsequent bleaching events are of primary interest for oceanic ecosystems functioning. In the present study, we compared the ultrastructure, morphology, symbiont density, photosynthetic capacities and respiration rates of colonial Collodaria exposed to a range of temperatures corresponding to natural conditions (21◦C), moderate (25◦C), and high (28◦C) thermal stress. We showed that symbiont density immediately decreased when temperature rose to 25◦C, while the overall Collodaria holobiont metabolic activity increased. When temperature reached 28◦C, the holobiont respiration nearly stopped and the host morphological structure was largely damaged, as if the host tolerance threshold has been crossed. Over the course of the experiment, the photosynthetic capacities of remaining algal symbionts were stable, chloroplasts being the last degraded organelles in the microalgae. These results contribute to a better characterization and understanding of temperature-induced bleaching processes in planktonic photosymbioses.

Villar, E., Dani, V., Bigeard, E., Linhart, T., **Mendez-Sandín, M**., Bachy, C., Six, C., Lombard, F., Sabourault, C., Not, F., 2018. *Symbiont Chloroplasts Remain Active During Bleaching-Like Response Induced by Thermal Stress in Collozoum pelagicum (Collodaria, Retaria)*. Front. Mar. Sci. 5, 1–11.

Estimating biogenic silica production of Rhizaria in the global ocean

Natalia Llopis Monferrer[1], Demetrio Boltovskoy[3], **Miguel Méndez Sandin**[2], Paul Tréguer[1], Fabrice Not[2], Aude Leynaert[1]

1-Marine Environmental Sciences Laboratory (LEMAR, UMR 6539) at the European Institute for Marine Studies (IUEM), Université de Bretagne Occidentale, CNRS, F-29280 Plouzané, France
2-UMR 7144 UPMC at the Station Biologique de Roscoff, France
3-University of Buenos Aires-CONICET · Institute of Ecology, Genetics and Evolution of Buenos Aires, Argentina

Submitted to: Global Biogeochemical Cycles

Abstract

Siliceous polycystine and phaeodarians are open-ocean planktonic protists found throughout the water column and are characterized by complex siliceous skeletons that are formed, at least partly, through the uptake of silicic acid. These protists contribute to marine carbon (C) and silica (Si) pools but little is known about their contribution to the Si biogeochemical cycle. Here we report the first measurements of the uptake rate of polycystine and phaeodarian cells using the 32Si based method from samples collected in the Mediterranean Sea. The elementary composition (biogenic silica, particulate organic carbon and nitrogen) of these organisms was also measured. Combining our results with published data on the distribution and abundance of Polycystina and Phaeodaria in the world ocean, we conclude that these organisms could contribute from 14 to 56% of the marine standing stock of bSi and from 2 to 16 % (5 to 47 Tmol Si yr-1) of the global oceanic bSi production. The implications for the global marine Si cycle are discussed.

Llopis-Monferrer N., Leynaert A., Tréguer P., Not F., **Sandín M.M**, l'Helguen S., Maguer J.F., Boltovskoy D. *Contribution of siliceous Rhizaria to the marine silica cycle*. (submitted). Global Biogeochemical Cycles

Five orders or six orders? - phylogenetic revision of Paleozoic relict radiolarians

Yasuhide Nakamura[1], **Miguel M. Sandin**[2], Noritoshi Suzuki[3], Akihiro Tuji[1] and Fabrice Not[4]

1-Department of Botany, National Museum of Nature and Science, Tsukuba 305–0005, Japan
2-Sorbonne Université, CNRS – UMR7144 – Ecology of Marine Plankton Group – Station Biologique de Roscoff, Roscoff 29682, France
3-Department of Earth Science, Graduate School of Science, Tohoku University, Sendai 980–8578, Japan
4-Station Biologique de Roscoff, Centre National de la Recherche Scientifique, Roscoff 29682, France

Submitted to: Protist

Abstract

A radiolarian order, Entactinaria, is defined by a specific skeletal structure called “initial spicular system”, and the oldest members of this group appeared in the Ordovician period. Extant entactinarians are supposed to be survivors of the most severe extinction event in the Earth, the Permian-Triassic Mass Extinction (PTME). However, our phylogenetic and morphological examinations revealed that entactinarians of today’s ocean are not monophyletic: Thalassothamnus should be classified under the subclass Phaeodaria of Cercozoa. Orosphaeridae are presumably related to the order Collodaria under Radiolaria. The families Hexalonchidae, Hexastylidae and Rhizosphaeridae would belong to the order Spumellaria. The results suggest that no true entactinarians survived the PTME, implying that the impact of this extinction event was more severe than previously expected.

Nakamura, Y., **Sandin, M.M.,** Suzuki N., Somiya R., Tuji A., Not, F. *Are they really survivors of the Permian-Triassic extinction? —phylogenetic revision of Paleozoic relict radiolarians.* (submited)

Molecular diversity of microalgae in symbiosis with Nassellaria and Spumellaria (Radiolaria)

Elsa Gadoin, **Miguel M. Sandin** and Fabrice Not

Master's thesis
Océanographie et Environments Marins, Sorbonne University

Abstract

Photosymbiosis, defined as the association of a photosynthetic partner with a heterotrophic host, plays a key role in marine ecosystems. Although it was widely studied in coral reefs, the diversity of symbionts associated with planktonic organisms remains unclear. However, many planktonic heterotrophic protists living in the photic layer, such as Radiolarian and Foraminifera, bear symbiotic microalgae. This mutualist association is essential for the maintenance of both partners, particularly in oligotrophic ecosystems. Radiolarians Nassellaria and Spumellaria are marine heterotrophic planktonic protists, belonging to the Rhizaria lineage, bearing symbiotic microalgae which diversity is poorly characterised. The use of the 18S gene as a genetic marker permitted the identification of Brandtodinium nutricula and Gymnoxanthella radiolariae as the main symbiotic dinoflagellate clades within Nassellaria and Spumellaria. The acquisition of new sequences allowed better characterize the intraspecific diversity of these two clades. The results of this study highlighted the low specificity of interactions between these two clades of symbionts and their different of nassellarians and spumellarians hosts. Brandtodinium and Gymnoxanthella are newly identified clades and their diversity and geographical distribution are still poorly characterized. The use of more specific genetic markers (ITS, 16S rDNA) could more specifically determine the internal diversity of these clades. In addition, the geographical distribution of these dinoflagellates, in a free-living or symbiotic stage, could be determined using detection methods such as fluorescent in situ hybridization (FISH).

He who possesses liberty otherwise than as an aspiration possesses it soulless, dead. One of the qualities of liberty is that, as long as it is being striven after, it goes on expanding.

Henrik Ibsen, 17 February 1871

Summary

Nassellaria and Spumellaria (Polycystines, Radiolaria) are planktonic amoeboid protists belonging to the Rhizaria lineage. They are widely distributed and abundant in the global ocean. Their silicified skeleton preserves very well in sediments, displaying an excellent and continuous fossil record dating back to the early Cambrian. Radiolarian fossil record is extremely valuable for paleo-environmental reconstruction studies. Radiolaria are difficult to maintain in culture preventing an accurate perception of their extant diversity and ecology in contemporary oceans, and most of it is inferred from the fossil record and sediment samples. Despite recent effort, their taxonomy and evolutionary history remains poorly known and controversial.

Here I explored the diversity and evolutionary patterns of Nassellaria and Spumellaria, based on a single cell integrative classification obtained from taxonomic marker genes (18S and 28S ribosomal DNA) and morphological characteristics. Our phylogenetic analyses established a morpho-molecular framework partly agree with the latest classifications relying essentially on the overall symmetry of the skeleton at Superfamily and Family level. This comprehensive morpho-molecular framework was integrated with recent phylogenetic studies of Acantharia and Collodaria in order to reconstruct the most extensive rDNA phylogenetic analysis of Radiolaria to date. The integrative classification of Radiolaria established the Acantharia, with Strontium sulphate skeleton, a sister clade of the Taxopodida and the Polycystines, both with silicate skeletons. Radiolarian evolutionary patterns were explored using a fossil calibrated molecular clock dating an origin of Radiolaria in the Early Neoproterozoic. Two major events characterized their diversification, the development of the skeleton in the Early Paleozoic and the establishment of the symbiosis in the Middle to Upper Jurassic, when oligotrophy and anoxia governed the oceans. A large environmental diversity was found associated with basal nodes, that following the morpho-molecular framework and evolutionary patterns led to hypothesize a large skeleton-less diversity related to Taxopodida (Rad-B) at basal positions in the radiolarian phylogenetic tree.

The intracellular genetic variability of Nassellaria and Spumellaria was explored finding an important taxonomic bias in both the variability and number of sequences when dealing with short read High-Throughput Sequencing (HTS) output. Sequencing platforms, such as Oxford Nanopore Technologies, provided interesting results for the full rDNA sequences, despite their high error rate. Our analyses allowed a better understanding of the global biodiversity and biogeography of Radiolaria, that was later explored through a metabarcoding approach across samples collected globally during Tara Oceans and Malaspina expeditions and regionally compared with the MOOSE-GE cruises in the western Mediterranean-sea. Radiolaria contributed about 9% of the total eukaryotic reads in the studied datasets. The colonial Collodaria was the group more abundant in large size fractions, adapted to oligotrophic and surface waters. Acantharia took the leading role in smaller size fractions and more productive waters. Spumellaria dominated the mesopelagic and bathypelagic followed by a big importance of environmental diversity believed to be associated to the skeleton-less Radiolaria.

This work brings a new comprehensive perspective of the evolutionary relationships and diversity of extant Radiolaria, highlighting their important planktonic role in both contemporary and past oceans.

Keywords: Nassellaria, Spumellaria, Radiolaria, Phylogenetics, Molecular evolution, Molecular clock, Metabarcoding