ORIGINAL PAPER

Towards an Integrative Morpho-molecular Classification of the Collodaria (Polycystinea, Radiolaria)

Tristan Biard^{a,b,1}, Loïc Pillet^{a,b}, Johan Decelle^{b,c}, Camille Poirier^d, Noritoshi Suzuki^e, and Fabrice Not^{a,b}

^aSorbonne Universités, Université Pierre et Marie Curie - Paris 6, UMR 7144, Team Diversity and Interactions in Oceanic Plankton, Station Biologique de Roscoff, CS90074, 29688 Roscoff cedex, France

^bCNRS, UMR 7144, Laboratoire Adaptation et Diversité en Milieu Marin, Place Georges Teissier, CS90074, 29688 Roscoff cedex, France

^cSorbonne Universités, Université Pierre et Marie Curie - Paris 6, UMR 7144, Team Evolution of Plankton and Pelagic Ecosystems, Station Biologique de Roscoff, CS90074, 29688 Roscoff cedex, France

^dMonterey Bay Aquarium Research Institute, 7700 Sandholdt road, Moss Landing CA 95039, United States

^eInstitute of Geology and Paleontology, Graduate School of Science, Tohoku University, 6-3 Aoba, Aramaki, Aoba-ku, Sendai, 980-8578 Japan

Submitted December 16, 2014; Accepted May 5, 2015

Monitoring Editor: Hervé Philippe

Running title: Morpho-molecular Classification of Collodaria

---

¹Corresponding author; fax +33 2 98 29 23 23
e-mail tristan.biard@sb-roscoff.fr (T. Biard).

Collodaria are ubiquitous and abundant marine radiolarian (Rhizaria) protists. They occur as either large colonies or solitary specimens, and, unlike most radiolarians, some taxa lack silicified structures. Collodarians are known to play an important role in oceanic food webs as both active predators and hosts of symbiotic microalgae, yet very little is known about their diversity and evolution. Taxonomic delineation of collodarians is challenging and only a few species have been genetically characterized. Here we investigated collodarian diversity using phylogenetic analyses of both nuclear small (18S) and large (28S) subunits of the ribosomal DNA, including 124 new sequences from 75 collodarians sampled worldwide. The resulting molecular phylogeny was compared to morphology-based classification. Our analyses distinguished the monophyletic clade of skeleton-less and spicule-bearing Sphaerozoidae from the sister clades Collosphaeridae (skeleton-bearing) and Collophidiidae (skeleton-less), while the Thalassicollidae was not retrieved as a monophyletic clade. Detailed morphological examination with electron microscopy combined with molecular analyses revealed many discrepancies, such as a mix between solitary and colonial species, co-existence of skeleton-less and skeleton-bearing specimens within the Collosphaeridae, as well as complex intraspecific variability in silicified structures. Such observations challenge a morphology-based classification and highlight the pertinence of an integrative taxonomic approach to study collodarian diversity.

Key words: Collodaria; integrative taxonomy; molecular phylogeny; polycystine; Radiolaria; single-cell.

Introduction

Radiolaria is a lineage of marine planktonic protists that appeared in the early Paleozoic and belongs to the supergroup Rhizaria (Adl et al. 2005; Nikolaev et al. 2004). Radiolarians are classified in five taxonomic orders, mainly distinguished by the morphology and composition of their mineral skeletons: strontium sulphate in Acantharia, silica in Taxopodida and in the polycystine Collodaria, Nassellaria and Spumellaria (Suzuki and Aita 2011). While many previous studies have investigated the past diversity and paleo-environmental signatures of radiolarians through their fossil record, less is known about the diversity and ecology of extant polycystines in marine ecosystems.

Collodarians are widespread in the oceans, exhibiting high abundances in calm and oligotrophic surface waters (Swanberg 1979). High densities of collodarian colonies have been reported in the Gulf of Aden (16,000 - 20,000 colonies per m³; Khmeleva 1967) and in the North Pacific Ocean (up to 30 colonies per m³; Dennett et al. 2002). In situ observations and culture experiments have described collodarians as active predators feeding on a broad range of prey (e.g. copepods, ciliates, phytoplankton or bacteria; Anderson 1978; Swanberg and Caron 1991), therefore contributing significantly to oceanic food webs. In addition to their heterotrophic behaviour, all known colonial collodarians harbour hundreds of endosymbiotic microalgae (Hollande and Enjume 1953), in most cases the dinoflagellate *Brandtodinium nutricula* (Probert et al. 2014). This makes collodarians significant contributors to primary production in oligotrophic surface waters (Michaels et al. 1995; Swanberg and Harbison 1980). The mixotrophic behaviour of collodarians, coupled with their wide distribution and abundance, emphasizes the ecological and biogeochemical significance of these uncultivated protists in oceanic waters (Anderson 1983; Dennett et al. 2002; Michaels et al. 1995). However, because of inadequate sampling and preservation procedures, they have often been neglected in environmental surveys and their ecological importance has very likely been underestimated.

Most Collodaria form colonies comprising tens to hundreds of individual radiolarian cells (i.e. central capsules) embedded in a gelatinous matrix that ranges from a few millimetres up to 3 meters long (Swanberg 1979). Large solitary species (i.e. one single radiolarian cell) of several millimetres have also been described within the Collodaria. Some species build a shell-like skeleton around their central capsule while others have siliceous spicules, similar to those in sponges, in the matrix and some lack mineral structures altogether. Taxonomically, collodarians have been grouped into three families based essentially on the presence or absence of colonial forms. The family Thalassicollidae is composed exclusively of solitary species that are classified based on the position of alveoli surrounding the central capsule or on the structure of the spicules (Anderson et al. 2002). The Collosphaeridae contains only colonial skeleton-bearing collodarians. As the siliceous shells of collodarians are preserved in sediments over large geological periods of time, micropaleontologists have described living and fossil species according to different features of the skeleton (shape and size; number and size of pores; position, size and number of openings; Strelkov and Reshtenyak 1971). The family Sphaeroidae includes skeleton-less and spicule-bearing colonial taxa. For the latter, spicule shape (size, number of radiate spines, presence of appendages) and positions are taxonomically informative characters (Brandt

1885; Haeckel 1887; Popofsky 1920). For the skeleton-less collodarians, taxonomic identification is more complex and transmitted light microscopy provides only very limited morphological information for their identification (Brandt 1885; Haeckel 1887). Swanberg (1979) noticed that the shapes of the colonies were not species-specific, but highlighted a set of morphological features (e.g. distribution of alveoli, gelatinous textures) that could potentially help in their identification. Fine structural analysis with transmission electron microscopy later allowed the separation of two skeleton-less colonial genera, *Collophidium* and *Collozoum*, within the Sphaerozoidae based on the shape of their central capsules and the density of cytoplasmic vacuoles (Anderson et al. 1999).

In light of the difficulty of achieving accurate morphological identification in uncultured protists, phylogenetic analyses with ribosomal genes have been shown to be valuable tools not only to assess the diversity and evolutionary patterns among these organisms, but also to define new taxonomic frameworks and identify life stages (Bachy et al. 2012; Decelle et al. 2012, 2013). For collodarians, phylogenetic studies based on the 18S rRNA gene have demonstrated that they form a distinct monophyletic group, included in the paraphyletic Nassellaria, and do not belong to the order Spumellaria as previously suggested (de Wever et al. 2001; Krabberød et al. 2011; Kunitomo et al. 2006). At the family level, the Collosphaeridae, Sphaerozoidae and Thalassicollidae form separate monophyletic clades, but their phylogenetic relationships remain unclear (Kunitomo et al. 2006; Zettler et al. 1999). Currently, with only 6 reference sequences, the family Thalassicollidae (i.e. the solitary species) constitutes the earliest diverging clade in most phylogenies (Ishitani et al. 2012; Krabberød et al. 2011; Kunitomo et al. 2006; Polet et al. 2004; Yuasa et al. 2005). The phylogenetic position of the Thalassicollidae and analyses of their nuclei led to the hypothesis that solitary cells represent the ancestral form for Collodaria (Suzuki et al. 2009). Recently, Ishitani et al. (2012) proposed a new phylogenetic scheme for the Collodaria, with a clade composed of sequences affiliated to the genus *Collophidium* and the informal proposition (International Code of Zoological Nomenclature rules were not respected) of a fourth collodarian family, the Collophidae. Yet, with only two sequences and weak support values, the robustness of this family remains to be assessed.

With only a few reference ribosomal sequences (22 18S rDNA and 2 28S rDNA sequences) available in GenBank (as of July 2014) relationships between collodarian families, genera and species that typically lack informative morphological characters, are still poorly resolved. In addition to clarifying these relationships, sequencing morphologically identified specimens collected in the environment will contribute to the construction of a robust

reference database of barcodes that will represent a powerful new tool to explore the ecology of collodarians at large spatio-temporal scales through metabarcoding approaches. The objective of the present study was to better understand the molecular diversity of Collodaria and the phylogenetic relationships between its taxa, by integrating morphological and molecular information.

Results

Molecular Phylogenies of Collodarian Families

A total of 75 collodarian specimens encompassing the four families, Sphaerozoidae, Collosphaeridae, Collophidiidae (formally described herein from the previously proposed Collophidae) and Thalassicollidae, were isolated from worldwide locations. All but nine of these specimens were colonial (Supplementary Material Fig. S1 and Table S1). From these specimens, 62 partial 18S rDNA and 62 partial 28S rDNA (D1-D2 regions) sequences were obtained (Supplementary Material Table S1). Phylogenetic analyses based on the concatenated dataset revealed 3 distinct clades, of which clades A and B, consisting of 18 and 11 sequences respectively, were the most closely related (Fig. 1). The third clade (clade C) contained most of the sequences ($n = 46$) and was basal to the other clades. Morphological identification of the analysed individuals showed that these monophyletic clades corresponded to three collodarian families: all skeleton-bearing (i.e. Collosphaeridae) occurred in clade A, the skeleton-less Collophidiidae grouped within clade B, while members of the Sphaerozoidae formed clade C. The fourth family of solitary Thalassicollidae was not retrieved as a monophyletic clade.

Within Collosphaeridae (Fig. 2), 18 novel and 3 previously available sequences revealed 6 distinct clades (A1 to A6) supported with moderate to high support values. Maximum values of ML bootstrap (BS) and posterior probability (PP) showed strong phylogenetic relationships between clades A4, A5 and A6, while weak support values suggested uncertain phylogenetic placement of clades A1, A2 and A3. The family Collophidiidae gathered 11 new and two reference sequences. Two clades (B1 and B2) contained 9 of the sequences whereas the other sequences could not be assigned to a specific clade. Both clades were supported with high support values but the phylogenetic relationships between them were not resolved. The Sphaerozoidae contained 46 new and 8 publicly

available sequences distributed in 11 clades (C1 to C11). Most of the clades were highly supported except for clades C2, C3 and C7 that had weak support, and clade C8 that was not supported at all. Phylogenetic relationships between these clades remain unclear.

Integrative Taxonomy of Collodaria

Based on morphological observations performed with light microscopy (LM) and Scanning Electron Microscopy (SEM) images, the Collosphaeridae mainly included skeleton-bearing specimens easily assigned to different genera according to the overall morphology of their shells: *Disolenia*, *Collosphaera* and *Siphonosphaera* (Figs 1 and 2). The *Disolenia* species had a polygonal shell with large irregular polygonal openings, 2 to 9 in number (Fig. 2A-G). All *Collosphaera* species had a crumpled sphere-like shell with polygonal to rounded pores (Fig. 2H, I). *Siphonosphaera* shells were mostly spherical with characteristic tubular projections pointing outwards (Fig. 2J). The molecular classification broadly matched morphological identifications, these three genera clearly belonging to different clades in the phylogenetic analyses (Fig. 1). However, these analyses also highlighted several discrepancies between molecular and morphology-based classifications. In 5 of the colonial specimens from clades A1, A2 and A5, no shells were observed in either LM or SEM images (Supplementary Material Fig. S1). Specifically, in clades A2 and A5 these skeleton-less forms were mixed with skeleton-bearing specimens and two of them (Pac 12 and Pac 17) were even genetically similar to one skeleton-bearing species (Sat 12). For some skeleton-bearing specimens, several shell morphotypes were found among genetically identical specimens or even within a single specimen. For instance, in specimens Sat 17 (Fig. 2B) and Sat 21 (Fig. 2D-D''), both identified as *Disolenia tenuissima*, two and three shell morphotypes coexisted, respectively. Finally, one specimen (Pac 3), a solitary Thalassicollidae morphologically identified as *Thalassicolla melacapsa*, was genetically placed among the 17 other Collosphaeridae specimens, which were exclusively colonial specimens.

In Collophidiidae (Fig. 1), all 10 colonial specimens were morphologically identified as belonging to the genus *Collophidium* (Supplementary Material Fig. S1). Clade B1 included an assemblage of different morpho-species while clade B2 was only composed of a single morphologically identified species (*C. serpentinum*). As for the Collosphaeridae, colonial and solitary specimens were mixed within a single clade. For instance, clade B1 included the solitary *Procyttarium prototypus*, that formally belong to the Thalassicollidae, and which was genetically identical to two colonial *Collophidium*. Regardless of this mix between colonial and solitary species, the Collophidiidae contained only naked specimens.

Seven genera from both solitary and colonial forms were represented in the Sphaerozoidae: the colonial *Collozoum*, *Rhaphidozoum* and *Sphaerouzoum*, and the solitary *Procyttarium*, *Thalassicolla*, *Thalassosphaera* and *Thalassophysa* (Fig. 1 and Supplementary Material Fig. S1). The skeleton-less genus *Collozoum* was highly polyphyletic as it was represented in clades C7, C8, C10 and C11. Although it was not supported, Clade C6 consistently gathered only the species *Collozoum inerme*. The clade was composed of a core of 5 sequences, supported with high support values, and 5 other sequences having a strong affinity for this core group. Several different morphologies (cell size, shape and organisation within the colony) were distinguished between clades, but affiliation to known species was unclear. The specimens forming clades C1-C6 and C9 were mainly spicule-bearing Sphaerozoidae (Fig. 3). Cell shrinkage caused by ethanol preservation disrupted the arrangement of spicules inside the colony leading to challenging taxonomical identification based on LM observation (Supplementary Material Fig. S1). The two C9 specimens representing the species *Rhaphidozoum acuferum* had clearly identifiable needle-like spicules (Fig. 3A, B), but the relation between clade C9 and other spicule-bearing clades was unclear. Clades C1-C6 contained the other spicule-bearing specimens, all belonging to the genus *Sphaerouzoum*, recognizable by their double tri-radiate spicules (Fig. 3C-K). Based on SEM images, detailed examination of all specimens in clade C6 (identified as *S. punctatum*) revealed numerous small spicules and a few larger ones (Fig. 3H, H'). Differences in spicule structure was also observed in other *Sphaerouzoum* species and ranged from a more complex spinose pattern to the appearance of a fourth radiate axis (Supplementary Material Fig S2). As for the two other families, while 39 specimens of the Sphaerozoidae were colonial, clades C3, C6, C10 and C11 also included solitary species from the Thalassicollidae (Fig. 1 and Supplementary Material Fig. S1).

Environmental Diversity of Collodaria

To evaluate the robustness of the coverage of our taxonomic sampling, the 92 environmental sequences affiliated to Collodaria and currently available in GenBank (July 2014) (Supplementary Material Table S2) were placed in the phylogenetic tree composed of our reference sequences (Fig. 4). The majority of environmental sequences (91%) grouped within clade B (Collophidiidae). More detailed placements mapped 1 and 2 sequences into clades B1 and B2, respectively. A majority of sequences (81) could not be precisely assigned to any known clade of Collophidiidae. Collosphaeridae represented 6.5% of the environmental

sequences (3 in clade A5, 1 in A1, and 2 not precisely assigned). The last family, Sphaerzoidae, only included one environmental sequence affiliated to clade C8.

Discussion

Morpho-Molecular Classification and Evolution of Collodaria

The morpho-molecular approach used in this study provides new insights into collodarian diversity and challenges the traditional classification and hypotheses on the evolutionary history of the group. From our phylogenetic analysis we recovered the monophyly of three of the four collodarian families (Collosphaeridae, Collophidiidae and Sphaerzoidae), while members of the polyphyletic Thalassicollidae were scattered throughout the tree (Fig. 1). The Thalassicollidae was historically created to group solitary specimens, but the co-occurrence we observed between solitary taxa and colonial species highlights a major discrepancy between molecular and morphological classification, and questions the validity of this character as a reliable taxonomic marker. Previous studies supported a phylogenetic separation of solitary and colonial species (Polet et al. 2004; Zettler et al. 1999), yet none of our solitary specimens, nor any environmental sequences, clustered with the previously described Thalassicollidae clade (data not shown). In addition, careful analysis of the DNA sequences of the Thalassicollidae reveals that the distinction and monophyly of this clade was essentially based on large ambiguous regions in the 18S rRNA gene (see Methods section). Although we cannot precisely determine their origin, these ambiguous regions may correspond to pseudo-genes, which have been identified in some marine protists and are generally highly divergent from native ribosomal sequences (Santos et al. 2003; Thornhill et al 2007). The inconsistency of the Thalassicollidae family in our analyses questions its basal position in the Collodaria and highlights the need to reconsider the hypothetical evolutionary history stating that Collodaria evolved from solitary Thalassicollidae to colonial Sphaerzoidae and then to skeleton-bearing colonial Collosphaeridae (Suzuki et al. 2009).

In our analyses, colonial and solitary specimens were genetically close or even identical in each of the three families, suggesting that they represent the same taxonomic entity and could be two distinct phases of the same life cycle. Knowledge about the life cycle in Collodaria is very limited and relies mainly on old studies. Several solitary forms have already been associated to colonial species (e.g. *Thalassophysa sanguinolenta* and *Collozoum*

pelagicum) (Brandt 1902). This association was later supported with phylogenetic analyses of both solitary and colonial forms (Polet et al. 2004; Zettler et al. 1999). Investigating reproductive mechanisms in Collodaria, Hollande and Enjumet (1953) pointed out the ability of solitary specimens to create “proto-colonies” through the simple budding of the cell. The release of flagellate swimmers was also observed in both colonial and solitary collodarian species (Anderson 1976, 1978; Hollande and Enjumet 1953), but the fate of these swimmers remains unknown to date.

The phylogenetic position of the Collophidiidae (clade B) as a sister-clade of the Collosphaeridae (clade A) is congruent with a previous phylogenetic study (Ishitani et al. 2012). The phylogenetic analyses clearly discriminate the Collophidiidae from the other skeleton-less family (i.e. Sphaerozoidae) as shown in Figure 1. The colonial genus *Collophidium*, originally included in the Sphaerozoidae, showed major ultrastructural and molecular differences with the genus *Collozoum* (Anderson et al. 1999; Zettler et al. 1999). Molecular divergence between *Collophidium* and the Sphaerozoidae was later suggested (Ishitani et al. 2012). Here, following the rule of the International Code of Zoological Nomenclature (ICZN 1999), we formally propose the erection of a new family: Collophidiidae Biard et Suzuki fam. nov. (with *Collophidium* as the type genus, *Collophidium serpentinum* as the type species; see taxonomic appendix below).

Collodarian shells have been suggested to have originated from the fusion of spicules in a spicule-bearing ancestor (Anderson and Swanberg 1981; Strelkov and Reshetnyak 1971). This hypothesis was later rejected because of the absence of congruent fossil records to link Sphaerozoidae and Collosphaeridae (Bjørklund and Goll 1979). In our morpho-molecular analyses, there is a clear separation between all skeleton-bearing (Collosphaeridae) and all spicule-bearing specimens (Sphaerozoidae) (Fig. 1). This suggests that skeleton-bearing and spicule-bearing collodarians have co-diversified and that the skeleton is not a derived character. Our results also rule out the possibility of the presence of both spicules and a shell during the life cycle of collodarians.

The Need for an Integrative Taxonomic Approach

In the present analyses, the monophyletic Sphaerozoidae (Fig. 1) contains three genera, *Collozoum*, *Sphaerozoum* and *Rhaphidozoum*, which is congruent with the traditional taxonomical scheme (Anderson 1976; Strelkov and Reshetnyak 1971). The high polyphyly of the genus *Collozoum*, which is distributed in clades C7, C8, C10 and C11, suggests that specific morphological characters need to be revised (Fig. 1). The name *Collozoum inerme*

has often been used as generic species name for nearly all skeleton-less collodarians, although in Haeckel's descriptions (1862) it is characterized by the presence of an oil-like droplet in the central capsule of the cell. In our specimens, such a structure was only found for members of clade C8, subsequently identified as *Collozoum inerme*. Specimens found in clade C7 and C10 presented distinct morphological features compared to clade C8, having more algal symbionts, a transparent layer around the endoplasm and spherical collodarian cells. Whether these morphological features are taxonomically reliable characters, implying that they are "fixed" during the life cycle and in different environmental conditions, remains to be determined.

In the Collosphaeridae, the co-existence of both skeleton-less and skeleton-bearing specimens occurred in two clades (Fig. 1). We were unable to identify the skeleton-less stages following the taxonomical scheme for Collosphaeridae. Haeckel (1862) first pointed out the absence of a skeleton in many colonies and the existence of skeleton-less cells during the early developmental stages of Collosphaeridae, which was later confirmed through culture attempts (Anderson and Gupta 1998; Anderson and Swanberg 1981). Since the traditional taxonomical scheme to identify Collosphaeridae is based on skeleton morphology, Anderson and Swanberg (1981) mentioned the impossibility to identify such cells in early stages of skeletal development. The existence of skeleton-less cells within a skeleton-bearing clade challenges morphological identification and emphasizes the requirement for novel morphological characters as well as molecular tools for accurate identification of different life stages.

In both Collosphaeridae and Sphaerzoidae, detailed morphological examinations revealed intraspecific variability in silicified structures (i.e. shell and spicules) (Figs 2 and 3, and Supplementary Material Fig. S2). Several shell morphotypes were found among genetically identical specimens or even within a single specimen (i.e. *Disolenia tenuissima*) (Fig. 2B-D''). Originally described with one or two, and occasionally three large openings (Hilmers 1906), *D. tenuissima* exhibited variations in shell pores that have been previously reported for other skeleton-bearing collodarians, and likely explained by the ontogeny of the silicified structure (Anderson and Swanberg 1981). These observations challenge Haeckel's classification based on shell features, as the existence of different morphotypes leads to uncertain taxonomic identification. As for the Collosphaeridae, several spicule-bearing Sphaerzoidae exhibited intraspecific variability of their spicules. The morphological taxonomic criteria classically used to discriminate between *Sphaerzoum* species include spinose patterns, arrangement and number of spicules, and the development state of an oil-

like droplet inside each cell (Anderson 1983). The co-existence of small and rare larger spicules has already been reported within colonies of *S. punctatum* and *S. fuscum* (Brandt 1881; Popofsky 1920). Whereas differences in spicule thickness can be explained by ontogeny (Brandt 1881), the range of variability in spicules encountered (different sizes or shapes) suggests more complex processes. Fusion between colonies, likely of the same species, is regularly observed (Hollande and Enjumet 1953; pers. observ.), but it is not clear whether these physical associations are temporary during the life cycle, or represent reproductive mechanisms, predatory relationships, or simply an experimental artefact (Hollande and Enjumet 1953; Huth 1913; Swanberg 1979). Therefore, we cannot exclude the possibility that a mix of shells or spicules within one specimen could be the consequence of such fusion, potentially blurring morphology-based taxonomic identifications.

Insights into the Diversity of Collodaria in the Environment

The present morpho-molecular framework for Collodaria allowed a more accurate taxonomic placement of sequences derived from environmental molecular diversity surveys available in public databases (Fig. 4 and Supplementary Material Table S2). Environmental sequences belonged to the three families highlighted previously (i.e. Collosphaeridae, Collophidiidae, Sphaerzoidae). Whereas most of the environmental sequences belonged to the Collophidiidae, the most represented family in our study was the Sphaerzoidae. This discrepancy could be explained by the fact that all specimens used in the present study were collected in surface waters and that most environmental sequences come from the entire water column in the Cariaco Basin, a very unusual ecosystem, or from deep samples (Edgcomb et al. 2011; Jungbluth et al. 2013; Sauvadet et al. 2010; Supplementary Material Table S2). Consequently, our sampling strategy could have introduced a bias against Collophidiidae, or other specific taxa, if they have marked ecological preferences impacting their distribution. A recent global environmental diversity survey based on metabarcodes of the V9 region of the 18S rRNA gene highlighted very high abundance, diversity and distribution of collodarian reads in the world oceans but could not identify the distribution of specific taxa (de Vargas et al. in press). Together with appropriate contextual data, the present morpho-molecular database sets the basis for a better understanding of the phylogeography and ecology of individual collodarian taxa.

Taxonomic Appendix

Family Collophidiidae Biard et Suzuki n. fam.

Type genus: *Collophidium* Haeckel, 1887 (type species: *Collozoum (Collophidium) serpentinum* Haeckel, 1887, designated by Campbell, 1954; raised to genus level by Anderson et al. 1999).
 Synonymy: Collophidae in Ishitani et al. (2012) [unavailable name; see remarks].
 Definition: Colonial Collodaria with a variable overall appearance of elongate, cylindrical, spherical form. Each colony comprises delicate gelatinous material encompassing scattered algal symbionts and string-like aggregations. A string-like aggregation consists of tens to a hundred collodarian cells in firm gelatinous material. This string-like aggregation ranges from several millimetres to several centimetres in length and approximate 0.5 mm in width. Algal symbionts are always distributed throughout fragile gelatinous material but not in string-like aggregations. The number of algal symbionts in a colony is variable. Each collodarian cell displays opaque inner protoplasm with surrounding transparent protoplasm. Both opaque inner and surrounding transparent protoplasm are defined as the endocapsulum. The outer transparent protoplasm consists of a layer of vacuoles whereas the inner opaque protoplasm contains organelles such as Golgi, nucleus, and mitochondria.
 Remarks: Molecular phylogenetic study performed by Ishitani et al. (2012) informally reported the separation of *Collophidium* at the family level. Here we formally established the family Collophidiidae following the rules of the International Code of Zoological Nomenclature (ICZN 1999).

Methods

Sample collection: Plankton samples were collected in the bay of Villefranche-sur-Mer (France, 43°41'10'' N, 7°18'50'' E) using a Regent net (680 µm mesh size), off Sesoko Island, Okinawa (Japan, 26°37'20'' N, 127°52'15'' E) by net tows (20 and 150 µm mesh sizes) and during the TARA Oceans expedition using a Bongo net (180 µm mesh size) or a hand net (Supplementary Material Table S1). Specimens were immediately handpicked individually from the plankton samples with autoclaved micropipettes, transferred into clean beakers filled with 0.2-1 µm filtered seawater and incubated at 18 °C. After 3-4 hours the specimens had self-cleaned and were transferred individually into clean containers filled with 0.2 µm filtered seawater. Images were then taken under a binocular microscope or an inverted microscope for higher magnification pictures. Each specimen was finally rinsed three times in 0.2 µm filtered seawater to avoid contamination and transferred into 1.5 ml Eppendorf tubes containing 50-150 µl absolute ethanol or 50 µl of Guanidine Isothiocyanate (GITC). Isolated specimens were stored at -20 °C until DNA extraction. Each isolated specimen was identified based on morphological criteria and images were compared to the original monographs.

Tentative affiliations to known species were carried out following the original descriptions of Brandt (1885), Haeckel (1887), Popofsky (1920), Strelkov and Resthetnyak (1971). Detailed information related to each of the samples used in this study can be found in the RENKAN database at <http://renkan.sb-roscoff.fr>.

DNA extraction, amplification and sequencing: Genomic DNA from ethanol fixed specimens was extracted using the MasterPure Complete DNA Purification kit (Epicentre) following the manufacturers instructions. For GITC fixed specimens, extraction was performed as described in Decelle et al. (2012). Each specimen was extracted individually in an autoclaved microtube to avoid cross-contamination between samples. Pellet debris from DNA extraction were eluted in milliQ water to recover collodarian skeletons and spicules, and subsequently stored at -20 °C. Complete Small SubUnit (18S-SSU) and partial (D1 - D2 regions) Large SubUnit (28S-LSU) of the rDNA were amplified using Collodaria-specific primers (Table 1). The sets of primers SA/S81col and S32col/V9R amplify the first 1500 positions and last 1200 positions of the complete 18S rRNA gene, respectively. The set of primers 28S Col-F/ITSa3 Col amplifies 640 positions of the D1-D2 regions of the 28S rRNA gene. PCR reaction mix contained 8.75 µl of sterile water, 1 µl of each primer (10 µM concentration), 0.75 µl dimethyl sulfoxide (DMSO), 12.5 µl of the Phusion® High-Fidelity PCR Master Mix (Finnzymes) and 1 µl of DNA template, in a 25 µl final volume. PCR amplification conditions were: 30 s initial denaturation at 98 °C followed by 35 cycles of 10 s denaturation at 98 °C, 30 s of annealing at 53-58 °C according to the primers set, 72 °C for 30 s elongation and a 10 min final elongation step at 72 °C. All PCR amplifications were conducted in a PCR workstation, using autoclaved microtubes and molecular grade water. Amplified products were visualized on a 1.5% agarose gel stained with ethidium bromide. Successfully amplified products were sent for sequencing at the GENOSCOPE (France). Sequencing was also performed locally on an ABI-PRISM 3100 Genetic Analyzer using the ABI BigDye Terminator v3.1 kit (Applied Biosystems) after a step of purification using the NucleoSpin Gel and PCR Clean-up kit (Macherey Nagel) following the manufacturer's instructions.

Phylogenetic analyses: Sequence quality was carefully checked on the chromatograms using the 4Peaks software. The forward and reverse fragments were then assembled using Seaview version 4.4.2. (Gouy et al. 2010) and the presence of chimeras was verified using the online Key DNA Tools (<http://keydnatools.com/>). All valid sequences were then compared to reference sequences using the BLAST tool and similar sequences identified on GenBank were integrated into our 18S and 28S databases. Careful examination of the

Thalassicollidae 18S rDNA reference sequences AF057741, AF057742, AF057743, AF057744, AF018160 and AY266297 revealed two ambiguous regions in all of them, located at positions 59-274 base pairs (bp) and 599-819 bp, totally different (0% identity) from all other collodarian sequences (except themselves) and from any other sequence published in GenBank. When removing these ambiguous regions from our alignment, these 5 sequences formed a clade where none of our single-cell sequences, neither any environmental sequences were retrieved (data not shown). These ambiguous 18S rDNA sequences were therefore removed from the subsequent phylogenetic analyses. According to previous studies, six nassellarian sequences (AB430759, AF382824, DQ386169, FJ032682, FJ032683, HQ651779) were used as outgroup (Krabberød et al. 2011). For each dataset, 18S (78 taxa; 1744 positions) and 28S (75 taxa; 666 positions), sequences were aligned using MAFFT version 7.213 (Kuraku et al. 2013) and ambiguous positions were manually removed. After the removal of ambiguous positions, 18S and 28S dataset included 1311 and 307 positions, respectively. Prior to phylogenetic analyses, the Perl script MrAIC 1.4.3 (Nylander 2004) in combination with PHYML v2.4.4 (Guindon and Gascuel 2003) was used to choose the best model of sequence evolution by the Akaike Information Criterion (AIC). Phylogenetic analyses were performed (as described below) independently on each gene marker (18S and 28S rDNA) and obtained topologies were compared in order to make sure that no paralogy issues exist and the two genes can be concatenated. As suggested in Decelle et al. 2012 or Krabberød et al. 2011, concatenation of both 18S and 28S rRNA genes in radiolarians phylogenies increase the phylogenetic resolution. Both 18S rDNA and 28S rDNA alignments were therefore concatenated using Sequence Matrix version 1.7.8. (Vaidya et al. 2011). Applying the obtained settings (GTR + Γ model) a Bayesian Inference (BI) method and a Maximum Likelihood (ML) method (Felsenstein 1981) were used to infer phylogeny. With the MrBayes program (Huelsenbeck and Ronquist 2001), two independent analyses were performed at the same time with four simultaneous chains (one cold and three heated) ran for 10 million generations, and sampled every 1000 generations. After discarded 2000 of the initial trees as burn-in, the consensus tree with the corresponding posterior probabilities was calculated for each data set. The ML method was implemented with the PhyML v3.0 software (Guindon et al. 2010) and the reliability of internal branches was assessed using the bootstrap method with 100 replicates (Felsenstein 1985). In parallel, another Bayesian analysis was performed using the PhyloBayes software (Lartillot et al. 2009), in order to test the CAT-GTR model of sequences evolution (Lartillot and Philippe 2004), allowing heterogeneity across sites and which has been shown to potentially improve phylogenetic inference (Tsagkogeorga

et al. 2009). This latter approach did not significantly improve the final topology and phylogenetic analyses inferred using the GTR + Γ model of sequences evolution were used. Bootstrap supports (BS) and posterior probabilities (PP) were associated to each node in the Bayesian topology. Final tree was visualized and edited in Fig Tree v 1.4.0. (Rambaut 2010). All sequences generated in the present study have been deposited in the GenBank database under accession numbers KR058196-KR058319. All alignments used in this study can be found online in the RENKAN database at <http://renkan.sb-roscoff.fr>.

Environmental diversity of Collodaria: For each family, closely related environmental 18S rDNA sequences available in GenBank (July 2014) were selected using BLAST in order to infer the environmental genetic diversity of Collodaria (Supplementary Material Table S2). Positions of these environmental sequences in our reference phylogenetic tree were determined using the pplacer software (Matsen et al. 2010) as earlier described in Decelle et al. (2012).

Scanning electron microscopy (SEM) observation: Eluted pellet debris containing collodarian skeletons or spicules recovered after DNA extractions were vortex mixed and sorted using an inverted microscope. Several skeletons and spicules were handpicked for each specimen and finally transferred into 0.2 ml Eppendorf tubes containing 50 μ l hydrogen peroxide. Each tube was heated at 70 °C for 10 min to remove residual organic matter. Several clean skeletons or spicules were then handpicked and placed on 0.2 μ m pore-size polycarbonate membrane filters, dried and stuck on SEM pin stub mounts. Filters were imaged with an FEI Phenom tabletop SEM (FEI Technologies).

Acknowledgements

This study was supported by the "DESIR" Project Emergence-UPMC from Université Pierre et Marie Curie, the JST-CNRS exchange program to Dr. Not and Dr. Suzuki, the Swiss National Science Foundation grant P2GEP3_148800, the "Bibliothèque du Vivant" network funded by the CNRS, the Muséum National d'Histoire Naturelle, the INRA and the CEA (Centre National de Séquençage). We are grateful to the Biogenouest Genomics and Genomer plate-forme core facility for its technical support. We would like to thank staff members of the Laboratoire d'Océanographie de Villefranche-sur-Mer (UPMC-CNRS, in particular John Dolan and Fabien Lombard), of the Station Biologique de Roscoff (UPMC-CNRS, in particular Sarah Romac) and of the Sesoko Marine Station (University of Ryukyus) as well as

the *Tara*-Oceans Expedition for providing sampling facilities. We are grateful to Ian Probert for valuable comments on English language.

References

Adl SM, Simpson AGB, Farmer MA, Andersen RA, Anderson OR, Barta JR, Bowser SS, Brugerolle G, Fensome RA, Fredericq S, James TY, Karpov S, Kugrens P, Krug J, Lane CE, Lewis LA, Lodge J, Lynn DH, Mann DG, Mccourt RM, Mendoza L, Moestrup Ø, Mozley-Standridge SE, Nerad TA, Shearer CA, Smirnov AV, Spiegel FW, Taylor MFJR (2005) The new higher level classification of eukaryotes with emphasis on the taxonomy of protists. *J Eukaryot Microbiol* **52**: 399–297

Anderson OR (1976) Fine structure of a collodarian radiolarian (*Sphaerozoum punctatum* Müller 1858) and cytoplasmic changes during reproduction. *Mar Micropaleontol* **1**:287–297

Anderson OR (1978) Light and electron microscopic observations of feeding behavior, nutrition, and reproduction in laboratory cultures of *Thalassicolla nucleata*. *Tissue Cell* **10**:401–412

Anderson OR (1983) Radiolaria. Springer-Verlag, New York, 363 p

Anderson OR, Gupta SM (1998) Evidence of binary division in mature central capsules of a collosphaerid colonial radiolarian: implications for shell ontogenetic patterns in modern and fossil species. *Paleontol Electron* **1**:1–13

Anderson OR, Swanberg NR (1981) Skeletal morphogenesis in some living collosphaerid radiolaria. *Mar Micropaleontol* **6**:385–396

Anderson OR, Gastrich MD, Zettler LA (1999) Fine structure of the colonial radiolarian *Collozoum serpentinum* (Polycystinea: Spumellaria) with a reconsideration of its taxonomic status and re-establishment of the genus *Collophidium* (Haeckel). *Mar Micropaleontol* **36**:81–89

- Anderson OR, Nigrini C, Boltovskoy D, Takahashi K, Swanberg NR** (2002) Class Polycystina. In Lee JJ, Leedale GF, Bardbury P (eds) Illustrated guide to the Protozoa. 2nd ed., Society of Protozoologists, Lawrence, pp 944–1022
- Bachy C, Gómez F, López-García P, Dolan JR, Moreira D** (2012) Molecular phylogeny of tintinnid ciliates (Tintinnida, Ciliophora). *Protist* **163**:873–887
- Bjørklund KR, Goll RM** (1979) Internal skeletal structures of *Collosphaera* and *Trisolenia*: a case of repetitive evolution in the Collosphaeridae (Radiolaria). *J Paleontol* **53**:1293–1326
- Brandt K** (1881) Untersuchungen an Radiolarien. Monatsber Kgl Preuss Akad Wiss Berlin Jahrg **1881**:388–404
- Brandt K** (1885) Die koloniebildenden Radiolarien (Spherozoëen) des Golfes von Neapel und der angrenzenden Meeresabschnitte. Monogr Fauna Flora Golf Neapel **13**:1–276
- Brandt K** (1902) Beiträge zur Kenntnis der Colliden. *Arch Protistenkd* **1**:59–88
- Brandt K** (1905) Zur Systematik der koloniebildenden Radiolarien. *Zool Jahrb (Suppl.)* **8**:311–352
- de Vargas C, Audic S, Henry N, Decelle J, Mahé F, Logares R, Lara E, Berney C, Le Bescot N, Probert I, Carmichael M, Poulain J, Romac S, Colin S, Aury JM, Bittner L, Chaffron S, Dunthorn M, Engelen S, Flegontova O, Guidi L, Horák A, Jaillon O, Lima-Mendez G, Lukeš J, Malviya S, Morard R, Mulot M, Scalco E, Siano R, Vincent F, Zingone A, Dimier C, Picheral M, Searson S, Kandels-Lewis S, Tara Oceans Coordinators, Acinas SG, Bork P, Bowler C, Gorsky G, Grimsley N, Hingamp P, Iudicone D, Not F, Ogata H, Pesant S, Raes J, Sieracki ME, Speich S, Stemmann L, Sunagawa S, Weissenbach J, Wincker P, Karsenti E** (in press) Eukaryotic plankton diversity in the sunlit ocean. *Science* (in press).
- de Wever P, Dumitrica P, Caulet JP, Nigrini C, Caridroit M** (2001) Radiolarians in the Sedimentary Record. Gordon and Breach Science Publishers, Amsterdam, 533 p

Decelle J, Suzuki N, Mahé F, de Vargas C, Not F (2012) Molecular phylogeny and morphological evolution of the Acantharia (Radiolaria). *Protist* **163**:435–450

Decelle J, Martin P, Paborstava K, Pond DW, Tarling G, Mahé F, de Vargas C, Lampitt R, Not F (2013) Diversity, ecology and biogeochemistry of cyst-forming Acantharia (Radiolaria) in the oceans. *PLoS ONE* **8**:e53598

Dennett MR, Caron DA, Micheals AF, Gallager SM, Davis CS (2002) Video plankton recorder reveals high abundance of colonial Radiolaria in surface waters of the central North Pacific. *J Plankton Res* **24**:797–805

Edgcomb VP, Kysela DT, Teske A, Gomez AV, Sogin ML (2002) Benthic eukaryotic diversity in the Guaymas Basin hydrothermal vent environment. *Proc Natl Acad Sci USA* **99**:7658–7662

Edgcomb V, Orsi W, Bunge J, Jeon S, Christen R, Leslin C, Holder M, Taylor GT, 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

Felsenstein J (1981) Evolutionary trees from DNA sequences: maximum likelihood approach. *J Mol Evol* **17**:368–376

Felsenstein J (1985) Confidence limits on phylogenetics: an approach using the bootstrap. *Evolution* **39**:783–791

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

Guindon S, Gascuel O (2003) A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst Biol* **52**:694–704

Guindon S, Dufayard JF, Lefort V, Anisimova M, Hordijk W, Gascuel O (2010) New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst Biol* **59**:307–321

- Haeckel E** (1862) Die Radiolarien (Rhizopoda Radiaria): Eine Monographie. Reimer, Berlin, 572 p
- Haeckel E** (1887) Report on the Radiolaria collected by H.M.S. Challenger during the years 1873–1876. Report on the Scientific Results of the Voyage of H.M.S. Challenger during the Years 1873–1876. Zoology **18**:1–1803
- Hilmers K** (1906) Zur Kenntnis der Collosphaeriden. Inaugural-Dissertation zur Erlangung der Doktorwürde der hohen philosophischen Fakultät der Königlichen Christian-Albrecht-Universität in Kiel. Druck von C. Schaidt, Kiel, 93 p
- Hollande A, Enjumet M** (1953) Contribution à l'étude biologique des Sphaerocollides (Radiolaires Collodaires et Radiolaires polycyttaires) et de leur parasites. Ann Sci Nat Zool **15**:99–183
- Huelsenbeck JP, Ronquist F** (2001) MRBAYES: Bayesian inference of phylogenetic trees. Bioinformatics **17**:754–755
- Huth W** (1913) Zur Entwicklungsgeschichte der Thalassicollen. Arch Protistenkd **30**:1–124
- ICZN (International Commission on Zoological Nomenclature)** (1999) International Code of Zoological Nomenclature, Fourth Edition. International Trust for Zoological Nomenclature, London, 306 p
- Ishitani Y, Ujiie Y, de Vargas C, Not F, Takahashi K** (2012) Phylogenetic relationships and evolutionary patterns of the order Collodaria (Radiolaria). PLoS ONE **7**:e35775
- Jungbluth SP, Grote J, Lin HT, Cowen JP, Rappé MS** (2013) Microbial diversity within basement fluids of the sediment-buried Juan de Fuca Ridge flank. ISME J **7**:161–172
- Khmeleva NN** (1967) Role of the radiolarians in the value of primary productivity in the Red Sea and Gulf of Aden. Dokl Akad Nauk SSSR **172**:1430–1433

- Krabberød AK, Brate J, Dolven JK, Ose RF, Klaveness D, Kristensen T, Bjorklund KR, Schachian-Tabrizi K** (2011) Radiolaria divided into Polycystina and Spasmaria in combined 18S and 28S rDNA phylogeny. *PLoS ONE* **6**:e23526
- 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
- Kuraku S, Zmasek CM, Nishimura O, Katoh K** (2013) aLeaves facilitates on-demand exploration of metazoan gene family trees on MAFFT sequence alignment server with enhanced interactivity. *Nucleic Acids Res* **41**:W22–W28
- Lartillot N, Philippe H** (2004) A Bayesian mixture model for across-site heterogeneities in the amino-acid replacement process. *Mol Biol Evol* **21**:1095–1109
- Lartillot N, Lepage T, Blanquart S** (2009) PhyloBayes 3: a Bayesian software package for phylogenetic reconstruction and molecular dating. *Bioinformatics* **25**:2286–2288
- López-García P, Philippe H, Gail F, Moreira D** (2003) Autochthonous eukaryotic diversity in hydrothermal sediment and experimental microcolonizers at the Mid-Atlantic Ridge. *Proc Natl Acad Sci USA* **100**:697–702
- Matsen FA, Kodner RB, Armbrust EV** (2010) pplacer: linear time maximum-likelihood and Bayesian phylogenetic placement of sequences onto a fixed reference tree. *BMC Bioinformatics* **11**:538
- Medlin L, Elwood HJ, Stickel S, Sogin ML** (1988) The characterization of enzymatically amplified eukaryotic 16S-like rRNA-coding regions. *Gene* **71**:491–499
- Michaels AF, Caron DA, Swanberg NR, Howse FA, Michaels CM** (1995) Planktonic sarcodines (acantharia, radiolaria, foraminifera) in surface waters near Bermuda - abundance, biomass and vertical flux. *J Plankton Res* **17**:131–163

- Nikolaev SI, Berney C, Fahrni JF, Bolivar I, Polet S, Mylnikov AP, Aleshin VV, Petrov NB, Pawlowski J** (2004) The twilight of Heliozoa and rise of Rhizaria, an emerging supergroup of amoeboid eukaryotes. *Proc Natl Acad Sci USA* **101**:8066–8071
- Not F, Gausling R, Azam F, Heidelberg JF, Worden AZ** (2007) Vertical distribution of picoeukaryotic diversity in the Sargasso Sea. *Environ Microbiol* **9**:1233–1252
- Nylander JAA** (2004) MrAIC.pl. 1.4.3. Program Distributed by the Author
- Polet S, Berney C, Fahrni J, Pawlowski J** (2004) Small-subunit ribosomal RNA gene sequences of Phaeodarea challenge the monophyly of Haeckel's Radiolaria. *Protist* **155**:53–63
- Popofsky A** (1920) Die Sphaerozoiden. Deutsche Südpolar-Exped 1901-1903, Zoologie II **16**: 541–587
- 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. *J Phycol* **50**:388–399
- Rambaut A** (2010) FigTree 1.3.1. [<http://tree.bio.ed.ac.uk/software/figtree/>]
- Santos SR, Kinzie RA, Sakai K, Coffroth MA** (2003) Molecular characterization of nuclear small subunit (18S) -rDNA pseudogenes in a symbiotic dinoflagellate (*Symbiodinium*, Dinophyta). *J Eukaryot Microbiol* **50**:417–421
- Sauvadet AL, Gobet A, Guillou L** (2010) Comparative analysis between protist communities from the deep-sea pelagic ecosystem and specific deep hydrothermal habitats. *Environ Microbiol* **12**:2946–2964
- Stoeck T, Taylor GT, Epstein SS** (2003) Novel eukaryotes from a permanently anoxic Cariaco Basin (Caribbean Sea). *Appl Environ Microbiol* **69**:5656–5663
- Stoeck T, Hayward B, Taylor GT, Varela R, Epstein SS** (2006) A multiple PCR-primer approach to access the microeukaryotic diversity in environmental samples. *Protist* **157**:31–43

703

704 **Strelkov AA, Reshetnyak VV** (1971) Kolonialnye radiolyarii Spumellaria Mikrovogo
 705 Okeana. Issledovaniya Faunyi Morey **9**:295–418
 706

707 **Suzuki N, Aita Y** (2011) Radiolaria: achievements and unresolved issues: taxonomy and
 708 cytology. Plankton Benthos Res **6**:69–91
 709

710 **Suzuki N, Ogane K, Aita Y, Kato M, Sakai S, Kurihara T, Matsuoka A, Ohtsuka S, Go**
 711 **A, Nakaguchi K, Yamaguchi S, Takahashi T, Tuji A** (2009) Distribution patterns of the
 712 radiolarian nuclei and symbionts using DAPI-fluorescence. Bull Natl Mus Nat Sci Ser B
 713 **35**:169–182
 714

715 **Swanberg NR** (1979) The Ecology of Colonial Radiolarians: Their Colony Morphology,
 716 Trophic Interactions and Associations, Behavior, Distribution, and the Photosynthesis of their
 717 Symbionts. PhD thesis. Woods Hole Oceanographic Institution and Massachusetts Institute of
 718 Technology
 719

720 **Swanberg NR, Caron DA** (1991) Patterns of sarcodine feeding in epipelagic oceanic
 721 plankton. J Plankton Res **13**:287–312
 722

723 **Swanberg NR, Harbison GR** (1980) The ecology of *Collozoum longiforme*, sp. nov., a new
 724 colonial radiolarian from the equatorial Atlantic Ocean. Deep-Sea Res **27**:715–731
 725

726 **Thornhill DJ, Lajeunesse TC, Santos SR** (2007) Measuring rDNA diversity in eukaryotic
 727 microbial systems: how intragenomic variation, pseudogenes, and PCR artifacts confound
 728 biodiversity estimates. Mol Ecol **16**:5326–5340
 729

730 **Tsagkogeorga G, Turon X, Hopcroft RR, Tilak M, Feldstein T, Shenkar N, Loya Y,**
 731 **Huchon D, Douzery EJP, Delsuc F** (2009) An updated 18S rRNA phylogeny of tunicates
 732 based on mixture and secondary structure models. BMC Evol Biol **9**:187
 733

734 **Vaidya G, Lohman DJ, Meier R** (2011) SequenceMatrix: concatenation software for the fast
 735 assembly of multi-gene datasets with character set and codon information. Cladistics **27**:171–
 736 180

Venter JC, Remington K, Heidelberg JF, Halpern AL, Rusch D, Eisen JA, Wu D, Paulsen I, Nelson KE, Nelson W, Fouts DE, Levy S, Knap AH, Lomas MW, Nealson K, White O, Peterson J, Hoffman J, Parsons R, Baden-Tillson H, Pfannkoch C, Rogers YH, Smith HO (2004) Environmental genome shotgun sequencing of the Sargasso Sea. *Science* **304**:66–74

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

Zettler LAA, Anderson OR, Caron DA (1999) Towards a molecular phylogeny of colonial spumellarians radiolaria. *Mar Micropaleontol* **36**:67–79

Tables and Figure Legends

Table 1. Oligonucleotide primers used for PCR amplifications and sequencing.

Figure 1. Molecular phylogeny of Collodaria inferred from the concatenated complete 18S and partial 28S (D1-D2 regions) rRNA genes (92 taxa and 1614 aligned positions). The tree was obtained by using a Bayesian method implemented using the GTR + Γ model of sequence evolution. PhyML bootstrap values (100 replicates) and MrBayes posterior probabilities are shown at the nodes when supports are higher than 60% and 0.7, respectively. Black circles indicate bootstrap support and posterior probabilities of 100% and 1.00, respectively. The three main clades, defined based on statistical support and morphological criteria, highlight the Collosphaeridae (A), Collophidiidae (B), Sphaerozoidae (C) families and their respective sub-clades. Morphological criteria are identified on the right for each specimen (dark grey: presence; white: absence): colonial or solitary, naked (lacking any silicified structure), spicule-bearing (one bar: simple spicules; two bars: radiate spicules) or skeleton-bearing. Reference sequences from GenBank are shown in bold. Six Nassellaria sequences were assembled as out-group (AB430759, AF382824, DQ386169, FJ032682, FJ032683, HQ651779). Branches with a double barred symbol are reduced to half-length for clarity.

Figure 2. Scanning Electron Microscopy images of Collosphaeridae skeletons. All of these images were obtained from the exact same specimens used for the molecular phylogenetic analysis in this study. (A) *Disolenia quadrata* (Sat 12). (B) *Disolenia tenuissima* (Sat 17). (C) *D. tenuissima* (Sat 19). (D-D'') *D. tenuissima* (Sat 21). (E) *Disolenia zanguebarica* (Sat 20). (F) *D. zanguebarica* (Sat 26). (G) *D. zanguebarica* (Sat 27). (H) *Collosphaera tuberosa* (Ind 44). (I) *C. tuberosa* (Ind 48). (J) *Siphonosphaera abyssi* (Pac 7); scale bars = 30 µm.

Figure 3. Scanning Electron Microscopy images of Sphaerozoidae spicules. All of these images were obtained from the exactly same specimens used for the molecular phylogenetic analysis in this study. (A) *Rhaphidozoum acuferum* (Ind 47). (B) *R. acuferum* (Pac 8). (C) *Sphaerouzoum fuscum* (Sat 24). (D) *Sphaerouzoum strigulosum* (Nat 6). (E) *Sphaerouzoum brandtii* (Pac 26). (F) *S. brandtii* (Pan 3). (G) *Sphaerouzoum armatum* (Pac 24). (H-H') *Sphaerouzoum punctatum* (Pac 21). (I) *Sphaerouzoum trigeminum* (Pan 19). (J) *S. trigeminum* (Sat 23). (K) *S. trigeminum* (Sat 18); scale bars = 20 µm.

Figure 4. pplacer phylogenetic placement of 92 environmental sequences onto a concatenated phylogenetic tree of Collodaria (complete 18S + partial 28S rRNA genes). Number of sequences affiliated to a clade (node) or sub-clade (branch) is shown inside green circles.

Table 1

GENE	PRIMER	Specificity	SEQUENCE 5'-3'	TYPE	REFERENCE
18S (1st part)	SA	Eukaryote	AAC CTG GTT GAT CCT GCC AGT	Forward	Medlin et al. 1988
	S81 Col	Collodaria	ATC ACA GAC CTG TYA TTG CWA	Reverse	This study
18S (2nd part)	S32 Col	Collodaria	TAT GCT AAC RWT GYT GCA	Forward	This study
	V9R	Eukaryote	CCT TCY GCA GGT TCA CCT AC	Reverse	Romac (unpub.)
28S	28S Col-F	Collodaria	TGG ACT TTC TAA GTA ATG GCG	Forward	This study
	ITSa3 Col	Collodaria	CAC CAT CTT TCG GGT CCC AGT	Reverse	This study

Figure 1

Phylogenetic tree showing the relationship between the Outgroup (Nassellaria) and the main clade. The tree is rooted on the left, with a branch labeled "Outgroup (Nassellaria)" leading to a node. The main clade is represented by a large, shaded, fan-shaped area.

Figure 2

Figure 3