Bridging the gap between morphological species and molecular barcodes – Exemplified by loricate choanoflagellates

Frank Nitsche, Helge Abuldhauge Thomsen, Daniel J. Richter

Bridging the gap between morphological species and molecular barcodes - exemplified by loricate choanoflagellates

Frank Nitsche[a1], Helge Abildhauge Thomsen[b] and Daniel J. Richter[c,d]

[a] Zoological Institute, General Ecology, University of Cologne, Zülpicher Str. 47b, 50674 Cologne, Germany

[b] National Institute of Aquatic Resources (DTU Aqua), Technical University of Denmark, Jægersborg allé 1, DK-2920 Charlottenlund, Denmark

[c] CNRS, UMR 7144, Station Biologique de Roscoff, Place Georges Teissier, 29680 Roscoff, France

[d] Sorbonne Universités, Université Pierre et Marie Curie (UPMC) Paris 06, UMR 7144, Station Biologique de Roscoff, Place Georges Teissier, 29680 Roscoff, France

[1]Corresponding author; e-mail FNitsche@uni-koeln.de (F. Nitsche)

Abstract

Translating the vast amounts of molecular barcodes from global surveys of microbial eukaryotes into ecological insight depends critically on a well-curated reference database with adequate taxonomic coverage. In this respect, the choanoflagellates resemble other eukaryotic lineages: reasonable coverage at higher taxonomic levels, but missing diversity at the species level. The acanthoecid (loricate) choanoflagellates are well-characterized morphologically, with over 115 species described, but less than 10% with any sequence data. Because lorica shape is species-specific, the acanthoecids represent an opportunity to link morphological with molecular data within a lineage of eukaryotes. To match morphospecies to sequences, we sampled the Kattegat and the Isefjord in Denmark in September 2014 and February 2015. We identified 45 morphospecies and sequenced ribosomal DNA of nine previously unsequenced species, roughly doubling the number of acanthoecid species with sequence data, including the first data representing five genera: *Bicosta*, *Calliacantha*, *Cosmoeca*, *Crinolina* and *Pleurasiga*. Our phylogenetic analysis is mainly congruent with morphology-based systematics. Five of the newly sequenced species match a previously unidentified barcode from *Tara* Oceans, providing access to the global distribution of species isolated from Danish waters. One species, *Calliacantha natans*, is the second most globally abundant choanoflagellate present in *Tara* Oceans. Our project translating new ribosomal DNA sequences to distributions of described species on a global scale supports the approach linking morphology to molecular barcodes for microbial eukaryote ecology.

Keywords: Acanthoecids, *Bicostata, Calliacantha, Diaphanoeca, Crinolina*, SSU, LSU

Introduction

The group of acanthoecid (loricate) choanoflagellates is compromised of at present more than 115 described morphospecies (Leadbeater 2015). In general they are ubiquitously found in brackish and marine waters, where they can reach high abundances of up to 3.5×10^6 cells per l^{-1}/Bering Sea (e.g. Sukhanova 2001). The systematics and taxonomy of this group of organisms have undergone substantial changes when analysed with molecular data, in particular the small and large subunit

(SSU and LSU) of the ribosomal DNA (rDNA) locus. In several cases within choanoflagellates (both loricates and craspedids), it was necessary to modify morphology based taxonomy based on results from molecular phylogeny (Carr et al. 2008). Within the acanthoecid choanoflagellates, two groups were established based on both developmental traits, such as the lorica assembly, and on confirming molecular data. In some cases, species described by morphology and assigned to the same genera do not form a monophyletic group when sequence data produced are analyzed in a phylogenetic tree. An example therefore is *Helgoeca nana,* which was transferred from the tectiform genus *Acanthocorbis* to the new, nudiform genus *Helgoeca* based on phylogenetic and morphologic studies (Leadbeater et al. 2008). In addition, only 10% of acanthoecid choanoflagellate species have thus far been analysed regarding their molecular identity, in particular with reference to one of the most frequently used and appropriate marker genes, the SSU of the rDNA locus (Pawlowski et al. 2012). This lack of molecular data prevents an assignment of next generation sequencing (NGS) data from ongoing molecular barcoding surveys at the species level. In particular, the vast majority of acanthoecid barcodes within the global dataset of the *Tara* Oceans expedition could not be assigned at the species level. Questions arising when analysing acanthoecid choanoflagellate sequences from public databases include: how many species are there? How variable is the analysed marker gene? How should a species be defined based on this data? This work is aimed to demonstrate the importance of combining morphological and molecular data to understand ecological processes occurring in the complex microbial food web.

Material and Methods

Sampling was done in Denmark, in September 2014 and February 2015 according to previous records of acanthoecids from this area (Thomsen 1976; Thomsen 1992; Thomsen et al. 2016).

In addition to surface samples from Isefjord (at Tempelkrogen and Kyndby), a Niskin bottle was used from board of the RV Ophelia to sample underneath the halocline at the field station of DTU Aqua at Helsingoer, The Sound (Table 1). Samples were transported back to the laboratory in a thermically isolated container and immediately processed. In the laboratory, about 500ml were concentrated on a 2µm Millipore filter using gravity for filtration. Half of this concentrate was transferred in to a petri dish for

single cell picking. A micromanipulator was used to pick and transfer the morphologically identified single cells to a PCR tube. The other half was further concentrated by centrifugation (5 min at 2000g) and the resulting pellet was fixed using osmium tetroxide vapour for 1 min on a cover slide. These samples were used for identification and microphotography. Schematic drawings were done based on drawings by Thomsen (1992) using Blender 2.77.

Single cells were amplified using two step PCR. For the first PCR the primer set 42F (5'-CTC AAR GAY TAA GCC ATG CA-3') and 1510R (5'- CCT TCY GCA GGT TCA CCT AC-3') were used, re-amplification was done using the primer set 82F (5'-GAA ACT GCG AAT GGC TC-3') and 1630R (5'- ACC TAC GGA AAC CTT GTT ACG-3') (final primer concentration 0.02µM). For both reactions the following program was applied: 2 min 98°C denaturation followed by 35 cycles of 30 sec at 95°C denaturation, 30 sec at 48°C and 30 sec at 52°C annealing, 2 min at 72°C elongation and a final elongation at 72°C for 7 min. For LSU rDNA single cells were amplified using the primer set NLF184/21(5'-ACC CGC TGA AYT TAA GCA TAT-3') and NLR3535/22 (5'- MRG GCT KAA TCT CAR YRG ATC G -3') in the first amplification and NLF1105/22 (5'- CCG AAG TTT CCC TCA GGA TAG C -3') and NLR2098/24 (5'- AGC CAA TCC TTW TCC CGA AGT TAC -3') for the reamplification. The PCR program was the same as for SSU, only elongation time was set to 4 min in the first PCR and 1.5 min for the reamplification.

PCR products were purified using the Qiagen PCR purification kit according to the manufacturer's instructions. Each purified SSU rDNA PCR product was initially sequenced directly with an ABI 3130xl DNA Sequence Analyzer using the primer C6F (5'- GAC TCA ACA CGG GRA AMC TYA CCA -3'). Positive sequences were sequenced further either by direct sequencing with three additional primers, 82F (5'- GAA ACT GCG AAT GGC TC -3'), V4F (5'- CCA GCA SCY GCG GTA ATT CC -3'), and 1630R (sequence given above), or in cases where the amount of DNA in the PCR product was insufficient for the multiple sequencing reactions that were necessary to produce a complete 18S assembly (*C. ventriosa*, *D. undulata* and *P. minima*) by cloning with the TOPO® TA Cloning® Kit into the pCRTM4-TOPO® vector from Life Technologies, according to the manufacturer's instructions, followed by sequencing with V4F and the standard vector-based M13F and M13R primers. LSU rDNA amplicons were treated accordingly only using the primers NLF184/21, NLF1105/22 and NLR2098/24 for sequencing.

Fragments were assembled manually using Bioedit (Hall 1999). The obtained sequences of SSU and LSU rDNA were used to build a concatenated two gene (SSU and LSU rDNA sequence Table 2), 4470bp (manually corrected) alignment by Kalign2 (gap open penalty 90, gap extention penalty 4, terminal gap penalties 3 and bonus score 0) (Lassmann et al. 2009). This alignment was analysed using Maximum likelihood (ML) and Bayesian inference (BI) methods. ML analysis was computed by RAxML (Stamatakis et al. 2008) using the RAxML GAMMA+P-Invar model of rate heterogeneity, as suggested by MrAic (Ronquist and Huelsenbeck 2003) and run using a GTR+I model and a four-category gamma distribution to model among-site rate variation, 1000 bootstrap replicates (rapid bootstrapping). The Bayesian inference search, computed by MrBayes 3.2.6 (Ronquist, Huelsenbeck, 2003) consisted of two parallel chain sets run at default temperatures with a sampling frequency of 10, run until the average standard deviation of split frequencies dropped below 0.01. The analysis consisted of 3,500,000 generations, with a burnin of 87,500, before calculating posterior probabilities. The choanoflagellates were rooted with a clade consisting of two ichthyosporean taxa (Nitsche 2014).

In order to identify *Tara* Oceans OTUs corresponding to the 9 species we sequenced, we extracted the V9 region of the 18S sequence for each species. We then searched for matching OTUs in the *Tara* Oceans data using a similar strategy to the one implemented in order to construct the OTUs (Mahé et al., 2014, de Vargas et al., 2015). We searched within the database of *Tara* Oceans V9 barcodes for either perfect matches or for matches with exactly one difference between our nine new sequences and a *Tara* Oceans V9 barcode using the alignment software Sumatra version 1.0.10 (https://git.metabarcoding.org/obitools/sumatra/wikis/home). We visualized these matches using network diagrams (Figure 4) with Cytoscape version 3.2.1 (Shannon et al., 2003). The idea of this approach is that many rare barcodes may be present in a sample simply due to PCR error, because of the massive amount of barcodes that were sequenced. If this is true, then when we build a network of barcodes, drawing connections between barcodes if there is exactly one DNA sequence difference between them, then the "true" barcode will be highly abundant, and will be surrounded by a "swarm" of much less abundant sequences that represent sporadic PCR errors (which may also be errors occurring on intermediate PCR products that already contained one or more errors). To create maps for each newly affiliated *Tara* Oceans OTU, we used the GPS coordinates of

each sampled station provided by the *Tara* Oceans Consortium (*Tara* Oceans Consortium, 2014, Pesant et al., 2015) and relative abundances for each OTU for the 0.8-5 µm and 5-20 µm size fractions (de Vargas et al., 2015). Although several of the species we describe may construct loricae measuring larger than 5 µm (or even larger than 20 µm), we found the majority of the abundance for each species within the 0.8-5 µm size fraction, and nearly all of the abundance within the 0.8-5 µm and 5-20 µm size fractions. We hypothesize that this occurs due to the filtration process, during which cells are likely pulled out of their loricae. There were only four cases in which a species was observed in either the 20-180 µm or 180-2000 µm size fractions, but not in the two smaller size fractions; in three of these cases, only 1 total barcode was observed outside the two smaller size fractions, and in one case, 4 barcodes were observed, amounts we considered to be below the threshold required to present the data on a map. We created the maps using R version 3.3.0 (R Core Team, 2016) using the following packages: data.table version 1.9.6 (Dowle et al., 2015), ggplot2 version 2.1.0 (Wickham, 2009) and maps version 3.1.0 (Brownrigg, 2016).

Results

Based on sampling from Danish waters, we were able to isolate, sequence and neotypify nine previously unsequenced morphospecies and five yet unsequenced genera of acanthoecid choanoflagellates: *Bicosta minor, Calliacantha longicaudata, C. natans, Cosmoeca ventricosa, Crinolina isefiordensis, Diaphanoeca sphaerica, D. undulata, Pleurasiga minima* and *P. reynoldsii.* All species could be maintained for a few days only in mixed cultures, containing environmental bacteria and other protists. An isolation and establishment of monocultures was not successful. For all species, schematic 3D drawings for clear lorica architecture (Fig 1A-I) and LM micrographs (2A-M) are provided alongside phylogenetic analysis based on SSU and LSU rDNA (Fig 3).

Bicosta minor (Reynolds) Leadbeater 1978; Fig. 1A
Additional references: Bérard-Therriault et al. 1999; Manton et al. 1980; Marchant 2005; Thomsen 1982; Tong 1997
Global distribution in *Tara* Oceans Data: no identical or nearly identical match in the *Tara* Oceans V9 database (Table 3).

6

Remarks: Marine tectiform acanthoecid, single cell isolated from a Kyndby sample collected 23 February 2015 (Table 1). Identification based on the original description, neotypified to relate a morphotype to genotype.

Emended diagnosis: Tectiform, cells solitary, lorica 11-45 μm in length, composed of seven costal strips arranged as two longitudinal costae, and by two costal strips each with tapered tips anteriorly and a posterior spine. (Fig 1A and 2A-C). The closest relative of the SSU rDNA sequence in NCBI GenBank, based on nucleotide BLAST results, is an uncultured choanoflagellate from the Scotia Sea (KC488357) at 98% identity.

Neotype sequence data: The SSU rDNA sequence of *B. minor* has been deposited in the GenBank database with the accession number KU587839.

Calliacantha longicaudata (Leadbeater) Leadbeater 1978; Fig. 1C

Additional references: Bérard-Therriault et al. 1999; Leadbeater 2015; Marchant 2005; McKenzie et al. 1997; Thomsen 1982; Thomsen 1992; Thomsen et al. 1997. Global distribution in *Tara* Oceans Data: no identical or nearly identical match in the *Tara* Oceans V9 database (Table 3).

Remarks: Marine tectiform acanthoecid, found at Kyndby (Table 1). Identification based on the original description, neotypified to relate a morphotype to genotype.

Emended diagnosis: Tectiform, cells solitary, lorica without stalk 11-18 μm in length, composed of five longitudinal costae each composed of two costal strips. Two transverse costae are linked to the anterior longitudinal strips, leaving the ultimate tips of the anterior costae as projecting spines. The pedicel is composed of a variable number of costal strips (Figs. 1C and 2E).

The closest relative of the SSU rDNA sequence in NCBI GenBank, based on nucleotide BLAST results, is an uncultured choanoflagellate from the Scotia Sea (KC488358) at 99.6% identity.

Neotype sequence data: The SSU and LSU rDNA sequences of *C. longicaudata* have been deposited in the GenBank database with the accession numbers KU587840 and KU587841, respectively.

Calliacantha natans (Grøntved) Leadbeater 1978; Fig. 1B

Additional references: Bérard-Therriault et al. 1999; Buck and Garrison 1983; Leadbeater 1972; Leadbeater 2015; Manton and Leadbeater 1978; Marchant 2005; Thomsen 1982; Thomsen et al. 1990; Throndsen 1993; Tong 1997

Global distribution in *Tara* Oceans Data: Fig. 5A (Table 3).

Remarks: Marine tectiform acanthoecid, found at Kyndby and Tempelkrogen (Table 1). Identification based on the original description, neotypified to relate a morphotype to genotype.

Emended diagnosis: Tectiform, cells solitary, lorica 20-60 µm in length, formed by six longitudinal costae and two transverse costae, one of which forms the anterior ring. Posterior end of lorica chamber tapers and joins with a single spine, while three spines emerge from the anterior transverse ring (Figs. 1B and 2F).

The closest relative of the SSU rDNA sequence in NCBI GenBank, based on nucleotide BLAST results, is an uncultured choanoflagellate from the Scotia Sea (KC488358) at 98% identity.

Neotype sequence data: The SSU and LSU rDNA sequences of *C. natans* have been deposited in the GenBank database with the accession numbers KU587842 and KU587843, respectively.

Cosmoeca ventricosa Thomsen 1984 in Thomsen and Boonruang 1984; Fig. 1D

Additional references: Bérard-Therriault et al. 1999; Booth 1990; Leadbeater 2015; Marchant 2005;Thomsen et al. 1990; Tong 1997

Global distribution in *Tara* Oceans Data: no identical or nearly identical match in the *Tara* Oceans V9 database (Table 3).

Remarks: Marine tectiform acanthoecid, found at Helsingør, 20m (Table 1). Identification based on the original description, neotypified to relate a morphotype to genotype.

Emended diagnosis: Tectiform, cells solitary, lorica 23-32 µm in length, barrel-shaped, composed of 3 transverse costae and 8-12 longitudinal costae. Two almost equally sized anterior transverse costae located at the anterior end of the lorica chamber and at the level of the joints between the second and the third longitudinal costal strip. Same number of strips in anterior transverse costae as there are longitudinal costae. Third transverse costa narrower, formed by 3-4 costal strips, positioned between the third and fourth longitudinal strip (Figs. 1D and 2G).

The closest relative of the SSU rDNA sequence in NCBI GenBank, based on nucleotide BLAST results, is an uncultured choanoflagellate from the Baltic Sea (FN690465) at 99% identity.

Neotype sequence data: The SSU rDNA sequence of *C. ventricosa* has been deposited in the GenBank database with the accession number KU587844.

Crinolina isefiordensis Thomsen 1976; Fig. 1E

Additional references: Bérard-Therriault et al. 1999; Leadbeater 2015; Thomsen 1982; Throndsen 1993; Tong 1997.

Global distribution in *Tara* Oceans Data: Fig. 5B (Table 3).

Remarks: Marine tectiform acanthoecid, found at Tempelkrogen (Table 1). Identification based on the original description, neotypified to relate a morphotype to genotype.

Emended diagnosis: Tectiform, cells solitary, lorica 25-30 µm in length, skirt-shaped, open anteriorly and posteriorly. 14-15 longitudinal costae composed of 6-7 costal strips. Anteriorly longitudinal spines from two costal strips protrude above anterior transverse ring, posteriorly longitudinal costae slightly project beyond the transverse ring (Figs. 1E and 2H). The closest relative of the SSU rDNA sequence in NCBI GenBank, based on nucleotide BLAST results, is an uncultured choanoflagellate from the Arctic Ocean, north of Barrow, Alaska (KJ762435) at 96% identity.

Neotype sequence data: The SSU rDNA sequence of *C. isefiordensis* has been deposited in the GenBank database with the accession number KU587845.

Diaphanoeca sphaerica Thomsen 1982; Fig. 1G

Additional references: Marchant 2005; Marchant and Perrin 1986.

Global distribution in *Tara* Oceans Data: no identical or nearly identical match in the *Tara* Oceans V9 database (Table 3).

Remarks: Marine tectiform acanthoecid, found at Tempelkrogen and Kyndby (Table 1). Identification based on the original description, neotypified to relate a morphotype to genotype.

Emended diagnosis: Tectiform, cells solitary, lorica 22-30 µm in length and 16-17 µm in diameter, bell-shaped, formed by 13-15 longitudinal costae composed of 9-10 strips each, and three transverse costae. In the two anterior transverse costae the number of strips equals the number of longitudinal costae, whereas the posterior

transverse costa contains fewer strips. Posterior lorica end with irregularly arranged costal strips. Large colony formations frequently found (Figs. 1G and 2I, J).

The closest relative of the SSU rDNA sequence in NCBI GenBank, based on nucleotide BLAST results, is *Diaphanoeca grandis* (AF084234) at 99% identity.

Neotype sequence data: The SSU and LSU rDNA sequences of *D. sphaerica* have been deposited in the GenBank database with the accession numbers KU587846 and KU587847, respectively.

Diaphanoeca undulata Thomsen 1982; Fig. 1F

Additional references: Thomsen 1992

Global distribution in *Tara* Oceans Data: Fig. 5C (Table 3).

Remarks: Marine tectiform acanthoecid, found at Helsingør, 20m (Table 1). Identification based on the original description, neotypified to relate a morphotype to genotype.

Emended diagnosis: Tectiform, cells solitary, lorica up to 30 µm in length without posterior pedicel, with posterior pedicel up to 50 µm long. 11-12 longitudinal costae, composed of 5 costal strips and three transverse costae form the lorica chamber. Anterior transverse costa crosses longitudinal costae at the junction between the first and second strip, leaving the first strips as free projecting spines. The middle transverse costa, which has an undulating appearance, crosses anterior to the joints between the second and the third longitudinal costal strip (Figs. 1F and 2K).

The closest relative of the SSU rDNA sequence in NCBI GenBank, based on nucleotide BLAST results, is an uncultured choanoflagellate from the Arctic Ocean, north of Barrow, Alaska (KJ762435) at 98% identity.

Neotype sequence data: The SSU rDNA sequence of *D. undulata* has been deposited in the GenBank database with the accession number KU587848.

Pleurasiga minima Throndsen 1970; Fig. 1H

Additional references: Bérard-Therriault et al. 1999; Hallegraeff 1983 ; Manton et al. 1976; Marchant 2005; Thomsen 1976; Thomsen 1982; Thomsen and Buck 1991; Tong 1997

Global distribution in *Tara* Oceans Data: Fig. 5D (Table 3).

Remarks: Marine tectiform acanthoecid, found in Tempelkrogen (Table 1). Identification based on the original description, neotypified to relate a morphotype to genotype.

Emended diagnosis: Tectiform, cells solitary, lorica 10-14 µm in height and 10-12 µm in width, amphora-shaped, two transverse, nearly equally wide costae, 7 longitudinal costae composed of 3 costal strips. Costal strip connections are, anteriorly, T-joints (Figs. 1H and 2L).

The closest relative of the SSU rDNA sequence in NCBI GenBank, based on nucleotide BLAST results, is an uncultured choanoflagellate from the Arctic Ocean, north of Barrow, Alaska (KJ763316) at 97% identity.

Neotype sequence data: The SSU and LSU rDNA sequences of *P. minima* have been deposited in the GenBank database with the accession numbers KU587849 and KU587850, respectively.

Pleurasiga reynoldsii Throndsen 1970; Fig.1I

Additional references: Marchant 2005; Thomsen 1982; Thomsen and Buck 1991; Throndsen 1993; Throndsen 1970

Global distribution in *Tara* Oceans Data: Fig. 5E (Table 3).

Remarks: Marine tectiform acanthoecid, found in Tempelkrogen (Table 1). Identification based on the original description, neotypified to relate a morphotype to genotype.

Emended diagnosis: Tectiform, cells solitary, lorica up to 23 µm in height and up to 22 µm in width, amphora-shaped, posterior transverse costae about 1/3 wider than anterior costae composed from 7 costal strips, 7 longitudinal costae between anterior and posterior costae composed by two costal strips, merging branch like into 4 terminal costae (Figs. 1I and 2M).

The closest relative of the SSU rDNA sequence in NCBI GenBank, based on nucleotide BLAST results, is an uncultured choanoflagellate from the Gulf Stream, North Atlantic (KJ759457) at 95% identity.

Neotype sequence data: The SSU and LSU rDNA sequences of *P. reynoldsii* have been deposited in the GenBank database with the accession numbers KU587851 and KU587852, respectively.

11

Species from the genus *Diaphanoeca* form a well-supported cluster in the concatenated SSU and LSU analysis (Fig. 3). *D. sphaerica* and *D. undulata* branch together with *D. grandis* and *D. spiralifurca*. *Crinolina isefiordensis*, a morphologically similar species, clusters within the same group, indicating a possible paraphyly or polyphyly (neither state is robustly supported) of *Diaphanoeca.*

The two species from the genus *Calliacantha*, *C. longicaudata* and *C. natans*, form a monophyletic group with high bootstrap support. The genus *Pleurasiga,* represented by *P. minima* and *P. reynoldsii*, also forms a monophyletic group with strong support. Finally, *Cosmoeca ventricosa* and *Bicosta minor* cluster together with *Acanthocorbis unguiculata*. Although this branching pattern has high bootstrap support, the different morphology of these species indicates that a subsequent incorporation of additional sequences representing each genus may cause the pattern to change.

We detected 5 operational taxonomic units (OTUs) within *Tara* Oceans V9 hypervariable region SSU data perfectly matching one of the 9 newly sequenced species (de Vargas et al., 2015). Each OTU was previously classified as a choanoflagellate, but lacked a species identification. These OTUs correspond to *C. natans, C. isefiordensis, D. undulata , P. minima and P. reynoldsii* (Table 3, Fig. 4). We did not detect perfect matches for *B. minor, C. longicaudata*, *C. ventricosa* or *D. undulata*, although in each case our new sequence was an equal or better match to the Tara OTU sequence than any existing sequences in GenBank. One of the newly identified species, *C. natans*, is the second most globally abundant choanoflagellate present within Tara Oceans V9 data. Its 18S V9 sequence is distantly related to any sequence currently in GenBank (the closest match is to *Didymoeca costata* at 88% identity). A network-based analysis comparing our new sequences to *Tara* V9 barcodes showed that, in all cases except for one, these sequences matched the most abundant barcode within the OTU identified. In the case of *P. reynoldsii*, it is possible that our new sequence might contain a single error, since there are no other sequences within 1 difference of any sequences within this OTU in the *Tara* V9 database (Fig. 4).

The five newly sequenced species present in *Tara* Oceans data display a range of distributions in a global context. *Calliacantha natans* (Fig. 5A), the second most abundant choanoflagellate in *Tara* Oceans data (Table 3), and *Diaphanoeca undulata*, the 16th most abundant (Fig. 5C) are found in nearly every station and ocean basin studied, with the exception of the northern Indian Ocean (where the

craspedid *Monosiga brevicollis* appears to be the dominant choanoflagellate; de Varga*s et* al., 2015). Both of these species are also found at their highest relative abundances in the cold water stations of the Antarctic, the South Atlantic, and the cold water coastal upwelling off the Horn of Africa (Supplementary Figs. S1 and S2). The two newly sequenced *Pleurasiga* species, *P. minima* and *P. reynoldsii*, are detected in the South Atlantic and the Antarctic, at the boundary between the Indian Gyre and the South Atlantic Gyre off the coast of South Africa, at an upwelling station off the west coast of South America, and in the Mediterranean (Figs. 5D, E). As is the case for *C. natans* and *D. undulata*, the two *Pleurasiga* species are also observed at their highest relative abundances in the cold water stations of the southern hemisphere (Supplementary Figs. S4, S5). *Crinolina isefiordensis*, by far the least abundant of the newly sequenced species, is found in only two stations world-wide (Fig. 5B, Supplementary Fig. S3).

Discussion

Within a short time and with comparably little effort we were able to roughly double the number of acanthoecid choanoflagellates with sequence data available in public databases. The genus *Diaphanoeca* is a good example of consensus between morphology and phylogenetic analysis, as all species form a well-supported cluster (Fig. 3). The presence of *Crinolina* within this cluster may be well justified by the similar morphology of the two genera. The genus *Crinolina*, in particular the type species *C. isefiordensis*, was characterized by the lorica being open posteriorly. Although it is closely related to *Diaphanoeca* (Thomsen 1976) in its morphology, it is well distinguishable. As, at present, there is no sequence from the type species *C. aperta* available, we refrain from transferring *C. isefiordensis* to the genus *Diaphanoeca*, as it might instead turn out to be a sister genus when additional sequence data become available. In addition the nested position of *C. isefiordiensis* lacks strong phylogenetic support.

Pleurasiga and *Calliacantha* also form highly supported clusters (Fig. 3). Thus, the phylogeny created with concatenated sequences of SSU and LSU rDNA confirmed the coherence of both genera. However, within our data set the monophyletic origin of the Stephanoecida cannot be considered robustly resolved on phylogenetic grounds. Adding housekeeping genes like HSP90 or α-tubulin might help to resolve this issue (Leadbeater et al. 2008; Nitsche et al. 2011), but these genes could not be

analysed in this study. Nevertheless, the nudiform species, Acanthoecida, are recovered as a monophyletic group. A future enhancement to our single-cell sequencing approach could include whole genome amplification of single cells to obtain sufficient DNA for the analysis of other marker genes (Krabberød et al. 2011). Despite the fact that our sampling was conducted exclusively in Danish waters, we detect an essentially global distribution for the species we sequenced with matching OTUs in the *Tara* Oceans expedition (which did not directly sample Danish waters). This result echoes previous observations of choanoflagellates as ubiquitous in aquatic environments, while providing evidence that single species are also found across all ocean basins. In addition, this highlights the fact that a local sampling approach can be translated via sequence barcodes to information on species distributions at a global level. Using these sequences, it should now be possible to link previous descriptions of species distribution using morphology directly to distributions from *Tara* and other surveys based on ribosomal sequence data. Within our sequences, we succeeded in providing an identification for the second most globally abundant choanoflagellate in *Tara* Oceans data, *Calliacantha natans*. *C. natans* and the other species we identify in *Tara* Oceans data display their highest relative abundances in cold water *Tara* sampling stations. The highest relative abundance for any newly sequenced species is that of *C. natans* in the 0.8-5 µm size fraction at the surface located off the coast of Argentine Patagonia, where it represents nearly 1% of the total abundance of all 2,828 OTUs detected, and is the seventh most abundant OTU. Thus, within our sample of only five newly sequenced loricate choanoflagellates, we are able to detect a global distribution for each species and to attest to the importance of individual species within individual microbial food webs.

In addition, our approach should also be applicable to other groups of microbial eukaryotes whose morphology is species-specific but which lack extensive associated molecular barcoding data. Depositing ribosomal sequence data, such as those we collected in this study, into curated databases like the Protist Ribosomal Reference database (Guillou et al. 2013) is a valuable and necessary addition to help resolve the flood of data acquired by High Throughput Sequencing technologies. Concluding we can answer the questions posed in the introduction - the number of described acanthoecid morphospecies, 115, and the number of acanthoecid OTUs from Tara Ocean data, 129 correlate well. This result indicates that about 10 % of

loricates have not been described yet. Therefore we find that the SSU rDNA is a suitable marker gene for species determination as shown in previous studies (Nitsche and Arndt 2015; Pawlowski et al. 2012). On the other hand we find that for species characterisation, morphology and sequence data still need to be supplemented with autecological data as salinity and temperature tolerance are required to interpret for example global distribution patterns.

Acknowledgements

References

Bérard-Therriault, L., Poulin, M., Bossé, L. 1999. Guide d'identification du phytoplancton marin de l'estuaire et du Golfe du Saint-Laurent, incuant également certains protozoaires. Les Presses Scientifique du CNRC, Ottawa.

Booth, B.C. 1990. Choanoflagellates from the Sub-Arctic North Pacific-Ocean, with Descriptions of 2 New Species. Can. J. Zool. 68, 2393 -2402.

Brownrigg, R., Becker, R.A., Wilks, A.R., Minka, T.P., Deckmyn, A. 2016. maps: Draw Geographical Maps. R package version 3.1.0. https://CRAN.R-project.org/package=maps

Buck, K.R., Garrison, D.L. 1983. Protists from the ice-edge region of the Weddell Sea. Deep-Sea Res. 30, 1261 -1277.

Carr, M., Leadbeater, B.S.C., Hassan, R., Nelson, M., Baldauf, S.L. 2008. Molecular phylogeny of choanoflagellates, the sister group to Metazoa. P. Natl. Acad. Scie. USA 105, 16641 -16646.

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, J.M., 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, S.G., Bork, P., Bowler, C., Gorsky, G., Grimsley, N., Hingamp, P., Iudicone, D., Not, F., Ogata, H., Pesant, S., Raes, J., Sieracki, M.E., Speich, S., Stemmann, L., Sunagawa, S., Weissenbach, J., Wincker, P., Karsenti, E. 2015. Eukaryotic plankton diversity in the sunlit ocean. Science 22, 348(6237):1261605. doi: 10.1126/science.1261605.

Dowle, M., Srinivasan, A., Short, T., Lianoglou, S., Saporta, R., Antonyan, E. 2015. data.table: Extension of Data.frame. R package version 1.9.6. https://CRAN.R-project.org/package=data.table.

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., Lara, E., Le Bescot, N., Logares, R., Mahe, 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.

Hall, T.A. 1999. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucleid Acids Symp. 41, 95 -98.

Hallegraeff, G.M. 1983. Scale-bearing and loricate nanoplankton from the East Australian current. Bot. Mar. 26, 493 -516.

Krabberød, A.K., Bråte, J., Dolven, J.K., Ose, R.F., Klaveness, D., Kristensen, T., Bjørklund, K.R., Shalchian-Tabrizi, K. 2011 Radiolaria divided into Polycystina and Spasmaria in combined 18S and 28S rDNA phylogeny. PLoS ONE 6, e23526. doi: 10.1371/journal.pone.0023526

Lassmann, T., Frings, O., Sonnhammer, E.L. 2009 Kalign2: high-performance multiple alignment of protein and nucleotide sequences allowing external features. Nuc. Acids Res. 37, 858 -865.

Leadbeater, B.S.C. 1972. Ultrastructural observations on some marine choanoflagellates from the coast of Denmark. Brit. Phycol. J. 7, 195 -211.

Leadbeater, B.S.C. 1978. Renaming of Salpingoeca Sensu Grontved. J. Mar. Bio. Assoc. UK 58, 511 -515.

Leadbeater, B.S.C. 2015. The Choanoflagellates - Evolution, Cell biology, Ecology. Camebridge University Press, Camebridge.

Leadbeater, B.S.C, Hassan, R., Nelson, M., Carr, M., Baldauf, S.L. 2008. A new genus, *Helgoeca gen. nov.*, for a nudiform choanoflagellate. Eur. J. Protistol. 44, 227 -237.

Mahé, F., Rognes, T., Quince, C., de Vargas, C., Dunthorn, M. 2014. Swarm: robust and fast clustering method for amplicon-based studies. PeerJ.,2:e593. doi: 10.7717/peerj.593.

Manton, I., Leadbeater, B.S.C. 1978 Some critical qualitative details of lorica construction in type species of *Calliacantha* Leadbeater (Choanoflagellata). P. Roy. Soc. Lond. B. Bio. 203, 49 -57.

Manton, I., Sutherland, J., Leadbeater, B.S.C. 1976. Further observations on fine-structure of marine collared flagellates (Choanoflagellata) from Arctic Canada and West Greenland - Species of *Parvicorbicula* and *Pleurasiga*. Can. J. Zool. 54, 1932 -1955.

Manton, I., Sutherland, J., Oates, K. 1980. A reinvestigation of collared flagellates in the Genus *Bicosta* Leadbeater with special reference to correlations with climate. Phil. T. Roy. Soc. B 290, 431 -447.

Marchant, H.J. 2005. Choanoflagellates. In: Scott FJ, Marchant HJ (Eds.), Antarctic Marine Protists. Australian Biological Resource Studies, Canberra and Australian Antarctic Division, Canberra and Hobbart, pp. 326-346.

Marchant, H.J., Perrin, R.A. 1986. Planktonic choanoflagellates from 2 Antarctic lakes including the description of *Spiraloecion didymocostatum gen. et sp. nov.* Polar Biol. 5, 207 -210.

McKenzie, C.H., Deibel, D., Thompson, R.J., MacDonald, B.A., Penney, R.W. 1997. Distribution and abundance of choanoflagellates (Acanthoecidae) in the coastal cold ocean of Newfoundland, Canada. Marine Biol. 129, 407 -416.

Nitsche, F. 2014. S*tephanoeca arndti spec. nov* - First cultivation success including molecular and autecological data from a freshwater acanthoecid choanoflagellate from Samoa. Eur. J. Protistol. 50, 412 -421.

Nitsche, F., Arndt, H. 2015. Comparison of similar Arctic and Antarctic morphotypes of heterotrophic protists regarding their genotypes and ecotypes. Protist 166, 42 -57.

Nitsche, F., Carr, M., Arndt, H., Leadbeater, B.S.C. 2011. Higher Level Taxonomy and Molecular Phylogenetics of the Choanoflagellatea. J. Eukaryot. Microbiol. 58, 452 -462.

Nylander, J.A.A. 2003. MrAIC.pl. Programm distributed by the author. Evolutionary Biology Centre, Uppsala University, Uppsala.

Pawlowski, J., Audic, S., Adl, S., Bass, D., Belbahri, L., Berney, C., 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., Nitsche, F., Romeralo, M., Saunders, G.W., Simpson, A.G., Smirnov, A.V., Spouge, J.L., Stern, R.F., Stoeck, T., Zimmermann, J., Schindel, D., de Vargas, C. 2012. CBOL Protist working group: Barcoding eukaryotic richness beyond the animal, plant, and fungal kingdoms. PLOS Biol. 10, doi:10.1371/journal.pbio.1001419.

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. 2015. Tara Oceans Consortium Coordinators. Open science resources for the discovery and analysis of Tara Oceans data. Scientific Data. 2, 150023. doi: 10.1038/sdata.2015.23.

R Core Team 2016. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Ronquist, F., Huelsenbeck, J.P. 2003. MrBayes 3: Bayesian phylogenetic inference under mixed models. Bioinformatics 19, 1572 -1574.

Shannon, P., Markiel, A., Ozier, O., Baliga, N.S., Wang, J.T., Ramage, D., Amin, N., Schwikowski, B., Ideker, T. 2003 Cytoscape: a software environment for integrated models of biomolecular interaction networks. Gen.Research 13, 2498-504.Stamatakis, A., Hoover, P., Rougemont, J. 2008. A Rapid Bootstrap Algorithm for the RAxML Web Servers. Syst. Biol. 57, 758 -771.

Sukhanova, I.N. 2001. Choanoflagellida on the southeastern Bering Sea shelf. Oceanology 41, 227 -231.

Tara Oceans Consortium, Coordinators; Tara Oceans Expedition, Participants (2014): Registry of selected samples from the Tara Oceans Expedition (2009-2013). doi:10.1594/PANGAEA.840721.

Thomsen, H.A. 1976. Studies on marine choanoflagellates II. Fine- structural observations on some silicified choanoflagellates from the Isefjord (Denmark), including the description of two new species. Norw. J. Bot. 23, 33 -51.

Thomsen, H.A. 1982. Planktonic Choanoflagellates from Disko Bugt, West Greenland, with a survey of the marine nanoplankton of the area. Commission for Scientific Research in Greenland.

Thomsen, H.A. 1992. Loricabærende choanoflagellater (Kraveflagellater). Plankton in Plankton i de indre danske farvande 11, 157 -194.

Thomsen, H.A., Buck, K.R., Coale, S.L., Garrison, D.I., Gowing, M.M. 1990. Loricate Choanoflagellates (Acanthoecidae, Choanoflagellida) from the Weddell Sea, Antarctica. Zool. Scr. 19, 367 -387.

Thomsen, H.A., Buck, K.R. 1991. Choanoflagellate diversity with particular emphasis on the Acanthoecidae. In: Patterson DJ, Larsen J (Eds.), The Biology of Free-Living Heterotrophic Flagellates. Clarendon Press, Oxford, pp. 259-184.

Thomsen, H.A., Garrison, D.L., Kosman, C. 1997. Choanoflagellates (Acanthoecidae, Choanoflagellida) from the Weddell Sea, Antarctica, taxonomy and community structure with particular emphasis on the ice biota; With preliminary remarks on choanoflagellates from Arctic Sea Ice (Northeast Water Polynya, Greenland). Arch. Protistenk. 148, 77 -114.

Throndsen, J. 1993. The planktonic marine flagellates. In: Tomas CR (Ed.), Marine Phytoplankton: a guide to naked flagellates and coccolithophorids. Academic Press, San Diego, pp. 7-145.

Throndsen, J. 1970. Marine planktonic Acanthoecaceans (Craspedophyceae) from Arctic waters. Nytt, Mag. Bot. 17, 103 -111.

Tong, S.M. 1997. Choanoflagellates in Southampton water, including the description of three new species. J. Mar. Biol. Assoc. UK 77, 929 -958.

Wickham, H 2009. ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York.

Figure 1. Schematic drawings of the species from this study. (A) *Bicosta minor,* (B) *Calliacantha natans,* (C) *C. longicaudata,* (D) *Cosmoeca ventricosa,* (E) *Crinolina isefiordensis,* (F) *Diaphanoeca undulata,* (G) *D. sphaerica,* (H) *Pleurasiga minima,* (I) *P. reynoldsii.* Scalebar 5µm

Figure 2. Light microscopical images of the species from this study. (A-C) *Bicosta minor,* (D-E) *Calliacantha longicaudata,* (F) *C. natans,* (G) *Cosmoeca ventriosa,* (H) *Crinolina isefiordensis,* (I-J) *Diaphanoeca sphaerica (*colony and single specimen)*, (K) *D. undulata,* (L) *Pleurasiga minima,* (M) *P. reynoldsii.* Scalebar for all images given in (F) with 5µm, except for (I) with different scale and scalebar at 5µm.

Figure 3. Molecular phylogeny of choanoflagellates based on a concatenated two-gene dataset. The tree shown was derived by Maximum likelihood analysis based on a combined SSU and LSU rDNA nucleotide sequence alignment. Branches are drawn proportional to the number of nucleotide substitutions per site as indicated by the scale bar at the lower left. Branches receiving 100% maximum likelihood bootstrap percentages (mlBP) and 1.00 Bayesian Inference posterior probabilities (biPP) support are denoted by an *. mlBP and biPP values are otherwise given above and below branches respectively, only values above 50% are shown.

Figure 4. Network diagrams of previously unidentified *Tara* Oceans V9 OTUs matching a newly sequenced species in this paper. Nodes represent individual barcodes and edges are present between barcodes if there is exactly one DNA sequence difference between them. The barcode within each OTU matching our new sequence is in red, and all other *Tara* Oceans barcodes are in blue. Node sizes are proportional to barcode global abundance.
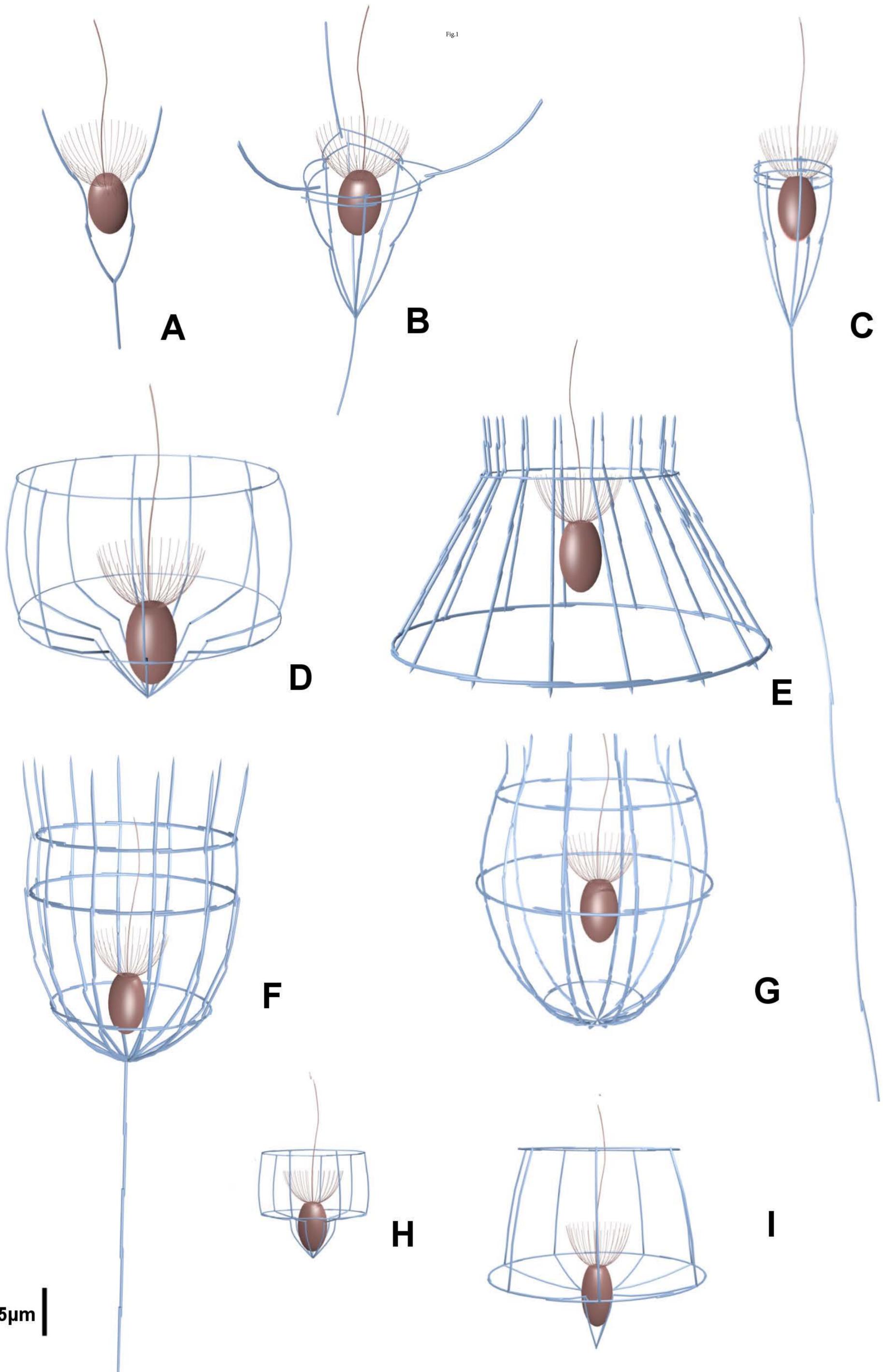
Figure 5 A-E. Presence (red) or absence (white) of newly sequenced species that were identified as matching a V9 OTU in the *Tara* Oceans data set within the 0.8-5 µm or 5-20 µm size fraction, either at the surface or at the deep chlorophyll maximum (DCM) for all *Tara* sampling stations with available data. A *Calliacantha natans*; B *Crinolina isefiordensis*; C *Diaphanoeca undulata*; D *Pleurasiga minima* and E

Pleurasiga reynoldsii. Although loricae for these species may measure larger than 5 or 20 µm, we display data for the 0.8-5 µm and 5-20 µm size fractions, since individual cells are expected to have been pulled out of their loricae during the filtration process that isolated individual size fractions.

Table 1. Relative abundances in percentage for the different sampling sites and species.

Table 2. Nucleotide sequences used to create the multigene phylogeny shown in Fig. 3. Reference Numbers are GenBank Accession.

Table 3. Matches of V9 18S barcode sequences in the *Tara* Oceans OTU data set. For each newly sequenced species, we searched the V9 portion of the 18S locus against the set of *Tara* Oceans V9 metabarcodes. *Tara* OTU Match %: the percent identity of the best match; *Tara* OTU Abundance Rank: the rank abundance of the matched *Tara* OTU within all choanoflagellate OTUs sequenced; *Tara* OTU GenBank Best Match: the best match within described choanoflagellate species in GenBank for the *Tara* V9 OTU; *Tara* OTU GenBank Match %: the percent identity of the match between the *Tara* OTU and the GenBank sequence.
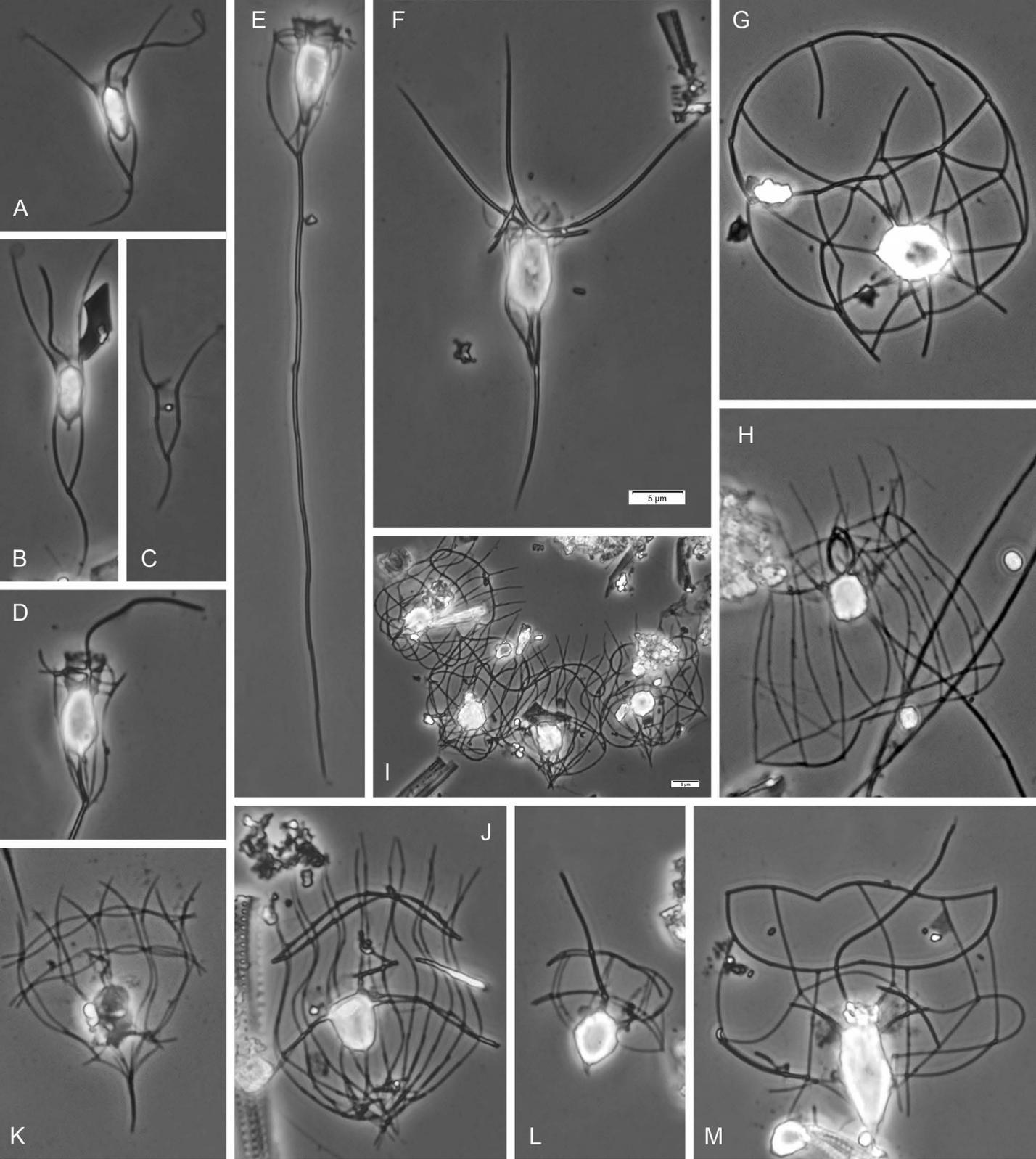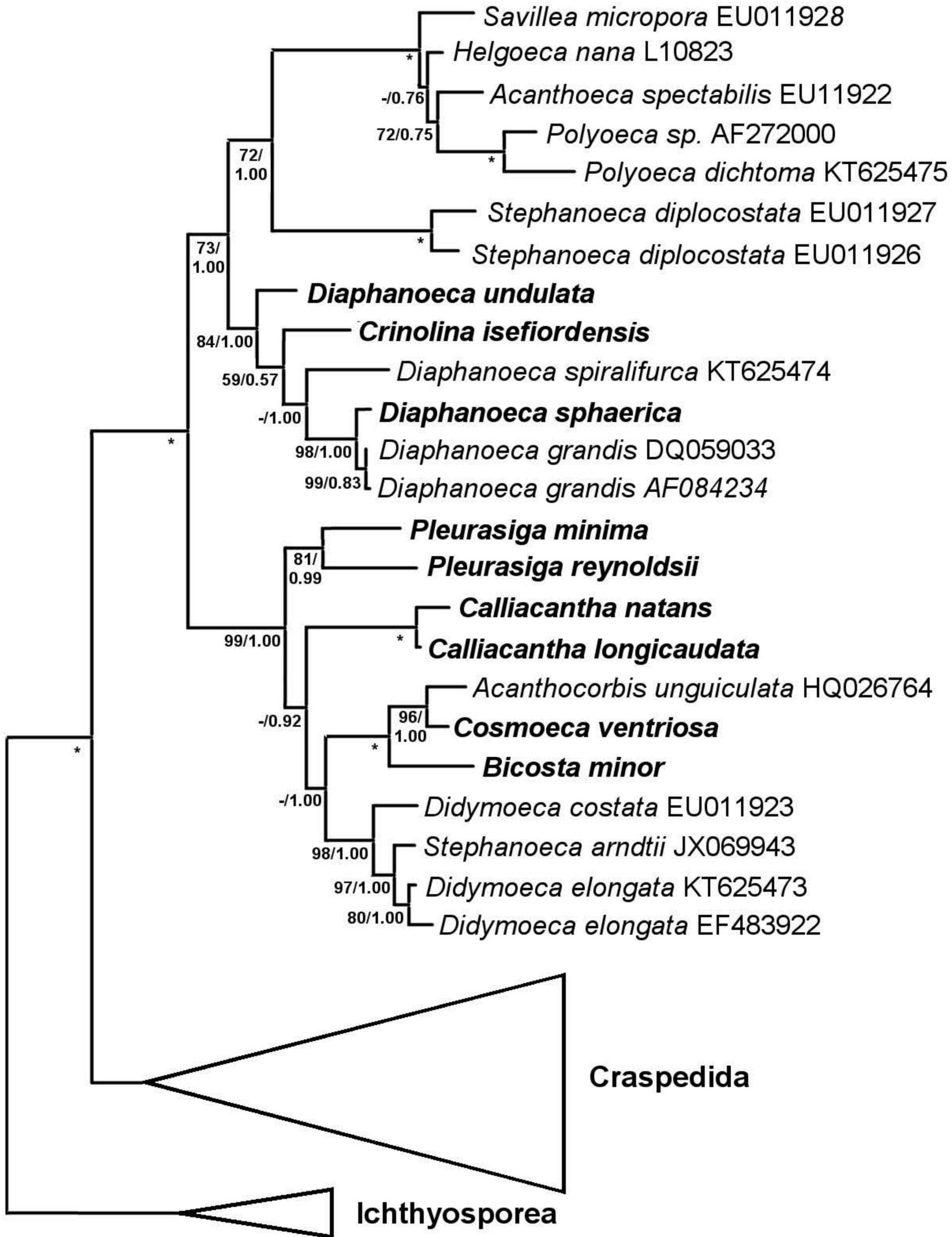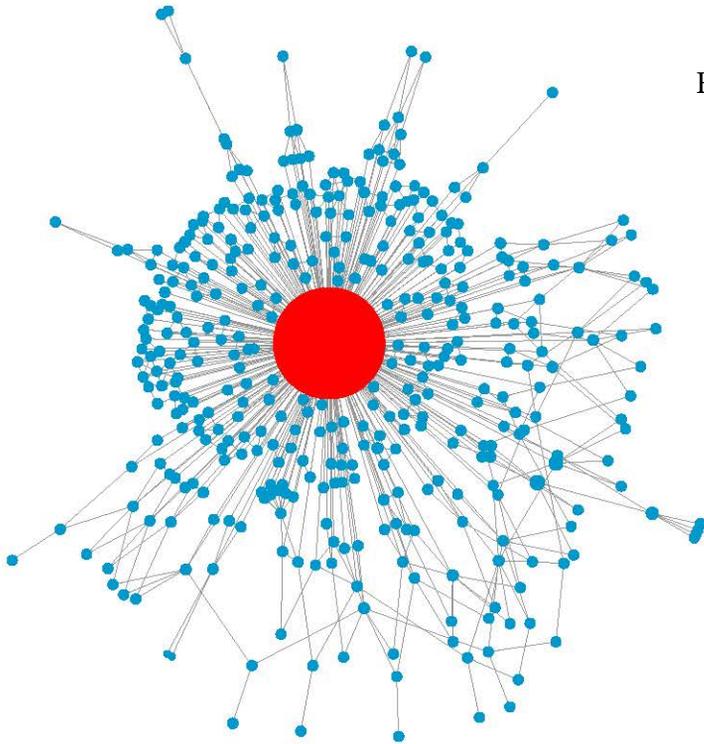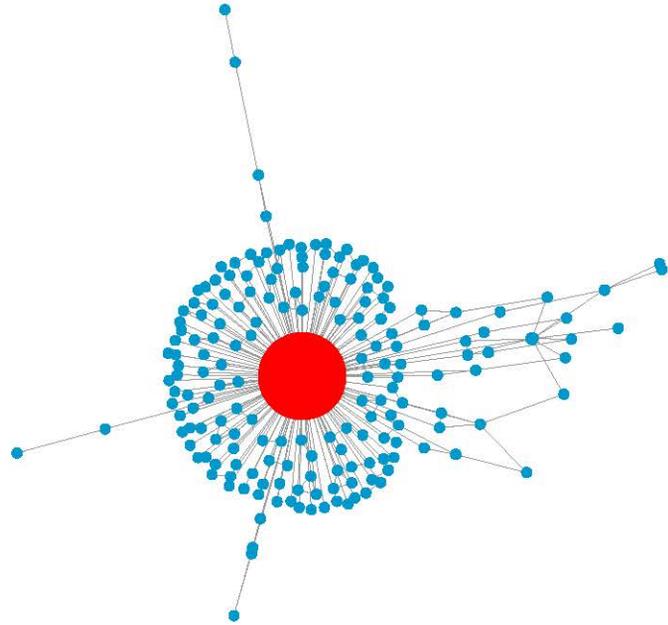
Fig.1

A

B

C

D

E

F

G

H

I

5μm

FIg.2

Fig.3

Savillea micropora EU011928
Helgoeca nana L10823
Acanthoeca spectabilis EU11922
Polyoeca sp. AF272000
Polyoeca dichtoma KT625475
Stephanoeca diplocostata EU011927
Stephanoeca diplocostata EU011926
Diaphanoeca undulata
Crinolina isefiordensis
Diaphanoeca spiralifurca KT625474
Diaphanoeca sphaerica
Diaphanoeca grandis DQ059033
Diaphanoeca grandis AF084234
Pleurasiga minima
Pleurasiga reynoldsii
Calliacantha natans
Calliacantha longicaudata
Acanthocorbis unguiculata HQ026764
Cosmoeca ventriosa
Bicosta minor
Didymoeca costata EU011923
Stephanoeca arndtii JX069943
Didymoeca elongata KT625473
Didymoeca elongata EF483922

Craspedida

Ichthyosporea

-/0.76
72/0.75
72/1.00
73/1.00
84/1.00
59/0.57
-/1.00
98/1.00
99/0.83
81/0.99
99/1.00
*
-/0.92
96/1.00
*
-/1.00
98/1.00
97/1.00
80/1.00
*

0.02

Fig.4

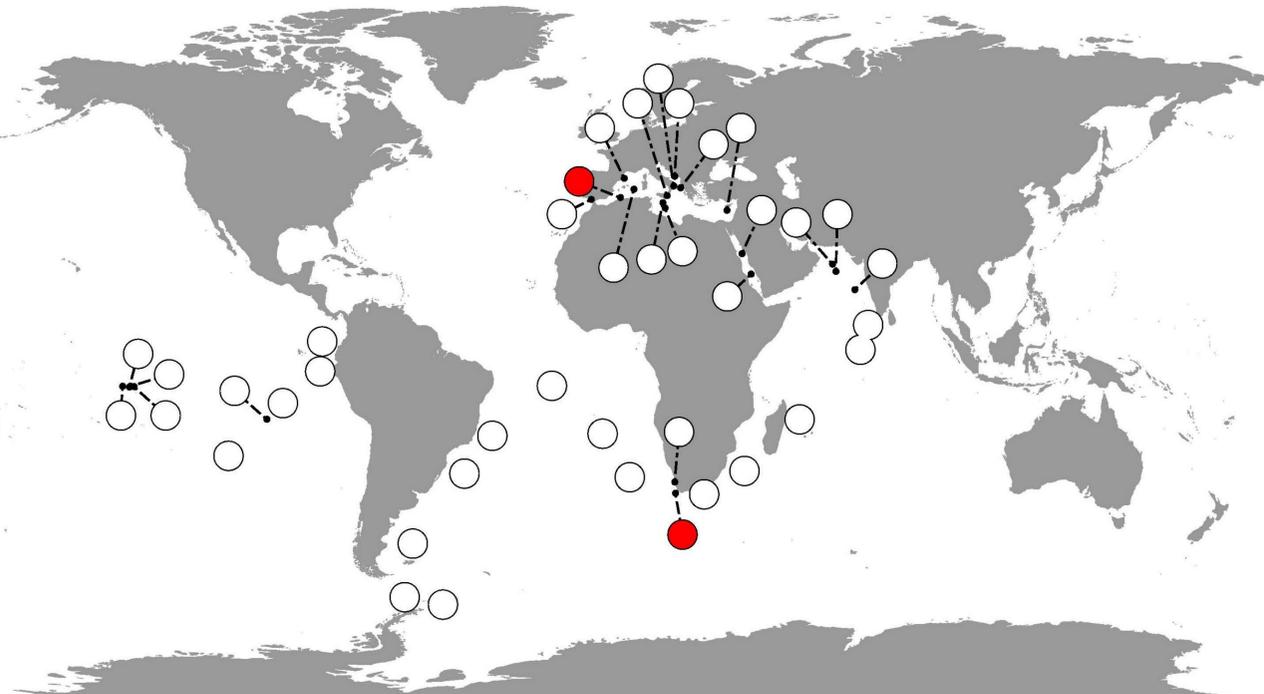*Calliacantha natans*

Diaphanoeca undulata

Pleurasiga minima

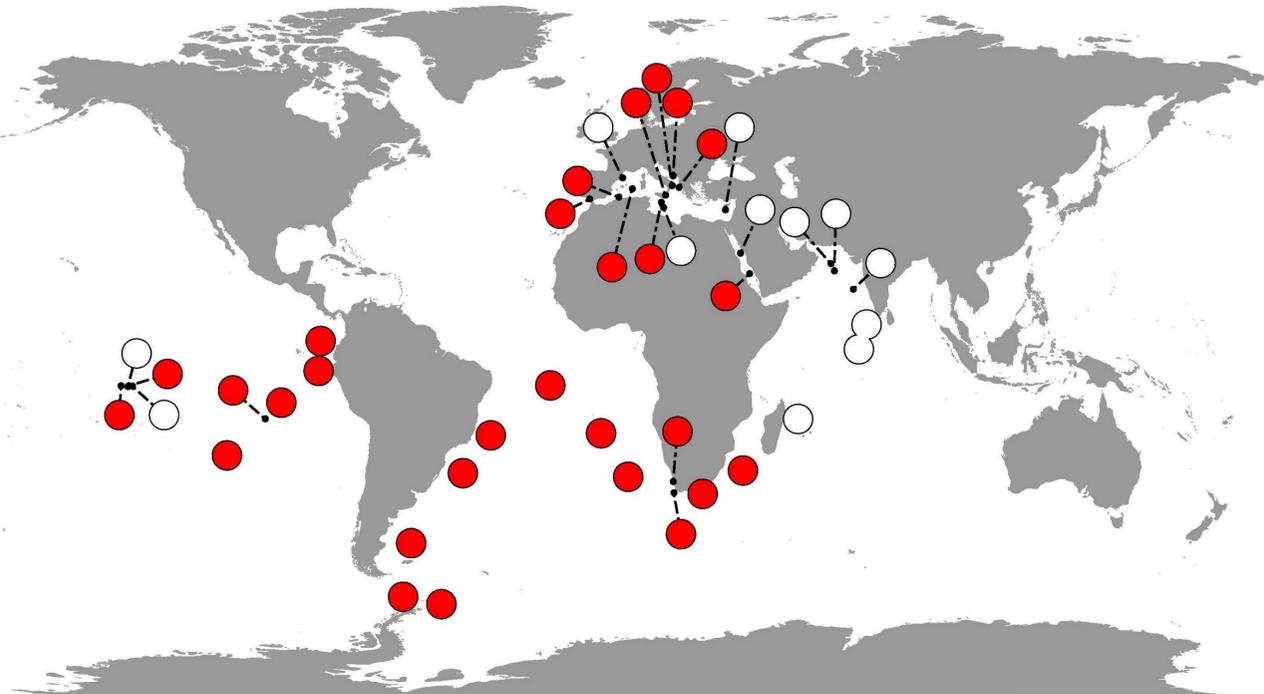*Pleurasiga reynoldsii*

Crinolina isefiordensis

Fig.5

A — *Calliacantha natans*

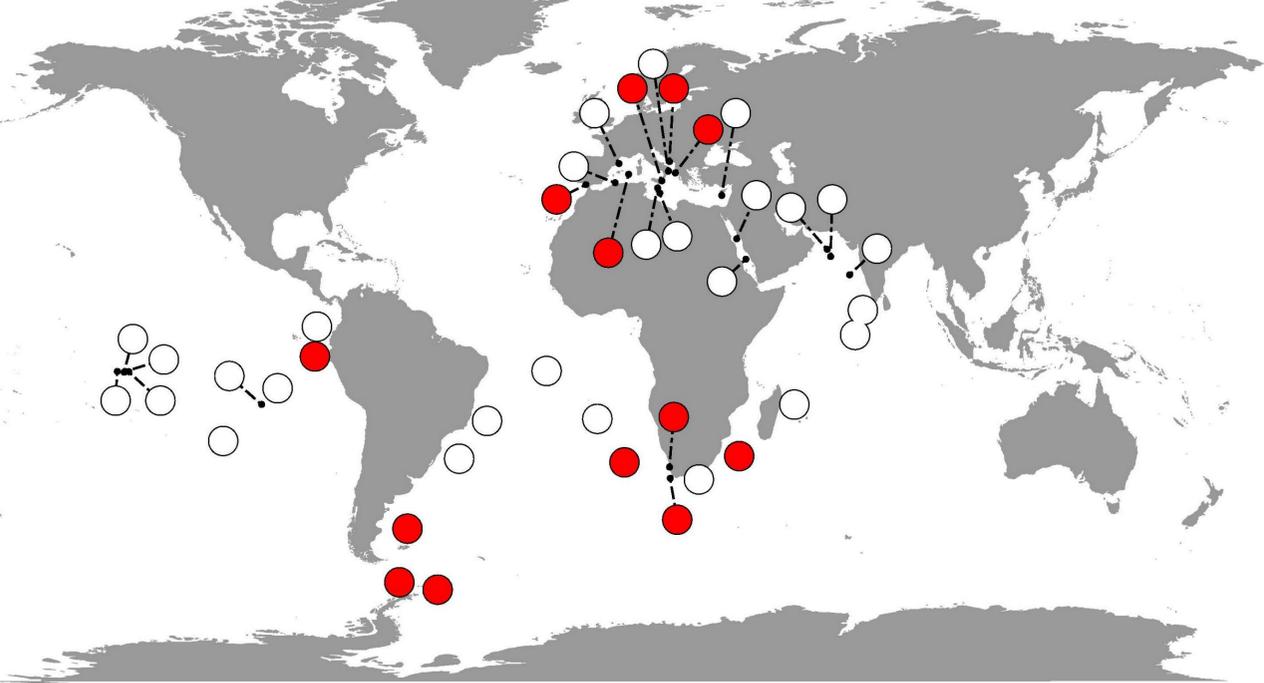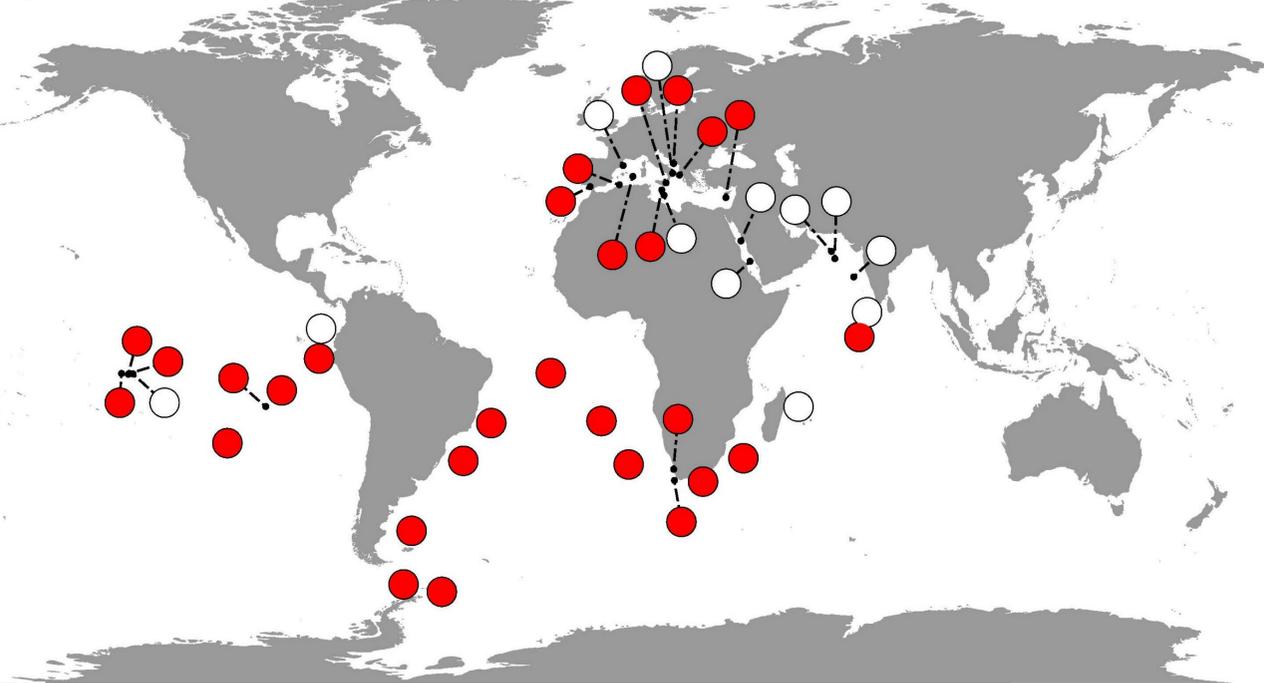**B** — *Crinolina isefiordensis*

C — *Diaphanoeca undulata*
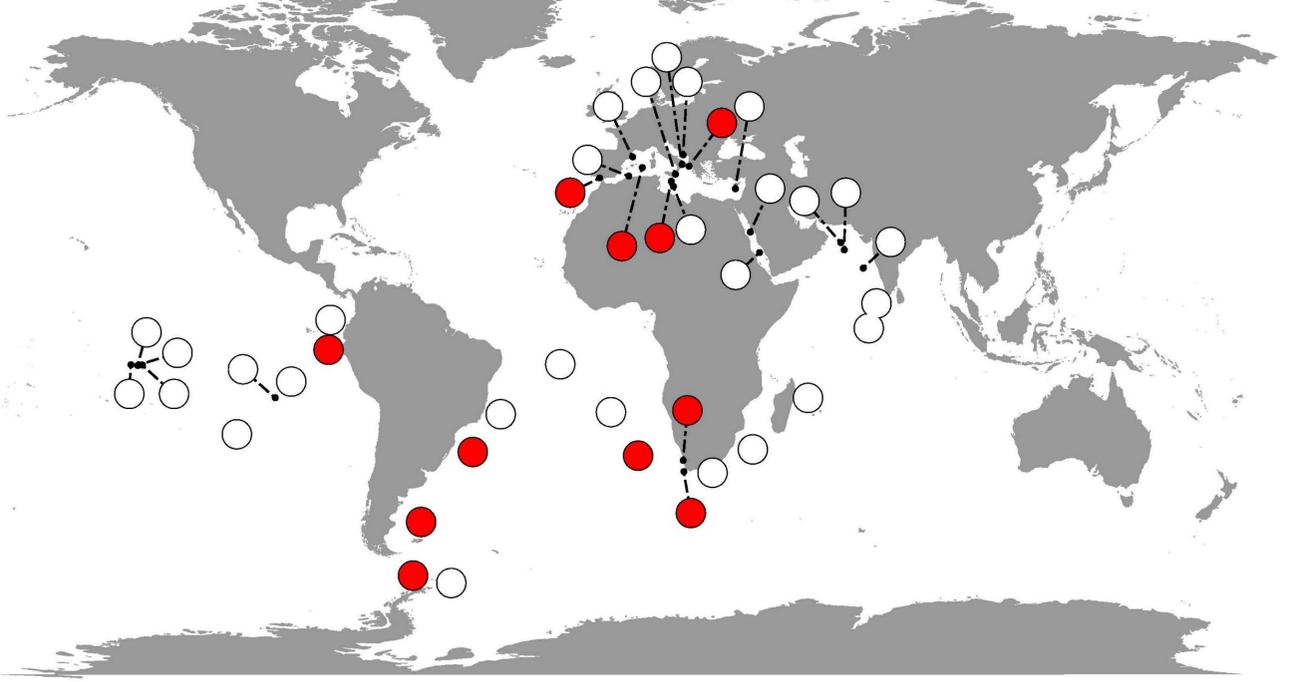
D — *Pleurasiga minima*

E — *Pleurasiga reynoldsii*

Table 1

	Station between Elsinore and Hälsingborg 24 Sep. 2014 20 m	Tempelkrogen 23 Feb. 2015 surface	Kyndby 23 Feb. 2015 surface
Bicosta minor (Reynolds 1976) Leadbeater 1978	2.0	14.9	**16.7**
Calliacantha longicaudata (Leadbeater 1975) Leadbeater 1978		13.3	**14.2**
Calliacantha natans (Grøntved 1956) Leadbeater 1978		**5.3**	**4.3**
Cosmoeca ventricosa Thomsen 1984	**20.3**		
Crinolina isefiordensis Thomsen 1976	3.7	**x**	
Diaphanoeca sphaerica Thomsen 1982		**3.2**	**4.3**
Diaphanoeca undulata Thomsen 1982	**2.4**		
Pleurasiga minima Throndsen 1970	1.2	**32.4**	33.3
Pleurasiga reynoldsii Throndsen 1970		**12.2**	6.8

Table 2

Species	SSU	LSU
Acanthoecida		
Acanthocorbis unguiculata	HQ026764	–
Acanthoeca spectabilis	AF084233	EU011933
Bicosta minor	**KU587839**	**–**
Calliacantha longicaudata	**KU587840**	**KU587841**
Calliacantha natans	**KU587842**	**KU587843**
Cosmoeca ventricosa	**KU587844**	**–**
Crinolina isefiordensis	**KU587845**	**–**
Diaphanoeca grandis	DQ059033	EU011939
Diaphanoeca grandis	AF084234	–
Diaphanoeca spiralifurca	KT625474	–
Diaphanoeca sphaerica	**KU587846**	**KU587847**
Diaphanoeca undulata	**KU587848**	**–**
Didymoeca costata	EU011923	EU011938
Didymoeca elongata	KT625473	–
Helgoeca nana	L10823	EU011934
Pleurasiga minima	**KU587849**	**KU587850**
Pleurasiga reynoldsii	**KU587851**	**KU587852**
Polyoeca sp.	AF272000	–
Polyoeca dichotoma	KT625475	–
Savillea micropora	EU011928	EU011944
Stephanoeca arndtii	JX069943	–
Stephanoeca diplocostata (Australia)	EU011926	EU011947
Stephanoeca diplocostata (France)	EU011927	EU011943
Craspedida		
Choanoeca perplexa	AF084232	EU011937
Codosiga gracilis	–	EU011935
Desmarella sp.	AF084231	–
Monosiga brevicollis	AF084618	AY026374
Monosiga sp. (ATCC 50635)	AF084230	EU011940
Salpingoeca abyssalis	DQ995808	–
Salpingoeca amphoridium	DQ059032	EU011942
Salpingoeca infusionum	AF100941	AY026380
Salpingoeca sp. (ATCC50938)	–	HQ026767
Salpingoeca sp. (ATCC50153)	EU011929	EU011945
Salpingoeca sp. (ATCC50929)	EU011930	EU011926
Salpingoeca rosetta (ATCC50818)	EU011924	EU011941
Salpingoeca sp. (ATCC50931)	HQ026770	HQ026771
Salpingoeca sp. (Mallorca)	HQ026772	–
Salpingoeca urceolata	EU011931	EU011948
Ichthyosporea		
Amoebidium parasiticum	Y19155	EU011932
Ichthyophonus hoferi	U25637	AY026370

Table 3

Species	*Tara* OTU Match %	*Tara* OTU Abundance Rank	*Tara* OTU GenBank Best Match	*Tara* OTU GenBank Match %
Bicosta minor	97	23	*Didymoeca costata*	83
Calliacantha longicaudata	96	7	*Didymoeca costata*	87
Calliacantha natans	100	2	*Didymoeca costata*	88
Cosmoeca ventricosa	99	15	*Didymoeca costata*	89
Crinolina isefiordensis	100	101	*Diaphanoeca grandis*	90
Diaphanoeca sphaerica	94	11	*Diaphanoeca grandis*	94
Diaphanoeca undulata	100	16	*Savillea micropora*	91
Pleurasiga minima	100	27	*Didymoeca costata*	92
Pleurasiga reynoldsii	100	39	*Didymoeca costata*	89