



University of HUDDERSFIELD

University of Huddersfield Repository

Carr, Martin, Richter, Daniel J., Fozouni, Parinaz, Smith, Timothy J., Jeuck, Alexandra, Leadbeater, Barry S.C. and Nitsche, Frank

A six-gene phylogeny provides new insights into choanoflagellate evolution

Original Citation

Carr, Martin, Richter, Daniel J., Fozouni, Parinaz, Smith, Timothy J., Jeuck, Alexandra, Leadbeater, Barry S.C. and Nitsche, Frank (2017) A six-gene phylogeny provides new insights into choanoflagellate evolution. *Molecular Phylogenetics and Evolution*, 107. pp. 166-178. ISSN 10557903

This version is available at <http://eprints.hud.ac.uk/30557/>

The University Repository is a digital collection of the research output of the University, available on Open Access. Copyright and Moral Rights for the items on this site are retained by the individual author and/or other copyright owners. Users may access full items free of charge; copies of full text items generally can be reproduced, displayed or performed and given to third parties in any format or medium for personal research or study, educational or not-for-profit purposes without prior permission or charge, provided:

- The authors, title and full bibliographic details is credited in any copy;
- A hyperlink and/or URL is included for the original metadata page; and
- The content is not changed in any way.

For more information, including our policy and submission procedure, please contact the Repository Team at: E.mailbox@hud.ac.uk.

<http://eprints.hud.ac.uk/>



Contents lists available at ScienceDirect

Molecular Phylogenetics and Evolution

journal homepage: www.elsevier.com/locate/ympev

A six-gene phylogeny provides new insights into choanoflagellate evolution



Martin Carr^{a,*}, Daniel J. Richter^{b,1,2}, Parinaz Fozouni^{b,3}, Timothy J. Smith^a, Alexandra Jeuck^c, Barry S.C. Leadbeater^d, Frank Nitsche^c

^aSchool of Applied Sciences, University of Huddersfield, Huddersfield HD1 3DH, UK

^bDepartment of Molecular and Cell Biology, University of California, Berkeley, CA 94720-3200, USA

^cUniversity of Cologne, Biocentre, General Ecology, Zulpicher Str. 47b, D-50674 Cologne, Germany

^dSchool of Biosciences, University of Birmingham, Edgbaston, Birmingham B15 2TT, UK

ARTICLE INFO

Article history:

Received 5 April 2016

Revised 12 October 2016

Accepted 17 October 2016

Available online 17 October 2016

Keywords:

Choanoflagellates

Craspedida

Genome evolution

Holozoa

Molecular phylogeny

Taxonomic revision

ABSTRACT

Recent studies have shown that molecular phylogenies of the choanoflagellates (Class Choanoflagellata) are in disagreement with their traditional taxonomy, based on morphology, and that Choanoflagellata requires considerable taxonomic revision. Furthermore, phylogenies suggest that the morphological and ecological evolution of the group is more complex than has previously been recognized. Here we address the taxonomy of the major choanoflagellate order Craspedida, by erecting four new genera. The new genera are shown to be morphologically, ecologically and phylogenetically distinct from other choanoflagellate taxa. Furthermore, we name five novel craspedid species, as well as formally describe ten species that have been shown to be either misidentified or require taxonomic revision. Our revised phylogeny, including 18 new species and sequence data for two additional genes, provides insights into the morphological and ecological evolution of the choanoflagellates. We examine the distribution within choanoflagellates of these two additional genes, EF-1A and EFL, closely related translation GTPases which are required for protein synthesis. Mapping the presence and absence of these genes onto the phylogeny highlights multiple events of gene loss within the choanoflagellates.

© 2016 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

1. Introduction

The choanoflagellates are a ubiquitous group of aquatic bacterivorous filter feeders (Arndt et al., 2000) and interest in their evolutionary biology has increased due to their recognized position as the sister-group to Metazoa in the eukaryotic supergroup Opisthokonta (Adl et al., 2012; Carr et al., 2008; Richter and King, 2013; Ruiz-Trillo et al., 2008). The opisthokonts are divided into two major lineages, one lineage being Holozoa, comprising Meta-

zoa and the protistan Choanoflagellata, Filasterea, Ichthyosporea plus Corallochytraea; the other lineage being Nucleotmycea (sometimes referred to as Holomycota) comprising Fungi and the nuclearioid amoebae (Adl et al., 2012).

1.1. Choanoflagellate taxonomy

It has long been acknowledged that the taxonomy of the choanoflagellates is in need of significant revision (Cavalier-Smith and Chao, 2003; Carr et al., 2008; Jeuck et al., 2014; Leadbeater et al., 2008; Medina et al., 2003; Nitsche et al., 2011; Stoupin et al., 2012). Choanoflagellate taxonomy has, in the past, been based upon morphological characters; in particular the external covering of the cell defined the three traditionally recognized families. Choanoflagellates possessing a solely organic cell cover were split into two families; Salpingoecidae Kent, which possessed a rigid theca, and Codonosigidae Kent, often called 'naked' choanoflagellates, which possessed a fine mucilaginous cover that is referred to as the glycocalyx. However, in molecular phylogenies neither group was recovered as monophyletic (Cavalier-Smith and Chao, 2003; Medina et al., 2003). Nitsche et al. (2011) showed that

* Corresponding author.

E-mail addresses: M.Carr@hud.ac.uk (M. Carr), daniel.j.richter@gmail.com (D.J. Richter), pfozouni@gmail.com (P. Fozouni), tim897@inbox.com (T.J. Smith), alexandra.jeuck@uni-koeln.de (A. Jeuck), b.s.c.leadbeater@bham.ac.uk (B.S.C. Leadbeater), FNitsche@uni-koeln.de (F. Nitsche).

¹ MC and DJR contributed equally to this work.

² Present addresses: 1 CNRS, UMR 7144, Station Biologique de Roscoff, Place Georges Teissier, 29680 Roscoff, France. 2 Sorbonne Universités, Université Pierre et Marie Curie (UPMC) Paris 06, UMR 7144, Station Biologique de Roscoff, Place Georges Teissier, 29680 Roscoff, France.

³ Present address: Medical Scientist Training Program, Biomedical Sciences Graduate Program, University of California, San Francisco, CA 94143, USA.

Codonosigidae is polyphyletic within Salpingoecidae and therefore synonymized the former with the latter within the order Craspedida Cavalier-Smith. The thecae of salpingoecids are found in a variety of morphologies; the most commonly observed are the flask (exemplified by *Choanoeca perplexa*, see Leadbeater, 1977), the cup (exemplified by *Salpingoeca rosetta*, see Dayel et al., 2011) and the tube (exemplified by *Salpingoeca tuba*, see Nitsche et al., 2011). Nitsche et al. (2011) also formally described two families of loricate choanoflagellates (order Acanthoecida) which produce cage-like silica baskets. The nudiform taxa were assigned to the Acanthoecidae Ellis *sensu* Nitsche et al. (2011), whilst tectiform taxa were assigned to a new family, Stephanoecidae Leadbeater.

We present here a molecular phylogeny containing 47 choanoflagellate species, created using a six-gene dataset. The six-genes are 18S small-subunit ribosomal RNA (SSU), 28S large-subunit ribosomal RNA (LSU), 90-kilodalton heat shock protein (hsp90), alpha-tubulin (tubA), elongation factor-1A (EF-1A, formerly EF-1 α) and elongation factor-like (EFL). The new phylogeny provides the basis to revise aspects of choanoflagellate taxonomy at the generic level; in particular we have amended the genus *Codosiga*. *Codosiga* currently comprises approximately 20 species of naked craspedids, which form stalked colonies. Most described taxa inhabit freshwater, with only four species (*Codosiga balthica* Wylezich and Karpov, *Codosiga cymosa* Kent, *Codosiga gracilis* (Kent) de Saedeleer, and *Codosiga minima* Wylezich and Karpov) recognized as marine (including brackish waters, and therefore defined as having a salinity >0.5 parts per thousand). On the basis of ribosomal RNA genes, Jeuck et al. (2014) recovered *Codosiga* species as being paraphyletic, with moderate to strong support, within Clade 1 craspedids, suggesting that the genus required taxonomic revision.

In addition to revising the taxonomy of *Codosiga*, we erect three additional new genera and formally describe the type species for each genus. The species are a naked craspedid erroneously deposited at the American Type Culture Collection (ATCC) under the name *Monosiga ovata* (ATCC 50635) and two distantly related, ovoid thecate craspedids. One of thecate species was isolated from the Mediterranean Sea for this study, whilst the second species was deposited at ATCC under the name *Salpingoeca* sp. (ATCC 50931). Nitsche et al. (2011) highlighted that a further four choanoflagellate species held in ATCC had been misidentified. We expand on this finding here and describe these four species, as well as five novel craspedid species.

1.2. Morphological, ecological and genomic evolution of choanoflagellates

The four-gene phylogenetic analysis of Carr et al. (2008) produced new insights into the evolution of choanoflagellates, but was hindered by only containing 16 species. The 47-taxon phylogeny presented here provides unprecedented insights into the morphological, ecological and genomic evolution of the choanoflagellates. At the genomic level, we concentrate on the inheritance of two paralogous GTPases, EF-1A and EFL. EF-1A is a major component of the eukaryotic protein synthesis machinery. Due to its importance in protein translation and its involvement in multiple additional pathways (Gaucher et al., 2001), EF-1A was considered an essential and ubiquitously distributed protein. It was therefore a considerable surprise when it was discovered that a number of eukaryotic taxa lacked EF-1A (Keeling and Inagaki, 2004). Those species which do not possess EF-1A have been shown to encode EFL, with very few eukaryotes known to possess both genes (Atkinson et al., 2014; Henk and Fisher, 2012; Kamikawa et al., 2013). The screening of a wide diversity of eukaryotic whole genomes has shown that very few organisms possess both genes.

EFL has a punctate distribution within eukaryotes and phylogenies based on EFL sequences are incongruent with accepted species phylogenies (Keeling and Inagaki, 2004; Noble et al., 2007). It has been speculated that both EFL and EF-1A were present in the genome of the eukaryotic last common ancestor (LCA) and that one of the genes has been subsequently retained in favour over the other in different eukaryotic lineages (Kamikawa et al., 2013). However, the incongruities between EFL and species phylogenies have led to the suggestion that EFL has undergone repeated rounds of lateral transfer into new hosts and, on occasion, replaced the endogenous EF-1A (Kamikawa et al., 2008, 2010a). EFL has previously been sequenced from representatives of Fungi, Choanoflagellata and Ichthyosporea (Keeling and Inagaki, 2004; Noble et al., 2007; Marshall and Berbee, 2010; Ruiz-Trillo et al., 2006), although each of these lineages also contains taxa which encode EF-1A. EFL appears to be absent from metazoans, with all studied species encoding EF-1A. Within the choanoflagellates EF-1A has been shown to be present in three freshwater craspedids, these being *Codosiga botrytis*, ATCC 50635 and ATCC 50153 (the latter two species whose names we revise as part of this work) (Atkinson et al., 2014; Paps et al., 2013; Steenkamp et al., 2006), whilst EFL has been found in *Monosiga brevicollis* ATCC 50154 and *Salpingoeca rosetta* ATCC 50818 (Atkinson et al., 2014; Noble et al., 2007). Combining newly generated and publicly available data, we identify the presence of EFL and EF-1A in 22 choanoflagellate species and speculate upon the inheritance of both genes within the group.

2. Materials and methods

2.1. Isolation of choanoflagellate species and rRNA Gene Sequencing

The species isolated and sequences generated for this study are listed in Table S1. For *Codosiga hollandica*, *Salpingoeca calixa*, *Salpingoeca oahu* and *Stagondoea pyriformis* DNA amplification was performed using single cell PCR (Nitsche and Arndt, 2008) and additionally applying the 28S large subunit ribosomal RNA (LSU) primers described in a previous study (Carr et al., 2008). The sequencing of LSU was performed using Big Dye-Terminator v3.1 Cycle Sequencing Kit (Applied Biosystems, Weiterstadt, Germany) in accordance with the manufacturer's instructions. Both strands obtained were tested for consistency.

For the remaining species, PCR of SSU was performed with universal eukaryotic ribosomal primers 18S_1F and 18S_1528R (Medlin et al., 1988). PCR products were separated by gel electrophoresis and extracted with the QIAquick Gel Extraction Kit from Qiagen, followed by cloning using the TOPO TA Cloning vector from Invitrogen, both following the manufacturer's protocol. A single clone was selected and Sanger sequencing reads were generated using two primers within the vector sequence: M13F (5' GTAAAACGACGGCCAGT 3') and M13R (5' CAGGAAACAGCTATGACC 3'). Internal sequence was generated using the pan-choanoflagellate SSU sequencing primers 18S_564F (5' AATTC-CAGCTCCAATAGC 3') and 18S_1205R (5' ATGTCTGGACCTGGTGAG 3'). Sequence reads were base called using phred version 0.0210425.c (Ewing and Green, 1998; Ewing et al., 1998) with default parameters, and aligned using FSA version 1.15.0 (Bradley et al., 2009) with default parameters.

Salpingoeca dolichothecata ATCC 50959 PCR, cloning and sequencing were performed using the same protocol with one modification. Additional SSU sequence was generated to bypass an internal polynucleotide repeat by sequencing the cloned PCR product with the following three ATCC 50959-specific primers: Sg_18S_1 (5' CCTTCAGACGAATCATCGAG 3'), Sg_18S_2 (5' TGA-GAACAAACCGCAAAG 3') and Sg_18S_3 (5' AATGCTGCTTTGAA-CACTC 3').

2.2. Transcriptome data

We augmented our data from PCR and sequencing by searching *de novo* transcriptome data from 19 choanoflagellate species (Richter, 2013). For SSU, LSU, hsp90 and tubA, we downloaded all available choanoflagellate data from GenBank and built multiple sequence alignments using FSA (Bradley et al., 2009) with default parameter values. HMM profiles were created for EF-1A and EFL using MAFFT 6.935 (Katoh et al., 2002) nucleotide alignments, each generated from eight genes (see Table S2).

Unaligned regions were removed using Gblocks (Talavera and Castresana, 2007) with allowed gap positions set to “half” and all other parameter values set to their most permissive. HMMs were created using hmmbuild from the HMMER 3.0 package (<http://hmmmer.org/>) with default parameter values. The assembled transcriptome for each species and its reverse complement were screened using hmmssearch of HMMER, with default parameter values. We chose the contig with the lowest E value as the representative sequence for that species. If there were multiple contigs with the same lowest E value, we chose the longest of those contigs.

2.3. Phylogenetic and molecular evolution analyses

The phylogeny of the choanoflagellates was analysed using SSU, LSU, hsp90, tubA, EFL and EF-1A (Table S3). For each gene, DNA sequences from all available species were aligned in MAFFT and then edited by eye to minimize insertion-deletion events. A single-gene EFL phylogeny presented a topology consistent with previously published choanoflagellate phylogenies, indicating the gene was present in the choanoflagellate LCA and confirming its suitability for phylogenetic reconstruction.

Highly divergent EF-1A transcripts were recovered from six species which also expressed EFL. These sequences show an elevated rate of evolution and loss of functional constraint (see Results), so they were not included in the multigene species phylogeny or the EF-1A protein phylogeny.

The concatenated 9767 bp alignment was analysed using maximum likelihood and Bayesian inference methods. For both analyses the alignment was divided into three separate partitions; the first for ribosomal RNA, the second being 1st and 2nd codon positions together and the final being 3rd codon positions. All parameters for the phylogenetic analyses were estimated by each program. The maximum likelihood analysis was performed with RAxML 7.2.6 (Stamatakis, 2006) using the GTRCAT model, as recommended by jmodeltest (Darriba et al., 2012). The analysis was initiated with 100 maximum parsimony trees and bootstrapped with 1000 replicates. The Bayesian analysis was performed using MrBayes 3.1.1 (Ronquist and Huelsenbeck, 2003) and run using a GTR + I + Γ model and a four-category gamma distribution to correct for among site rate variation. The search consisted of two parallel chain sets run at default temperatures with a sample frequency of 1000 and run so that the average standard deviation of split frequencies dropped below 0.01. The analysis consisted of 5,000,000 generations, with a burnin of 1250 before calculating posterior probabilities. The choanoflagellate phylogeny was rooted with a two-taxa ichthyosporean clade and an eight-taxa metazoan clade.

Additional analyses were undertaken in order to investigate marine:freshwater transitions. Publicly available SSU sequences from the 10 freshwater clades (Table S4) in the del Campo and Ruiz-Trillo (2013) study were incorporated into the 6-gene dataset, with the SSU partition re-aligned in MAFFT, in order to identify freshwater incursions. 100 bootstrap replicates were used for the RAxML analysis, otherwise the 6-gene alignments were analysed following the protocols for the main phylogeny. In addition to the full dataset, additional analyses were performed with

alignments with length cut-offs of 750 and 1000 aligned and edited nucleotides in the SSU alignment.

Predicted amino acid sequences of EFL and EF-1A (Table S5) were recovered from GenBank using both keyword searches and BLASTp similarity searches with conceptual choanoflagellate protein sequences. Sequence recovery for EF-1A was restricted to Opisthokonta, whereas EFL sequences were recovered from all available eukaryotic groups. Alignments for each protein family were created using MAFFT and edited by eye. ProtTest 3.2.2 (Abascal et al., 2005) indicated that the LG + I + Γ + F (Le and Gascuel, 2008) was the most appropriate amino acid substitution model for both EF-1A and EFL. Maximum likelihood phylogenies for both protein families were created using RAxML GUI 1.3 (Michalak, 2012). Each analysis was performed with 25 rate categories, initiated with 100 parsimony trees and bootstrapped with 1000 replicates. Bayesian Inference phylogenies for both families were created using MrBayes 7.2.7 on the Cipres Science Gateway 3.3 (Miller et al., 2010). The searches used a mixed amino acid model and consisted of two parallel chain sets run at default temperatures with a sample frequency of 100. The analysis consisted of 5,000,000 generations, with a burnin of 12,500, before calculating posterior probabilities.

To determine nucleotide substitution rates at synonymous and non-synonymous sites for EF-1A and EFL, nucleotide alignments were created in MAFFT using all identified choanoflagellate sequences for each gene. The alignments were edited to remove all codons which contained gaps and nucleotide substitution rates were determined using the online KaKs Calculation Tool housed by the Computational Biology Unit, University of Bergen (<http://services.cbu.uib.no/tools/kaks>). Phylogenetic trees, using the topology of our six-gene phylogeny, were provided, but otherwise default parameters were used.

All alignments are available from the corresponding author upon request. All new sequences have been deposited into GenBank under the Accession Numbers KT757415–KT757519, KT768096–KT768098 and KT988065.

3. Results

3.1. Species misidentification and taxonomic revisions within *Craspedida*

Nitsche et al. (2011) highlighted 23 choanoflagellate species misidentifications both within culture collections and DNA databases, however no attempt was made in that work to revise choanoflagellate taxonomy at the species or generic level. In addition to describing four new craspedid genera and five new species, we take the opportunity here to clarify the taxonomic descriptors of a further ten craspedid species (Fig. 1, Table 1 and Taxonomic Summary).

3.2. Phylogenetic analyses of 47 choanoflagellate species

It has previously been speculated that EFL may have undergone horizontal transfer between eukaryotic groups (Gile et al., 2006; Keeling and Inagaki, 2004). Single gene, EFL nucleotide ML and Bayesian inference phylogenies were created in order to identify any possible horizontal transfer events and determine the suitability of the gene for choanoflagellate phylogenetic reconstruction. The resulting phylogenies were consistent with previously published choanoflagellate phylogenies (data not shown), indicating that the gene has been inherited vertically throughout the choanoflagellate radiation and has not undergone horizontal transfer within the group. This finding confirms the suitability of EFL for choanoflagellate phylogenetics.

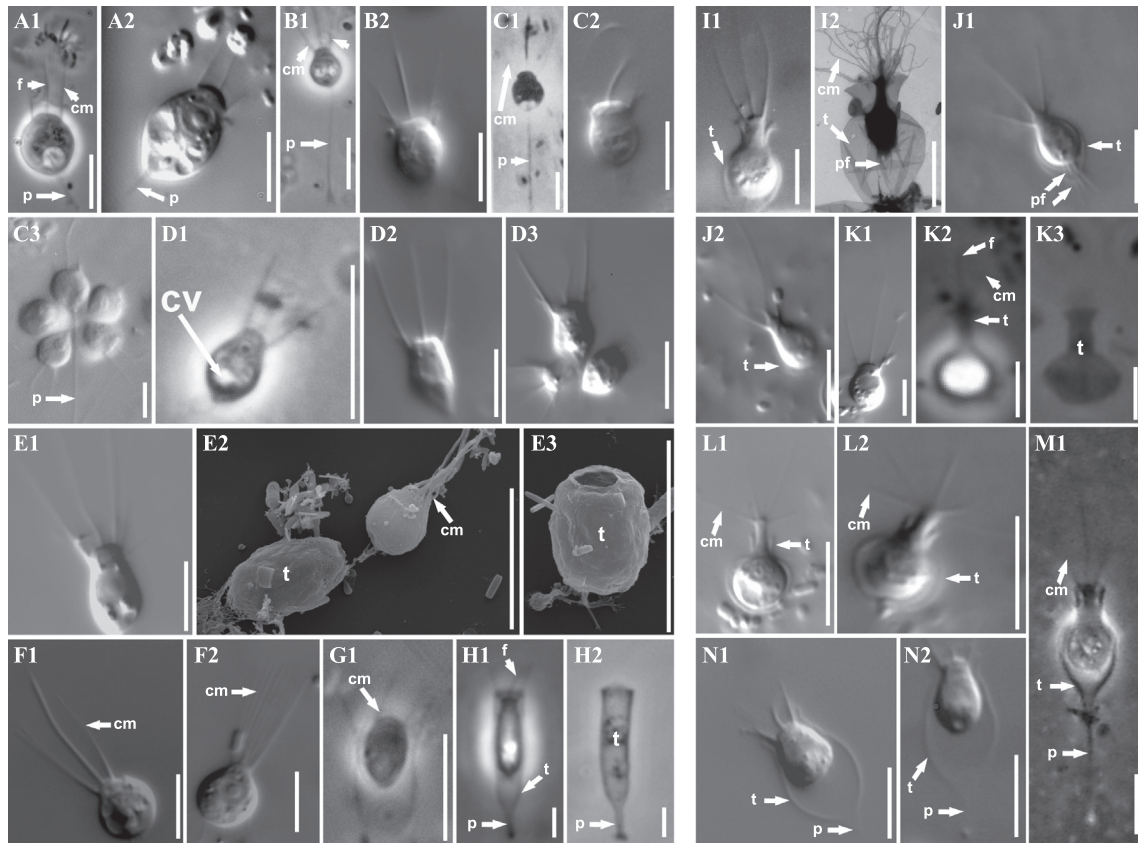


Fig. 1. Choanoflagellate species described in this study. Extracellular structures are indicated where they may otherwise be difficult to discern, using the following abbreviations: 'cm' = collar microvilli, 'cv' = contractile vacuole, 'f' = flagellum, 'p' = peduncle, 'pf' = posterior filopodia. All scale bars are 5 μ m. (A) *Codosiga hollandica*. A1: single cell on peduncle, phase contrast, A2: single cell on peduncle, DIC. (B) *Hartaetosiga balthica*. B1: single cell on long peduncle, phase contrast (image originally published in Wylezich et al. (2012)), B2: single cell without peduncle, DIC. (C) *Hartaetosiga gracilis*. C1: single cell on long peduncle, phase contrast (image originally published in Leadbeater and Morton (1974)), C2: single cell without peduncle, DIC, C3: colony of cells on a single long peduncle, DIC. (D) *Hartaetosiga minima*. D1: single cell without peduncle, phase contrast (image originally published in Wylezich et al. (2012)), D2: single cell without peduncle, DIC, D3: top view, colony of cells on a single peduncle, DIC. (E) *Microstomoeca roanoka*. E1: non-thecate single cell, DIC, E2: single cell and empty theca, SEM, E3: empty theca, SEM. (F) *Mylnosiga fluctuans*. F1 and F2: single cells, DIC. (G) *Salpingoeca calixa*. G1: thecate single cell, phase contrast. (H) *Salpingoeca dolichothecata*. H1: thecate single cell, phase contrast, H2: empty theca, phase contrast. (I) *Salpingoeca helianthica*. I1: thecate single cell, DIC, I2: thecate single cell, shadowcast whole mount TEM. (J) *Salpingoeca kvevrii*. J1: thecate single cell, DIC, J2: thecate single cell, DIC. (K) *Salpingoeca macrocollata*. K1: non-thecate single cell, DIC, K2: thecate single cell, phase contrast, K3: empty theca, phase contrast. (L) *Salpingoeca punica*. L1 and L2: thecate single cells lacking flagella, DIC. (M) *Salpingoeca oahu*. M1: thecate single cell on long peduncle, phase contrast. (N) *Stagondoeca pyriformis*. N1 and N2: thecate single cells, DIC.

The newly generated gene sequences were incorporated into a six-gene phylogenetic framework, in an alignment with sequences from 57 holozoan taxa (of which 47 were choanoflagellates). The resulting phylogeny is shown in Fig. 2. Consistent with previous studies, the choanoflagellates were recovered as monophyletic with strong support (1.00 Bayesian inference posterior probability (biPP) and 99% maximum likelihood bootstrap percentage (mlBP)), as were both Craspedida (1.00 biPP, 91% mlBP) and Acanthoecida (1.00 biPP, 100% mlBP).

Carr et al. (2008) showed that the root of the acanthoecids was recovered at different positions depending on whether the third positions of codons were included or omitted from the analysed dataset. With the inclusion of third codon positions, the acanthoecid root was recovered between the nudiform and tectiform groups, whereas if third positions are excluded an alternative root within the tectiform taxa was recovered. The greatly expanded dataset here, in terms of both gene and species number, recovers the same alternative root when third positions are excluded (Fig. S1), albeit with lower support values. An amino acid phylogeny created with concatenated sequences of Hsp90, TubA, EFL and EF-1A also recovered the nested position of the nudiform species within the tectiform species (data not shown). The nudi-

form species are recovered as monophyletic in all datasets; the monophyly of the tectiform species however cannot be considered robustly resolved on phylogenetic grounds.

3.3. Resolution of the polyphyly of *Codosiga*

The SSU sequence of ATCC 50964 (deposited as *Monosiga gracilis*) showed 99.8% identity to the published SSU of *Codosiga balthica*, isolated from the Gotland Deep, Baltic Sea (Wylezich et al., 2012). This demonstrates that ATCC 50964 is a North American isolate of *Codosiga balthica* and challenges the previously proposed endemism of this taxon (Wylezich et al., 2012).

The nine species currently attributed to *Codosiga* are found in two distinct positions within Craspedida. The marine *Codosiga balthica*, *Codosiga gracilis* and *Codosiga minima* are found in a group which corresponds to Clade 1 of Carr et al. (2008) and form a monophyletic group with strong support (1.00biPP, 100%mlBP, Fig. 2). The freshwater *Codosiga* species form a monophyletic group (1.00biPP, 100%mlBP), which is robustly nested deeply within the Clade 2 craspedids. Thus, the two groups of *Codosiga* taxa are shown to be distantly related and are separated from each other by nine branches in the Fig. 2 phylogeny. The genus *Codosiga* is

Table 1
Taxa described in this study.

Species name	ATCC/NCBI identifier	Previous identifier
<i>Codosiga hollandica</i> Carr, Richter and Nitsche	ATCC PRA-388/1742608	Newly described species
<i>Hartaetosiga balthica</i> (Wylezich and Karpov) Carr, Richter and Nitsche	ATCC 50964/1194294	<i>Codosiga balthica</i> Wylezich and Karpov; <i>Monosiga gracilis</i>
<i>Hartaetosiga cymosa</i> (Kent) Carr, Richter and Nitsche	n/a	<i>Codosiga cymosa</i> Kent
<i>Hartaetosiga gracilis</i> (Kent) Carr, Richter and Nitsche	ATCC 50454/216892	<i>Codosiga gracilis</i> Kent
<i>Hartaetosiga minima</i> (Wylezich and Karpov) Carr, Richter and Nitsche	n/a/1194293	<i>Codosiga minima</i> Wylezich and Karpov
<i>Microstomoeca roanoka</i> Carr, Richter and Nitsche	ATCC 50931/1051735	<i>Salpingoeca</i> sp. (ATCC 50931)
<i>Mylnosiga fluctuans</i> Carr, Richter and Nitsche	ATCC 50635/81526	<i>Monosiga ovata</i>
<i>Salpingoeca calixa</i> Carr, Richter and Nitsche	n/a/1742604	Newly described species
<i>Salpingoeca dolichothecata</i> Carr, Richter and Nitsche	ATCC 50959/1742605	<i>Salpingoeca gracilis</i>
<i>Salpingoeca helianthica</i> Carr, Richter and Nitsche	ATCC 50153/1042118	<i>Salpingoeca napiformis</i>
<i>Salpingoeca kvevrii</i> Carr, Richter and Nitsche	ATCC 50929/473812	<i>Salpingoeca pyxidium</i>
<i>Salpingoeca macrocollata</i> Carr, Richter and Nitsche	ATCC 50938/1009460	<i>Salpingoeca minima</i>
<i>Salpingoeca oahu</i> Carr, Richter and Nitsche	n/a/1742606	Newly described species
<i>Salpingoeca punica</i> Carr, Richter and Nitsche	ATCC 50788/334868	<i>Salpingoeca amphoridium</i>
<i>Stagondoeca pyriformis</i> Carr, Richter and Nitsche	n/a/1742607	Newly described species

therefore clearly not recovered as monophyletic, with the polyphyly of the genus being a more parsimonious explanation than *Codosiga* paraphyly (2 unweighted parsimony steps rather than 8 unweighted parsimony steps).

The type species, *Codosiga botrytis*, is a member of the freshwater clade and accordingly the freshwater species retain the generic name. The marine taxa should no longer be considered as members of *Codosiga*. The marine species are naked, stalked craspedids, nested within thecate taxa, therefore they cannot be placed into any of the genera of closely related species. We consequently erect a new genus, *Hartaetosiga*, to accommodate them (see Taxonomic Diagnoses).

Finally, an undescribed freshwater, naked, craspedid, isolated from the Atlantic island of Madeira, is recovered with strong support (1.00biPP, 100%mlBP, Fig. 2) with the freshwater *Codosiga* species. We describe the species as *Codosiga hollandica* (see Fig. 1A1, Taxonomic Diagnoses).

3.4. Resolution of the polyphyly of *Monosiga*

The two species attributed to the genus *Monosiga*, namely *Monosiga brevicollis* and *Monosiga* sp. ATCC 50635, are recovered as distantly related to each other in clades 1 and 2 of the craspedids. ATCC 50635, which was deposited under the name *Monosiga ovata*, is a marine species. In contrast, *Monosiga ovata* as described by Kent (1880–82) is a freshwater species, with a different periplast morphology, and is unlikely to be the same species (see Taxonomic Diagnoses). ATCC 50635 is a naked craspedid, nested within thecate species; therefore it does not possess the diagnostic characteristics of any closely related genera. Accordingly, we attribute ATCC 50635 to a new genus, *Mylnosiga*, as the type species *Mylnosiga fluctuans*, in order to clarify the taxonomy of *Monosiga*.

3.5. Classification of novel thecate choanoflagellates

ATCC 50931 (deposited under the name *Salpingoeca* sp.), isolated in 2000 from a Virginian saltmarsh, possesses a novel ovoid thecal morphology with a small, circular aperture at the top of the periplast (Fig. 1E2 and E3). Due to the distinctive theca morphology, which is unlike those of previously described genera, ATCC 50931 has been allocated its own genus, *Microstomoeca*, and is described under the name *Microstomoeca roanoka* (see Taxonomic Diagnoses).

Salpingoeca tuba and ATCC 50959 (deposited under the name *Salpingoeca gracilis*) are the first tube thecate species included in a multi-gene framework. ATCC 50959 was previously shown to have been misidentified (Nitsche et al., 2011) and we describe it here under the name *Salpingoeca dolichothecata* (see Fig. 1H1 and H2, Taxonomic Diagnoses). The two tube thecate species form a robust grouping (1.00biPP, 100%mlBP, Fig. 2) and are recovered at the base of Craspedida, as a sister-group to the other craspedids with strong support (1.00biPP, 91%mlBP). We propose the informal label, Clade 3 craspedids, for the group of tube thecate species.

A marine thecate species from Bálós Lagoon, Crete was shown to have an ovoid thecal morphology (see Fig. 1N1 and N2). In contrast to the ovoid theca of *M. roanoka* the theca has a short neck, somewhat similar to those of flask bearing species, and narrow anterior aperture from which the collar and flagellum extend. At the posterior pole, the theca tapers into a short peduncle. The ovoid morphology of the theca, which exhibits a neck, does not resemble that of any other genus; therefore a new genus, *Stagondoeca*, is erected to accommodate this species. Due to the theca having a droplet-like shape, this species has been described under the name *Stagondoeca pyriformis* (see Taxonomic Diagnoses). *Stagondoeca pyriformis* does not cluster with any other genera in the six-gene phylogeny and is recovered at the base of the Clade 1 craspedids with strong support (1.00biPP, 99%mlBP).

Partial fragments of both SSU and LSU were sequenced from a novel freshwater, flask species. The species was isolated from O'ahu, Hawaii and has been named as *Salpingoeca oahu* (see Fig. 1M1, Taxonomic Diagnoses). The species falls into a large paraphyletic group of freshwater species, which is recovered with moderate to strong support in our phylogeny (1.00biPP, 74%mlBP, Fig. 2). The SSU sequence of *Salpingoeca oahu* shows 99.6% identity to an environmental SSU sequence (EU860868) amplified from groundwater in the Netherlands (Fig. S2E), suggesting that this species may have a cosmopolitan freshwater distribution.

Furthermore, a thecate species from McKenzie Bay, New Zealand, which possesses a high-walled cup, is described under the name *Salpingoeca calixa* (see Fig. 1G1, Taxonomic Diagnoses). *Salpingoeca calixa* is recovered with strong support (1.00biPP, 83%mlBP, Fig. 2) with two other cup thecate species (*Salpingoeca infusionum* and *Salpingoeca longipes*); all three of these species produce thecae with high walls.

3.6. Evolution of marine and freshwater choanoflagellates

The phylogeny illustrated in Fig. 2 contains 17 freshwater and 30 marine species of choanoflagellate. Sixteen of the freshwater taxa fall into a single, paraphyletic group (1.00biPP, 74%mlBP) within the Clade 2 craspedids of the phylogeny (see also Fig. 3). A single marine species, *Salpingoeca macrocollata*, is robustly nested deep within the freshwater species and appears to be a revertant to the marine environment.

The current phylogeny requires a minimum of three freshwater:marine transitions to explain the distribution of sampled species and the phylogeny is consistent with most freshwater species being the descendants of a single major invasion by the Clade 2 craspedids. A further freshwater incursion has occurred

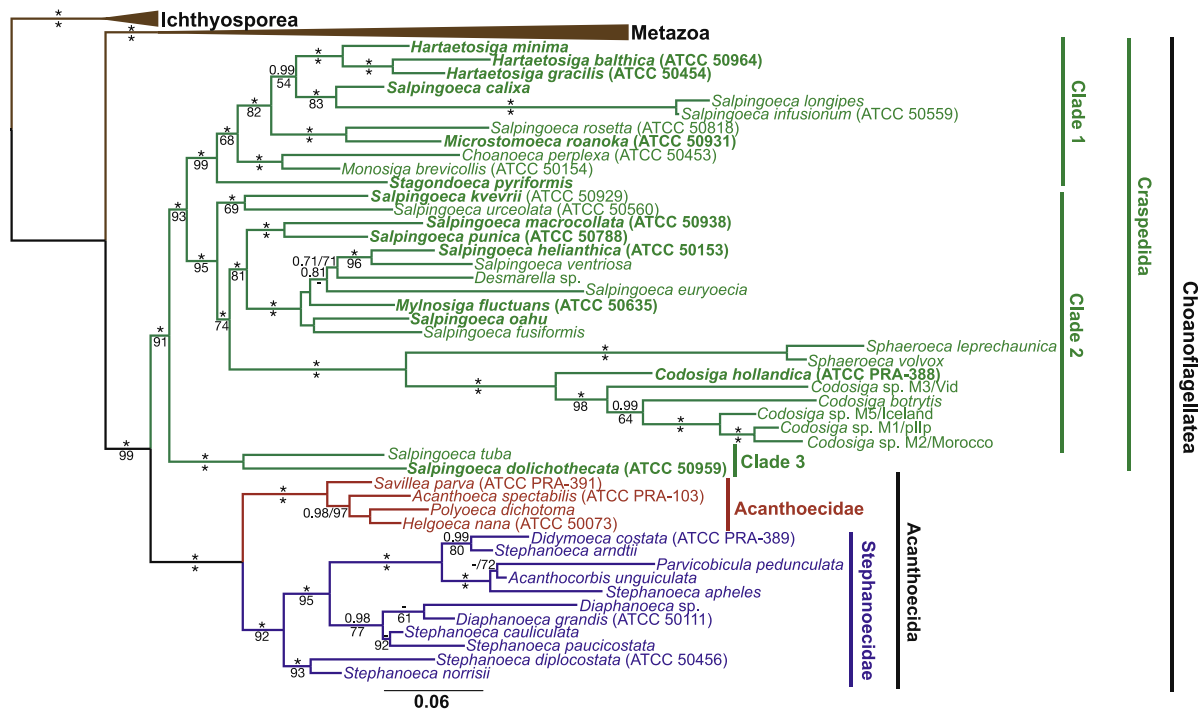


Fig. 2. Maximum likelihood six-gene phylogeny of the choanoflagellates. The phylogeny is based upon 9767 aligned nucleotides positions from partial sequences of the genes SSU, LSU, hsp90, tubA, EFL and EF-1A. Branches are drawn proportional to the number of nucleotide substitutions per site as indicated by the scale bar at the lower left. Values of 1.00 biPP and 100% mlBP support are denoted by an *, biPP and mlBP values are otherwise given above and below branches respectively. Values are omitted from weakly supported branches (i.e. biPP < 0.70, mlBP < 50%). Species described here are written in bold font.

in the loricate species (*Stephanoeca arndtii*) (Nitsche, 2014). However, *Stephanoeca arndtii* is nested deeply within marine species and has closely related marine relatives, indicating that it is a descendant of a relatively recent, minor freshwater transition.

Most observed freshwater diversity is found in a single group of Clade 2, however it is clear that the freshwater environment has been invaded by choanoflagellates on multiple occasions. In addition to the colonization of freshwater highlighted here, del Campo and Ruiz-Trillo (2013) performed a meta-analysis of environmental SSU sequences which recovered ten putative clades containing freshwater choanoflagellates. As with previously published choanoflagellate SSU phylogenies (Carr et al., 2008; Cavalier-Smith and Chao, 2003; Nitsche et al., 2011) the deeper branches within the choanoflagellates were poorly resolved and two of the putative freshwater clades (Clade L and FRESCHO3) were recovered outside the diversity of known choanoflagellate species. Additional Bayesian Inference and ML phylogenetic analyses were performed here using the dataset from the del Campo and Ruiz-Trillo (2013) study incorporated into the six-gene dataset used to generate the phylogeny in Fig. 2. When the complete del Campo and Ruiz-Trillo dataset was used, sequences from their FRESCHO2, FRESCHO4, Napiformis and *Salpingoeca* clades fall within the previously identified main freshwater group in Clade 2 (Figs. S2A and S2D), rather than forming discrete, independent clades. Of the ten freshwater clades present in the del Campo and Ruiz-Trillo study, only the FRESCHO3, *Lagenoeca* and *Napiformis* groupings are recovered as monophyletic. Four putative novel freshwater groups are recovered in the phylogeny; however, only one of the additional groups has a robust position within the tree, with a strongly supported branch (biPP \geq 0.97, mlBP \geq 75%) separating the group from the main group recovered in the six-gene phylogeny. The one strongly supported novel group is made up by a single species, *Salpingoeca* sp. (Vietnam), which was isolated from the Mekong River in Vietnam (Nitsche et al., 2011). This species could represent a third freshwater incursion, however it

was isolated from a reach of the Mekong noted for tidal inflow and saltwater intrusion (Mekong River Commission, 2005); therefore *Salpingoeca* sp. (Vietnam) may be a transient migrant from the sea. This latter possibility is reinforced by the identification of almost identical oceanic sequences (KC488361, 99.9% nucleotide identity, and AB275066, 99.5% nucleotide identity) which appear to be from marine populations of the same species.

In an attempt to limit noise in the phylogenetic signal caused by short sequences, two further datasets were constructed. Length cut-offs for edited alignments of 750 bp and 1000 bp were applied to the dataset. Due to its short length *Salpingoeca* sp. (Vietnam) was excluded from the additional analyses, but both edited datasets (Figs. S2B and C and S2E and F) recover the same additional three other freshwater groupings as the full SSU dataset, albeit with fewer sequences. The groups consist of a singleton, a monophyletic freshwater group and a paraphyletic freshwater cluster and correspond to sequences from the Clade L, FRESCHO1, FRESCHO2, FRESCHO3, FRESCHO4 and *Salpingoeca* clades of del Campo and Ruiz-Trillo (2013). The singleton sequence from clay sand sediment and the paraphyletic freshwater grouping do not have robustly supported positions and therefore neither provides strong evidence of further freshwater incursions. FRESCHO1, at the base of Clade 1, is separated from the main freshwater group by a single strongly supported branch in the 750 bp and 1000 bp datasets, raising the possibility that this group highlights the presence of a third freshwater invasion. Adding the FRESCHO1 sequences alone to the six-gene dataset recovers the same position (data not shown). Although the position of FRESCHO1 is separated from the main freshwater group by one strongly supported branch (e.g. biPP \geq 0.97, mlBP \geq 75%) in the 750 bp, 1000 bp and FRESCHO1-only datasets, the presence of the sequences in the phylogenetic analyses cause a loss of support values in the backbone of the phylogeny. The maximum likelihood bootstrap support for the branches between FRESCHO1 and the Clade 2 freshwater group are reduced by 12–48% compared to Fig. 2 phylogeny. This loss of

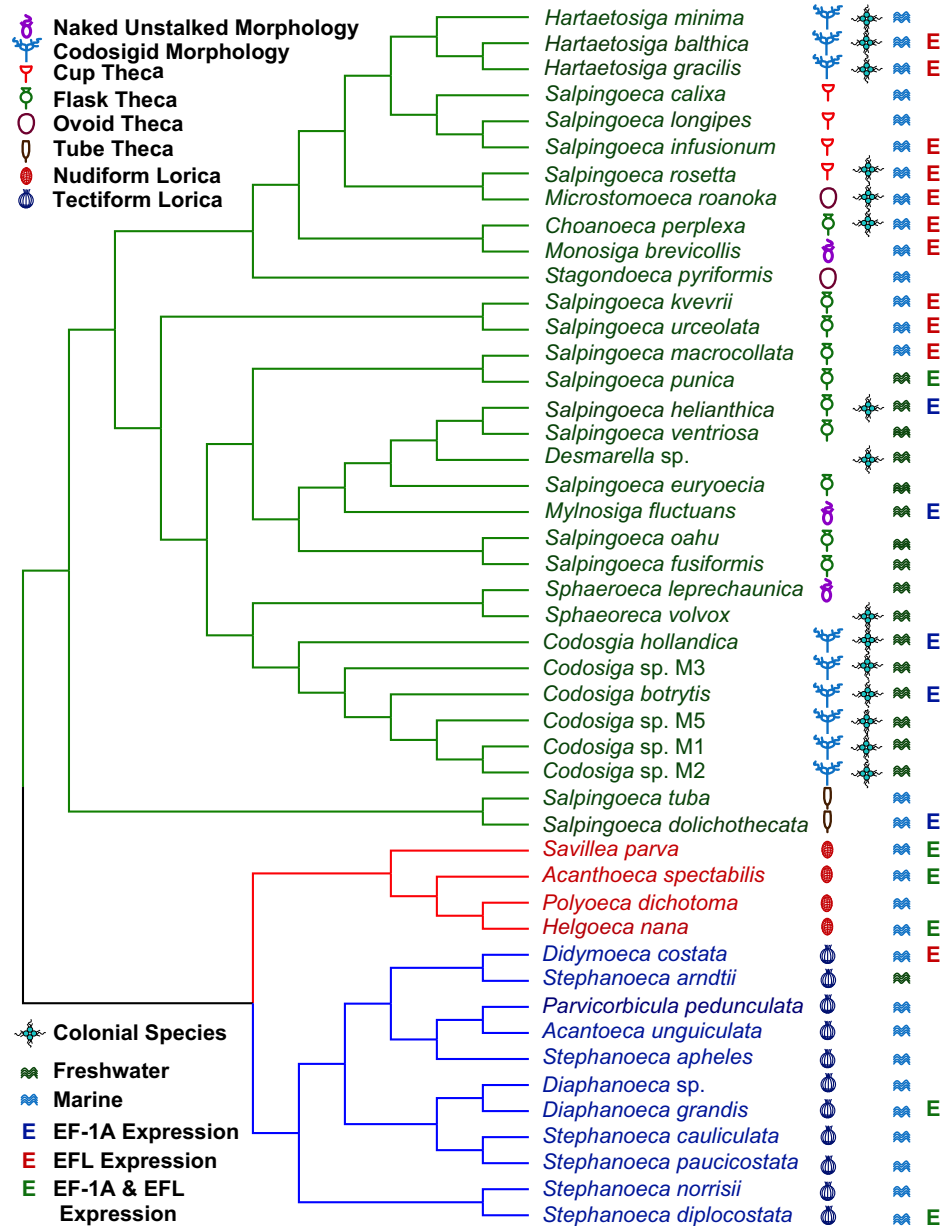


Fig. 3. Morphological, ecological and genomic traits of the choanoflagellates. The representative phylogeny highlights traits for each species present in the main phylogeny. Periplast morphology is shown in the first column and the identity of known colonial species shown in the second column. The third column identifies species which are found in a freshwater or saline environment and the final column shows which species are known to express EF-1A and EFL. Gaps represent missing data.

support points to conflicting phylogenetic signals within the datasets and means the trees containing environmental data should be considered with caution.

3.7. Evolution of EF-1A and EFL within choanoflagellates

Transcriptome data from 19 species of choanoflagellate were screened with hidden Markov model profiles of both EF-1A and EFL. Combining the transcriptome data with publicly available data shows that, of 22 choanoflagellate species, 11 species were only shown to express EFL, five species only presented EF-1A in their transcriptomes and six species expressed both genes (Figs. 3, S3). The loss of either gene can be inferred by its absence from the transcriptome sequences; however this cannot be considered conclusive proof, as the genes may be expressed at low levels or in stages not present in the cultures used to generate the transcrip-

omes. A number of factors point against either of the alternative explanations to gene loss; specifically, in other studied species the two genes are highly expressed, with EF-1A making up 1–2% of total protein content (Kamikawa et al., 2008; Kang et al., 2012; Liu et al., 1996), and stage-specific expression has yet to be reported for either gene. PCR screening of genomic DNA, using primers devised by Baldauf and Doolittle (1997), failed to amplify EF-1A in *Acanthoeca spectabilis*, *Choanoeca perplexa*, *Diaphanoeca grandis*, *Didymoeca costata*, *Hartaetosiga gracilis*, *Helgoeca nana*, *Monosiga brevicollis*, *Salpingoeca infusionum*, *Salpingoeca kvevrii*, *Salpingoeca rosetta*, *Salpingoeca urceolata* and *Stephanoeca diplocostata*. The absence of EF-1A from the whole genome sequences of both *Monosiga brevicollis* and *Salpingoeca rosetta* confirms that gene loss has occurred within choanoflagellate genomes. Given these findings, we conclude that it is likely that the absence of expression is due to the loss of the gene from a species' genome.

The six species shown to express both proteins (*Acanthoeca spectabilis*, *Diaphanoeca grandis*, *Helgoeca nana*, *Savillea parva*, *Salpingoeca punica* and *Stephanoeca diplocostata*) all possess a conserved EFL protein and the protein phylogeny is broadly consistent with the species phylogeny (Fig. S4). The EF-1A genes in these six species are highly divergent with elevated rates of nucleotide substitutions at both synonymous (Ks) and non-synonymous sites (Ka) (Table S6). All 11 sequenced EF-1A genes show values of Ka/Ks less than one in their terminal branches, indicating purifying selection operating on the amino acid sequences. The strength of purifying selection on the EF-1A amino acid sequence, based upon mean Ka/Ks values, is significantly weaker (t -test, $p = 0.002$) in those species that show co-expression, suggesting a loss of functional constraint. In contrast, mean Ka/Ks values for EFL genes show no significant difference when the gene is co-expressed or expressed alone in transcriptomes (t -test, $p = 0.413$). EF-1A in dual-expressing species also shows weaker biases in codon usage. Highly expressed genes, such as EF-1A, tend to show non-random codon usage and a bias toward a set of preferred codons, whilst genes with lower expression levels employ a broader range of codons (Sharp and Li, 1986). The 'effective number of codons' statistic (N_c) is a measure of codon bias (Wright, 1990), with values ranging from 20 (highly biased) to 61 (totally random codon usage). Past studies have shown strong codon bias ($N_c \leq 30$) for EF-1A (Lafay and Sharp, 1999; Sharp and Cowe, 1991) and similar values are observed here for species shown only to express EF-1A (Table S7). Consistent with a weakening of selection on codon usage in co-expressed EF-1A genes, mean N_c values are significantly higher in EF-1A in co-expressing species than when EF-1A is expressed alone (t -test, $p = 0.0004$). For EFL, N_c values are similar to previously published results for EF-1A and do not significantly differ when the gene is co-expressed or EF-1A is absent (t -test, $p = 0.305$).

A maximum likelihood phylogeny of EF-1A within opisthokont species (Fig. S5) is broadly consistent with the vertical inheritance of the gene throughout the opisthokont radiation (all branches $<0.50\text{biPP}$, $<25\text{mlBP}$). The choanoflagellate LCA appears to have possessed EF-1A with four potential, independent losses of the gene within the choanoflagellates (Fig. S3).

An amino acid phylogeny of 133 EFL sequences clustered the choanoflagellate proteins together; the group was recovered as paraphyletic since sequences from three ichthyosporeans (*Creolimax fragrantissima*, *Sphaeroforma arctica* and *Sphaeroforma tapetis*) were nested within it (Fig. S4), albeit without strong phylogenetic support ($\text{biPP} < 0.97$, $\text{mlBP} < 75\%$). The topology of the choanoflagellate EFL sequences is broadly consistent with the six-gene phylogeny shown in Fig. 2, indicating that the gene was present in the choanoflagellate LCA and has been inherited vertically during the choanoflagellate radiation. Mapping the observed presence and absence of EFL onto the choanoflagellate phylogeny shows three putative independent losses within the choanoflagellates, all of which have occurred within Craspedida (Fig. S3). The holozoan EFL sequences are recovered together with strong support (1.00biPP , 77mlBP) and the phylogeny is consistent with the presence of the gene in the LCA of choanoflagellates and ichthyosporeans.

4. Discussion

4.1. Taxonomic revisions of the choanoflagellates

The multigene phylogeny presented here provides sufficient grounds to revise and clarify aspects of the taxonomy of the choanoflagellates. In addition to the considerable phylogenetic distance between *Codosiga* and *Hartaetosiga*, species from the two genera also show numerous morphological and ecological differ-

ences further justifying their separation. All known isolates of the genus *Codosiga* are freshwater species, whereas all known isolates of *Hartaetosiga* are marine. Furthermore, the cell covering of *Codosiga* has two appressed layers and has been shown to be more substantial than that of the single-layered investment of *Hartaetosiga* (Hibberd, 1975; Leadbeater and Morton, 1974; Wylezich et al., 2012). Leadbeater (1977) noted that the single layered cell coat of *Hartaetosiga gracilis* is similar to that observed in motile cells of *Choanoeca perplexa* Ellis, which is also a Clade 1 craspedid (Fig. 2). The cell body extends into the periplast stalk in *Codosiga*, but does not in *Hartaetosiga*. In stalked colonies, the cells of *Codosiga botrytis* do not fully undergo cytokinesis and are connected in some isolates by cytoplasmic bridges (Hibberd, 1975); however such bridges have not been reported in the stalked colonies of *Hartaetosiga* species. The newly described *Hartaetosiga* genus is nested within a grouping of thecate species, indicating that the naked *Hartaetosiga* species have lost their ancestral theca.

We have begun to establish a degree of order within the taxonomy of the craspedids by splitting taxa previously assigned to *Codosiga* into two phylogenetically, morphologically and ecologically coherent genera. Three further new genera, *Microstomoeca*, *Mylnosiga* and *Stagondoeca* have been erected in order to accommodate phylogenetically and morphologically distinct craspedids. Furthermore, the taxonomy of 7 misidentified species present in culture collections has been resolved (Taxonomic Summary). Obvious problems within craspedid taxonomy remain with the paraphyletic *Salpingoeca*, which can be considered a 'dustbin' genus that shows no rationale on a phylogenetic, ecological or morphological level. Unfortunately, DNA sequences are not available for the type species, *Salpingoeca gracilis* James-Clark, which has not been deposited in a culture collection. *Salpingoeca gracilis* is a freshwater, tube thecate species, therefore it would be preferable, when sufficient evidence is available, for future taxonomic revisions to assign non-tube thecate species to other genera.

One other remaining contentious area of choanoflagellate phylogeny is the root of the acanthoecid taxa. The position of the root has profound implications for determining the evolution of the nudiform and tectiform morphologies. A sister-relationship (Fig. 2) between the two groups makes it difficult to determine the morphology and development of the common ancestor of nudiforms and tectiforms; however, the alternative topology (Fig. S1) indicates the tectiform morphology is ancestral and that the nudiform morphology is derived. The rooting of the acanthoecids between the nudiform and tectiform groups (Fig. 2) is not simply due to a phylogenetic signal present in third codon positions, as ribosomal RNA phylogenies can also recover this topology (Carr et al., 2008; Jeuck et al., 2014). Carr et al., (2008) noted that codon third positions were not saturated with mutations in choanoflagellates and, using parsimony criteria, that the monophyly of both the nudiform and tectiform groups was strongly supported over tectiform paraphyly on morphological, developmental and ecological grounds. Furthermore, in a series of experiments with the tectiform *Stephanoeca diplocostata*, Leadbeater (reviewed in Leadbeater, 2015) showed that cells deprived of silica became 'naked' and lacked loricae. When silica was re-introduced to the medium, the new loricae were initially produced in a nudiform-like manner. This phenomenon is difficult to reconcile with the Fig. S1 topology, which indicates that the nudiform morphology is derived and evolved after the nudiform lineage diverged from the ancestor of the nudiforms and *Stephanoeca diplocostata*.

The plastic nature of cell coverings, as well as the ability of species to develop multiple morphologies presents practical problems for choanoflagellate taxonomy. Future work may require a consensus within the choanoflagellate community whether to take a 'lumper' or 'splitter' approach to the taxonomy of the group.

4.2. Evolutionary trends within the choanoflagellates

The phylogeny presented here places 47 choanoflagellate species into a phylogenetic context and therefore provides an unprecedented opportunity to evaluate the evolution of morphological, ecological and genomic traits within the group. Seven major traits within the choanoflagellates are discussed in greater detail below (Fig. 3).

4.2.1. Flask-theca morphology

The flask-theca is perhaps the most structurally complex of the known thecate morphologies. This is because, apart from its precise shape, it possesses a ‘scalloped’ flange on the inner surface of the neck which attaches to the anterior end of the cell (Leadbeater, 1977). On the outer surface the neck is often decorated with a pattern of narrow ridges. It is the only thecal morphology that is present in both clades 1 and 2 of Craspedida within the phylogeny (Fig. 3). One possible explanation for the distribution of the flask-thecates is that the flask was the ancestral thecal morphology of both clades 1 and 2 craspedids. An alternative scenario, of convergent evolution, seems unlikely due to the remarkable similarity of the complex morphology observed in both Clade 1 and Clade 2 flask-thecates (described in Carr et al., 2008).

4.2.2. Tube-theca morphology

Salpingoeca dolichothecata and *Salpingoeca tuba* are the first species with tube thecae to be placed in a multi-gene phylogenetic framework. The two tube thecate species form a strongly supported monophyletic group (1.00biPP, 100%mlBP), consistent with a single origin of the tube morphology. The tube species are recovered within the craspedids (1.00biPP, 91%mlBP) with strong support, as the earliest branching lineage, and form Clade 3 of Craspedida, a previously unidentified clade, within this taxon (Fig. 2). The tube morphology shows similarities to the flask theca, as the theca of *Salpingoeca gracilis* has longitudinal ridges on its outer surface (Thomsen, 1977). It is unknown whether the common ridge morphology is a case of convergent evolution or a result of common ancestry.

4.2.3. Cup-theca morphology

Four species in the phylogeny possess cup-theca morphologies (*Salpingoeca rosetta*, *Salpingoeca calixa*, *Salpingoeca infusionum* and *Salpingoeca longipes*) and all are present in Clade 1 of Craspedida. Furthermore, an additional cup-thecate, *Salpingoeca abyssalis*, has previously been shown to fall within Clade 1 (Nitsche et al., 2011). The cup species fall into two groups, which exhibit different thecal morphologies. *Salpingoeca calixa*, *Salpingoeca infusionum* and *Salpingoeca longipes* form a well-supported clade (1.00biPP, 83% mlBP) and develop thecae with high sided-walls (Figs. 1G, and 2 in Jeuck et al., 2014); in contrast, *Salpingoeca rosetta* forms a separate cup lineage and produces a shallow theca with short-sided walls (Dayel et al., 2011; Dayel and King, 2014).

The phylogeny is congruent with two independent origins of the cup morphology in closely related taxa; however, a single origin for the cup morphology followed by two losses (in the naked *Hartaetosiga* species and the ovoid thecate *Microstomoeca*) cannot be discounted. Morphologically the short-walled, cup theca of *S. rosetta* appears distinct from the high-walled thecae of *Salpingoeca calixa*, *Salpingoeca infusionum* and *Salpingoeca longipes*, which may be viewed as being consistent with an independent evolutionary origin.

4.2.4. Naked craspedids

Species that produce ‘naked’ cells are found in both Clade 1 and Clade 2 of the craspedids. The naked species appear to be polyphyletic (6 unweighted parsimony steps, against 9 unweighted

steps for their paraphyly) and nested within the thecate craspedids. The ‘naked’ appearances of the codosigid and monosigid morphologies are almost certainly derived states. The current phylogeny indicates that the codosigid (multiple cells on a peduncle) morphology has evolved on at least two occasions, with species previously assigned to *Codosiga* being recovered as polyphyletic (Fig. 3). Three species, namely *Monosiga brevicollis*, *Mylnosiga fluctuans* and *Sphaeroeca leprechaunica*, have naked, free-floating mature cells that lack a peduncle for attachment. The three species are distantly related in clades 1 and 2 (Figs. 2 and 3); therefore, their shared morphologies appear to be a case of convergent evolution.

In addition to convergent evolution between naked species, taxa with the capacity to develop multiple cells on a single peduncle may be mistaken as monosigid (a single cell on a peduncle) species when they initially settle onto a surface prior to cell division (Leadbeater and Morton, 1974; Wylezich et al., 2012). The number of cells per peduncle therefore appears to be a plastic morphological trait and unreliable for choanoflagellate taxonomy.

4.2.5. Coloniality

The ability of choanoflagellates to form ephemeral colonies has long been recognized (Fromentel, 1874; Stein, 1878) and a possible evolutionary link between coloniality in choanoflagellates and multicellularity in metazoans has previously been speculated upon (Carr et al., 2008; Dayel et al., 2011; Levin et al., 2014; Richter and King, 2013). Colonies are found in a variety of different morphologies (Carr and Baldauf, 2011; Leadbeater, 1983) and recent work has shown that individual species are capable of developing multiple colonial morphologies (Dayel et al., 2011). This important finding casts further doubt on the reliance of morphological traits in the taxonomy of craspedid choanoflagellates, as colonial forms attributable to *Desmarella*, *Proterospongia* and *Sphaeroeca* have been found in clonal cultures of the same species. For this reason, we have avoided proposing new names for choanoflagellate genera based on the appearance of the colonial forms of the species contained within those genera, and we propose that future researchers subscribe to the same logic.

Coloniality has been observed in 15 of the 32 craspedid species present in the current phylogeny (Fig. 3). Coloniality cannot be excluded in any of the other craspedids, as most have poorly studied life cycles; however it is unlikely that it is a trait of some species, such as *Monosiga brevicollis* or *Mylnosiga fluctuans*, which have been intensively studied. The common structure of intercellular cytoplasmic bridges present in Clade 1 (*Salpingoeca rosetta*, see Dayel et al., 2011) and Clade 2 (*Desmarella moniliformis* and *Codosiga botrytis*, see Hibberd, 1975; Leadbeater and Karpov, 2000) choanoflagellates either suggests that such bridges were present early in craspedid evolution, or that there has been a remarkable level of convergent evolution within the group. Similar cytoplasmic bridges are also present between metazoan cells (Fairclough et al., 2013), suggesting such bridges may have much greater antiquity. To date the only acanthoecid species showing colony formation via connections between protoplasts is *Diaphanoeca sphaerica*.

4.2.6. Freshwater-marine transitions

The available data point to freshwater:marine transitions being rare events in choanoflagellate evolution, as is the case in many protistan groups (Logares et al., 2009). Only two groups of freshwater species are present in our main phylogeny, with a single, major freshwater group in the Clade 2 craspedids accounting for all but one of the freshwater species.

The use of environmental SSU sequences within our multigene framework (Fig. S2) indicates the possibility of greater choanoflagellate diversity than is observed in Fig. 2 tree and provides equivocal evidence for further freshwater invasions. These data should

be treated with a degree of caution, as, in many cases, the environmental freshwater sequences do not have strongly supported phylogenetic relationships with marine choanoflagellates and SSU is known to produce unstable and unresolved phylogenies in the choanoflagellates (Nitsche et al., 2011). One putative group (FRESCHO1) does have strong phylogenetic support, albeit with only a single strongly supported branch. However, the presence of FRESCHO1 sequences cause a major loss of bootstrap support between the group and the major freshwater grouping, while other branches in the tree are largely unaffected. The loss of support is indicative of conflicting phylogenetic signals, such as those that result in long-branch attraction (Baldauf, 2003), in the environmental datasets. Those species represented in the environmental groups will require additional sequenced genes in future in order to provide their robust phylogenetic placement.

4.2.7. Distribution and evolution of EFL and EF-1A

Genomes possessing a single elongation factor, as well as dual-encoding species are present within the 22 choanoflagellate species studied. Mapping the observed presence of the two genes onto the Fig. 2 phylogeny points to both elongation factors being present in the genome of the choanoflagellate LCA, with gene loss occurring within the group. Screening whole-genome sequences has shown that very few studied eukaryotes possess both EFL and EF-1A within their genomes (Atkinson et al., 2014). Specifically, within the choanoflagellates, it can be seen that both *Monosiga brevicollis* and *Salpingoeca rosetta* have EFL within their genomes, but have lost EF-1A.

Choanoflagellate EF-1A proteins reveal that, in contrast to the canonical proteins, the divergent proteins in co-expressing species do not show conservation in nucleotide binding sites or those amino acids required for interacting with EF-1A's guanine exchange factor, EF-1B (see Fig. S6). These factors suggest that the divergent EF-1A proteins are no longer capable of functioning in protein synthesis, with that role most probably taken by EFL. Similar losses of translation function for EF-1A in dual-encoding species have previously been speculated upon in a diverse range of eukaryotes (Kamikawa et al., 2013). Atkinson et al. (2014) proposed a mutational ratchet to explain the loss of either EF-1A or EFL in dual-encoding species and the findings here appear to show this process in action within the choanoflagellates. In the six dual-expressing species EFL appears to have remained conserved and functional, whilst EF-1A has accumulated mutations that are likely to have resulted in a loss of translation function. The level of functional constraint on EFL does not differ when the gene is either co-expressed or expressed alone; however K_a/K_s and N_c values are significantly higher for EF-1A when it is co-expressed than when it is expressed alone.

The deeper branches of the EFL tree presented here are poorly resolved, which has also been observed in previously published phylogenies (Kamikawa et al., 2010a, 2010b; Noble et al., 2007). As a result, it is currently difficult to speculate on the evolutionary origin of the gene within the holozoans. The EFL phylogeny is consistent with the gene being present, along with EF-1A, in the LCA of the choanoflagellates and ichthyosporeans. If this scenario is correct, then both genes were also present in the LCA of choanoflagellates and metazoans, with EFL apparently being lost in a stem-group metazoan. These data highlight a major difference in the cell biology of choanoflagellates and metazoans, as EF-1A is universal in metazoan protein synthesis (Kamikawa et al., 2010b) whilst the majority of choanoflagellates appear to employ EFL.

5. Conclusions

The revised six-gene phylogeny presented here greatly increases the number and diversity of named choanoflagellate spe-

cies placed into a phylogenetic framework. Importantly, tube and ovoid thecate species are added to the phylogeny, with the tube species recovered as monophyletic and the ovoid thecate species appearing as polyphyletic. The phylogeny highlights the presence of a major freshwater radiation in the Clade 2 craspedids, with a second minor incursion in the loricate species; however, analyses of environmental SSU sequences raise the possibility of greater freshwater diversity and additional freshwater incursions.

Finally, transcriptome data provide strong evidence that the choanoflagellate LCA possessed both EF-1A and EFL in its genome. Both genes appear to have subsequently been lost on multiple occasions, with more studied species possessing EFL rather than EF-1A.

6. Taxonomic summary

All species described here are members of the Order Craspedida Cavalier-Smith (see Nitsche et al., 2011 for a general diagnosis of the group).

Codosiga (James-Clark) emend. Carr, Richter and Nitsche

Sedentary cells have a two layered, extracellular investment that extends into a peduncle. The protoplast extends posteriorly into the peduncle. Mature cells do not produce a rigid theca. Sessile mature cells can divide longitudinally to form stalked colonies. Adjacent cells within a colony may be connected by means of cytoplasmic bridges. Mitochondrial cristae flattened. All known species are found in freshwater.

Type species: *Codosiga botrytis* (Ehrenberg) emend. Leadbeater

***Codosiga hollandica* sp. nov. Carr, Richter and Nitsche**

Subjective synonym: ATCC PRA-388

Cell body is 7–11 μm in length and 6–8 μm in width. Collar is 2–9 μm in length. Protoplast almost spherical, with a slight pinched waist below the collar. Peduncle shorter than diameter of protoplast. Fig. 1A1 and A2. Paratypes are held as clonal cultures in the ATCC (ATCC PRA-388).

Type locality: Freshwater pond, Madeira

Etymology: The SSU sequence for this species is essentially identical to a number of uncultured eukaryotic sequence clones found at a water treatment plant in the Netherlands (roughly 30 clones with the identifier BSF-B-15euk (Valster et al., 2010)). The species is named after Holland, a common colloquial name used as a *pars pro toto* to refer to the Netherlands.

Hartaetosiga gen. nov. Carr, Richter and Nitsche

Cell body possesses a distinctive waist behind the attachment of the collar microvilli. Posterior region of cell body enclosed in a delicate single-layered organic investment subtended by a peduncle. The cell body does not extend into the peduncle. Sessile mature cells can divide to form stalked colonies of cells. Mitochondrial cristae flat, saccular or tubular. All known species are marine or brackish.

Etymology: The name is derived from the Latin *Hartaetus* meaning kite, since the appearance and movement of cells on long stalks is reminiscent of kites flying on lines.

Type species: *Hartaetosiga* (*Codosiga*) *gracilis* (Kent) Carr, Richter and Nitsche

***Hartaetosiga* (*Codosiga*) *gracilis* (Kent) Carr, Richter and Nitsche**

Synonyms: *Monosiga gracilis* Kent, 1880, *Codosiga pyriformis* Kent, *Codosiga gracilis* (Kent, 1880) de Saedeleer, 1927

Basionym: *Codosiga gracilis* (Kent, 1880) de Saedeleer, 1927.

Subjective synonym: ATCC 50454

Cell body is 4–8 μm in length and 3–7 μm in width. Collar is 8–20 μm in length. Sedentary mature cells produce a peduncle of 8–40 μm in length. Cell body tapers toward, but does not extend into, the peduncle. Fig. 1C1–C3. Paratypes are held as clonal cultures in the ATCC (ATCC 50454).

***Hartaetosiga (Codosiga) balthica* (Wylezich et Karpov) Carr, Richter and Nitsche**

Basionym: *Codosiga balthica* Wylezich et Karpov 2012

Subjective synonym: ATCC 50964

Sedentary stalked solitary cells: 3–5 µm in length and 2 µm in width. Collar is 3–5 µm in length. Sedentary cells may produce stalked colonies of 2–4 cells. Adult sedentary protoplast present in delicate extracellular investment which produces a peduncle 9–14 µm in length. Protoplast globular to pyriform in shape. Mitochondrial cristae tubular or saccular. Fig. 1B1 and B2. Paratypes are held as clonal cultures (strain IOW94) in the laboratory of the Leibniz Institute for Baltic Sea Research in Rostock-Warnemünde and the ATCC (ATCC 50964).

Note. This species was deposited at ATCC under the name *Monosiga gracilis*. *Monosiga gracilis* was originally described by Kent (1880–82), but later synonymized with *Codosiga gracilis* by de Saedeleer (1927). The DNA sequences of *Hartaetosiga balthica* and *Hartaetosiga gracilis* show a high level of divergence (95% identity between their SSU genes) indicating that ATCC 50964 is not an isolate of *Hartaetosiga gracilis*.

***Hartaetosiga (Codosiga) cymosa* (Kent) Carr, Richter and Nitsche**

Basionym: *Codosiga cymosa* Kent, 1880

Cell body symmetrically ovate. Single cell per individual peduncle in sedentary adult cells. Pedicels arranged on an elaborate corymbiform peduncle which may be composed of over 100 individuals.

***Hartaetosiga (Codosiga) minima* (Wylezich and Karpov) Carr, Richter and Nitsche**

Basionym: *Codosiga minima* (Wylezich and Karpov)

Sedentary stalked solitary cells: 2–3 µm in length and 1 µm in width. Collar is 2–3 µm in length. Sedentary cells may produce stalked colonies of 2–4 cells. Adult sedentary protoplast present in delicate extracellular investment that extends into a peduncle 7 µm in length. Protoplast vase shaped with a prominent neck. Mitochondrial cristae oval. Fig. 1D1–D3.

***Microstomoeca* gen. nov. Carr, Richter and Nitsche**

Cell body enclosed in a robust organic theca subtended by a peduncle. The theca has an ovoid, or droplet, shaped morphology, without a neck, that possesses a small aperture.

Etymology: The name is derived from the Greek *mikros* (= small) and *stoma* (= mouth), since the thecae of mature cells possess a small aperture through which the collar and flagellum protrude.

Type species: *Microstomoeca roanoka* Carr, Richter and Nitsche

***Microstomoeca roanoka* sp. nov. Carr, Richter and Nitsche**

Subjective synonym: ATCC 50931

Cell 4–7 µm in length, collar 2–5 µm in length and flagellum 3–10 µm in length. Cell body enclosed in a robust organic theca subtended by a peduncle. The theca has an ovoid shaped morphology, without a neck, that possesses a small aperture. Ephemeral colonial stage of swimming cells, with outward facing collars and flagella can be produced. Fig. 1E1–E3. Paratypes are held as clonal cultures in the ATCC (ATCC 50931).

Type locality: Saltmarsh, Virginia, USA

Etymology: This species is named after the failed English colony of Roanoke, due both to its ability to form transient colonies and its geographical location. The original settlement of Roanoke was established on what later became the Virginia Colony, which encompassed the current US state of Virginia (the collection site of the environmental sample from which the culture was isolated).

***Mylnosiga* gen. nov. Carr, Richter and Nitsche**

Cell body is 4–8 µm in length and 4–7 µm in width. Collar is 4–11 µm in length. Ovate protoplast present in a delicate extracellular investment, which lacks a pedicel. Mature cells are not sessile, but may drift in water currents. Sedentary cells may produce pos-

terior microvilli for attachment. Long collar and flagellum. Freshwater.

Etymology: The genus is named in honour of AP Mylnikov, who isolated the type culture from a freshwater pond in Yaroslavl, Russia in 1979.

Type species: *Mylnosiga fluctuans* Carr, Richter and Nitsche

***Mylnosiga fluctuans* sp. nov. Carr, Richter and Nitsche**

Subjective synonym: ATCC 50635

Cell 3–6 µm in length, collar 4–7.8 µm in length and flagellum 6–12 µm in length. Ovate protoplast present in a delicate extracellular investment, which does not extend to a pedicel. Freshwater. Fig. 1F1 and F2. Paratypes are held as clonal cultures in the ATCC (ATCC 50635).

Type locality: Freshwater pond in Yaroslavl, Russia

Note. This species was deposited at ATCC in 1979 under the name *Monosiga ovata*. *Monosiga ovata*, as originally described by Kent (1880–82), is a sedentary marine organism which possesses a short peduncle, whereas *Mylnosiga fluctuans* was isolated from a freshwater pond and does not produce a peduncle. Based upon the morphological and ecological differences ATCC 50635 and *Monosiga ovata* appear to be different species.

Etymology: The name is derived from the Latin *fluctuans* (= floating), as the species is freely suspended in the water column.

***Salpingoeca calixa* sp. nov. Carr, Richter and Nitsche**

Ovoid adult sedentary cells possess a typical, high-walled cup-shaped theca, 4–8 µm in length and 2–4 µm in width from which the collar and flagellum of the protoplast emerges. Flagellum longer than the collar. Theca tapers gradually into a peduncle of 19–27 µm in length. Marine. Fig. 1G1.

Type locality: McKenzie Bay, Rangetoto Island, New Zealand

Etymology: The name is derived from the Latin *calix* (=cup), which refers to the morphology of the theca in sedentary cells.

***Salpingoeca dolichothecata* sp. nov. Carr, Richter and Nitsche**

Subjective synonym: ATCC 50959

Highly elongated protoplast of sedentary cells contained in a long, straight sided, tubular theca, 18–20 µm in length and 5 µm in width. Collar and flagellum extend above the theca. Base of theca tapers into a short peduncle, approximately 5 µm in length. Marine. Fig. 1H1 and H2. Paratypes are held as clonal cultures in the ATCC (ATCC 50959).

Type locality: Great Marsh, Delaware, USA

Note. This species was deposited at ATCC under the name *Salpingoeca gracilis*. *Salpingoeca gracilis*, as originally described by James-Clark (1867), is a freshwater organism which possesses a long peduncle, whereas *Salpingoeca dolichothecata* is a marine organism and produces a short peduncle. Based upon the morphological and ecological differences ATCC 50959 and *Salpingoeca gracilis* appear to be different species.

Etymology: The name is derived from the Greek *dolicho* (= long), which refers to the extended theca of sedentary cells.

***Salpingoeca helianthica* sp. nov. Carr, Richter and Nitsche**

Subjective synonym: ATCC 50153

Adult, sedentary cells present in flask-theca, with short, broad neck, 8–11 µm in length and 5–7 µm in width. Theca extends into a short peduncle. Ephemeral colonial stage in life cycle of swimming cells with outward facing collars and flagella. Freshwater. Fig. 1I1 and I2. Paratypes are held as clonal cultures in the ATCC (ATCC 50153).

Type locality: Aquarium water, Rockville, Maryland, USA

Note. This species was deposited at ATCC under the name *Salpingoeca napiformis*. *Salpingoeca napiformis*, as originally described by Kent (1880–82), is a marine organism, whereas *Salpingoeca helianthica* was isolated from a freshwater aquarium. Based upon the ecological difference ATCC 50153 and *Salpingoeca helianthica* appear to be different species.

Etymology: The name is from the Latin *helianthus*, for sunflower, since the colonial life stage resembles a sunflower, with a dark circular center surrounded by radially symmetrical cell bodies forming the colony.

***Salpingoeca kvevrii* sp. nov. Carr, Richter and Nitsche**

Subjective synonym: ATCC 50929

Cell 4–6 μm in length, collar 2–4 μm in length and flagellum 1–10 μm in length. Adult sedentary cells possess a flask shaped theca with short neck. The top of the theca is broad and tapers toward a base, which lacks a stalk. Marine. Fig. 1J1 and J2. Paratypes are held as clonal cultures in the ATCC (ATCC 50929).

Type locality: Saltmarsh, Hog Island, Virginia, USA

Note. This species was deposited at ATCC in 1999 under the name *Salpingoeca pyxidium*. *Salpingoeca pyxidium*, as originally described by Kent (1880–82), is a freshwater species, whereas *Salpingoeca kvevrii* was isolated from a saltmarsh. Based upon the ecological difference ATCC 50929 and *Salpingoeca pyxidium* appear to be different species.

Etymology: The name is taken due to the similarity in shape between the theca of sedentary cells and kvevri wine jars.

***Salpingoeca macrocollata* sp. nov. Carr, Richter and Nitsche**

Subjective synonym: ATCC 50938

Cell body is 4–7 μm in length and 3–7 μm in width. Collar is 3–9 μm in length. Globular, sedentary, adult cells present in flask-morphology theca with long straight neck (Fig. 1K3). Theca neck height greater than diameter of main body. Marine. Fig. 1K1 and K2. Paratypes are held as clonal cultures in the ATCC (ATCC 50938).

Type locality: Saltmarsh, Hog Island, Virginia, USA

Note. This species was deposited at ATCC in 2001 under the name *Salpingoeca minuta*. *Salpingoeca minuta*, as originally described by Kent (1880–82), is a freshwater organism with a short, broad neck within its theca. In contrast, *Salpingoeca macrocollata* was isolated from a saltmarsh and possesses a long, narrow neck. Based upon the ecological and morphological differences ATCC 50938 and *Salpingoeca macrocollata* appear to be different species.

Etymology: The name is derived from the Greek *macro-* and Latin *-colla* which refers to the long neck of the theca in sedentary cells.

***Salpingoeca oahu* sp. nov. Carr, Richter and Nitsche**

Adult sedentary cells possess a typical flask-theca, 11.5–14.5 μm in length and 4.5–6 μm in width from which the collar and flagellum of the protoplast emerges. Flagellum considerably longer than the collar. Theca opening at the top of a short and broad neck. Theca tapers gradually into a long peduncle of 19–27 μm in length. Freshwater. Fig. 1M1.

Type locality: Freshwater pond, O'ahu, Hawaii

Etymology: The species is named after the island of O'ahu, where the species was first identified.

***Salpingoeca punica* sp. nov. Carr, Richter and Nitsche**

Subjective synonym: ATCC 50788

Cell 4–6 μm in length, collar 3–4 μm in length and flagellum 1–1.5 μm in length. Adult sedentary cells possess a globular flask-theca, from which the collar and flagellum of the protoplast emerges. Theca opening at the top of a very short, broad neck. Freshwater. Fig. 1L1 and L2. Paratypes are held as clonal cultures in the ATCC (ATCC 50788).

Note. This species was deposited at ATCC under the name *Salpingoeca amphoridium*. *Salpingoeca amphoridium*, as originally described by James-Clark (1867), possessed a theca with a long, narrow neck. In contrast, the theca of *Salpingoeca punica* produces a short, broad neck. Based upon the morphological difference ATCC 50788 and *Salpingoeca punica* appear to be different species.

Etymology: The theca morphology resembles the shape of a pomegranate and thus the name is derived from the Latin *Punica*, which is the genus name of the pomegranate.

***Stagondoea* gen. nov. Carr, Richter and Nitsche**

Type species: *Stagondoea pyriformis* Carr, Richter and Nitsche

Small, uninucleate protists with a single, centrally positioned, anterior flagellum, which is surrounded by collar of long, actin-supported microvilli. Phagotrophic. Cell body enclosed in a robust organic theca from which a pedicel extrudes. The theca has an ovoid, or droplet, shaped morphology without a neck.

Etymology: The name is derived from the Greek Stagondion (= droplet), since the thecae of mature cells develop a droplet-shaped morphology.

***Stagondoea pyriformis* sp. nov. Carr, Richter and Nitsche**

Ovoid cell body is 3–5 μm in length and 3–4.5 μm in width. Collar is 8–12 μm in length and surrounds a flagellum of \sim 9 μm in length. Sedentary cells produce a pyriform theca 9–11 μm in length and 5–6.5 μm in width which tapers into a short peduncle. Cell body tapers toward, but does not extend into, the pedicel.

Etymology: The name is taken from the Latin *pyriformis*, which refers to the pear-like morphology of the theca in sedentary cells.

Author Contributions

Conceived and designed the experiments: MC DJR BSCL FN. Performed the experiments: DJR FN AJ PF. Analysed the data: MC DJR TJS FN. Contributed reagents/materials: DJR FN AJ PF. Wrote the paper: MC DJR PF TJS AJ BSCL FN.

Acknowledgments

The authors thank Sergey Karpov, Paul Sharp, Cédric Berney, and Gemma Atkinson for helpful discussion on the manuscript, Alberto Stolfi and Michael Eisen for suggestions of new choanoflagellate taxonomic names, and the choanoflagellate research community for feedback throughout the course of this work. We also thank two anonymous reviewers for their comments on the manuscript. We acknowledge Kayley Hake and Heather Middleton for assistance in sending choanoflagellate cultures between laboratories. DJR was supported by a National Defense Science and Engineering Graduate fellowship from the United States Department of Defense, a National Science Foundation Central Europe Summer Research Institute Fellowship, a Chang-Lin Tien Fellowship in Environmental Sciences and Biodiversity and a postdoctoral fellowship from the Conseil Régional de Bretagne, and the French Government “Investissements d’Avenir” program OCEANOMICS (ANR-11-BTBR-0008). PF was supported by NSF/EDEN IOS Grant Number 0955517.

Appendix A. Supplementary material

Supplementary data associated with this article can be found, in the online version, at <http://dx.doi.org/10.1016/j.ympev.2016.10.011>.

References

- Abascal, F., Zardoya, R., Posada, D., 2005. ProtTest: selection of best-fit models of protein evolution. *Bioinform* 21, 2104–2105.
- Adl, S.M., Simpson, A.G.B., Lane, C.E., Lukeš, J., Bass, D., et al., 2012. The revised classification of eukaryotes. *J. Eukaryot. Microbiol.* 59, 429–493.
- Arndt, H., Dietrich, D., Auer, B., Cleven, E.-J., Gräfenhan, T., Weitere, M., Mylnikov, A. P., 2000. Functional diversity of heterotrophic flagellates in aquatic ecosystems. In: Leadbeater, B.S.C., Green, J. (Eds.), *Flagellates: Unity, Diversity and Evolution*. Taylor & Francis, London, pp. 240–268.
- Atkinson, G.C., Kuzmenko, A., Chicherin, I., Soosaar, A., Tenson, T., Carr, M., Kamenski, P., Haurlyuk, V., 2014. An evolutionary ratchet leading to loss of elongation factors in eukaryotes. *BMC Evol. Biol.* 14, 35.
- Baldauf, S.L., 2003. Phylogeny for the faint of heart: a tutorial. *Trends Genet.* 19, 345–351.
- Baldauf, S.L., Doolittle, W.F., 1997. Origin and evolution of the slime moulds (Mycetozoa). *Proc. Natl. Acad. Sci. U.S.A.* 94, 12007–12012.
- Bradley, R.K., Roberts, A., Smoot, M., Juvekar, S., Do, J., Dewey, C., Holmes, I., Pachter, L., 2009. Fast statistical alignment. *PLoS Comput. Biol.* 5 (e1), 000392.

- Carr, M., Baldauf, S.L., 2011. The protistan origins of animals and fungi. In: Pöggeler, S., Wöstemeyer, J. (Eds.), *The Mycota*, vol. XIV. Springer Press, pp. 3–23.
- Carr, M., Leadbeater, B.S.C., Hassan, R., Nelson, M., Baldauf, S.L., 2008. Molecular phylogeny of choanoflagellates, the sister group to Metazoa. *Proc. Natl. Acad. Sci. U.S.A.* 105, 16641–16646.
- Cavalier-Smith, T., Chao, E.E.Y., 2003. Phylogeny of Choanozoa, Apusozoa, and other protozoa and early eukaryotic megaevolution. *J. Mol. Evol.* 56, 540–563.
- Darriba, D., Taboada, G.L., Doallo, R., Posada, D., 2012. JModelTest2: more models, new heuristics and parallel computing. *Nat. Methods* 9, 772.
- Dayel, M.J., Alegado, R.A., Fairclough, S.R., Levin, T.C., Nichols, S.A., McDonald, K., King, N., 2011. Cell differentiation and morphogenesis in the colony-forming choanoflagellate *Salpingoeca rosetta*. *Dev. Biol.* 357, 73–82.
- Dayel, M.J., King, N., 2014. Prey capture and phagocytosis in the choanoflagellate *Salpingoeca rosetta*. *PLoS ONE* 9, e95577.
- de Saedeleer, H., 1927. Notes de protistologie: I. Craspédomonadines; matériel systématique. *Ann. Soc. Roy. Zool. Belg* 58, 242–288.
- del Campo, J., Ruiz-Trillo, I., 2013. Environmental survey meta-analysis reveals hidden diversity among unicellular opisthokonts. *Mol. Biol. Evol.* 30, 802–805.
- Ewing, B., Green, P., 1998. Base-calling of automated sequencer traces using phred. II. Error probabilities. *Genome Res.* 8, 186–194.
- Ewing, B., Hillier, L., Wendl, M.C., Green, P., 1998. Base-calling of automated sequencer traces using phred. I. Accuracy assessment. *Genome Res.* 8, 175–185.
- Fairclough, S.R., Chen, Z., Kramer, E., Zeng, Q., Young, S., et al., 2013. Premetazoan genome evolution and the regulation of cell differentiation in the choanoflagellate *Salpingoeca rosetta*. *Genome Biol.* 14, R15.
- Fromentel, E. de, 1874. *Études Sur Les Microzoaires Ou Proprement Dits*. G. Masson, Paris.
- Gaucher, E.A., Miyamoto, M.M., Benner, S.A., 2001. Function-structure analysis of proteins using covariances-based evolutionary approaches: elongation factors. *Proc. Natl. Acad. Sci. U.S.A.* 98, 547–552.
- Gile, G.H., Patron, N.J., Keeling, P.J., 2006. EFL GTPase in cryptomonads and the distribution of EFL and EF-1 α in chromalveolates. *Protist* 157, 435–444.
- Henk, D.A., Fisher, M.C., 2012. The gut fungus *Basidiobolus ranarum* has a large genome and different copy numbers of putatively redundant elongation factor genes. *PLoS ONE* 7, e31268.
- Hibberd, D.J., 1975. Observations on the ultrastructure of the choanoflagellate *Codosiga botrytis* (Ehr.) Saville-Kent with special reference to the flagellar apparatus. *J. Cell Sci.* 17, 191–219.
- James-Clark, H., 1867. On the Spongiae Ciliatae as Infusoria Flagellata: or observations on the structure, animality and relationship of *Leucosolenia botryoides* Bowerbank. *Mem. Boston Soc. Nat. History* 1, 305–340.
- Jeuck, A., Arndt, H., Nitsche, F., 2014. Extended phylogeny of the Craspedida. *Eur. J. Protistol.* 50, 430–443.
- Kamikawa, R., Brown, M.W., Nishimura, Y., Sako, Y., Heiss, A.A., Yubuki, N., Gawryluk, R., Simpson, A.G.B., Roger, A.J., Hashimoto, T., Inagaki, Y., 2013. Parallel re-modelling of EF-1 α function: divergent EF-1 α genes co-occur with EFL genes in diverse distantly related eukaryotes. *BMC Evol. Biol.* 13, 131.
- Kamikawa, R., Inagaki, Y., Sako, Y., 2008. Direct phylogenetic evidence for lateral transfer of elongation factor-like gene. *Proc. Natl. Acad. Sci. U.S.A.* 105, 6965–6969.
- Kamikawa, R., Yabuki, A., Nakayama, T., Ishida, K.-I., Hashimoto, T., Inagaki, Y., 2010a. Cercozoa comprises both EF-1-containing and EFL-containing members. *Eur. J. Protistol.* 47, 24–28.
- Kamikawa, R., Sakaguchi, M., Matsumoto, T., Hashimoto, T., Inagaki, Y., 2010b. Rooting for the root of elongation factor-like protein phylogeny. *Mol. Phylo. Evol.* 56, 1082–1088.
- Kang, L.-K., Tsui, F.-H., Chang, J., 2012. Quantification of diatom gene expression in the sea by selecting uniformly transcribed mRNA as the basis for normalization. *Appl. Environ. Microbiol.* 78, 6051–6058.
- Katoh, K., Misawa, K., Kuma, K.-I., Miyama, T., 2002. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res.* 30, 3059–3066.
- Keeling, P.J., Inagaki, Y., 2004. A class of eukaryotic GTPase with a punctate distribution suggesting multiple functional replacements of translation elongation factor 1 α . *Proc. Natl. Acad. Sci. U.S.A.* 101, 15380–15385.
- Kent, W.S., 1880–82. *A Manual of the Infusoria*, vols. 1–3. D. Bogue, London.
- Lafay, B., Sharp, P.M., 1999. Synonymous codon usage variation among *Giardia lamblia* genes and isolates. *Mol. Biol. Evol.* 16, 1484–1495.
- Le, S.Q., Gascuel, O., 2008. An improved general amino acid replacement matrix. *Mol. Biol. Evol.* 25, 1307–1320.
- Leadbeater, B.S.C., 1977. Observations on the life-history and ultrastructure of the marine choanoflagellate *Choanoeca perplexa* Ellis. *J. Mar. Biol. Assoc. UK* 57, 285–301.
- Leadbeater, B.S.C., 1983. Life history and ultrastructure of a new marine species of *Proterospongia* (Choanoflagellida). *J. Mar. Biol. Assoc. UK* 63, 135–160.
- Leadbeater, B.S.C., 2015. *The Choanoflagellates: Evolution*. Cambridge University Press, Biology and Ecology.
- Leadbeater, B.S.C., Karpov, S.A., 2000. Cyst formation in a freshwater strain of the choanoflagellate *Desmarella moniliformis* Kent. *J. Eukaryot. Microbiol.* 47, 433–439.
- 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.
- Leadbeater, B.S.C., Morton, C., 1974. A microscopical study of a marine species of *Codosiga* James-Clark (Choanoflagellata) with special reference to the ingestion of bacteria. *Biol. J. Linn. Soc.* 6, 337–347.
- Levin, T.C., Greaney, A.J., Wetzel, L., King, N., 2014. The *rosetteless* gene controls development in the choanoflagellate *S. rosetta*. *eLife* 3, e04070.
- Liu, G., Tang, J., Edmonds, B.T., Murray, J., Levin, S., Condeelis, J., 1996. F-actin sequesters elongation factor 1a from interaction with aminoacyl-tRNA in a pH-dependent reaction. *J. Cell Biol.* 135, 953–963.
- Logares, R., Bråte, J., Bertilsson, S., Clasen, J.L., Shalchian-Tabrizi, K., Rengefors, K., 2009. Infrequent marine-freshwater transitions in the microbial world. *Trends Microbiol.* 17, 414–422.
- Marshall, W.L., Berbee, M.L., 2010. Population-level analyses indirectly reveal cryptic sex and life history traits of *Pseudoperkinsus tapetis* (Ichthyosporidia, Opisthokonta): a unicellular relative of the animals. *Mol. Biol. Evol.* 27, 2014–2026.
- Medina, M., Collins, A.G., Taylor, J.W., Valentine, J.W., Lipps, J.H., Amaral-Zettler, L., Sogin, M.L., 2003. Phylogeny of Opisthokonta and the evolution of multicellularity and complexity in Fungi and Metazoa. *Int. J. Astro* 2, 203–211.
- Medlin, L., Elwood, H.J., Stickle, S., Sogin, M.L., 1988. The characterization of enzymatically amplified eukaryotic 16S-like rRNA-coding regions. *Gene* 71, 491–499.
- Mekong River Commission, 2005. *Overview of the Hydrology of the Mekong Basin*. Red Plough International Co. Ltd., Mekong River Commission, Vientiane.
- Michalak, S., 2012. RaxmlGUI: a graphical front-end for RAXML. *Org. Divers. Evol.* 12, 335–337.
- Miller, M.A., Pfeiffer, W., Schwartz, T., 2010. Creating the CIPRES Science Gateway for inference of large phylogenetic trees. In: *Proceedings of the Gateway Computing Environments Workshop (GCE)*, 14 Nov. 2010, New Orleans, LA, pp. 1–8.
- Nitsche, F., 2014. *Stephanoeca arndtii* 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., 2008. A new choanoflagellate species from Taiwan: morphological and molecular biological studies of *Diplothecca elongata* nov. spec. and *D. costata*. *Eur. J. Protistol.* 44, 220–226.
- Nitsche, F., Carr, M., Arndt, H., Leadbeater, B.S.C., 2011. Higher level taxonomy and molecular phylogenies of the Choanoflagellata. *J. Eukaryot. Microbiol.* 58, 452–462.
- Noble, G.P., Rogers, M.B., Keeling, P.J., 2007. Complex distribution of EFL and EF-1a proteins in the green algal lineage. *BMC Evol. Biol.* 7, 82.
- Paps, J., Medina-Chacón, L.A., Marshall, W., Suga, H., Ruiz-Trillo, I., 2013. Molecular phylogeny of unikonts: new insights into the position of apusomonads and ancyromonads and the internal relationships of opisthokonts. *Protist* 164, 2–12.
- Richter, D.J., 2013. *The gene content of diverse choanoflagellates illuminates animal origins* PhD Thesis. University of California, Berkeley. Permalink: <http://escholarship.org/uc/item/7xc2p94p>.
- Richter, D.J., King, N., 2013. The genomic and cellular foundations of animal origins. *Annu. Rev. Genet.* 47, 509–537.
- Ronquist, F., Huelsenbeck, J.P., 2003. MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinform. B* 19, 1572–1574.
- Ruiz-Trillo, I., Lane, C.E., Archibald, J.M., Roger, A.J., 2006. Insights into the evolutionary origin and genome architecture of the unicellular opisthokonts *Capsaspora owczarzakii* and *Sphaeroforma arctica*. *J. Eukaryot. Microbiol.* 53, 379–384.
- Ruiz-Trillo, I., Roger, A.J., Burger, G., Gray, M.W., Lang, B.F., 2008. A phylogenomic investigation into the origin of Metazoa. *Mol. Biol. Evol.* 25, 664–672.
- Sharp, P.M., Cowe, E., 1991. Synonymous codon usage in *Saccharomyces cerevisiae*. *Yeast* 7, 657–678.
- Sharp, P.M., Li, W.-H., 1986. Codon usage in regulatory genes in *Escherichia coli* does not reflect selection for 'rare' codons. *Nucleic Acids Res.* 14, 7737–7749.
- Stamatakis, A., 2006. RAXML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* 22, 2688–2690.
- Steenkamp, E.T., Wright, J., Baldauf, S.L., 2006. The protistan origins of animals and fungi. *Mol. Biol. Evol.* 23, 93–106.
- Stein, F., 1878. *Der Organismus Der Infusionsthiere III. Der Organismus Der Flagellaten Oder Geisselinfusorien* Leipzig. W. Englemann.
- Stoupin, D., Kiss, A.K., Arndt, H., Shatilovich, A.V., Gilichinsky, D.A., Nitsche, F., 2012. Cryptic diversity within the choanoflagellate morphospecies complex *Codosiga botrytis* – Phylogeny and morphology of ancient and modern isolates. *Eur. J. Protistol.* 47, 263–273.
- Talavera, G., Castresana, J., 2007. Improvement of phylogenies after removing divergent and ambiguously aligned blocks from protein sequence alignments. *System. Biol.* 56, 564–577.
- Thomsen, H.A., 1977. External morphology of the choanoflagellate *Salpingoeca gracilis* James-Clark. *J. Mar. Biol. Assoc. UK* 57, 629–634.
- Valster, R.M., Wullings, B.A., van der Kooij, D., 2010. Detection of protozoan hosts for *Legionella pneumophila* in engineered water systems by using a biofilm batch test. *Appl. Environ. Microbiol.* 76, 7144–7153.
- Wright, F., 1990. The 'effective number of codons' used in a gene. *Gene* 87, 23–29.
- Wylezich, C., Karpov, S.A., Mylnikov, A.P., Anderson, R., Jürgens, K., 2012. Ecologically relevant choanoflagellates collected from hypoxic water masses of the Baltic Sea have untypical mitochondrial cristae. *BMC Microbiol.* 12, 271.