University of HUDDERSFIELD

Marine:freshwater transitions in choanoflagellates

Author:

Bernardo de Haro Reyes

A thesis submitted to the University of Huddersfield in partial fulfilment of the requirements for the degree of Doctor in Philosophy.

The University of Huddersfield

School of Applied Sciences

Supervisor:

Dr. Martin Carr

Co-supervisors:

Prof. Paul Humphreys

Dr. Frank Nitsche

December 18, 2020

Copyright© statement

Declaration of authorship

I, Bernardo de Haro Reyes, declare that this thesis titled, “Marine:freshwater transitions in choanoflagellates” and the work presented in it are my own. I confirm that:

·This work was done wholly or mainly while in candidature for a research degree at this University.

·Where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated.

·Where I have consulted the published work of others, this is always clearly attributed.

·Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work.

·I have acknowledged all main sources of help.

·Where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself.

University of Huddersfield

Abstract

Department of Biological and Geographical Sciences

School of Applied Sciences

Doctor in Philosophy

Marine:freshwater transitions in choanoflagellates

by Bernardo de Haro Reyes

Choanoflagellates are ubiquitous and one of the most important groups of heterotrophic pico- and nanoplankton. They are the closest relatives to metazoans and are considered a very valuable group for the study of the early evolution, in animals, of some key life traits as important as sexual reproduction or multicellularity. The phylogeny of freshwater choanoflagellates remains poorly understood, with only one major group containing most of the diversity of freshwater cultured species. Conversely, previous studies have estimated a high hidden biodiversity amongst freshwater choanoflagellates that have led to the emergence of clades composed solely of environmental sequences. Such disparities between the known and estimated diversity of freshwater choanoflagellates pose a major barrier towards the understanding of the evolutionary history of marine:freshwater transitions in choanoflagellates.

By a combination of exhaustive phylogenetic analysis, ancestral sequence reconstruction, and phylogenetic placement of environmental sequences, the present study has sought to investigate these discrepancies, in order to unravel the evolutionary history of marine:freshwater transitions in choanoflagellates. Furthermore, the present study has focused on the assessment of the tolerance to salinity changes in several species of choanoflagellates, in order to identify candidates for future studies focused on the molecular mechanisms of adaptation to varying salinities. Additionally the present study has involved extensive efforts towards the isolation of new species of choanoflagellates species from rare freshwater habitats, more specifically acidic and hyperalkaline environments.

As a result of the analysis carried out, the present study has revealed that three independent colonisations of freshwater environments have occurred throughout the evolutionary history of Clade 2 choanoflagellates. Similarly, the results have identified three major hotspots of hidden diversity within choanoflagellates. Such hotspots are located at basal positions of Acanthoecida, Craspedida, and within Clade 2 salpingoecids. The estimated hidden diversity within Clade 2 salpingoecids is of marine origin and indicates that, within this group, freshwater colonisations events might be much more frequent than previously thought. The estimated hidden diversity of the other two hotspots is of both marine and freshwater origin indicating that multiple lineages of both marine and freshwater choanoflagellates remain unknown and that marine:freshwater transitions have occurred independently several times throughout the evolutionary history of choanoflagellates. The results from the autoecology of *Microstomoeca roanoka* revealed that this species is a prime candidate for the study of the molecular mechanisms of adaptation to varying salinities in choanoflagellates. Lastly, the present study contains the description of four morphospecies of choanoflagellates, two from hyperalkaline and two from acidic environments.

Acknowledgements

First and foremost I would like to thank my supervisor Dr. Martin Carr, for all the support, the trust and, more importantly, for all the knowledge, scientific or otherwise, that has shared with me throughout these past years. You have been the best supervisor I could have ever asked for, supportive and, at the same time, always knowing how much rope you should give me. For that you have my most sincere gratitude.

My gratitude, also, to all the past and present members of the Carr lab that helped me on the field trips to fish for choanoflagellates.

I am also grateful to my co-supervisor Prof. Paul Humphreys, without whom I would not have had access to the samples from Hanson Aggregates, nor the equipment for the isolation of choanoflagellates.

I would also like to thank my co-supervisor form the University of Cologne, Dr. Frank Nitsche and his student Sabine Schiwitza. For all your help and friendship and for taking care of my cultures once I had to leave Huddersfield, I thank you.

Special thanks to everyone in the office and to all those that, not sharing the office, shared their time with us. These past years wouldn’t have been the same without all of you around. May the friendship that we share last for many years.

I would, of course, also like to thank my family for their unconditional support throughout my whole life because without them none of this would have been possible. One can only hope to be worthy of so much love.

Finally, a special mention to the person that has been by my side all this years, María Luisa Martín Cerezo. I will always be grateful for all your love, your patience, your support and for everything you have shared with me for the last thirteen years. This work is dedicated to you and to little Luna.

Contents.

List of figures.

List of tables.

Chapter I: Introduction.

Choanoflagellates: an overview.

Choanoflagellates are a group of heterotrophic, aquatic, single-celled organisms characterised by a distinctive morphology: a spherical to ovoid cell with a single flagellum surrounded by a collar-like structure of microvilli (Karpov and Leadbeater, 1998; Leadbeater, 2015). They have been known since the early 19th century, but it was not until 1867 that they were, unequivocally, described by James-Clark, who also noted their morphological resemblance to the choanocytes of sponges (James-Clark, 1867). Choanoflagellates are widely distributed and can be found in most aquatic habitats, such as marine waters, freshwaters, hypersaline lakes (Schiwitza *et al.*, 2018; Schiwitza *et al.*, 2019) alkaline lakes (Couradeau *et al.*, 2011; Antony *et al.*, 2013; Hake, 2019) and karstic lakes (Miracle *et al.*, 1992).

Choanoflagellates are considered one of the major groups of heterotrophic nanoplankton and of great importance for the recycling of organic matter. Through filtering the water body, they prey on bacteria and other picoplanktonic autotrophs becoming one of the major drivers of bacterial population dynamics (Fenchel, 1982 a-c; Azam *et al.*, 1983). Furthermore, some species have been reported as able to subsist solely feeding on dissolved organic carbon (Gold *et al.*, 1970; Marchant and Scott, 1993). Their feeding habits confer them a crucial role in the microbial loop of energy recycling and make them one of the primary means by which much of the organic carbon becomes available to higher trophic levels (Fenchel, 1982a; Azam *et al.*, 1983).

Choanoflagellates and the origin of animals.

Due to the evidence supporting a close relationship with metazoans (Carr *et al.*, 2008b), the study of choanoflagellates has received an unprecedented attention in recent years. Numerous studies have used them for the study of the early evolution, in animals, of some key life traits as important as sexual reproduction, multicellularity or animal-bacteria interactions (King *et al.*, 2003; King, 2004; Snell *et al.*, 2006; Abedin and King, 2008; Alié and Manuel, 2010; Alegado *et al.*, 2012; Levin and King, 2013;

Hoffmeyer and Burkhardt, 2016; Woznica *et al.*, 2016; Richter *et al.*, 2018; Laundon *et al.*, 2019; Hake, 2019; Booth and King, 2020).

James-Clark (1867) described the morphology and named the first two genera of choanoflagellates, *Codosiga* and *Salpingoeca.* Later Kent (1880-1882) also reported the capability of *Proterospongia (Protospongia) haeckeli* to produce colonies. Colony formation in choanoflagellates involves several cell adhesion and cell-to-cell signalling proteins and pathways that were previously believed to be present only in metazoans (Figure 1.1) (King *et al.*, 2003; King, 2004; Snell *et al.*, 2006; Abedin and King, 2008; Hoffmeyer and Burkhardt, 2016). The analysis of the transcriptomes of 19 species of choanoflagellates (Richter *et al.*, 2018) corroborated that ~372 gene families previously considered animal-specific are also present in choanoflagellates. Moreover, colony formation in *S. rosetta* and *Salpingoeca monosierra* is related to phosphotyrosine signalling, which is critical in many processes in animals, such as development, homeostasis, cell differentiation, proliferation and ageing (Hake, 2019). Colony formation in the model choanoflagellate *Salpingoeca rosetta* is triggered and regulated by the presence of sphingolipids produced by *Algoriphagus machipongonensis* (Alegado *et al.*, 2012; Woznica *et al.*, 2016).

Similarly, the life cycle of choanoflagellates was considered to be limited to asexual reproduction through cell division. For this reason, it remained poorly studied for a long time. However, the presence of a bimodal size distribution, retrotransposons in the genome and genes involved in meiosis suggested the hypothesis that choanoflagellates might be able to sexually reproduce (Thomsen *et al.*, 1990; Carr *et al.*, 2008a; Carr *et al.*, 2010). This hypothesis was lately confirmed by Levin and King (2013). Similarly to colony formation, sexual reproduction in *S. rosetta* is triggered by a bacterial cue, the EroS enzyme produced by *Aliivibrio fischeri* (Beijernick) Urbanczyk (Woznica *et al.*, 2017; Woznica and King, 2018).

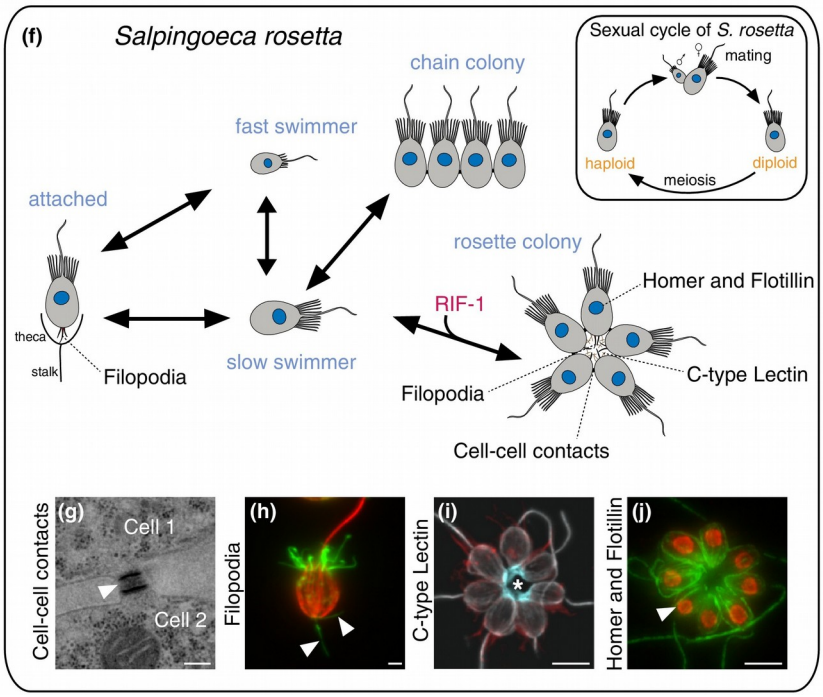

Figure 1.1: Life cycle of *S. rosetta* (f). Detail of a cytoplasmatic bridge between two cells in a *S. rosetta* colony (g). Filipodia in an attached *S. rosetta* cell (h). C-type Lectin located in the centre of a *S. rosetta* colony (i). The postsynaptic protein homologue Homer localised to the nucleus of *S. rosetta*. Scale bars: 200 nm (g), 1 µm (h), 5 µm (i, j). Modified from Hoffmeyer and Burkhardt (2016).

The fact that both colony formation and sexual reproduction in *S. rosetta* are triggered and regulated by bacterial cues suggest that the relationship between choanoflagellates and bacteria go beyond predator-prey relationships. This has been recently confirmed by the study of *S. monosierra* and its related microbiome. Colonies of *S. monosierra* exhibit a circular blastula-like architecture that contains an extracellular matrix in the centre of the colony. Such matrix is inhabited by a cohort of symbiotic bacteria that differs from the microbiome present surrounding the colony and resembles that of animal gut (Hake, 2019).

The phylogeny of choanoflagellates.

The first systematic approach to the taxonomy of choanoflagellates was carried out by Kent (1880-1882), who erected the Order Choanoflagellata, sub-divided in three families, namely Codonosigidae, Salpingoecidae and Phalansteriidae. Subsequent revision moved Phalansteriidae out of the choanoflagellates and added a third family, Acanthoecaceae (Norris, 1965; Cavalier-Smith, 2004).

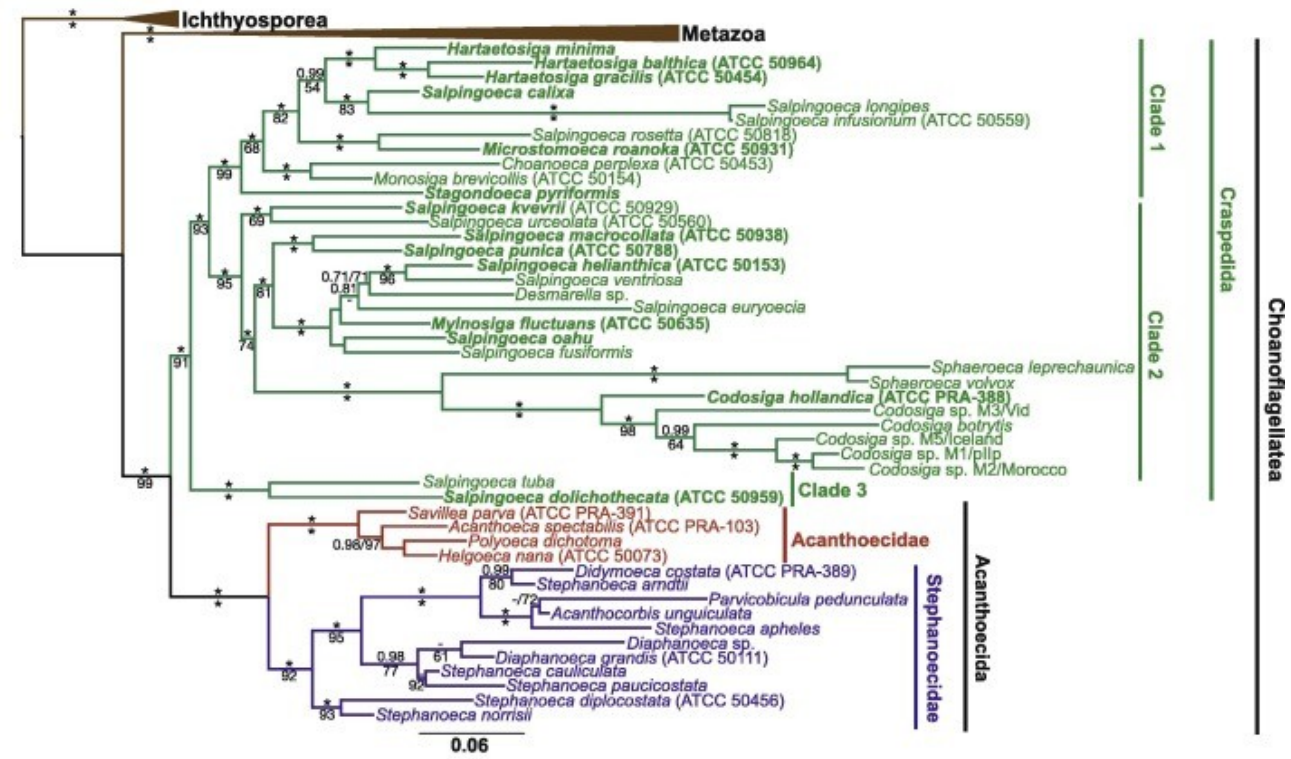

Figure 1.2: Phylogeny of choanoflagellates based on six concatenated genes. Choanoflagellates form a monophyletic group with a sister relationship to metazoans. Craspedida is divided into three clades. Acanthoecida is divided into two clades: Acanthoecidae and Stephanoecidae. From Carr *et al.* (2017).

Further revisions (Figure 1.2, Figure 1.3) using phylogenetic approaches (Carr *et al.*, 2008b; Nitsche *et al.*, 2011; Carr *et al.*, 2017) confirmed that the Class Choanoflagellatea forms a monophyletic group, divided into two orders: Craspedida and Acanthoecida. Order Craspedida is divided into three clades.

Clade 1 and Clade 3 are composed solely of marine species, while Clade 2 contains both marine and freshwater species. Order Acanthoecida, also known as loricates, is divided into two groups: Acanthoecidae, also known as nudiforms, and Stephanoecidae, also known as tectiforms. These two clades differ on the ontogeny of the lorica (Leadbeater, 2008b; Leadbeater *et al.*, 2008). These same studies have shown choanoflagellates to be the sister group of Metazoa, the group containing all the animal. Nevertheless, conflicting topologies for some key sections of the phylogenetic tree have been recovered in other studies, placing the root of nudiforms within the tectiforms (Medina *et al.*, 2003; Carr *et al.*, 2008b; López-Escardó *et al.*, 2019) and the species *Salpingoeca kvevrii* and *Salpingoeca urceolata*, previously placed in Clade 2, as basal to Clade 3 (López-Escardó *et al.*, 2019).

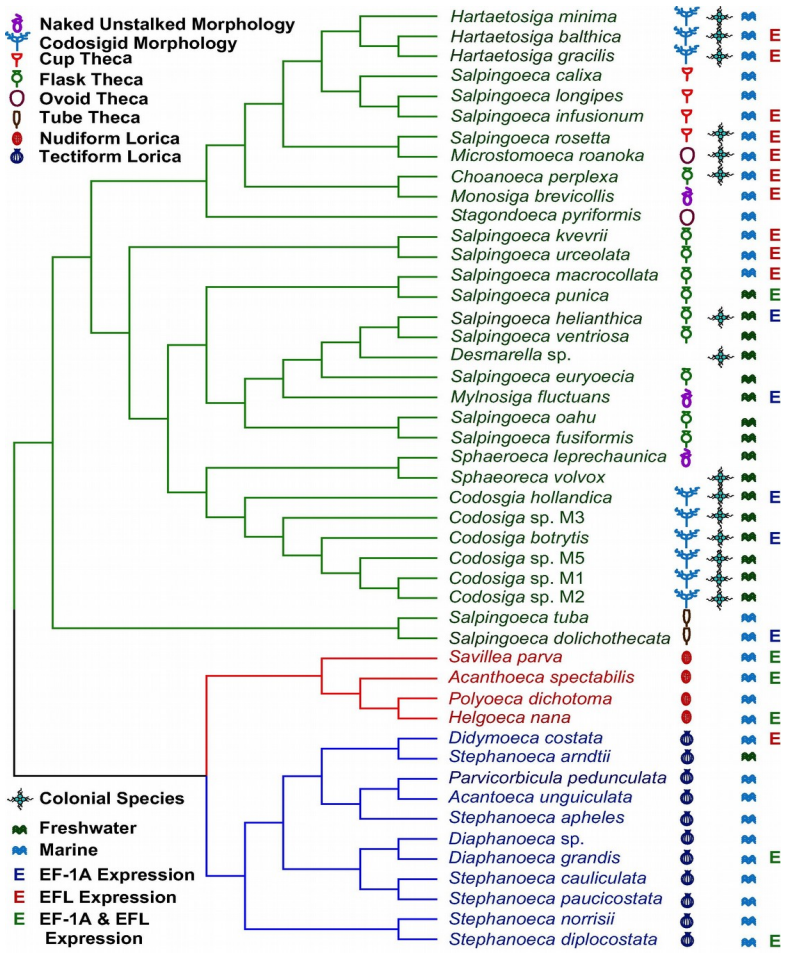

Figure 1.3: Cladogram of the phylogeny of choanoflagellates based on six concatenated genes. Most of the freshwater species are contained in one clade. Columns at the right of the name indicate cell morphology, ability to form colonies, ecology and EFL/EF-1a expression. From Carr et al. (2017).

Ancestral sequence reconstructions.

Ancestral sequence reconstruction is a methodology that allows to infer the sequences of last common ancestors, extrapolating back in time, from the sequences of extant species. In the context of biology, ancestral reconstructions allow not only to recover the nucleotides or amino acids sequences of taxa that are not present in the sampling, but also to recover other ancestral states such as gene order, phenotypic states, and geographic range of ancestral populations or species. Ancestral sequence reconstructions have gained traction over the last decade due to its application to protein engineering (Zakas *et al.*, 2016; Gumulya and Gillam, 2017; Gumulya *et al.*, 2018; Furukawa *et al.*, 2020; Laursen *et al.*, 2021). Nevertheless, the use of this methodology in the context of phylogenetics was first described by Dobzhansky and Sturtevant (1938).

From a phylogenetic perspective, ancestral sequence reconstructions allow to study the evolutionary relationships between extant species (tree leaves) and their last common ancestors (internal nodes). This methodology is used to infer the nucleotide sequences of the internal nodes of a given phylogeny. In order to do that, the analysis applies certain evolutionary model to a given phylogeny and estimates the sequence of the internal nodes of the tree. At a certain depth all extant species of a clade are included in the last common ancestor of that particular clade. Ancestral sequence reconstructions can be carried out using different approaches and algorithms that can be generally classified in: Maximum Parsimony, Maximum Likelihood and Bayesian Inference. Studies addressing the accuracy and systemic bias affecting such approaches have been previously showed that both ML and Bayesian methodologies outperform Maximum Parsimony reconstructions. Furthermore regardless of the approach, accuracy decreases at deeper nodes and that, for both ML and Bayesian analysis, model selection and specifically rate variation are key to reduce systemic bias and increase the accuracy (Krishnan *et al.*, 2004; Hanson-Smith *et al.*, 2010; Matsumoto *et al.*, 2015). A benchmark study carried out by Randall *et al* (2016) found out that both DNA and codon-based reconstructions carried out by empirical Bayesian approaches outperformed amino acid sequences reconstructions and were able to correctly infer >98% of the sites.

By allowing the comparison between ancestral and derived states, the analysis of the phylogenetic relationships between the extant species and the reconstructed ancestral sequences can help unravel the evolutionary history of the organisms present in a given phylogeny.

Marine:freshwater transitions: an overview.

Salinity is one of the major factors affecting osmotic pressure in aquatic environments. In protists, diverse cell functions such as photosynthesis, osmoregulation, cell growth, and silification are affected by salinity (González-Moreno *et al.*, 1997; Roubeix and Lancelot, 2008; Shabala *et* al., 2009). In stramenopiles, the response to salt stress involves chaperones, Na+/H+ transporters, lipid metabolisms, carbohydrate metabolism, and transcription factors (Harding *et al.*, 2017). Although species have different degrees of tolerance to salinity changes (Finley, 1930; Loefer, 1939; James et al., 2003), and euryhalinity has been reported in various to occur in phylogenetically distant species (Fenchel, 1987; Jones and Gates, 1994; Prokina and Mylnikov, 2017; Schiwitza *et al.*, 2019), marine-to-freshwater transitions in unicellular eukaryotes are generally regarded as infrequent events that occurred early in the evolution of many phyla (Logares *et al.*, 2009, Heger *et al.*, 2009). This is due to the fact that the extensive changes, at proteomic level, required to regulate all cell functions affected by salinity, require long evolutionary times (Cabello-Yeves and Rodriguez-Valera, 2019). However, several recent studies have reported that some eukaryotic phyla have transitioned multiple times between marine and freshwater environments (Bråte *et al.*, 2010; Annenkova *et al.*, 2015; Dittami *et al.*, 2017). Some phyla, such as dinoflagellates, have transitioned over ten times in the last 140 million years, most of them during the last 40 million years (Žerdoner Čalasan *et al.*, 2019). These young frequent transitions seem contradicting, considering the extensive genetic changes the species need undergo in order to overcome such environmental barrier. However, different salinity ranges exert different levels of adaptative pressure (Filker *et al.*, 2017). Some freshwater protists, like diatoms diversify and turn over faster than their marine counterparts, especially in benthic communities (Nakov *et al.*, 2019).

Due to its effects over cell functions and over communities diversification and turn over rates, salinity is one of the main drivers of evolution in protists. Despite their importance on the evolutionary history and current diversity of choanoflagellates, marine:freswater transitions remain mostly unexplored. This

is a consequence to the fact that all but one species of known freshwater choanoflagellates belong to a single clade (Clade 2 craspedids). The current phylogeny of choanoflagellates explains the current diversity of freshwater craspedids as derived from a single, large-scale, colonisation of freshwater habitats.

Phylogenetic placement of environmental sequences and its implications over marine:freshwater transitions in choanoflagellates evolutionary history.

The development of next-generation sequencing (NGS) techniques has largely been proved as a very useful tool for the amplification of DNA sequences from environmental samples (Roh *et al.*, 2010; Shokralla *et al.*, 2012; Taberlet *et al.*, 2012). The application of NGS techniques on environmental samples was first used and optimised for studying bacterial communities' composition (Tringe *et al.*, 2004; Woyke *et al.*, 2006; Sogin *et al.*, 2006) and lately adapted to the study of other organisms (Hajibabaei *et al.*, 2011; del Campo and Masana, 2011). In recent years, the availability of this techniques has led to an unprecedented effort in the survey of global biodiversity and many big projects and expeditions, such as TaraOceans and Malaspina 2010 expeditions, have gathered extensive amounts of eDNA sequences from seas around the globe (Pesant *et al.*, 2015; Duarte, 2015).

Previous studies have used data from environmental samples (eDNA) to estimate high values of hidden diversity amongst opisthokonts, and especially amongst choanoflagellates (del Campo and Masana, 2011; del Campo and Ruiz-Trillo, 2013; del Campo *et al.*, 2015; Arroyo *et al.*, 2018; Venter *et al.*, 2018). Such studies have shown that most choanoflagellates sequences retrieved from environmental samples belong to uncultured or unknown species. Phylogenetic reconstructions that included these environmental sequences have led to the proposal of new clades within the phylogeny of choanoflagellates composed solely of environmental sequences (del Campo and Ruiz-Trillo, 2013). One of those environmental clades, FRESCHO3, clustered with another environmental clade, Clade L, have been placed in a basal branch of the tree previous to the split between Craspedida and Acanthoeca (Figure 1.4). It was proposed that both clades represent an early divergence of choanoflagellates for which no representative species has been yet identified. Further revisions (Carr *et al.*, 2017), moved both FRESCHO3 and Clade L clades to a position basal to Acanthoecida (Figure 1.5). From those

clades for which cultured representatives exist, Acanthoecida is the most abundant in marine samples (del Campo *et al.*, 2015), while Craspedida is the most abundant in freshwater samples. However, several sequences related to Acanthoecida have been detected in the latest freshwater surveys, suggesting that the diversity of freshwater loricates might be higher than previously thought (Arroyo *et al.*, 2018). Moreover, a high number of choanoflagellates related sequences have been detected in soil samples (Venter *et al.*, 2018). Such sequences are placed all along the phylogenetic tree of choanoflagellates, some of them related to cultured species and others forming clusters that could not be related to any cultured species, suggesting the existence of entire groups of unknown soil choanoflagellates.

Such studies, however, relied on phylogenetic reconstructions based only in 18S rDNA sequences, which are not completely reliable when the recovery of the choanoflagellates phylogeny is attempted. The phylogenetic reconstruction obtained from this gene only shows important differences, although not strongly supported, to the currently accepted phylogeny of choanoflagellates (Carr *et al.*, 2008b; Carr *et al.*, 2017).

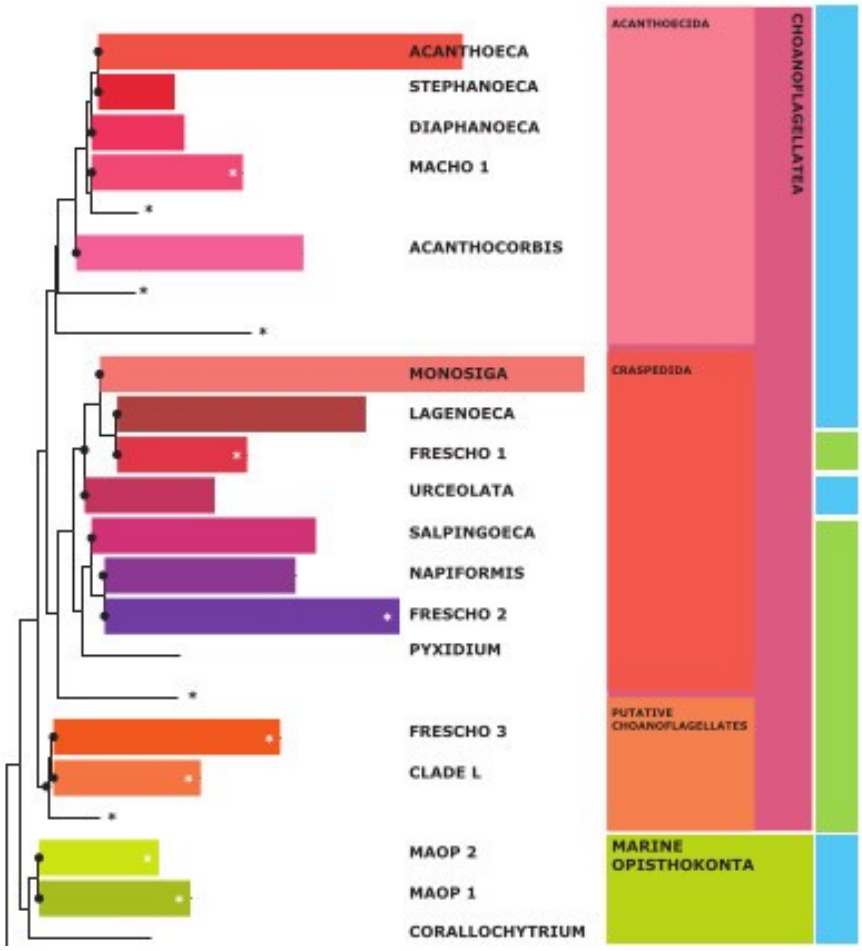

Figure 1.4: Maximum Likelihood (ML) phylogeny of choanoflagellates, including environmental sequences, based on 18S rDNA sequences. Clades marked with asterisk (*) indicate clades composed solely of environmental sequences. Last bar to the right indicates marine (blue) or freshwater (green) origin of the environmental sequences. Modified from del Campo and Ruiz-Trillo (2013).

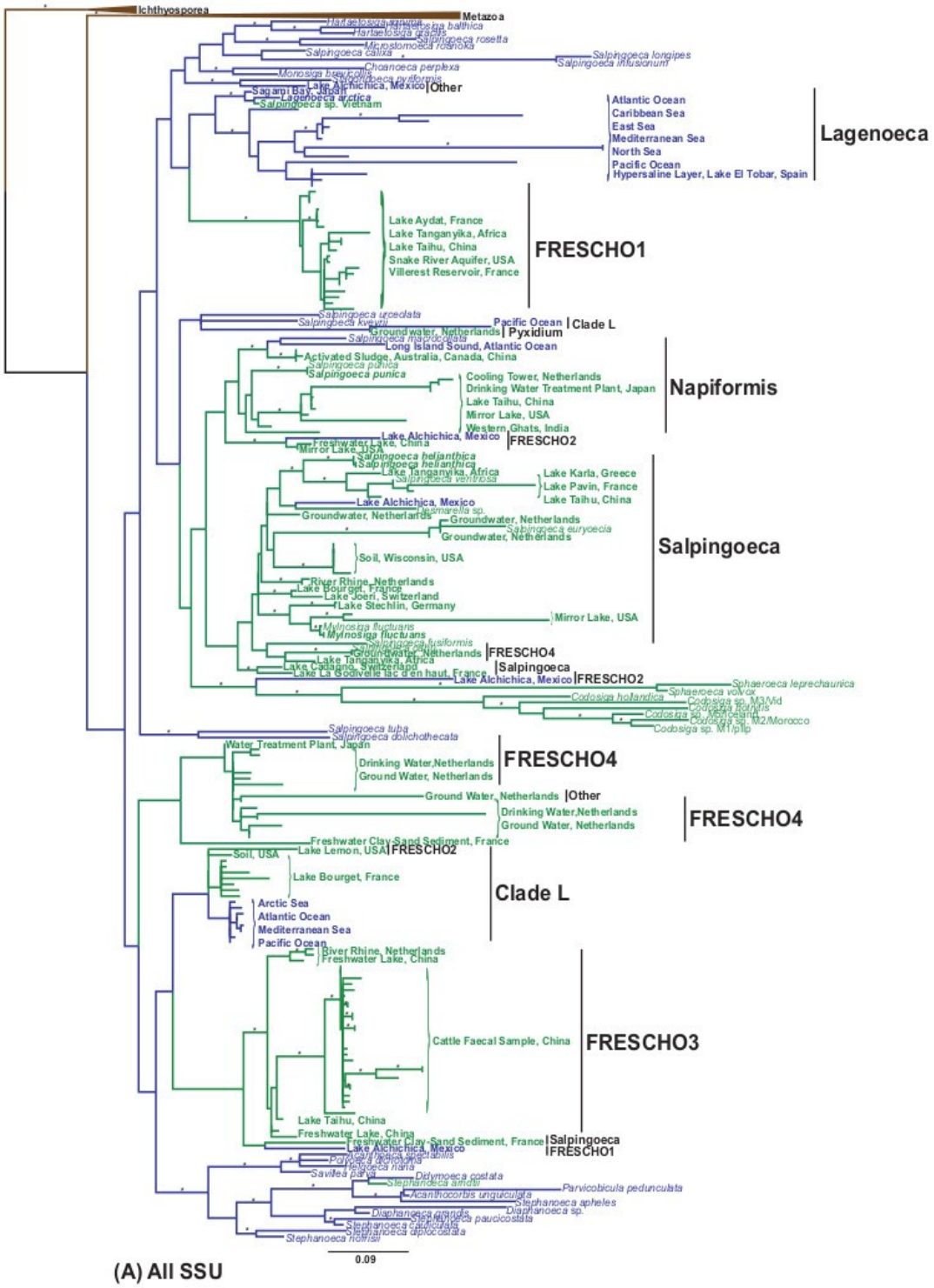

Figure 1.5: Phylogeny of the choanoflagellates including sequences from environmental samples. Branches coloured to indicate marine origin (blue) and freshwater origin (green). From Carr *et al.*, (2017).

Traditionally the phylogenetic placement of eDNA sequences has involved the clustering of sequences into Operational Taxonomic Units (OTUs) based on sequence similarity (Stackebrandt and Goebel, 1994; Creer *et al.*, 2010; Chariton *et al.*, 2015). This approach is highly dependent on the target organism and parameter selection (Patin *et al.*, 2013; Clare et al., 2016; Alberdi *et al*., 2017) and can produce a reduction in the resolution of the analysis that can lead to the loss of much of the hidden diversity (Venter *et al.*, 2018). Bioinformatic tools such as EPA-ng (Barbera *et al.*, 2019) and Pplacer (Matsen *et al*., 2010) use a clustering-free approach that avoids these problems resulting in a more reliable phylogenetic placement eDNA sequences.

Despite the efforts made towards the study of marine:freshwater transitions in choanoflagellates in recent years, the phylogenetic relationships between the different groups of freshwater choanoflagellates is not yet fully understood. Moreover, when only cultured species are considered only one major clade (Clade 2) contains all the freshwater species known, except for the loricates *Acanthocorbis mongolica* Paul and *Stephanoeca arndti* Nitsche (Figure 1.3). The currently accepted hypothesis is that most freshwater diversity in choanoflagellates is derived from a single, large freshwater invasion in Craspedida.

Autoecology of choanoflagellates.

Salinity tolerance experiments often involve the culture of organisms under different salt concentrations. They constitute a quick and cost-effective methodology for the determination of the osmotic tolerance of cultured organisms and provide very useful information that can have deep evolutionary implications (Knoll and Bauld, 1989; Hauer and Rogerson, 2005; Lowe et al., 2005; Koch and Ekelund, 2005). Despite its importance for the understanding of the autoecology and evolutionary history of organisms, salinity tolerance studies, in choanoflagellates, have received little regard and the available information on the salinity tolerance of choanoflagellates scarce and mostly is limited to a few closely related species (Rodriguez and Julian, 2009; Jeuck, 2014; Nitsche, 2014; Schiwitza, 2018; Schiwitza, 2019). A recent study confirmed the euryhalinity, the capacity to tolerate a wide range of salinities, in the nudiform *Enibas tolerabilis* (Schiwitza *et al.*, 2019). This species, isolated from a

hypersaline lake in the Atacama Desert, is able to grow in salinities ranging from 0 PSU (Practical Salinity Units) to 100 PSU.

Scope of the project.

The phylogenetic relationship between choanoflagellates and metazoans makes choanoflagellates an excellent model for the study of the origin of animals. One of the major factors affecting organisms in aquatic environments in the salinity of the environment. In that regard, there is a noticeable disparity between animals, with all major groups of aquatic animals but ctenophores having colonised freshwater environments, and choanoflagellates, with most of the freshwater diversity having derived from a single marine-to-freshwater transition. Nevertheless, the evolutionary history of marine:freshwater transitions in choanoflagellates remains poorly understood. The limited, but strongly supported, diversity of freshwater choanoflagellates observed in cultured species highly contrast with the greater, albeit weakly supported, diversity of freshwater species estimated from eDNA analysis. The goal of the present study is to fill the gap in the knowledge of freshwater choanoflagellates and to investigate the discrepancies between the diversity of cultured and uncultured freshwater choanoflagellates species, in order to unravel the evolutionary history of marine:freshwater transitions and provide a clear insight into the origin of freshwater groups. To achieve this goal, this study focused on the reconstruction of an updated and reliable phylogeny of choanoflagellates based on six genes. This phylogeny was then used as a reference for the phylogenetic placement of reconstructed ancestral sequences related to the freshwater groups and environmental sequences from a wide range on environments. Furthermore, the present study focused on the assessment of the tolerance to salinity changes in several species throughout the phylogenetic tree of choanoflagellates, in order to identify candidates for future studies focused on the transcriptomic changes occurring in choanoflagellates as a result of the adaptation to varying salinities. Additionally the present study involved extensive efforts towards the isolation of new choanoflagellates species from freshwater habitats, with special regard to both acidic and hyperalkaline environments. The content of the present study can be summarised as follows:

-Chapter II: The evolutionary history of marine:freshwater transitions in choanoflagellates from a phylogenetic perspective.

This chapter focuses on the study of the evolutionary history of marine:freshwater transitions in choanoflagellates from a phylogenetic perspective. The chapter subdivided in three sections:

-Section I: The phylogeny of choanoflagellates.

This section focuses on the reconstruction of an updated, reliable phylogeny of choanoflagellates. Through extensive phylogenetic analysis, the phylogeny, based on six genes, of all cultured choanoflagellates is here reviewed.

-Section II: Unravelling the evolutionary history of marine:freshwater transitions in choanoflagellates I: ancestral sequence reconstructions within Clade 2.

This section focuses on the reconstruction of ancestral sequences within Clade 2, the phylogenetic placement within the phylogeny of choanoflagellates and the implications of such placements in the evolutionary history of marine:freshwater transitions in choanoflagellates.

-Section III: Unravelling the evolutionary history of marine:freshwater transitions in choanoflagellates II: phylogenetic placement of environmental sequences.

This section focuses on the phylogenetic placement of environmental sequences within the phylogeny of choanoflagellates and the implications of such placements in the evolutionary history of marine:freshwater transitions in choanoflagellates.

-Chapter III: The evolutionary history of marine:freshwater transitions in Choanoflagellates from an ecological perspective.

This chapter focuses on the study of the evolutionary history of marine:freshwater transitions in choanoflagellates from an ecological perspective. The chapter subdivided in three sections:

-Section I: Field samplings, samples processing, choanoflagellates isolation and cultures maintenance.

This section focuses on the description of the efforts taken towards the isolation of new choanoflagellates and the maintenance of cultures of choanoflagellates in the laboratory. It includes the description of four newly isolated choanoflagellates morphospecies.

-Section II: Salinity tolerance in choanoflagellates.

This section focuses on the experiments carried out to assess the salinity tolerance of several species of choanoflagellates and to obtain samples for future studies on the transcriptomic changes necessary for the adaptation to varying salinities.

-Section III: pH tolerance in choanoflagellates.

This section focuses on the experiments carried out to assess the pH tolerance of the four newly isolated choanoflagellates morphospecies.

-Chapter IV The evolutionary history of marine:freshwater transitions in Choanoflagellates. Current knowledge and future perspectives.

This chapter focuses on summarising all the evidence gathered during the present work and discussing it from an evolutionary perspective. It also explores the future perspectives of the study on marine:freshwater transitions and the phylogeny of choanoflagellates.

Chapter II: The evolutionary history of marine:freshwater transitions in Choanoflagellates from a phylogenetic perspective.

-Section I: The phylogeny of choanoflagellates.

Introduction

Choanoflagellates are ubiquitous and one of the most important groups of heterotrophic pico- and nanoplankton. Their feeding habits confer them a crucial role in the microbial loop of energy recycling and make them one of the primary means by which much of the organic carbon becomes available to higher trophic levels (Fenchel, 1982a; Azam *et al.*, 1983). They are the closest relatives to metazoans (Carr *et al.*, 2017) and are considered a very valuable group for the study of the early evolution, in animals, of some key life traits as important as sexual reproduction or multicellularity (King *et al.*, 2003; King, 2004; Snell *et al.*, 2006; Abedin and King, 2008; Alegado *et al.*, 2012; Woznica *et al.*, 2016; Laudon *et al.*, 2019).

The phylogeny of freshwater choanoflagellates remains poorly understood. Recent studies (Carr *et al.*, 2017; Schiwitza *et al.*, 2018; Schiwitza *et al.*, 2019; López-Escardó *et al.*, 2019) have recovered conflicting phylogenetic placements of key species closely related to Clade 2 craspedids, the group containing most of the freshwater species know. The currently accepted hypothesis is that most freshwater diversity in choanoflagellates is derived from a single, large freshwater invasion in Craspedida. In order to test such hypothesis, an exhaustive phylogenetic study was carried out.

Material and methods.

Whole genome amplification, PCR amplification, cloning and sequencing.

DNA extractions from single cells of *Paramonosiga thecata*, *Salpingoeca amphora* and *Salpingoeca steinii* were carried out, following the protocol described in Nitsche (2014), and kindly donated for the present study by Frank Nitsche. Whole genome amplification was performed using REPLI-g Single Cell Kit (QIAgen), following the protocol for the amplification of purified genomic DNA proposed by the manufacturer.

Amplified whole-genome DNA was used as starting material to attempt amplification, by PCR, of both 18S and 28S ribosomal subunits (SSU and LSU, respectively), α-tubulin (tubA), elongation factor 1-A (EF-1A) and elongation factor-like (EFL) genes (details on the primers for each reaction can be found below). All PCR amplifications consisted of a denaturing step at 94°C for 2 min, followed by 30 cycles of denaturing at 94°C for 15 s, annealing at varying temperatures (further explained below) for 15 s and extension at 72°C for 1 min, and a final extension step at 72°C for 10 min. PCR products were analysed on 1% agarose gels and later purified using the QIAquick Gel Extraction kit (Qiagen, UK), following the protocol proposed by the manufacturer with a minor modification. The modification consisted in reducing the volume of elution buffer, from 25µl to 10µl, in the first elution step, in order to increase DNA concentration.

SSU amplification.

SSU was amplified using the universal primers 1F and 1528R developed by Medlin et al. (1988) using an annealing temperature of 57.8°C. PCR mastermix contained 10µl DNA template, 10µl 5x MyTaq Red Reaction Buffer (Bioline, UK), 1.2µl of each forward and reward primers, 1.2µl MyTaq Red DNA Polymerase (Bioline, UK) and 26.4µl deionised water.

LSU amplification.

LSU was amplified in four fragments following the protocol from Carr et al. (2008). PCR mastermix contained 10µl DNA template, 10µl 5x MyTaq Red Reaction Buffer (Bioline, UK), 1.2µl of each forward and reward primers, 1.2µl MyTaq Red DNA Polymerase (Bioline, UK) and 26.4µl deionised water. PCR was run with an annealing temperature of 58°C.

tubA amplification.

The protocol for the amplification of tubA consisted in two nested PCR. The primers developed by Carr et al. (2008b) were used for both steps. The initial reaction was carried out using the primers atF1 and atR3, with a 10°C touchdown step reducing the annealing temperature from 60°C to 50°C, over 10 cycles. The second PCR was performed using the primer pairs atF3 + atR2 or atF3 + atR1, with an annealing temperature of 58°C. PCR master mixes contained 10µl DNA template, 10µl 5x MyTaq Red Reaction Buffer, 1µl of each forward and reward primers, 1µl MyTaq Red DNA Polymerase and 27µl deionised water and 1µl DNA template, 10µl 5x MyTaq Red Reaction Buffer, 1µl of each forward and reward primers, 1 µl MyTaq Red DNA Polymerase and 36µl deionised water, for first and second reactions respectively.

EFL and EF-1A amplification.

For the amplification of EFL and EF-1A a set of new primers were developed (Table 2.1.1). The protocol for the amplification consisted in two nested PCR. The initial reaction performed was common for both genes and used a touchdown approach, reducing the annealing temperature in 5°C, from 60°C to 55°C, over 5 cycles. The primer pairs for the initial reaction were EFL_F1 + EFL_R1 and EF1A_F1 + EF1A_R1, for EFL and EF-1A respectively. The primer pairs used in the following nested PCR, for the amplification of EFL, were EFL_F2 + EFL_R2 or EFL_F2 + EFL_R3, with a 5°C touchdown step reducing the annealing temperature, over 5 cycles, from 56°C to 51°C. For the amplification of EF-1A, the primers used for the nested PCR were EF1A_F2 + EF1A_R2 or EF1A_F3 + EF1A_R3, with a 5°C touchdown step reducing the annealing temperature, over 5°C, from 58.1°C to

53.1°C and from 58.8°C to 53.8°C respectively. PCR master mixes were similar to those used for the amplification of tubA sequences.

Two approaches were used for the sequencing of amplified PCR products. For SSU and LSU genes, Sanger sequencing was carried out from purified PCR amplicons. Amplified SSU fragments were sequenced in 3'-5' direction while LSU fragments were sequenced in both 3'-5' and 5'-3' directions. For tubA, EF-1A and EFL genes, purified PCR products were cloned into the plasmid pGEM-T Easy (Promega) and transformed into subcloning Efficiency DH5α competent cells (Invitrogen), following the protocols developed by the manufacturers. Plasmid DNA was extracted from 6 ml overnight cultures using the QIAprep Miniprep kit (Qiagen), following the protocol proposed by the manufacturer. Purified plasmid DNA was then used for Sanger sequencing. All sequencing was carried out by Macrogen (Macrogen, The Netherlands). Upon receipt, sequences were manually checked and low quality tails were removed using 4Peaks v1.8 (Griekspoor and Groothuis). LSU fragments were aligned against reference LSU sequences from our dataset and merged in a single sequence using MEGA v7 (Kumar et al., 2016).

Table 2.1.1: Primers designed for this study. Primers were designed using conserved sites from alignments of EF-1A or and EFL sequences from Carr et al. (2017)

Gene	Primer	Sequence
EFL	EFL_F1	5’-TGCGGCCACGTCGATGCYGG-3’
	EFL_F2	5’-TTCATCAAGAACATGATCACYGG-3’
	EFL_R1	5’-GTTNCCYTCCATRATGGCRA-3’
	EFL_R2	5’-TCNCCGACRCGNGGCATGTT-3’
	EFL_R3	5’-TCNCCRACACCCTTGATCTTG-3’
EFL1A	EF1A_F1	5’-GYGAGCGYGGTATCACCATCG-3’
	EF1A_F2	5’-GGTGAGTTCGAGGCTGGTATC-3’
	EF1A_F3	5’-GGTGAGTTCGAGGCTGGTATC-3’
	EF1A_R1	5’-GCRATGTGRGCRGTGTGGCA-3’
	EF1A_R2	5’-CCACGGCGRATRTCCTTSAC-3’
	EF1A_R3	5’-TCGTGGTGCATCTCNACAGAC-3’

Phylogenetic reconstruction.

The most comprehensive, to our knowledge, phylogenetic reconstruction of choanoflagellates up until today, was reconstructed for this project. This phylogeny comprises sequences from 63 choanoflagellates species and 8 outgroup species (Table A2.1.1, appendix). The dataset includes both compete and partial sequences from 6 genes, namely: SSU and LSU ribosomal subunits, heatshock protein 90kD (HSP90), tubA, EF1-A and EFL. Due to unavailability, most of the species included in the dataset contained missing data, with only 10 species containing sequences for the complete set of genes.

The subsets of ribosomal sequences contained rDNA sequences from direct Sanger sequencing. On the other hand, the subsets from HSP90, tubA, EF-1A and EFL contain exclusively coding sequences (CDS) from these genes. Compared to previously published phylogenies (Carr et al, 2017) *Sphaeroeca volvox* and *Desmarella* sp. were excluded from the analysis due to the ambiguity in the origins of their

sequences. Since the description and sequencing of these species, all cultures have been lost and, therefore, it is impossible to confirm the validity of their sequences.

Reference alignment.

In order to build the multi-genes reference alignment each of the subsets was aligned independently. Aligned subsets were then concatenated into two multigene multiple sequences alignments (MSA). The first MSA contained the concatenated alignments of both ribosomal subunits, while the second MSA contained the concatenated alignment of the remaining four genes.

Ribosomal subunits subsets.

Due to the difficulties encountered when trying to produce high quality alignments for SSU and LSU subsets, derived from the presence of hypervariable regions in them, two main approaches were tested to build these alignments. The first approach was building the alignments using several different software independently. Ribosomal subunits sequences' alignment was built independently using MAFFT v7.310 (Katho and Standley, 2013), Muscle v3.5.11(Edgar, 2004a), HMMER v3.2 (Mistry et al., 2013), Clustal Omega v1.2.4 (Sievers et al., 2011) and SSU-Align v0.1.1 (Nawrocki, 2009). Alignments were manually checked for quality assessment using AliView v1.21 (Larsson, 2014). The second approach was a two steps alignment method. On a first step, SSU and LSU alignments were built using the most accurate iterative refinement method, L-INS-i, of MAFFT v7.310, that incorporates pairwise alignment information, using default gap opening and gap extension values. On a second step, upon visual inspection, the poorly aligned regions, corresponding to the hypervariable regions of the ribosomal subunits, were realigned in AliView v1.26 (Larsson, 2014) using the built-in MUSCLE v3.8.425 (Edgar, 2004a; Edgar 2004b) extension with default parameters. Realigned blocks contained the hypervariable region and flanking regions of at least 5 conserved nucleotides. Resulting alignments were trimmed using trimAl v1.2rev59 (Capella-Gutierrez *et al*, 2009), using the heuristic selection of the automatic method based on similarity statistics. As stated by the authors, this method is optimised for ML phylogenetic tree reconstructions.

Protein coding subsets.

The alignments of the protein coding subsets, HSP90, tubA, EF-1A and EFL; were built using AliView v1.26. DNA sequences were translated into amino acids sequences, aligned using the built-in MUSCLE v3.8.425 extension with default parameters, trimmed using AliView v1.26 built-in trimming extension and finally translated back to the original nucleotide sequences. This process avoids the break-up of codons, thus retaining the phylogenetic information contained in the proteins' sequences similarity.

Data partitioning and model selection.

For the reconstruction of the choanoflagellates phylogenetic tree, several data partitioning schemes were tested (Table 2.1.2). The most simple partitioning schemes comprised two partitions, one containing the ribosomal subunits sequences and one containing the remaining four genes, and the most complex schemes comprised six partitions, one for every individual gene.

Prior to the reconstruction of the phylogenetic tree, the best-fitting substitution model was estimated for each partition, following the Akaike Information Criterion corrected for small sample sizes (AICc), using ModelFinder (Kalyaanamoorthy et al., 2017) in IQ-TREE v1.6.12 (Nguyen, 2015; Chernomor *et al.*, 2016) with an extended model selection that allows FreeRate models (Yang,1995; Soubrier et al., 2012). Each partitioning scheme was run three times, using two different types of molecular data, nucleotide sequences and amino acids sequences, as inputs. For the ribosomal subunits (SSU and LSU), only the nucleotides sequences were used as input data, while for the remaining genes (HSP90, tubA, EF-1A and EFL), both nucleotides and amino acids sequences were used as input data. Furthermore, when using nucleotides sequences as input data for the protein coding genes, the best-fitting substitution models were estimated using two different sets of models, nucleotides and codon substitution models. Codon substitution models consider both the changes in the nucleotide sequence and their effect over the amino acid sequence. These substitution models retain the phylogenetic information contained in the proteins' sequences since they are able to discriminate between synonymous and non-synonymous changes in the DNA sequence. In addition to this partitioning

schemes, another partitioning scheme was also used which split the nucleotide sequence of all protein coding genes into three partitions, one for each nucleotide position of the codons.

Table 2.1.2: Partitioning schemes tested for the reconstruction of choanoflagellates' phylogeny. Partitions describe the subsets of sequences included on each partition. Substitution models indicate the type of substitution models (nucleotides, amino acids or codon) that were tested for each partition.

Partitioning scheme	Number of partitions	Partitions	Substitution Models
2-PartitionsA	2	1: Ribosomal (SSU+LSU) 2: Coding (HSP90, tubA, EF-1A, EFL)	Nucleotides
2-PartitionsB	2	1: Ribosomal (SSU+LSU) 2: Coding (HSP90, tubA, EF-1A, EFL)	Partition 1: Nucleotides Partition 2: Amino acid
2-PartitionsC	2	1: Ribosomal (SSU+LSU) 2: Coding (HSP90, tubA, EF-1A, EFL)	Partition 1: Nucleotides Partition 2: Codon
4-PartitionsA	4	1: Ribosomal (SSU+LSU) 2: HSP90 3: tubA 4: EF-1A+EFL	Nucleotides
4-PartitionsB	4	1: Ribosomal (SSU+LSU) 2: HSP90 3: tubA 4: EF-1A+EFL	Partition 1: Nucleotides Partitions 2-4: amino acids
4-PartitionsC	4	1: Ribosomal (SSU+LSU) 2: HSP90 3: tubA 4: EF-1A+EFL	Partition 1: Nucleotides Partitions 2-4: Codon
4-PartitionsD	4	1: Ribosomal (SSU+LSU) 2: 1st nucleotide of the codon (HSP90+tubA+EF-1A+EFL) 3: 2nd nucleotide of the codon (HSP90+tubA+EF-1A+EFL) 4: 3rd nucleotide of the codon (HSP90+tubA+EF-1A+EFL)	Nucleotides
6-PartitionsA	6	1: SSU 2: LSU 3: HSP90 4: tubA 5: EF-1A 6: EFL	Nucleotides

Partitioning scheme	Number of partitions	Partitions	Substitution Models
6-PartitionsB	6	1: SSU 2: LSU 3: HSP90 4: tubA 5: EF-1A 6: EFL	Partitions 1-2: Nucleotides Partitions 3-6: amino acids
6-PartitionsC	6	1: SSU 2: LSU 3: HSP90 4: tubA 5: EF-1A 6: EFL	Partitions 1-2: Nucleotides Partitions 3-6: Codon

Phylogenetic tree reconstruction.

Maximum-likelihood (ML) phylogenetic trees of choanoflagellates were reconstructed, in IQ-TREE, using the above described partitioning schemes, using the best-fitting substitution models estimated by ModelFinder. In all cases edge-proportional partition models (Chernomor *et al.*, 2016) were used. According to IQ-TREE developers, this partition model is the recommended for typical analysis since it accommodates different evolutionary rates between partitions and each partition has its own partition specific evolutionary rate that rescales all its branch lengths. Maximum-likelihood (ML) values were estimated using 100000 ultrafast bootstrap replicates (Minh *et al.*, 2013). The unrooted consensus tree was the rooted using the Ichthyosporeans *Ichthyophonus hoferi* and *Amoebidium parasiticum* as outgroups and the graphical representation of the tree was produced using Archaeopteryx v0.9920 (Zmasek and Eddy, 2001).

From the resulting ten phylogenetic trees, an extended majority-rule consensus tree was built using IQ-TREE. Such trees are built following the topology recovered from the majority of the trees used to their reconstruction. Splits support of these trees is calculated as the percentage of trees from which the same topology was recovered.

During the early stages of the phylogenetic analyses, alternative phylogenetic reconstructions were carried out in RAxML v8.2.9 (Stamatakis, 2018) and MrBayes v3.2.6 (Ronquist *et* al., 2012). However, due to the lack of some of the substitution models present in IQ-TREE, especially, but not limited to, codon substitution models, the results from these software were not fully comparable with the results obtained from IQ-TREE, and therefore the construction of these alternative phylogenies was discontinued.

Results.

Whole genome amplification, PCR amplification, cloning and sequencing.

From the amplicons sent for sequencing, only complete SSU and LSU sequences from *S. steinii* and partial LSU sequences from *S. amphora* were obtained and included in further analysis.

Phylogenetic reconstruction.

Reference alignment.

After concatenation and sequence trimming, the best alignment produced by the single-step approach was 4161bp long, obtained by aligning the SSU and LSU subsets on SSU-Align and MAFFT, respectively. In comparison, the alignment produced by the two-steps approach was ~13.5% longer, with a final length, after concatenation and sequence trimming, of 4722bp.

Data partitioning and model selection.

2-PartitionsA.

According to AICc, the best-fitting substitution models estimated by ModelFinder were GTR+F+R10 for the first partition and GTR+F+R4 for the second partition (Table 2.1.3). The results were also supported by Akaike Information Criterion (AIC). On the other hand, Bayesian Information Criterion (BIC) supported the same substitution model for partition 2, while the best-fitting substitution model for partition 1, according to this criterion, was TIM2+F+R5 (Table 2.1.4). The accumulative likelihood (w-AIC, w-AICc and w-BIC) of the best-fitting model was higher for BIC than for AIC/AICc (Table 2.1.4).

Table 2.1.3: Results of the model selection for partitioning scheme 2-PartitionsA estimated by ModelFinder. Log-Likelihood (LogL) and degrees of freedom (df) were used to calculate the accumulative likelihood of the best models under Akaike Information Criterion (AIC), AIC corrected for small sample sizes (AICc) and Bayesian Information Criterion (BIC) (w-AIC, w-AICc and w-BIC respectively).

Partition	Best-fitting model (AIC/AICc)	LogL	df	w-AICc	w-AIC	w-BIC
Partition 1	GTR+F+R10	-72895.7550	27	0.6829	0.6887	7.3281×10^{-13}
Partition 2	GTR+F+R4	-63297.9622	15	0.6635	0.6610	0.8650

Table 2.1.4: Comparison of model selection for partitioning scheme 2-PartitionsA estimated, by ModelFinder, under AIC/AICc criteria and BIC criterion.

Partition	Best-fitting model (AIC/AICc)	Best-fitting model (BIC)	w-AIC/w-AICc	w-BIC
Partition 1	GTR+F+R10	TIM2+F+R5	0.6887/0.6829	0.9204
Partition 2	GTR+F+R4	GTR+F+R4	0.6610/0.6635	0.8650

2-PartitionsB.

According to AICc, the best-fitting substitution models estimated by ModelFinder were GTR+F+R10 for the first partition and LG+F+R4 for the second partition (Table 2.1.5). The results were also

supported by AIC. On the other hand, best-fitting substitution models according to BIC were TIM2+F+R5 and LG+F+I+G4 for partitions 1 and 2 respectively (Table 2.1.6). For the best-fitting substitution model in both partitions, w-AIC/w-AICc was higher than w-BIC (Table 2.1.6).

Table 2.1.5: Results of the model selection for partitioning scheme 2-PartitionsB estimated by ModelFinder. Log-Likelihood (LogL) and degrees of freedom (df) were used to calculate the accumulative likelihood of the best models under Akaike Information Criterion (AIC), AIC corrected for small sample sizes (AICc) and Bayesian Information Criterion (BIC) (w-AIC, w-AICc and w-BIC respectively).

Partition	Best-fitting model (AIC/AICc)	LogL	df	w-AICc	w-AIC	w-BIC
Partition 1	GTR+F+R10	-71787.7015	27	0.9312	0.9326	$9.3841x10^{-6}$
Partition 2	LG+F+R4	-25478.8216	26	0.8858	0.8800	0.0504

Table 2.1.6: Comparison of model selection for partitioning scheme 2-PartitionsB estimated, by ModelFinder, under AIC/AICc criteria and BIC criterion.

Partition	Best-fitting model (AIC/AICc)	Best-fitting model (BIC)	Best-fitting model accumulative likelihood (AIC/AICc)	Best-fitting model accumulative likelihood (BIC)
Partition 1	GTR+F+R10	TIM2+F+R5	0.9326/0.9312	0.5918
Partition 2	LG+F+R4	LG+F+I+G4	0.8800/0.8858	0.8395

2-PartitionsC.

According to AICc, the best-fitting substitution models estimated by ModelFinder were GTR+F+R10 for the first partition and KOSI07+F+R4 for the second partition (Table 2.1.7). The results were also supported by AIC. On the other hand, BIC supported the same substitution model for partition 2, while the best-fitting substitution model for partition 1, according to this criterion, was GTR+F+R4 (Table 2.1.8). For the first partition, best-fitting substitution model's w-AIC/w-AICc was higher than w-BIC, while for Partition 2 all three values were equal (Table 2.1.8).

Table 2.1.7: Results of the model selection for partitioning scheme 2-PartitionsC estimated by ModelFinder. Log-Likelihood (LogL) and degrees of freedom (df) were used to calculate the accumulative likelihood of the best models under Akaike Information Criterion (AIC), AIC corrected for small sample sizes (AICc) and Bayesian Information Criterion (BIC) (w-AIC, w-AICc and w-BIC respectively).

Partition	Best-fitting model (AIC/AICc)	LogL	df	w-AICc	w-AIC	w-BIC
Partition 1	GTR+F+R10	-72406.7300	27	0.9287	0.9303	$1.7979x10^{-6}$
Partition 2	KOSI07+F+R4	-54595.2036	67	0.9999	0.9999	0.9999

Table 2.1.8: Comparison of model selection for partitioning scheme 2-PartitionsC estimated, by ModelFinder, under AIC/AICc criteria and BIC criterion.

Partition	Best-fitting model (AIC/AICc)	Best-fitting model (BIC)	Best-fitting model accumulative likelihood (AIC/AICc)	Best-fitting model accumulative likelihood (BIC)
Partition 1	GTR+F+R10	GTR+F+R4	0.9303/0.9287	0.5574
Partition 2	KOSI07+F+R4	KOSI07+F+R4	0.9999/0.9999	0.9999

4-PartitionsA.

According to AICc, the best-fitting substitution models estimated by ModelFinder were GTR+F+R9 for the first partition, GTR+F+R4 for the second partition, GTR+F+R4 for the third partition and GTR+F+R3 for the fourth partition (Table 2.1.9). The results were also supported by AIC. However, the support for GTR+F+R9 as the best-fitting substitution model was weak for both AIC and AICc (w-$AIC_{Partition1}$=0.3447, w-$AICc_{Partition1}$=0.3412). On the other hand, BIC supported the same substitution models for Partition 2 only, while the best-fitting substitution models for partitions 1, 3 and 4, according to this criterion, were TIM2+F+R5, TIM+F+I+G4 and GTR+F+I+G4 respectively (Table 2.1.10). For Partition 3 and 4, for the best-fitting substitution model, w-AIC/w-AICc was higher than w-BIC, while the opposite is true for Partition 1 (Table 2.1.10).

Table 2.1.9: Results of the model selection for partitioning scheme 4-PartitionsA estimated by ModelFinder. Log-Likelihood (LogL) and degrees of freedom (df) were used to calculate the accumulative likelihood of the best models under Akaike Information Criterion (AIC), AIC corrected for small sample sizes (AICc) and Bayesian Information Criterion (BIC) (w-AIC, w-AICc and w-BIC respectively).

Partition	Best-fitting model (AIC/AICc)	LogL	df	w-AICc	w-AIC	w-BIC
Partition 1	GTR+F+R9	-72952.4730	27	0.3412	0.3447	$1.8749x10^{-11}$
Partition 2	GTR+F+R4	-24647.5023	15	0.8897	0.8788	0.6036
Partition 3	GTR+F+R4	-11133.8906	15	0.8228	0.8170	0.0664
Partition 4	GTR+F+R3	-27251.8253	13	0.8041	0.8023	0.0465

Table 2.1.10: Comparison of model selection for partitioning scheme 4-PartitionsA estimated, by ModelFinder, under AIC/AICc criteria and BIC criterion.

Partition	Best-fitting model (AIC/AICc)	Best-fitting model (BIC)	Best-fitting model accumulative likelihood (AIC/AICc)	Best-fitting model accumulative likelihood (BIC)
Partition 1	GTR+F+R9	TIM2+F+R5	0.3447/0.3412	0.8631
Partition 2	GTR+F+R4	GTR+F+R4	0.8788/0.8897	0.6036
Partition 3	GTR+F+R4	TIM+F+I+G4	0.8170/0.8228	0.2249
Partition 4	GTR+F+R3	GTR+F+I+G4	0.8023/0.8041	0.5267

4-PartitionsB.

According to AICc, the best-fitting substitution models estimated by ModelFinder were GTR+F+R10 for the first partition, LG+R4 for the second partition, mtART+F+R3 for the third partition and LG+F+I+G4 for the fourth partition (Table 2.1.11). The results were also supported by AIC. On the other hand, BIC only supported the same substitution models for partition 4, while the best-fitting substitution models for partitions 1, 2 and 3, according to this criterion, were TIM2+F+R5, LG+I+G4 and LG+R3 respectively (Table 2.1.12). In the cases when discrepancies between AIC/AICc and BIC existed, best-fitting substitution model's w-AIC/w-AICc was always higher than w-BIC (Table 2.1.12).

Table 2.1.11: Results of the model selection for partitioning scheme 4-PartitionsB estimated by ModelFinder. Log-Likelihood (LogL) and degrees of freedom (df) were used to calculate the accumulative likelihood of the best models under Akaike Information Criterion (AIC), AIC corrected for small sample sizes (AICc) and Bayesian Information Criterion (BIC) (w-AIC, w-AICc and w-BIC respectively).

Partition	Best-fitting model (AIC/AICc)	LogL	df	w-AICc	w-AIC	w-BIC
Partition 1	GTR+F+R10	-71800.1295	27	0.9162	0.9180	$4.1115x10^{-6}$
Partition 2	LG+R4	-9484.2694	7	0.8813	0.8755	0.1567
Partition 3	mtART+F+R3	-2496.7117	24	0.9027	0.8785	$5.4307x10^{-12}$
Partition 4	LG+F+I+G4	-13228.9161	22	0.6194	0.5607	0.9461

Table 2.1.12: Comparison of model selection for partitioning scheme 4-PartitionsB estimated, by ModelFinder, under AIC/AICc criteria and BIC criterion.

Partition	Best-fitting model (AIC/AICc)	Best-fitting model (BIC)	Best-fitting model accumulative likelihood (AIC/AICc)	Best-fitting model accumulative likelihood (BIC)
Partition 1	GTR+F+R10	TIM2+F+R5	0.9180/0.9162	0.5850
Partition 2	LG+R4	LG+I+G4	0.8755/0.8813	0.8232
Partition 3	mtART+F+R3	LG+R3	0.8785/0.9027	0.8330
Partition 4	LG+F+I+G4	LG+F+I+G4	0.5607/0.6194	0.9461

4-PartitionsC.

According to AICc, the best-fitting substitution models estimated by ModelFinder were GTR+F+R10 for the first partition, SCHN05+F+G4 for the second partition, KOSI07+F+R3 for the third partition and KOSI07+F+R4 for the fourth partition (Table 2.1.13). The results were also supported by Akaike Information Criterion (AIC). On the other hand, Bayesian Information Criterion (BIC) supported the same substitution models for partitions 2 and 3, while the best-fitting substitution models for partitions 1 and 4, according to this criterion, were GTR+F+R4 and KOSI07+F+I+G4 respectively (Table 2.1.14). For the best-fitting substitution model estimated for each partition w-AIC/w-AICC was only higher than w-BIC in Partition 1 (Table 2.1.14).

Table 2.1.13: Results of the model selection for partitioning scheme 4-PartitionsC estimated by ModelFinder. Log-Likelihood (LogL) and degrees of freedom (df) were used to calculate the accumulative likelihood of the best models under Akaike Information Criterion (AIC), AIC corrected for small sample sizes (AICc) and Bayesian Information Criterion (BIC) (w-AIC, w-AICc and w-BIC respectively).

Partition	Best-fitting model (AIC/AICc)	LogL	df	w-AICc	w-AIC	w-BIC
Partition 1	GTR+F+R10	-72344.1025	27	0.8908	0.8930	$1.6104x10^{-9}$
Partition 2	SCHN05+F+G4	-20621.0081	62	0.7733	0.7311	0.9622
Partition 3	KOSI07+F+R3	-9456.1719	65	0.8797	0.7728	0.9951
Partition 4	KOSI07+F+R4	-24055.5112	67	0.9815	0.9753	0.0824

Table 2.1.14: Comparison of model selection for partitioning scheme 4-PartitionsC estimated, by ModelFinder, under AIC/AICc criteria and BIC criterion.

Partition	Best-fitting model (AIC/AICc)	Best-fitting model (BIC)	Best-fitting model accumulative likelihood (AIC/AICc)	Best-fitting model accumulative likelihood (BIC)
Partition 1	GTR+F+R10	GTR+F+R4	0.8930/0.8908	0.7424
Partition 2	SCHN05+F+G4	SCHN05+F+G4	0.7311/0.7733	0.9622
Partition 3	KOSI07+F+R3	KOSI07+F+R3	0.7728/0.8797	0.9951
Partition 4	KOSI07+F+R4	KOSI07+F+I+G4	0.9753/0.9815	0.9176

4-PartitionsD.

According to AICc, the best-fitting substitution models estimated by ModelFinder were GTR+F+R10 for the first partition, GTR+F+R3 for the second partition, GTR+F+I+G4 for the third partition and GTR+F+R4 for the fourth partition (Table 2.1.15). The results were also supported by AIC. On the other hand, BIC supported the same substitution models for partitions 3 and 4, while the best-fitting substitution models for partitions 1 and 2, according to this criterion, were TIM2+F+R4 and GTR+F+I+G4 respectively (Table 2.1.16). For partition 1, w-AIC/w-AICc of the best-fitting substitution model was higher w-BIC, while the opposite is correct for partition 2 (Table 2.1.16).

Table 2.1.15: Results of the model selection for partitioning scheme 4-PartitionsD estimated by ModelFinder. Log-Likelihood (LogL) and degrees of freedom (df) were used to calculate the accumulative likelihood of the best models under Akaike Information Criterion (AIC), AIC corrected for small sample sizes (AICc) and Bayesian Information Criterion (BIC) (w-AIC, w-AICc and w-BIC respectively).

Partition	Best-fitting model (AIC/AICc)	LogL	df	w-AICc	w-AIC	w-BIC
Partition 1	GTR+F+R10	-73146.6174	27	0.7958	0.7996	$3.5258x10^{-11}$
Partition 2	GTR+F+R3	-15699.2245	13	0.6984	0.6998	0.0132
Partition 3	GTR+F+I+G4	-11640.8188	11	0.6725	0.6661	0.9938
Partition 4	GTR+F+R4	-33299.7514	15	0.8770	0.8734	0.8439

Table 2.1.16: Comparison of model selection for partitioning scheme 4-PartitionsD estimated, by ModelFinder, under AIC/AICc criteria and BIC criterion.

Partition	Best-fitting model (AIC/AICc)	Best-fitting model (BIC)	Best-fitting model accumulative likelihood (AIC/AICc)	Best-fitting model accumulative likelihood (BIC)
Partition 1	GTR+F+R10	TIM2+F+R4	0.7996/0.7958	0.5961
Partition 2	GTR+F+R3	GTR+F+I+G4	0.6998/0.6984	0.9868
Partition 3	GTR+F+I+G4	GTR+F+I+G4	0.6661/0.6984	0.9938
Partition 4	GTR+F+R4	GTR+F+R4	0.8734/0.8770	0.8439

6-PartitionsA.

According to AICc, the best-fitting substitution models estimated by ModelFinder were TIM2+F+R9 for the first partition, GTR+F+R4 for Partitions 2, 3 and 4, GTR+F+R3 for the fifth partition and GTR+F+I+G4 for the sixth partition (Table 2.1.17). The results were also supported by AIC. However, the support for the best-fitting substitution models for Partitions 1, 5 and 6 was weak for both AIC and AICc (w-$AIC_{Partition1}$=0.1635, w-$AICc_{Partition1}$=0.1583; w-$AIC_{Partition5}$=0.3701, w-$AICc_{Partition5}$=0.3617; w-$AIC_{Partition6}$=0.5002, w-$AICc_{Partition5}$=0.5079). On the other hand, BIC did not supported any of these substitution models for any of the partitions. Instead, the best-fitting substitution models, according to BIC, were TN+F+R6 for the first partition, TIM2+F+R4 for the second partition, GTR+F+I+G4 for the third partition, TIM+F+R4 for the fourth partition, TN+F+I+G4 for the fifth partition and TIM3+F+I+G4 for the sixth partition (Table 2.1.18). Only when the best-fitting substitution model was

weakly supported (Partition 1, 5 and 6) the best-fitting substitution model's w-BIC was higher than w-AIC/w-AICc (Table 2.1.18).

Table 2.1.17: Results of the model selection for partitioning scheme 6-PartitionsA estimated by ModelFinder. Log-Likelihood (LogL) and degrees of freedom (df) were used to calculate the accumulative likelihood of the best models under Akaike Information Criterion (AIC), AIC corrected for small sample sizes (AICc) and Bayesian Information Criterion (BIC) (w-AIC, w-AICc and w-BIC respectively).

Partition	Best-fitting model (AIC/AICc)	LogL	df	w-AICc	w-AIC	w-BIC
Partition 1	TIM2+F+R9	-22589.5658	23	0.1583	0.1635	$1.0557x10^{-7}$
Partition 2	GTR+F+R4	-50148.7866	15	0.7949	0.7916	0.4927
Partition 3	GTR+F+R4	-24560.1786	15	0.8901	0.8794	0.3459
Partition 4	GTR+F+R4	-11119.6064	15	0.8252	0.8186	0.1131
Partition 5	GTR+F+R3	-15297.1865	13	0.3617	0.3701	$6.4512x10^{-5}$
Partition 6	GTR+F+I+G4	-11987.3408	11	0.5079	0.5002	0.0485

Table 2.1.18: Comparison of model selection for partitioning scheme 6-PartitionsA estimated, by ModelFinder, under AIC/AICc criteria and BIC criterion.

Partition	Best-fitting model (AIC/AICc)	Best-fitting model (BIC)	Best-fitting model accumulative likelihood (AIC/AICc)	Best-fitting model accumulative likelihood (BIC)
Partition 1	TIM2+F+R9	TN+F+R6	0.1635/0.1583	0.3558
Partition 2	GTR+F+R4	TIM2+F+R4	0.7916/0.7949	0.5068
Partition 3	GTR+F+R4	GTR+F+I+G4	0.8794/0.8901	0.6480
Partition 4	GTR+F+R4	TIM+F+R4	0.8186/0.8252	0.3870
Partition 5	GTR+F+R3	TN+F+I+G4	0.3701/0.3617	0.4933
Partition 6	GTR+F+I+G4	TIM3+F+I+G4	0.5002/0.5079	0.9312

6-PartitionsB.

According to AICc, the best-fitting substitution models estimated by ModelFinder were TIM3+F+R10 for the first partition, GTR+F+R5 for the second partition, LG+R4 for the third partition,

mtART+F+R3 for the fourth partition, LG+F+I+G4 for Partitions 5 and 6 (Table 2.1.19). The results were also supported by AIC. However, the support for the best-fitting substitution models for Partitions 1 and 6 was weak for both AIC and AICc (w-$AIC_{Partition1}$=0.2789, w-$AICc_{Partition1}$=0.2705; w-$AIC_{Partition6}$=0.5778, w-$AICc_{Partition5}$=0.6374). On the other hand, BIC did not supported any of these substitution models for any of the partitions. Instead, the best-fitting substitution models, according to BIC, were TN+F+R6 for the first partition, TIM2+F+R5 for the second partition, LG+I+G4 for the third partition, LG+R3 for the fourth partition and WAG+I+G4 for Partitions 5 and 6 (Table 2.1.20). For Partitions 4 and 6, for the best-fitting substitution models, w-AIC/w-AICc was higher than w-BIC, while the opposite was true for the remaining partitions (Table 2.1.20).

Table 2.1.19: Results of the model selection for partitioning scheme 6-PartitionsB estimated by ModelFinder. Log-Likelihood (LogL) and degrees of freedom (df) were used to calculate the accumulative likelihood of the best models under Akaike Information Criterion (AIC), AIC corrected for small sample sizes (AICc) and Bayesian Information Criterion (BIC) (w-AIC, w-AICc and w-BIC respectively).

Partition	Best-fitting model (AIC/AICc)	LogL	df	w-AICc	w-AIC	w-BIC
Partition 1	TIM3+F+R10	-22215.6906	25	0.2705	0.2789	$4.7823x10^{-8}$
Partition 2	GTR+F+R5	-49454.0054	17	0.8671	0.8650	0.1196
Partition 3	LG+R4	-9501.7193	7	0.8766	0.8709	0.0731
Partition 4	mtART+F+R3	-2495.1597	24	0.9024	0.8781	$6.7128x10^{-12}$
Partition 5	LG+F+I+G4	-8052.7911	22	0.8222	0.8260	$5.2839x10^{-11}$
Partition 6	LG+F+I+G4	-5107.1249	22	0.6374	0.5778	$1.7256x10^{-8}$

Table 2.1.20: Comparison of model selection for partitioning scheme 6-PartitionsB estimated, by ModelFinder, under AIC/AICc criteria and BIC criterion.

Partition	Best-fitting model (AIC/AICc)	Best-fitting model (BIC)	Best-fitting model accumulative likelihood (AIC/AICc)	Best-fitting model accumulative likelihood (BIC)
Partition 1	TIM3+F+R10	TN+F+R6	0.2789/0.2705	0.4897
Partition 2	GTR+F+R5	TIM2+F+R5	0.8650/0.8671	0.8798
Partition 3	LG+R4	LG+I+G4	0.8709/0.8766	0.8957
Partition 4	mtART+F+R3	LG+R3	0.8781/0.9024	0.8374

Partition	Best-fitting model (AIC/AICc)	Best-fitting model (BIC)	Best-fitting model accumulative likelihood (AIC/AICc)	Best-fitting model accumulative likelihood (BIC)
Partition 5	LG+F+I+G4	WAG+I+G4	0.8260/0.8222	0.8905
Partition 6	LG+F+I+G4	WAG+I+G4	0.5778/0.6374	0.5765

6-PartitionsC.

According to AICc, the best-fitting substitution models estimated by ModelFinder were TIM3+F+R9 for the first partition, GTR+F+R5 for the second partition, SCHN05+F+G4 for the third partition, KOSI07+F+R3 for the fourth partition, KOSI07+F+R6 for the fifth partitions and KOSI07+F+G4 for the sixth partition (Table 2.1.21). The results were also supported by AIC, with the exemption of partition 1, where the best-fitting substitution model according to AIC was GTR+F+R10. However, the support for the best-fitting substitution models for Partitions 1 and 2 was weak for both AIC and AICc ($w\text{-}AIC_{Partition1}$=0.1928, $w\text{-}AICc_{Partition1}$=0.1899; $w\text{-}AIC_{Partition2}$=0.6441, $w\text{-}AICc_{Partition2}$=0.6492). These results were supported by BIC for Partitions 3-6, while the best-fitting substitution models, according to this criterion, were TN+F+R7 and TIM2+F+R5 for the Partitions 1 and 2 respectively (Table 2.1.22). w-AIC/w-AICc of the best-fitting substitution model was higher than w-BIC for Partitions 4 and 6, while the opposite was true for the remaining partitions (Table 2.1.22).

Table 2.1.21: Results of the model selection for partitioning scheme 6-PartitionsC estimated by ModelFinder. Log-Likelihood (LogL) and degrees of freedom (df) were used to calculate the accumulative likelihood of the best models under Akaike Information Criterion (AIC), AIC corrected for small sample sizes (AICc) and Bayesian Information Criterion (BIC) (w-AIC, w-AICc and w-BIC respectively). (*) Indicates the model is not supported as the best-fitting model by AIC.

Partition	Best-fitting model (AIC/AICc)	LogL	df	w-AICc	w-AIC	w-BIC
Partition 1	TIM3+F+R9*	-22438.4562	23	0.1899	0.1840	$8.0067x10^{-6}$
Partition 2	GTR+F+R5	-49913.6963	17	0.6492	0.6441	0.2004
Partition 3	SCHN05+F+G4	-20726.7491	62	0.7728	0.7305	0.9615
Partition 4	KOSI07+F+R3	-9498.7305	65	0.8967	0.8014	0.9959
Partition 5	KOSI07+F+R6	-13696.6305	71	0.9425	0.8725	0.9517

Partition	Best-fitting model (AIC/AICc)	LogL	df	w-AICc	w-AIC	w-BIC
Partition 6	KOSI07+F+G4	-10083.9560	62	0.7919	0.7300	0.9553

Table 2.1.22: Comparison of model selection for partitioning scheme 6-PartitionsC estimated, by ModelFinder, under AIC/AICc criteria and BIC criterion. (*) Indicates the model is not supported as the best-fitting model by AIC.

Partition	Best-fitting model (AIC/AICc)	Best-fitting model (BIC)	Best-fitting model accumulative likelihood (AIC/AICc)	Best-fitting model accumulative likelihood (BIC)
Partition 1	TIM3+F+R9*	TN+F+R7	0.1840/0.1899	0.9131
Partition 2	GTR+F+R5	TIM2+F+R5	0.6441/0.6492	0.7982
Partition 3	SCHN05+F+G4	SCHN05+F+G4	0.7305/0.7728	0.9615
Partition 4	KOSI07+F+R3	KOSI07+F+R3	0.8014/0.8967	0.9959
Partition 5	KOSI07+F+R6	KOSI07+F+R6	0.8725/0.9425	0.9517
Partition 6	KOSI07+F+G4	KOSI07+F+G4	0.7300/0.7919	0.9553

Phylogenetic tree reconstruction.

The monophyly of choanoflagellates and its major groups Acanthoecida and Craspedida.

Consistently with previously published phylogenies (Carr *et al.*, 2017; Schiwitza *et al.*, 2018; Schiwitza *et al.*, 2019), and regardless of the partitioning scheme used, the ML phylogenetic reconstructions performed recovered choanoflagellates as a monophyletic group with a sister relationship to metazoans (100% Maximum-likelihood Bootstrap Support (mlBP)) (Figures 2.1.1-2.1.11).

Similarly, the major group within choanoflagellates Acanthoecida was always recovered as monophyletic (100% mlBP). The second major group, Craspedida, however, was only recovered as monophyletic when codon substitution models were used for the partitions containing protein coding genes, regardless of the partitioning scheme used (Figures 2.1.8-2.1.10). All reconstructions based on partitioning schemes that used either nucleotides or amino acids substitution models for protein coding genes recovered *Salpingoeca prava* as basal to Acanthoecida, thus recovering Craspedida as

paraphyletic (Figures 2.2-2.7 and 2.11). Furthermore, phylogenies based on partitioning schemes 4-PartitionsA and 6-PartitionsA recovered the main clade of marine craspedids (Clade 1) as a sister clade to the clade formed by the acanthoecids and *S. prava* (Figures 2.1.3 and 2.1.4).

The extended majority-rule consensus tree recovered *S. prava* as basal to Acanthoecida (70% support). Similarly, the clade formed by *S. dolichothecata* and *S. tuba* were recovered as basal to all the remaining craspedids (70% support). Freshwater craspedids were recovered in a single clade (50% support) further subdivided in two clades, one containing all freshwater codosigids (100% support) and the other containing all freshwater salpingoecids and the marine salpingoecids *S. kvevrii*, *S. macrocollata* and *S. urceolata* (40% support) (Figure 2.1.1).

The relationship between nudiforms and tectiforms.

Only reconstructions based on partitioning schemes 2-PartitionsA, 4-PartitionsA and 6-PartitionA supported a sister relationship between nudiforms and tectiforms, while the rest of the reconstructions supported the placement of the root of the nudiforms within the tectiforms (Figures 2.1.2-2.1.4).

The relationship within Clade 2 craspedis.

Reconstructions based on partitioning schemes 2-PartitionsB, 4-PartitionsB and 6-PartitionsB (Figures 2.1.5-2.1.7) supported a sister relationship between the freshwater codosigids and the clade formed by *Salpingoeca dolichothecata* and *Salpingoeca tuba* (Clade 3) (71% mlBP, 55% mlBP and 54% mlBP, for reconstructions based on partitioning schemes 2-PartitionsB, 4-PartitionsB and 6-PartitionsB respectively), while the remaining reconstructions recovered them as a part of Clade 2 craspedids, alongside the remaining freshwater craspedids and the marine species *Salpingoeca kvevrii*, *Salpingoeca macrocollata* an *Salpingoeca urceolata*. Consistently with previous studies (Carr *et al.*, 2017; Schiwitza *et al.,* 2018), all reconstructions recovered the marine choanoflagellates *S. macrocollata* as sister to the fresh-water species *Salpingoeca punica* (Figures 2.1.1-2.1.11). The marine choanoflagellates *S. kvevrii* and *S. urceolata* were recovered within the fresh-water craspedids on reconstructions based on partitioning schemes 4-PartitionsA, 4-PartitionsC, 6-PartitionsA and 6-

PartitionsC (Figures 2.1.3, 2.1.4, 2.1.8 and 2.1.9). Additionally, the reconstruction based on partitioning scheme 4-PartitionsA recovered the clade containing the hypersaline species *Salpingoeca crinita*, *Salpingoeca huasca* and *Salpingoeca surira* as sister to the fresh-water codosigids (Figure 2.1.3).

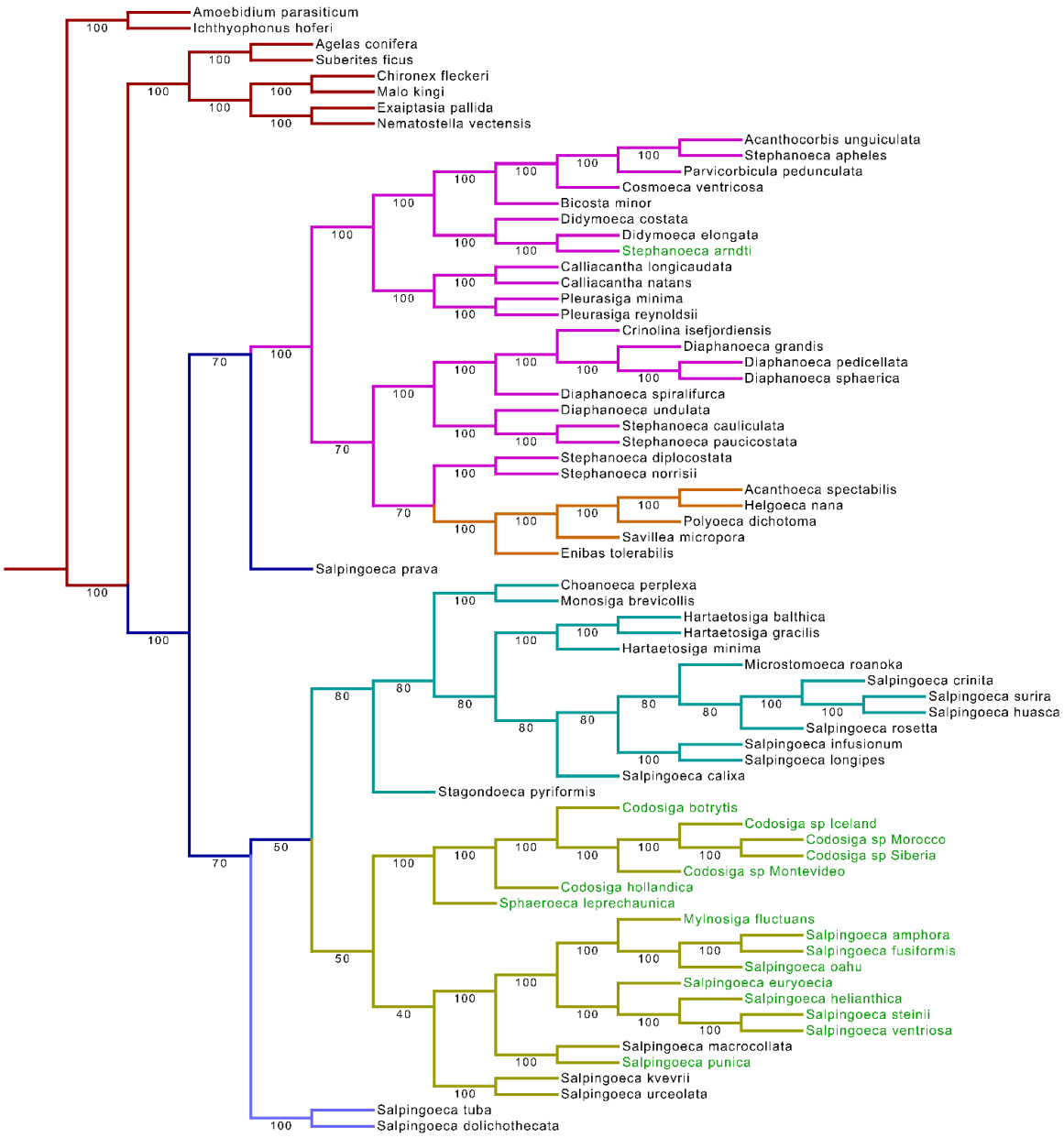

Figure 2.1.1: Extended majority-rule consensus tree of choanoflagellates, obtained from 10 independent phylogenetic reconstructions. Branches are colour-coded indicating the different groups of species as follows: outgroup (red), basal choanoflagellates (blue), tectiforms (magenta), nudiforms (orange), Clade 1 craspedids (cyan), Clade 2 craspedids (yellow), Clade 3 craspedids (light blue). Species names in green indicate freshwater species. Support values, calculated by percentage of the trees supporting such topology, are indicated below each branch.

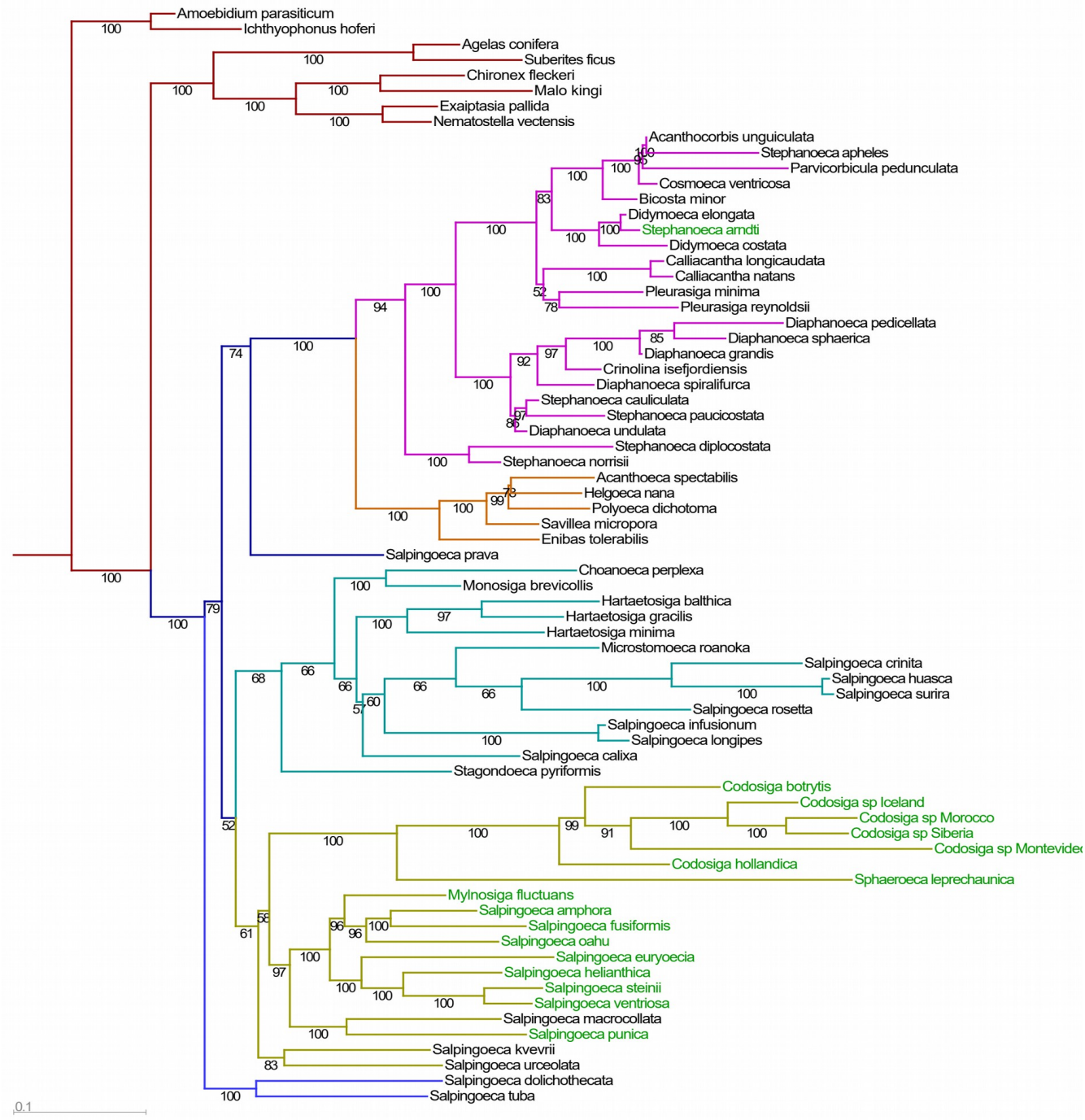

Figure 2.1.2: Phylogeny of choanoflagellates based on the partitioning scheme 2-PartitionsA. Branches are colour-coded indicating the different groups of species as follows: outgroup (red), basal choanoflagellates (blue), tectiforms (magenta), nudiforms (orange), Clade 1 craspedids (cyan), Clade 2 craspedids (yellow), Clade 3 craspedids (light blue). Species names in green indicate freshwater species. Branch lengths are drawn proportional to the number of substitutions per site as indicated by the scale. Ultrafast mlBP support values are indicated below each branch.

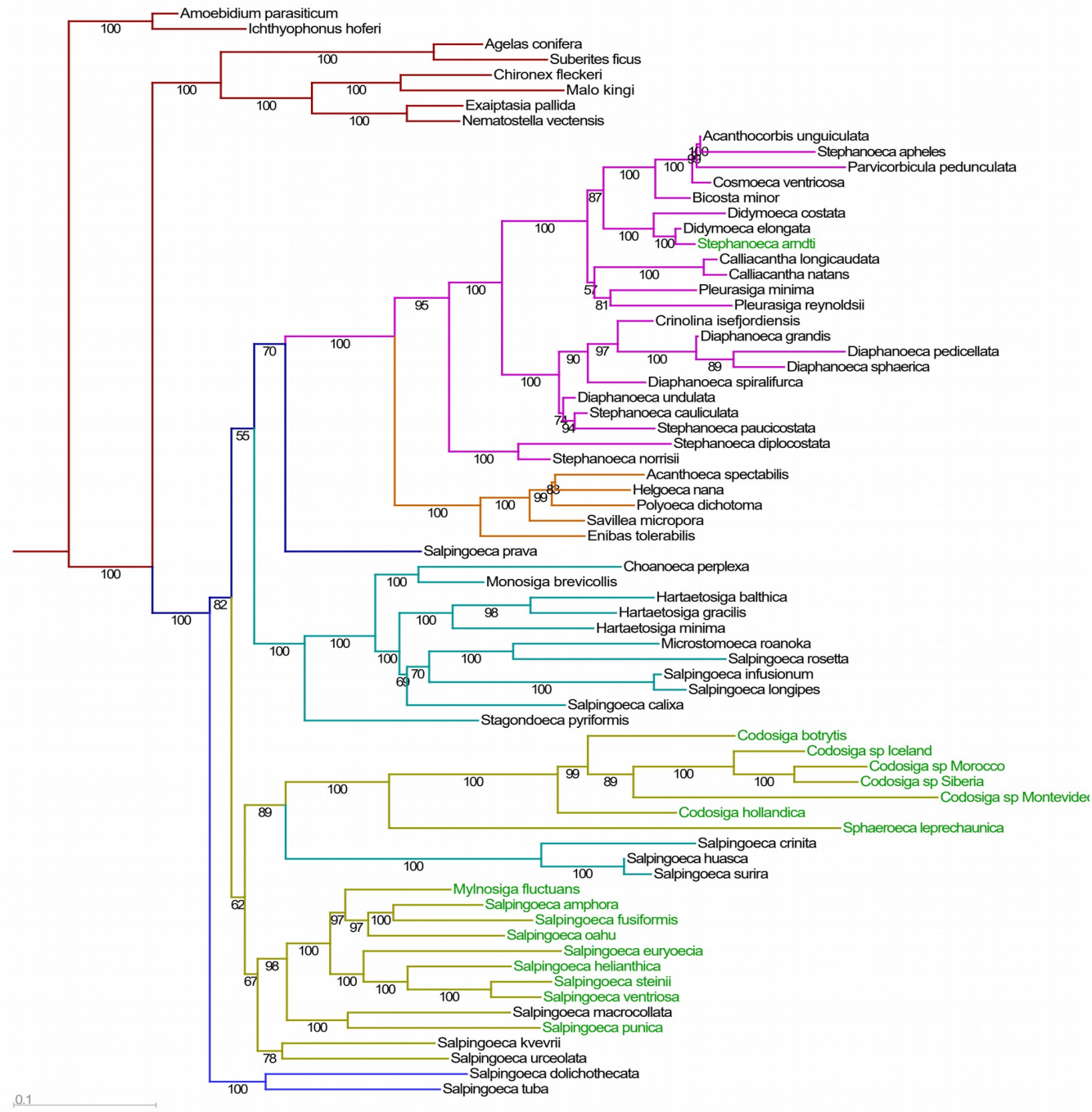

Figure 2.1.3: Phylogeny of choanoflagellates based on the partitioning scheme 4-PartitionsA. Branches are colour-coded indicating the different groups of species as follows: outgroup (red), basal choanoflagellates (blue), tectiforms (magenta), nudiforms (orange), Clade 1 craspedids (cyan), Clade 2 craspedids (yellow), Clade 3 craspedids (light blue). Species names in green indicate freshwater species. Branch lengths are drawn proportional to the number of substitutions per site as indicated by the scale. Ultrafast mlBP support values are indicated below each branch.

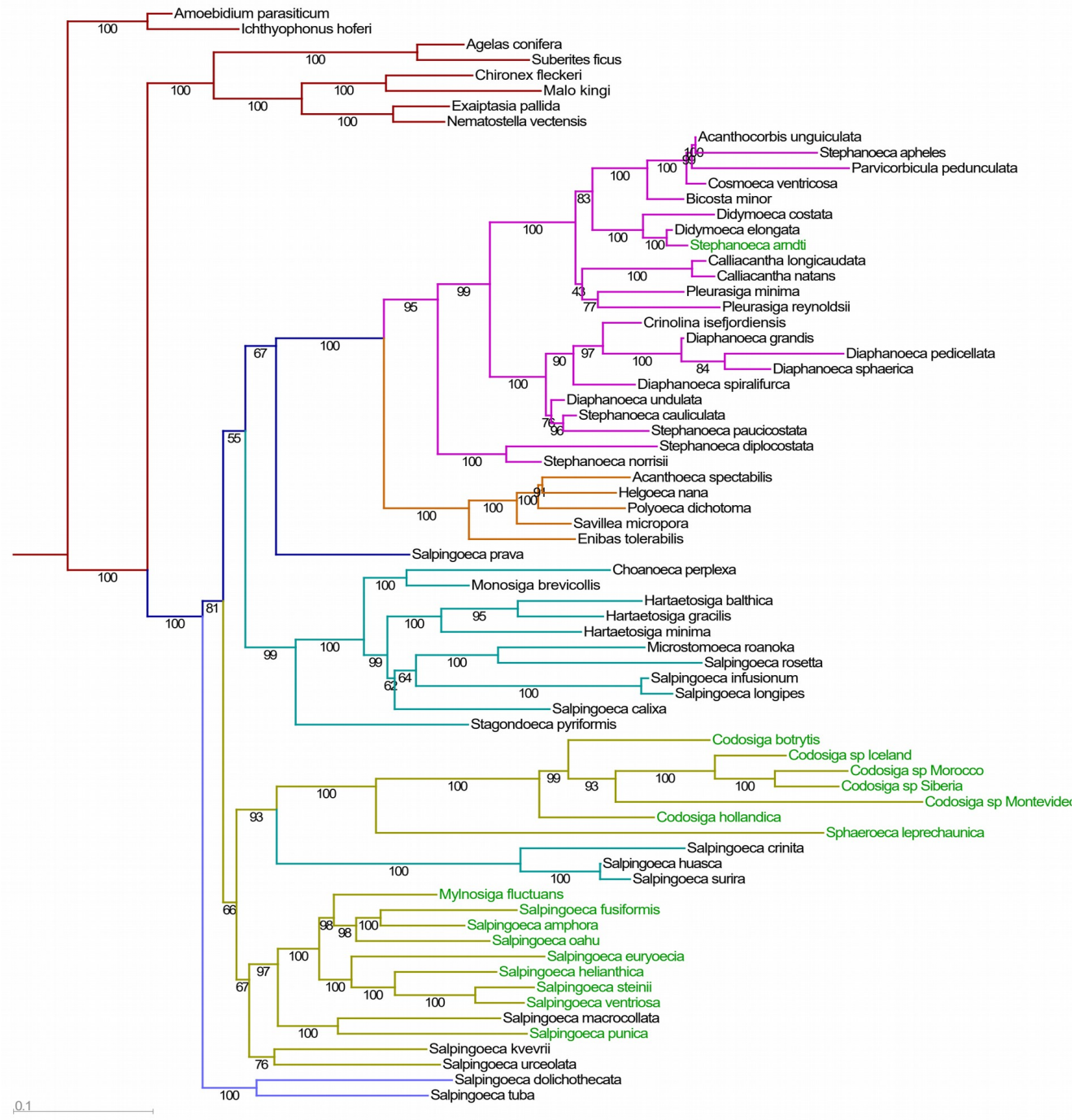

Figure 2.1.4: Phylogeny of choanoflagellates based on the partitioning scheme 6-PartitionsA. Branches are colour-coded indicating the different groups of species as follows: outgroup (red), basal choanoflagellates (blue), tectiforms (magenta), nudiforms (orange), Clade 1 craspedids (cyan), Clade 2 craspedids (yellow), Clade 3 craspedids (light blue). Species names in green indicate freshwater species. Branch lengths are drawn proportional to the number of substitutions per site as indicated by the scale. Ultrafast mlBP support values are indicated below each branch.

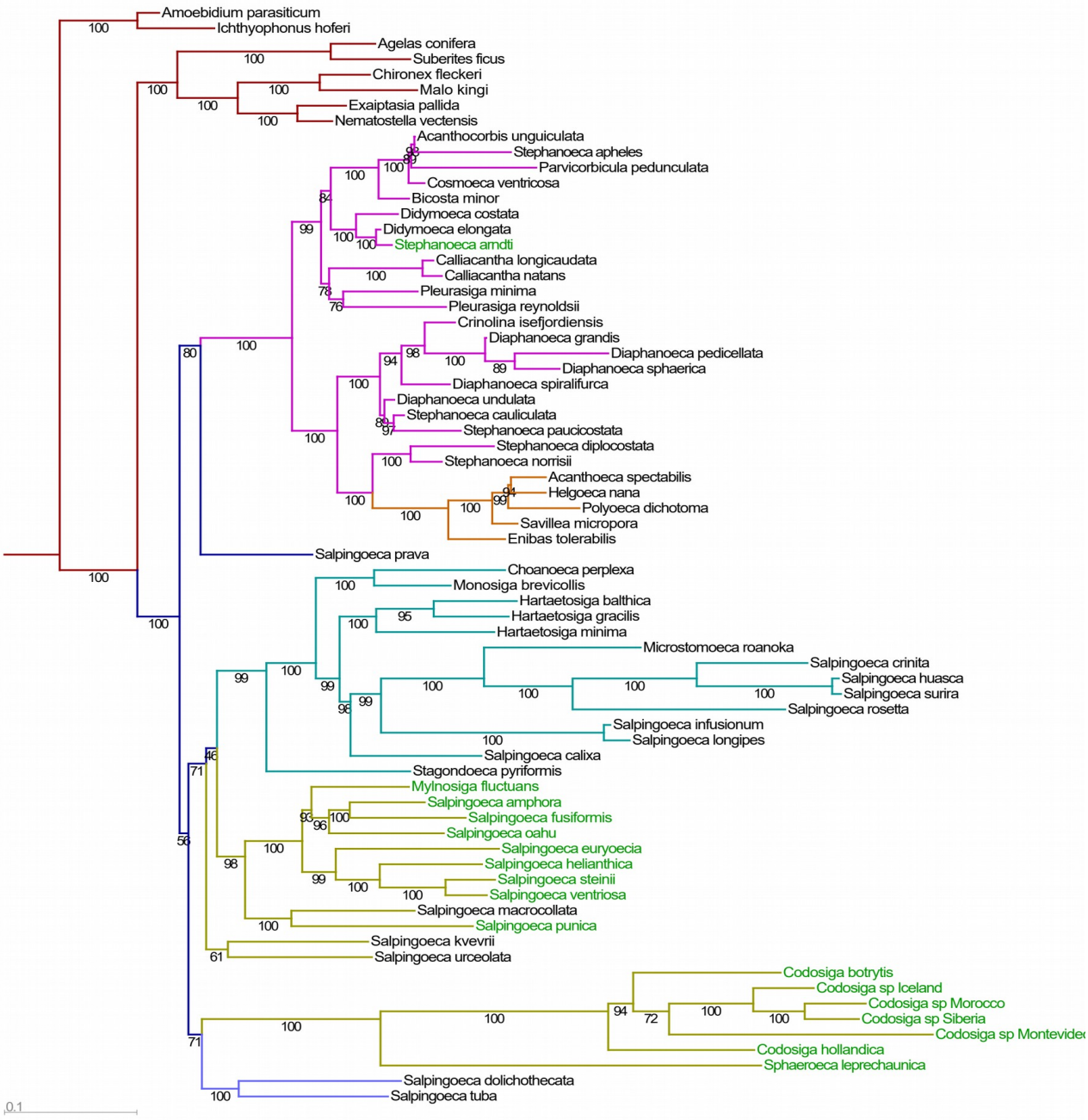

Figure 2.1.5: Phylogeny of choanoflagellates based on the partitioning scheme 2-PartitionsB. Branches are colour-coded indicating the different groups of species as follows: outgroup (red), basal choanoflagellates (blue), tectiforms (magenta), nudiforms (orange), Clade 1 craspedids (cyan), Clade 2 craspedids (yellow), Clade 3 craspedids (light blue). Species names in green indicate freshwater species. Branch lengths are drawn proportional to the number of substitutions per site as indicated by the scale. Ultrafast mlBP support values are indicated below each branch.

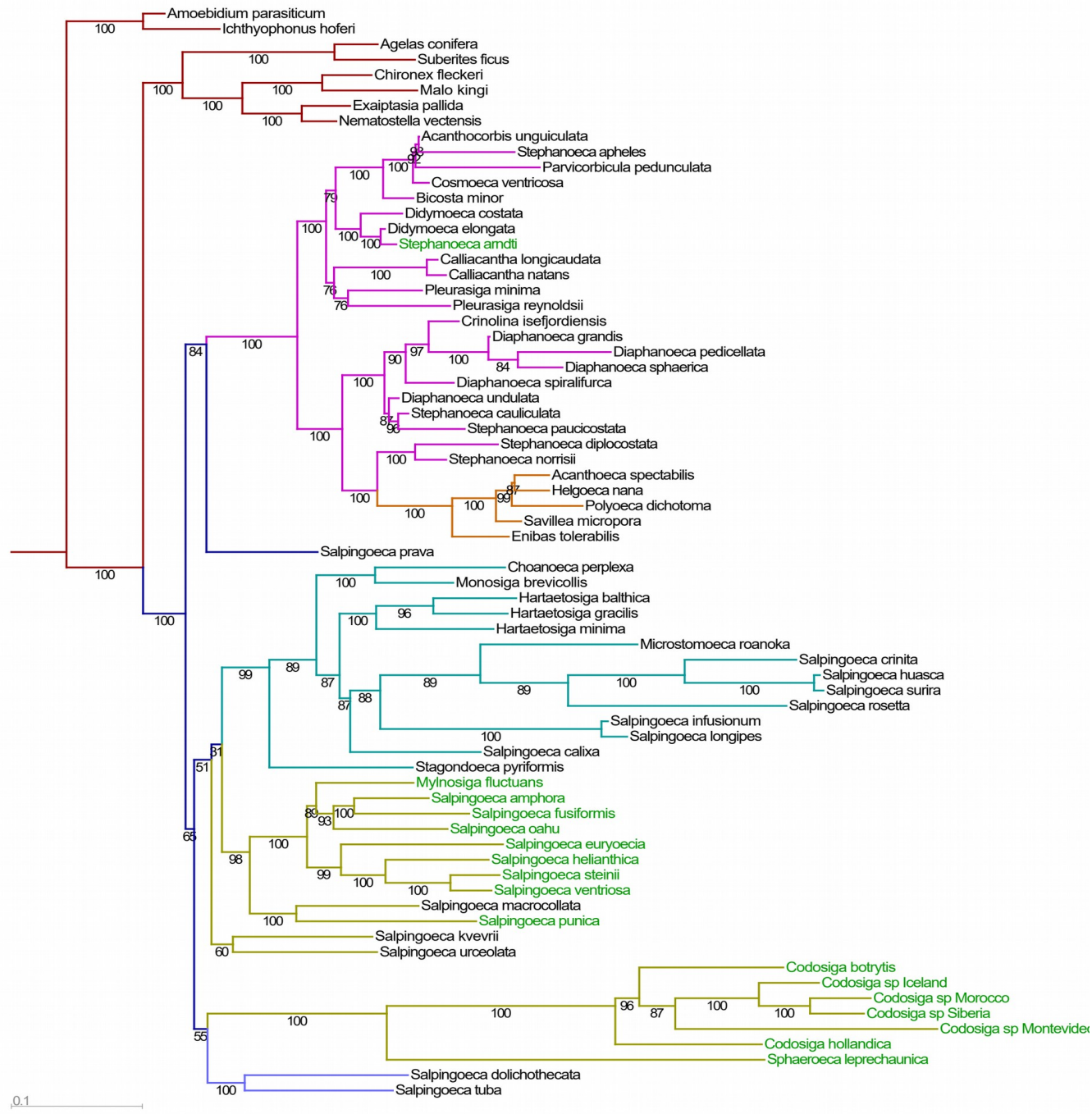

Figure 2.1.6: Phylogeny of choanoflagellates based on the partitioning scheme 4-PartitionsB. Branches are colour-coded indicating the different groups of species as follows: outgroup (red), basal choanoflagellates (blue), tectiforms (magenta), nudiforms (orange), Clade 1 craspedids (cyan), Clade 2 craspedids (yellow), Clade 3 craspedids (light blue). Species names in green indicate freshwater species. Branch lengths are drawn proportional to the number of substitutions per site as indicated by the scale. Ultrafast mlBP support values are indicated below each branch.

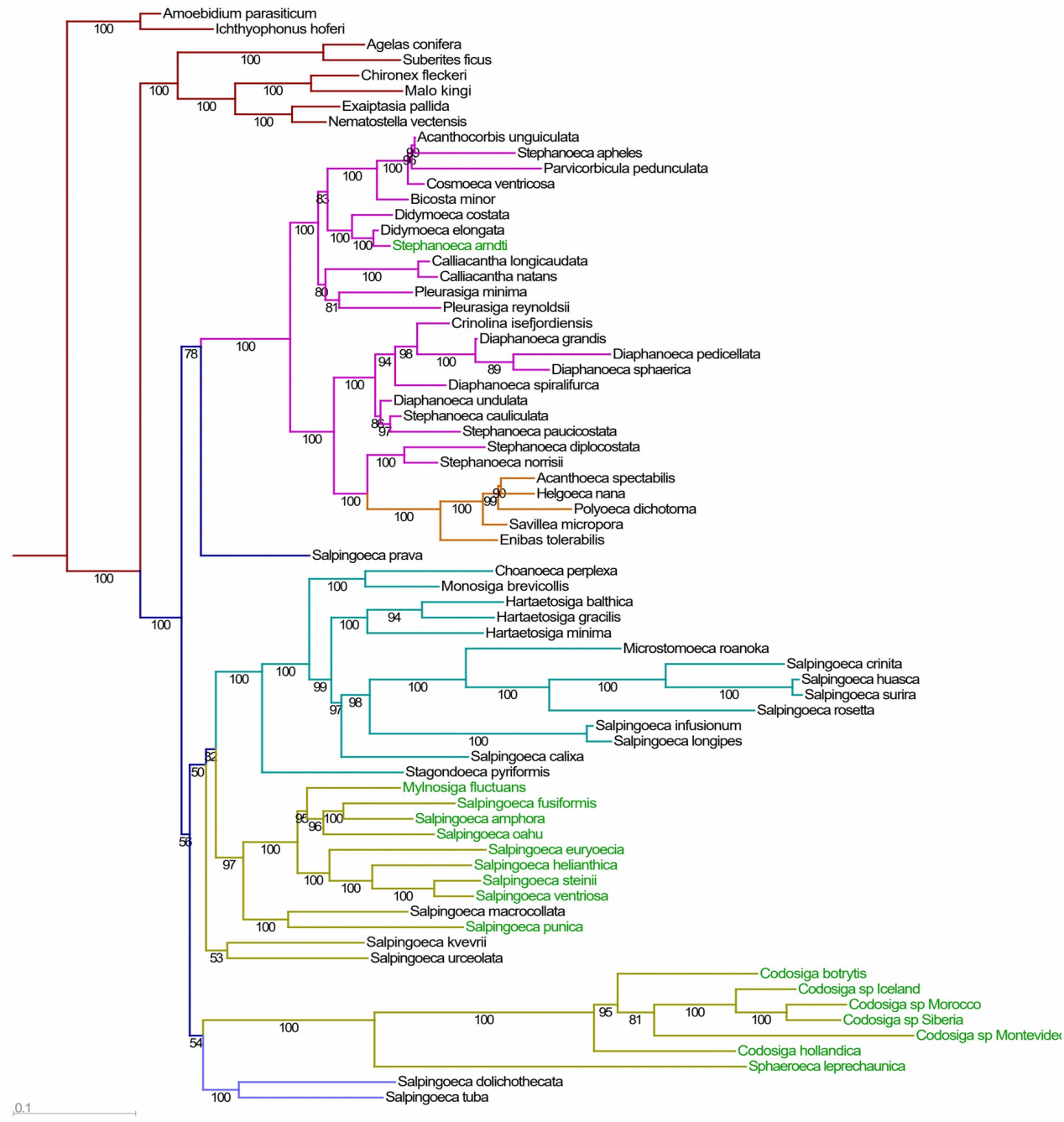

Figure 2.1.7: Phylogeny of choanoflagellates based on the partitioning scheme 6-PartitionsB. Branches are colour-coded indicating the different groups of species as follows: outgroup (red), basal choanoflagellates (blue), tectiforms (magenta), nudiforms (orange), Clade 1 craspedids (cyan), Clade 2 craspedids (yellow), Clade 3 craspedids (light blue). Species names in green indicate freshwater species. Branch lengths are drawn proportional to the number of substitutions per site as indicated by the scale. Ultrafast mlBP support values are indicated below each branch.

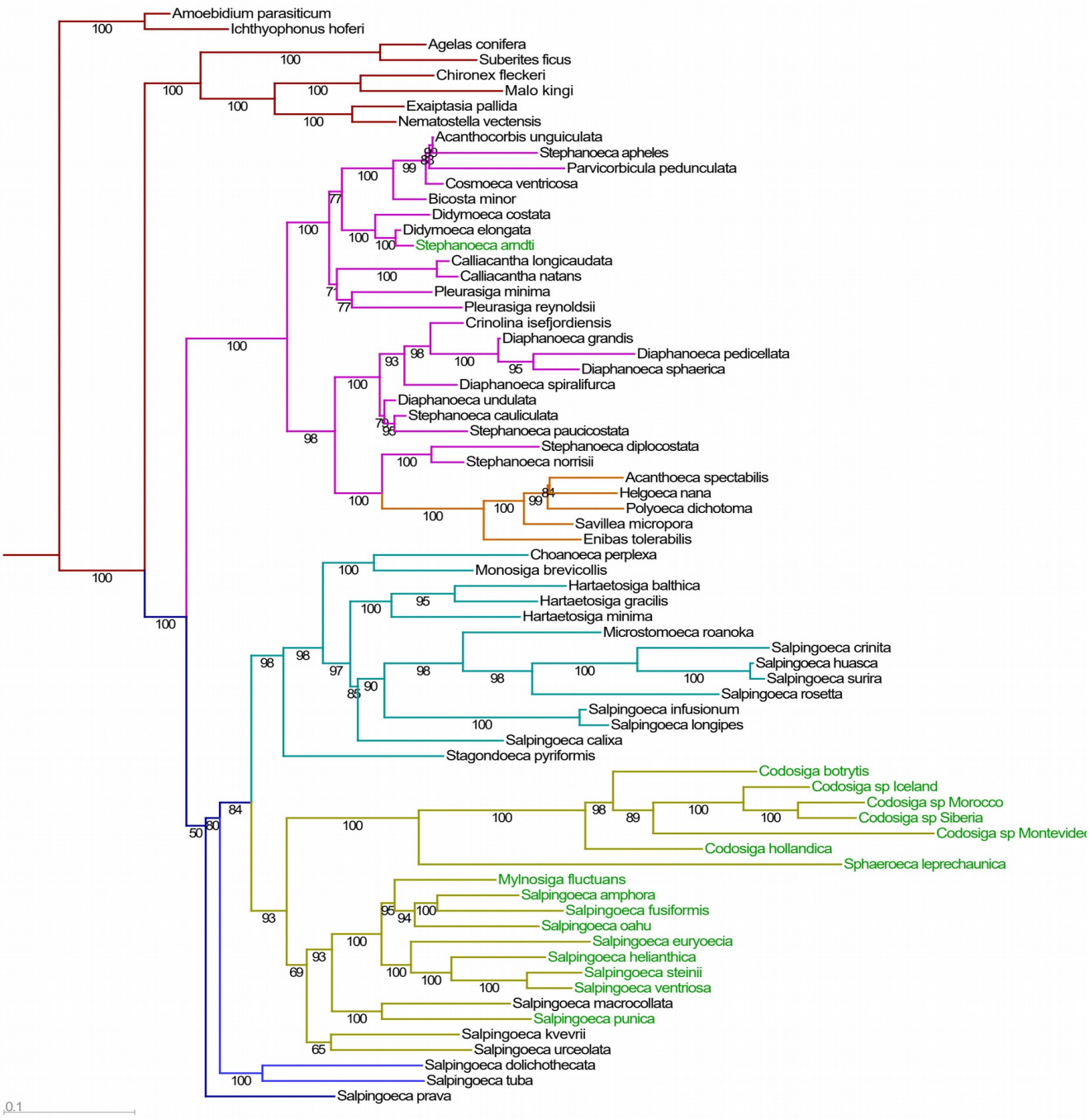

Figure 2.1.8: Phylogeny of choanoflagellates based on the partitioning scheme 4-PartitionsC. Branches are colour-coded indicating the different groups of species as follows: outgroup (red), basal choanoflagellates (blue), tectiforms (magenta), nudiforms (orange), Clade 1 craspedids (cyan), Clade 2 craspedids (yellow), Clade 3 craspedids (light blue). Species names in green indicate freshwater species. Branch lengths are drawn proportional to the number of substitutions per site as indicated by the scale. Ultrafast mlBP support values are indicated below each branch.

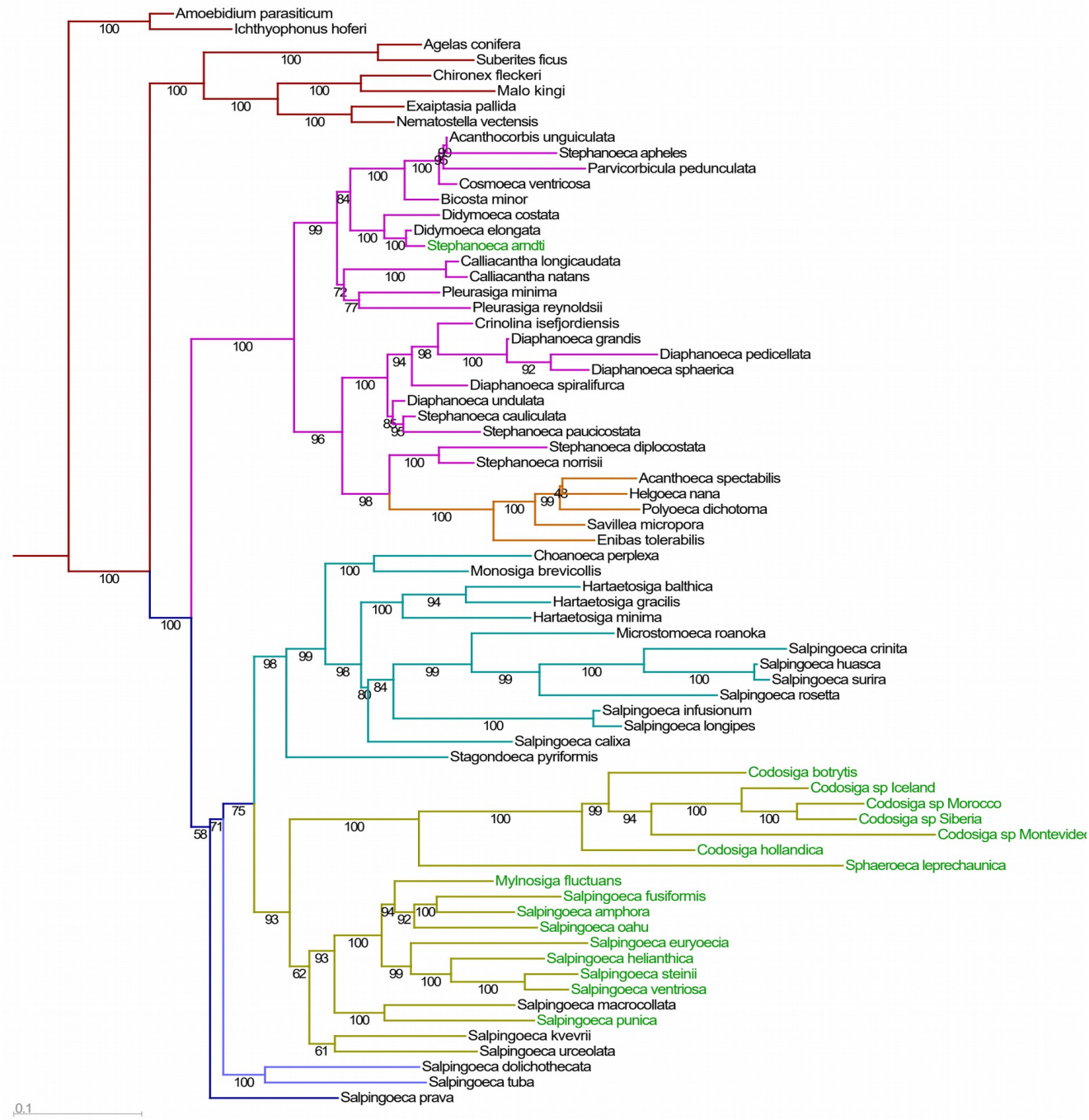

Figure 2.1.9: Phylogeny of choanoflagellates based on the partitioning scheme 6-PartitionsC. Branches are colour-coded indicating the different groups of species as follows: outgroup (red), basal choanoflagellates (blue), tectiforms (magenta), nudiforms (orange), Clade 1 craspedids (cyan), Clade 2 craspedids (yellow), Clade 3 craspedids (light blue). Species names in green indicate freshwater species. Branch lengths are drawn proportional to the number of substitutions per site as indicated by the scale. Ultrafast mlBP support values are indicated below each branch.

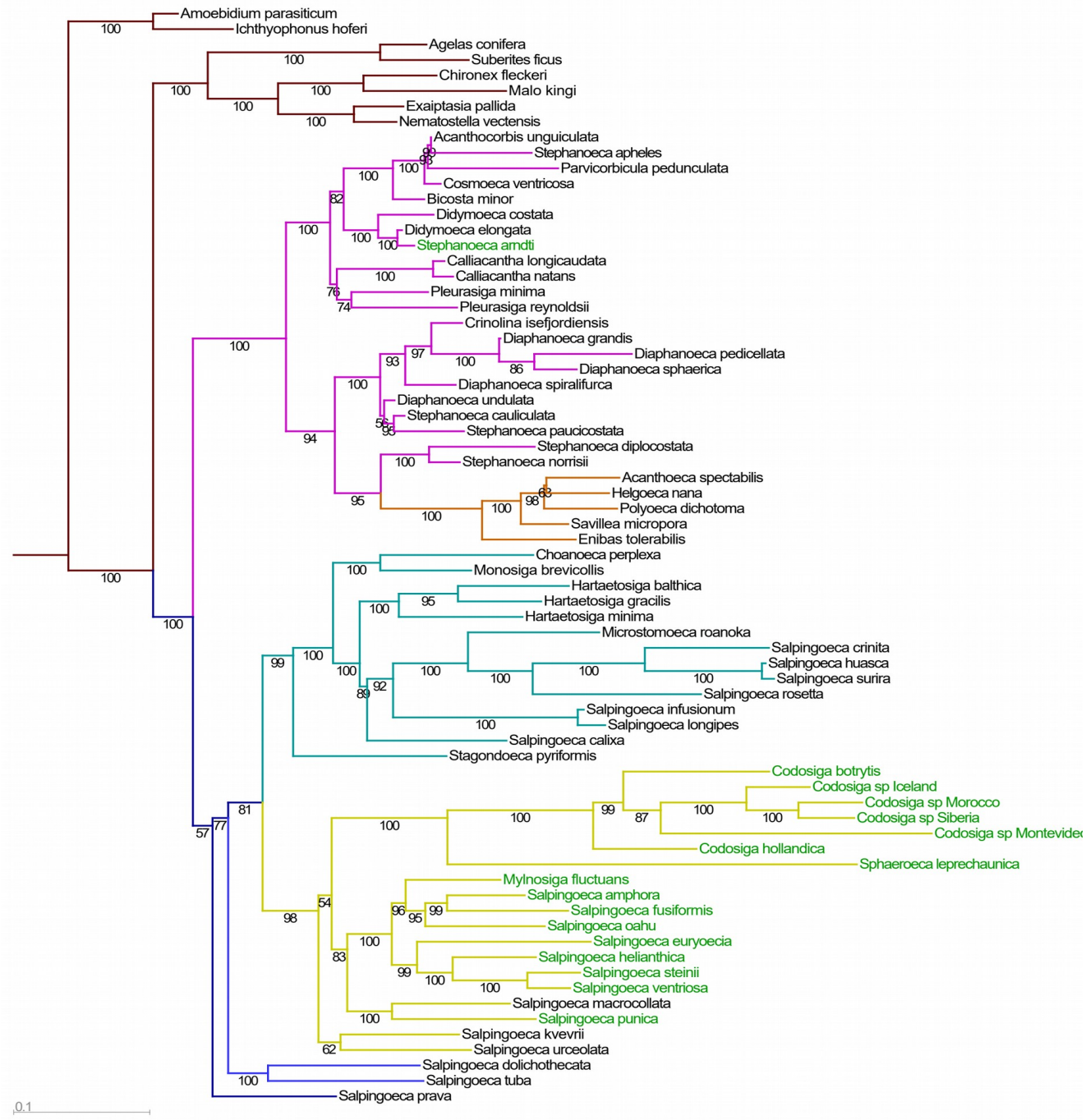

Figure 2.1.10: Phylogeny of choanoflagellates based on the partitioning scheme 2-PartitionsC. Branches are colour-coded indicating the different groups of species as follows: outgroup (red), basal choanoflagellates (blue), tectiforms (magenta), nudiforms (orange), Clade 1 craspedids (cyan), Clade 2 craspedids (yellow), Clade 3 craspedids (light blue). Species names in green indicate freshwater species. Branch lengths are drawn proportional to the number of substitutions per site as indicated by the scale. Ultrafast mlBP support values are indicated below each branch.

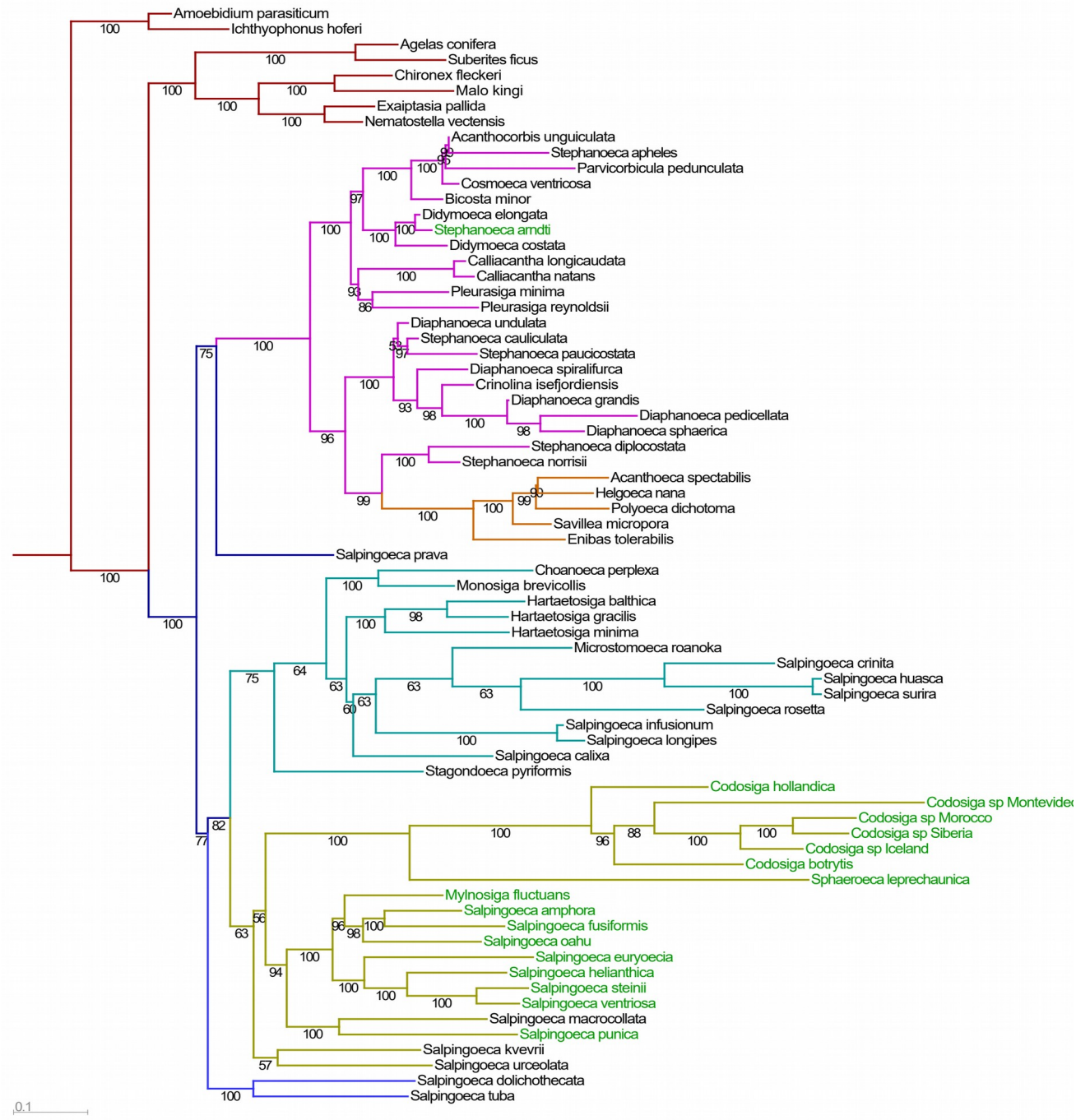

Figure 2.1.11: Phylogeny of choanoflagellates based on the partitioning scheme 4-PartitionsD. Branches are colour-coded indicating the different groups of species as follows: outgroup (red), basal choanoflagellates (blue), tectiforms (magenta), nudiforms (orange), Clade 1 craspedids (cyan), Clade 2 craspedids (yellow), Clade 3 craspedids (light blue). Species names in green indicate freshwater species. Branch lengths are drawn proportional to the number of substitutions per site as indicated by the scale. Ultrafast mlBP support values are indicated below each branch.

Discussion.

Phylogenetic reconstruction.

Reference alignment.

When attempting to align the ribosomal subunits subsets, none of the software tested was able to produce high quality alignments independently (data not shown). This was especially evident for SSU, due to the presence of a higher number of hypervariable regions in the sequences. Such hypervariable regions were poorly aligned in every case, which caused the alignments to contain a high number of mismatches in those regions and the flanking regions.

Poorly aligned regions are a source of noise for downstream analysis, since they introduce sequence dissimilarity that is not the result of evolutionary processes but of computational artefacts (Lake, 1991; Morrison and Ellis, 1997; Castresana, 2000). A common practice to minimise the noise or even eliminate the noise introduced by these regions in downstream analyses is to manually curate the alignments and subsequently trim these regions out from the alignment prior to the phylogenetic reconstruction. Nevertheless, this practice should be avoided for two main reasons. Firstly, this practice reduces the reproducibility of the analyses, since manual curation is performed at the researcher's discretion (Morrison and Ellis, 1997; Castresana 2000). Secondly, hypervariable regions in general, and SSU and LSU hypervariable regions specifically, are of especial importance for phylogenetic analysis, since there are regions subjected to adaptation and they contain very strong phylogenetic signals that are directly related to the organisms evolutionary history. Particularly, the phylogenetic signal of hypervariable regions of the SSU, such as v2, v4, v5, v7 and v9, is so strong that they are widely used as primary target regions for biodiversity characterization analyses of eDNA samples (Ribera *et al.*, 2002; Bass *et al.*, 2007; Ki, 2011; Jo *et al.*, 2019; Cordier *et al.*, 2019; Apothéloz Perret Gentil *et al.*, 2020; Cowart *et al.*, 2020). Therefore, when trimming these regions out, valuable phylogenetic data is lost in the process. Since, as previously shown, the dataset used for the phylogenetic reconstruction

contained 42 species for which only ribosomal sequences are available, retaining as much of the phylogenetic information contained in SSU and LSU sequences was deemed of maximum importance.

The two-steps alignment method, on the other hand, produced high quality alignments throughout, reducing the number of mismatches in both hypervariable and flanking regions. When comparing both strategies, the two-steps alignment method produced longer alignments with substantially less poorly aligned regions. Since the trimming method used in both approaches is automated and was run under the same parameters, the increase in the retained sequence length is a direct consequence of a higher quality in the produced alignment. Thus, the alignment produced by the two-steps approach was kept for further analyses.

Data partitioning and model selection.

Protein coding genes as a single evolutionary unit.

The simplest of the partitioning schemes tested, 2-PartitionsA-C, are an over-simplified interpretation of the evolutionary history of choanoflagellates genomes. They are based on the prior that all protein coding genes are equally subjected to adaptative pressure and are evolving as a single unit. The results from partitioning schemes that split the protein coding genes by loci did not support such hypothesis, since none of them supported a single substitution model as the best-fitting for all the partitions containing protein coding genes. These partitioning schemes are also based on the prior that ribosomal subunits evolution is coupled, acting as a single evolutionary unit (further discussed below).

The partitioning scheme 4-PartitionsD can be considered as an especial extension of 2-PartitionsA-C. On one hand, it does account for different evolution rates for the different nucleotides regarding which position in the codon they belong to, which allows the model to have different evolutionary speeds for synonymous and non-synonymous changes on the nucleotide sequence. On the other hand, this scheme is based on the same over-simplistic prior than 2-PartitionsA-C partitioning schemes, that all protein coding genes are equally subjected to adaptative pressure and are evolving as a single unit. This partitioning scheme was designed to attempt the integration of all the phylogenetic information

contained in both the nucleotides and the amino acids sequences while avoiding the use of parameter-rich substitution models such as codon substitution models. Although the results from this partitioning scheme supported the hypotheses than every individual position of the codon evolves at different rates, the support for the best-fitting substitution model for the partitions containing the codon positions that produce non-synonymous changes, partitions 2 and 3, was weak ($w\text{-}AICc_{Partition\ 2}$=0.6984, $w\text{-}AICc_{Partition\ 3}$=0.6725). w-AICc represent a measurement of the accumulative likelihood of a model (likelihood of a model to be the best model among all models considered). Typically, values <0.75 are considered as low or non-supporting when an alternative hypothesis retains nearly as much or most of the remaining likelihood. For both partitions, the second best-fitting substitution models retained most of the remaining likelihood (w-AICc=0.2004, ≈66% of the remaining likelihood and w-AICc=0.2888, ≈88% of the remaining likelihood, for partitions 2 and 3 respectively). Furthermore, the averaged w-AICc for the three partitions that contained the sequences of protein coding genes, partitions 2, 3 and 4, was lower than the w-AICc for the partition that contained the same data and used codon substitution models, partition 2, under the partitioning scheme 2-PartitionsC ($w\text{-}AICc_{Partition\ 1+Partition\ 2+Partition\ 3}$/3=0.7493, $w\text{-}AICc_{Partition\ 2}$=0.9999, for 4-PartitionsD and 2-PartitionsC, respectively). Therefore this partitioning scheme represented no benefit compared to, 2-PartitionsC, a comparable partitioning scheme that used codon substitution models.

One gene, one partition schemes.

On the contrary, the most complex partitioning schemes tested, 6-PartitionsA-C, consider that every gene is evolving at different rates and as individual units. Although this is partly supported by the results of model selection, 6-PartitionsC found that all the different partitions are evolving under different substitution models, there are two major drawback affecting these partitioning schemes. Firstly, this partitioning schemes considers both ribosomal subunits as individual evolutionary units. Although the best-fitting substitution models are different for both ribosomal subunits in all three partitioning schemes, the AICc support for the best-fitting model of Partition 1 is weak in all three cases ($w\text{-}AICc_{Partition\ 1}$<0.3, for the best-fitting substitution model). The second best-fitting substitution models retained nearly as much of the likelihood as the best-fitting substitution model (w-AICc=0.1380, w-AICc=0.1627, w-AICc=0.1728; for the second best-fitting substitution models for 6-PartitionsA, 6-PartitionsB and 6-PartitionsC respectively) (further discussion on ribosomal subunits

evolution below). Secondly, this partitioning scheme considers both EF-1A and EFL as individual evolutionary units. This prior is supported by the result of both partitioning schemes 6-PartitionsA and 6-PartitionsC. However, when considering only the amino acids sequence, the results from 6-PartitionsB supported the best-fitting model, for both EF-1A and EFL, being the same (LG+F+I+G4). Moreover, both 6-PartitionsA and 6-PartitionsC supported that both genes are indeed evolving under similar substitution models (GTR+F and KOSI07+F, for 6-PartitionsA and 6-PartitionsC respectively) that differ only in the type and number of parameters of the model (R3 vs I+G4 and R6 vs G4, for 6-PartitionsA and 6-PartitionsC respectively).

From the species, in the dataset, for which EF-1A or EFL sequences were available (52.11% of the species in the dataset), EF-1A sequences were available for 51.35% of the species and EFL sequences were available for 48.65% of the species. Out of those species, only 5 species (13.51%) have both EF-1A and EFL and, when this happens, EF-1A is highly modified. Such modifications consisted on elevated rates of nucleotides substitutions at both synonymous and non-synonymous sites, reduced strength of purifying selection and weaker biases in codon usage, suggesting a loss of functional constraint (Carr *et al*, 2017) Elongation factors are essential components of the for protein biosynthesis processes. More specifically, EF-1A is a key component of the translational machinery that binds to aminoacylated tRNA and mediates its transport to the A-site of the ribosome. This is a crucial step during the elongation phase of translation. The fact that this gene is not present in most of the species for which EFL sequence are available and that, when its present, it seems to have lost its functional constrains, suggests that EFL is replacing EF-1A and acting as mediator in the transport of aminoacylated tRNA to the ribosome. Such assumption is in accordance with previously published studies (Kamikawa *et al.*, 2013; Atkinson *et al.*, 2014; Carr *et al.*, 2017). Since both genes seem to be fulfilling the same role it is expected that they are subjected to the same adaptative pressures. Hence, the possibility that the differences in parameters between the best-fitting substitution models for both partitions are the result of numerical artefacts, due to the incompleteness of the datasets, rather than the result both genes evolving differently cannot be easily discarded.

Schemes combining ribosomal subunits and elongation factors into evolutionary units.

The remaining three partitioning schemes, 4-PartitionsA-C are, on the one hand, based on the prior that both ribosomal subunits are evolving coupled and under the same evolution model. In eukaryotes, all three ribosomal subunits 18S, 5.8S and 28S are organised into a single polycistronic rDNA transcription unit, which also contains two external transcribed spacers in both 5' and 3' ends (5'ETS and 3'ETS) and two internal transcribed spacers that separate 18S from 5.8S and 5.8S from 28S subunits (ITS1 and ITS2 respectively) (Perry *et al.*, 1970; Srivastava and Schlessinger, 1991; Russell and Zomerdijk, 2005). All this subunits conform the rDNA transcription unit, which is transcribed as a whole and later cleaved during rRNA maturation. Since the sequences of SSU and LSU are two parts of a single transcription unit and are an integral part of the cell's transcription machinery, they can be considered as a single evolutionary unit and it is expected that both parts are subjected to the same adaptative pressures, thus evolving under a single model. Such prior is supported by the fact that all but one of the partitioning schemes that considered both ribosomal subunits as a single evolutionary unit, supported GTR+F+R10 as the best-fitting substitution model for Partition 1. Moreover, out of those, the only partitioning schemes that supported an alternative substitution model as best-fitting, 4-PartitionsA, did so with weak support ($w\text{-}AICc_{Partition\ 1}=0.3412$).

On the other hand, these partitioning schemes are based on the prior that both EF-1A and EFL are evolving as a single unit. As above discussed, the data suggests that both genes are fulfilling the same role and therefore may be evolving under the similar evolutionary forces, thus they can be considered as a single evolutionary unit. Similarly to partitioning schemes 6-PartitionsA-C, partitioning schemes 4-PartitionsA-C are based on the prior that each evolutionary unit is evolving individually and can evolve under different models depending on the adaptative pressures they are subjected to. Because of the priors these partitioning schemes are based on, they are here interpreted as the most realistic ones of the schemes used for the phylogenetic reconstructions.

Treating the same data differently: nucleotides vs amino acids vs codons substitution models.

The different partitioning schemes tested differed not only in the number of partitions they used and the priors behind such numbers, but also on the type of sequences that were used as input for the partitions containing the protein coding genes (nucleotides or amino acids sequences) and the substitution models that were tested over such data.

Partitioning schemes that considered only nucleotides substitution models (with the exception of the already discussed 4-PartitionsD) do not account for the differences in the evolution models between nucleotides that produce synonymous and non-synonymous changes in the resulting proteins. Synonymous changes in the DNA sequence do not produce changes in the resulting proteins and therefore are not subjected to the same adaptative pressures as non-synonymous changes. While non-synonymous changes result in different proteins over which natural selection can act upon (e.g. proteins that cannot fulfil their role, adaptation of enzymes to new substrates), synonymous changes result in the same proteins and therefore natural selection cannot act differently over the resulting proteins. Nevertheless, nucleotides producing synonymous changes are still under selective pressures such as preferential codon usage. Some choanoflagellates exhibit strong a bias towards the preferential use of specific nucleotides in the 3rd position of the codons (Carr *et al.*, 2017, Southworth *et al.*, 2018). Such bias towards optimal codon usage can be related to two, non mutually exclusive, mechanisms: translational efficiency (faster translation and protein biosynthesis, and more accurate translation) and mutational pressure (Sharp *et* al., 1995; Smith and Eyre-Walker, 2001). Since synonymous and non-synonymous changes are subjected to different adaptative pressures it can expected that nucleotides producing both types of changes are evolving differently. The results from partitioning scheme 4-PartitionsD support such hypothesis since the estimated best-fitting substitution models are different for every position of the codon. Nevertheless, it is worth noting that partitioning schemes that considered only nucleotides substitution models (4-PartitionsA, 4-PartitionsD and 6-PartitionsA), did supported different variants of GTR+F as the best-fitting model for all partitions containing protein coding genes. This seems to indicate that, even though genes are subjected to different adaptative pressures and therefore are evolving at different rates, choanoflagellates' DNA sequences are evolving following a general pattern that fits the GTR substitution model.

Similarly, partitioning schemes that considered only amino acids substitution models do not account for synonymous changes in the DNA sequence. As discussed above, such changes are indeed subjected to selective pressures. Moreover, in a neutral evolution scenario, such mutations tend to accumulate over time and are the main force of speciation. Therefore, synonymous changes contain phylogenetic information that is of especial importance to accurately reconstruct the phylogenetic relationships between closely related species.

Finally, partitioning schemes that considered codon substitution models integrate the phylogenetic signals contained in both the changes in the nucleotide sequence and their effect over the resulting amino acid sequence. These substitution models are able to discriminate between synonymous and non-synonymous changes in the DNA sequence and define different probabilities for these changes to occur. They represent, therefore, a much more accurate way to model the molecular evolution of datasets that contain CDS sequences since they account for all the phylogenetic signal contained in such datasets, as opposed to nucleotides and amino acids substitution models.

Phylogenetic tree reconstruction.

Topology shared across all phylogenies: the monophyly of choanoflagellates and Acanthoecida.

As above discussed the different partitioning schemes used are based upon different priors and treated the dataset differently. As a result, different substitution models were estimated as the best-fitting substitution model for each partition. Substitution models are at the same time used as priors for hypothesis construction when reconstructing a phylogeny and, therefore, have a deep impact in the resulting phylogenetic reconstruction. Thus, the fact that the reconstructed phylogenies differed from each other was expected as they are based upon different priors. From all the partitioning schemes used for the phylogenetic reconstructions, 4-PartitionsC uses the most realistic priors and the most accurate substitution models and, therefore, the phylogeny recovered from this scheme is here considered as the most accurate reconstruction of them all (Figure 2.1.12).

Nevertheless, regardless of the partitioning scheme used, all phylogenies recovered choanoflagellates as a monophyletic group with a sister relationship to metazoans. This results are consistent with previously published studies (Carr *et al.*, 2017; Schiwitza *et al.*, 2018; Schiwitza *et al.*, 2019). Similarly, and in accordance with this same study, Acanthoecida, one of the two major groups of choanoflagellates, was recovered as monophyletic.

The placement of S. prava and its effects over the topology of the phylogenies.

The second major group of choanoflagellates, Craspedida, was only recovered when the most accurate substitution models were used (2-PartitionsC, 4-PartitionsC and 6-PartitionsC). The remaining reconstructions recovered *S. prava* as a basal species with a sister relationship to Acanthoecida. Many of the discrepancies in the topology of the phylogenies reconstructed are consequence of the placement of this species within the phylogenetic trees.

The support of *S. prava* as basal to all craspedids is weak (<75% mlBP) in all three cases. Although *S. prava* presents an atypical morphology that differs from most craspedis due to the presence of a life stage with no visible collar, the presence of a theca and the absence of lorica have been interpreted as evidence supporting a closer relationship to Craspedida than to Acanthoecida (Schiwitza *et al.*, 2018). Moreover, this same study recovered *S. prava* as sister to Clade 3 (71% mlBP). In the current analysis, the alternative placement for *S. prava* as a sister species to all acanthoecids is also weak (<75% mlBP) when only nucleotide substitution models are used, with the exception of 4-PartitionsD (75% mlBP). The only results that supported the placement of *S. prava* as basal to all acanthoecids with a support >75% mlBP were those based on partitioning schemes that used amino acids substitution models for partitions containing the sequences of protein coding genes (2-PartitionsB, 4-PartitionsB and 6-PartitionsB). Nevertheless, such phylogenetic reconstructions were the only ones that supported a sister relationship between the fresh-water codosigids and the clade formed by *S. dolichothecata* and *S. tuba*. Such relationship was weakly supported (<75%) in all three cases. Moreover, that relationship is inconsistent with previously published studies and it is most likely an artefact caused by the attraction of the long branches of the codosigids towards, the basal position of *S. dolichothecata* and *S. tuba*.

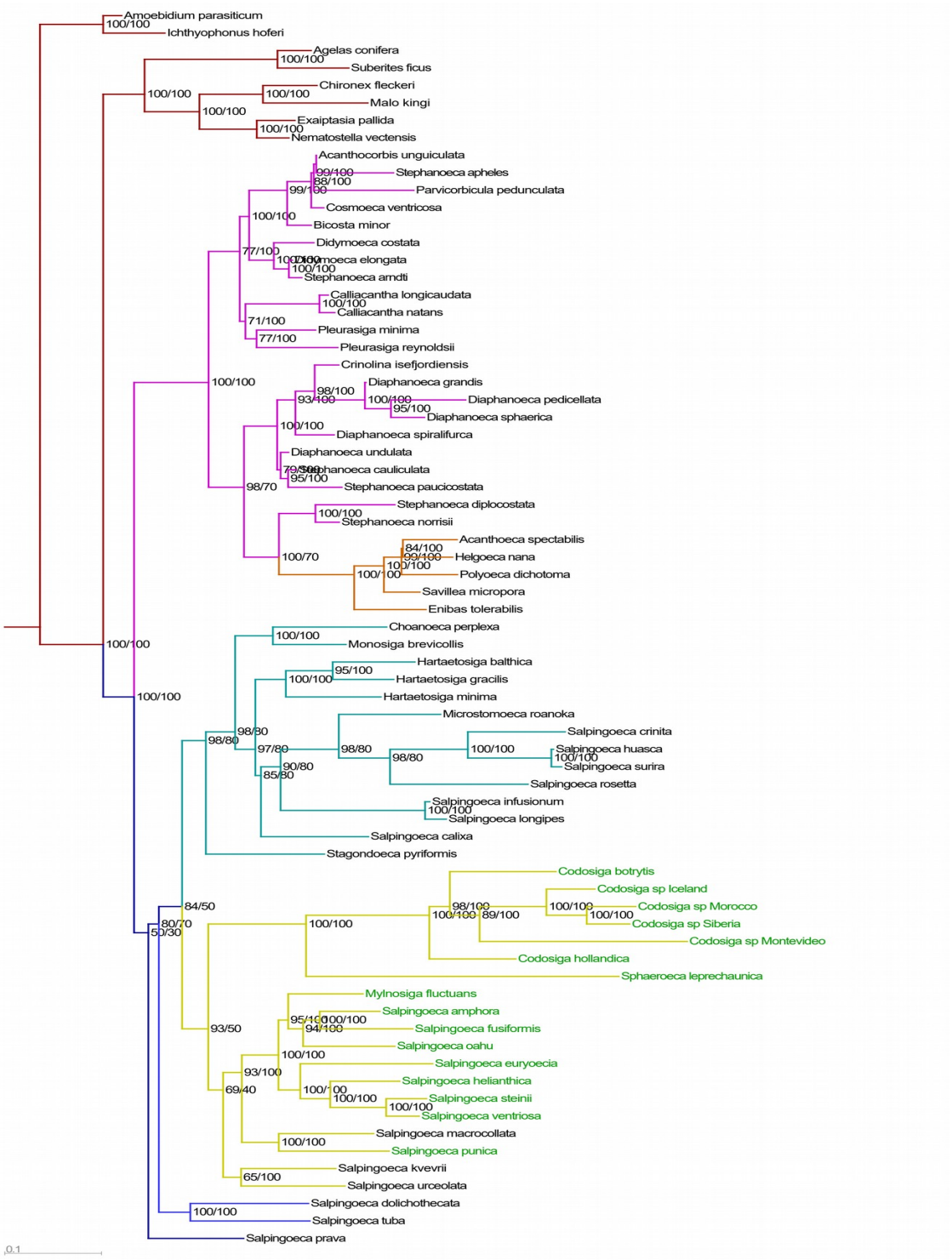

Figure 2.1.12: Consensus support tree for the phylogeny of choanoflagellates based on the partitioning scheme 4-PartitionsC. Branches are colour-coded indicating the different groups of species as follows: outgroup (red), basal choanoflagellates (blue), tectiforms (magenta), nudiforms (orange), Clade 1 craspedids (cyan), Clade 2 craspedids (yellow), Clade 3 craspedids (light blue). Species names in green indicate freshwater species. Branch lengths are drawn proportional to the number of substitutions per site as indicated by the scale. Support values indicate Ultrafast mlBP estimated for the phylogeny based on this partitioning scheme (left) and percentage of the phylogenetic reconstructions supporting such topology (right).

The placement of the codosigids in a basal position within the craspedids increases the dissimilarity between *S. prava* and the rest of the craspedids, thus increasing the support for its placement as basal to the acanthoecids.

From the remaining four reconstructions that weakly supported the placement of *S. prava* as basal to all the acanthoecids, two of them, 4-PartitionsA and 6-PartitionsA, also recovered the remaining craspedids as paraphyletic. Such topology is weakly supported in both cases (55% and 70% mlBP for the split between Clade 1 craspedis and *S. prava* and for the split between *S. prava* and the acanthoecids; 55% and 67% for the same splits, for phylogenies based on 4-PartitionsA and 6-PartitionsA respectively) and inconsistent with previously published studies (Carr *et al.*, 2008b; Carr *et al.*, 2017). This topologies are most likely the result of the attraction of the Clade 1 craspedids towards the acanthoecids due to sequence similarities with *S. prava* rather than with the acanthoecids. Similarly, these phylogenies recovered Clade 2 craspedids as paraphyletic, supporting a sister relationship between the hypersaline species *S. crinita*, *S. huasca* and *S surira* and the codosigids. This placement is inconsistent with Schiwitza *et al.* (2018) which similarly to the rest of the phylogenies recovered in the present study supported the placement of these hypersaline species in Clade 1 with a sister relationship to *S. rosetta*. This placement is most likely caused an artefact of long branch attraction effect, since both the hypersaline clade and the codosigids present the longest branches of the phylogenetic tree.

From the remaining two reconstructions supporting the placement of *S. prava* as basal to all acanthoecids, the reconstruction based on 2-PartitionsA recovered the splits between Clade 1 and Clade 2, the most basal splits within Clade 2 and most of the splits within Clade 1 with weak support (<0.70% mlBP). The same is true for the reconstruction based on 4-PartitionsD, with the exception of the split between Clade 1 and Clade 2 craspedids. Conversely, the phylogenies that recovered *S. prava* as basal to all craspedids, based on partitioning schemes 2-PartitionsC, 4-PartitionsC and 6-PartitionsC, consistently recovered a topology of Craspedida that is in congruence with previously published studies with strong support for most splits. Therefore, the placement of *S. prava* as a basal craspedid is here interpreted as the correct one, despite the results from the extended majority-rule consensus tree. Moreover, this results are consistent with previously published phylogenies (Schiwitza *et al.*, 2018).

The sequences of *S. prava* included in both the present study and Schiwitza *et al* (2018) were obtained independently, which combined with the fact that the strategies followed for the tree reconstruction vary between both studies can explain the differences obtained for the exact placement of *S. prava*. Whether the exact placement of this species in the phylogeny of choanoflagellates is as the most basal of the craspedids, as the results from the present study suggest, or as a basal Clade 3 choanoflagellate, as the results from Schiwitza *et al.* (2018) suggest, is open to interpretation, since in both cases the support for such specific placement is weak. The interpretation here exposed is that the phylogenetic reconstruction of the present study is more exhaustive and, for the arguments above discussed, more accurate than the one performed in Schiwitza *et al.* (2018) and, therefore, so is the placement of *S. prava*.

The placement of the nudiforms and its effects over the phylogeny of Acanthoecida.

Acanthoecida, also known as loricates, is an order of choanoflagellates that exhibit a characteristic morphological feature that is not found in any other group of choanoflagellates. The species belonging to this group have developed an extracellular basket-like siliceous structure, the lorica, that have allowed them to successfully colonise a wide range of marine environments. The lorica is composed of siliceous costae made up of rod-shaped costal strips, that are deposited and stored in bundles extracellularly. Once a full set of this costae has been produced, they are arranged in two layers forming the lorica (Leadbeater, 1994; Leadbeater, 2008b). There are two major variations of such process.

On one hand, in nudiforms when mature lorica-bearing cells divide, produces a ‘naked’ flagellated cell that swims away from the parent cell lorica and assembles its own lorica (Leadbeater, 2008a). On the other hand, in the tectiforms, when cells divides, prior to cell division a complete set of costae is produced and stored at the top of the collar. Once the complete set is produced, the cell divides and one of the cells is rotated upside down and pushed backwards, taking with it the bundles of strips accumulated on the top of the parent cell. The expelled cell then uses these bundles of strips to arrange a new lorica (Leadbeater, 2010).

The seemingly simplest cell division mechanisms, alongside the now discarded absence of transversal costae (Schiwitza *et al.*, 2019) has led to the interpretation that the nudiform condition is ancestral to Acanthoecidae (Carr *et al.*, 2017). Furthermore the difference on species abundance (only a handful of nudiform species have been described against >100 tectiform species) has been interpreted as the result of a major radiation occurring within the tectiforms as a result of the evolutionary advantages of this condition (Leadbeater, 2010). Finally, the fact that cells from *Stephanoeca diplocostata*, a tectiform species have been reported to be able to 'revert' a nudiform condition due to silica starvation (Leadbeater, 1989; Leadbeater and Cheng, 2010) has been interpreted as evidence supporting that the tectiform condition is derived from the nudiform condition.

Previously published phylogenies (Carr *et al.*, 2017; Schiwitza *et al.*, 2019) have recovered both major groups of Acanthoecida as sister monophyletic clades, supporting such interpretation and suggesting that the split between nudiforms and tectiforms occurred early in the evolution of Acanthoecida. However, the present study only recovered such topology when reconstructions were based on partitioning schemes that use only nucleotides substitution. In congruence with López-Escardó *et al.* (2019), the phylogenetic reconstructions based on partitioning schemes that use codon substitution models, interpreted in this study as the most accurate ones, recovered the nudiforms within the tectiforms, suggesting that the nudiform condition is derived from the tectiform condition.

It is here hypothesised that these results might be caused by the incompleteness of the dataset used. While nudiforms are represented in the present study by 5 species, the tectiforms are represented by 20 species. In contrast, protein coding genes sequences are available for 4 nudiforms (80% of the nudiforms) and only 2 tectiforms (10% of the tectiforms). Such discrepancies in the number of taxa and the completeness of the datasets combined with the richness in parameters of the codon substitution models might be over-fitting the model to the dataset for this specific section of the tree, causing the incorrect placement of the nudiforms within the tectiforms.

Nevertheless, due to the absence of any lorica-bearing species, living or fossil, sister to Acanthoecida, the hypothesis that nudiforms are derived from the tectiforms cannot be fully dismissed at the moment. Although that hypothesis can be regarded as less parsimonious, in order to fully resolve the relationship

between both groups, a more comprehensive dataset, both in number of taxa and in sequences availability for each taxa, is required.

The placement of S. kvevrii and S. urceolata and its effects over the evolutionary history of marine:freshwater transitions.

The phylogenetic reconstructions that recovered Clade 2 craspedids as a monophyletic clade differed on the exact placement of the marine species *S. kvevrii* and *S. urceolata*. The placement of these species is crucial to the unravelling of the evolutionary history of fresh-water invasions within Craspedida. Previous studies have placed this species as basal to the rest of the species of Clade 2 (Carr *et al.*, 2017; Schiwitza *et al.*, 2018). This placement is in congruence with the results from the phylogenetic reconstructions based on the partitioning schemes 2-PartitionsA, 2-PartitionsC and 4-PartitionsD. As discussed above, this three schemes have in common the prior that all protein coding genes are evolving as a single evolutionary unit. In all three cases, this placement is weakly supported since the support for the split between codosigids and fresh-water salpingoecids is <60% mlBP. Moreover, only the phylogeny based on 2-PartitionsC recovered the split between *S. kvevrii* and *S. urceolata* and the rest of the Clade 2 with strong support (98% mlBP).

On the other hand, the remaining phylogenies, based on the schemes 4-PartitionsA, 4-PartitionsC, 6-PartitionsA and 6-PartitionsD, recovered the placement of *S. kvevrii* and *S. urceolata* as sister to the rest of the Clade 2 salpingoecids, thus placing the codosigids as the most basal group of Clade 2 craspedids. In all four cases, the support for the split between *S. kvevrii* and *S. urceolata* and the rest of Clade 2 salpingoecids is weak (<70% mlBP). From these phylogenies, only those based on partitioning schemes that used codon substitution models (4-PartitionsC and 6-PartitionsC) strongly support the split between the codosigids and the rest of the Clade 2 species (93% mlBP in both cases).

Such results are indicating that phylogenetic reconstructions based solely on nucleotides substitution models are not good enough to properly unravel the relationships between species within Clade 2, as the support for the monophyly of Clade 2 is low. Nevertheless, although phylogenies using codon substitution models for protein coding genes strongly support the monophyly of Clade 2, they differ on

the exact placement of *S. kvevrii* and *S. urceolata*. When protein coding genes are considered a single evolutionary unit (2-PartitionsC), these species are recovered as the most basal to all Clade 2 craspedids, while when genes are considered separate evolutionary units (4-PartitionsC and 6-PartitionsC), the codosigids are recovered as the most basal Clade 2 craspedids. More importantly, the support for both alternative topologies is low, and if low supported splits (<75% mlBP) are collapsed, the resulting topology is the same. Such topology contains a polytomy with four groups at the base of Clade 2, namely: codosigids, freshwater salpingoecids, *S. kvevrii* and *S. urceolata*.

As mentioned above, the exact placement of *S. kvevrii* and *S. urceolata* within Clade 2 craspedids has a deep impact on the interpretation of the evolutionary history of craspedids' freshwater invasions. The resulting unresolved topology leads to three hypotheses that can explain the current distribution of freshwater species within Clade 2 craspedids.

The first of these hypothesis supports a scenario in which only one marine-to-freshwater colonisation took place early in the history of Clade 2 craspedids and two subsequent freshwater-to-marine reversions took place independently leading to the origin of the marine within Clade 2 craspedids. The second hypothesis supports a scenario in which two marine-to-freshwater colonisations took place independently. The first colonisation lead to the diversification of codosigids, while the other one lead to the diversification of freshwater salpingoecids. This second colonisation was subsequently followed by a freshwater-to-marine reversion that lead to the origin of *S. macrocollata*. Lastly, the third hypotheses supports a scenario in which three marine-to-freshwater took place independently leading to the diversification of codosigids, *S. punica*, and the remaining freshwater salpingoecids.

Based on the prior that invasions from marine to freshwater habitats are as likely to happen as reversions from freshwater to marine habitats, these three hypotheses are equally parsimonious. Although marine-to-freshwater transitions in unicellular eukaryotes are generally regarded as infrequent events that occurred early in the evolution of many phyla (Logares *et al.*, 2009), several recent studies have showed that multiple transitions have happened within closely related groups of the same phylum (Bråte *et al.*, 2010; Annenkova *et al.*, 2015; Dittami *et al.*, 2017; Žerdoner Čalasan *et al.*, 2019), some of them as recently as 40 MYA. Even more, euryhalinity has been confirmed in several

non freshwater choanoflagellates such as *Enibas tolerabilis* (Schiwitza *et al.*, 2019), *Microstomoeca roanoka* (present study, further discussed in Chapter III, Section II), *S. rosetta* and *M. brevicollis* (Nitsche, personal communication). Similarly, freshwater habitats have been traditionally considered as absorbing habitats and marine-to-freshwater transitions regarded as irreversible. Previous phylogenies of choanoflagellates (Carr *et al.*, 2017) interpreted the ecology of *S. macrocollata* as an example of freshwater-to-marine reversion. Moreover, Nakov *et al*. (2019) proved that marine:freshwater transition rates, in either direction, are lineage dependant in diatoms.

There is, therefore, no conclusive argument to reject the prior that marine:freshwater transitions are equally likely to occur in either direction. Thus, the three hypotheses remain equally parsimonious. In order to further explore the evolutionary history of marine:freshwater transitions, the present study used a combination of ancestral sequences reconstruction and phylogenetic placement of environmental sequences. Such approaches and their implications in unravelling the evolutionary history of marine:freshwater transitions in choanoflagellates are presented and discussed in the following sections of this chapter.

A note on the taxonomy of choanoflagellates.

Although it is not the main focus of the present study, it is clear from the results here obtained that the taxonomy of choanoflagellates requires urgent revision. The monophyly of several genera of choanoflagellates is not supported by any of the phylogenies here recovered and therefore such genera call for immediate revision. The results of the phylogenetic reconstructions here exposed support the need for the revision of the genus *Salpingoeca*, proposed by Carr *et al.* (2017), and the genus *Stephanoeca*, proposed by Schiwitza *et al.* (2019). In addition, the results here exposed consistently support the placement of *Criolina isefjordiensis* within the genus *Diaphanoeca* and, therefore, it is here interpreted that the taxonomy of this species should be reviewed, with the most likely outcome of its inclusion within the genus *Diaphanoeca*.

Conclusions.

Through exhaustive phylogenetic analyses, the results here exposed support the monophyly of choanoflagellates and the major groups Acanthoecida and Craspedida. The results suggests that nudiforms have derived from tectiforms, although it has been here hypothesised that the incompleteness of the dataset might be affecting the accuracy of the analyses around this specific section of the phylogenetic tree of choanoflagellates. The recent publication of transcriptomes from 19 species of choanoflagellates (Richter *et* al., 2018) provides an unprecedented amount of data that could be used in future studies to identify highly conserved, slowly evolving proteins. The inclusion of such sequences might help to further resolve those sections of the phylogenetic tree for which, currently, conflicting topologies are being recovered. The results here exposed support the early split between freshwater codosigids and salpingoecids, opening three equally parsimonious scenarios for the evolutionary history of marine:freshwater transitions within Clade 2 choanoflagellates. The first of these scenarios is that one marine-to-freshwater colonisation took place early in the history of Clade 2 craspedids and was subsequently followed by two freshwater-to-marine reversions that led to the origin of the marine species within the group. The second scenario is that two marine-to-freshwater colonisations took place independently and a later freshwater-to-marine reversion led to the origin of *S. macrocollata*. The third scenario is that three independent marine-to-freshwater colonisation took place leading to the origin of the different freshwater lineages of the group.

-Section II: Unravelling the evolutionary history of marine:freshwater transitions in choanoflagellates I: ancestral sequence reconstructions within Clade 2.

Introduction.

As discussed in the previous section of this chapter, the current evolutionary history of marine:freshwater transitions within Clade 2 craspedids can be explained by three equally parsimonious hypothesis. The first of these hypothesis (Scenario 1) is that only one marine-to-freshwater colonisation took place early in the history of Clade 2 craspedids and two subsequent freshwater-to-marine reversions took place independently leading to the origin of the marine taxa within the group. The second hypothesis (Scenario 2) is that two marine-to-freshwater colonisations took place independently leading to the diversification of codosigids and freshwater salpingoecids. This second colonisation was subsequently followed by a freshwater-to-marine reversion that lead to the origin of *S. macrocollata*. The third hypothesis (Scenario 3) is that three marine-to-freshwater colonisations took place independently leading to the diversification of codosigids, *S. punica*, and the remaining freshwater salpingoecids.

In order to further explore these scenarios, a phylogenetic reconstruction was carried out including the reconstructed ancestral sequence of the last common ancestor of Clade 2 craspedids. Sequences of the last common ancestor of Clade 2 craspedids were reconstructed under three evolutionary scenarios and subsequently placed in the phylogeny of choanoflagellates reconstructed in the previous section of this chapter. The study of the phylogenetic relationships between the different subclades and the reconstructed sequences was used to detect the phylogenetic signals of marine:freshwater transitions within Clade 2 craspedids.

Material and methods.

The sequences of both SSU and LSU ribosomal subunits were reconstructed for the last common ancestor of Clade 2 craspedids using IQ-TREE v1.6.12. This software uses an empirical Bayesian approach to reconstruct ancestral sequences. For a given phylogeny and following a given substitution model, it estimates the posterior probability for each nucleotide to each position of the sequence. On each site, the nucleotide with the highest posterior probability is kept as the ancestral state. Although this approach assumes that the evolutionary model parameters and the phylogenetic tree are known without error, and the size and complexity of the data can make this assumption unrealistic (Joy et al., 2016), it has proven an effective methodology for both reconstructing functional proteins (Gumulya et al., 2018; Furukawa et al., 2020) and exploring divergent evolution of a protein-protein interaction (Laursen et al., 2021).

The reconstruction was carried out under three different scenarios. Under the fist scenario, the sequence ASRFW was reconstructed from the aligned sequences of all fresh-water Clade 2 craspedids. Under the second scenario, the sequence ASRM was reconstructed from the aligned sequences of all marine Clade 2 craspedids. Finally, under the third scenario, the sequence ASRFWM was reconstructed from the all the aligned sequences of all Clade 2 craspedids.

In order to maximise the congruence between the ancestral sequence reconstruction and the previously reconstructed phylogeny, the reconstruction was performed following the topology of the tree reconstructed from the phylogeny based on the partitioning scheme 4-PartitionsC, interpreted in the present study as the most accurate phylogeny. In order to minimise computational burdens the constrain tree and the reference alignments used contained only the sequences used for the ancestral sequence reconstruction and the SSU and LSU sequences from *Stagondoeca pyriformis*, the most basal species of the sister clade to Clade 2 craspedids. *S. pyriformis* was used to as an outgroup to root the constrain phylogenetic tree and GTR+F+R10, the best-fitting substitution model estimated during the phylogenetic reconstruction, was used as the substitution model for the reconstruction.

The reconstructed ancestral sequences were included in the MSA and a new round of phylogenetic reconstructions was performed using IQ-TREE v1.6.12. In order to maximise the congruence between the current phylogenetic reconstruction and the previous phylogeny, the partition scheme and the substitution models used for every partition were kept the same as in the analysis based on 4-PartitionsC. Moreover, the phylogenetic tree recovered from the afore mentioned analysis was used as a constrain to the topology of the current reconstruction. Maximum-likelihood (ML) values were estimated using 100000 ultrafast bootstrap replicates (Minh *et al.*, 2013). The graphical representation of the tree was produced using Archaeopteryx v0.9920.

Results.

Three ancestral sequences were reconstructed, each containing the concatenated sequence of both ribosomal subunits, and included into the MSA. The phylogenetic reconstruction, based on the partitioning scheme 4-PartitionsC, recovered all three of them within Clade 2 craspedids. ASRM was recovered as sister to *S. urceolata*, but not to *S. kvevrii* (100% mlBP). ASRFWM was recovered as basal to the clade containing all freshwater salpingoecids and *S. macrocollata* (97% mlBP). ASRFW was recovered as basal to the clade containing all freshwater salpingoecids but *S. punica* (73% mlBP). None of the reconstructed sequences was basal to the codosigids (Figure 2.2.1).

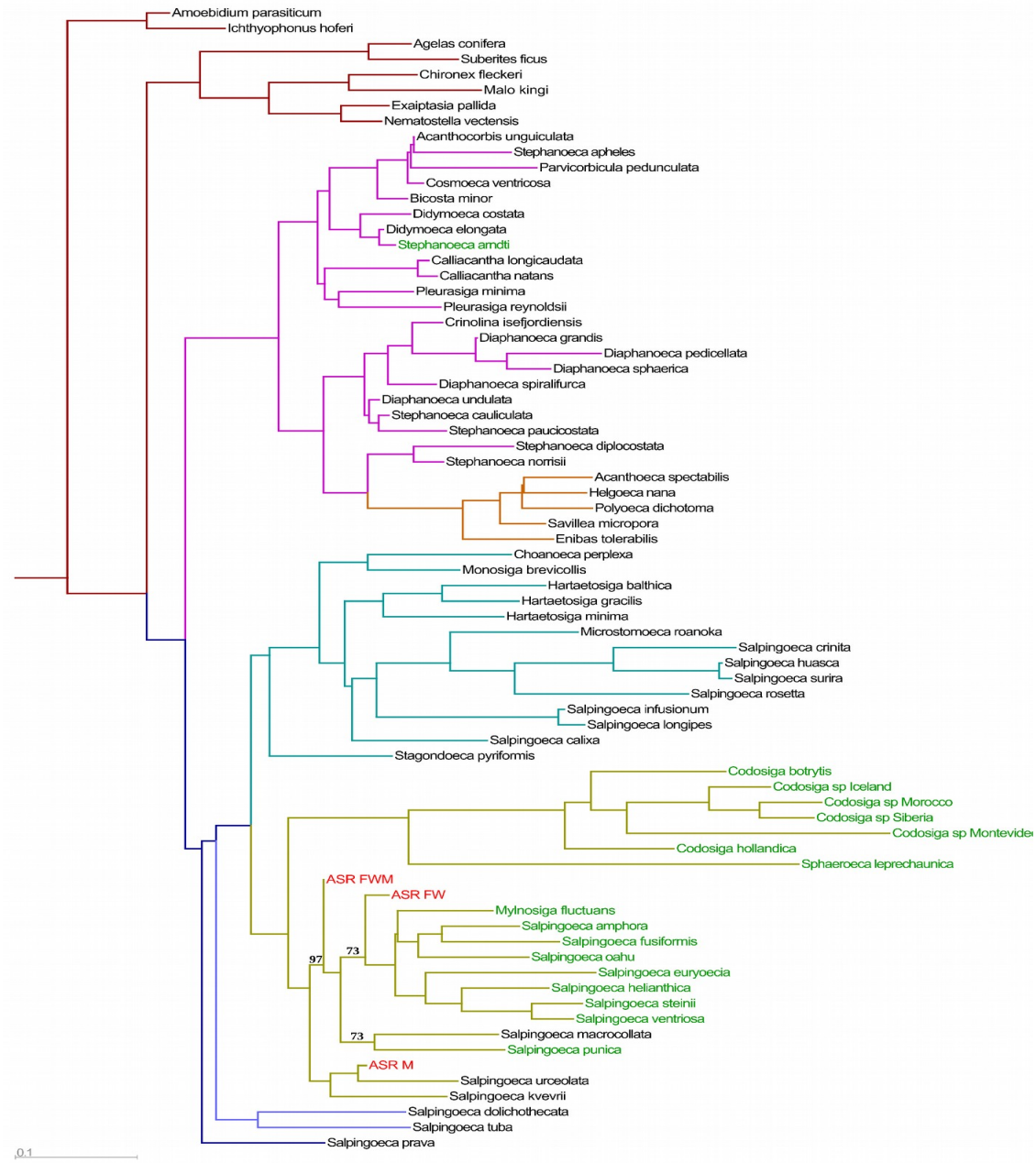

Figure 2.2.1: Phylogenetic placement of the reconstructed ancestral sequences on the phylogeny, based on the partitioning scheme 4-PartitionsC. Branches are colour-coded indicating the different groups of species as follows: outgroup (red), basal choanoflagellates (blue), tectiforms (magenta), nudiforms (orange), Clade 1 craspedids (cyan), Clade 2 craspedids (yellow), Clade 3 craspedids (light blue). Species names in green indicate freshwater species. Names in red indicate reconstructed ancestral sequences. Branch lengths are drawn proportional to the number of substitutions per site as indicated by the scale. Ultrafast mlBP support values are 100% unless otherwise stated on top of the branch.

Discussion.

All three reconstructed ancestral sequences placed at different depths within the Clade 2 salpingoecids. The support for the different placements was strong for ASRM and ASRFWM (100% and 97% mlBP respectively) but weak for ASRFW (73%). (Figure 2.2.1). These three sequences have been reconstructed using three different subsets of data which as deep implications on the assumed autoecology of the reconstructed last common ancestor.

ASRFW is based only on freshwater species and therefore it can be assumed that it corresponds to a freshwater state. In contrast, ASRM is based only on marine samples, and therefore it can be assumed that ASRM corresponds to a marine state. This methodology of subsampling extant species has been proved to have minor effects on the reconstructed sequences (Randall *et al.,* 2016). Lastly, ASRFWM in based on both marine and freshwater species. Since according to the phylogeny used by the analysis to infer the ancestral sequences, the marine state is ancestral to the freshwater state, it can be assumed that the ancestral state of ASRFWM is marine.

The placement of ASRM and the limitations of its reconstruction.

ASRM was placed as sister to *S. urceolata* but not to *S. kvevrii*. Such placement might be seen as contradicting the expectation that ASRM should be placed basal to all marine Clade 2 craspedids. Nevertheless, the interpretation for this placement is that ASRM is affected by two of the main limitations of ancestral sequence reconstructions. Firstly, as mentioned before, the empirical Bayesian approach used to reconstruct the sequences assumes that the evolutionary model parameters and the tree used to reconstruct the sequence are true (Joy *et al.*, 2016). The evolutionary model parameters used to reconstruct the sequence were exactly the same as the parameters used for the phylogenetic reconstruction. However, the phylogenetic tree provided for the reconstruction of ASRM, although is not wrong, does not contain a good representation of the diversity contained within Clade 2, since it only contains marine species. As previously mentioned, subsampling has been proved to have minor

effects on the reconstructed sequences (Randall *et al.*, 2016). Secondly, ancestral sequence reconstructions loose accuracy with node depth (Joy *et al.*, 20016; Randall *et al.*, 20016). When reconstructing ASRM, the last common ancestor of Clade 2 represents an internal node close to the external branches of the tree. Nevertheless, when the full diversity of Clade 2 is integrated during the subsequent phylogenetic reconstruction the reconstructed last common ancestor represents a much deeper internal node on the phylogeny. The priors that ASRM retains a good representation of the diversity of Clade 2, and that the increase in depth associated with the integration of the full diversity of Clade 2 has no effect over the accuracy of ASRM reconstruction, are unrealistic. Thus, the phylogenetic and evolutionary signal of ASRM are here interpreted as non meaningfull. This phenomenon was only observed for the placement of ASRM, while both ASRFW and ASRFWM retained their original placements within internal nodes of Clade 2.

The placement of ASRFW and ASRFWM and it implications on the evolutionary history of marine-to-freshwater colonisations in Clade 2 craspedids.

The resulting phylogeny strongly supports the placement of the codosigids as a sister group to the salpingoecids, with both ASRFW and ASRFWM being recovered as salpingoecids. The fact that none of the ancestral sequences are recovered as basal to the codosigids is here interpreted as suporting an early freshwater colonisation event at the origin of codosigids. Such event lead to a evolutionary radiation that quickly diverged from the rest of the Clade 2 salpingoecids. The placement of ASRFWM as sister to all freshwater salpingoecids and *S. macrocollata* marks the evolutionary moment at which Clade 2 salpingoecids had not yet colonised freshwater environments. Most of the diversity of freshwater salpingoecids originated after a single freshwater colonisation that occurred at the placement of ASRFW. This colonisation has led to the diversification of all freshwater salpingoecids but *S. punica*. The fact that ASRFW was not recovered basal to *S. punica* is here interpreted as suporting the hypotesis that the colonisation event that lead to the diversification of *S. punica* is independent from the event that lead to the diversification of the rest of freshwater Clade 2 salpingoecids. Therefore, these results are interpreted as supporting a Scenario 3 hypothesis, in which, at least, 3 independent colonisations of freshwater environments have taken place within Clade 2 craspedids. Such interpretation is in congruence with previously published studies that have described similar, multiple independent freshwater colonisation events occurring within closely related species in other groups of

unicellular eukaryotes (Bråte *et al.*, 2010; Annenkova *et al.*, 2015; Dittami *et al.*, 2017; Nakov *et al.* 2019; Žerdoner Čalasan *et al.*, 2019).

Comparing the phylogenetic signal produced by the freshwater invasion of codosigds and freshwater Clade 2 salpingoecids.

When comparing codosigids and freshwater Clade 2 salpingoecids, disimilarities can be appreciated on the phylogenetic signal produced by the freshwater colonisation events. In codosigids, freshwater colonisation lead to a rapid evolutionary radiation that resulted in a highly divergent group. The presence of basal, long branches within codosigids is here interpreted as a conequence of such rapid radiation. This phylogenetic signal is usually interpreted as an ancient colonisation of a new habitat (Logares *et al.*, 2009). The broad range of open niche space present at the newly colonised habitat triggers a rapid radiation of the organisms able to colonise it.

Contrastingly, in freshwater Clade 2 salpingoecids the main freshwater colonisation event lead to the diversification of closely related species. The short branches present throughout the diversity of this groups can be interpreted as the result of the combined effect of two processes.

Firstly, this colonisation event is younger than the event leading to the diversification of the codosigids. Due to the shorter time frame bigger differences have not yet had time to accumulate and as a result Clade 2 salpingoecids are less divergent from their closest marine relatives than codosigids from theirs. Nevertheless, the lack of a fossile record impedes the proper callibration of the reconstructed phylogeny thus it is not possible to estimate the time elapsed between both colonisation events.

Secondly, since freshwater Clade 2 salpingoecids colonised freshwater habitats after codosigids had already done it, the diversification of Clade 2 salpingoecids that followed the colonisation event can be regulated by competitive constrains. Such competitive constrains are derived from the direct competition with the codosigids that occupy the same niche. In order for a colonisation event to result into an adaptative radiation under such constrains it requires either a genetic innovation that confers an

evolutive advantage over the other organisms competing for the same niche, or to take place within ecosystems where the local diversity of competitors is depauperated. When the colonisation event takes place within ecosystems that have a depauperated local diversity of competitors the colonisation event produces an adaptative radiation of closely related species that occupy the available niches. In metazoans, competitive constrains have been prooved to regulate adaptative radiations at both regional and community scales (Betancur-R. e*t al.*, 2012). Previous studies have proved that unicellular eukaryotes such as diatoms have higher diverification and turn over rates in freshwater than in marine environments (Nakov *et al.*, 2019), which combined with the evolutionary constrains imposed by competition can lead to multiple, independent adaptative radiations occurring at regional scales (Betancur-R. e*t al.*, 2012).

The combination of both the age of the colonisation events, and the competitive constrains that regulate the subsequent adaptative radiations are is here interpreted as the cause of the disimilarites between the phylogenetic signals produced by the freshwater colonisation events that led to the diversification of codosigids and freshwater Clade 2 salpingoecids. Since all the freshwater Clade 2 salpingoecids, but *S. punica*, were recovered as a monophyletic group, the phylogenetic reconstruction carried out on the previous section of this chapter does not recovered any signal of adaptative radiations occurring at a regional scale. Nevertheless, although the geographical distance between freshwater Clade 2 salpingoecids covers a wide distribution area, it is here recognised that the taxon sampling around this section of the tree might not be deep enough to fully recover such signal. A wider taxon sampling could recover the phylogenetic signal of adaptative radiations occurring at regional scales and reveal multiple, independent marine:freshwater transitions across craspedids and specially freshwater Clade 2 salpingoecids, following a pattern similar to the one observed in other eukaryotes such as dinophytes (Žerdoner Čalasan *et al.*, 2019). Similarly, inclusion of new taxa around codosigids could deeply impact their phylogenetic placement and their phylogenetic relationships with the remaining Clade 2 craspedids, thus impacting the evolutionary history of marine:freshwater transitions in choanoflagellates.

Conclusions.

Through the analysis of the phylogenetic relationships between extant species and the reconstructed sequences of the last common ancestor of Clade 2 choanoflagellates, the results here exposed support the scenario of at least three independent marine-to-freshwater transitions within the Clade 2 choanoflagellates. Similarly, the results support that the diversity of freshwater codosigids derived from a single freshwater colonisation event that led to an adaptative radiation. Such event took place early in the evolution of Clade 2. The results here exposed also support that the diversity of all freshwater Clade 2 salpingoecids but S. punica is derived from a single freshwater colonisation. However, the phylogenetic signal left by the freshwater colonisation events is different for codosigids and Clade 2 salpingoecids. It has been here hypothesised that such differences are consequence of the combined effects of both the age of the colonisation events and the regulation by competitive constrains. The phylogenetic analyses carried out in both the present and the previous section of this chapter do not recovered any signal of adaptative radiations occurring at a regional scale. Nevertheless, it has been here proposed that a wider taxon sampling could reveal the phylogenetic signal of such processes and this could have deep implications on the evolutionary history of marine:freshwater transitions in choanoflagellates.

-Section III: Unravelling the evolutionary history of marine:freshwater transitions in choanoflagellates II: phylogenetic placement of environmental sequences.

Introduction

As discussed in the previous section of this chapter, the results from the phylogenetic placement of reconstructed ancestral sequences within Clade 2, supports that multiple independent marine-to-freshwater transitions have occurred within this clade. Moreover, the phylogenetic signal left by these transitions is different for codosigids and Clade 2 salpingoecids. In order to further explore the evolutionary history of marine:freshwater transitions within choanoflagellates, an analysis of phylogenetic placement of environmental sequences was carried out. This analysis was used to explore the distribution of the hidden biodiversity within choanoflagellates, in both marine and freshwater habitats, and to compare such distributions between both habitats. This results were subsequently compared to the phylogenetic signals detected on the previous sections of this chapter. Furthermore, the distribution of hidden freshwater diversity was used to revise the phylogenetic position of the currently proposed environmental freshwater clades of choanoflagellates.

Material and methods.

For such analysis, all publicly available data, from amplicon sequencing of DNA corresponding to the size fraction of protist, was downloaded from Tara Oceans' database (Pesant *et al*., 2015). This dataset includes SSU sequences of 1170 eDNA samples from 153 sites across the world oceans at 3 different depths. Paired-end sequences were merged from the original single-end sequences and filtered by quality score and sequence length using OBITools package (Boyer *et al.*, 2014). Sequences with either quality score <40.00 or sequence length <100 bp were discarded. In addition, 891 sequences of SSU, identified as eukaryotes, opisthokonts and choanoflagellates, were retrieved from SILVA SSU Parc

138.1 database (Pruesse *et al.*, 2007). SILVA SSU Parc 138.1 is a curated database and only includes sequences of >300 bp length.

A total of 67281 SSU sequences (full dataset available from the author upon request) were aligned using PaPaRa v2.5 (Berger and Stamatakis, 2012). The untrimmed version of the SSU alignment used for the phylogenetic reconstructions, after realignment of poorly aligned regions in AliView v1.26 and before sequence trimming in trimAl v1.2rev59, and the reconstructed phylogeny based on partitioning scheme 4-PartitionsC were used as reference alignment and reference tree respectively. Due to the lack of SSU sequences in the dataset, *Salpingoeca amphora* was removed from the reference MSA and the reference tree for this analysis.

Aligned sequences were placed into the reconstructed phylogeny using RAxML Evolutionary Placement Algorithm (RAxML-EPA) in RAxML v8.2.12 (Stamatakis, 2018), with default parameters. Due to FreeRate models for heterogeneity across sites (+R) not been implemented in RAxML, GTR+Γ was used as the substitution model for the analysis. Placements uncertainty was estimated by expected distance between placement locations (EDPL) using Gappa v0.6.1 (Czech *et al.*, 2020). Query sequences that contained multiple placements were accumulated towards the root until a threshold of 0.95 likelihood weight ratio was reached, using Gappa v0.6.1, and the resulting placements were converted into a heat phylogenetic tree, where branch colour represented the number of query sequences placed in it. The graphical representation of such tree was produced using Archaeopteryx v0.9920.

Results.

Average EDPL, a measurement of overall placement uncertainty, was 0.0362 (σ=0.0341) (expected number of substitutions per site). The average branch length of the reference tree was 0.0829 (σ=0.0876) expected number of substitutions per site. 91.6% of the placements had an EDPL <0.0829 (Figure 2.3.1).

The phylogenetic placement analysis recovered sequences placed all along the reference tree, with sequences identified as belonging to all three major groups of the reference tree: Craspedida, Acanthoecida and outgroup (Figure 2.3.2).

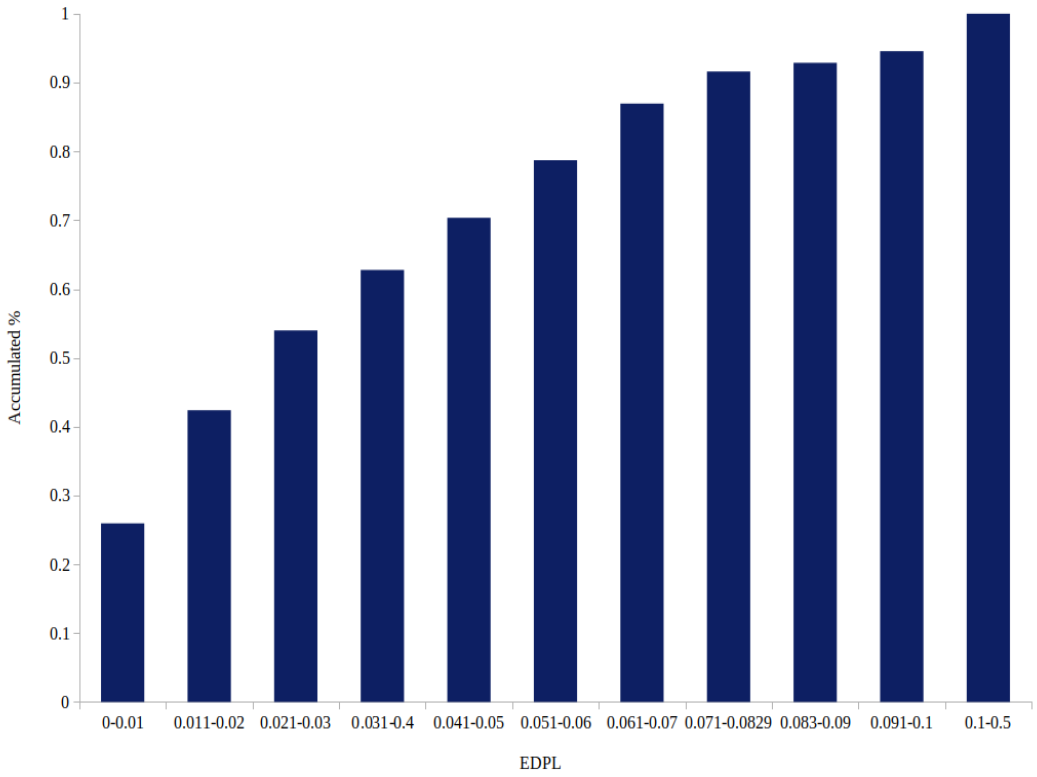

Figure 2.3.1: Phylogenetic placements uncertainty. Accumulated percentage of the query sequences per EDPL range, of the phylogenetic placement of 67281 SSU eDNA sequences in the phylogeny of choanoflagellates, based on the partitioning scheme 4-PartitionsC.

Within Acanthoecida, the branches containing the higher number of placements are the most basal branch of all acanthoecids (~5000 sequences placed), the branch leading to the split between the lineages of genus *Calliacantha* and *Pleurasiga* and the lineages of genus *Didymoeca* and its closely related species (>5000 sequences placed) and the branch leading to the split between the lineages of genus *Didymoeca* and *Stephanoeca arndti* and the rest of their closely related species (~2500 sequences placed).

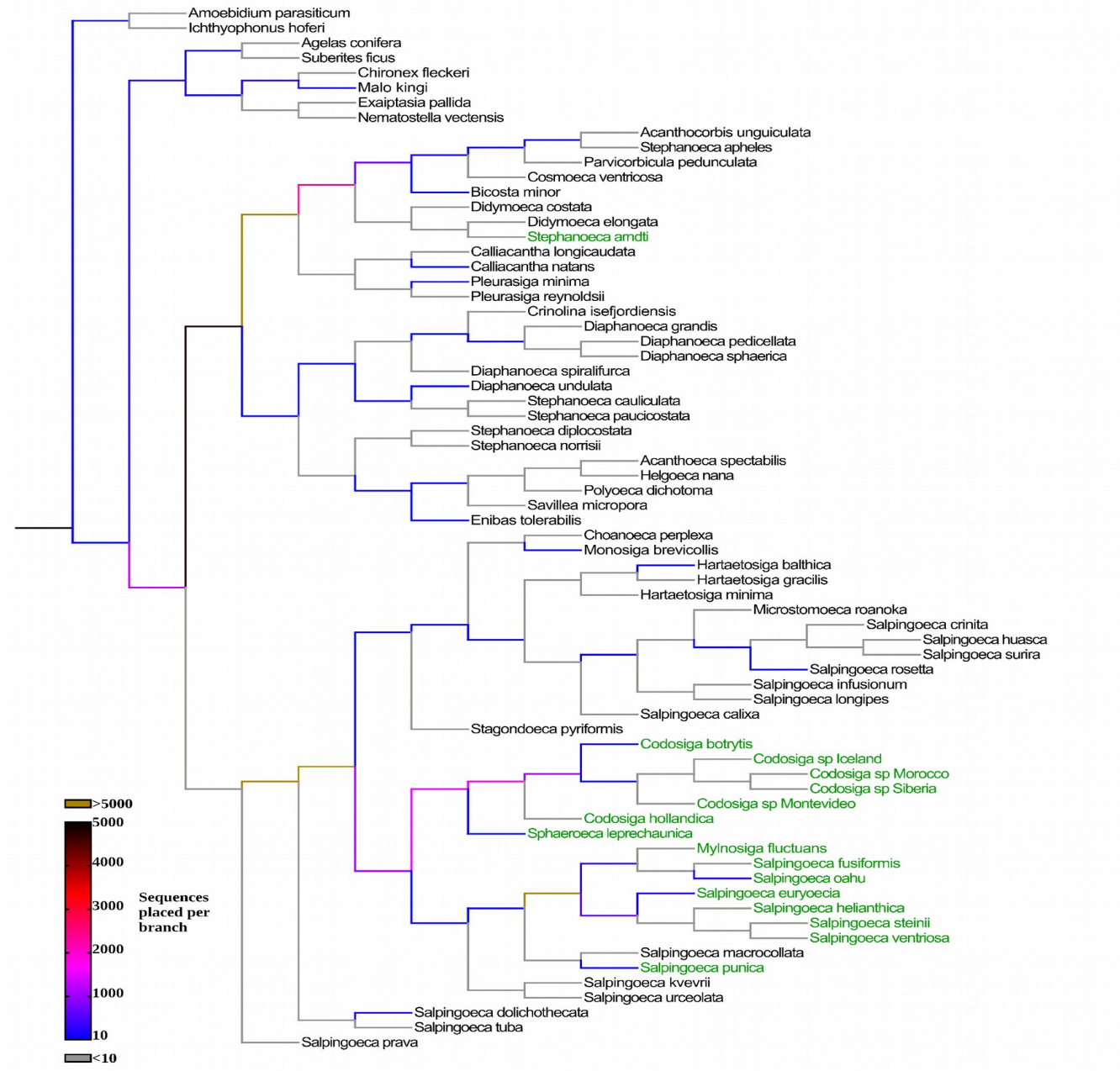

Figure 2.3.2: Heatmap tree of the distribution of environmental sequences in the phylogenetic tree of choanoflagellates. Based on the phylogenetic placement of 67281 SSU eDNA sequences into the phylogenetic tree of choanoflagellates reconstructed, based on the partitioning scheme 4-PartitionsC. Branches are colour-coded to represent the number of query sequences placed on them as follows: <10 placements (grey); gradient 10 placements (blue) to ~2000 placements (magenta), to ~3500 placements (red) to 5000 placements (black); >5000 placements (yellow). Species names in green indicate freshwater species.

Within Craspedida, the branches containing the higher number of placements were the branch leading to the split between Clade 3 and the clade formed by Clade 1 and Clade 2 (>5000 sequences placed), the branch leading to the split between Clade 1 and Clade 2 (>5000 sequences placed), the most basal branch to all freshwater Clade 2 salpingoecids but *S. punica* (>5000 sequences placed) and the two most internal branches within the codosigid lineage (~1500-2000 sequences placed)(Figure 2.3.3). In addition, the most basal branch of Clade 2 craspedids and the most basal branch to all choanoflagellates also contained ~1500 sequences placed on each of them. The remaining branches contained <1000 sequences placed on each of them.

Phylogenetic placement of sequences from freshwater origin.

The dataset contained 145 sequences of samples with confirmed freshwater origin. Out of those, 57.241% (83 sequences) were recovered within Craspedida, 41.379% (60 sequences) were recovered within Acanthoecida and the remaining 1.379% (2 sequences) were recovered within the outgroup (Figure 2.3.4, Table A2.3.1, appendix). Within craspedids, Clade 2 was the most abundant group, representing 71.084% (59 sequences) of the sequences placed within Craspedida, followed by Clade 1 (13.253%, 11 sequences) and the split between Clade 1 and Clade 2 (12.048%, 10 sequences). The least abundant group of craspedids was basal craspedids, which represented 3.861% (3 sequences) of the sequences placed within Craspedida. Within acanthoecids, the placement of environmental sequences from freshwater origin followed a pattern similar to that of sequences from marine origin, with most of the sequences placed at one of the most abundant placement locations (Figure 2.3.5). No sequences were recovered as basal choanoflagellates.

Figure 2.3.3: Placement of environmental sequences within Craspedida. Placement locations of the five main groups of craspedids. The analyses revealed a small hot spot of hidden freshwater diversity basal to Calde 1 (|A) and three main hot spots of hidden diversity within craspedids (|B, |D and |E). Furthermore a small pocket of hidden diversity was also identified basal to Clade 2 craspedids (|C). Branches are colour-coded indicating the different groups of species as follows: outgroup (red), basal choanoflagellates (blue), acanthoecids (magenta), Clade 1 craspedids (cyan), Clade 2 craspedids (yellow), Clade 3 craspedids (light blue). Grey branches indicate environmental sequences. Species names in green indicate freshwater species.

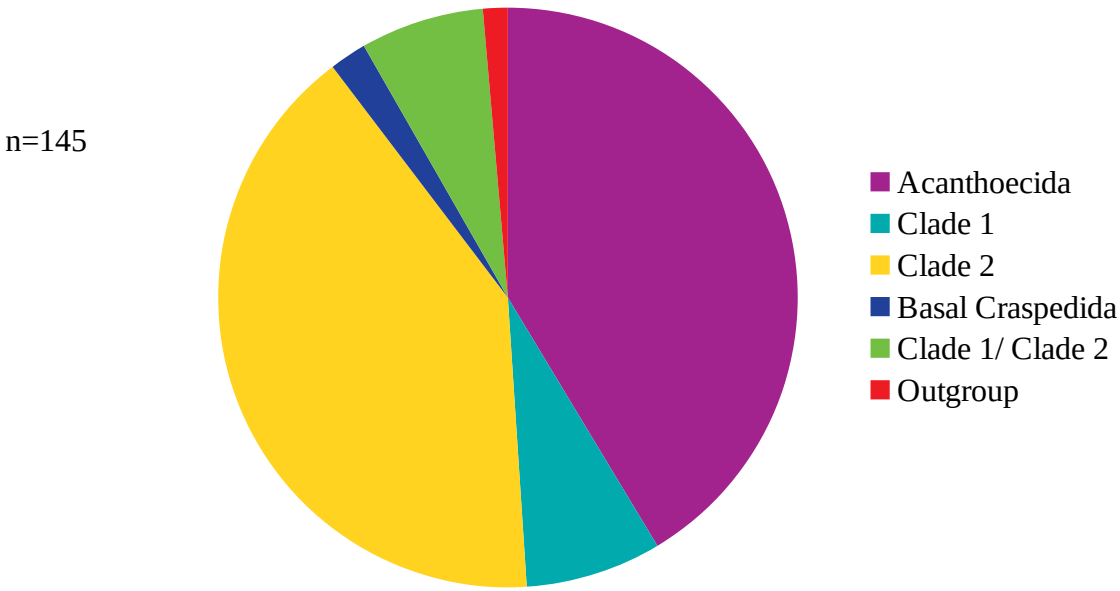

Figure 2.3.4: Placement of sequences from freshwater origin. Group of placement of the 145 sequences with confirmed freshwater origin. Groups are drawn proportional to the number of sequences placed. Colours indicate: Acanthoecida (magenta), Clade 1 craspedids (cyan), Clade 2 craspedids (yellow), basal craspedids (blue), the split between Clade 1 and Clade 2 craspedids (green) and outgroup (red).

Figure 2.3.5: Placement of sequences from freshwater origin within Acanthoecida. Placement locations of the three main groups of freshwater acanthoecids. The distribution of placements location of freshwater samples (current figure) follows the same pattern than marine samples (Figure 2.3.2). Branches are colour-coded indicating the different groups of species as follows: outgroup (red), craspedids (blue), tectiforms (magenta), nudiforms (orange). Grey branches indicate environmental sequences. Species names in green indicate freshwater species.

Placement of sequences from soil samples.

The dataset contained 23 sequences of samples with confirmed soil origin. Out of those, 43.478% (10 sequences) were recovered within Craspedida, 26.087% (6 sequences) were recovered within Acanthoecida, 17.391% (4 sequences) were recovered as basal choanoflagellates and the remaining 13.043% (3 sequences) were recovered within the outgroup (Figure 2.3.6, Table A2.3.2, appendix). Within craspedids, Clade 2 was the most abundant group, representing 80% (8 sequences) of the sequences placed within Craspedida, followed by basal choanoflagellates and split between Clade 1 and Clade 2 (10%, 1 sequence per group).

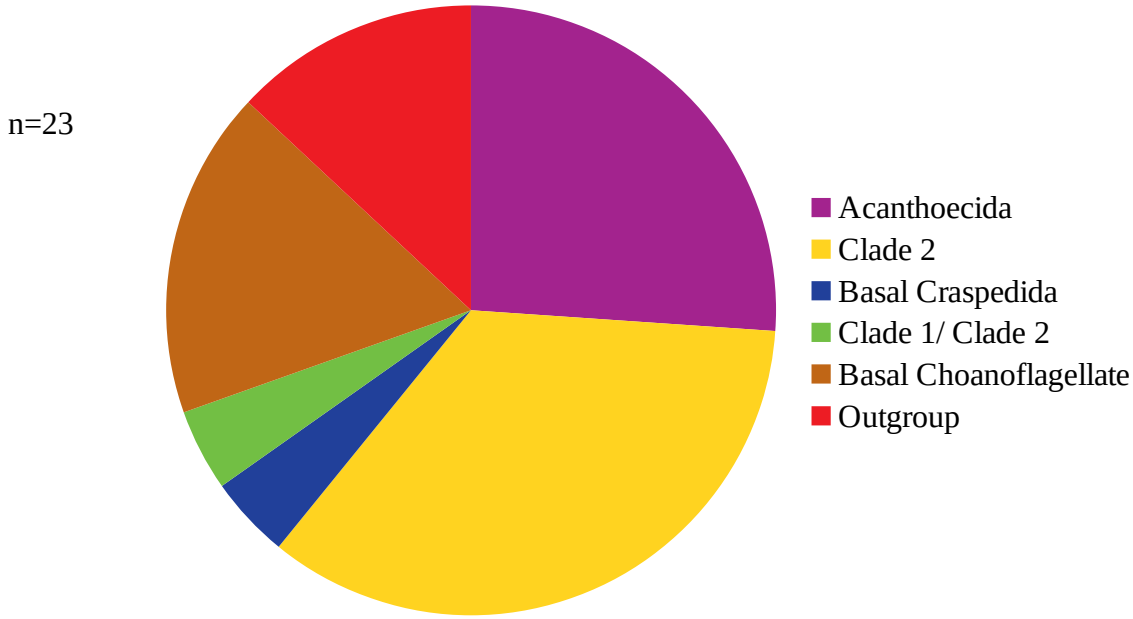

Figure 2.3.6: Placement of sequences from soil samples. Group of placement of the 23 sequences with confirmed soil origin. Groups are drawn proportional to the number of sequences placed. Colours indicate: Acanthoecida (magenta), Clade 2 craspedids (yellow), basal craspedids (blue), the split between Clade 1 and Clade 2 craspedids (green), basal choanoflagellates (brown) and outgroup (red).

Placement of the sequences originally used to define environmental clades of choanoflagellates.

The dataset contained 92 sequences (92%) of the 100 sequences originally used to establish the clades, of freshwater choanoflagellates, composed solely of environmental sequences FRESCHO1 (22 sequences, 88% of the sequences originally used), FRESCHO2 (3 sequences, 60% of the sequences originally used), FRESCHO3 (33 sequences, 100% of the sequences originally used), FRESCHO4 (19 sequences, 86.364% of the sequences originally used), FRESCHO6 (1 sequence, 100% of the sequences originally used) and Clade L (17 sequences, 94.444% of the sequences originally used) (del Campo and Ruiz-Trillo, 2013) (Table A2.3.3, appendix).

21 sequences of FRESCHO1 (95.455%) were recovered as Craspedida, while the remaining sequence was recovered as a basal acanthoecid (Figure 2.3.7-a). From FRESCHO2, 2 sequences (66.667%) were recovered as Clade 2 craspedids and the remaining sequence was recovered as a basal acanthoecid (Figure 2.3.7-b). All but two sequences from FRESCHO3 (31, 93.939%) were recovered as Acanthoecida, while the remaining sequences were recovered as *S. prava* (Figure 2.3.7-c). Within FRESCHO3 acanthoecids, the most abundant group was basal acanthoecids, representing 48.387% (15 sequences) of the sequences placed within Acanthoecida, followed by the split between genera *Pleurasiga* and *Didymoeca* lineages (41.935%, 13 sequences) and the split between genera *Didymoeca* and *Bicosta* lineages (9.677%, 3 sequences). From FRESCHO4, 71.684% (14 sequences) were recovered as Acanthoecida and the remaining sequences were recovered as Craspedida (Figure 2.3.7-d). Within FRESCHO4 acanthoecids, 1 sequence was recovered as a basal acanthoecid and the remaining 13 sequences were recovered at the splits between genera *Pleurasiga* and *Didymoeca* (6 sequences) between genera *Didymoeca* and *Bicosta* lineages (7 sequences). Within FRESCHO4 craspedids, 4 sequences (80%) were recovered as Clade 2 and the remaining sequence was recovered as a basal craspedid. The sequence from FRESCHO6 was recovered as a basal Clade 1 craspedid. All but one Clade L sequences were recovered as basal acanthoecids, while the remaining sequence was recovered at the split between Clade 1 and Clade 2 craspedids (Figure 2.3.7-e).

Discussion.

The overall placement uncertainty, measured as the average EDPL, is less than half the average branch length, (0.0362 and 0.0829, expected number of substitutions per site). Therefore, on average, the placements cannot jump from their branch to the neighbouring branches. However, only 906 placements (1.3466% of the total) have an EDPL=0, and therefore the sequence has only been placed in one branch of the tree. The remaining placements have an EDPL>0, which means that the sequence has been placed at least in two branches. The fact that the average EDPL values are less than half the average branch length and that EDPL<0.0829 for >90% of the sequences, indicates that the sequences are being placed in closely related short branches, rather than scattered all around the tree. Considering this and in order to guarantee that the interpretation of the results was as conservative as possible, the different placements of the each query sequence was accumulated towards the root of the tree until the placement in a particular branch had reached 0.95 likelihood weight ratio.

The phylogenetic placement of environmental sequences within a phylogenetic tree is often regarded as a way of estimating the hidden biodiversity amongst a group of organisms (de Campo and Masana, 2011; del Campo and Ruiz-Trillo, 2013; del Campo *et al.*, 2015; Arroyo *et al.*, 2018; López-Escardó *et al.*, 2018; Venter *et al.*, 2018; del Campo *et al.*, 2019; Mitsi *et al.*, 2019). Such hidden diversity usually indicates the presence of species that either have not been sampled at all, or cryptic species that have been sampled but, due to erroneous identification, have not been included on the phylogenetic reconstruction. When sequences are placed within the internal branches of the tree, it indicates that the tree is missing taxa that have a sister relationship to the branch they have been placed on, while placements on the external branches indicate that a certain species is present in the environmental samples.

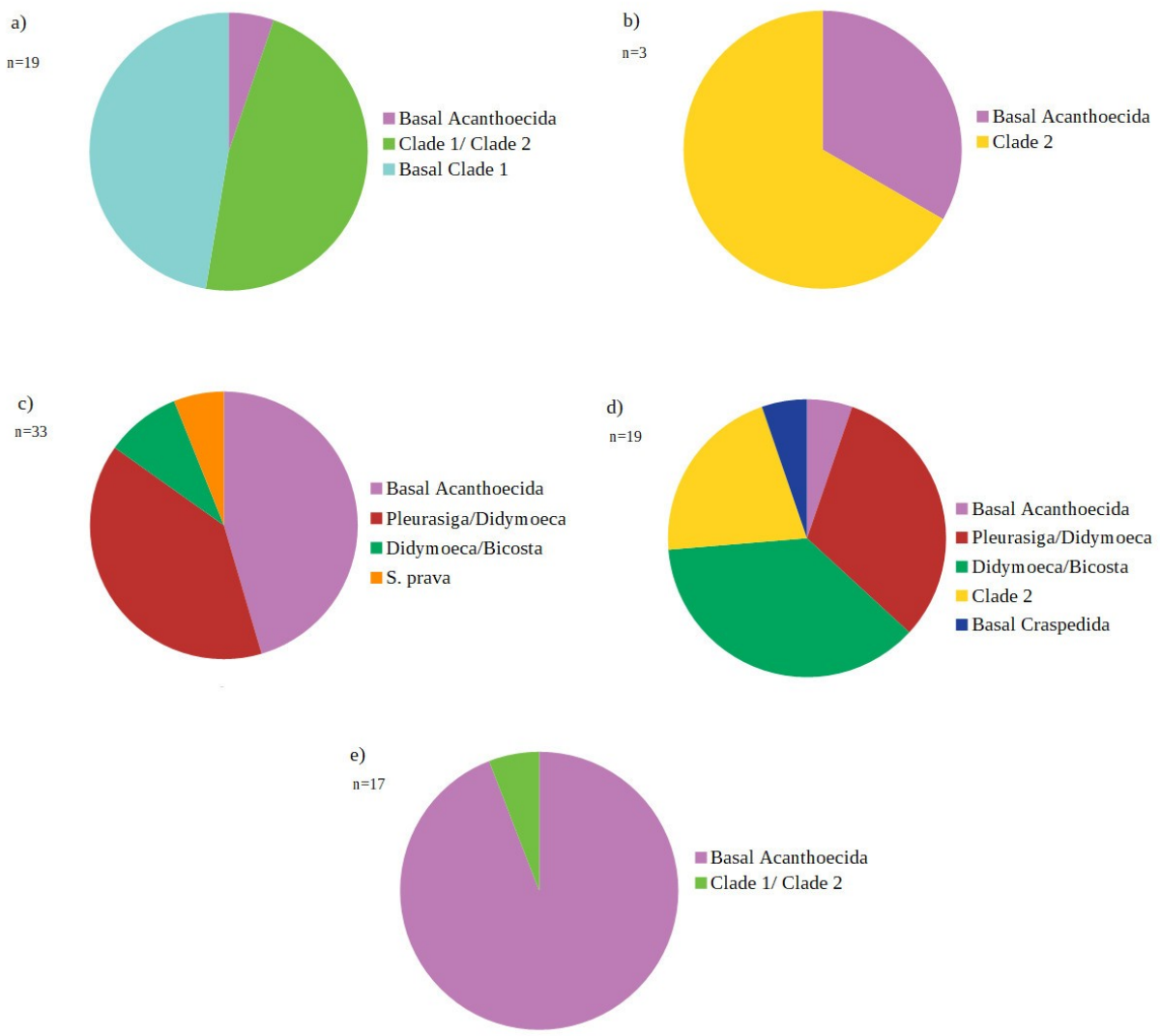

Figure 2.3.7: Placement of the sequences originally used to define environmental clades of choanoflagellates. Group of placement of the 19 sequences of FRESCHO1 (a), 3 sequences of FRESCHO2 (b), 33 sequences of FRESCHO3 (c), 19 sequences of FRESCHO4 (d) and 17 sequences of Clade L (e). Groups are drawn proportional to the number of sequences placed. Colours indicate: Basal acanthoecids (violet), Clade 1 craspedids (cyan), Clade 2 craspedids (yellow), basal craspedids (blue), the split between Clade 1 and Clade 2 craspedids (green), the split between genera *Pleurasiga* and *Didymoeca* lineages (dark red), *Salpingoeca prava* (orange) and the split between genera *Didymoeca* and *Bicosta* lineages (dark green).

Three major hotspots of hidden diversity within choanoflagellates.

Consistently with previously published studies (del Campo and Masana, 2011; del Campo and Ruiz-Trillo, 2013; del Campo *et al.*, 2015; Arroyo *et al.*, 2018; Venter *et al.*, 2018), the results of the phylogenetic placement of eDNA sequences suggests a high hidden biodiversity amongst choanoflagellates. From these results three hotspots of hidden biodiversity amongst choanoflagellates can be inferred.

The first of these hotspots is located basal to Acanthoecida. Such hidden biodiversity amongst the acanthoecids is expected and can be easily explained. Even though the phylogenetic reconstruction carried out in the present study is the most comprehensive, to our knowledge, phylogeny of choanoflagellates up until today, from the >100 species of acanthoecids described it only includes 27 species. Furthermore, since acanthoecids are the most common group of choanoflagellates inhabiting marine environments (del Campo *et al.,* 2015), and since most of the sequences used for this analysis came from the marine surveys carried out by the Tara Oceans expeditions, a high abundance of acanthoecids within this environmental sequences can be expected. Moreover, the analysis of the placement of samples with a confirmed freshwater and soil origin reveal a high abundance of acanthoecids in these samples. Such results are in accordance with previous studies that have estimated a high hidden diversity of freshwater acanthoecids (Arroyo *et al.*, 2018). The highest number sequences placed on a single branch is on the branch leading to the split between the lineages of genera *Pleurasiga* and *Didymoeca* and their closely related species. This indicates the absence of several taxa and maybe even complete linages within or closely related to this clade. The low number of sequences placed around the internal nodes of the nudiforms suggest that this group is indeed composed of very few species. This contrasts with the much higher number of tectiforms both known and estimated. The differences in the diversity of nudiforms and tectiforms have been previously interpreted as a consequence of the adaptative advantages derived from the immediate inheritance, following division, of a lorica by the juvenile cell in tectiforms (Leadbeater, 2010).

The second hotspot is located on the branches leading to the split between Clade 3 and the clade formed by Clade 1 and Clade 2 and to the split between Clade 1 and Clade 2, within Craspedida. Such hidden biodiversity suggests the underrepresentation of the lineages at the at the most basal positions within the craspedids. The high number of sequences placed in the branch sister to *S. prava* supports the interpretation of the phylogenetic reconstructions (previously discussed in the first section of this chapter) that *S. prava* is indeed an early branching craspedid. As previously stated query sequences that contained multiple placements were accumulated towards the root until a threshold of 0.95 likelihood weight ratio was reached. If the placement of *S. prava* was incorrect, query sequences related to this species would have multiple placements around the branch of the genuine position of *S. prava* due to sequence similarity with the acanthoecids and, therefore, this accumulation would have led to most of this sequences being placed deeper into the tree, on the branch basal to all choanoflagellates. On the other hand, the high number of sequences placed in the branch leading to the splits of Clade 1 and Clade 2 craspedids, rather than in the branch leading to the split of *S. dolichothecata* and *S. tuba*, suggests that such hidden biodiversity is more closely related to Clade 1 and Clade 2 than to Clade 3. In order for a sequence to be placed in this branch, it either represents a species that is indeed basal to both Clade 1 and Clade 2, or it represents a species that has been placed in both Clade 1 and Clade 2 at the same time and later accumulated towards this branch. In either case, that suggests the absence of several taxa and maybe even complete lineages related to both Clade 1 and Clade 2, which could affect the current reconstructed relationship between both groups. At the same time, the low number of sequences placed in the branch leading to the split between *S. dolichothecata* and *S. tuba* suggests that Clade 3 is most likely an early branching lineage composed of two to a very few species. It could be argued that this might be an artefact of the sampling and that many low-abundance species could still be present within this group, however the sampling methodology of Tara Oceans comprised sampling operations in over 200 sampling stations around the world at three different depths during different periods of the year. Moreover the sequences retrieved from SILVA database comprised many different studies around the world and not restricted to marine environments. Thus, the lack of sequences related to Clade 3 is here interpreted as the result of the groups low diversity.

The third hotspot is located in the branch basal to all freshwater Clade 2 salpingoecids but *S. punica*. Since most of the sequences included in the analysis come from marine environments, such hidden diversity suggests the existence of marine taxa more closely related to this lineage than to *S.*

macrocollata and *S. punica*. This results further support the interpretation, exposed in the previous section of this chapter, that, the freshwater colonisation event that led to the diversification of this lineage is independent from the event that led to the diversification of *S. punica*. Moreover, the moderately high number of sequences placed on the internal branches of the codosigids suggest the existence of marine taxa more closely related to this group than to the rest of the Clade 2 species. Such results are in congruence with the same interpretation, since they further support that the freshwater colonisation event that led to the diversification of the codosigids is independent from the two events that led to the diversification of *S. punica* and the rest of the freshwater Clade 2 salpingoecids.

Finally, the number of sequences placed on the basal branches of both Clade 2 and all the choanoflagellates suggest that a low hidden diversity is exist around both branches. Nevertheless this hidden diversity is much lower than that of the above discussed hotspots, which indicates that most of the absent taxa belong to any of the three hotspots.

Hidden diversity of freshwater choanoflagellates.

The results of the phylogenetic placement of environmental sequences revealed a that most of the diversity of freshwater choanoflagellates was related to Acanthoecida and Clade 2 craspedids. This results are in partial agreement with previous studies (Arroyo *et al.*, 2018) that have estimated a high diversity of freshwater acanthoecids from freshwater samples. Such result would usually be interpreted as estimating a high diversity of FRESCHO 3, FRESCHO 4 and Clade L diversity. However, a detailed analysis of such placements (further discussed below) revealed results that are conflicting with these freshwater environmental clades as defined by del Campo and Ruiz-Trillo (2013). The high proportion of sequences related to Clade 2 matched the expectations since most of the known freshwater choanoflagellates are contained in this clade. The placement of sequences basal to Clade 1 suggests the existence of unknown freshwater species related to this clade, probably as an early branching radiation within Clade 1. This findings are in concordance with the environmental clade FRESCHO 1 defined by del Campo and Ruiz-Trillo (2013) (further discussed below). The placements of freshwater sequences at the split between Clade 1 and Clade 2 and as basal craspedids are in accordance with the second hotspot of hidden diversity within choanoflagellates above described.

Hidden diversity of soil choanoflagellates.

The result of the phylogenetic placement of environmental sequences revealed that most of the diversity of choanoflagellates in soil samples is related to Craspedida. Nevertheless, this results suggest the presence of acanthoecids in these same samples. Most of the placements are related to one of the three major hotspots of hidden diversity above discussed. However, a fourth group of basal choanoflagellates can be identified in this placements. This group is contains four sequences that were recovered at the split between Craspedida and Acanthoecida. There is no sufficient data to discuss whether these sequences represent a genuine clade of soil choanoflagellates that diverged from the rest of choanoflagellates before the split of Craspedida and Acanthoecida or it represents a set of highly divergent sequences that have been placed in this location as a result of analytical artefacts. Previous studies that have analysed the diversity of choanoflagellates in soil samples (Venter *et al.*, 2018) did not include Acanthoecida in their analysis and therefore the results of the analysis of the present study cannot be fully compared. The presence of sequences related to Acanthoecida and to basal choanoflagellates in soil samples falls beyond the scope of the present study and require additional investigation.

Phylogenetic placement of the sequences originally used to define environmental clades of choanoflagellates.

The phylogenetic placement of the sequences originally used to define environmental clades of choanoflagellates revealed several inconsistencies across the clades FRESCHO 2, FRESCHO 3 and FRESCHO 4.

Firstly, the sequences used to define FRESCHO 3, originally defined as basal to all choanoflagellates (del Campo and Ruiz-Trillo, 2013) and subsequently moved to a position basal to Acanthoecida (Carr *et al.*, 2017), are placed at three different locations within Acanthoecida. Although basal acanthoecids represents the most abundant placement position of this group (15 sequences), an almost as abundant (13 sequences) placement position is located at the split between the lineages of genera *Pleurasiga* and *Didymoeca*, and their related species. Moreover a less abundant (3 sequences) third placement position

is located at the split between the lineages of genera *Didymoeca* and *Bicosta*, and their related species. Finally, a last placement position (2 sequences) is located at *Salpingoeca prava*. These results indicate that more than half of the sequences originally used to define FRESCHO 3 are not basal to Acanthoecida, but rather a part of the acanthoecid hotspot of hidden diversity above described. Moreover, the inclusion of sequences of *S. prava*, as a basal craspedid (as discussed in Section I of the present chapter), indicates the polyphyly of this group. This has deep implications when using this groups as a reference for phylogenetic reconstructions based on OTUs. Since OTUs are constructed based on sequence similarity, the polyphyly of this group can lead to the misplacement of sequences related to *S. prava* and other potential basal craspedids as related to Acanthoecida instead of as related to Craspedida. Moreover, due to the presence of sequences related to the splits between the lineages of genera *Pleurasiga* and *Didymoeca* and between the lineages of genera *Didymoeca* and *Bicosta* within FRESCHO3, phylogenetic reconstructions based upon OTUs constructed by sequence similarity to this clade can potentially lead to the underestimation of the diversity within Acanthoecida.

Secondly, the sequences used to define FRESCHO 4, originally defined as basal to Acanthoecida, subsequently proved polyphyletic (Carr *et al.*, 2017), and finally reduced to several species basal to Acanthoecida, are placed at three different locations within Acanthoecida. Similarly to FRESCHO 3 these locations correspond to the split between the lineages of genera *Didymoeca* and *Bicosta* (7 sequences), the split between the lineages of genera *Pleurasiga* and *Didymoeca* (6 sequences) and as basal to Acanthoecida (1 sequence). As in the case of FRESCHO 3, the results indicate that most of the sequences, related to Acanthoecida, originally used to define FRESCHO 4 are not basal to Acanthoecida, but rather a part of the acanthoecid hotspot of hidden diversity above described. Similarly to FRESCHO3, the use of this clade in the construction of OTUs can potentially lead to the underestimation of the diversity within Acanthoecida.

In addition to the inconsistencies above described, and in accordance with the revision of the freshwater clades from environmental sequences (Carr *et al.*, 2017), the phylogenetic placement of the sequences used to define FRESCHO 2 revealed the polyphyly of this clade. Although the number of sequences included in the dataset was low (3 sequences), the phylogenetic placement recovered the clade in three locations within the phylogenetic tree. Consistently with Carr *et al.* (2017), one sequence

was recovered as a basal acanthoecid, while the other two were recovered as Clade 2. From the later, one sequence was recovered as basal to codosigids and the other as basal to freshwater Clade 2 salpingoecids. These results indicate that not only the group is polyphyletic, due to the presence of sequences related to both Craspedida and Acanthoecida, but also that when only sequences related to Clade 2 are considered it contains sequences related to both major groups within the clade (salpingoecids and codosigids). As a result of the phylogenetic diversity contained within this clade, its use in analysis based on sequence similarity can potentially lead to the underestimation of the diversity within Clade 2.

Minor inconsistencies were also detected in the environmental clades FRESCHO 1 and Clade L. In both cases, the presence of a single sequence broke the monophyly of the clades. In the case of FRESCHO 1, the sequences of this clade were, with the exception of one sequence, recovered in two locations. The first location corresponds to basal to the split between Clade 1 and Clade 2 craspedids (9 sequences) and the second location as basal to Clade 1 craspedids (9 sequences). The first location corresponds to the second hotspot of hidden diversity above described, while the second corresponds to the original placement of FRESCHO 1 (del Campo and Ruiz-Trillo, 2013). Consistently with previous interpretations of the phylogenetic position of FRESCHO 1 (del Campo and Ruiz-Trillo, 2013; Carr *et al.,* 2017; Arroyo *et al.*, 2018), these results suggests the existence of a group of freshwater choanoflagellates that diverged early in the evolution of Clade 1 as a consequence of a marine-to-freshwater transition, following a pattern similar to that occurred within Clade 2 and that led to the origin of the codosigids. Moreover, the results also suggest the presence of freshwater choanoflagellates closely related to the split between Clade 1 and Clade 2. Such results, in combination with the presence of a hidden diversity hotspot around the same locations of the tree suggests that marine:freshwater transitions around Clade 1 and Clade 2 might be much a more common phenomenon than previously thought. It is here hypothesised that the inclusion of new taxa around this part of the phylogeny of choanoflagellates might have a deep impact over the current view of the phylogenetic relationship between both clades.

Similarly to the case of FRESCHO 1, all sequences of Clade L, with the exception of one sequence, were recovered as basal to Acanthoecida. The results from the phylogenetic placements, regardless of

the origin of the sample, have consistently recovered sequences placed at this location. In congruence with previous studies (del Campo *et al.*, 2015; Arroyo *et al.*, 2018), these results in combination with the first hotspot of hidden diversity above described, suggest that most of the diversity of Acanthoecida remains poorly explored. It is here hypothesised that the inclusion of new taxa around this part of the phylogeny of choanoflagellates might have a deep impact over the current view of the phylogenetic relationship between nudiforms and tectiforms. In this regard, the inclusion of highly conserved protein-coding genes from species placed around this hotspot of hidden diversity might help on resolving the currently conflicting topologies of Acanthoecida.

Conclusions.

The results here exposed identify three major hotspots of hidden diversity within choanoflagellates, related to basal positions of Acanthoecida, Craspedida and the freshwater salpingoecids. The hotspots related to basal positions of Acanthoecida and Craspedida indicate that the deepest branches of the choanoflagellates phylogeny remain mostly unexplored. This is especially noticeable in Acanthoecida, which in addition to the hidden diversity of marine choanoflagellates shows high levels of hidden diversity in both freshwater and soil samples. Similarly, positions basal to Craspedida, especially around the split between Clade 1 and Clade 2, show high levels of hidden diversity both in freshwater and marine environments, indicating that marine:freshwater transitions, in this section of the phylogenetic tree, might be a much more frequent phenomena than previously thought. Moreover, the high levels of hidden diversity of marine choanoflagellates within Clade 2 indicates the existence of numerous marine species related to this clade. Such results further support the hypotheses that multiple independent freshwater colonisation events have taken place throughout the evolutionary history of Clade 2 craspedids. In addition to this, the phylogenetic placement of the sequences originally used to define environmental clades of choanoflagellates revealed several inconsistencies across the clades FRESCHO 2, FRESCHO 3 and FRESCHO 4. As a result of such inconsistencies, it is here recommended that the use of this clades in the analysis of environmental sequences, through phylogenetic reconstructions based upon OTUs, is discontinued. Such analyses can potentially lead to the misplacement of sequences related to *S. prava* and other potential basal craspedids and to the underestimation of the diversity of choanoflagellates within Acanthoecida and Clade 2. In this regard,

there are not enough evidences to recommend the discontinuance of the use of environmental clades FRESCHO 1 and Clade L, nevertheless due to the high hidden diversity estimated around the location of both groups, further addition of new taxa to the phylogeny of choanoflagellates might require the split of this clades into smaller subclades.

Chapter III: The evolutionary history of marine:freshwater transitions in Choanoflagellates from an ecological perspective.

-Section I: Field samplings, samples processing, choanoflagellates isolation and cultures maintenance.

Introduction.

Choanoflagellates are widely distributed and can be found in most aquatic habitats. Within the approximately around 300 described species, representatives can be found living under a wide range of salinities. Despite some of the first descriptions of choanoflagellates included freshwater representatives, as originally described by James-Clark (1867), and despite the efforts in recent years on the study of choanoflagellates, only a handful of freshwater species have been described up to date. Most of these freshwater species, as shown in the Chapter II of the present study, are closely related.

Freshwater habitats, therefore, represent an underexplored source for new choanoflagellates species. In an attempt to increase the existing knowledge on freshwater choanoflagellates, throughout the course of the present study several sampling campaigns were carried out, focusing on freshwater habitats. Moreover, additional efforts were taken in order to include in these sampling campaigns other underrepresented habitats such as hypersaline environments, brackish waters, hot spring, hyperalkaline environments and acidic environments.

This section describes all sampling campaigns that were carried out, as well as the methodologies used to process such samples, the isolation techniques used and the culturing media production protocols used to isolate, maintain and carry out several experiments in choanoflagellates.

Material and methods.

Artificial media production for choanoflagellates cultures.

For the culture of choanoflagellates in the laboratory several different artificial culture media were produced. Two basic culture media were produced, WC media (Guillard and Lorenzen, 1972) and artificial sea water, for freshwater and marine choanoflagellates respectively. From further on and unless otherwise stated, artificial media refers to these two media depending on whether the choanoflagellates cultured are considered freshwater or marine.

WC media was prepared, in Schott bottles, in 1000ml batches and subsequently stored at 4ºC (a full description of the protocol can be found in Protocol A3.1.1, appendix). Artificial sea water (ASW) media was produced by the addition of 30g/l Instant Ocean® sea salt mix (Instant Ocean, USA) to utrapure MILLI-Q® water (Merc KgaA, Germany). The salinity of the media was adjusted to 30ppt by the addition of NaCl and the pH was adjusted to pH8.0. After mixture, ASW media was sterilised and stored at 4ºC (the typical salt composition of Instant Ocean® sea salt mix can be found in Table A3.1.1, appendix).

These basic culture media were further modified to adapt them to the specific requirements of each of the choanoflagellates cultured (Table 3.1.1). For each basic media a Nutrients Supplement was produced by the addition of 1g/l wheat grass to the media. After sterilization, the media was filtered through 30µm ø (Whatman™ Qualitative Filter Papers Grade 113, GE Healthcare Life Sciences) filter paper to remove the excess of particulate organic matter and the filtrate was sterilised again by autoclaving. Supplemented media was produced by the replacement of 10ml/l of utrapure MILLI-Q® water for the same volume of Nutrients Supplement during the production the media. For the use culture of freshwater choanoflagellates originally sampled from the hyperalkaline (Hanson Aggregates 1-2) and acidic (Strensall Common 1-5) water bodies and the subsequent pH tolerance experiments, several pH-adjusted versions of WC media were produced by the addition of NaOH or HCl, for basic

and acidic pH adjustment respectively, to the original media before sterilisation. For the salinity experiments several salinity-adjusted versions of ASW media were produced by either dilution on WC media or addition of NaCl, for lower and higher salinity levels respectively, before sterilisation.

Table 3.1.1: Summary of the modified versions of artificial culture media prepared during the present study. WC indicates WC media (Guillard and Lorenzen, 1972), ASW indicates artificial sea water prepared with Instant Ocean® sea salt mix, plus sign (+) indicates media supplemented with wheat grass a carbon source, asterisk (*) indicate both supplemented and non-supplemented versions of the media were produced.

Modified media name	Basic Media	Modification
WC+ nutrients supplement	WC	Addition of 1g/l wheat grass to WC media
WC+ Media	WC	Addition of 10ml/l of WC+ nutrients supplement to WC media
WC pH3.4 Media*	WC	Adjusted to pH3.4 by the addition of HCl to the original media.
WC pH4.3 Media*	WC	Adjusted to pH4.3 by the addition of HCl to the original media.
WC pH5.2 Media*	WC	Adjusted to pH5.2 by the addition of HCl to the original media.
WC pH6.5 Media*	WC	Adjusted to pH6.5 by the addition of HCl to the original media.
WC pH9 Media*	WC	Adjusted to pH9 by the addition of NaOH to the original media.
WC pH10 Media*	WC	Adjusted to pH10 by the addition of NaOH to the original media.
WC pH10.5 Media*	WC	Adjusted to pH10.5 by the addition of NaOH to the original media.
WC pH11 Media*	WC	Adjusted to pH11 by the addition of NaOH to the original media.
ASW+ nutrients supplement	ASW	Addition of 1g/l wheat grass to ASW media
ASW+ Media	ASW	Addition of 10ml/l of ASW+ nutrients supplement to ASW media
ASW5 Media*	ASW	Salinity adjusted to 5 PSU by dilution of ASW media on WC media.
ASW10 Media*	ASW	Salinity adjusted to 10 PSU by dilution of ASW media on WC media.
ASW15 Media*	ASW	Salinity adjusted to 15 PSU by dilution of ASW media on WC media.
ASW20 Media*	ASW	Salinity adjusted to 20 PSU by dilution of ASW media on WC media.
ASW25 Media*	ASW	Salinity adjusted to 25 PSU by dilution of ASW media on WC media.
ASW35 Media*	ASW	Salinity adjusted to 35 PSU by addition of NaCl to ASW media.
ASW40 Media*	ASW	Salinity adjusted to 40 PSU by addition of NaCl to ASW media.
ASW50 Media*	ASW	Salinity adjusted to 50 PSU by addition of NaCl to ASW media.
ASW60 Media*	ASW	Salinity adjusted to 60 PSU by addition of NaCl to ASW media.
ASW70 Media*	ASW	Salinity adjusted to 70 PSU by addition of NaCl to ASW media.
ASW100 Media*	ASW	Salinity adjusted to 100 PSU by addition of NaCl to ASW media.

In order to minimise the risk of cross-contamination, for every media, both basic and modified versions, separated batches were produced to work under wet laboratory (samples processing and choanoflagellates isolations) and aseptic conditions (culture maintenance, salinity tolerance experiments and pH tolerance experiments).

Field Sampling.

In an effort to isolate new choanoflagellates species, several field samplings were carried out at multiple locations throughout the present study (Table 3.1.2). Sampling sites were selected as to include a wide range of both salinities and pH. For such samplings, variable volumes of water and sediments were collected in falcon tubes and plastic bottles. Sediments were collected from the bottom or the walls of the ponds by scrapping the surface with the containers. When plastic non-sterile containers were used, containers were disinfected with a solution of commercial bleach (~5% NaOCl). As recommended by Laramie *et al.* (2015), bottles were rinsed 3 times with commercial bleach over a period of 10 min, then rinsed with distilled water 3 times and another 3 times with sample water on site. Such procedure avoids both the contamination of the water sample with organisms that are not genuinely present in the water body and the removal of the disinfecting agent. In addition to the samplings focused on the isolation of choanoflagellates, a sampling campaign was carried out to obtain samples for eDNA extraction and 18S metagenome sequencing. For this purpose, four 4 litres samples were taken from the shore of the water body. Two of the samples were taken from the top part of the water column and two from the bottom part of the water column, including part of the substrate.

Table 3.1.2: Summary of the sampling sites. Localities and the GPS coordinates (indicated between brackets), type of water body. Freshwater indicate salinities of ~0 PSU, brackish water indicates salinities of ~10 PSU and hypersaline indicate salinities of >70 PSU. Neutral indicates pH6.5-pH8.5, acidic indicates pH3.4-pH5.2 and hyperalkaline indicates pH10.2-pH11.3. eDNA indicates whether or not samples for eDNA extraction were taken.

Sampling site	Locality (GPS)	Type	Salinity	pH	eDNA
Greenhead Park	Huddersfield, UK (53.648, -1.796)	Permanent artificial pond	Freshwater	Neutral	No
Gledholt Woods 1	Huddersfield, UK (53.645, -1.800)	Temporary pond	Freshwater	Neutral	No
Gledholt Woods 2	Huddersfield, UK (53.645, -1.800)	Permanent artificial pond	Freshwater	Neutral	No
CP Creek	Huddersfield, UK (53.639, -1.815)	Temporary creek	Freshwater	Neutral	No
River Colne 1	Huddersfield, UK (53.639, -1.815)	River	Freshwater	Neutral	No
River Colne 2	Marsden, UK (53.601, -1.927)	River	Freshwater	Neutral	No
River Colne 3	Huddersfield, UK (53.643, -1.772)	River	Freshwater	Neutral	No
Canal	Huddersfield, UK (53.642, -1.776)	Canal	Freshwater	Neutral	No
Blackmoorfoot reservoir	Blackmoorfoot, UK (53.608, -1.849)	Permanent artificial pond	Freshwater	Neutral	No
Torrevieja Lagoon	Torrevieja, Spain (30.020, -0.718)	Seminatural Coastal Lagoon	Hypersaline	Neutral	Yes
La Mata Lagoon	Torrevieja, Spain (38.022, -0.667)	Seminatural Coastal Lagoon	Hypersaline	Neutral	Yes
Font Salada	Oliva, Spain (38.887, -0.079)	Hot spring	Brackish water	Neutral	Yes
Beniarrés reservoir	Beniarrés, Spain (38.808, -0.352)	Permanent artificial lake	Freshwater	Neutral	Yes

Sampling site	Locality (GPS)	Type	Salinity	pH	eDNA
Tibi reservoir	Tibi, Spain (38.503, -0.560)	Permanent artificial lake	Freshwater	Neutral	Yes
Buttermere Lake	Buttermere, UK (54.536, -3.274)	Mountain Lake	Freshwater	Neutral	Yes
Crummock Water Lake	Buttermere, UK (54.025, -3.287)	Mountain Lake	Freshwater	Neutral	Yes
Stoney Middleton	Stoney Middleton, UK (53.276, -1.653)	Hot spring	Freshwater	Neutral	No
Cologne University	Cologne, Germany (50.925, 6.936)	Permanent artificial pond	Freshwater	Neutral	No
Hanson Aggregates 1	Settle, UK (54.146,-2.303)	Permanent artificial pond	Freshwater	Hyperalkaline	No
Hanson Aggregates 2	Settle, UK (54.143,-2.306)	Permanent artificial pond	Freshwater	Hyperalkaline	No
Hanson Aggregates 3	Settle, UK (54.143,-2.306)	Permanent artificial pond	Freshwater	Neutral	No
Strensall Common 1 (POND 4)	York, UK (54.038, -1.000)	Permanent pond	Freshwater	Acidic	No
Strensall Common 2 (POND 8)	York, UK (54.036, -0.993)	Permanent pond	Freshwater	Acidic	No
Strensall Common 3 (POND 9)	York, UK (54.036, -0.990)	Permanent pond	Freshwater	Acidic	No
Strensall Common 4 (POND 10)	York, UK (54.035, -0.989)	Permanent pond	Freshwater	Acidic	No
Strensall Common 5 (DOG POND)	York, UK (54.041, -1.009)	Permanent pond	Freshwater	Acidic	No

Samples processing.

Sample processing for eDNA extraction.

Upon arrival to the laboratory, samples were stored at 4ºC and processed within 24h. Processing consisted on two filtrations steps, followed by DNA extraction from the filters. On the first step, water was filtrated through a 30µm ø (pore diameter) filter paper. This step retains most of the particulate matter and substrate to which choanoflagellates can be attached and allow the easy removal of bigger sediments, mainly stones, that might cause problems in further steps. On the second step, water was filtrated through a 0.2nm ø filter paper (Whatman™ Qualitative Filter Papers Grade 602H GE Healthcare Life Sciences) in order to retain the free swimming organisms present in the water that have not been retained in the previous filtering step. In order to facilitate and speed-up the processing of the samples, the filtration was assisted by a vacuum pump. Subsamples were poured in a Büchner funnel containing the filter papers and filtrate was collected in 2l Büchner flasks. After the second filtration, the filtrate was discarded and the filter papers were used to extract DNA following the protocol proposed by Kaevska and Slana (2015). The protocol consists on four steps, namely: elution of the filter content in PBS buffer with 0.2% Tween 80, addition of glass beads and vortexing for 5 minutes, centrifugation at 7000x g for 10 minutes and extraction of DNA using using DNeasy Blood & Tissue kit (QIAGEN, Germany). After extraction, eDNA was quantified by spectrophotometry using Nanodrop 2000 (Thermo Fisher Scientific, USA) and stored at 4ºC for short term storage or -20ºC for long term storage.

Sample processing for choanoflagellates isolation.

Upon arrival to the laboratory, samples were stored at room temperature and protected from light. Within the following 24h samples were split in two halves. The first half was used for choanoflagellates isolation and the second half was treated to obtain culturing media.

Sample processing for culturing media.

Subsamples for culturing media were subjected to a 2-step filtering process. On the first step samples are filtered through 30µm ø filter paper to remove the higher sized fraction of organic and inorganic particulate matter as well as the biggest fractions of planktonic organisms (macro-, meso- and most microplankton). Subsequently, subsamples were filtered through <2µm ø filter paper (Whatman™ Qualitative Filter Papers Grade 602H GE Healthcare Life Sciences) to remove most of the particulate organic and inorganic matter as well as the remaining micro- and nanoplanktonic organisms. In order to facilitate and speed-up the processing of the subsamples, the filtration was assisted by a vacuum pump. Subsamples were poured in a Büchner funnel containing the filter papers and filtrate was collected in 2l Büchner flasks. After filtration, subsamples were split in two halves. One half was autoclaved and stored at 4ºC for use under sterility conditions. The other half was stored at room temperature for use under wet-lab conditions.

This process allows the obtention of culturing media that maintains the physico-chemical conditions of the field, thus allowing to replicate to a great extent the environmental conditions happening in the field at the time of the sampling. On one hand, unsterilised media contains the same femto- and picoplanktonic organisms that are present in the field at the time of sampling, thus avoiding the need for nutrient and food supplementing. Since the relationship between this organisms and choanoflagellates is not limited to predator-prey relationships, but also can have implications in processes such as colony formation and sexual reproduction (Alegado *et al.*, 2012; Levin *et al.*, 2014; McFall-Ngai, 2014; Woznica *et al.*, 2016; Woznica *et al.*, 2017; Booth *et al.*, 2018; Hake, 2019) the presence of this organisms can affect the viability of the cultures in ways that are unknown or have not been accounted for. On the other hand, sterilised media is less prone to changes in the nutrients balance and, generally speaking, the physico-chemical conditions of the media, as most organisms have been filtered out or inactivated.

Sample processing for choanoflagellates isolation.

Subsamples for choanoflagellates isolation were kept at either at room temperature or at 4ºC depending on the temperature of the water body they were sampled from. Subsamples were split into several Petri dishes. After a settling period of 24h, Petri dishes containing the sample were slightly agitated by hand and carefully washed and refreshed with unsterilised culture media prepared as described above. This process led to the establishment of mixed cultures of prokaryotes and eukaryotes present in the water samples. In order to maximise the probability of cultures establishment and due to the high turbidity of some samples, a subset of the cultures of each sample were refreshed with artificial media instead of unsterilised filtered water media. Amongst all the above-described culture media, the medium employed to refresh each culture was the one with the most similar conditions to the field conditions (Table 3.1.3).

Table 3.1.3: Summary of the artificial culture media (as described in Table 3.1.1) employed to refresh mixed cultures and isolations form each sampling site.

Sampling site	Culture Media
Greenhead Park	WC+
Gledholt Woods 1	WC+
Gledholt Woods 2	WC+
CP Creek	WC+
River Colne 1	WC+
River Colne 2	WC+
River Colne 3	WC+
Canal	WC+
Blackmoorfoot reservoir	WC+
Torrevieja Lagoon	ASW+100
La Mata Lagoon	ASW+70
Font Salada	ASW+10
Beniarrés reservoir	WC+
Tibi reservoir	WC+
Buttermere Lake	WC+
Crummock Water Lake	WC+
Stoney Middleton	WC+

Sampling site	Culture Media
Cologne University	WC+
Hanson Aggregates 1	WC+ pH9
Hanson Aggregates 2	WC+ pH9
Hanson Aggregates 3	WC+
Strensall Common 1	WC+ pH3.4
Strensall Common 2	WC+ pH4.3
Strensall Common 3	WC+ pH4.3
Strensall Common 4	WC+ pH4.3
Strensall Common 5	WC+ pH5.2

After a settling period of ~2h, cultures were screened by direct observation under an inverted microscope at 600x magnification and isolation was attempted as described below.

Isolation and establishment of new choanoflagellates cultures from environmental samples.

Screenings were carried out for periods of varying time (up to 3 months) depending on the presence of choanoflagellates on the culture. Cultures were refreshed, with unsterilised media or artificial media accordingly, every 1-3 weeks depending on the total abundance of organisms in the culture and the overall growth of the culture established by direct observation.

On those mixed cultures where choanoflagellates were observed isolation by single-cell picking was attempted. In order to do so, glass capillaries (Hirschmann Laborgeräte, Germany) were heated and pulled to produce a tip of a diameter slightly bigger that the cell size of the choanoflagellates observed. Capillaries were then attached to a CellTram® 4r Oil, manual hydraulic microinjector (Eppendorf, Germany) mounted on a Burleigh PCS-5000 micromanipulator (EXFO Photonic Solutions Inc, Canada) and used to pick single choanoflagellates cells. Each cell was then transferred to a 15ml Nunc™ EasYFlask™ culture flask (Thermo Fisher) containing the appropriated culture media, either sterilised filtered water or artificial media. Culture flasks were stored at room temperature.

After a period of incubation of 1 week, cultures were screened for the presence of choanoflagellates by direct observation under an inverted microscope. Cultures where the presence of choanoflagellates could not be confirmed by direct observation where incubated for another week. Cultures that underwent three unsuccessful rounds of incubation and screening were considered as failed and discarded. Cultures where only choanoflagellates were detected were deemed as successful isolations and treated as such (see description bellow). Cultures where choanoflagellates were present alongside other organisms were incubated for varying periods of time (up to 3 months) and screened by direct observation on a weekly basis. Further isolation rounds were attempted by either serial dilutions, single-cell picking or a combination of both strategies.

The first approach, serial dilutions, was used when mixed cultures exhibited low to medium abundances of choanoflagellates and a similar or higher abundance of non-choanoflagellates organisms. In these cases, cultures were incubated for 1-3 weeks, culture flasks were then vigorously shaken by hand and flask walls were scrapped, culture media was discarded to eliminate as many of the free-swimming organisms present in the culture as possible, refreshed with fresh media and subsequently split into 3-5 culture flasks (depending on choanoflagellates abundance determined by direct observation). After one week of incubation, cultures were screened again by direct observation under the microscope and, if non-choanoflagellates organisms were still present in the culture, a second round of dilutions were carried out. This process was repeated a minimum of three times before any further action was taken. When the abundance of choanoflagellates was very low on the mixed cultures, the choanoflagellates present exhibited low to none flagellar movement or morphological abnormalities (e.g. absence of flagellum, damaged cell membranes or reduced luminosity and bacterial colonies formation attached to the cell membrane) isolation was deemed unsuccessful and cultures were discarded. On the other hand, cultures where only choanoflagellates were present were considered successful isolations and kept as isolated choanoflagellates cultures.

The second approach, single-cell picking, was used when cultures exhibited medium to high abundances of choanoflagellates and a lower abundance of non-choanoflagellates organisms. In these cases, after the initial period of incubation, culture flasks were vigorously shaken by hand and poured into a Petri dish. Flask were subsequently refilled with fresh media and the process was repeated for

another two rounds. Petri dishes containing the washed off media were used to attempt single-cell picking isolation as previously described. After the third round of washing the culture flasks were refreshed with fresh media and kept for further examination. After an incubation period of 1-3 weeks those flasks where the presence of choanoflagellates could not be confirmed by direct observation were discarded while the flasks that still contained choanoflagellates were kept for further isolation attempts or as isolated choanoflagellates cultures depending on whether or not non-choanoflagellates organisms were still present on the culture.

The third approach, serial dilution followed by single-cell picking, was used when cultures that underwent the serial dilution process exhibited medium to high abundances of choanoflagellates but non-choanoflagellates organisms were still present on the culture. In these cases, after the isolation by serial dilution process was repeated for three times, cultures were used for single-cell picking isolation as described above.

Cultures considered as successful isolations were stored at room temperature, screened and refreshed every 2-3 weeks. In order to minimise possible sources of contamination, isolated cultures were refreshed only with artificial culture media. Amongst all the above-described culture media, the media employed to refresh each culture was the one with the most similar conditions to the field conditions. During the screening process, cultures where contamination from non-choanoflagellates eukaryotes was detected were either discarded or used for further rounds of isolation attempts depending on the abundance of non-choanoflagellates present in the culture.

Cultures maintenance.

Maintenance of choanoflagellates cultures from collections.

Choanoflagellates cultures of *Diaphanoeca grandis* Ellis (ATCC 50111), *Mylnosiga fluctuans* Carr *et al.* (ATCC 50635) and *Microstomoeca roanoka* Carr *et al.* (ATCC 50931) were acquired from American Type Culture Collection (ATCC). Cultures of *D. grandis* were received as a growing culture

and were immediately subcultured in 25ml culture flasks with fresh chilled ASW media and stored at 4ºC. The remainder of the cultures were received as frozen ampoules. Ampoules were stored for two days at -80ºC prior to culture initiation in order to prepare all necessary media and materials for culture initiation. Culture initiation followed the following protocol:

Step 1. One day before thawing the ampoule, culture media was inoculated with *K.* pneumoniae or *E. aerogenes*, for marine and freshwater species respectively, with a bacterial loop.

Step 2. Frozen ampoules were thawed at 35ºC in a water bath without agitation.

Step 3. Immediately after thawing, material was aseptically transferred to 25ml cultures flasks containing either ASW or WC media, for marine and freshwater species respectively.

Step 4. Flask were incubated at 25ºC.

Step 5. After 24h hours incubation cultures were screened by direct observation under an inverted microscope.

Step 6. After 1 week incubation each culture was split in 5 25ml cultures flask containing the appropriated culture media. 2 Cultures were stored at room temperature and the remaining 3 cultures were stored at 25ºC.

Step 7. After 1 week incubation, subcultures were screened by direct observation and from those with a higher cell density 3 0.5ml aliquots were extracted and used for cryopreservation at 80ºC. Aliquots were mixed by pipetting with 0.5ml glycerol and incubated for 10-15min in dry-ice before being stored at -80ºC.

After cultures were established, they were stored at 17ºC (FOC120I Cooled Incubator, Velp Scientifica™, Italy), with the exception of cultures from *D. grandis* that were stored at 4ºC, and maintained by subculturing in fresh media every 2 weeks. Cultures were screened for cell density by direct observation.

Bacterial cultures production and maintenance.

Alongside the media above described, several bacterial cultures media were produced for the necessary culture of bacterial strains (Table 3.1.4). Four different bacterial strains were cultured in the laboratory, namely: *Pseudomonas aeruginosa* (Schroeter) Migula (ATCC 10145), *Klebsiella aerogenes* Tindall *et al.* (ATCC 13048), *Raoultella planticola* (Bagley *et al.*) Drancourt *et al.* (ATCC 700831) and *Escherichia coli* DH5α Subcloning Efficiency™ Competent Cells (Invitrogen, USA).

Cultures of *P. aeruginosa* were grown in LB broth liquid media produced from LB Broth Base (Invitrogen) following manufacturer's instructions. After inoculation, the culture was incubated at 37ºC for 24h and then stored at 4ºC. Once established, the culture was refreshed 24h prior to be used.

Similarly, cultures of *K. aerogenes* and *R. planticola* were grown on Nutrient agar plates produced from Nutrient Broth (Invitrogen) and SELECT Agar (Invitrogen) following manufacturer's instructions. After inoculation, cultures were incubated at 37ºC for 24h and then stored at 37ºC for immediate use or resuspended in 1:1 Nutrient broth:glycerol solution and stored at -80ºC for long term storage.

Upon arrival to the laboratory, *E. coli* competent cell cultures were stored at -80ºC, following manufacturer's specifications. Cultures were grown on X-gal (Cambridge Bioscience, UK) LB agar plates (50µg/ml X-gal) and Ampicillin (Cambridge Bioscience) LB agar plates (100µg/ml Ampicillin) produced from LB Agar (Invitrogen) following manufacturer's instructions. (for detailed protocol on *E. coli* DH5α transformation refer to manufacturer's protocol).

Table 3.1.4: Summary of the culture conditions for the bacterial strains used during the present study.

Bacterial strain	Culture media	Culture media typical composition*	Storage conditions	
			Short term storage	Long term storage
P. aeruginosa	LB broth	10g/l Peptone 5g/l Yeast extract 5g/l NaCl	Liquid culture stored at 4ºC	Liquid culture stored at 4ºC
K. aerogenes *R. planticola*	Nutrient agar	15g/l Agar + 5g/l NaCl 5g/l Peptone 2g/l Yeast extract 1g/l 'Lab-Lemco' powder	Plates stored at 37ºC	1:1 Nutrient broth:glycerol solution stored at -80ºC
E. coli DH5α**	X-gal LB agar Ampicillin LB agar	10g/l Peptone 5g/l Yeast extract 5g/l NaCl + X-gal 50µg/ml or Ampicilin100 µg/ml	-	Frozen cells stored at -80ºC

Results.

Samples processing: eDNA extractions.

According to the results of the spectrophotometric analysis, the eDNA extraction protocol performed poorly (Table 3.1.5). Concentration yields varied between 0 and 47.4 ng/µl. Even though some extractions showed a clear absorbance peak around 260 nm, none of the extractions presented values of A260/A280 and A260/A230 consistent with highly purified DNA (1.7-2.0 and >1.5 respectively).

Table 3.1.5: Results of eDNA quantification in Nanodrop2000. DNA Yield (ng/µl), absorbance measured at 260nm and 280nm (A260 and A280 respectively), ratio of absorbance at 260nm/ absorbance at 280nm (A260/A280) and ratio of absorbance at 260nm/absorbance at 230nm (A260/230).

Sample	DNA Yield (ng/µl)	A260	A280	A260/A280	A260/A230
Torrevieja 1	0.2	0.005	0.004	1.19	0.01
Torrevieja 2	21.8	0.437	0.075	5.8	0.09
Tibi 1	32.1	0.642	0.415	1.55	1.1
Tibi 2	47.4	0.949	0.683	1.39	0.69
Beniarrés 1	10.3	0.206	0.133	1.54	1.56
Beniarrés 2	23.5	0.47	0.209	2.25	0.15
Font Salat 1	-0.1	-0.003	0.038	-0.07	-0.01
Font Salat 2	25.8	0.515	0.215	1.58	0.59
La Mata 1	8.2	0.164	-0.159	-1.03	0.04
La Mata 2	28.4	0.568	-0.472	-1.2	0.06
La Mata 2b	5.2	0.105	-0.042	-2.5	0.06
Torrevieja 1’	-3.8	-0.077	-0.021	3.67	0.53
Torrevieja 2’	11.4	0.228	0.058	3.96	0.1
Tibi 1’	10.6	0.212	0.192	1.11	3.23
Tibi 2’	27.2	0.544	0.426	1.28	0.82
Beniarrés 1’	-1.2	-0.025	0.029	-0.85	0.09
Beniarrés 2’	30.7	0.614	0.442	1.39	0.45
Font Salat 1’	-7.1	-0.141	-0.044	3.21	0.44
Font Salat 2’	20.3	0.406	0.283	1.44	0.48
La Mata 1’	1.5	0.029	-0.116	-0.25	0.01
La Mata 2’	9.3	0.185	-0.123	-1.51	0.06
La Mata 2b’	1	0.02	-0.08	-0.25	0.01

Isolation and establishment of new choanoflagellates cultures from environmental samples.

Successful isolation was achieved for four choanoflagellates morphospecies. Two morphospecies, ST-AC1 and ST-AC2, were isolated from Strensall Common 1 samples, although they were also present in Strensall Common 2, Strensall Common 3 and Strensall Common 4 sites. The pH of the samples from

Strensall Common 1-5 sites, measured on site ranged from pH4.2 to pH5.2. Upon arrival to the laboratory pH measurements varied ranging from pH3.4, for samples from Strensall Common 1 site, to pH5.2, for samples from Strensall Common 5 site. Since ST-AC1 and ST-AC2 were isolated from Strensall Common 1 site, the artificial media used for the establishment of isolated cultures was WC+ pH3.4.

The other two morphospecies, YW-ALK1 and YW-ALK2, were isolated in mixed cultures from Hanson Aggregates 1 samples. Attempts to obtain isolated cultures for each morphospecies were unsuccessful. The pH measurements, on site, indicated a pH11.3 for this site. This measurements were later confirmed upon arrival of the samples to the laboratory. However, during the time that took for the presence of choanoflagellates in the sample to be confirmed, approximately 2 weeks, the pH of the sample dropped to pH8.5. The artificial media used for the establishment of isolated cultures was WC+ pH9.

The provisional description of the morphospecies, pending of molecular characterisation and confirmation of the morphological characters through electron microscopy are as follow:

Morphospecies ST-AC1: Cells ~3-5µm in length, collar approximately the same length as the protoplast and flagellum around 1.5 times the length of the collar. Round protoplast enclosed in a robust organic theca subtended by a peduncle approximately the same length as the protoplast. Theca has a flask shape, round to slightly conical in the base, a marked neck. The top section of the theca extends covering a third to half of the collar length (Figure 3.1.1).

Isolation site: Strensall Common 1 (POND 4), York, UK (54.038,-1.000). Permanent acidic pond (pH 3.4) surrounded by peat bog. Volume of the water body subjected to high fluctuations depending on the cycles of draught and flood of the area.

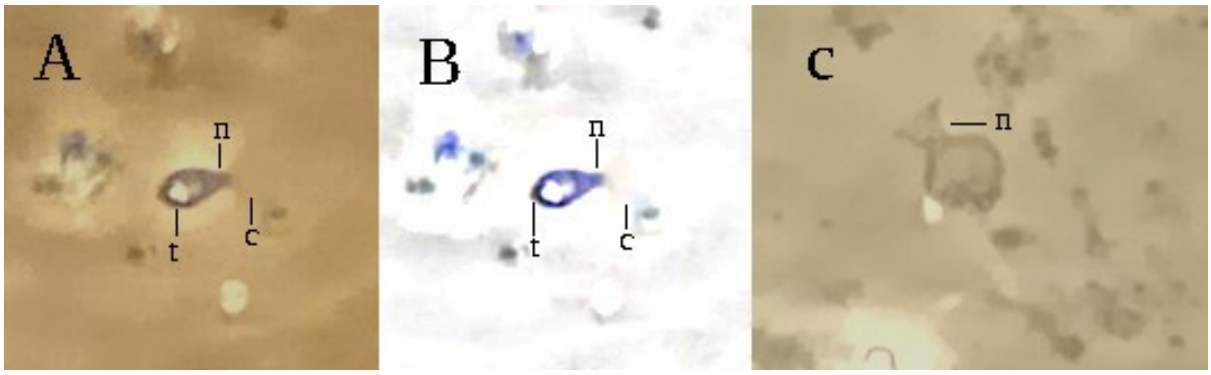

Figure 3.1.1: Images of ST-AC1 from phase contrast microscopy. Living cell (A-B) and empty theca (C) from mixed cultures. Digitally enhanced contrast through edge-detection by Difference of Gaussian (B). Morphological features indicated as follow: theca in the living cell (t), neck of the theca (n) and collar (c).

Morphospecies ST-AC2: Cells ~5-7µm in length, collar around 1.3 times the length of the protoplast and flagellum around 2.5-3 times the length of the protoplast. Ovoid protoplast enclosed in a robust organic theca subtended by a short peduncle. Theca has a vase shape, conical in the base and without neck (Figure 3.1.2).

Isolation site: Strensall Common 1 (POND 4), York, UK (54.038,-1.000). Permanent natural acidic pond (pH3.4) surrounded by peat bog. Volume of the water body subjected to high fluctuations depending on the cycles of draught and flood of the area.

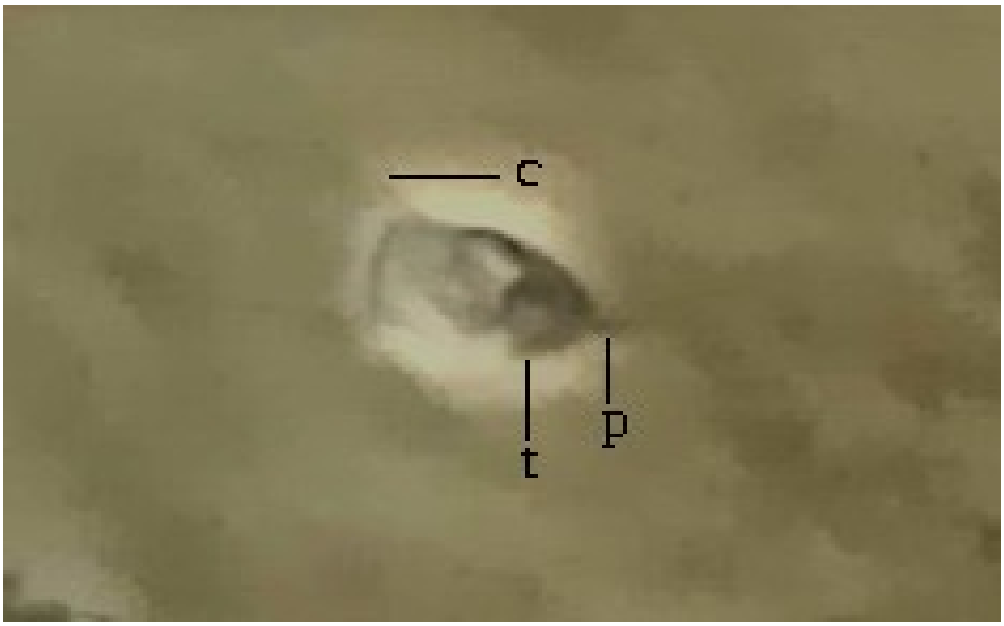

Figure 3.1.2: Image of ST-AC2 from phase contrast microscopy. Living cell from mixed cultures. Morphological features indicated as follow: theca in the living cell (t), peduncle (p) and collar (c).

Morphospecies YW-ALK1 and YW-ALK2: Cells 7-12μm in length, collar ~1/3-2/3 (1/2-2/3 in YW-ALK1 and 1/3-1/2 in YW-ALK2, respectively) the length of the protoplast and flagellum around 1.2 times the length of the collar. Elongated protoplast enclosed in a robust organic theca subtended by a peduncle (around ~1/4-1/3 the length of the theca in YW-ALK1 and <1/5 of the length of the theca in YW-ALK2). Theca has a tube shape, round in the base. The protoplast retracts inside the theca in response to mechanical stimuli, occupying the basal section of the theca (up to ~1/3-1/2 of the theca in YW-ALK1 and ~1/2-2/3 of the theca in YW-ALK2). When retracted the collar and flagellum are protected inside the theca (Figure 3.1.3).

Isolation site: Hanson Aggregates 1, Settle, UK (54.146,-2.303). Permanent artificial hyperalkaline pond (pH11.3). The elevated pH of the site is linked to the disposal of lime kiln wastes from an adjacent limestone quarry during the 1940's.

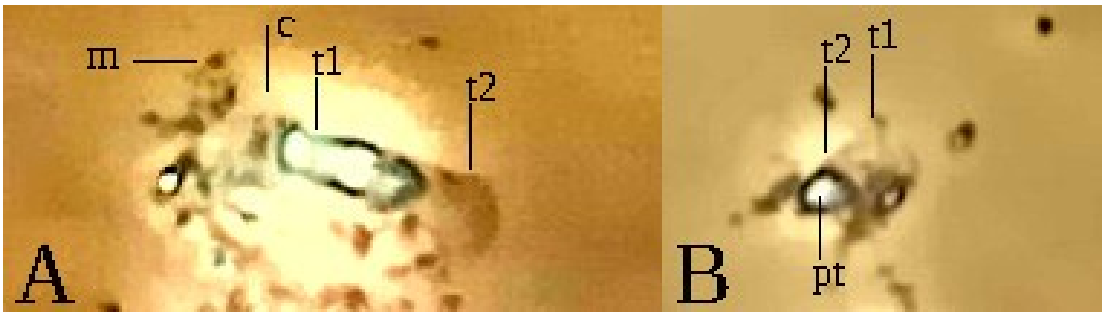

Figure 3.1.3: Images of YW-ALK1/YW-ALK2 from phase contrast microscopy. Living cell (A) and retracted living cell (B) from mixed cultures. Morphological features indicated as follow: particulate matter trapped by the collar (m), collar (c), anterior section of the theca (t1), posterior section of the theca (t2), protoplast retracted within the theca (pt).

Morphospecies YW-ALK1 and YW-ALK2 are highly similar, the main differences being YW-ALK2 has a higher aspect ratio (width:length) and shorter collar and flagellum. Slightly bigger cells have been observed for YW-ALK2 than for Morphospecies YW-ALK1. Their high similarity and the fact that this two morphospecies were never isolated from one another suggests they might not be two different species but rather the reflection of interspecific variation.

Successfully isolated cultures were shipped to Dr. Frank Nitsche's group, who kindly collaborated with the present study, at the Zoology Department from the University of Cologne for further morphological and genetic analyses. Unfortunately, the package containing ST-AC1 and ST-AC2 was damaged during the transit and the delivery company destroyed the package, thus destroying all available isolates. YW-ALK1 and YW-ALK2 are currently being maintained by Dr. Frank Nitsche and Sabine Schiwitza. Unfortunately, due to the current global pandemic the morphological and genetic analyses are being held up *sine die*. Preliminary results of ribosomal SSU and LSU sequencing (data not shown) suggest that both morphospecies might be different species (Nitsche, personal communication).

Discussion.

Samples processing: eDNA extractions.

The low quality performance of the eDNA extraction protocol can be produced by several factors related to the water body from which samples have been taken and the protocol itself. On one hand, it is important to consider the biological productivity potential of the water body. In this regard, the eDNA concentration in highly productivity environments (i.e. eutrophic and hypereutrophic environments) is higher due to the highest abundance of microorganisms present in this environments (Paul *et al*., 1989; Siuda and Chróst, 2000). Similarly, eDNA extractions from water bodies with higher concentrations of suspended particulate organic matter will yield higher eDNA concentrations. The second factor to be considered is the position of the sample in the water column. In this regard, eDNA concentrations are higher in the samples containing substrate since, beside the organisms present in the water column, they contain the microbiota associated to the sediments and the deposition of organic matter coming from the water column (Turner *et al.*, 2015). However, the concentration of DNA adsorbing compounds, such as humic acids or metals, in this samples are higher due to the presence of organisms decomposing the particulate organic matter and the deposition of sediments. The DNA adsorption ratio of the soil is highly variable depending on soil composition which, simultaneously, affects eDNA availability for microorganisms (Nielsen *et al.*, 2006; Nielsen *et al.*, 2007; Pietramellara *et al.*, 2009). These compounds are known to interfere with DNA extraction and subsequent analysis (Steffan *et al.*, 1988; Tebbe and Vahjen, 1993; Harry *et al.*, 1999). It is also important to consider the chemical composition of the water, since DNA extraction is highly dependant on the ionic balance of the reactions. In this regard, high concentrations of salts within the sample can interfere with the DNA extraction process.

However, the protocol failed to obtain high purity DNA regardless of the sample it was used on, even though the selected sampling sites were highly variable regarding the above mentioned conditions. This suggests that the quality of the DNA extractions was independent of the sample and rather a failure of

the protocol itself. In this regard, alternative protocols were considered. Nevertheless, the volume of sample required for such protocols exceeded the capacity of transport and sample processing of the available facilities. Due to the technical problems encountered and the availability of data, from public databases such as those from Tara Oceans and SILVA, for the analyses carried out in the present study, the obtention of eDNA samples was discontinued.

Isolation and establishment of new choanoflagellates cultures from environmental samples.

The choanoflagellates isolated during the present study represent a clear example of the ubiquity of choanoflagellates. On one hand ST-AC1 and ST-AC2 are examples of acidophile choanoflagellates naturally occurring in water bodies with pH values ranging from pH3.4 to pH5.2. On the other hand YW-ALK1 and YW-ALK2, represent examples of alkaliphile choanoflagellates naturally occurring at pH11.3.

The existence of acidophile and alkaliphile choanoflagellates has been previously detected, by molecular characterisation of environmental samples, in several studies. SILVA database contains 18S sequences related to choanoflagellates from studies focused on the microbial communities of acidic environments and similarly several studies have reported the presence of choanoflagellates related sequences on environmental samples from alkaline lakes (Couradeau *et al.*, 2011; Antony *et al.*, 2013).

Nevertheless, only one species of alkaliphile choanoflagellates, *Salpingoeca monosierra*, has been isolated (Hake, 2019). This species was isolated from samples of Mono Lake (CA, USA) naturally occurring at pH10. Contrary to YW-ALK1 and YW-ALK2, for which colony formation has no been observed, *S. monosierra* is a colony forming species. Although *S. monosierra* has been reported in a single celled stage, culturing under laboratory conditions has always led to colony formation. Similarly, to the best of our knowledge, ST-AC1 and ST-AC2 represent the first successful isolation from acidophile choanoflagellates.

Conclusions.

As above mentioned ST-AC1 and ST-AC2 have been found in several sampling sites at varying pH values. Similarly YW-ALK1 and YW-ALK2 were isolated from a sample that varied in pH from hyperalkaline, at the time of sampling, to neutral, at the time of isolation. This seems to suggest that all four morphospecies are able to tolerate certain degrees of variation in pH, what makes them excellent candidates for the study of pH tolerance in choanoflagellates.

-Section II: Salinity tolerance in choanoflagellates.

Introduction.

Choanoflagellates are widely distributed and can be found in most aquatic habitats. Representatives can be found living under a wide range of salinities, from freshwater species as originally described by James-Clark (1867) to hypersaline lakes (Schiwitza *et al.*, 2018; Schiwitza *et al.*, 2019).

Salinity is one of the major factors affecting osmotic pressure in aquatic environments and its effects over diverse biological processes have been widely studied in protists (Finley, 1930; Loefer, 1939; James et al., 2003; Roubeix and Lancelot, 2008). Salinity tolerance experiments constitute a quick and cost-effective methodology for the determination of the osmotic tolerance of cultured organisms and provide very useful information that can have deep evolutionary implications (Knoll and Bauld, 1989; Hauer and Rogerson, 2005; Lowe et al., 2005; Koch and Ekelund, 2005).

Despite the efforts in recent years on the study of choanoflagellates, and despite its importance for the understanding of the evolutionary history of organisms, the knowledge about the salinity tolerance in choanoflagellates remains is scarce. Only a few studies have focused on the characterisation of choanoflagellates' autoecology, most of them restricted to closely related species (Rodriguez and Julian, 2009; Nitsche, 2014; Jeuck, 2014; Nitsche and Arndt, 2015; Schiwitza *et al.*, 2018; Schiwitza *et al.*, 2019).

In order to further expand the current knowledge on the salinity tolerance in choanoflagellates a series of three experiments were carried out.

Firstly, a pilot study was performed to assess the tolerance to abrupt salinity changes of three species of choanoflagellates, namely: *Diaphanoeca grandis*, *Mylnosiga fluctuans* and *Microstomoeca roanoka*.

This species were selected so that all three major groups of choanoflagellates were included. *D. grandis* is a representative of Acanthoecida, while *M. fluctuans* and *M. roanoka* are both representatives of Craspedida, more precisely Clade 2 and Clade 1 respectively. Moreover, this selection includes a wide range of salinity, since *D. grandis* is a marine species, *M. fluctuans* is a freshwater species and *M. roanoka* inhabits brackish waters.

Following the results of this pilot study a second experiment was carried out to quantify the tolerance to abrupt salinity changes of *M. roanoka*. This experiment consisted on a replicate of the pilot study but used an increased number of cultures and focused on the quantification of cultures' performance by including cell counts.

Lastly, a third experiment was carried out to assess the salinity tolerance range of *M. roanoka*. In this case, the experiment focused on a more detailed assessment of the whole salinity range at which this species can grow. Moreover, this last experiment focused on the obtention of RNA samples that will allow to further investigate the molecular mechanisms involved in salinity tolerance in choanoflagellates.

Material and methods.

Assessment of the tolerance to abrupt salinity changes in choanoflagellates.

For this experiment, 3 cultures of each species were grown under three salinity conditions, namely: High, Medium and Low salinity. All culture media was prepared as described in Section I of this chapter. High salinity treatment consisted on ASW+ media (30 PSU, pH 8.0), Medium salinity treatment consisted on ASW+15 media (15 PSU, pH 8.0), and Low salinity treatment consisted on WC+ media (0 PSU, pH 8.0).

At the beginning of the experiment, each of the cultures for each species was subcultured under each condition. Subculturing consisted on the extraction of 5ml aliquots from the culture, after agitation, and its transfer to new 25ml cultures flasks containing fresh media. After 1h settling period cultures were screened for the presence of choanoflagellates and subsequently incubated. Cultures of *M. fluctuans* and *M. roanoka* were incubated at 17ºC, while cultures of *D. grandis* were incubated at 4ºC. During the following 15 days cultures were screened daily and culture growth and performance was established by direct observation.

Quantification of the tolerance to abrupt salinity changes in *M. roanoka.*

This experiment consisted on replicating the experiment carried out during the pilot study, with an increased number of cultures and focusing on the quantification of the differences in culture growth observed. During this experiment, cultures of *M. roanoka* were grown under the same three treatments established during the pilot study, namely: High (30 PSU), Medium (15 PSU) and Low (0 PSU) salinity.

At the beginning of the experiment five cultures of *M. roanoka* that had been growing at 30 PSU, were subcultured under each condition. Subculturing was carried out by vigorously shaking the cultures followed by the extraction of three 1ml aliquots. Each aliquot was then transferred to a new 25ml culture flaks containing the appropriated media. After 1h settling period, an initial screening was carried out and cultures were subsequently incubated at 17ºC. Culture growth was monitored on a daily basis and the number of viable cells present in five microscope fields at 600x magnification was counted at 0h, 24h, 48h, 72h, 96h, 144h, 168h and 238h.

Assessment of the salinity tolerance range in *M. roanoka.*

During this last experiment cultures of *M. roanoka* were subjected to a gradual change of salinity. From the whole set of running cultures, five 25ml cultures were selected for this experiment. Cultures were selected in an attempt to minimise variability of cell density between cultures and to avoid cultures that

exhibited a high number of cyst-like structures. All cultures selected had been grown under the same growth conditions, cultured in ASW Media (30 PSU) and room temperature for a month before the experiment started.

At the beginning of the experiment, three subcultures were initiated from each of the five starting cultures. For this, cultures were vigorously shaken by hand and the walls of the culture flask containing the culture were scrapped with a rubber bacterial loop. Immediately after, three 5ml aliquot were extracted from the culture and transferred to new 25ml cultures flasks containing fresh ASW media. After 1 week incubation at room temperature, each subculture was again split into three 15ml subcultures following the above described procedure and subcultures were again incubated for a week at room temperature. At the end of this initial phase of the experiment nine 15ml subcultures were established from each 25ml culture. Out of those nine subcultures, three were subjected to a gradual decrease in salinity, another three were subjected to a gradual increase of salinity and the remaining three were kept at the same initial conditions for the whole experiment and were used as manipulation controls to determine whether or not cultures manipulation affected viability of the cultures.

Gradual salinity decrease was carried out in six steps, from 30 PSU to 0 PSU salinities, at a rate of 5 PSU decrease per step. Gradual salinity increase was also carried out in 6 steps from 30 PSU to 100 PSU salinities, with an initial 5 PSU increase step, followed by 4 10 PSU increase steps and a final 30 PSU. This phase of the experiment lasted for 18 days, six blocks of three days. Every first day cultures remained at incubation at room temperature. Every 2nd day cultures were screened by direct observation paying especial attention to cell size, membrane integrity, cysts density and colony formation. Finally, every 3rd day a 1ml aliquot was transferred from the previous step to a new flask containing media adjusted to the salinity of the next step. For the manipulation control treatment, 5ml aliquots were transferred from the previous cultures into new flasks containing fresh ASW Media. All intermediate steps cultures were kept and refreshed on the 18th day.

The last phase of this second experiment consisted on a 1 week incubation of all the cultures, followed by a final screening of the cultures to assess cultures viability by direct observation. After the final screening the best two performing cultures at each extreme of the salinity tolerance range were selected

for RNA extraction. RNA extraction was carried out using the TRIzol™ Reagent (Invitrogen), following the protocol specified by the manufacturer. Extracted RNA samples were stored at -80ºC. Results.

Results.

Assessment of the tolerance to abrupt salinity changes in choanoflagellates.

From the three species tested, only *M. roanoka* was capable of growing under variable salinity conditions, while the other two species were only capable of growing under their optimal salinity conditions (Table 3.2.1). Culture death, happened at variable times and was followed by bacterial rapid bacterial growth.

Table 3.2.1: Summary of cultures' performance during the experiment "Assessment of the tolerance to abrupt salinity changes in choanoflagellates." Variation in sizes, damages to the outer membranes and other observations were assessed by direct observation. Treatments indicate the treatment type and salinity in PSU (between brackets). Variation in cell size indicates the direction and magnitude of the changes. Outer membranes damage indicates de time at which outer membrane damages were first reported. Other observations summarises other relevant notes reported throughout the experiment. Culture viability indicates either that the culture was viable throughout the experiment (Viable) or the time at which culture death was reported.

Species	Treatment (Salinity)	Variation in cell size	Outer membranes damage	Other observations	Culture viability
D. grandis	Low (0 PSU)	Increase, filling all the cavity of the lorica	1h	No cell lysis reported	1h
	Medium (15 PSU)	Increase, up to 2/3 of the cavity of the lorica	24h	No cell lysis reported. Reduced flagellar motility at 24h.	48h
	High (30 PSU)	None	None	No apparent culture growth	Viable

Species	Treatment (Salinity)	Variation in cell size	Outer membranes damage	Other observations	Culture viability
M. fluctuans	Low (0 PSU)	None	None	None	Viable
	Medium (15 PSU)	~50% decrease	None	High bacterial density attached to the cells at 72h	96h
	High (30 PSU)	~50% decrease	None	High bacterial density attached to the cells at 72h	96h
M. roanoka	Low (0 PSU)	None	None	Low flagellar motility at 24h. Encystment reported at 72h.	72-96h
	Medium (15 PSU)	None	None	Encystment reported at 72h.	Viable
	High (30 PSU)	None	None	Encystment reported at 72h	Viable

Cultures from *D. grandis* were only capable of growing under High salinity treatment. Cell death was reported immediately after the initial settling period, 1h after culture establishment, in all replicates subjected to Low salinity treatment. Although cell lysis was not directly observed, cells were reported to grow in size, expanding until filling the whole cavity of the lorica. Structural damages on the outer cell membranes were appreciated, primary consisting on the thinning of membranes and loss of uniformity around the posterior and central sections of the cell. During the second screening and thereafter, 24h from the beginning of the experiment, no living cells of *D. grandis* were observed in any of the replicates growing under Low salinity treatment. During the initial screening, *D. grandis* cells growing under Medium salinity treatment presented increased cell size, occupying up to 2/3 of the inner lorica cavity, and diminished flagellar motility. No cell death was reported at this stage. After the initial incubation, after 24h, cultures presented a high number of dead cells and living cells were

reported to have very reduced flagellar motility and damages in the outer membranes. From 48h, and thereafter, no living cells were present in any of the cultures growing under Medium salinity treatment. Living cells were reported throughout the experiment in cultures of *D. grandis* growing under High salinity treatment. No apparent damage in the cell membranes and no abnormalities in cell size were reported in any of the cultures. Nevertheless, cultures of *D. grandis* growing under High salinity treatment exhibited low growth rates, with no apparent culture density changes reported throughout the experiment.

Cultures of *M. fluctuans* were only capable of growing under Low salinity treatment. No apparent cell damage or affected flagellar motility were reported after the initial 1h settling period. In cultures growing under Medium and High salinity treatments, diminished cell sizes were reported after 24-48h of experiment. Moreover, bacterial density increased rapidly during the same period of time. After 72h, cells of *M. fluctuans* growing under Medium and High salinity treatments presented a diminished cell size of ~1/2 compared to the initial cell size and the cell size of the cultures growing under Low salinity treatment. At this time, bacterial flocks and biofilms have been formed and were reported both growing on the surface of the culture flasks and floating in the media. Similarly, high bacterial densities were observed growing attached to the surface of *M. fluctuans* outer cell membranes. After 96h, and thereafter no living *M. fluctuans* cells were observed growing under High salinity treatment and bacterial biofilms were observed completely covering most of the dead cells. The same was reported to occur in cultures growing under Medium salinity treatments on the following 24-48h.

Cultures of *M. roanoka* were capable of growing under Medium and High salinity treatments. No apparent cell damage or changes in cell size were reported throughout the experiment for cultures growing under any of the treatments. Cells of *M. roanoka* growing under Low salinity treatment presented low flagellar motility after the initial 24h, and cultures were overgrown by bacterial biofilms after 72-96h, similarly to what has been described above for cultures of *M. fluctuans* growing under Medium and High salinity treatments. Cyst-like structures formation was reported in both Medium and High salinity treatments. The formation of this structures consisted in the thickening of the outer membranes of the cells and the apparent loss of the feeding organelles. Cells that had undergone the complete encysting process presented 2-2.5 times thicker outer membranes, no sign of flagella or

microvilli collar and no observable cell organelles. This process was reported to start at 72-96h. At the end of the experiment, ~3/4 of the cells observed under High salinity conditions were reported as cyst-like structures. For cultures growing under Medium salinity treatment this structures accounted for ~50% of the culture. Although encystment was observed under Low salinity treatment, this cultures were rapidly overgrown by bacteria after 72-96h. Although this species has been reported to form ephemeral colonies, no colony formation was observed during the experiment.

Quantification of the tolerance to abrupt salinity changes in *M. roanoka.*

Throughout the experiment, 10 days, no cell growth was observed in cultures under Low salinity treatment. Over the same time, both High salinity and Medium salinity treatments produced a steady culture growth (Figure 3.2.1; Table A3.2.1a-c, appendix). Highly significant differences ($p<0.001$) in cell growth were found after 48h between High and Medium salinity treatments, being cultures under Medium salinity conditions those showing faster culture growths (Table 3.2.2). Even though cysts are not unviable *per se* encysted cells were not included in cell counts. The reason for such decision is that encysted cells do not reproduce while in this stage, and therefore they do not contribute to culture growth.

Table 3.2.2: Summary of cultures' performance, measured as average viable cells counted per microscope field at 600x magnification, during the during the experiment "Quantification of the tolerance to abrupt salinity changes in *M. roanoka*.". Hour indicates time passed since culture establishment, treatments indicate the treatment type and salinity in PSU (between brackets) and p indicates probability of Student's t-test between High and Medium treatments. Highly significant differences ($p<0.001$) were found after 2 days.

	Treatment			
Hour	High (30 PSU)	Medium (15 PSU)	Low (0 PSU)	p
0	1.6	1.12	0.84	0.327
24	1.92	2.92	0.28	0.079
48	10.52	34.12	1.2	<0.001
72	26.48	82.6	1.44	<0.001
96	28.76	84.68	0.84	<0.001

		Treatment		
144	50.84	126.6	0.36	<0.001
168	57.44	172.36	0.12	<0.001
240	75.16	227.48	0.44	<0.001

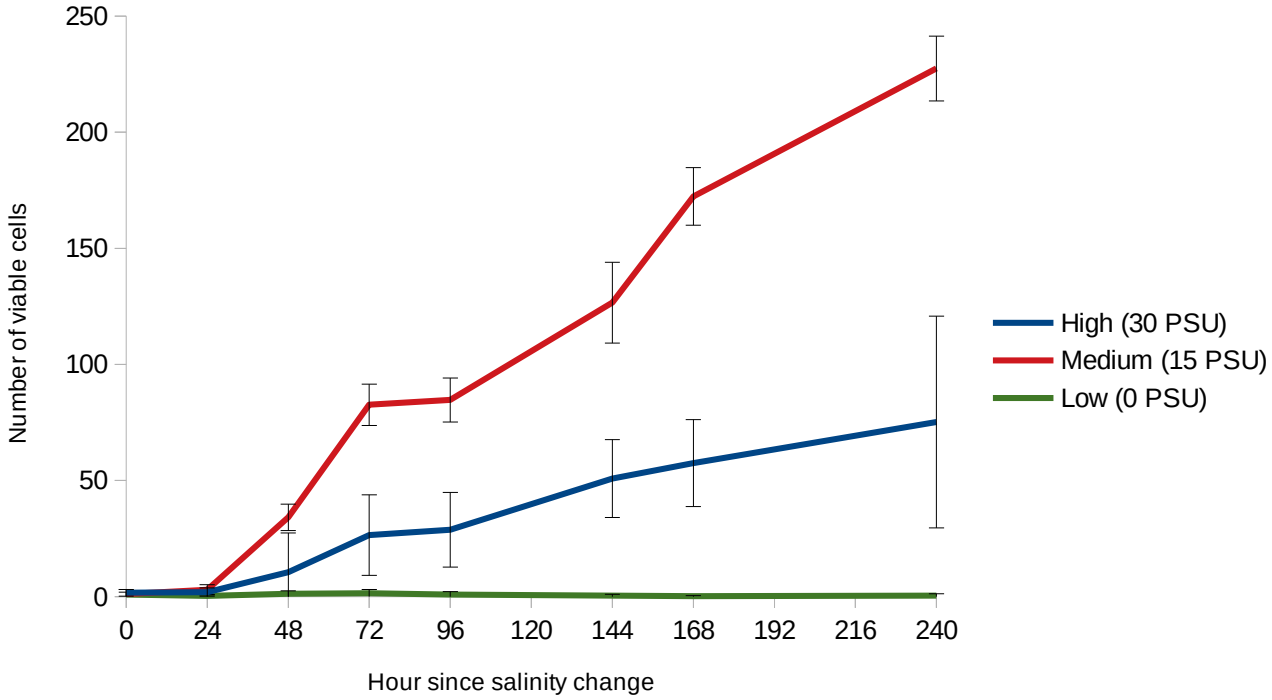

Figure 3.2.1: Tolerance to abrupt salinity changes in *M. roanoka.* Culture growth, measured as average number of viable cells per microscope field at 600x magnification, throughout the experiment "Quantification of the tolerance to abrupt salinity changes in *M. roanoka*.". Coloured lines corresponding to High (Blue), Medium (Red) and Low (Green) salinity treatments.

Assessment of the salinity tolerance range in *M. roanoka.*

Cultures of *M. roanoka* were grown in a range of salinity between 0-5 PSU and 100 PSU (Table 3.3.3). Only four cultures were capable of surviving at 0 PSU, while all of them grew at 5 PSU. Damages on the outer membranes of the cell were reported at 0-5 PSU, although they were not observed in every culture and not all cells within the affected cultures exhibited such damages. The proportion of cells

presenting damages on the outer membranes was higher in cultures growing at 0 PSU than in cultures growing at 5 PSU (estimated at ~80% and of the cells in cultures growing at 0 PSU and at 60% in the most affected cultures growing under 5 PSU). No changes in cell size were observed in any salinity condition. Bacterial overgrowth only happened at 0 PSU and only in those cultures that were unable to grow under such condition. The density of encysted cells did not vary noticeably throughout the salinity variation phase of the experiment and encysted cells never surpassed 10% of the total cells present in the culture. Colony formation was not reported in any culture.

Table 3.2.3: Summary of cultures' performance, assessed by direct observation, during the experiment "Assessment of the salinity tolerance range in *M. roanoka*." From 5 starting cultures (A-E), 3 subcultures were obtained (X1-3), and further subcultured into 3 subcultures (Xn.1-3) and subjected to gradual changes in salinity. Final salinity of viable cultures (Final salinity) were registered and percentage of cells exhibiting outer membranes damages at the end of the experiment was estimated by direct observation (Outer membranes damage).

Culture	Treatment	Final salinity	Outer membranes damage
A1.1	Decrease	0 PSU	~75%
A2.1	Decrease	5 PSU	None
A3.1	Decrease	0 PSU	~90%
A1.2	Control	30 PSU	None
A2.2	Control	30 PSU	None
A3.2	Control	30 PSU	None
A1.3	Increase	100 PSU	None
A2.3	Increase	100 PSU	None
A3.3	Increase	100 PSU	None
B1.1	Decrease	5 PSU	~20%
B2.1	Decrease	5 PSU	None
B3.1	Decrease	5 PSU	None
B1.2	Control	30 PSU	None
B2.2	Control	30 PSU	None
B3.2	Control	30 PSU	None
B1.3	Increase	100 PSU	None
B2.3	Increase	100 PSU	None

Culture	Treatment	Final salinity	Outer membranes damage
B3.3	Increase	100 PSU	None
C1.1	Decrease	5 PSU	None
C2.1	Decrease	5 PSU	None
C3.1	Decrease	0 PSU	~80%
C1.2	Control	30 PSU	None
C2.2	Control	30 PSU	None
C3.2	Control	30 PSU	None
C1.3	Increase	100 PSU	None
C2.3	Increase	100 PSU	None
C3.3	Increase	100 PSU	None
D1.1	Decrease	5 PSU	~50%
D2.1	Decrease	0 PSU	~80%
D3.1	Decrease	5 PSU	~10%
D1.2	Control	30 PSU	None
D2.2	Control	30 PSU	None
D3.2	Control	30 PSU	None
D1.3	Increase	100 PSU	None
D2.3	Increase	100 PSU	None
D3.3	Increase	100 PSU	None
E1.1	Decrease	5 PSU	~60%
E2.1	Decrease	5 PSU	~60%
E3.1	Decrease	5 PSU	~10%
E1.2	Control	30 PSU	None
E2.2	Control	30 PSU	None
E3.2	Control	30 PSU	None
E1.3	Increase	100 PSU	None
E2.3	Increase	100 PSU	None
E3.3	Increase	100 PSU	None

Due to the damages observed on the outer cell membranes and in order to avoid the sequencing of poorly performing cultures and dying cells, 5 PSU was established as the lower boundary for RNA sequencing. RNA was extracted from a total of 4 cultures, the two best performing cultures at 5 PSU

and the two best performing cultures at 100 PSU. Since the transcriptome of *M. roanoka* cultures growing under intermediate salinity levels of this range is already publicly available (Richter *et al.*, 2018), no RNA extraction from cultures growing at intermediate salinities was carried out.

Discussion.

From the three species tested, only *M. roanoka* was capable of tolerating abrupt changes in salinity. Moreover this species is capable of growing under a wide range of salinities, form 0-5 PSU to 100 PSU. These results indicate that *M. roanoka* is able to adapt to sudden changes in salinities affecting the population within a single generation, as well as to gradual changes affecting the population within several generations.

Cultures of *D. grandis* quickly died after being cultured under salinities lower than 30 PSU. *D. grandis* is a member of the Acanthoecida, a group adapted to marine a pelagic environments (Leadbeater, 2008b). Although a high diversity of freshwater acanthoecids has been estimated in the current study and previously published studies (Arroyo *et al.*, 2018), only two species, *Stephanoeca arndti* and *Acanthocorbis mongolica,* have been described inhabiting freshwater ecosystems (Paul, 2012; Nitsche, 2014). The increase in cell size observed suggests a massive input of water in the cell as a consequence of the reduced salinity of the media. Previously published studies (Nitsche and Arndt, 2015), reported a salinity tolerance range from ~10 PSU up to 42 PSU and 71 PSU (for Southern and Northern Hemisphere clones respectively). Such range implies *D. grandis* is able to grow under salinity concentrations lower than those here used as Medium salinity treatment during the first experiment. Such growth was accomplished by gradual decrease of the salinity by 1 PSU every 24h. *D. grandis* is with a circumpolar distribution, present in both polar regions in the Northern and Souther Hemispheres and capable of reaching subtropical latitudes (Nitsche and Arndt, 2015). Marine environments around polar regions are subjected to gradual, seasonal changes in salinity due to the seasonal cycles of ice formation and melting. Species that inhabit such environments are either capable to cope with these changes, amongst other changes related to this same cycles such as chlorophyll α concentration, or suffer seasonal variations in their abundances. In this regard, *D. grandis* has been reported to occur throughout the year in sites near Davis, Antarctica (Marchant and Perrin, 1990). This naturally

occurring seasonal changes in salinity are gradual and involve several generations for a population to acclimate from one extreme to the other of the salinity tolerance range. The combined previous knowledge on *D. grandis* autoecology, the seasonal abundance in its natural habitats and the results of the first experiment indicate that this species is able to cope with gradual but not to abrupt changes in the salinity of their environment.

Cultures of *M. fluctuans* under High and Medium salinity treatments were overgrown by bacteria at ~96h. This suggests that even though no damage on the outer membranes was observed, the choanoflagellates present in the culture were not able to overcome the bacterial growth, probably due to a reduced bacterial consumption. The smaller cell size reported suggest the dehydration of the cells due to the increased salinity of the media. In the case of *M. fluctuans*, this species was isolated from a freshwater pond in Yaroslav, Russia. Although it can be argued that continental water bodies can be subjected to varying salinities as a consequence of seasonal cycles of evaporation and precipitation, there is not sufficient data about the hydrology of the isolation site and the seasonal abundance of *M. fluctuans* at the site to argue whether or not *M. fluctuans*, as *D. grandis*, is capable to cope with gradual changes in salinity. The results of the first experiment, on the other hand, clearly indicate that *M. fluctuans* is not able to adapt to sudden changes in salinity in the water body.

Cultures of *M. roanoka* were capable of growing after abrupt changes in salinity from 15 to 30 PSU. Such abrupt changes occur naturally in coastal marshlands habitats, like the site where this species was originally isolated from (Carr *et al.*, 2017), as a consequence of tidal flooding and retraction cycles. The results from the second experiment showed that populations of *M. roanoka* grow faster at 15 PSU than at 30 PSU, indicating that this species is better adapted to grow in brackish waters environments, such as marshlands, than in marine environments. Marshland environments are also subjected to abrupt drops in salinity as a consequence of freshwater flooding from precipitations taking place upstream. The differences observed in the tolerance between abrupt increases and decreases in salinity are here interpreted as a consequence of the frequency and predictability of such events. On one hand, tidal flooding are highly predictable events with a frequency of approximately 12h, thus they represent a selective pressure to which organisms can adapt. On the other hand, flooding from inland precipitations are much less predictable events that can occur at varying frequencies depending on other

environmental factors. Even though the cultures under Low salinity treatment were not capable of growing, since encystment was observed to happen in this condition as well, it is here hypothesised that *M. roanoka* might be capable of surviving during periods of low salinity through a cryptobiotic stage. Such periods will, most likely, produce a change in the protist community composition, maintaining the predatory pressure over the bacterial community and avoiding the here observed bacterial overgrowth.

Cyst formation has previously been reported in choanoflagellates (Atkins *et al.*, 1998; Leadbeater and Karpov, 2000). The temporal patterns of cyst-like structures formation here reported followed those observed in *Desmarella moniliformis* Kent (Leadbeater and Karpov, 2000). Cyst-like structures appeared after ~72h and grew in density until becoming preponderant in around 15 days. Although differences in encysted cells density were observed between Medium and High salinity treatments, encysting happened in all salinity treatments which seems to indicate that this process is unrelated to the salinity of the media. Moreover, Leadbeater and Karpov (2000) reported no variation in cell encystment rate of cultures of *D. moniliformis* subjected to cold treatment, changes in salinity or aged culture media, which could indicate this process is independent to physico-chemical factors and related to culture density or other environmental factors that have not been accounted for. The fact that during the salinity variation phase of the third experiment the density of encysted cells remained constant seems to support the hypotheses that encystment rate is related to cell density. During this experiment cultures were subsampled every 72h, thus cell density was kept low throughout this phase of the experiment, even though growth rates are expected to be different as the results of the second experiment indicate. Finally, the results of the third experiment indicate that *M. roanoka* can grow under a wide range of salinities, when variations in salinity occur gradually. Such results matched the expectations since coastal marshlands are subjected to gradual, seasonal variations in salinity related to evaporation and inland flooding cycles.

In combination the results from the three experiments suggest that *M. roanoka* is extremely well adapted to marshland and estuarine environments, with a wide range of salinity tolerance, presenting an optimal growth at around 15 PSU, and being able to cope with abrupt changes in salinity, including surviving harsh conditions though cryptobiotic stages. Although the precise regulatory mechanisms by which *M. roanoka* can adapt to such variable salinities remains unknown, it is expected that the RNA

samples extracted from cultures growing in both extremes and of the range and the comparison with the publicly available transcriptome provide useful information to deepen the current knowledge in that regard. Unfortunately, due to the current global pandemic such analysis have been held up *sine die*.

Conclusions.

From the three species used in this experiment, only *M. roanoka* is capable to cope with abrupt changes in salinity from 15 to 30 PSU. This same species showed a wide range of salinity tolerance that seems in accordance with the environmental of the habitat it was originally isolated from. The combination of its wide salinity tolerance range and ability to tolerate abrupt changes in salinity makes this species an excellent candidate for the study of the molecular and regulatory mechanisms involved in marine:freshwater transitions in choanoflagellates. In that sense, the analysis of RNA samples obtained from cultures growing at both extremes of the range and its comparison with the publicly available data is expected to provide useful information in the near future.

-Section III: pH tolerance in choanoflagellates.

Introduction.

As mentioned in the previous section of this chapter, the autoecology of many species of choanoflagellates remains completely unknown with only a few studies focusing on this topic, most of them restricted to closely related species (Rodriguez and Julian, 2009; Nitsche, 2014; Jeuck, 2014; Nitsche and Arndt, 2015; Schiwitza *et al.*, 2018; Schiwitza *et al.*, 2019). To our knowledge, no previous studies have investigated the pH tolerance of choanoflagellates.

Both acidic and hyperalkaline habitats represent extreme environments in which the harsh conditions require organisms to be highly adapted. In order for organisms to survive in these environments, they must develop mechanisms that guarantee the correct osmoregulation of the cell and the maintenance of the homeostasis (Baker-Austin and Dopson, 2007; Horikoshi, 2016). Moreover, low pH increases cationic metals solubility in water which results in high concentrations of heavy metals (Johnson, 1998).

As described in the first section of this chapter, ST-AC1, ST-AC2, YW-ALK1 and YW-ALK2 represent ideal candidates for the study of pH tolerance in choanoflagellates.

ST-AC1 and ST-AC2, the first isolates of acidophile choanoflagellates to our knowledge, were isolated from Strensall Common 1 sampling site, although they were observed also in Strensall Common 2-4 sampling sites. The pH of this sites varied from pH3.4 to pH5.2. Similarly, YW-ALK1 and YW-ALK2 were isolated in a mixed culture from Hanson Aggregates 1 sampling site. The pH of this site varied from pH11.3, at collection time, to pH8.5, at isolation time.

Following the methodology employed for the experiment “Assessment of the tolerance to abrupt salinity changes in choanoflagellates”, two experiment was conducted to asses the tolerance to abrupt pH changes in ST-AC1, ST-AC2, YW-ALK1 and YW-ALK2.

Material and methods.

Assessment of the tolerance to abrupt pH changes in ST-AC1 and ST-AC2.

Two cultures of ST-AC1 and ST-AC2, one of each, growing at pH3.4 and room temperature were grown under four pH conditions, namely: High pH, Medium High pH, Medium Low pH and Low pH. All treatments consisted on modified versions of WC+ culture media as described in Section I of this chapter. High pH treatment consisted on WC+ pH6.5 media, Medium High pH treatment consisted on WC+ pH5.2, Medium Low pH consisted on WC+ pH4.3 media and Low pH treatment consisted on WC+ pH3.4 media.

At the beginning of the experiment, each culture was subcultured under each condition. Subculturing consisted on the extraction of 1ml aliquots from the culture, after scrapping the walls of the culture flask and vigorously shaking it, and its transfer to new 15ml cultures flasks containing fresh media. After 1h settling period cultures were screened for the presence of choanoflagellates and subsequently incubated at room temperature. During the following 15 days cultures were screened daily and culture growth and performance was established by direct observation.

Assessment of the tolerance to abrupt pH changes YW-ALK1 and YW-ALK2.

One mix culture containing both YW-ALK1 and YW-ALK2, growing at pH9 was grown under five treatments, namely: Neutral pH, Low pH, Medium Low pH, Medium High pH and High pH. All treatments consisted on modified versions of WC culture media as described in Section I of this chapter. Neutral pH treatment consisted on WC+ media, Low pH treatment consisted on WC+ pH9

media, Medium Low pH consisted on WC+ pH10 media, Medium High pH treatment consisted on WC+ pH10.5 media and High pH treatment consisted on WC+ pH11 media.

At the beginning of the experiment, the culture was subcultured under each condition. Subculturing consisted on the extraction of 1ml aliquots from the culture, after scrapping the walls of the culture flask and vigorously shaking it, and its transfer to new 15ml cultures flasks containing fresh media. After 1h settling period cultures were screened for the presence of choanoflagellates and subsequently incubated at room temperature. Due to the low culture growth rate exhibited by these two morphospecies, the experiment time was extended and lasted for 75 days. During this time cultures were screened twice a week and culture growth and performance was established by direct observation. Culture media was refreshed biweekly.

Results.

Assessment of the tolerance to abrupt pH changes in ST-AC1 and ST-AC2.

Cultures of ST-AC1 and ST-AC2 were capable of growing under all but High pH treatments (Table 3.3.1). No apparent cell damage or changes in cell size were reported throughout the experiment for cultures growing under any of the treatments. Flagellar motility was only affected under High pH treatment. No colony or cyst-like structures formation were reported.

Table 3.3.1: Summary of cultures' performance during the experiment "Assessment of the tolerance to abrupt pH changes in ST-AC1 and ST-AC2.". Variation in sizes, damages to the outer membranes and other observations were assessed by direct observation. Treatments indicate the treatment type and pH (between brackets). Variation in cell size indicates the direction and magnitude of the changes. Outer membranes damage indicates de time at which outer membrane damages were first reported. Other observations summarises other relevant notes reported throughout the experiment. Culture viability indicates either that the culture was viable throughout the experiment (Viable) or the time at which culture death was reported.

Morphospecies	Treatment (pH)	Variation in cell size	Outer membranes damage	Other observations	Culture viability
ST-AC1	Low (pH3.4)	None	None	None	Viable
	Medium Low (pH4.3)	None	None	None	Viable
	Medium High (pH5.2)	None	None	None	Viable
	High (pH6.5)	None	None	No flagellar motility and high bacterial density attached to the cells at 48h	196h
ST-AC2	Low (pH3.4)	None	None	None	Viable
	Medium Low (pH4.3)	None	None	None	Viable
	Medium High (pH5.2)	None	None	None	Viable
	High (pH6.5)	None	None	No flagellar motility and high bacterial density attached to the cells at 48h	196h

In cultures growing under High pH treatment, bacterial flocks and biofilms have growing on the surface of the culture flasks and floating in the media were reported after 48h. Similarly, high bacterial densities were observed growing attached to the surface of the outer cell membranes and bacterial flocks were reported around the collar of the cells. At this time, no flagellar motility was appreciated.

After, 168h cultures under High pH treatment were completely overgrown by bacteria, with bacterial biofilms completely covering most of the dead cells. At this stage, only empty thecas of ST-AC1 were reported in the cultures. Both ST-AC1 and ST-AC2 morphospecies showed similar responses to the treatments.

Assessment of the tolerance to abrupt changes in pH in YW-ALK1 and YW-ALK2.

Mixed cultures of YW-ALK1 and YW-ALK2 were capable of growing under all treatments. No apparent cell damage, changes in size or flagellar motility, and no colony or cyst-like structures formation were reported under any treatment. Cultures growth rate and cell density remained very low, similar to the growth rate and cell density of the original culture, throughout the experiment for cultures growing under all treatments. Similarly bacterial cell density remained low in all treatments throughout the experiment.

Table 3.3.2: Summary of cultures' performance during the experiment "Assessment of the tolerance to abrupt pH changes in YW-ALK1 and YW-ALK2.". Variation in sizes, damages to the outer membranes and other observations were assessed by direct observation. Treatments indicate the treatment type and pH (between brackets). Variation in cell size indicates the direction and magnitude of the changes. Outer membranes damage indicates de time at which outer membrane damages were first reported. Other observations summarises other relevant notes reported throughout the experiment. Culture viability indicates either that the culture was viable throughout the experiment (Viable) or the time at which culture death was reported.

Morphospecies	Treatment (pH)	Variation in cell size	Outer membranes damage	Other observations	Culture viability
YW-ALK1	Neutral (pH8.0)	None	None	Low growth rate and cell density throughout the experiment. No bacterial overgrowth.	Viable
YW-ALK1	Low (pH9)	None	None	Low growth rate and cell density throughout the experiment. No bacterial overgrowth.	Viable

Morphospecies	Treatment (pH)	Variation in cell size	Outer membranes damage	Other observations	Culture viability
YW-ALK2	Medium Low (pH10)	None	None	Low growth rate and cell density throughout the experiment. No bacterial overgrowth.	Viable
	Medium High (pH10.5)	None	None	Low growth rate and cell density throughout the experiment. No bacterial overgrowth.	Viable
	High (pH11)	None	None	Low growth rate and cell density throughout the experiment. No bacterial overgrowth.	Viable
	Neutral (pH8.0)	None	None	Low growth rate and cell density throughout the experiment. No bacterial overgrowth.	Viable
	Low (pH9)	None	None	Low growth rate and cell density throughout the experiment. No bacterial overgrowth.	Viable
	Medium Low (pH10)	None	None	Low growth rate and cell density throughout the experiment. No bacterial overgrowth.	Viable
	Medium High (pH10.5)	None	None	Low growth rate and cell density throughout the experiment. No bacterial overgrowth.	Viable
	High (pH11)	None	None	Low growth rate and cell density throughout the experiment. No bacterial overgrowth.	Viable

Discussion.

Assessment of the tolerance to abrupt pH changes in ST-AC1 and ST-AC2.

Cultures of both ST-AC1 and ST-AC2 were capable of growing after abrupt changes in pH within the range of pH the were isolated from. This suggests that both morphospecies present a wide range of pH tolerance with an optimal around acidic pH values. As the results from the experiments of the salinity tolerance in *M. roanoka* showed, this abrupt changes only represent a partial vision of the real tolerance exhibited by the organisms and further experiments are necessary to fully assess the pH tolerance range of both morphospecies. Due to the loss of the original cultures, new sampling and isolation campaigns will be required to further investigate these choanoflagellates and to elucidate the mechanisms involved in the adaptation to acidic conditions exhibited by them as well as their placement in the phylogenetic tree of choanoflagellates.

Assessment of the tolerance to abrupt changes in pH in YW-ALK1 and YW-ALK2.

Mixed cultures of YW-ALK1 and YW-ALK2 were capable of growing after abrupt changes in pH from hyperalkaline to neutral conditions. This indicates these morphospecies have a present a wide range of pH with an optimal around alkaline pH values. These results are in congruence with previous knowledge that alkaliphile organisms can be found in neutral and even acidic environments exhibiting low growth rates (Horikoshi, 2004). The low growth rate exhibited by the cultures of these morphospecies requires further investigation to clarify whether this species naturally presents a much longer life cycle than other choanoflagellates or such growth rate is the consequence of environmental factors. Similarly, further studies in which cultures of these morphospecies are subjected to gradual pH changes, rather than abrupt changes are required to fully assess their pH tolerance. It is here hypothesised that since the sample these morphospecies were isolated from suffered a gradual variation in pH from sample collection to choanoflagellates isolation of 2.8 pH units, and in accordance with Horikoshi (2004), the salinity tolerance range of this morphospecies might be much wider than

reflected by the experiment here performed. Lastly, it is expected that comparative transcriptomic analysis from the ongoing cultures and its comparison with transcriptomic data from *S. monosierra* provides useful information to investigate the molecular mechanisms involved in the tolerance to alkaline environments in choanoflagellates.

Conclusions.

All four morphospecies were able to cope with abrupt changes in pH, being the hyperalkaline morphospecies YW-ALK1 and YW-ALK2 the ones showing a wider tolerance. As ST-AC1 and ST-AC2 represent, to the best of our knowledge, the first successfully isolated acidophile choanoflagellates it is highly important that further efforts are taken to reisolate them from their original sampling sites. Similarly, further studies on the ongoing cultures of YW-ALK1 and YW-ALK2 are required to determine the relationship between both morphospecies, their phylogenetic placement within the choanoflagellates and the molecular mechanisms of the adaptation to varying pH, from hyperalkaline to neutral, that they exhibit.

Chapter IV: The evolutionary history of marine:freshwater transitions in Choanoflagellates. Current knowledge and future perspectives.

The phylogeny of choanoflagellates.

The study of choanoflagellates has received unprecedented attention in recent years. Due to their phylogenetic relationship with metazoans (Carr *et al.*, 2008b), choanoflagellates have emerged as model organisms for the study of the origin of multicellularity in animals (King *et al.*, 2003; King, 2004; Snell *et al.*, 2006; Abedin and King, 2008; Alié and Manuel, 2010; Alegado *et al.*, 2012; Levin and King, 2013; Hoffmeyer and Burkhardt, 2016; Woznica *et al.*, 2016; Richter *et al.*, 2018; Laundon *et al.*, 2019; Hake, 2019; Booth and King, 2020). In order to better understand the processes that have driven the evolutionary history of choanoflagellates, a robust and reliable phylogeny is essential.

The phylogeny of choanoflagellates has been previously reviewed throughout the last eighteen years (Medina *et al.*, 2003; Carr *et al.*,2008b, Nitsche *et al.*, 2011; Carr *et al.*, 2017; López-Escardó *et al.*, 2019). The currently most accepted phylogenies of choanoflagellates recovered Acanthoecida and Craspedida as sister orders. These same phylogenies have recovered all but one freshwater species in a single clade, Clade 2 Craspedids (Carr *et al.*, 2017; Schiwitza *et al.*, 2018; Schiwitza *et al.*, 2019). As a result, current phylogenies recovered the diversity of freshwater craspedids as derived from a single major freshwater invasion (Carr *et al.*, 2017).

Nevertheless, eDNA studies have estimated high values of hidden diversity amongst choanoflagellates (del Campo and Masana, 2011; del Campo and Ruiz-Trillo, 2013; del Campo *et al.*, 2015; Arroyo *et al.*, 2018; Venter *et al.*, 2018). Such estudies have revealed that most of the sequences retrieved from environmental samples belong to uncultured or unknown species. The phylogenetic analysis of environmental samples has been used to propose clades composed solely of environmental sequences, some of them freshwater specific (del Campo and Ruiz-Trillo, 2013). These studies suggest that the, currently most accepted, phylogenies are not representative of the diversity of choanoflagellates and

that the evolutionary history of marine:freshwater transitions in choanoflagellates might be more complex than recovered by the phylogenies.

The first section of the second chapter, of the present work, has focused on carrying out the most exhaustive phylogenetic analysis of choanoflagellates up to date, to the best of our knowledge. This reconstruction was carried out to provide a robust and reliable phylogeny that was later used as backbone for the study of the evolutionary history of marine:freshwater transitions in choanoflagellates. These analysis included a revision of diverse methodological approaches related to every step of the phylogenetic reconstruction: sequence alignment, data partitioning, substitution models selection and phylogenetic reconstruction.

The analysis revealed that the resulting phylogenenies are greatly affected by parameter selection throughout the reconstruction process. The inconsitencies across different reconstructions suggests that the dataset is incomplete around some sections of the phylogeny (basal choanoflagellates, acanthoecids and codosigids). Based on the nature of the data here analysed and the evolutionary priors that affect parameter selection, the current study recovered the most accurate phylogeny of choanoflagellates. The phylogeny recontructed in the present study differed with previous reconstructions (Carr *et al.*, 2017; Schiwitza *et al.*, 2018; Schiwitza *et al.*, 2019) in several sections of the tree. Major changes afected both the phylogenetic relationship between nudiforms and tectiforms, and the phylogenetic relationship between codosigids and the rest of Clade 2 craspedids.

Molecular evidence of multiple independent marine:freshwater transitions in choanoflagellates.

The placement of codosigids as an early branching group within Clade 2 significantly affects the evolutionary history of marine:freshwater invasions within choanoflagellates. Under the reconstruction presented, at least three independent marine:freshwater transitions are required to explain the current diversity of freshwater Clade 2 choanoflagellates. Furthermore, there are three equally parsimonious scenarios that explain the recovered diversity of freshwater choanoflagellates. Under the first scenario,

only one marine-to-freshwater colonisation took place early in the history of Clade 2 craspedids and two subsequent freshwater-to-marine reversions took place independently leading to the origin of the marine species within Clade 2 craspedids. Under the second scenario two marine-to-freshwater colonisations took place independently. The first colonisation lead to the diversification of codosigids, while the other one led to the diversification of freshwater salpingoecids. This second colonisation was subsequently followed by a freshwater-to-marine reversion that lead to the origin of *S. macrocollata*. Lastly, under the third scenario three marine-to-freshwater took place independently leading to the diversification of codosigids, *S. punica*, and the remaining freshwater salpingoecids.

The second section of the second chapter, of the present work, has focused on reconstructing the sequences of the last common ancestor of Clade 2. The phylogenetic relationship between the reconstructed sequences and the sequences of extant species was analysed to further investigate the evolutionary history of marine:freshwater transitions within Clade 2. The reconstruction was carried out for three subsets of data that implied different evolutionary histories. The analysis revealed that the reconstructed last common ancestors were more similar to Clade 2 salpingoecids than to codosigids. Such results in combination with the tree topology recovered for codosigids in the present study, long basal branches followed by a cluster of short branches, suggest that codosigids diverged early in the evolution of Clade 2 as a result of an early, independent, freshwater colonisation that led to an adaptative radiation. Similarly, the analysis revealed that *S. punica* colonised freshwater environments independently from the rest of the freshwater Clade 2 salpingoecids. The results, therefore, support multiple independent marine-to-freshwater transitions within Clade 2. The phylogenetic signal left by the freshwater colonisation events is different for codosigids and Clade 2 salpingoecids. Such differences can be interpreted as a consequence of the combined effects of both the age of the colonisation events and the regulation by competitive constrains. Nevertheless, the results suggest that taxon sampling is not deep enough to recover the phylogenetic signals of adaptative radiations occurring at regional scales. It is here hypothesised that the addition of new taxa could have deep implications on the evolutionary history of marine:freshwater transitions in choanoflagellates.

The third section of the second chapter, of the present work, has focused on the phylogenetic placement of environmental sequences within the phylogeny of choanoflagellates. This analysis was carried out to

explore the hidden biodiversity of choanoflagellates, to compare the patters of hidden biodiversity between marine and freshwater environments, and to further explore the evolutionary history of marine:freshwater transitions in choanoflagellates. These results revealed the existence of three major hotspots of hidden biodiversity within choanoflagellates, related to basal positions of Acanthoecida, Craspedida and within freshwater Clade 2 salpingoecids. These major hotspots of hidden biodiversity contained data from both marine and freshwater environments. Furthermore, a much smaller hotspot of hidden freshwater choanoflagellates, consistent with the previously proposed FRESCHO 1 clade (del Campo and Ruiz-Trillo, 2013), was identified. The placement of these sequences suggest that a marine-to-freshwater transition occurred early in the evolution of Clade 1 craspedids that derived in an adaptative radiation. This patter is similar to that observed in Clade 2 and the codosigids. Furthermore, the results support the existence of high abundance of marine hidden diversity within Clade 2. Such hidden diversity concentrates around the last common ancestor of all freshwater Clade 2 salpingoecids but *S. punica*, indicating the presence of numerous marine species related to this clade. Moreover, the prensence of sequences from marine origin related to codosigids suggests the existence of marine species related to this early branching Clade 2 choanoflagellates. These results further support the hypotheses that multiple independent freshwater colonisation events occurred throughout the evolutionary history of Clade 2 craspedids.

Throughout the second chapter of the present work evidence has been gathered supporting that mutiple, independent, marine:freswater transitions have occurred within the evolutionary history of choanoflagellates. Phylogenetic analysis support that the evolutionary history of marine:freshwater transitions in choanoflagellates is much more complex than previously thought. The results exposed support that at least three independent marine-to-freshwater transitions have occurred within Clade 2 craspedids. Moreover, the analises identified three major hotspots of hidden diversity within choanoflagellates. The diversity of acanthoecids, basal craspedids and marine Clade 2 choanoflagellates is underrepresented in the reconstructed phylogeny. In addition, a much smaller fourth hotspot of hidden freshwater diversity, corresponding to the position of the environmental clade FRESCHO 1, was also identified. In agreement with previous studies (del Campo and Masana, 2011; del Campo and Ruiz-Trillo, 2013; del Campo *et al.*, 2015; Arroyo *et al.*, 2018; Venter *et al.*, 2018), the results exposed suggest the absence, in the reconstructed phylogeny, of entire lineages of both marine and freshwater choanoflagellates. A wider taxon sampling around Clade 2 salpingoecids could reveal

patterns of regional adaptative radiations. Nevertheless, due to the placement of hidden diveristy around basal positions of both Acanthoecida and Craspedida, the inclusion of these taxa on future phylogentic reconstructions could dramatically change the phylogenetic relationship between freshwater clades and species, and provide a clearer picture of the evolutionary history of marine:freshwater transitions in choanoflagellates.

Autoecological evidence of salinity resistance in choanoflagellates.

Despite the efforts in recent years on the study of choanoflagellates, and despite its importance for the understanding of the evolutionary history of organisms, salinity tolerance in choanoflagellates remains mostly unexplored. Only a few studies have focused on the characterisation of choanoflagellates' autoecology, most of them restricted to closely related species (Rodriguez and Julian, 2009; Nitsche, 2014; Jeuck, 2014; Nitsche and Arndt, 2015; Schiwitza *et al.*, 2018; Schiwitza *et al.*, 2019).

The second section of the third chapter, of the present work, has focused on the identification of a good candidate species for the study of salinity tolerance in choanoflagellates. Through a series of three experiments, and based on its autoecology, *M. roanoka* was selected as a prime candidate to study the molecular mechanisms of salinity tolerance in choanoflagellates. The species is able to grow at after abrupt changes in salinity from 15 to 30 PSU. Furthermore, the species has been observed to produce cyst, that can act as cryptobiotic stages. Eventhough cultures were unable to grow after abrupt changes in salinity from 15 to 0 PSU, cyst formation was observed to occur, indicating that the species migth be able to survive unfavourable environmental conditions through cryptobiosis. Moreover, *M. roanoka* showed a wide salinity tolerance being able to grow from 0-5 to 100 PSU when salinity varied gradually. This species was originally isolated from a marshland and it is extrimely well adapted to marshland and estuarine environments. This transitional environments are of especial importance for the study of the evolutionary history of marine:freshwater transitions in choanoflagellates, since they represent the link between the high diversification and turnover rates of freshwater environments and the much bigger pool of marine diversity.

RNA samples of *M. roanoka* have been isolated from cultures growing at both extremes of the range of salinity and will be sequenced in future studies. Comparative transcriptomic approaches will be used to allow the identification of differentialy expressed genes under different salinity conditions and the selection for candidate genes related to salinity tolerance in choanoflagellates. Such genes could, subsequently, be mined from publicaly available data to further improve the robustness of future revisions of the evolutionary history of marine:freshwater transitions in choanoflagellates.

Observation and autoecology of four new morfospecies of extremophiles in choanoflagellates.

Despite the recent efforts made towards the study of choanoflagellates, phylogenetic data supports that big portions of the diversity of choanoflagellates remain unexplored. The diversity of known extremophyles in choanoflagellates is restricted to the alkaliphile *S. monosierra* (Hake, 2019). Nevertheless, sequences related to choanoflagellates have been previously reported in eDNA samples from both acidic and alkaline environments (Pruessen *et al.*, 2007; Couradeau *et al.*, 2011; Antony *et al.*, 2013).

At the first section of the third chapter, of the present work, four new morphospecies of, previously unknown to science, choanoflagellates extremophyles have been reported: ST-AC1, ST-AC2, YW-ALK1, YW-ALK2. ST-AC1 and ST-AC2 were isolated from a permanent natural acidic pond (pH3.4) in Strensall Common, York, UK (54.038,-1.000). The area is contains several permanent acidic ponds, pH3.4 to pH5.2 embebed in a matrix of peat bog. The volume of the water bodies is subjected to high fluctuations depending on the cycles of draught and flood of the area. Additionally, YW-ALK1 and YW-ALK2 were isolated from a permanent artificial hyperalkaline pond (pH11.3) in Hanson Aggregates, Settle, UK (54.146,-2.303). The elevated pH of the site is linked to the disposal of lime kiln wastes from an adjacent limestone quarry during the 1940’s. To our knowledge, ST-AC1 and ST-AC2 are the first reported isolates of acidophile choanoflagellates, and, similarly, YW-ALK1 and YW-ALK2 are the second reported isolates of alkaliphile choanoflagellates, the first non-colonial alkaliphiles. Both isolation sites represent rare freshwater environments. In order to survive the harsh environmental conditions, organisms are required to develop mechanisms that guarantee the correct

osmoregulation of the cell and the maintenance of the homeostasis (Baker-Austin and Dopson, 2007; Horikoshi, 2016). The biodiversity of choanoflagellates in rare freshwater environments remains completely unexplored and the its characterisation is expected to continue revealing previously unknown species, as supported by the phylogenetic analysis exposed throughout the second chapter of the present work.

The third section of the third chapter, of the present work, focused on the characterisation of the tolerance to abrupt pH changes of the newly described morphoespecies. Prior to the experiments, all four morphospecies had been reported to exhibit tolerance to certain changes in pH. Morphospecies ST-AC1 and ST-AC2 were reported to naturally occur in samples at pH3.4 to pH5.2,while morphospecies YW-ALK1, and YW-ALK2 were isolated from a sample that changed from pH11.3, at collection time, to pH8.5, at isolation time. Both ST-AC1 and ST-AC2 morphospecies can grow after abrupt changes in pH from pH3.4 to pH5.2, while, similarly, both YW-ALK1 and YW-ALK2 can grow after abrupt changes in pH from pH9 to pH8 and from pH9 to pH11. The original isolates ST-AC1 and ST-AC2 have been lost and it is higly important that further efforts are taken to reisolate them from their original sampling sites. Similarly, further studies on the ongoing cultures of YW-ALK1 and YW-ALK2 are required to determine the relationship between both morphospecies, their phylogenetic placement within the choanoflagellates and the molecular mechanisms of the adaptation to varying pH, from hyperalkaline to neutral, that they exhibit.

Going around to move forward. Future perspectives in marine:freshwater transitions and phylogenetics in choanoflagellates.

Throughout the present work, multiple evidences have been presented supporting that large portions of the biodiversity of choanoflagellates remain unexplored. The analyses carried out throughout the present work have taken into account the problems related to the size and completeness of the datasets used in order to ensure the maximum accuracy and robustness possible. Nevertheless, the results support the existence of large hotspots of hidden diversity in basal positions of the tree. The inclusion of such taxa could dramatically increase the diversity of freshwater choanoflagellates, thus reshaping the evolutionary history of marine:freshwater transitions in choanoflagellates. The results of the present

work suggest that marine:freshwater transitions in choanoflagellates are much more common than previously thought and that the only factor preventing this hypothesis from receiving phylogenetic support is the lack of representative taxa. Further efforts are needed in order to characterise the biodiversity of choanoflagellates. A much wider taxon sampling is necessary to increase the accuracy and robustness of the phylogenetic reconstructions. Such phylogenies provide the necessary evolutionary context for the study of other key life traits such as habitat preference or resistance to environmental stressors.

Furthermore, data about the autoecology of choanoflagellates is scarce. Further studies focused on the characterization of the autoecology and tolerance to environmental stressors of choanoflagellates are required in order to approach the study of ecological and geographical patterns of choanoflagellates diversity. Moreover, the characterisation of the autoecology of choanoflagellates can help identifying model candidates for the study of the molecular mechanisms related to particular traits, as proved on the present work for *M. roanoka*, YW-ALK1 and YW-ALK2. Such molecular studies can have a wide range of applications, from the fundamental questions on the study of the origin of animals, to protein biosynthesis and drug design. Additionally, the identified relevant genes can then be included into further phylogenetic reconstructions in order to integrate, in the reconstruction, the phylogenetic signal of that trait's evolutionary history. As an example, the characterisation of the salinity tolerance of *M. roanoka*, led to the isolation of RNA samples of *M. roanoka* growing under a wide range of salinities. Comparative transcriptomic analysis will be used to identify differentially expressed genes and select candidate genes that are directly related to salinity tolerance mechanisms. Such genes will then be mined from previously published data and incorporated into the phylogeny of choanoflagellates.

Overall, the study of choanoflagellates has gained traction over the last years, however fundamental drivers of choanoflagellates diversity, such as habitat preference, remain mostly unexplored. In order to fully assess the evolutionary implications of newly discovered aspects of choanoflagellates' cell biology, it is necessary to review them through the perspective of phylogenetics. Most of the molecular evidence, gathered in recent years, exploring the biology of choanoflagellates are based on *S. rosetta*. The potential for the use of this species as a model organism is unquestionable and it will allow to further understand the biology of choanoflagellates. Nevertheless, from an evolutive perspective, *S.*

rosetta alone does not fully integrate the biodiversity of choanoflagellates. Infact, no single species is likely to provide a good representation of the diversity of a group as diverse as choanoflagellates. The traits and features exhibited by one species might be limited to that species or a handful of closely related species. It is, therefore, necessary to review such traits and features throughout the phylogeny of choanoflagellates, before creating hypotheses on their evolutionary implications. In order to provide a reliable backbone for the evolutionary hypotheses built, the phylogeny of choanoflagellates requires continuous revision by addition of new species and completion of the current data sets. Such efforts towards the characterisation of choanoflagellates diversity will, simultaneously, provide new model organisms and new opportunities to explore the biology of choanoflagellates.

Bibliography.

Abedin, M. and King, N., 2008. The premetazoan ancestry of cadherins. *Science*, *319*(5865), pp.946-948.

Alberdi, A., Aizpurua, O., Gilbert, M.T.P. and Bohmann, K., 2018. Scrutinizing key steps for reliable metabarcoding of environmental samples. *Methods in Ecology and Evolution*, *9*(1), pp.134-147.

Alegado, R.A., Brown, L.W., Cao, S., Dermenjian, R.K., Zuzow, R., Fairclough, S.R., Clardy, J. and King, N., 2012. A bacterial sulfonolipid triggers multicellular development in the closest living relatives of animals. *elife*, *1*, p.e00013.

Alié, A. and Manuel, M., 2010. The backbone of the post-synaptic density originated in a unicellular ancestor of choanoflagellates and metazoans. *BMC evolutionary biology*, *10*(1), pp.1-10.

Annenkova, N.V., Hansen, G., Moestrup, Ø. and Rengefors, K., 2015. Recent radiation in a marine and freshwater dinoflagellate species flock. *The ISME journal*, *9*(8), pp.1821-1834.

Antony, C.P., Kumaresan, D., Hunger, S., Drake, H.L., Murrell, J.C. and Shouche, Y.S., 2013. Microbiology of Lonar Lake and other soda lakes. *The ISME journal*, *7*(3), pp.468-476.

Apothéloz Perret Gentil, L., Bouchez, A., Cordier, T., Cordonier, A., Guéguen, J., Rimet, F., Vasselon, V. and Pawlowski, J., 2020. Monitoring the ecological status of rivers with diatom eDNA metabarcoding: a comparison of taxonomic markers and analytical approaches for the inference of a molecular diatom index. *Molecular Ecology*.

Arroyo, A.S., López-Escardó, D., Kim, E., Ruiz-Trillo, I. and Najle, S.R., 2018. Novel diversity of deeply branching Holomycota and unicellular holozoans revealed by metabarcoding in Middle Paraná River, Argentina. *Frontiers in Ecology and Evolution, 6*, p.99.

Atkins, M.S., Anderson, O.R. and Wirsen, C.O., 1998. Effect of hydrostatic pressure on the growth rates and encystment of flagellated protozoa isolated from a deep-sea hydrothermal vent and a deep shelf region. *Marine ecology progress series, 171*, pp.85-95.

Atkinson, G.C., Kuzmenko, A., Chicherin, I., Soosaar, A., Tenson, T., Carr, M., Kamenski, P. and Hauryliuk, V., 2014. An evolutionary ratchet leading to loss of elongation factors in eukaryotes. *BMC evolutionary biology*, *14*(1), pp.1-9.

Azam, F., Fenchel, T., Field, J.G., Gray, J.S., Meyer-Reil, L.A. and Thingstad, F., 1983. The ecological role of water-column microbes in the sea. *Marine ecology progress series*, pp.257-263.

Barbera, P., Kozlov, A.M., Czech, L., Morel, B., Darriba, D., Flouri, T. and Stamatakis, A., 2019. EPA-ng: massively parallel evolutionary placement of genetic sequences. *Systematic biology*, *68*(2), pp.365-369.

Bass, D., Howe, A., Brown, N., Barton, H., Demidova, M., Michelle, H., Li, L., Sanders, H., Watkinson, S.C., Willcock, S. and Richards, T.A., 2007. Yeast forms dominate fungal diversity in the deep oceans. *Proceedings of the Royal Society B: Biological Sciences, 274*(1629), pp.3069-3077.

Baker-Austin, C. and Dopson, M., 2007. Life in acid: pH homeostasis in acidophiles. *Trends in microbiology, 15*(4), pp.165-171.

Berger, S.A. and Stamatakis, A., 2012. PaPaRa 2.0: a vectorized algorithm for probabilistic phylogeny-aware alignment extension. *Heidelberg Institute for Theoretical Studies*.

Betancur R, R., Ortí, G., Stein, A.M., Marceniuk, A.P. and Alexander Pyron, R., 2012. Apparent signal of competition limiting diversification after ecological transitions from marine to freshwater habitats. *Ecology Letters*, *15*(8), pp.822-830.

Booth, D.S., Szmidt-Middleton, H. and King, N., 2018. Transfection of choanoflagellates illuminates their cell biology and the ancestry of animal septins. *Molecular biology of the cell*, *29*(25), pp.3026-3038.

Booth, D.S. and King, N., 2020. Genome editing enables reverse genetics of multicellular development in the choanoflagellate Salpingoeca rosetta. *Elife*, *9*, p.e56193.

Boyer, F., Mercier, C., Bonin, A., Le Bras, Y., Taberlet, P. and Coissac, E., 2016. obitools: A unix inspired software package for DNA metabarcoding. *Molecular ecology resources*, *16*(1), pp.176-182.

Bråte, J., Klaveness, D., Rygh, T., Jakobsen, K.S. and Shalchian-Tabrizi, K., 2010. Telonemia-specific environmental 18S rDNA PCR reveals unknown diversity and multiple marine-freshwater colonizations. *BMC microbiology*, *10*(1), pp.1-9.

Cabello-Yeves, P.J. and Rodriguez-Valera, F., 2019. Marine-freshwater prokaryotic transitions require extensive changes in the predicted proteome. *Microbiome*, *7*(1), p.117.

Capella-Gutiérrez, S., Silla-Martínez, J.M. and Gabaldón, T., 2009. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics*, *25*(15), pp.1972-1973.

Carr, M., Nelson, M., Leadbeater, B.S. and Baldauf, S.L., 2008a. Three families of LTR retrotransposons are present in the genome of the choanoflagellate Monosiga brevicollis. *Protist*, *159*(4), pp.579-590.

Carr, M., Leadbeater, B.S., Hassan, R., Nelson, M. and Baldauf, S.L., 2008b. Molecular phylogeny of choanoflagellates, the sister group to Metazoa. *Proceedings of the National Academy of Sciences*, *105*(43), pp.16641-16646.

Carr, M., Leadbeater, B.S. and Baldauf, S.L., 2010. Conserved meiotic genes point to sex in the choanoflagellates. *Journal of Eukaryotic Microbiology*, *57*(1), pp.56-62.

Carr, M., Richter, D.J., Fozouni, P., Smith, T.J., Jeuck, A., Leadbeater, B.S. and Nitsche, F., 2017. A six-gene phylogeny provides new insights into choanoflagellate evolution. *Molecular Phylogenetics and Evolution, 107*, pp.166-178.

Castresana, J., 2000. Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Molecular biology and evolution*, *17*(4), pp.540-552.

Cavalier-Smith, T., Chao, E.E.Y. and Oates, B., 2004. Molecular phylogeny of Amoebozoa and the evolutionary significance of the unikont Phalansterium. *European Journal of Protistology, 40*(1), pp.21-48.

Chariton, A.A., Stephenson, S., Morgan, M.J., Steven, A.D., Colloff, M.J., Court, L.N. and Hardy, C.M., 2015. Metabarcoding of benthic eukaryote communities predicts the ecological condition of estuaries. *Environmental pollution, 203*, pp.165-174.

Chernomor, O., Von Haeseler, A. and Minh, B.Q., 2016. Terrace aware data structure for phylogenomic inference from supermatrices. *Systematic biology, 65*(6), pp.997-1008.

Clare, E.L., Chain, F.J., Littlefair, J.E. and Cristescu, M.E., 2016. The effects of parameter choice on defining molecular operational taxonomic units and resulting ecological analyses of metabarcoding data. *Genome, 59*(11), pp.981-990.

Cordier, T., Frontalini, F., Cermakova, K., Apothéloz-Perret-Gentil, L., Treglia, M., Scantamburlo, E., Bonamin, V. and Pawlowski, J., 2019. Multi-marker eDNA metabarcoding survey to assess the environmental impact of three offshore gas platforms in the North Adriatic Sea (Italy). *Marine environmental research, 146*, pp.24-34.

Couradeau, E., Benzerara, K., Moreira, D., Gerard, E., Kaźmierczak, J., Tavera, R. and López-García, P., 2011. Prokaryotic and eukaryotic community structure in field and cultured microbialites from the alkaline Lake Alchichica (Mexico). *PloS one*, *6*(12), p.e28767.

Cowart, D.A., Matabos, M., Brandt, M.I., Marticorena, J. and Sarrazin, J., 2020. Exploring environmental DNA (eDNA) to assess biodiversity of hard substratum faunal communities on the lucky strike vent field (Mid-Atlantic Ridge) and investigate recolonization dynamics after an induced disturbance. *Frontiers in Marine Science*, *6*, p.783.

Creer, S., Fonseca, V.G., Porazinska, D.L., Giblin Davis, R.M., Sung, W., Power, D.M., Packer, M., Carvalho, G.R., Blaxter, M.L., Lambshead, P.J.D. and Thomas, W.K., 2010. Ultrasequencing of the meiofaunal biosphere: practice, pitfalls and promises. *Molecular Ecology*, *19*, pp.4-20.

Czech, L., Barbera, P. and Stamatakis, A., 2020. Genesis and Gappa: processing, analyzing and visualizing phylogenetic (placement) data. *Bioinformatics*, *36*(10), pp.3263-3265.

del Campo, J. and Massana, R., 2011. Emerging diversity within chrysophytes, choanoflagellates and bicosoecids based on molecular surveys. *Protist*, *162*(3), pp.435-448.

del Campo, J. and Ruiz-Trillo, I., 2013. Environmental survey meta-analysis reveals hidden diversity among unicellular opisthokonts. *Molecular biology and evolution*, *30*(4), pp.802-805.

del Campo, J., Mallo, D., Massana, R., de Vargas, C., Richards, T.A. and Ruiz Trillo, I., 2015. Diversity and distribution of unicellular opisthokonts along the European coast analysed using high throughput sequencing. *Environmental microbiology*, *17*(9), pp.3195-3207.

del Campo, J., Heger, T., Rodríguez-Martínez, R., Worden, A.Z., Richards, T.A., Massana, R. and Keeling, P.J., 2019. Assessing the diversity and distribution of apicomplexans in host and free-living environments using high-throughput amplicon data and a phylogenetically informed reference framework. *Frontiers in microbiology, 10*, p.2373.

Dittami, S.M., Heesch, S., Olsen, J.L. and Collén, J., 2017. Transitions between marine and freshwater environments provide new clues about the origins of multicellular plants and algae. *Journal of Phycology, 53*(4), pp.731-745.

Dobzhansky, T. and Sturtevant, A.H., 1938. Inversions in the chromosomes of Drosophila pseudoobscura. *Genetics*, *23*(1), p.28.

Duarte, C.M., 2015. Seafaring in the 21st century: the Malaspina 2010 circumnavigation expedition.

Edgar, R.C., 2004a. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic acids research*, *32*(5), pp.1792-1797.

Edgar, R.C., 2004b. MUSCLE: a multiple sequence alignment method with reduced time and space complexity. *BMC bioinformatics*, *5*(1), p.113.

Fenchel, T., 1982a. Ecology of heterotrophic microflagellates. I. Some important forms and their functional morphology. *Marine ecology progress series. Oldendorf, 8*(3), pp.211-223.

Fenchel, T., 1982b. Ecology of heterotrophic microflagellates. 11. Bioenergetics and growth. *Marine ecology progress series. Oldendorf, 8*(22523), p.1.

Fenchel, T., 1982c. Ecology of heterotrophic microflagellates. IV. Quantitative occurrence and importance as bacterial consumers. *Marine ecology progress series. Oldendorf, 9*(3), p.5.

Fenchel, T., 1987. Ecology of protozoa: The biology of free-living phagotrophic protists–Science Technical. *Madison, Wisconsin*.

Filker, S., Forster, D., Weinisch, L., Mora Ruiz, M., González, B., Farías, M.E., Rosselló Móra, R. and Stoeck, T., 2017. Transition boundaries for protistan species turnover in hypersaline waters of different biogeographic regions. *Environmental microbiology*, *19*(8), pp.3186-3200.

Finley, H.E., 1930. Toleration of fresh water protozoa to increased salinity. *Ecology, 11*(2), pp.337-347.

Furukawa, R., Toma, W., Yamazaki, K. and Akanuma, S., 2020. Ancestral sequence reconstruction produces thermally stable enzymes with mesophilic enzyme-like catalytic properties. *Scientific reports*, *10*(1), pp.1-13.

Gold, K., Pfister, R.M. and Liguori, V.R., 1970. Axenic Cultivation and Electron Microscopy of Two Species of Choanoflagellida 1. *The Journal of Protozoology, 17*(2), pp.210-212.

González Moreno, S., Gómez Barrera, J., Perales, H. and Moreno Sánchez, R., 1997. Multiple effects of salinity on photosynthesis of the protist Euglena gracilis. *Physiologia Plantarum*, *101*(4), pp.777-786.

Guillard, R.R. and Lorenzen, C.J., 1972. Yellow-green algae with Chlorophyllide C 1, 2. *Journal of Phycology*, *8*(1), pp.10-14.

Gumulya, Y. and Gillam, E.M., 2017. Exploring the past and the future of protein evolution with ancestral sequence reconstruction: the ‘retro’approach to protein engineering. *Biochemical Journal*, *474*(1), pp.1-19.

Gumulya, Y., Baek, J.M., Wun, S.J., Thomson, R.E., Harris, K.L., Hunter, D.J., Behrendorff, J.B., Kulig, J., Zheng, S., Wu, X. and Wu, B., 2018. Engineering highly functional thermostable proteins using ancestral sequence reconstruction. *Nature Catalysis*, *1*(11), pp.878-888.

Hajibabaei, M., Shokralla, S., Zhou, X., Singer, G.A. and Baird, D.J., 2011. Environmental barcoding: a next-generation sequencing approach for biomonitoring applications using river benthos. *PLoS one*, *6*(4), p.e17497.

Hake, K.H., 2019. The microbiome of a colonial choanoflagellate from Mono Lake, CA. *University of California, Berkeley*.

Hanson-Smith, V., Kolaczkowski, B. and Thornton, J.W., 2010. Robustness of ancestral sequence reconstruction to phylogenetic uncertainty. *Molecular biology and evolution*, *27*(9), pp.1988-1999.

Harding, T., Roger, A.J. and Simpson, A.G., 2017. Adaptations to high salt in a halophilic protist: differential expression and gene acquisitions through duplications and gene transfers. *Frontiers in microbiology*, *8*, p.944.

Harry, M., Gambier, B., Bourezgui, Y. and Garnier-Sillam, E., 1999. Evaluation of purification procedures for DNA extracted from rich organic samples: interference with humic substances. *Analusis*, *27*(5), pp.439-441.

Hauer, G. and Rogerson, A., 2005. Remarkable salinity tolerance of seven species of naked amoebae (gymnamoebae). *Hydrobiologia*, *549*(1), pp.33-42.

Heger, T.J., Mitchell, E.A., Todorov, M., Golemansky, V., Lara, E., Leander, B.S. and Pawlowski, J., 2010. Molecular phylogeny of euglyphid testate amoebae (Cercozoa: Euglyphida) suggests transitions between marine supralittoral and freshwater/terrestrial environments are infrequent. *Molecular Phylogenetics and Evolution*, *55*(1), pp.113-122.

Hoffmeyer, T.T. and Burkhardt, P., 2016. Choanoflagellate models—Monosiga brevicollis and Salpingoeca rosetta. *Current opinion in genetics & development*, *39*, pp.42-47.

Horikoshi, K., 2004. Alkaliphiles. *Proceedings of the Japan Academy, Series B*, *80*(4), pp.166-178.

Horikoshi, K., 2016. Alkaliphiles. In *Extremophiles* (pp. 53-78). Springer, Tokyo.

James, K.R., Cant, B. and Ryan, T., 2003. Responses of freshwater biota to rising salinity levels and implications for saline water management: a review. *Australian Journal of Botany*, *51*(6), pp.703-713.

James-Clark, H., 1867. Sponge ciliatae as infusoria flagellata. *Memoirs Boston Society of Natural History*, *1*, pp.305-340.

Jeuck, A.C., 2014. *Extended Phylogeny and Morphology of Marine and Freshwater Choanoflagellates with Additional Methodological Studies on Heterotrophic Flagellates* (Doctoral dissertation, Universitäts-und Stadtbibliothek Köln).

Johnson, D.B., 1998. Biodiversity and ecology of acidophilic microorganisms. *FEMS microbiology ecology*, *27*(4), pp.307-317.

Jones, T.C. and Gates, M.A., 1994. A morphometric study of euryhalinity in marine populations of the ciliate genus Euplotes. *Journal of Eukaryotic Microbiology*, *41*(4), pp.303-316.

Joy, J.B., Liang, R.H., McCloskey, R.M., Nguyen, T. and Poon, A.F., 2016. Ancestral reconstruction. *PLoS computational biology*, *12*(7), p.e1004763.

Kaevska, M. and Slana, I., 2015. Comparison of filtering methods, filter processing and DNA extraction kits for detection of mycobacteria in water. *Annals of Agricultural and Environmental Medicine*, *22*(3).

Krishnan, N.M., Seligmann, H., Stewart, C.B., De Koning, A.J. and Pollock, D.D., 2004. Ancestral sequence reconstruction in primate mitochondrial DNA: compositional bias and effect on functional inference. *Molecular Biology and Evolution*, *21*(10), pp.1871-1883.

Kalyaanamoorthy, S., Minh, B.Q., Wong, T.K., von Haeseler, A. and Jermiin, L.S., 2017. ModelFinder: fast model selection for accurate phylogenetic estimates. *Nature methods*, *14*(6), pp.587-589.

Kamikawa, R., Brown, M.W., Nishimura, Y., Sako, Y., Heiss, A.A., Yubuki, N., Gawryluk, R., Simpson, A.G., Roger, A.J., Hashimoto, T. and Inagaki, Y., 2013. Parallel re-modeling of EF-1α function: divergent EF-1α genes co-occur with EFL genes in diverse distantly related eukaryotes. *BMC evolutionary biology*, *13*(1), p.131.

Karpov, S.A. and Leadbeater, B.S., 1998. Cytoskeleton structure and composition in choanoflagellates. *Journal of Eukaryotic Microbiology*, *45*(3), pp.361-367.

Katoh, K. and Standley, D.M., 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Molecular biology and evolution*, *30*(4), pp.772-780.

Kent, W.S., 1880-1882. *A Manual of the Infusoria: Including a Description of All Known Flagellate, Ciliate, and Tentaculiferous Protozoa, British and foreign, and an Account of the Organization and the Affinities of the Sponges* (Vol. 1-3). David Bogue.

Ki, J.S., 2012. Hypervariable regions (V1–V9) of the dinoflagellate 18S rRNA using a large dataset for marker considerations. *Journal of Applied Phycology, 24*(5), pp.1035-1043.

King, N., Hittinger, C.T. and Carroll, S.B., 2003. Evolution of key cell signaling and adhesion protein families predates animal origins. *Science, 301*(5631), pp.361-363.

King, N., 2004. The unicellular ancestry of animal development. *Developmental cell*, *7*(3), pp.313-325.

Knoll, A.H. and Bauld, J., 1989. The evolution of ecological tolerance in prokaryotes. *Earth and Environmental Science Transactions of The Royal Society of Edinburgh*, *80*(3-4), pp.209-223.

Koch, T.A. and Ekelund, F., 2005. Strains of the heterotrophic flagellate Bodo designis from different environments vary considerably with respect to salinity preference and SSU rRNA gene composition. *Protist*, *156*(1), pp.97-112.

Kumar, S., Stecher, G. and Tamura, K., 2016. MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Molecular biology and evolution*, *33*(7), pp.1870-1874.

Lake, J.A., 1991. The order of sequence alignment can bias the selection of tree topology. *Molecular biology and evolution*, *8*(3), pp.378-385.

Laramie, M.B., Pilliod, D.S., Goldberg, C.S. and Strickler, K.M., 2015. *Environmental DNA sampling protocol-filtering water to capture DNA from aquatic organisms* (No. 2-A13). US Geological Survey.

Larsson, A., 2014. AliView: a fast and lightweight alignment viewer and editor for large datasets. *Bioinformatics*, *30*(22), pp.3276-3278.

Laundon, D., Larson, B.T., McDonald, K., King, N. and Burkhardt, P., 2019. The architecture of cell differentiation in choanoflagellates and sponge choanocytes. *PLoS biology, 17*(4), p.e3000226.

Laursen, L., Čalyševa, J., Gibson, T.J. and Jemth, P., 2021. Divergent Evolution of a Protein–Protein Interaction Revealed through Ancestral Sequence Reconstruction and Resurrection. *Molecular biology and evolution, 38*(1), pp.152-167.

Leadbeater, B.S.C., 1989. Developmental studies on the loricate choanoflagellate *Stephanoeca diplocostata* Ellis VI. Effects of silica replenishment on silica impoverished cells. *Protoplasma, 153*(1-2), pp.71-84.

Leadbeater, B.S., 1994. Developmental studies on the loricate choanoflagellate *Stephanoeca diplocostata* Ellis: VII. Dynamics of costal strip accumulation and lorica assembly. *European journal of protistology*, *30*(1), pp.111-124.

Leadbeater, B.S. and Karpov, S.A., 2000. Cyst formation in a freshwater strain of the choanoflagellate *Desmarella moniliformis* Kent. *Journal of Eukaryotic Microbiology*, *47*(5), pp.433-439.

Leadbeater, B.S., 2008a. Choanoflagellate lorica construction and assembly: the nudiform condition. I. *Savillea species*. *Protist, 159*(2), pp.259-268.

Leadbeater, B.S., 2008b. Choanoflagellate evolution: the morphological perspective. *Protistology*, *5*(4).

Leadbeater, B.S., Henouil, M. and Berovic, N., 2008. Choanoflagellate lorica construction and assembly: the nudiform condition. II. *Acanthoeca spectabilis* Ellis. *Protist, 159*(3), pp.495-505.

Leadbeater, B.S., 2010. Choanoflagellate lorica construction and assembly: the tectiform condition. *Volkanus costatus* (= *Diplotheca costata*). *Protist*, *161*(1), pp.160-176.

Leadbeater, B.S. and Cheng, R., 2010. Costal strip production and lorica assembly in the large tectiform choanoflagellate *Diaphanoeca grandis* Ellis. *European journal of protistology, 46*(2), pp.96-110.

Leadbeater, B.S., 2015. The choanoflagellates. Cambridge University Press.

Levin, T.C. and King, N., 2013. Evidence for sex and recombination in the choanoflagellate *Salpingoeca rosetta*. *Current Biology, 23*(21), pp.2176-2180.

Levin, T.C., Greaney, A.J., Wetzel, L. and King, N., 2014. The rosetteless gene controls development in the choanoflagellate *S. rosetta*. *Elife, 3*, p.e04070.

Loefer, J.B., 1939. Acclimatization of fresh-water ciliates and flagellates to media of higher osmotic pressure. *Physiological Zoology, 12*(2), pp.161-172.

Logares, R., Bråte, J., Bertilsson, S., Clasen, J.L., Shalchian-Tabrizi, K. and Rengefors, K., 2009. Infrequent marine–freshwater transitions in the microbial world. *Trends in microbiology, 17*(9), pp.414-422.

López-Escardó, D., Grau-Bové, X., Guillaumet-Adkins, A., Gut, M., Sieracki, M.E. and Ruiz-Trillo, I., 2019. Reconstruction of protein domain evolution using single-cell amplified genomes of uncultured choanoflagellates sheds light on the origin of animals. *Philosophical Transactions of the Royal Society B, 374*(1786), p.20190088.

Lowe, C.D., Day, A., Kemp, S.J. and Montagnes, D.J., 2005. There are high levels of functional and genetic diversity in *Oxyrrhis marina*. *Journal of Eukaryotic Microbiology, 52*(3), pp.250-257.

Marchant, H.J. and Perrin, R.A., 1990. Seasonal variation in abundance and species composition of choanoflagellates (Acanthoecideae) at Antarctic coastal sites. *Polar Biology, 10*(7), pp.499-505.

Marchant, H.J. and Scott, F.J., 1993. Uptake of sub-micrometre particles and dissolved organic material by Antarctic choanoflagellates. *Marine Ecology-Progress Series, 92*, pp.59-59.

Matsen, F.A., Kodner, R.B. and Armbrust, E.V., 2010. pplacer: linear time maximum-likelihood and Bayesian phylogenetic placement of sequences onto a fixed reference tree. *BMC bioinformatics, 11*(1), p.538.

Matsumoto, T., Akashi, H. and Yang, Z., 2015. Evaluation of ancestral sequence reconstruction methods to infer nonstationary patterns of nucleotide substitution. *Genetics*, *200*(3), pp.873-890.

McFall-Ngai, M., 2014. Divining the essence of symbiosis: insights from the squid-vibrio model. *PLoS Biology*, *12*(2), p.e1001783.

Medina, M., Collins, A.G., Taylor, J.W., Valentine, J.W., Lipps, J.H., Amaral-Zettler, L. and Sogin, M.L., 2003. Phylogeny of Opisthokonta and the evolution of multicellularity and complexity in Fungi and Metazoa. *International Journal of Astrobiology, 2*(3), pp.203-211.

Medlin, L., Elwood, H.J., Stickel, S. and Sogin, M.L., 1988. The characterization of enzymatically amplified eukaryotic 16S-like rRNA-coding regions. *Gene, 71*(2), pp.491-499.

Minh, B.Q., Nguyen, M.A.T. and von Haeseler, A., 2013. Ultrafast approximation for phylogenetic bootstrap. *Molecular biology and evolution, 30*(5), pp.1188-1195.

Miracle, M.R., Vicente, E. and Pedrós-Alió, C., 1992. Biological studies of Spanish meromictic and stratified karstic lakes. *Limnetica, 8*, pp.59-77.

Mistry, J., Finn, R.D., Eddy, S.R., Bateman, A. and Punta, M., 2013. Challenges in homology search: HMMER3 and convergent evolution of coiled-coil regions. *Nucleic acids research*, *41*(12), pp.e121-e121.

Mitsi, K., Arroyo, A.S. and Ruiz-Trillo, I., 2019. A global metabarcoding analysis expands molecular diversity of Platyhelminthes and reveals novel early-branching clades. *Biology letters*, *15*(9), p.20190182.

Morrison, D.A. and Ellis, J.T., 1997. Effects of nucleotide sequence alignment on phylogeny estimation: a case study of 18S rDNAs of Apicomplexa. *Molecular biology and evolution*, *14*(4), pp.428-441.

Nakov, T., Beaulieu, J.M. and Alverson, A.J., 2019. Diatoms diversify and turn over faster in freshwater than marine environments. *Evolution*, *73*(12), pp.2497-2511.

Nawrocki, E., 2009. Structural RNA homology search and alignment using covariance models.

Nguyen, L.T., Schmidt, H.A., Von Haeseler, A. and Minh, B.Q., 2015. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Molecular biology and evolution*, *32*(1), pp.268-274.

Nielsen, K.M., Calamai, L. and Pietramellara, G., 2006. Stabilization of extracellular DNA and proteins by transient binding to various soil components. In *Nucleic acids and proteins in soil* (pp. 141-157). Springer, Berlin, Heidelberg.

Nielsen, K.M., Johnsen, P.J., Bensasson, D. and Daffonchio, D., 2007. Release and persistence of extracellular DNA in the environment. *Environmental biosafety research*, *6*(1-2), pp.37-53.

Nitsche, F., Carr, M., Arndt, H. and Leadbeater, B.S., 2011. Higher level taxonomy and molecular phylogenetics of the Choanoflagellatea. *Journal of Eukaryotic Microbiology*, *58*(5), pp.452-462.

Nitsche, F., 2014. *Stephanoeca arndti* spec. nov.–First cultivation success including molecular and autecological data from a freshwater acanthoecid choanoflagellate from Samoa. *European journal of protistology*, *50*(4), pp.412-421.

Nitsche, F. and Arndt, H., 2015. Comparison of similar Arctic and Antarctic morphotypes of heterotrophic protists regarding their genotypes and ecotypes. *Protist*, *166*(1), pp.42-57.

Norris, R.E., 1965. Neustonic marine craspedomonadales (choanofiagellates) from Washington and California. *The Journal of Protozoology*, *12*(4), pp.589-602.

Patin, N.V., Kunin, V., Lidström, U. and Ashby, M.N., 2013. Effects of OTU clustering and PCR artifacts on microbial diversity estimates. *Microbial ecology*, *65*(3), pp.709-719.

Paul, M., 2012. *Acanthocorbis mongolica* nov. spec.–Description of the first freshwater loricate choanoflagellate (Acanthoecida) from a Mongolian lake. *European journal of protistology*, *48*(1), pp.1-8.

Paul, J.H., Jeffrey, W.H., David, A.W., DeFlaun, M.F. and Cazares, L.H., 1989. Turnover of extracellular DNA in eutrophic and oligotrophic freshwater environments of southwest Florida. *Applied and environmental microbiology*, *55*(7), pp.1823-1828.

Perry, R.P., Cheng, T.Y., Freed, J.J., Greenberg, J.R., Kelley, D.E. and Tartof, K.D., 1970. Evolution of the transcription unit of ribosomal RNA. *Proceedings of the National Academy of Sciences*, *65*(3), pp.609-616.

Pesant, S., Not, F., Picheral, M., Kandels-Lewis, S., Le Bescot, N., Gorsky, G., Iudicone, D., Karsenti, E., Speich, S., Troublé, R. and Dimier, C., 2015. Open science resources for the discovery and analysis of Tara Oceans data. *Scientific data*, *2*(1), pp.1-16.

Pietramellara, G., Ascher, J., Borgogni, F., Ceccherini, M.T., Guerri, G. and Nannipieri, P., 2009. Extracellular DNA in soil and sediment: fate and ecological relevance. *Biology and Fertility of Soils*, *45*(3), pp.219-235.

Prokina, K.I. and Mylnikov, A.P., 2017. Heterotrophic flagellates of Sphagnum bogs and lakes in Usman pine forest, Voronezh oblast. *Inland Water Biology*, *10*(2), pp.182-191.

Pruesse, E., Quast, C., Knittel, K., Fuchs, B.M., Ludwig, W., Peplies, J. and Glöckner, F.O., 2007. SILVA: a comprehensive online resource for quality checked and aligned ribosomal RNA sequence data compatible with ARB. *Nucleic acids research*, *35*(21), pp.7188-7196.

Randall, R.N., Radford, C.E., Roof, K.A., Natarajan, D.K. and Gaucher, E.A., 2016. An experimental phylogeny to benchmark ancestral sequence reconstruction. *Nature communications*, *7*(1), pp.1-6.

Ribera, I., Hogan, J.E. and Vogler, A.P., 2002. Phylogeny of hydradephagan water beetles inferred from 18S rRNA sequences. *Molecular Phylogenetics and Evolution*, *23*(1), pp.43-62.

Richter, D.J., Fozouni, P., Eisen, M.B. and King, N., 2018. Gene family innovation, conservation and loss on the animal stem lineage. *Elife*, *7*, p.e34226.

Rodriguez, S.S. and Julian, D., 2009. Survival of a Choanoflagellate Following Acute Exposure to Abiotic Stressors. *The FASEB Journal 23*, p778

Roh, S.W., Abell, G.C., Kim, K.H., Nam, Y.D. and Bae, J.W., 2010. Comparing microarrays and next-generation sequencing technologies for microbial ecology research. *Trends in biotechnology*, *28*(6), pp.291-299.

Roubeix, V. and Lancelot, C., 2008. Effect of salinity on growth, cell size and silicification of an euryhaline freshwater diatom: *Cyclotella meneghiniana* Kütz. *Transitional waters bulletin*, *2*(1), pp.31-38.

Russell, J. and Zomerdijk, J.C., 2005. RNA-polymerase-I-directed rDNA transcription, life and works. *Trends in biochemical sciences*, *30*(2), pp.87-96.

Schiwitza, S., Arndt, H. and Nitsche, F., 2018. Four new choanoflagellate species from extreme saline environments: Indication for isolation-driven speciation exemplified by highly adapted Craspedida from salt flats in the Atacama Desert (Northern Chile). *European Journal of Protistology*, *66*, pp.86-96.

Schiwitza, S., Arndt, H. and Nitsche, F., 2019. First description of an euryoecious acanthoecid choanoflagellate species, *Enibas tolerabilis* gen. et sp. nov. from a salar in the Chilean Andes based on morphological and transcriptomic data. *European Journal of Protistology*, *67*, pp.106-113.

Shabala, L., McMeekin, T. and Shabala, S., 2009. Osmotic adjustment and requirement for sodium in marine protist thraustochytrid. *Environmental Microbiology*, *11*(7), pp.1835-1843.

Sharp, P.M., Averof, M., Lloyd, A.T., Matassi, G. and Peden, J.F., 1995. DNA sequence evolution: the sounds of silence. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, *349*(1329), pp.241-247.

Shokralla, S., Spall, J.L., Gibson, J.F. and Hajibabaei, M., 2012. Next generation sequencing technologies for environmental DNA research. *Molecular ecology*, *21*(8), pp.1794-1805.

Sievers, F., Wilm, A., Dineen, D., Gibson, T.J., Karplus, K., Li, W., Lopez, R., McWilliam, H., Remmert, M., Söding, J. and Thompson, J.D., 2011. Fast, scalable generation of high quality protein multiple sequence alignments using Clustal Omega. *Molecular systems biology, 7*(1), p.539.

Siuda, W. and Chrost, R.J., 2000. Concentration and susceptibility of dissolved DNA for enzyme degradation in lake water some methodological remarks. *Aquatic Microbial Ecology*, *21*(2), pp.195-201.

Smith, N.G. and Eyre-Walker, A., 2001. Why are translationally sub-optimal synonymous codons used in *Escherichia coli*?. *Journal of molecular evolution, 53*(3), pp.225-236.

Snell, E.A., Brooke, N.M., Taylor, W.R., Casane, D., Philippe, H. and Holland, P.W., 2006. An unusual choanoflagellate protein released by Hedgehog autocatalytic processing. *Proceedings of the Royal Society B: Biological Sciences, 273*(1585), pp.401-407.

Sogin, M.L., Morrison, H.G., Huber, J.A., Welch, D.M., Huse, S.M., Neal, P.R., Arrieta, J.M. and Herndl, G.J., 2006. Microbial diversity in the deep sea and the underexplored “rare biosphere”. *Proceedings of the National Academy of Sciences, 103*(32), pp.12115-12120.

Soubrier, J., Steel, M., Lee, M.S., Der Sarkissian, C., Guindon, S., Ho, S.Y. and Cooper, A., 2012. The influence of rate heterogeneity among sites on the time dependence of molecular rates. *Molecular Biology and Evolution, 29*(11), pp.3345-3358.

Southworth, J., Armitage, P., Fallon, B., Dawson, H., Bryk, J. and Carr, M., 2018. Patterns of ancestral animal codon usage Bias revealed through Holozoan Protists. *Molecular biology and evolution, 35*(10), pp.2499-2511.

Srivastava, A.K. and Schlessinger, D., 1991. Structure and organization of ribosomal DNA. *Biochimie*, *73*(6), pp.631-638.

Stackebrandt, E. and Goebel, B.M., 1994. Taxonomic note: a place for DNA-DNA reassociation and 16S rRNA sequence analysis in the present species definition in bacteriology. *International journal of systematic and evolutionary microbiology*, *44*(4), pp.846-849.

Steffan, R.J., Goksøyr, J., Bej, A.K. and Atlas, R.M., 1988. Recovery of DNA from soils and sediments. *Applied and environmental microbiology*, *54*(12), pp.2908-2915.

Taberlet, P., Coissac, E., Hajibabaei, M. and Rieseberg, L.H., 2012. Environmental DNA. *Molecular ecology*, *21*(8), pp.1789-1793.

Tebbe, C.C. and Vahjen, W., 1993. Interference of humic acids and DNA extracted directly from soil in detection and transformation of recombinant DNA from bacteria and a yeast. *Applied and environmental microbiology*, *59*(8), pp.2657-2665.

Thomsen, H.A., Buck, K.R., Coale, S.L., Garrison, D.L. and Gowing, M.M., 1990. Loricate choanoflagellates (Acanthoecidae, Choanoflagellida) from the Weddell Sea, Antarctica. *Zoologica scripta*, *19*(4), pp.367-387.

Tringe, S.G., Von Mering, C., Kobayashi, A., Salamov, A.A., Chen, K., Chang, H.W., Podar, M., Short, J.M., Mathur, E.J., Detter, J.C. and Bork, P., 2005. Comparative metagenomics of microbial communities. *Science*, *308*(5721), pp.554-557.

Turner, C.R., Uy, K.L. and Everhart, R.C., 2015. Fish environmental DNA is more concentrated in aquatic sediments than surface water. *Biological Conservation*, *183*, pp.93-102.

Venter, P.C., Nitsche, F. and Arndt, H., 2018. The hidden diversity of flagellated protists in soil. *Protist*, *169*(3), pp.432-449.

Woyke, T., Teeling, H., Ivanova, N.N., Huntemann, M., Richter, M., Gloeckner, F.O., Boffelli, D., Anderson, I.J., Barry, K.W., Shapiro, H.J. and Szeto, E., 2006. Symbiosis insights through metagenomic analysis of a microbial consortium. *Nature*, *443*(7114), pp.950-955.

Woznica, A., Cantley, A.M., Beemelmanns, C., Freinkman, E., Clardy, J. and King, N., 2016. Bacterial lipids activate, synergize, and inhibit a developmental switch in choanoflagellates. *Proceedings of the National Academy of Sciences*, *113*(28), pp.7894-7899.

Woznica, A., Gerdt, J.P., Hulett, R.E., Clardy, J. and King, N., 2017. Mating in the closest living relatives of animals is induced by a bacterial chondroitinase. *Cell*, *170*(6), pp.1175-1183.

Woznica, A. and King, N., 2018. Lessons from simple marine models on the bacterial regulation of eukaryotic development. *Current opinion in microbiology*, *43*, pp.108-116.

Yang, Z., 1995. A space-time process model for the evolution of DNA sequences. *Genetics*, *139*(2), pp.993-1005.

Zakas, P.M., Brown, H.C., Knight, K., Meeks, S.L., Spencer, H.T., Gaucher, E.A. and Doering, C.B., 2017. Enhancing the pharmaceutical properties of protein drugs by ancestral sequence reconstruction. *Nature biotechnology*, *35*(1), pp.35-37.

Zmasek, C.M. and Eddy, S.R., 2001. ATV: display and manipulation of annotated phylogenetic trees. *Bioinformatics*, *17*(4), pp.383-384.

Žerdoner Čalasan, A., Kretschmann, J. and Gottschling, M., 2019. They are young, and they are many: dating freshwater lineages in unicellular dinophytes. Environmental microbiology, 21(11), pp.4125-4135.

Appendix.

Chapter II, Section I.

Table A2.1.1: Summary of the sequences included in the phylogenetic reconstruction of choanoflagellates. Alphanumeric code indicates NCBI accession numbers, minus sign (-) indicate missing data, crosses (X) indicate newly generated sequences.

Species name	SSU	LSU	HSP90	tubA	EF-1A	EFL
Acanthocorbis unguiculata	HQ026764.1	-	-	-	-	-
Acanthoeca spectabilis	EU011922.1	KT757416.1	KT757419.1	KT757420.1	KT757417.1	KT757418.1
Agelas conifera	AY734443.1	KC869634.1	-	-	JQ606684.1	-
Amoebidium parasiticum	Y19155.1	EU011932.1	-	-	AY582828.1	-
Bicosta minor	KU587839.1	-	-	-	-	-
Calliacantha longicaudata	KU587840.1	KU587841.1	-	-	-	-
Calliacantha natans	KU587842.1	KU587843.1	-	-	-	-
Chironex fleckeri	GQ849073.1	GQ849051.1	-	-	FJ460241.1	-
Choanoeca perplexa	AY149898.1	EU011937.1	KT757435.1	EU011964.1	-	KT757434.1
Codosiga botrytis	JF706243.1	KT757422.1	-	-	HQ896019.1	-
Codosiga hollandica	KT757430.1	KT757431.1	KT757433.1	KT757436.1	KT757432.1	-
Codosiga sp. Iceland	JF706239.1	KT757443.1	-	-	-	-
Codosiga sp. Montevideo	JF706242.1	KT757442.1	-	-	-	-
Codosiga sp. Morocco	JF706236.1	KT757441.1	-	-	-	-
Codosiga sp. Siberia	JF706237.1	KT757440.1	-	-	-	-
Cosmoeca ventricosa	KU587844.1	-	-	-	-	-
Crinolina isefjordiensis	KU587845.1	-	-	-	-	-
Diaphanoeca grandis	AF084234.1	EU011939.1	KT757450.1	EU011966.1	KT768098.1	KT757449.1
Diaphanoeca pedicellata	HQ237460.1	-	-	-	-	-
Diaphanoeca sphaerica	KU587846.1	KU587847.1	-	-	-	-
Diaphanoeca spiralifurca	KT625474.1	-	-	-	-	-
Diaphanoeca undulata	KU587848.1	-	-	-	-	-
Didymoeca costata	EU011923.1	EU011938.1	KT757446.1	EU011965.1	-	KT757445.1
Didymoeca elongata	KT625473.1	-	-	-	-	-
Enibas tolerabilis	MH687869.1	MH687870.1	MK041919.1	MK041920.1	-	MK041921.1
Exaiptasia pallida	KR186056.1	KP761327.1	-	-	-	-
Hartaetosiga balthica	KT757421.1	JQ034425.1	KT757424.1	KT757425.1	-	KT757423.1
Hartaetosiga gracilis	AY149897.1	EU011935.1	KT757428.1	EU011963.1	-	KT757427.1
Hartaetosiga minima	JQ034422.1	JQ034423.1	-	-	-	-
Helgoeca nana	KT757452.1	KT757453.1	KT768096.1	KT768097.1	KT757454.1	KT757455.1
Ichthyophonus hoferi	U43712.1	AY026370.1	-	-	-	-
Malo kingi	GQ849084.1	GQ849061.1	-	-	EU878256.1	-
Microstomoeca roanoka	KT757502.1	KT757503.1	KT757505.1	KT757506.1	-	KT757504.1
Monosiga brevicollis	AF100940.1	AY026374.1	AY226081.1	XM_001743353.1	AY026073.1	-

Species name	SSU	LSU	HSP90	tubA	EF-1A	EFL
Mylnosiga fluctuans	AF271999.1	EU011940.2	KT757458.1	KT757459.1	KT757457.1	-
Nematostella vectensis	AF254382.1	KJ483089.1	AY226090.1	AY226056.1	AB126336.1	-
Parvicorbicula pedunculata	HQ026765.1	-	-	-	-	-
Pleurasiga minima	KU587849.1	KU587850.1	-	-	-	-
Pleurasiga reynoldsii	KU587851.1	KU587852.1	-	-	-	-
Polyoeca dichotoma	KT625475.1	-	-	-	-	-
Salpingoeca amphora	-	X	-	-	-	-
Salpingoeca calixa	KT757470.1	KT757471.1	-	-	-	-
Salpingoeca crinita	MH490945.1	-	-	-	-	-
Salpingoeca dolichothecata	KT757472.1	KT757473.1	KT757475.1	KT757476.1	KT757474.1	-
Salpingoeca euryoecia	KJ631038.1	KJ631045.1	-	-	-	-
Salpingoeca fusiformis	KJ631039.1	KJ631044.1	-	-	-	-
Salpingoeca helianthica	KT757487.1	KT757488.1	KT757464.1	KT757469.1	KT757489.1	-
Salpingoeca huasca	MH490950.1	-	-	-	-	-
Salpingoeca infusionum	AF100941.1	KT757478.1	KT757480.1	KT757481.1	-	KT757479.1
Salpingoeca kvevrii	EU011930.1	EU011946.1	KT757497.1	KT757498.1	-	KT757496.1
Salpingoeca longipes	KJ631040.1	KJ631046.1	-	-	-	-
Salpingoeca macrocollata	KT757482.1	KT757483.1	KT757485.1	KT757486.1	-	KT757484.1
Salpingoeca oahu	KT757492.1	KT757493.1	-	-	-	-
Salpingoeca prava	MH490946.1	MH490947.1	-	-	-	-
Salpingoeca punica	KT757460.1	KT757461.1	KT757490.1	KT757491.1	KT757462.1	KT757463.1
Salpingoeca rosetta	EU011924.1	XR_001137148.1	KT757501.1	XM_004995459.1	XM_004996684.1	-
Salpingoeca steinii	X	X	-	-	-	-
Salpingoeca surira	MH490948.1	MH490949.1	-	-	-	-
Salpingoeca tuba	HQ026774.1	KT757507.1	-	-	-	-
Salpingoeca urceolata	EU011931.1	KT757515.1	KT757517.1	KT757518.1	-	KT757516.1
Salpingoeca ventriosa	KJ631041.2	KT757519.1	-	-	-	-
Savillea micropora	EU011928.1	KT757495.1	KT757465.1	KT757466.1	-	KT757468.1
Sphaeroeca leprechaunica	KJ631042.1	KJ631047.1	-	-	-	-
Stagondoeca pyriformis	KT757499.1	KT757500.1	-	-	-	-
Stephanoeca apheles	EF523336.1	-	-	-	-	-
Stephanoeca arndti	JX069943.1	-	-	-	-	-
Stephanoeca cauliculata	HQ026766.1	-	-	-	-	-
Stephanoeca diplocostata	AY149899.1	EU011943.1	KT757512.1	KT757513.1	KT757510.1	KT757511.1
Stephanoeca norrisii	HQ026768.1	-	-	-	-	-
Stephanoeca paucicostata	HQ026769.1	-	-	-	-	-
Suberites ficus	AF100947.1	AY026381.1	-	AY226051.1	-	-

Chapter II, Section III.

Table A2.3.1: Placement of sequences from freshwater origin. Accession number indicates SILVA/NCBI accession number of the sequence. Placement indicates clade of placement as used in Figure 2.3.3. Placement Location indicate exact placement within the phylogenetic tree.

Accession number	Placement	Placement Location
AB238174	Acanthoecida	*Pleurasiga/Didymoeca*
AB920848	Acanthoecida	*Didymoeca/Bicosta*
AB996621	Clade 2	*Sphaeroeca leprechaunica*
AB996660	Clade 2	*Sphaeroeca leprechaunica*
AB996687	Clade 2	*Sphaeroeca leprechaunica*
DQ104581	Craspedida	Clade1/Clade2
DQ104582	Craspedida	Clade1/Clade2
DQ104583	Clade 1	Marine salpingoecid
DQ104587	Craspedida	Clade1/Clade2
DQ104589	Craspedida	Clade1/Clade2
DQ104590	Craspedida	Clade1/Clade2
DQ104591	Craspedida	Clade1/Clade2
DQ104597	Craspedida	Clade1/Clade2
DQ409084	Clade 1	Basal Clade 1
EF196686	Clade 2	*Sphaeroeca leprechaunica*
EF196724	Acanthoecida	Basal Acanthoecida
EF196742	Acanthoecida	Basal Acanthoecida
EF196744	Acanthoecida	Basal Acanthoecida
EF196771	Acanthoecida	Basal Acanthoecida
EF196774	Acanthoecida	Basal Acanthoecida
EF196791	Acanthoecida	Basal Acanthoecida
EF196793	Acanthoecida	Basal Acanthoecida
EU144002	Craspedida	Clade1/Clade2
EU144008	Clade 2	*Salpingoeca punica/Salpingoeca macrocollata*
EU860443	Clade 2	Basal Codosigid
EU860468	Acanthoecida	*Pleurasiga/Didymoeca*
EU860494	Acanthoecida	Basal Acanthoecida
EU860510	Acanthoecida	*Pleurasiga/Didymoeca*
EU860513	Acanthoecida	*Pleurasiga/Didymoeca*
EU860518	Acanthoecida	*Pleurasiga/Didymoeca*
EU860526	Acanthoecida	*Pleurasiga/Didymoeca*
EU860777	Acanthoecida	*Didymoeca/Bicosta*
EU860781	Craspedida	Basal Craspedida
EU860845	Acanthoecida	*Didymoeca/Bicosta*
EU860868	Clade 2	*Salpingoeca oahu*

EU860873	Acanthoecida	*Didymoeca/Bicosta*
EU860874	Acanthoecida	*Didymoeca/Bicosta*
EU860941	Clade 2	Basal Codosigid
EU860943	Clade 2	Basal Codosigid
EU860949	Acanthoecida	*Didymoeca/Bicosta*
EU860967	Acanthoecida	*Didymoeca/Bicosta*
FJ157338	Craspedida	Clade1/Clade2
FJ410526	Acanthoecida	Basal Acanthoecida
FJ410582	Acanthoecida	Basal Acanthoecida
FJ410600	Acanthoecida	Basal Acanthoecida
FJ410659	Acanthoecida	Basal Acanthoecida
FJ482775	Acanthoecida	Basal Acanthoecida
FN865341	Outgroup	Outgroup
FN866056	Outgroup	Outgroup
GQ844405	Clade 2	*Codosiga botrytis* *Sodosiga helianthica/Salpingoeca*
GQ844406	Clade 2	*steinii+Salpingoeca ventriosa*
GQ844407	Clade 2	*Codosiga sp. Montevideo/Codosiga spp.*
GQ844411	Clade 2	*Codosiga sp. Montevideo/Codosiga spp.*
GQ844419	Acanthoecida	*Didymoeca/Bicosta*
GQ844420	Clade 2	*Codosiga sp. Montevideo/Codosiga spp.*
GQ844425	Clade 2	*Codosiga sp. Montevideo/Codosiga spp.*
GQ844577	Clade 2	*Codosiga sp. Montevideo/Codosiga spp.*
GU067794	Clade 2	*Sphaeroeca leprechaunica*
GU290066	Clade 1	Basal Clade 1
GU290074	Clade 2	Basal Codosigid
GU290082	Clade 2	*Salpingoeca steinii/Salpingoeca ventriosa*
GU290096	Clade 2	*Salpingoeca oahu*
GU290101	Clade 2	Basal Codosigid
GU290104	Clade 2	Basal Codosigid
GU297614	Craspedida	Clade1/Clade2
GU647170	Acanthoecida	Basal Acanthoecida
GU647175	Clade 2	*Salpingoeca punica*
GU647190	Clade 2	Basal FW salpingoecid
GU647193	Clade 2	*Mylnosiga fluctuans*
GU647194	Clade 2	*Salpingoeca punica*
GU647195	Clade 2	*Salpingoeca punica*
GU969087	Acanthoecida	*Enibas tolerabilis*
GU969091	Acanthoecida	*Enibas tolerabilis*
GU969092	Clade 1	Basal Clade 1
GU969095	Acanthoecida	*Enibas tolerabilis*
GU970254	Acanthoecida	*Didymoeca/Bicosta*
HQ219345	Clade 1	Basal Clade 1
HQ219352	Clade 1	Basal Clade 1
HQ219353	Clade 1	Basal Clade 1
HQ219358	Clade 1	Basal Clade 1

HQ219439	Clade 1	Basal Clade 1
HQ219444	Clade 1	Basal Clade 1
HQ219446	Clade 1	Basal Clade 1
JF260899	Clade 2	*Codosiga botrytis*
JF720681	Acanthoecida	*Didymoeca/Bicosta*
JF720720	Clade 2	*Sphaeroeca leprechaunica*
JF774961	Acanthoecida	*Basal Acanthoecida*
JF774993	Acanthoecida	*Didymoeca/Bicosta*
JF775071	Clade 2	FW salpingoecid
JF775098	Clade 2	FW salpingoecid
JF775104	Clade 2	FW salpingoecid
JN090872	Acanthoecida	*Enibas tolerabilis*
JN090879	Clade 2	*Salpingoeca steinii/Salpingoeca ventriosa*
JN547276	Clade 2	*Sphaeroeca leprechaunica*
JQ692231	Clade 2	*Codosiga botrytis*
JX069943	Acanthoecida	*Stephanoeca arndti*
JX426907	Clade 2	*Sphaeroeca leprechaunica*
JX426911	Acanthoecida	Basal Acanthoecida
JX426912	Acanthoecida	Basal Acanthoecida
JX426914	Acanthoecida	Basal Acanthoecida
JX426925	Clade 2	*Sphaeroeca leprechaunica*
JX426928	Acanthoecida	*Pleurasiga/Didymoeca*
JX426930	Clade 2	*Sphaeroeca leprechaunica*
JX426932	Acanthoecida	*Didymoeca/Bicosta*
JX426933	Acanthoecida	Basal Acanthoecida
JX426945	Acanthoecida	*Didymoeca/Bicosta*
JX426946	Acanthoecida	*Didymoeca/Bicosta*
JX426952	Acanthoecida	Basal Acanthoecida
JX426954	Acanthoecida	Basal Acanthoecida
JX426955	Acanthoecida	*Didymoeca/Bicosta*
JX426962	Acanthoecida	Basal Acanthoecida
JX426971	Clade 2	*Sphaeroeca leprechaunica*
JX426974	Acanthoecida	Basal Acanthoecida
JX426976	Acanthoecida	Basal Acanthoecida
JX426977	Acanthoecida	Basal Acanthoecida
JX426982	Acanthoecida	Basal Acanthoecida
JX426988	Acanthoecida	Basal Acanthoecida
JX426989	Clade 2	*Sphaeroeca leprechaunica*
JX427000	Clade 2	*Sphaeroeca leprechaunica*
JX427004	Clade 2	Basal clade2
JX427015	Acanthoecida	Basal Acanthoecida
JX427017	Acanthoecida	Basal Acanthoecida
JX427019	Clade 2	*Mylnosiga fluctuans*
JX427020	Clade 2	*Sphaeroeca leprechaunica*
KC440271	Clade 2	*Sphaeroeca leprechaunica*
KC440528	Clade 2	*Sphaeroeca leprechaunica*
KC440618	Clade 2	*Sphaeroeca leprechaunica*

KC440660	Clade 2	*Sphaeroeca leprechaunica*
KC440694	Clade 2	*Sphaeroeca leprechaunica*
KC440700	Clade 2	*Sphaeroeca leprechaunica*
KC440704	Clade 2	*Sphaeroeca leprechaunica*
KC440737	Clade 2	*Sphaeroeca leprechaunica*
KC440740	Clade 2	*Sphaeroeca leprechaunica*
KC440754	Clade 2	*Sphaeroeca leprechaunica*
KC440765	Clade 2	*Sphaeroeca leprechaunica*
KC440769	Clade 2	*Sphaeroeca leprechaunica*
KC440772	Clade 2	*Sphaeroeca leprechaunica*
KC575498	Craspedida	Basal Craspedida
KC575499	Craspedida	Basal Craspedida
KF733570	Acanthoecida	*Didymoeca costata*
KM261803	Clade 2	Basal Codosiga
KP059194	Clade 2	*Salpingoeca fusiformis*
KP708814	Acanthoecida	*Didymoeca/Bicosta*
KP708837	Acanthoecida	*Didymoeca/Bicosta*
KU196154	Acanthoecida	*Enibas tolerabilis*

Table A2.3.2: Placement of sequences from soil samples. Accession number indicates SILVA/NCBI accession number of the sequence. Placement indicates clade of placement as used in Figure 2.3.4. Placement Location indicate exact placement within the phylogenetic tree.

Accession number	Placement	Placement Location
AB534446	Basal Choanoflagellate	Basal Choanoflagellate
DQ511121	Acanthoecida	Basal Acanthoecida
DQ512283	Outgroup	Outgroup
DQ511300	Outgroup	Outgroup
EU527031	Basal Choanoflagellate	Basal Choanoflagellate
AB970211	Clade 2	*Salpingoeca hollandica*
GU568135	Basal Choanoflagellate	Basal Choanoflagellate
FN394928	Clade 2	*Salpingoeca leprechaunica*
FN394949	Clade 2	*Salpingoeca leprechaunica*
FN394947	Clade 2	*Salpingoeca leprechaunica*
AB970445	Acanthoecida	Basal Acanthoecida
AB970208	Clade 2	*Salpingoeca hollandica*
JN007829	Clade 2	Basal FW salpingoecid
JN007848	Clade 2	*Mylnosiga fluctuans*
JN007836	Clade 2	*Mylnosiga fluctuans*
JX290171	Acanthoecida	*Pleurasiga/Didymoeca*
JN696788	Acanthoecida	Basal Acanthoecida
KF996262	Basal Craspedida	Basal Craspedida
KC664537	Basal Choanoflagellate	Basal Choanoflagellate
KP137398	Clade 1/Clade 2	Clade1/Clade2

AY969234	Acanthoecida	Basal Acanthoecida
DQ510996	Outgroup	Outgroup
EF024885	Acanthoecida	Basal Acanthoecida

Table A2.3.3: Placement of the sequences originally used to define environmental clades of freshwater choanoflagellates. Accession number indicates SILVA/NCBI accession number of the sequence. Environmental Clade indicates the Clade the sequence was originally asigned to. Placement indicates clade of placement as used in Figure 2.3.5. Placement Location indicate exact placement within the phylogenetic tree.

Accession number	Environmental Clade	Placement	Placement Location
DQ104581	FRESCHO 1	Craspedida	Clade1/Clade2
DQ104582	FRESCHO 1	Craspedida	Clade1/Clade2
DQ104587	FRESCHO 1	Craspedida	Clade1/Clade2
DQ104589	FRESCHO 1	Craspedida	Clade1/Clade2
DQ104591	FRESCHO 1	Craspedida	Clade1/Clade2
DQ104597	FRESCHO 1	Craspedida	Clade1/Clade2
DQ409084	FRESCHO 1	Clade 1	Basal Clade 1
EU144002	FRESCHO 1	Craspedida	Clade1/Clade2
GU290066	FRESCHO 1	Clade 1	Basal Clade 1
GU297614	FRESCHO 1	Craspedida	Clade1/Clade2
GU297691	FRESCHO 1	Craspedida	Clade1/Clade2
HQ219345	FRESCHO 1	Clade 1	Basal Clade 1
HQ219352	FRESCHO 1	Clade 1	Basal Clade 1
HQ219353	FRESCHO 1	Clade 1	Basal Clade 1
HQ219358	FRESCHO 1	Clade 1	Basal Clade 1
HQ219439	FRESCHO 1	Clade 1	Basal Clade 1
HQ219444	FRESCHO 1	Clade 1	Basal Clade 1
HQ219446	FRESCHO 1	Clade 1	Basal Clade 1
JN825671	FRESCHO 1	Acanthoecida	Basal Acanthoecida
AF372736	FRESCHO 2	Acanthoecida	Basal Acanthoecida
GU647190	FRESCHO 2	Clade 2	Basal FW salpingoecid
JN825691	FRESCHO 2	Clade 2	Basal Codosigid
FJ410526	FRESCHO 3	Acanthoecida	Basal Acanthoecida
FJ410600	FRESCHO 3	Acanthoecida	Basal Acanthoecida
FJ410659	FRESCHO 3	Acanthoecida	Basal Acanthoecida
FJ848455	FRESCHO 3	Acanthoecida	*Pleurasiga/Didymoeca*
FJ848459	FRESCHO 3	Acanthoecida	Basal Acanthoecida
FJ848461	FRESCHO 3	Acanthoecida	Basal Acanthoecida
FJ848462	FRESCHO 3	Acanthoecida	*Pleurasiga/Didymoeca*
FJ848464	FRESCHO 3	Acanthoecida	Basal Acanthoecida
FJ848466	FRESCHO 3	Acanthoecida	*Pleurasiga/Didymoeca*
FJ848474	FRESCHO 3	Acanthoecida	*Pleurasiga/Didymoeca*
FJ848478	FRESCHO 3	Acanthoecida	*Pleurasiga/Didymoeca*
FJ848479	FRESCHO 3	Acanthoecida	Basal Acanthoecida

FJ848480	FRESCHO 3	Acanthoecida	Basal Acanthoecida
FJ848481	FRESCHO 3	Acanthoecida	*Pleurasiga/Didymoeca*
FJ848482	FRESCHO 3	Craspedida	*Salpingoeca prava*
FJ848483	FRESCHO 3	Acanthoecida	*Pleurasiga/Didymoeca*
FJ848484	FRESCHO 3	Acanthoecida	*Pleurasiga/Didymoeca*
FJ848485	FRESCHO 3	Acanthoecida	*Pleurasiga/Didymoeca*
FJ848486	FRESCHO 3	Craspedida	*Salpingoeca prava*
FJ848490	FRESCHO 3	Acanthoecida	*Pleurasiga/Didymoeca*
FJ848492	FRESCHO 3	Acanthoecida	Basal Acanthoecida
FJ848494	FRESCHO 3	Acanthoecida	Basal Acanthoecida
FJ848496	FRESCHO 3	Acanthoecida	Basal Acanthoecida
FJ848499	FRESCHO 3	Acanthoecida	Basal Acanthoecida
FJ848502	FRESCHO 3	Acanthoecida	Basal Acanthoecida
FJ848504	FRESCHO 3	Acanthoecida	*Pleurasiga/Didymoeca*
FJ848505	FRESCHO 3	Acanthoecida	Basal Acanthoecida
FJ848509	FRESCHO 3	Acanthoecida	*Pleurasiga/Didymoeca*
FJ848511	FRESCHO 3	Acanthoecida	*Pleurasiga/Didymoeca*
GQ844419	FRESCHO 3	Acanthoecida	*Didymoeca/Bicosta*
JF720681	FRESCHO 3	Acanthoecida	*Didymoeca/Bicosta*
JF774961	FRESCHO 3	Acanthoecida	Basal Acanthoecida
JF774993	FRESCHO 3	Acanthoecida	*Didymoeca/Bicosta*
AB238174	FRESCHO 4	Acanthoecida	*Pleurasiga/Didymoeca*
EU860443	FRESCHO 4	Clade 2	Basal Codosigid
EU860468	FRESCHO 4	Acanthoecida	*Pleurasiga/Didymoeca*
EU860494	FRESCHO 4	Acanthoecida	Basal Acanthoecida
EU860510	FRESCHO 4	Acanthoecida	*Pleurasiga/Didymoeca*
EU860513	FRESCHO 4	Acanthoecida	*Pleurasiga/Didymoeca*
EU860518	FRESCHO 4	Acanthoecida	*Pleurasiga/Didymoeca*
EU860526	FRESCHO 4	Acanthoecida	*Pleurasiga/Didymoeca*
EU860777	FRESCHO 4	Acanthoecida	*Didymoeca/Bicosta*
EU860781	FRESCHO 4	Craspedida	Basal Craspedida
EU860845	FRESCHO 4	Acanthoecida	*Didymoeca/Bicosta*
EU860868	FRESCHO 4	Clade 2	*Salpingoeca oahu*
EU860873	FRESCHO 4	Acanthoecida	*Didymoeca/Bicosta*
EU860874	FRESCHO 4	Acanthoecida	*Didymoeca/Bicosta*
EU860941	FRESCHO 4	Clade 2	Basal Codosigid
EU860943	FRESCHO 4	Clade 2	Basal Codosigid
EU860949	FRESCHO 4	Acanthoecida	*Didymoeca/Bicosta*
EU860967	FRESCHO 4	Acanthoecida	*Didymoeca/Bicosta*
GU970254	FRESCHO 4	Acanthoecida	*Didymoeca/Bicosta*
AM179824	FRESCHO 6	Clade 1	Basal Clade 1
AY969234	Clade L	Acanthoecida	Basal Acanthoecida
DQ120003	Clade L	Craspedida	Clade1/Clade2
EF024885	Clade L	Acanthoecida	Basal Acanthoecida
EF196724	Clade L	Acanthoecida	Basal Acanthoecida
EF196742	Clade L	Acanthoecida	Basal Acanthoecida
EF196744	Clade L	Acanthoecida	Basal Acanthoecida

EF196771	Clade L	Acanthoecida	Basal Acanthoecida
EF196774	Clade L	Acanthoecida	Basal Acanthoecida
EF196791	Clade L	Acanthoecida	Basal Acanthoecida
EF196793	Clade L	Acanthoecida	Basal Acanthoecida
GU647170	Clade L	Acanthoecida	Basal Acanthoecida
HM369560	Clade L	Acanthoecida	Basal Acanthoecida
HQ394139	Clade L	Acanthoecida	Basal Acanthoecida
HQ866504	Clade L	Acanthoecida	Basal Acanthoecida
HQ870125	Clade L	Acanthoecida	Basal Acanthoecida
HQ870522	Clade L	Acanthoecida	Basal Acanthoecida
HQ870602	Clade L	Acanthoecida	Basal Acanthoecida

Chapter III, Section I.

Protocol A3.1.1: WC media production protocol.

Step1. Dilution of the following solutes in 700ml ultrapure MILLI-Q® water (Merc KgaA, Germany):

Solute	Concentration (g/l)
Tris base	0.0606
EDTA	0.0146
NaCl	0.03

Step 2. Addition of 1ml of the following solutions:

Solution	Initial concentration of the solution (g/l)
$NaNO_3$	42.5
$MgSO_4x7H_2O$	36.97
$CaCl_2x2H_2O$	36.76
K_2HPO_4	4.35
$NaHCO_3$	31.5
$Na_2O_3Six5H_2O$	21.2
$FeCl_3x6H_2O$	3.15

Step 3. Addition of 1ml of the following solution: premixed trace elements stock

Solute	Initial concentration of the solute (g/l)
Na_2EDTA	4.36
$CuSO_4x5H_2O$	0.01
$ZnSO_4x7H_2O$	0.22
$CoCl_2x6H_2O$	0.01

Solute	Initial concentration of the solute (g/l)
$MnCl_2x4H_2O$	0.18
$NaMoO_4x2H_2O$	0.006
H_3BO_3	0.13

Step 4. Addition of utrapure MILLI-Q® water water to a final volume of 1000ml.

Step 5. pH adjustment at pH8.0.

Step 6. Sterilisation by autoclaving.

Step 7. Aseptic addition of 1ml of the following solution: premixed vitamins stock.

Solute	Initial concentration of the solute (g/l)
Thiamin HCl	0.1
Biotin	0.0005
Cyanocobalamin (Vitamin B12)	0.0005

Table A3.1.1: Typical composition of Instant Ocean® sea salt mix. Concentration of the ions expressed in parts per million (ppm).

Ion	Concentration (ppm)
Chloride	19290
Sodium	10780
Sulfate	2660
Magnesium	1320
Potassium	420
Calcium	400
Carbonate/bicarbonate	200
Bromide	56
Strontium	8.8
Boron	5.6
Fluoride	1
Lithium	0.3

Ion	Concentration (ppm)
Iodide	0.24
Barium	<0.04
Iron	<0.04
Manganese	<0.025
Chromium	<0.015
Cobalt	<0.015
Copper	<0.015
Nickel	<0.015
Selenium	<0.015
Vanadium	<0.015
Zinc	<0.015
Molybdenum	<0.01
Aluminum	<0.006
Lead	<0.005
Arsenic	<0.004
Cadmium	<0.002

Chapter III, Section II.

Table A3.2.1a: Culture growth, under High salinity treatment (30 PSU), throughout the experiment "Quantification of the tolerance to abrupt salinity changes in *M. roanoka*". Number of viable cells per culture and microscope field at 600x magnification at different times (t) (in hours).

	High (30 PSU)				
t (hours)	H1	H2	H3	H4	H5
0	3	0	0	0	0
	3	1	5	0	2
	2	0	4	0	4
	0	3	1	2	0
	7	0	1	0	2
24	2	1	2	4	4
	5	1	0	7	2
	0	0	0	2	0
	1	4	0	0	0
	2	0	7	2	2
48	17	4	19	13	13
	24	7	9	10	6
	18	11	9	7	12
	19	3	8	10	6
	16	2	10	4	6
72	23	16	8	39	26
	25	34	25	25	43
	28	28	17	16	45
	26	21	23	24	40
	28	19	36	21	26
96	19	27	31	14	46
	17	29	28	27	26
	13	19	36	29	42
	23	35	48	25	41
	23	26	31	24	40
144	44	52	36	47	46
	41	77	22	65	50
	44	91	37	57	44
	55	75	21	68	37

	49	83	48	44	38
	68	66	81	53	51
	81	57	58	44	53
168	52	66	51	54	38
	49	75	53	33	47
	67	74	49	63	53
	80	104	85	87	68
	51	90	83	84	71
238	72	88	68	82	72
	55	66	74	77	89
	46	89	73	73	52

Table A3.2.1b: Culture growth, under Medium salinity treatment (15 PSU), throughout the experiment "Quantification of the tolerance to abrupt salinity changes in *M. roanoka*". Number of viable cells per culture and microscope field at 600x magnification at different times (t) (in hours).

	Medium (15 PSU)				
t (hours)	M1	M2	M3	M4	M5
	0	0	0	1	1
	2	0	1	0	1
0	0	0	1	0	2
	4	3	0	1	0
	5	0	4	0	2
	2	2	4	1	5
	5	3	3	0	4
24	4	0	2	4	5
	7	1	4	3	2
	2	0	4	2	4
	50	36	35	22	30
	50	34	33	23	34
48	61	24	31	21	27
	90	20	39	14	23
	58	21	34	22	21
	98	77	83	67	62
	112	76	81	87	63
72	99	70	62	89	72
	118	82	66	69	96
	118	103	65	78	72
96	91	80	83	96	62
	127	70	81	108	69
	89	65	79	94	66
	87	78	96	94	85

	106	54	95	77	85
	120	134	113	165	153
	146	110	113	133	126
144	115	95	111	145	143
	143	104	114	125	136
	115	116	120	136	134
	169	170	164	184	164
	164	163	167	197	168
168	157	157	165	166	169
	171	180	154	210	143
	209	211	143	191	173
	292	243	215	254	206
	322	220	208	214	164
238	296	205	215	226	168
	334	221	185	200	206
	288	223	183	200	199

Table A3.2.1c: Culture growth, under Low salinity treatment (0 PSU), throughout the experiment "Quantification of the tolerance to abrupt salinity changes in *M. roanoka*". Number of viable cells per culture and microscope field at 600x magnification at different times (t) (in hours).

	Low (0 PSU)				
t (hours)	L1	L2	L3	L4	L5
	0	0	3	0	0
	0	2	1	0	1
0	0	2	1	0	1
	0	2	3	0	0
	1	2	2	0	0
	1	0	0	0	0
	0	0	0	0	0
24	2	0	0	0	0
	0	0	0	0	2
	1	1	0	0	0
	0	2	1	0	1
	3	2	3	0	2
48	1	2	2	1	1
	1	1	0	1	0
	0	5	1	0	0
	0	2	4	0	0
	1	3	4	0	0
72	1	0	5	2	0
	1	5	1	0	0
	2	2	0	1	2
96	0	3	1	1	0

	0	3	0	1	0
	1	1	2	0	0
	0	1	3	0	0
	0	0	0	0	4
	0	1	1	0	0
	0	1	0	0	0
144	0	1	0	0	1
	1	0	2	0	0
	0	0	1	0	0
	0	0	1	0	0
	0	0	0	0	0
168	0	0	0	0	0
	0	0	0	0	0
	0	0	0	1	1
	1	0	0	0	1
	0	1	0	3	0
238	0	0	0	0	0
	0	1	0	0	1
	0	1	1	0	1