The University of South Bohemia in České Budějovice
Faculty of Science

Exploration of contaminations in *Mesodinium* species and its impact on phylogenomic analysis.

Bachelor Thesis

Thisara Nugapitiya

Advisor: Dr. Martin Kolisko

České Budějovice 2024

Bibliography:

Nugapitiya, T., 2024: Exploration of contaminations in *Mesodinium* species and its impact on phylogenomic analysis. BC. Thesis, in English. – p 29, Faculty of Science, University of South Bohemia, České Budějovice, Czech Republic.

Annotation:

Mesodinium species has been a challenging species to position on phylogenomic tree due to its complex genomic structure and mixotrophic nutritional modes. This study aims to clarify this species true phylogenetic placement by categorizing genes with gene trees having contradictory positioning of *Mesodinium* sp. and then reconstructing phylogenetic trees with each group to observe if the position varies.

Declaration:

I declare that I am the author of this qualification thesis and that in writing it I have used the sources and literature displayed in the list of used sources only.

Klagenfurt, Austria, 09/12/2024

.....

Place, date

.....

Student's signature

Table of Contents

Abstract 1

Introduction 2

Methodology 16

Results 17

Discussion 19

References 21

Abstract

Mesodinium sp. belonging to phylum *Ciliophora*, can be responsible for creating dense marine algal blooms, that are sometimes large enough to be seen from satellites and plays an important role in nutrient cycling. They possess a complex genetic make up due to their mixotrophic lifestyle and kleptoplasty. However, accurate phylogenomic placement of this species has been a challenge as it has been placed either at the base of the ciliates or as a crown species. In different phylogenomic studies it has been hypothesized that its deep branching position is due to contamination by other organisms. The aim of this research is to figure out which one of these two placements is the most likely, by identifying the genes that branch either at the base or outside of the ciliates. These genes have then been assembled into another taxon within the phylogenomic matrix. We expect the two alternative *Mesodinium* datasets will either be placed as a crown species or remain as a true deep branching ciliate lineage. This will allow us to pinpoint or rule out the source of difference between the two conflicting topologies. By examination of these two distinct trees we would be able to determine if *Mesodinium* is placed as a crown species or as a basal lineage. These analyses will help to clarify the phylogenetic positioning and reveal if contaminations or true genetic divergence influence *Mesodinium's* place within ciliates.

Introduction

Phylogenomics

Since the invent of DNA sequencing, the use of sequence data for phylogenetics (1) have became a major tool for inferring relationships between organism. Initially, this has been done using single, usually highly conserved gene markers. Gene encoding the small subunit ribosomal RNA was and still is used predominantly to classify living organisms (2). Over time, as more different genes were analyzed, more topological incongruences were observed between phylogenies which were based on single gene data. Most importantly, single gene data were often insufficient to provide well resolved statistically supported phylogeny (3). The next generation sequencing methods, that made it cheap and easy to collect large amount of sequence data from almost any organism of interest, led to the "invention" of phylogenomics. Phylogenomics aim to combine hundreds of genes into one large dataset for inferring phylogenies (4). The basic assumption of phylogenomics is that each gene carries little bit of true phylogenetic signal and some mostly random noise and that by combining multiple genes together will provide sufficient "true" signal (and still lots of random noise) that will then lead to well resolved phylogenetic tree (5).

A phylogenetic tree, otherwise known as a cladogram is a visual representation in which various phylogenetic and evolutionary relationships between biological taxa are denoted based on their morphological characteristics or genetics (6-8). Initially, homologous DNA (or protein) sequences are collected from public databases (like GenBank, DDBJ or EMBL) or through the performed experiments. Sequence alignment is then performed and there are multitude of accurate methods for multiple sequence alignment (9). Trimming is essential before tree inference is performed to filter out any unreliable regions of the alignment that might affect the true outcome (10). Trimming more than what is required can eliminate true evolutionary signal from the final analysis while insufficient post alignment editing could introduce noise (11, 12). Selection of an appropriate algorithm for tree reconstruction is important (13). The two major categories of techniques for tree inferences (14) are distance-based methods (like NJ and UPGMA methods) (15) and character-based methods (parsimony and likelihood methods). Distance based methods measure the evolutionary distance between organisms by computing the sequence differences in the alignment (16). While the character based methods aim to search

all possible (or its approximation through heuristic search) trees and choose the one with best score based on the criterion(17). Today, maximum likelihood-based methods are generally always used as the methods of choice as they are best at encompassing the complexity of sequence evolution. As for the reliability of phylogenies based on morphological characters and or single genes depend on the suitability of the data and the reconstruction method. A reliable gene can be considered as genes that have undergone minimal change over time. Multiple changes create noise (homoplasy) hiding true phylogenetic signal (18).

In most phylogenomic studies the two main steps of classical phylogenetic inference are retained. Which is the identification of homologous characters (finding genes and constructing their alignments) and tree reconstruction.

In the Phylogenomic sequence based methods, multiple sequence alignments are used for tree reconstruction. Once the alignments are completed, tree reconstruction can be done in two ways. Via the so-called supermatrix or supertree approach. For super matrix approach, all of the alignments are concatenated into one large "super alignment", which is then analyzes as "one gene". The supermatrix methods is the most straight forward approach in "adding" the weak phylogenetic signal in individual genes for tree reconstruction. As "supermatrix" is made of multiple concatenated genes, many taxa might be missing from certain genes. The sequences of genes that are not present in an individual species are coded as dashes or sometimes as question marks. Many recent studies (19, 20, 21, 22) have shown that the effect of the missing data is relatively well tolerated during tree reconstruction. The reason that the impact of missing data is minimal, is likely that most taxa are still largely represented in other genes (19). The supermatrix approach is very effective for phylogenetic studies because it can handle missing data well. This makes it possible to create phylogenomic datasets cheaply by using information from existing databases (20, 23) or sequencing specific genes through PCR (24, 25, 26). As a result, this method allows researchers to include many different species in their analysis, rather than being limited to only those with fully sequenced genomes.

The second approach creates a "supertree" by combining all trees constructed from individual genes. Various methods for supertree construction have been proposed (27). But due to the inherent simplicity of the approach, matrix representation using parsimony (MRP) method(28, 29) is being used often(30). Examples of supertree methods in phylogenomics are reconstructing

the evolutionary relationships of bacteria (31) and model eukaryotic species with fully sequenced genomes (32). Another popular method using individual gene trees as input is ASTRAL (33; Accurate species tree algorithm). Astral is very effective at resolving conflict between individual genes cause by incomplete lineage sorting (ILS) (34). ILS is phenomenon when time between speciation events is shorter relatively to the size of population (35). To counter ILS the coalescent based approach ASTRAL (Accurate species tree algorithm) takes groups of four taxa (quartets) derived from individual gene trees, and then identify and resolve conflicts in gene tree topologies putatively caused by ILS. This method explicitly accounts for ILS since it is based on a multi-species coalescent model, which takes into account how genes coalesce back to their common ancestors over time. ASTRAL has been found out to be more accurate than the other methods under simulated conditions. The relative accuracy of ASTRAL as well as other methods like concatenation depend on the amount of ILS, where ASTRAL possesses an advantage when the ILS levels are at least. This approach also enables analysis with unrooted gene trees as input trees broadening it's application (33).

Another approach to genome-wise phylogenetic analysis is genome feature based method, that avoids the requirement of multiple sequence alignment step. But it still depends on the correct assessment of homology and orthology for accurate inferences. In this method the presence or absence of genes in a genome is noted and analyzed to infer evolutionary relationships between organisms. Each gene or a set of genes is assigned a binary character which indicates its presence or absence in the genome. Changes in gene content are considered to be robust since it involves numerous combinations of genes presence and absence across a genome. This leads to a less likelihood of Homoplasy caused by convergence or reversal(36), potentially making them good phylogenetic markers (37). Once the gene content matrix is generated for the organisms of interest, standard phylogenetic tree reconstruction methods like distance (37-43) and parsimony(43-46) are generally used. A phenomenon known as "big genome attraction"(47) hinders the use of this method where organisms with large genomes tend to get grouped together due to the sheer number of shared gene absences rather than due to true evolutionary relationships. Gene order (the rearrangement of genes within a genome) was identified as a valuable marker for evolutionary studies since it reflected rearrangements in the genome across time.(48) The mathematically complex nature of this approach where the estimation if the evolutionary distance between genomes based on the number of rearrangements required to

transform one genome into another.(49) Current methods only consider inversions but new methods are being developed to include a broader range of rearrangement events such as duplications, insertions and deletions into phylogenetic reconstruction (50).

The main prerequisite for any concatenated phylogenomic analyses is “congruence” of the phylogenetic signal between individual genes that are being used for tree reconstruction. There are three main sources of incongruent signal in phylogenomics: paralogy, laterally transferred genes and incomplete lineage sorting (see above). Paralogy originates via gene duplications. Because gene can duplicate at any point within its evolutionary history, paralogs can exist in all studies organisms (i.e.: in all eukaryotes), only within a specific group (i.e.: paralogs only within animals or vertebrates) and also only within one species (i.e.: multiple copies of individual gene only within *Tetrahymena thermophila*). Combining gene sequences from multiple species that are paralogous to each other will lead to false signal (Fig 1). Therefore, only orthologous genes (genes derived from an ancestor from speciation events) should be included in the analyses and careful paralog/ortholog assessment is crucial in phylogenomic studies. This assessment is traditionally carried out manually through inspection of individual gene datasets (51). The

movement of genetic material between species (horizontal gene transfer) can also disrupt the assumption that gene similarity reflects orthology (52).

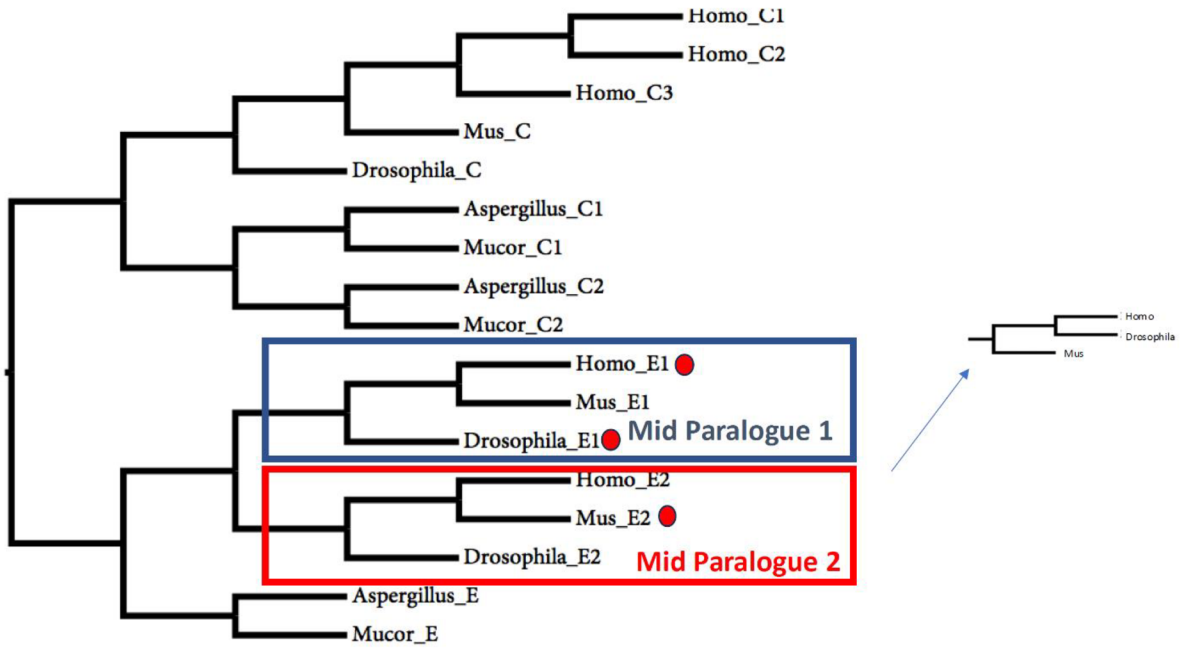

Figure 1 False phylogenetic interpretation caused by the inclusion of multiple paralogous genes.

Phylogenomic studies focusing Eukaryotes where sequence based approach was used at a larger scale, have confirmed the monophyly of many phyla which were defined using morphological features and rRNA data (53). Although, the clarification of the eukaryotic tree of life is often held back due to the lack of genomic data from several key species (*Rhizaria, Cryptophytes, Haptophytes and Jakobids*). Therefore, confirmation of the hypothesis of the six proposed Eukaryotic supergroups (*Opisthokonta, Amoebozoa, Plantae, Chromalveolata, Rhizaria, Excavata*) (54, 55) stemming from decades worth of molecular and ultrastructural studies has

been difficult. But follow up phylogenomic studies (56) leads to a partially reshuffled tree with 7-8 major supergroups and still several lineages without proper home (Fig2).

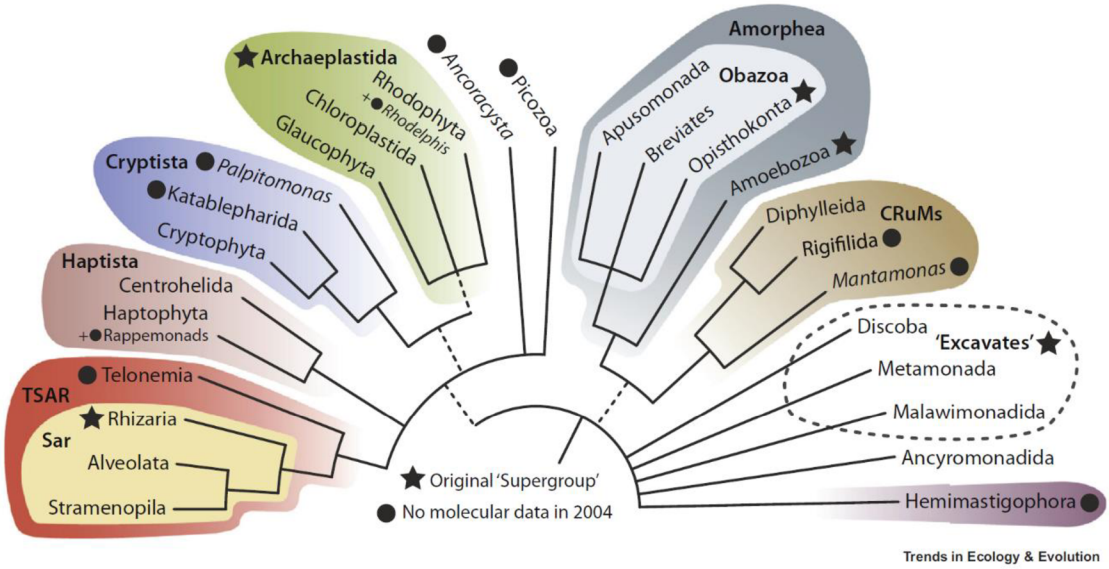

Figure 2 A summarization of the recent phylogenomic studies. Colored groups represent the supergroups. Bifurcated or multifurcated lines represent unresolved branching pattern amongst lineages. Dashed lines showcase minor uncertainties about the monophyly of some groups. The star annotate the supergroups which were also considered as supergroups in the previous version. The circular dots denote lineages with no molecular data at the time of the previous model. Modified from ADD NAME and et al.(56)

Another concern for phylogenomics are systematic errors. A systematic errors mean that, as more data of the same type are included, the stronger the signal for false results. The causal cause for a systematic errors is the failure of the assumptions made by the method about evolutionary processes to account for the properties of the data. In other words, the phylogenetic methods tend to oversimplify the evolutionary processes on the sequence level. This is main reason why methods like Maximum Parsimony, simple distance methods or even Maximum Likelihood with simple models are exceptionally susceptible. Systematic errors are amongst one of the earliest issues faced by phylogenetics (57). The most common sources of systematic errors are long branch attraction artifact and compositional bias. Long branch attraction is a common issue when a particular species evolves at a faster rate than the rest (57, 58). This is also called heterotachy. These fast-evolving species can be “attracted” to each other in phylogenies. This attraction is caused by the fact that as they mutate very fast any real

phylogenetic signal is overwritten by multiple mutations. As a result, these organisms often end up grouping falsely together even though they are not closely related (57). Compositional bias has similar effect, here species with biased composition of their sequences might appear closely related simply due to the fact that they are for example GC rich. While large datasets, like phylogenomic ones, generally result in an overall enhancement of the resolution of the phylogenetic trees by generating high statistical probabilities from either bootstrap analysis (59) or Bayesian posterior probabilities (60), they are not immune to these systematic errors. It is therefore important to use state of the art most realistic models, like for example mixture models and perform checks for presence of systematic errors like fast evolving site removal or amino acid recoding (61, 62).

Ciliophora and *Mesodinium*

Phylum *Ciliophora* being one of the most diverse of lineages amongst unicellular eukaryotes has been long recognized as a monophyletic group. Ciliates have been thriving since a billion years ago in a myriad of ecosystems, and this was long before the emergence of multicellular organisms. These are one of the most complex unicellular organisms both genetically and morphologically.

Ciliates are recognized as a taxonomic phylum related to two other sister phyla which are *Apicomplexa* and *Dinozoa*. This relation is due to these phyla sharing membrane bound alveoli and pellicles in common. These three phyla groups together are called Alveolates. The name Alveolates is due to the fact that they possess subpellicular alveoli. Small subunit ribosomal RNA genes (SSU-rRNA) also shows monophyly of these groups. The ciliates consist of eleven lineages (classes). These categorizations were initially made by factoring in the differences in somatic kinetid patterns, and sometimes oral structure patterns and cell division (63, 64). But morphological overlaps can be observed amongst these different classes. Subphylum *Postciliodesmatophora* has been confirmed from the gene sequence data of both SSU and LSU rRNA samples, which also suggests two main sub phyletic radiations (65, 66). The subphylum *Intramacronucleata* divides their macronuclei using an assembly of microtubules inside the

macronuclear envelope, hence the so-called intra macronuclear microtubules (65, 66). Unlike in *Postciliodesmatophora* with some similarities in the somatic kinetid arrangement across classes, *Intramacronucleata* shows a wide diversity in molecular phylogenies (Fig 5). Therefore, *Intramacronucleata* classes are categorized mostly based on the somatic kinetid arrangements, oral region features and oral development patterns (64, 67, 68). Ciliates exhibit both heterotrophic and autotrophic nutritional modes, like for example, *Mesodinium* that exhibits autotrophic nutritional mode by integrating prey chloroplasts within their cytoplasm.

The phylum Ciliophora can be characterized by three main features (6):

1.Ciliates exhibit nuclear dimorphism. A cellular condition in which two different kinds of nuclei are present in the cytoplasm. The first type is called the macronucleus, is transcriptionally active. This produces messenger ribonucleic acid (RNA) to maintain cellular functions and it usually contains many copies of the ciliate genome. Micronucleus which is diploid meaning it contains two sets of chromosomes serving as the germline deoxyribonucleic acid (DNA), undergoing meiosis during conjugation. Typically, the macronucleus contains many copies of the ciliate genome.

2. Conjugation is the second unique feature where through the temporary fusion by cytoplasmic bridge, two ciliates exchange gametic nuclei derived from the mitotic division of the micronuclei of each partner. Genetic recombination events might occur leading to greater genetic diversity with the phylum.

3. Ciliates possess cilia at least at one point of their life cycle. These hair-like structures are flagella that may envelope the entire cell, in some species forming specialized organelles called “cirri”. In some sessile species of ciliates, cilia can be localized to specific regions like the oral and the cilia appearing on the body or somatic regions only during dispersal. Other sessile ciliates might lack cilia in all stages except during the dispersal stage as observed in suctorians.

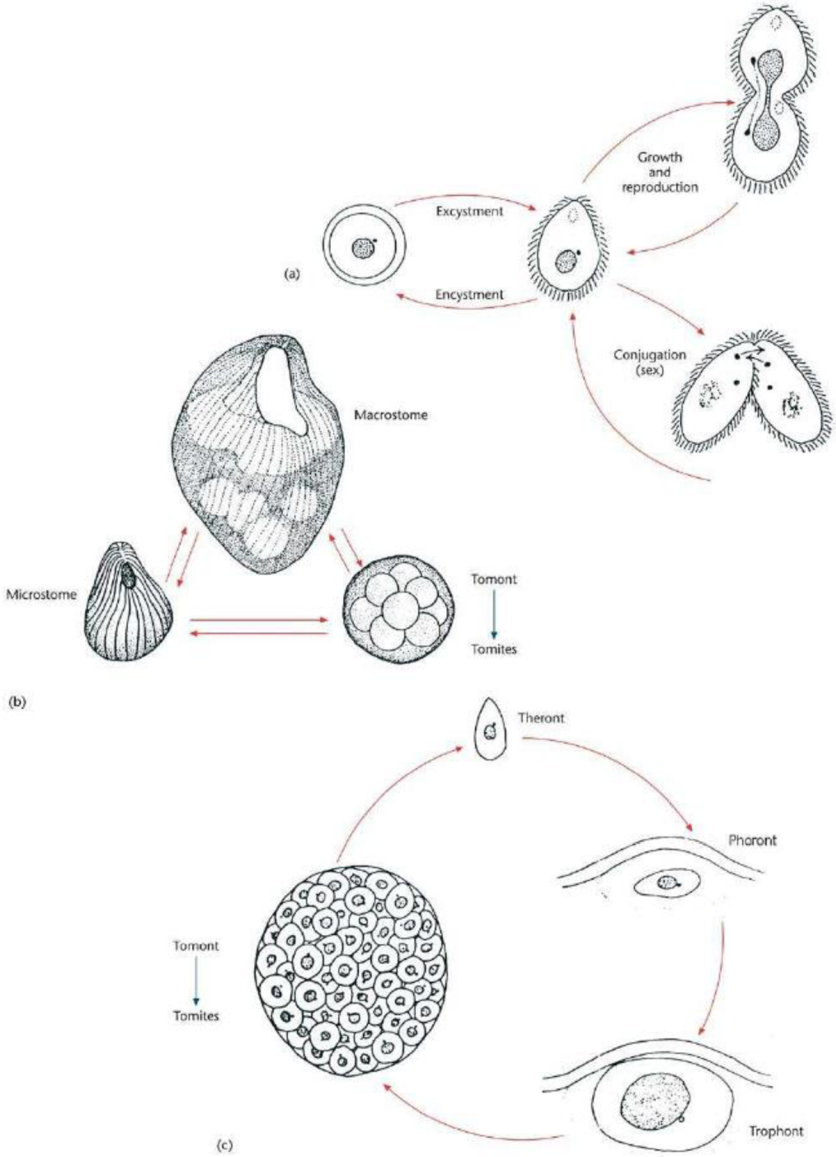

Figure 3 Depicts the life cycle of ciliated protists. (a) Shows the general 3 phases of the life cycle.(b) Life cycle of the hymenostome Tetrahymena patula which possesses a microstome phase which feeds on bacteria, a macrostome phase to feed on microstomes when bacterial food supply is low and a dividing phase (tomonts) and the tormites escape the cyst when bacteria are abundant to feed on. (c) Life cycle of hymenostome Ichthyophthirius multifiliis is a parasitic ciliate which cause the white spot disease on the skin of fish. The small cell (Theront) burrows under the skin of fish and converts into a Phoront that feeds. And a division stage similar to (b) (based on 64, 71) (69).

The survival strategies that have been adapted by the ciliate decide the sort of life cycle it has. The typical life cycle of a ciliate can be divided into three categories as in Fig 3a. The presence of prey organisms leads to ciliate growth and then reproduction. Reproduction is usually accomplished by transverse binary fission following bi-partitioning of the macronucleus and mitosis of the micronucleus (70). During nutrient scarce conditions, ciliates are often signaled to be sexually active entering the conjugation phase (Fig 3a). Researchers exploit this for studying the process in the laboratory to gain insight into the particular cellular and molecular processes occurring during the phase. When no sexual partners are available, ciliates undergo encystment (formation of a protective wall around the cell preventing desiccation) which is also seen in majority of the protists. When favorable conditions arise, the ciliate reverts back to trophozoites undergoing excysment. Dispersal of these ciliates to distant habitats most likely occur during the cyst stage. As observed in (Fig 3b) and (Fig 3c), there are many variations to this basic life cycle.

There is a clear observable diversity in the morphology across different species of ciliates, which has led to the majority of taxonomic classifications prior to phylogenetics. However, the basic morphological features can be generalized (Fig 4). The cilia can be separated into two main categories as oral and somatic. The somatic cilia reside on the surface arranged in longitudinal lines called kineties. Each kinety is composed of multiple kinetids which is composed of either one kinetosome and its cilium (monokinetids) or two kinetosomes and their cilia (dikinetids, Fig 4). These somatic cilia function as propellers for locomotion in ciliates. Some perioral kinetids modify into a more complex arrangement near the oral region assisting in food capturing (Fig 4). Oral cilia are arranged as dikinetids around the cytostome (cell mouth). Some of these dikinetids can form paroral kinety, which functions as a food filtration technique in water which are usually restricted to the right side of the oral region. Food particles are filtered by the compound movement of these oral cilia and formed into food vacuoles which are passed across the cytoplasm. After digestion, it is excreted via a cytoproct (permanent cell anus) (Fig 4). Water from the food vacuoles which is continuously phasing across the cell membranes is expelled through the contractile vacuole pore. Plasmalamella (cell membrane) covers the cell surface of ciliates. Underlying this layer are alveoli (Fig 4). Beneath these alveoli, a fibrous layer called the pellicle is formed (Fig 4). These fibrillar layers provide structural support for the

ciliates maintaining its unique cell shape. Together with the substructures of kinetids and pellicles, the ciliate cortex is formed.

There is a variation in the fibrillar arrangements in somatic kinetids for each group of ciliates. These variations have led to the characterizations of ciliate lineages in various studies (64, 67,71).

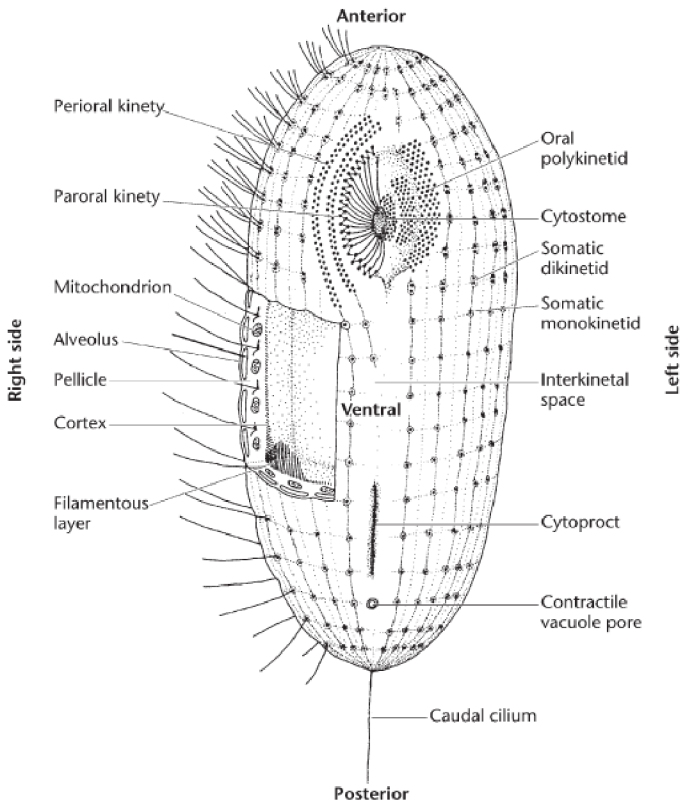

Figure 4 A generalized morphological image of Ciliates. (based on Lynn and Small 1989)(69).

There are about 8000 known species of ciliates where around two third are free living and the remaining are symbionts (69), but the true number could be drastically larger. The free living species are globally distributed (72) and can be commonly found in benthos or sediments in marine and freshwater ecosystems (72), in environmental soil (73), in marine plankton (74) and in fresh water (75). And they even have been adapted to living in extreme habitats like hot water springs.

Usually, ciliates are the apex species of microbial food webs, feeding on flagellates, phytoplankton and bacteria. In return these ciliates are consumed by jellyfish, zooplanktons and small fish species.

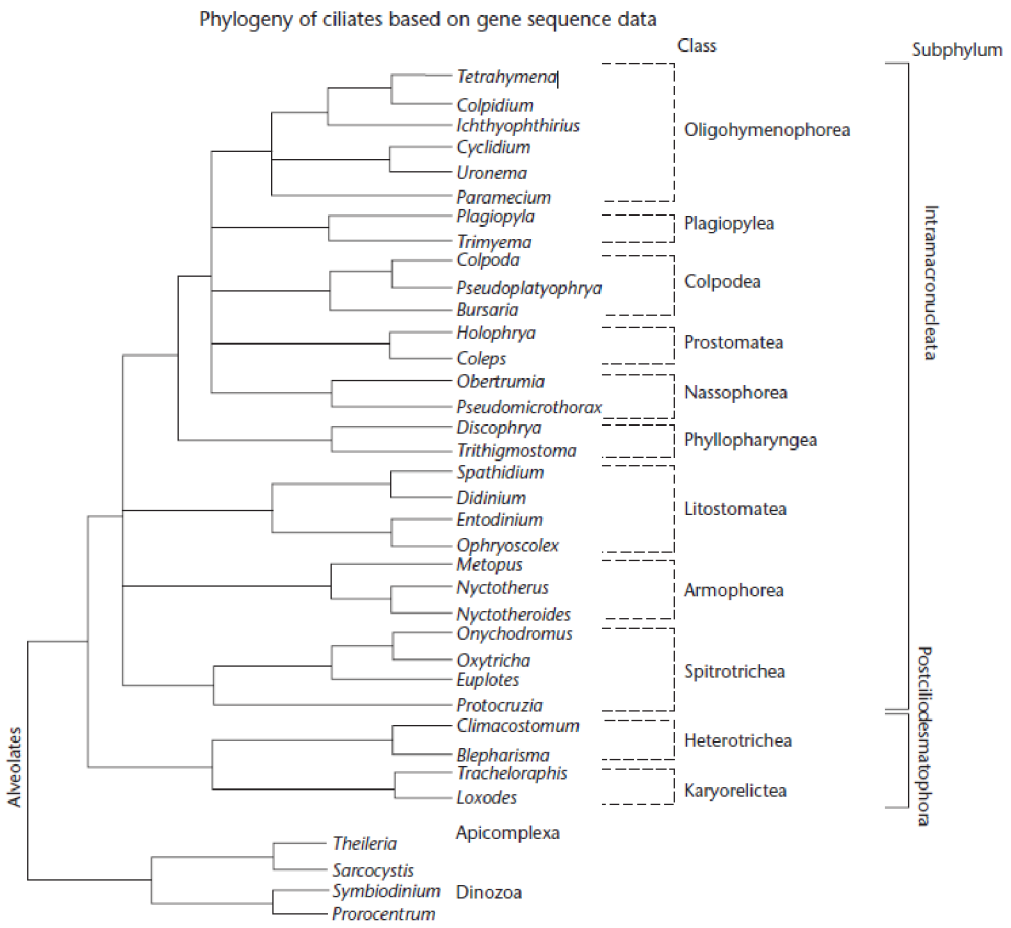

Figure 5 Phylogenies based on SSU-rRNA gene comparisons. The two other alveolate phyla Apicomplexa and Dinozoa serve as outgroups to root the ciliate section. (69)

In this study, we focused on the phylogenetic position of the genus *Mesodinium*. This genus as most ciliates has a global distribution, residing in fresh water and marine habitats. *Mesodinium* possesses a distinctive double banded arrangement around the midsection of the cell (Fig 6). The first *Mesodinium* sp. which was discovered was *Mesodinium rubrum* (by H. Lohmann, in 1908). This species has been the focus of many studies due to it being one of the main marine primary producers (76) and in certain aquatic ecosystems creating dense massive red algal

blooms (forming red tides) (77). Several other members of the genus have been identified since the discovery of *M. rubrum*. These form distinct individual phylogenetic lineages. But a categorization could be done according to the manner in which these species handle their prey. *M. rubrum* and *M. major* receive most of the required energy by photosynthesis via stolen prey chloroplasts, while *M. pulex* and *M. pupula* are exclusively heterotrophic while the two other lineages which are *M. chamaeleon* and *M. coatsi* being hybrid.

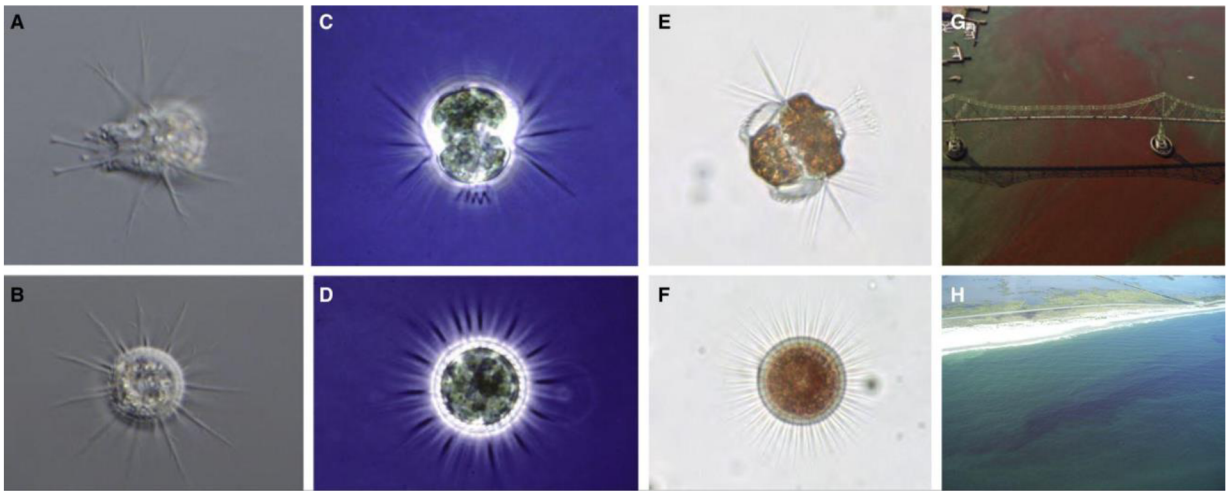

Figure 6 (A,B) Lateral and apical view of M. pulex with ingested prey. (C,D) Lateral and apical view of M. chamaeleon with phycocyanin plastids. (E,F) Lateral and apical view of M. rubrum with phycoerythrin plastids. (G) An M. rubrum bloom in the Columbia River, (H) outer banks of North Carolina, near Oregon, USA (78).

The possession of cryptophytic endosymbionts makes *Mesodinium* interesting from an evolutionary point of view (79, 80, 81). The morphological characteristics of this species have also not been verified due to the lack of complete descriptions of the intraciliary structures and silverline system (82). In some phylogenetic studies *Mesodinium* occupies an inconclusive phylogenetic position and was considered to cluster within *Litostomatea* in the order *Cyclotrichida* (76, 83). While some studies (84) defining *Mesodinium* as a species that branches early on the evolutionary timescale. *Mesodinium* is a phylogenetically hard to position taxon taxon, likely prone to artifacts as long branch attraction, at least in the small and large subunit rDNA gene trees (77, 82, 85, 86, 87, 88, 89, 90). It is a common practice to exclude *Mesodinium* in phylogenetic analysis due to their rapid evolutionary rates (82, 85, 91). Even though SSU rDNA sequences representing *Mesodinium rubrum* and *Mesodinium pulex* share three similarities with other *Litostomatea*, it still possesses additional deletions, substitutions and insertions which are not shared by other ciliates (77, 82, 92, 93). Morphological classification

has equally been problematic. In one study, light microscopic analysis of the species identified them as a well defined group due to their unique overall morphology (94). While in one, being placed within the class *Litostomatea* based on mouth positioning and inconspicuous oral ciliature (95). Or either being localized to an uncertain taxonomic status due to these unique characters and their somatic ciliature (67, 96). More recent analyses suggests that the only common feature between *Mesodiniidae* and *Litostomatea* was the ciliary transition region (81, 94, 97, 98) with the phylogenetic position of *Mesodinium* still being inconclusive.

A variety of nutritional modes are being adapted by *Mesodinium*. This mixotrophic nature has made *Mesodinium* (example- *M. rubrum* and *M. major*) to not only sequester the cryptophyte plastids and mitochondria, but the kleptokaryon as well. This is kept transcriptionally active for their own metabolism (79, 99, 100). As a resul, mixotrophic *Mesodinium* possess prey genetic material within their own. This mixing of *Mesodinium* and cryptophytic genetic material leads to conflicting phylogenetic placements of the genus.

In this study, the aim is to answer this contradictory placements of *Mesodinium*. Studies suggest that the placement of *Mesodinium* outside the Litostomatea is likely due to the data being mixed in with contaminating sequences (86, 89, 90). If the basal deep branching position of *Mesodinium* is due to a contamination, it would likely stem from genes introduced by an organism that is either a deep branching ciliate or a lineage in close proximity to ciliates. By removing genes associated with this basal signal we hypothesized that *Mesodinium* should either retain its basal position or shift to a crown species further verifying its phylogenetic positioning. This will allow us to gain further insight into ciliate evolution and to further clarify the Eukaryotic tree of life.

Methodology

The dataset of Rotterova et al. (101) was used as a starting point. The raw trees, which included all considered sequences for the phylogenomic analyses (i.e.: single gene datasets prior to ortholog and paralog selection), were manually parsed again using the software Parasorter (102) with special attention being paid to the position of *Mesodinium*. The genes were then split into two groups: one where *Mesodinium* branched basal to the Ciliates and another where it convincingly branched within Ciliates.

Next, in each final single gene dataset (i.e.: clean final single gene dataset just prior to concatenation) the *Mesodinium* were renamed to "DEEP" and "CROWN" based on whether *Mesodinium* branched basal to or within other Ciliates, respectively. The individual genes were then concatenated using *alvert.py* software from BARREL-o-MONKEYS (http://roger lab.biochemistryandmolecularbiology.dal.ca/Software/Software.htm#Monkeybarrel). A phylogenomic tree was constructed from the resulting dataset using Maximum Likelihood with C60+LG+G as implemented in the program IQ-Tree with 2000 ultrafast bootstraps.

To test for long branch attraction, the outgroup sequences and long branching *Protocruzia* ciliate were then manually removed (using the program Aliview) from the original final single gene alignments (i.e.: with *Mesodinium* not split into DEEP and CROWN). The phylogenetic tree was then recomputed using Maximum Likelihood with C60+LG+G as implemented in IQ-Tree with 2000 ultrafast bootstraps. The trees were visualized and saved as pdfs in the program TreeView.

Results

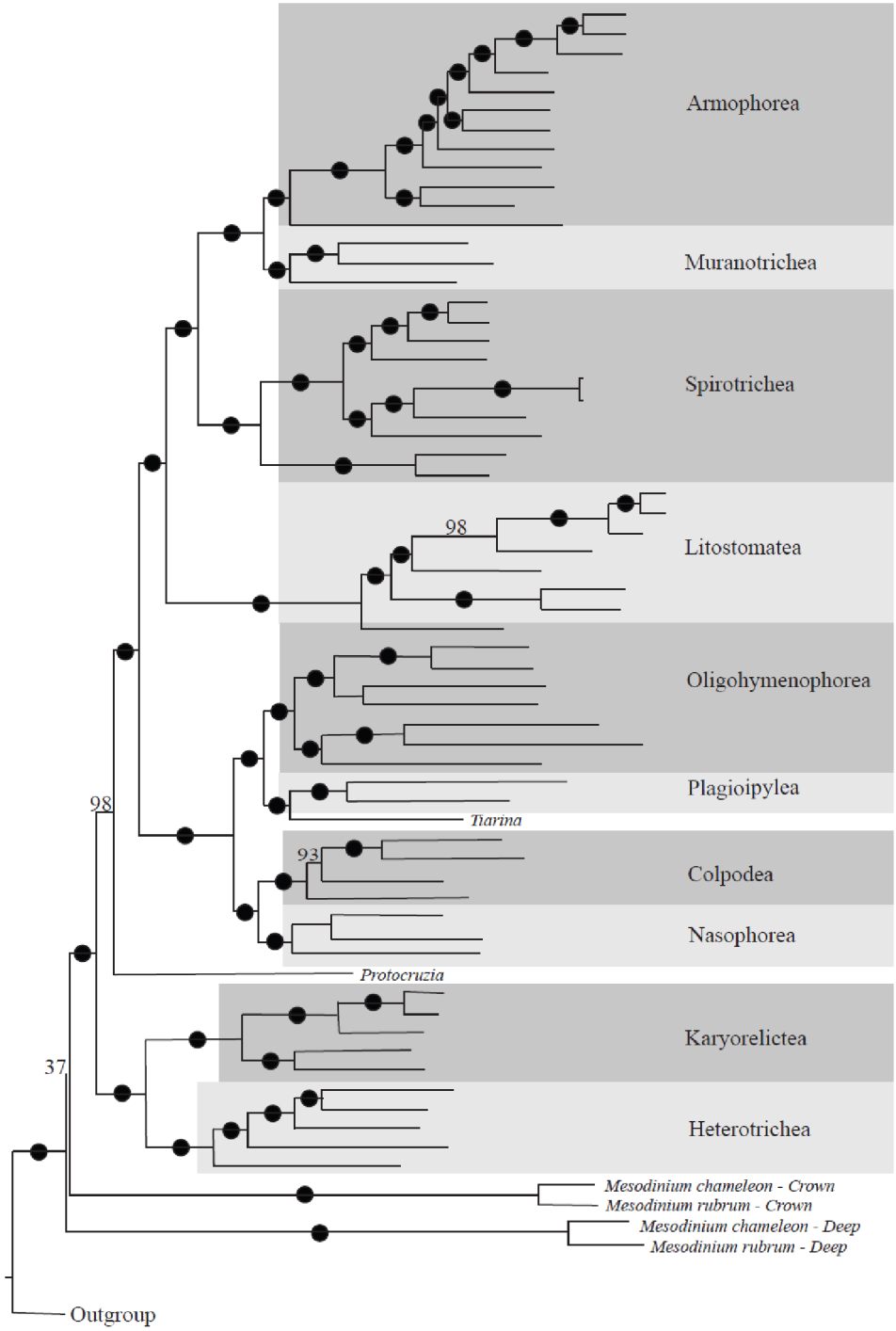

Figure 7 Phylogenomic tree with each Mesodinium species being split into two groups based on their position in single gene trees. The tree was computed in IQ-Tree using LG+C60+G model and 2000 ultrafast bootstraps. Black dots denote BP support 100.

Figure 8 Close up image of the branching pattern of the two groups of Mesodinium which where categorized as IN for the ones branching clearly within ciliates and OUT for the ones branching outside ciliates. (As also seen in figure 7).

Out of 206 genes, 174 show both *Mesodinium* species branching clearly within ciliates, while 36 had both *Mesodinium* species branching as basal taxa. After these genes were grouped into independent taxa marked as "DEEP" and "CROWN", the phylogenomic tree was reconstructed. Both groups of *Mesodinium* branched in a similar deep branching position, as basal taxa, and not with *Litostomatea (103)* (Figure 7 and 8). This was followed by removal of the outgroup and *Protocruzia*, the longest branching taxon, from the dataset. The resulting tree also showed *Mesodinium* as a deep branching taxon.

Discussion

Almost all previously published phylogenomic studies that included *Mesodinium* showed it to be a deep branching ciliate taxon (84, 101). This has been contradicted in 2019 by Lasek-Nesselquist et al (103), where it was suggested that this position of *Mesodinium* is artificial. The authors assume this is the result of contamination of *Mesodinium*, which might stem from food sources or happen during RNA isolation or sequencing. They carefully parsed *Mesodinium*'s position in all of their single gene trees and removed *Mesodinium* when it appeared as a contamination (i.e. not branching within ciliates), which resulted in *Mesodinium* branching within *Litostomatea* (103). However, during recent analyses done in collaboration between the laboratories of Dr. Martin Kolisko and Dr. Ivan Cepicka, *Mesodinium* still kept branching as a deep branching taxon even after rigorous contamination filtering (101). Even though *Mesodinium* was not included in the final analyses (101), it has posed an interesting phylogenomic problem. During preparation of the datasets, the authors included both *Mesodinium* sequences that branched convincingly inside the Ciliate monophylum as well as *Mesodinium* sequences that branched as the most basal Ciliate lineage and could in some extreme case represent a contamination. Here, we aimed to distinguish whether the position of *Mesodinium* in the Roterrova et al. (2024) (101) dataset is caused by inclusion of genes where *Mesodinium* branches at the base of ciliates and might represent contamination from some other ciliate group. Splitting the individual genes into two groups based on the phylogenetic position of *Mesodinium* showed no difference in the results (Fig 7).

Our results have clearly revealed that the position of *Mesodinium* in the Roterrova et al. (101) datasets has not been caused by insufficient removal of *Mesodinium* contaminations. While this is positive news for the future usage of the Rotterova et al. dataset, it does not explain the difference with the results of Lasek-Nesselquist et al. (103).

Another explanation could be long branch attraction (LBA) artifact, since *Mesodinium* forms fairly long branches on the phylogenetic tree (77, 82, 85, 86, 87, 88, 89, 90) and could be attracted to the long branching *Protocruzia* and outgroup taxa. Therefore, as the next step, we have removed the outgroup and ciliate *Protocruzia* from the dataset and recomputed the tree.

The constructed tree shows no change in *Mesodinium's* position, as it remains in a position consistent with the tree with full taxon sampling. It can therefore be concluded that the problematic position of *Mesodinium* in many phylogenomic datasets might not have only been caused by the inclusion of contaminations (even though in some early studies this might also have been a cause), but that the conflicting positions are either caused by gene sampling, taxon sampling, or some other biases in the preparation of the data.

The placement of *Mesodinium* as an early branching lineage of ciliates has a few implications. It might be considered additional evidence for a previous morphological study that identified *Mesodinium* as its own group (94). Early branching lineages can also provide valuable insights for understanding the evolution of traits of the group. As a continuation of our study, it would be interesting to perform additional analyses, such as including more *Mesodinium* species in our dataset or including genes used by Lasek-Nesselquist et al. (103) and performing full LBA analyses on both ours and the Lasek-Nesselquist et al. datasets. Additionally, further expanding the taxonomic sampling could help researchers clarify the precise placement of *Mesodinium*. Further studies of Mesodinium using functional genomics to investigate the genomic features which contributed to *Mesodinium*'s unique traits like kleptoplasty could reveal the evolutionary events which led to its complex genome. Comparative studies of *Mesodinium* to other basal lineages and crown ciliates could advance our knowledge of evolutionary traits across ciliates in the tree of life.

References

(1) Zuckerkandl, E. & Pauling, L. Molecules as documents of evolutionary history. J. Theor. Biol. 8, 357-366 (1965).

(2) Woese, C. R. & Fox, G. E. Phylogenetic structure of the prokaryotic domain: The primary kingdoms. Proc. Natl. Acad. Sci. USA 74, 5088-5090 (1977).

(3) Frédéric Delsuc, Henner Brinkmann, Hervé Philippe. Phylogenomics and the reconstruction of the tree of life.. Nature Reviews Genetics, 2005, 6 (5), pp.361-75. ff10.1038/nrg1603ff. ffhalsde-00193293f

(4) Eisen, J. A. & Fraser, C. M. Phylogenomics: intersection of evolution and genomics. Science 300, 1706-1707 (2003).

(5) Phylogenomic analysis: AJ Roger, M Kolisko, AGB Simpson - Evolution of virulence in eukaryotic microbes, (2012)(pg 44-69)

(6) Sanderson, M.J.; Driskell, A.C. The challenge of constructing large phylogenetic trees. Trends Plant Sci. 2003, 8, 374–379.

(7) Hug, L.A.; Baker, B.J.; Anantharaman, K.; Brown, C.T.; Probst, A.J.; Castelle, C.J.; Butterfield, C.N.; Hernsdorf, A.W.; Amano, Y.; Ise, K.; et al. A new view of the tree of life. Nat. Microbiol. 2016, 1, 16048.

(8) Abaza, S. What is and why do we have to know the phylogenetic tree? Parasitol. United J. 2020, 13, 68–71.

(9) Ashkenazy, H.; Sela, I.; Levy Karin, E.; Landan, G.; Pupko, T. Multiple Sequence Alignment Averaging Improves Phylogeny Reconstruction. Syst. Biol. 2019, 68, 117–130. [CrossRef]

(10) Francis, W.R.; Canfield, D.E. Very few sites can reshape the inferred phylogenetic tree. PeerJ 2020, 8, e8865.

(11) Talavera, G.; Castresana, J. Improvement of phylogenies after removing divergent and ambiguously aligned blocks from protein sequence alignments. Syst. Biol. 2007, 56, 564–577.

(12) Williams, T.A.; Heaps, S.E. An introduction to phylogenetics and the tree of life. In Methods in Microbiology; Elsevier: Amsterdam, The Netherlands, 2014; Volume 41, pp. 13–44.

(13) Desper, R.; Gascuel, O. Theoretical foundation of the balanced minimum evolution method of phylogenetic inference and its relationship to weighted least-squares tree fitting. Mol. Biol. Evol. 2004, 21, 587–598. [CrossRef]

(14) Wang, Z.; Sun, J.; Gao, Y.; Xue, Y.; Zhang, Y.; Li, K.; Zhang, W.; Zhang, C.; Zu, J.; Zhang, L. Fusang: A framework for phylogenetic tree inference via deep learning. Nucleic Acids Res. 2023, 51, 10909–10923. [CrossRef]

(15) Balaban, M.; Jiang, Y.; Roush, D.; Zhu, Q.; Mirarab, S. Fast and accurate distance-based phylogenetic placement using divide and conquer. Mol. Ecol. Resour. 2022, 22, 1213–1227.

(16) Vaz, C.; Nascimento, M.; Carriço, J.A.; Rocher, T.; Francisco, A.P. Distance-based phylogenetic inference from typing data: A unifying view. Brief. Bioinform. 2021, 22, bbaa147.

(17) Ojha, K.K.; Mishra, S.; Singh, V.K. Computational molecular phylogeny: Concepts and applications. In Bioinformatics; Academic Press: New York, NY, USA, 2022; pp. 67–89.

(18) Gribaldo, S. & Philippe, H. Ancient phylogenetic relationships. Theor. Popul. Biol.61, 391-408 (2002).

(19) Philippe, H. et al. Phylogenomics of eukaryotes: Impact of missing data on large alignments. Mol. Biol. Evol. 21, 1740-1752 (2004).

(20) Driskell, A. C. et al. Prospects for building the tree of life from large sequence databases. Science 306, 1172-1174 (2004).

(21) Gatesy, J., Matthee, C., DeSalle, R. & Hayashi, C. Resolution of a supertree/supermatrix paradox. Syst. Biol. 51, 652-664 (2002).

(22) Wiens, J. J. Missing data, incomplete taxa, and phylogenetic accuracy. Syst. Biol.52, 528-538 (2003).

(23) Sanderson, M. J., Driskell, A. C., Ree, R. H., Eulenstein, O. & Langley, S. Obtaining maximal concatenated phylogenetic data sets from large sequence databases. Mol. Biol. Evol. 20, 1036-1042 (2003).

(24) Qiu, Y. L. et al. The earliest angiosperms: Evidence from mitochondrial, plastid andnuclear genomes. Nature 402, 404-407 (1999).

(25) Madsen, O. et al. Parallel adaptive radiations in two major clades of placental mammals. Nature 409, 610-614 (2001).

(26) Murphy, W. J. et al. Molecular phylogenetics and the origins of placental mammals. Nature 409, 614-618 (2001).

(27) Bininda-Emonds, O. R. P., Gittleman, J. L. & Steel, M. A. The (Super)tree of life: Procedures, problems, and prospects. Annu. Rev. Ecol. Syst. 33, 265-289 (2002).

(28) Baum, B. Combining trees as a way of combining data sets for phylogenetic inference, and the desirability of combining gene trees. Taxon 41, 3-10 (1992).

(29) Ragan, M. A. Phylogenetic inference based on matrix representation of trees. *Mol. Phylogenet. Evol.* **1**, 53-58 (1992).

(30) Bininda-Emonds, O. R. P. The evolution of supertrees. Trends Ecol. Evol. 19, 315- 322 (2004).

(31) Daubin, V., Gouy, M. & Perrière, G. A phylogenomic approach to bacterial phylogeny: Evidence of a core of genes sharing a common history. Genome Res. 12, 1080-1090 (2002).

(32) Philip, G. K., Creevey, C. J. & McInerney, J. O. The Opisthokonta and the Ecdysozoa may not be clades: Stronger support for the grouping of plant and animal than for animal and fungi and stronger support for the Coelomata than Ecdysozoa. Mol. Biol. Evol. Published online Feb 9 (2005).

(33) S. Mirarab, R. Reaz, Md. S. Bayzid, T. Zimmermann, M. S. Swenson, T. Warnow, ASTRAL: genome-scale coalescent-based species tree estimation, Bioinformatics, Volume 30, Issue 17, September 2014, Pages i541–i548, https://doi.org/10.1093/bioinformatics/btu462

(34) J.F.C. Kingman, The coalescent, Stochastic Processes and their Applications, Volume 13, Issue 3, 1982, Pages 235-248, ISSN 0304-4149,https://doi.org/10.1016/0304-4149(82)90011-4.

(35) Wayne P. Maddison, Gene Trees in Species Trees, Systematic Biology, Volume 46, Issue 3, September 1997, Pages 523–536, https://doi.org/10.1093/sysbio/46.3.523.

(36) Van Niel, C. B. in *Perspectives and Horizons in Microbiology* 3-12 (Rutgers University Press, New Brunswick, 1955).

(37) Gribaldo, S. & Philippe, H. Ancient phylogenetic relationships. Theor. Popul Biol. 61, 391-408 (2002).

(38) Snel, B., Bork, P. & Huynen, M. A. Genome phylogeny based on gene content. Nat. Genet. 21, 108-110 (1999).

(39) Tekaia, F., Lazcano, A. & Dujon, B. The genomic tree as revealed from whole proteome comparisons. Genome Res. 9, 550-557 (1999).

(40) Clarke, G. D., Beiko, R. G., Ragan, M. A. & Charlebois, R. L. Inferring genome trees by using a filter to eliminate phylogenetically discordant sequences and distance matrix based on mean normalized BLASTP scores. J. Bacteriol. 184,2072-2080 (2002).

(41) Korbel, J. O., Snel, B., Huynen, M. A. & Bork, P. SHOT: A web server for the construction of genome phylogenies. Trends Genet. 18, 158-162 (2002).

(42) Dutilh, B. E., Huynen, M. A., Bruno, W. J. & Snel, B. The consistent phylogenetic signal in genome trees revealed by reducing the impact of noise. J. Mol. Evol. 58, 527-539 (2004).

(43) Lin, J. & Gerstein, M. Whole-genome trees based on the occurrence of folds and orthologs: implications for comparing genomes on different levels. Genome Res. 10, 808-818 (2000).

(44) Wolf, Y. I., Rogozin, I. B., Grishin, N. V., Tatusov, R. L. & Koonin, E. V. Genome trees constructed using five different approaches suggest new major bacterial clades. BMC Evol. Biol. 1, 8 (2001).

(45) Fitz-Gibbon, S. T. & House, C. H. Whole genome-based phylogenetic analysis of free-living microorganisms. Nucleic Acids Res. 27, 4218-4222 (1999).

(46) House, C. H. & Fitz-Gibbon, S. T. Using homolog groups to create a whole genomic tree of free-living organisms: An update. J. Mol. Evol. 54, 539-547 (2002).

(47) Lake, J. A. & Rivera, M. C. Deriving the genomic tree of life in the presence of horizontal gene transfer: Conditioned reconstruction. Mol. Biol. Evol. 21, 681-690 (2004).

(48) Sankoff, D. et al. Gene order comparisons for phylogenetic inference: Evolution of the mitochondrial genome. Proc. Natl. Acad. Sci. USA 89, 6575-6579 (1992).

(49) Hannenhalli, S. & Pevzner, P. A. Transforming cabbage into turnip: Polynomial algorithm for sorting signed permutations by reversals. J. ACM 46, 1-27 (1999).

(50) Moret, B., Tang, J. & Warnow, T. in Mathematics of Evolution and Phylogeny (ed. Gascuel, O.) 321-352 (Oxford Univ. Press, Oxford, 2005).

(51) Koski, L. B. & Golding, G. B. The closest BLAST hit is often not the nearest neighbor. J. Mol. Evol. 52, 540-542 (2001).

(52) Philippe, H. & Douady, C. J. Horizontal gene transfer and phylogenetics. Curr. Opin. Microbiol. 6, 498-505 (2003).

(53) Bapteste, E. et al. The analysis of 100 genes supports the grouping of three highly divergent amoebae: Dictyostelium, Entamoeba, and Mastigamoeba. Proc. Natl. Acad. Sci. USA 99, 1414-1419 (2002).

(54) Stechmann, A. & Cavalier-Smith, T. Rooting the eukaryote tree by using a derived gene fusion. Science 297, 89-91 (2002).

(55) Simpson, A. G. & Roger, A. J. The real 'kingdoms' of eukaryotes. Curr. Biol. 14, R693-R696 (2004).

(56) The New Tree of Eukaryotes: Burki, Fabien et al. Trends in Ecology & Evolution, Volume 35, Issue 1, 43 – 55

(57) Felsenstein, J. Cases in which parsimony or compatibility methods will be positively misleading. Syst. Zool. 27, 401-410 (1978).

(58) Hendy, M. & Penny, D. A framework for the quantitative study of evolutionary trees. Syst. Zool. 38, 297-309 (1989).

(59) Felsenstein, J. Confidence limits on phylogenies: An approach using the bootstrap. Evolution 39, 783-791 (1985).

(60) Huelsenbeck, J. P., Ronquist, F., Nielsen, R. & Bollback, J. P. Bayesian inference of phylogeny and its impact on evolutionary biology. Science 294, 2310-2314 (2001).

(61) Lartillot, N. & Philippe, H. A Bayesian mixture model for across-site heterogeneities in the amino-acid replacement process. Mol. Biol. Evol. 21, 1095-1109 (2004).

(62) Pagel, M. & Meade, A. A phylogenetic mixture model for detecting pattern heterogeneity in gene sequence or character-state data. Syst. Biol. 53, 571-581 (2004).

(63) Foissner W (1996) Ontogenesis in ciliated protozoa, with emphasis on stomatogenesis. In: Hausmann K and Bradbury PC (eds) Ciliates. Cells as Organisms, pp. 95–177. Stuttgart: Gustav Fischer.

(64) Lynn DH and Small EB (2002) Phylum Ciliophora. In: Lee JJ, Bradbury PC and Leedale GF (eds) An Illustrated Guide to the Protozoa, pp. 371– 656. Lawrence, KS: Society of Protozoologists.

(65) Lynn DH (1996b) My journey in ciliate systematics. Journal of Eukaryotic Microbiology 43: 253–260.

(66) Lynn DH (2004) Morphology or molecules: how do we identify the major lineages of ciliates (Phylum Ciliophora)? European Journal of Protistology 39: 356–364.

(67) Lynn DH and Corliss JO (1991) Ciliophora. In: Harrison FW and Corliss JO (eds) Microscopic Anatomy of Invertebrates. Protozoa, pp. 333– 467. New York: Wiley-Liss.

(68) Lynn DH and Small EB (1997) A revised classification of the Phylum Ciliophora Doflein, 1901. Revista de la Sociedad Mexicana de Historia Natural 47: 65–78.

(69) Ciliophora-Denis H Lynn, University of Guelph, Ontario, Canada(doi: 10.1038/npg.els.0004264)

(70) Raikov IB (1996) Nuclei of ciliates. In: Hausmann K and Bradbury PC (eds) Ciliates. Cells as Organisms, pp. 221–242. Stuttgart: Gustav Fischer.

(71) Lynn DH and Small EB (1989) Phylum Ciliophora. In: Margulis L, Corliss JO, Melkonian M and Chapman DJ (eds) Handbook of Protoctista, pp. 498–523. Boston: Jones and Bartlett Publishers

(72) Finlay BJ, Esteban GF and Fenchel T (1998) Protozoan diversity: Converging estimates of the global number of free-living ciliate species. Protist 149: 29–37.

(73) Foissner W (1987) Soil protozoa: fundamental problems, ecological significance, adaptations in ciliates and testaceans, bioindicators, and guide to the literature. In: Corliss JO and Patterson DJ (eds) Progress in Protistology, vol. 2, pp. 69–212. Bristol: Biopress Ltd.

(74) Pierce RW and Turner JT (1992) Ecology of planktonic ciliates in marine food webs. Reviews in Aquatic Sciences 6: 139–181.

(75) Fenchel T (1987) Ecology of Protozoa. The Biology of Free-living Phagotrophic Protists. Berlin: Springer.

(76) Lynn, D. H. 2008. Ciliated Protozoa: Characterization, Classification, and Guide to the Literature. Springer Verlag, Dordrecht. 605 pp.

(77) Johnson, M. D., Tengs, T, Oldach, D. W., Delwiche, C. F. and Stoecker, D. K. 2004. Highly divergent SSU rRNA genes found in the marine ciliates Myrionecta rubra and Mesodinium pulex. Protist, 155: 347–359.

(78) Mesodinium; Moeller, Holly V. et al. (Current Biology, Volume 33, Issue 7, R249-R250)

(79) Gustafson, D. E., Jr., Stoecker, D. K., Johnson, M. D., Van Heukelem, W.F. and Sneider, K. 2000. Cryptophyte algae are robbed of their organelles by the marine ciliate Mesodinium rubrum. Nature, 405: 1049–1 052.

(80) Lindholm, T., Lindroos, P. and Mork, A. C. 1988. Ultrastructure of the photosynthetic ciliate Mesodinium rubrum. Biosystems, 21: 141–149.

(81) Moestrup, Ø., Garcia-Cuetos, L., Hansen, P. J. and Fenchel, T. 2012. Studies on the genus Mesodinium I: ultrastructure and description of Mesodinium chamaeleon n. sp., a benthic marine species with green or red chloroplasts. Journal of Eukaryotic Microbiology, 59: 20–39.

(82) Strüder-Kypke, M. C., Wright, A. D., Foissner, W., Chatzinotas, A. and Lynn, D. H. 2006. Molecular phylogeny of litostome ciliates (Ciliophora, Litostomatea) with emphasis on free-living haptorian genera. Protist, 157: 261–278.

(83) Adl, S. M., Simpson, A. G., Lane, C. E., Lukes, J., Bass, D., Bowser, S. S., Brown, M. W., Burki, F., Dunthorn, M., Hampl, V. et al. 2012. The revised classification of eukaryotes. Journal of Eukaryotic Microbiology, 59: 429–493.

(84) *Zoological Systematics*, 40(1): 31–40 (January 2015), DOI:10.11865/zs.20150102

(85) Vd'acny P, BourlandWA, OrsiW, Epstein SS, FoissnerW. 2011. Phylogeny and classification of the Litostomatea (Protista, Ciliophora), with emphasis on free-living taxa and the 18S rRNA gene. Mol Phylogenet. Evol. 59:510–522.

(86) Chen X, Ma HG, Al-Rasheid KA, Miao M. 2015. Molecular data suggests the ciliate Mesodinium (Protista: Ciliophora) might represent an undescribed taxon at class level. Zool Syst. 40:31–40.;

(87) Chen X, Zhao X, et al. 2015. Phylogenomics of non-model ciliates based on transcriptomic analyses. Protein Cell 6(5):373–385.

(88) Gao F, et al. 2016. The all-data-based evolutionary hypothesis of ciliated protists with a revised classification of the phylum Ciliophora (Eukaryota, Alveolata). Sci Rep. 6(1):24874.

(89) Lynn DH, Kolisko M. 2017. Molecules illuminate morphology: phylogenomics confirms convergent evolution among "oligotrichous" ciliates. Int J Syst Evol Microbiol. 67:3676–3682.

(90) Lynn DH, Kolisko M, BourlandW. 2018. Phylogenomic analysis of Nassula variabilis n. sp., Furgasonia blochmanni, and Pseudomicrothorax dubius confirms a nassophorean clade. Protist 169(2):180–189.

(91) Gao F, Katz LA. 2014. Phylogenomic analyses support the bifurcation of ciliates into two major clades that differ in properties of nuclear division. Mol Phylogenet Evol. 70:240–243.

(92) Bass D, et al. 2009. A molecular perspective on ecological differentiation and biogeography of cyclotrichiid ciliates. J Eukaryot Microbiol. 56(6):559–567.

(93) Herfort L, et al. 2011. Myrionecta rubra population genetic diversity and its cryptophyte chloroplast specificity in recurrent red tides in the Columbia River estuary. Aquat Microb Ecol. 62(1):85–97.

(94) Garcia-Cuetos L, Moestrup Ø, Hansen PJ. 2012. Studies on the genus Mesodinium II. Ultrastructural and molecular investigations of five marine species help clarifying the taxonomy. J Eukaryot Microbiol. 59(4):374–400.

(95) Corliss JO. 1979. The ciliated protozoa: characterization, classification, and guide to the literature. London/New York: Pergamon Press.

(96) Adl SM, et al. 2019. Revisions to the classification, nomenclature, and diversity of eukaryotes. J Eukaryot Microbiol. 66(1):4–119.

(97) Nam SW, ShinW, Coats DW, Park JW, YihW. 2012. Ultrastructure of the oral apparatus of Mesodinium rubrum from Korea. J Eukaryot Microbiol. 59(6):625–636.

(98) Nam SW, Shin W, Kang M, Yih W, Park MG. 2015. Ultrastructure and molecular phylogeny of Mesodinium coatsi sp. nov., a benthic marine ciliate. J Eukaryot Microbiol. 62(1):102–120.

(99) Johnson MD, Oldach D, Delwiche CF, Stoecker DK. 2007. Retention of transcriptionally active cryptophyte nuclei by the ciliate Myrionecta rubra. Nature 445(7126):426–428.

(100) Kim M, Drumm K, Daugbjerg N, Hansen PJ. 2017. Dynamics of sequestered cryptophyte nuclei in Mesodinium rubrum during starvation and refeeding. Front Microbiol. 8:423.

(101) Rotterová J, Pánek T, Salomaki ED et al. Single cell transcriptomics reveals UAR codon reassignment in Palmarella Salina (Metopida, Armophorea) and confirms Armophorida belongs to APM clade. Mol Phylogenet Evol 2024;191:107991. https://doi. org/10.1016/j.ympev.2023.107991

(102) Tice AK,Zˇihala D, Pa´nek T, Jones RE, Salomaki ED, Nenarokov S, et al. (2021) PhyloFisher: A phylogenomic package for resolving eukaryotic relationships. PLoS Biol 19(8): e3001365. https://doi.org/10.1371/journal. pbio.3001365

(103) Erica Lasek-Nesselquist, Matthew D Johnson, A Phylogenomic Approach to Clarifying the Relationship of *Mesodinium* within the Ciliophora: A Case Study in the Complexity of Mixed-Species Transcriptome Analyses, *Genome Biology and Evolution*, Volume 11, Issue 11, November 2019, Pages 3218–3232, https://doi.org/10.1093/gbe/evz233