9-1-2011

Phylogenics and Patterns of Molecular Evolution in Amoebozoa

Daniel J.G. Lahr
dlahr@cns.umass.edu

PHYLOGENETICS AND PATTERNS OF MOLECULAR EVOLUTION IN AMOEBOZOA

A Dissertation Presented

by

DANIEL J. G. LAHR

Submitted to the Graduate School of the
University of Massachusetts Amherst in partial fulfillment
of the requirements for the degree of

DOCTOR OF PHILOSOPHY

September 2011

Program in Organismic and Evolutionary Biology

PHYLOGENETICS AND PATTERNS OF MOLECULAR EVOLUTION IN AMOEBOZOA

A Dissertation Presented

by

DANIEL J. G. LAHR

Approved as to style and content by:

_____
Laura A. Katz, Chair

_____
Benjamin B. Normark, Member

_____
Michael E. Hood, Member

_____
O. Roger Anderson, Member

_____
Elizabeth R. Dumont, Director
Organismic and Evolutionary Biology

DEDICATION

To Paula and my family.

"Without trace of nerve elements, and without definite, fixed organs of any kind, internal or external, the Rhizopod,—simplest of all animals, a mere jelly-speck,—moves about with the apparent purposes of more complex creatures. It selects and swallows its appropriate food, digests it, and rejects the insoluble remains. It grows and reproduces its kind. It evolves a wonderful variety of distinctive forms, often of the utmost beauty, and, indeed, it altogether exhibits such marvelous attributes, that one is led to ask the question in what consists the superiority of animals usually regarded as much higher in the scale of life."

Joseph Leidy *in* Freshwater Rhizopods of North America, 1879, p. 5

ACKNOWLEDGEMENTS

I am grateful to all the incredible people that supported me during the long years of crafting this dissertation. The many lessons I learned from Laura Katz while under her supervision cannot be easily tallied; above all she has been and will continue to be a great friend and inspiration. I am indebted to all the very talented colleagues and students with whom I had the pleasure of sharing the laboratory with: Laure Bellec, Mary Doherty, Micah Dunthorn, Jessica Grant, Allie Kovner, Jian Hua Lin, Truc Nguyen, Maiko Tamura and Yonas Tekle. They've all been fundamental since the training days up to the very end of my journey. I especially thank my long time office colleague Laura W. Parfrey with whom I had much fruitful discussions and collaborations.

I am also indebted to my external collaborators, who helped expand my view of the world and understanding of different types of science: Edward Mitchell and Enrique Lara at the University of Neuchatel, Switzerland; Tanja Bosak at MIT; Francis MacDonald and Andrew Knoll at Harvard. David "Paddy" Patterson at MBL deserves special mention as he has been one of the most important influences on my scientific career. There are additionally many individuals at Smith College who were great colleagues and also helped me in many ways: Sara Pruss at the Geology Department; Rob Dorit at the Biological Sciences Department; Dick Briggs and Judith Wopereis at the Microscopy Center.

A heartfelt thank you to my dissertation committee: Benjamin Normark, Michael Hood and O. Roger Anderson. These were incredibly smart and experienced people who had a great impact on the directions my research took. I would also like to thank the other great committee members I had while at the Organismic and Evolutionary Biology

program: Joe Elkinton, Paul Sievert and Duncan Irschick have been helpful with literature suggestions and the like. The work of Betsy Dumont and Beth Jakobs as directors, and of Penny Jaques as program manager were equally extraordinary and very helpful to me.

Finally I'd like to thank the friends and family who were supportive and put up with my often overly long conversations about evolutionary biology (and how everything else is meaningless). Firstly, the Americans who welcomed me into their country and lives: the Crand family (John, Sue, Gigi, Hannah and Mike) and the Reyda family (Florian, Laura, Sonja and Sam). Then, all the Brazilians who visited and supported us while in our "exile": Felipão; Lamas; Vitão; Fernando; Tim; Ana Paula and Thomás; Fernanda and Fábio; Verônica. Alas, I could not have done this without the love and support from my family and especially Paula Turrini.

ABSTRACT

PHYLOGENETICS AND PATTERNS OF MOLECULAR EVOLUTION IN AMOEBOZOA

SEPTEMBER 2011

DANIEL J. G. LAHR, B. Sc., UNIVERSITY OF SÃO PAULO, BRAZIL

M. Sc., UNIVERSITY OF SÃO PAULO, BRAZIL

Ph.D., UNIVERSITY OF MASSACHUSETTS AMHERST

Directed by: Professor Laura A. Katz

My dissertation explores several aspects of the relationship between morphological and molecular evolution in amoeboid lineages:

Chapter 1 – *General Introduction:* This chapter provides an overview of the most pressing issues in Amoebozoa phylogeny that are dealt with in the remainder of the thesis.

Chapter 2 - *Reducing the impact of PCR-mediated recombination in molecular evolution and environmental studies using a new generation high fidelity DNA polymerase:* This chapter addresses the methodological difficulty in the study of large gene families, the generation of artifactual sequences by recombination during PCR

Chapter 3 - *Evolution of the actin gene family in testate lobose amoebae (Arcellinida) is characterized by two distinct clades of paralogs and recent independent*

expansions: This chapter explores intriging patterns of evolution in the actin gene families of testate amoebae.

Chapter 4 - *Comprehensive phylogenetic reconstruction of Amoebozoa based on concatenated analysis of SSU-rDNA and actin genes*: A deep phylogenetic analyses of the Amoebozoa, enables exploration of well supported taxonomic units within the group.

Chapter 5 - *Interpreting the evolutionary history of the Tubulinea (Amoebozoa), in light of a multigene phylogeny*: This chapter explores a more restrict taxonomic unit within the Amoebozoa – the Tubulinea – based on an expanded sample of genes and taxa.

Chapter 6 - *The chastity of amoebae: re-evaluating evidence for sex in amoeboid organisms*: This chapter asks whether the null-hypothesis that amoebae are asexual is consistent with current phylogenetic evidence.

TABLE OF CONTENTS

Page

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

GENERAL INTRODUCTION

Theoretical expectations in evolutionary biology are largely guided by principles established from the study of macro-organisms (e.g. plants, animals and fungi). Molecular biology has challenged some of these expectations. Phylogenetic trees based on data for ribosomal genes revealed that life is separated into three large domains (Archaea, Bacteria and Eukarya) as opposed to the traditionally perceived five Kingdoms (Monera, Protista, Plantae, Fungi and Animalia). However, classical and synthetic evolutionary theories have established a number of fundamental concepts that remain largely unchallenged. For instance the change in classification exemplified by the switch from 5 Kingdoms to 3 Domains does not challenge the fundamental expectation that there is homology (common descent) between characters, and that analyses of these homologies enable us to reconstruct a historical diagram of relationships (phylogenetic tree). However, as we start to explore the molecular evolution of microbial eukaryotes, some fundamental aspects of general theories are inconsistent. Here, I discuss specific examples of molecular and morphological incongruence that are emerging from the study of diverse amoeboid organisms.

One fundamental prediction (or assumption) of evolutionary theory is that diversity of morphological traits should be consistent with reconstructions of relationships based on molecular data. After all, observation of descent with

modification of morphological traits is the foundation of evolutionary biology. In the early days of molecular biology, phylogenetic reconstructions did not agree with the most accepted taxonomic systems for several biological groups. The issue prompted great discussion, but two trivial reasons were responsible for the incongruence: 1) most early phylogenetic reconstructions were wrong (early molecular reconstruction methods were problematic); 2) some morphological predictions were wrong and ended up better interpreted with aid of molecular trees. Most puzzles generated by early molecular reconstructions were resolved and no major dents were put into the canon of evolutionary theory.

The Amoebozoa reveal considerable discord between molecular and morphological data. A growing body of evidence shows that at multiple hierarchical levels, morphology does not accurately predict phylogeny for amoeboid organisms. There are multiple possible causes, which may be caused by systematic methodological errors or may be truly biological. For instance, incomplete sampling of taxa can produce an incorrect estimate of relationships; in this case, a systematic error in the methodological approach could be creating the pattern of discordance. On the other hand, there may be widespread morphological convergence happening at much deeper levels than is known for other organisms.

Recognized as one of the major eukaryotic “supergroups”, the Amoebozoa are a collection of amoeboid organisms initially detected based on analyses of the small sub-unit ribosomal DNA (SSU-rDNA) gene. Not all amoebae are Amoebozoa though. Included in the Amoebozoa are only those organisms that generally produce lobose pseudopods, which are rounded, semi-cylindrical protrusions of cytoplasm with both a

granular internal compartiment (endoplasm) and a hyaline, external part (ectoplasm). There are however other amoeboid organisms scattered across the tree of eukaryotes. A large number of them are in the "supergroup" Rhizaria. These generally produce pseudopods of the filopodial type. Filopods are thin pseudopods composed solely of ectoplasmic material, with no granular inclusions. At this deep level of divergence, there aren't many morphological or intracellular characters to be compared. However, we still expect that in less-inclusive levels the morphology will be informative for phylogeny as is the case in other groups.

Within the Amoebozoa, a number of morphologically well-defined lineages are consistently recovered in molecular reconstructions. Examples of these are the Amoebidae (which includes the text-book favorite *Amoeba proteus*), the Dictyostellidae (including the cellular slime mold *Dictyostelium discoideum*, a favorite example for multicellularity apart from plants, animals and fungi), the Centramoebidae (containing the opportunistic pathogen *Acanthamoeba* spp.) and about 15 other groups. Interestingly the relationships between these well-defined groups are largely unknown: there are proposed phylogenies and a working hypothesis of classification, but uncertainty about relationships in the Amoebozoa is higher than in other groups. Although this lack of resolution may be attributed to insufficient data, it is worth noting that similar amounts of data were sufficient to resolve relationships in other high level groups. For example, single gene analyses of SSU-rDNA are able to separate the Alveolata into their three main components: dinoflagelates, ciliates and apicomplexans.

The inconsistency between morphology and molecules starts to appear when trying to group the well-defined lower level relationships within the higher group

Amoebozoa. For instance, the Gracilipoda comprises an interesting group of three distinct genera that share a characteristic that is fundamentally distinct from any other Amoebozoa: they produce filose pseudopodia. The general body shape (a flattened network) and locomotion characteristics are much more similar to organisms in the Rhizaria than to organisms in the Amoebozoa. Another interesting example is the genus *Phalansterium*. These are spherical organisms with a protruding flagellum, with no apparent amoeboid movement that although falls firmly within Amoebozoa, does not at the moment have a well-supported relationship with any other group. Both examples may be interpreted as cases of convergence, even at this deep level.

The most compelling cases of morphological/molecular incongruency are at the level of genera and species. Across the Amoebozoa, a growing number of genera and species (these, the building blocks of systematic knowledge) turn out to be non-monophyletic upon deeper inspection. *Rhizamoeba*, *Hartmannella*, *Amoeba*, *Nebela*, *Heleopera*, *Difflugia* are all examples of easily identifiable amoebozoan taxa that are non-monophyletic in molecular trees. In fact, around 60% of genera for which more than a single species had molecular sequences sampled are non-monophyletic. Were morphologists simply wrong more than half the time? Even more puzzling, some species display a high level of cryptic genetic diversification, as is the case of *Hyalosphenia papilio* and *Vannella symplex*. These results point to a fairly problematic conclusion: morphology does not indicate species cohesion neither can be immediately used to reconstruct relationships.

Of fundamental concern is the incongruency between morphology and molecules in the Arcellinida. The arcellinids are amoebae with tests (shells). Tests are an important

feature as they not only provide morphological identity, but they also provide a fossilization potential. Fossil testate amoebae enable the calibration of the ancient Ameobozoa that are otherwise completely devoid of a fossil record. This fundamentally useful feature of testate amoebae is under significant threat if we cannot confidently attribute a specific morphology to a monophyletic group of organisms. One example is the Lesquereusiidae. This Family comprises all members of Arcellinida that can biomineralize silica particles to be added in their shells. Three genera included here were: *Lesquereusia, Netzelia* and *Quadrulella* but current molecular evidence places *Quadrulella* within the Nebelidae with high support, and statistical tests reject the possibility of monophyly of Lesquereusiidae. Hence the biomineralization of silica particles, a character that can be identified in fossil forms, has emerged multiple times in the Arcellinida, making calibration of the fossil record more difficult.

The emerging pattern is only visible now because of increasing interest in microbial eukaryotes. As most of the diversity of microbial eukaryotes is neither harmful nor beneficial to humans, they have often been neglected, and thus theories and interpretations are spotty. Additionally, amoebae are intrinsically difficult to study; microscopes and culturing are a necessity, which require much training. Only recently molecular biology techniques enable objective and repeatable exploration of deep relationships within and among amoeboid organisms. The intriguing inconsistency between morphological and molecular data presented here should spark a renewed interest in amoebae as subjects of evolutionary studies. If these observations are even partially correct, they have the potential of deeply modifying our understanding of

evolutionary processes. If they are incorrect, they will likely lead the way to the improvement of analytical methods.

CHAPTER 2

REDUCING THE IMPACT OF PCR-MEDIATED RECOMBINATION IN MOLECULAR EVOLUTION AND ENVIRONMENTAL STUDIES USING A NEW GENERATION HIGH FIDELITY DNA POLYMERASE

2.1 Abstract

PCR-mediated recombination can greatly impact estimates of diversity, both in environmental studies and in analyses of gene family evolution. Here we measure chimera (PCR-mediated recombinant) formation by analyzing a mixture of eight partial actin sequences isolated from the amoeba *Arcella hemisphaerica* amplified under a variety of conditions that mimic standard laboratory situations. We further compare a new generation proofreading processivity-enhanced polymerase to both a standard proofreading enzyme and previously published results. Proofreading polymerases are preferred over other polymerases in instances where evolutionary inferences must be made. Our analyses reveal that reducing the initial template concentration is as critical as reducing the number of cycles in order to decrease chimera formation and improve accuracy. Furthermore, assessing the efficiency of recovery of original haplotypes demonstrates that multiple PCR reactions are required to capture the actual genetic diversity of a sample. Finally, the experiments confirm that processivity-enhanced polymerases enable substantial decrease of PCR-mediated recombination through reducing starting template concentration, without compromising the robustness of PCR reactions.

2.2 Introduction

Polymerase chain reaction (PCR)-based methods are the norm in molecular evolution studies of non-model taxa and in explorations of environmental DNAs. For example, degenerate PCR is often used in systematic studies where numerous diverse taxa are to be sampled (Baldauf 2003; Bapteste et al. 2002; Grant et al. 2009; Nikolaev et al. 2004; Tekle et al. 2008; Yoon et al. 2008). Despite advances and dropping costs in mega- and meta-genomic sequencing techniques (Allen and Banfield 2005; Holt and Jones 2008; Keller and Zengler 2004), PCR methods remain key in hypothesis-driven environmental studies (Barns et al. 1996; Costas et al. 2007; Dawson and Pace 2002; Doherty et al. 2007; Edgcomb et al. 2002). PCR is used in such studies mainly because of its reproducibility that enables the targeting of specific genes of interest from diverse taxa.

One worrisome aspect of PCR-based studies is the phenomenon of PCR-mediated recombination, or chimera formation (Brakenhoff et al. 1991; Meyerhans et al. 1990). Chimeras are formed when incompletely extended DNA fragments anneal to closely related sequences generating recombinants between starting templates (Bradley and Hillis 1997; Judo et al. 1998; Kanagawa 2003). It can be difficult to differentiate original haplotypes from chimeras, leading to overestimation of biological diversity in environmental studies (Berney et al. 2004; Hugenholtz and Huber 2003; von Wintzingerode et al. 1997). Interpretations about the fate of genes in molecular evolution studies can also be compromised by the presence of chimeras, as has been shown in tests of positive selection in the Major Histocompatibility Complex (*MHC*) in sticklebacks (Lenz and Becker 2008).

Most experimental work on PCR-mediated recombination has used traditional enzymes such as *Taq* polymerase to determine rates of chimera formation under conditions normally used in studies of environmental microbial samples (*e.g.* bacterial and archaeal 16s SSU-rDNA surveys (Acinas et al. 2005; Liesack et al. 1991; Qiu et al. 2001; Speksnijder et al. 2001; Suzuki et al. 1998; Suzuki and Giovannoni 1996; von Wintzingerode et al. 1997; Wang and Wang 1997; Yu et al. 2006). These studies determined that chimera formation can be reduced for most DNA polymerases when the cycle number is lowered and extension time increased (Judo et al. 1998; Kanagawa 2003; Kurata et al. 2004; Yu et al. 2006; Zaphiropoulos 1998). The recommendation is that the lowest number of cycles be determined experimentally, which should be around 20 cycles or fewer. These suggestions can be easily followed in experiments with high quality DNA from organisms of known genome complexity (Lenz and Becker 2008; Qiu et al. 2001). However, when dealing with DNA extracted from organisms that may have highly complex genomes or preparations with chemical compounds that are not completely removed (e.g. environmental DNAs from sediments), it is more difficult to optimize PCR for downstream applications such as cloning and sequencing (Acinas et al. 2005; Speksnijder et al. 2001; von Wintzingerode et al. 1997; Wang and Wang 1997). Additionally, in molecular evolution studies where single nucleotide polymorphisms are important for inferring evolutionary processes (*e.g.* population studies, analyses of rare biosphere) the high error rate of *Taq* polymerase is not desirable. To address such difficulties, a new generation of DNA polymerases has emerged that combine proofreading capabilities with enhanced DNA binding motifs (Wang et al. 2004), including Phusion (Finnzymes, Finland); PfuUltra (Stratagene, CA) and Pfx50

(Invitrogen, CA). These enzymes have not yet been analyzed for dynamics of chimera formation.

Our goal is to understand the formation of PCR-mediated recombinants when many closely related sequences are present in the same reaction, and when a high number of cycles is required to generate robust products. Low primer to target amplicon ratio is assumed to be the main reason for mismatch pairing in later cycles, which leads to chimera formation (Acinas et al. 2005; Brakenhoff et al. 1991; Judo et al. 1998; Meyerhans et al. 1990), thus we also surveyed different initial DNA concentrations. Varying DNA concentrations is also relevant because in genomic DNA extractions, the absolute number of genome copies varies according to genome size and the subsequent high copy number of members of large gene families could lead to increased PCR-recombination (Lenz and Becker 2008).

Here we analyze the formation of chimeras from a set of eight paralogous protein-coding genes by comparing the following experimental conditions: 1) a processivity-enhanced, proofreading polymerase to a traditional proofreading polymerase; 2) high cycle number to standard cycle number and 3) a range of initial template concentrations. These sets of conditions are relevant to numerous research areas as parameters fall within recommendations and are likely to be used in standard laboratory practice.

2.3 Methods

2.3.1 Origin of templates

We chose to investigate a set of eight paralogous haplotypes of the actin gene extracted from the testate amoeba *Arcella hemisphaerica*. The eight haplotypes differ

from 2.4-20.5% in nucleotide sequence. Actin clones were obtained from previous work in *A. hemisphaerica* as described in Tekle et. al (Tekle et al. 2008), except that resulting clones were purified using the PureLink kit (Invitrogen, CA, USA). To generate templates for the experiment (Figure 1.1), we eliminated the vector by diluting each purification to 25 ng/μl and amplified them separately using *Arcella*-specific degenerate primers designed from an alignment with over 30 actin paralogs from this taxon: AhemAct-F (5' GARGARCAYCCYGTYTTGTTGAC 3') and AhemAct-R (5' TAYTTYCTYTCDGGRGGAGCAAT 3'). Phusion Hot Start polymerase (New England Biolabs, USA, Cat. No. F540) was used in the following conditions: 98°C denaturing for 15s, 56°C annealing for 15s, 72°C extension for 45s, 35 cycles. These primers yield an actin fragment that is 670 base pairs long. We performed these experiments using appropriate negative controls and the amplified products were sequenced to check for quality. Each amplified product was then purified using Microclean (The Gel Company, CA, USA). Finally, all haplotypes were individually diluted to 1 ng/μl and mixed (Figure 1.1).

2.3.2 Conditions

The conditions surveyed varied across a gradient of template concentrations at a high cycle number and a low cycle number (Table 1.1). We performed amplifications using both $Vent_R$ polymerase (New England Biolabs, MA, USA, Cat. No. M0254) and Phusion Hot Start polymerase (New England Biolabs, MA, USA, Cat. No. F540). To assess varying template concentrations, we started with the mixture of eight haplotypes, each at 1ng/μl measured in a NanoDrop (NanoDrop Products, Wilminton DE, USA),

which is equivalent to $1.4x10^9$ molecules/μl of the amplified 670 base pair (bp) fragment. We then proceeded to dilute this mixture in five consecutive 1:10 solutions of 10mM Tris-HCl and the final dilution 1:2. Hence we obtained the following concentrations: 1 ng/μl; 10^{-1}ng/μl; 10^{-2} ng/μl; 10^{-3}ng/μl; 10^{-4}ng/μl; 10^{-5}ng/μl and $5x10^{-6}$ng/μl. These concentrations correspond respectively to the following amounts of template molecules/μl: $1.4x10^9$; $1.4x10^7$; $1.4x10^5$; $1.4x10^3$ and $6.8x10^2$ (Table 1.1).

We chose to use 30 cycles as a reasonable number for standard PCR in molecular evolution and environmental studies. For a high cycle condition, 50 cycles were used as an upper extreme boundary where effects of high cycling would certainly be obtained (see Lenz et. al (2008) for a brief discussion on commonly used cycle numbers). For Phusion polymerase, every template concentration yielded enough products in the 50 cycle condition, but the 30 cycle condition did not present sufficient yield for downstream analysis in the lowest template concentration ($5x10^{-6}$ng/μl or $6.8x10^2$ molecules/μl). We decided to analyze the data from the 50 cycle condition to the lowest dilution we could get, thus for both cycle numbers we analyzed the lowest possible dilution. For $Vent_R$ polymerase, we only obtained satisfactory yields for downstream processing in the two highest concentrations analyzed (1 ng/μl and 10^{-1}ng/μl; Table 1.1).

Except for number of cycles and initial template concentrations, all reactions were performed using the same cycling parameters. The recommended extension times (30s/kilobase for Phusion, 60s/kilobase for $Vent_R$) were increased three-fold, taking into account previous claims that longer extension times decreases chimera formation (Judo et al. 1998; Kanagawa 2003; Kurata et al. 2004; Meyerhans et al. 1990; Wang and Wang 1997). For experiments on Phusion polymerase, the concentrations for amplification

mixtures followed manufacturer's protocol (1X HF buffer, 1.5mM of $MgCl_2$, 0.2 mM of each dNTP, 0.5 μM of each primer and 0.01 U/μl of polymerase), and the cycling conditions were 95°C for 3 min, followed by 30 or 50 cycles at 98°C for 15s, 56°C for 15s, 72°C for 90s and then a final extension at 72°C for 5min. For $Vent_R$ polymerase the concentration for amplification mixtures were per manufacturer's recommendations (1X Thermopol buffer, 1.5 mM of Mg_2SO_4, 0.2 mM of each dNTP, 0.5 μM of each primer and 0.005 U/μl of polymerase). Reaction mixtures were incubated at 95°C for 3min, then 30 or 50 cycles at 95°C for 15s, 56°C for 15s, 72°C for 3 min and a final extension of 72°C for 5 min. In both Phusion and $Vent_R$ reaction mixtures, the final concentrations of DNA were $1.1x10^{-4}$ μM, $1.1x10^{-6}$ μM, $1.1x10^{-8}$ μM, $1.1x10^{-10}$ μM and $5.7x10^{-11}$ μM corresponding to the experimental dilutions, respectively (in molecules/μl): $1.4x10^{9}$; $1.4x10^{7}$; $1.4x10^{5}$; $1.4x10^{3}$ and $6.8x10^{2}$.

2.3.3 Cloning

Each amplification reaction was run into 1% Seakem GTG agarose gel (Cambrex Bio Science, ME, USA) made with modified TA.E buffer (40 mM Tris-acetate, pH8.0, 0.1 mM Na_2EDTA). The 670 bp band was visualized by staining with SYBR Safe (Invitrogen, CA, USA) at dilution $1:10^4$. We then excised the band from the gel and isolated DNA from agarose with the Millipore UltraFREE DA (Millipore Corp., MA, USA). The obtained product was further purified using Microclean (The Gel Company, CA, USA). The purified products were then ligated using Zero Blunt TOPO cloning kit (Invitrogen, CA, USA) and transformed into One Shot Competent Cells (Invitrogen, CA, USA) per manufacturers instructions. Cloned cells were plated in Luria-

Bertani/Kanamycin plates and colonies were screened for inserts by direct PCR using Ampli*Taq* Gold Polymerase (Invitrogen, CA, USA). Positive colonies were then purified in a 96-well format using PureLink kit (Invitrogen, CA, USA), per manufacturers instructions. Sequencing reactions were performed in a 96-well format in an ABI 3100 automated sequencer at the PennState University Nucleic Acid Facility (University Park, PA, USA). We aimed to sequence 24 clones of each amplification/cloning event for comparative reasons, but the rate of sequencing failure varied across conditions (see Table 1.1).

2.3.4 Replicates and controls

In order to avoid stochastic effects, we replicated the experiment by: 1) diluting and mixing the original templates two times independently and 2) repeating a subset of conditions for each experiment (Table 1.1). For the Phusion experiments, templates were made two times independently, starting from the first amplification of haplotypes. Each condition was amplified up to two times for each independent making of templates. For $Vent_R$ experiments, templates were made once and amplifications replicated twice. Negative controls were used throughout the experiment, and no contamination was detected. We also used a positive control for contamination by randomly choosing one of the original haplotypes to run through the whole protocol (dilution, amplification, gel isolation, cloning and sequencing) side-by-side with the experimental mixed haplotypes. We sequenced at least 4 positive clones for each experiment and no cross-contamination was detected, as all sequenced clones were identical to the original haplotype. This

positive control indicates that we did not have cross-tube contamination, which could cause concentration errors or template bias.

2.3.5 Determining PCR recombination events

We analyzed sequences for each experiment individually. The sequences were initially scanned for quality using SeqMan (DNASTAR 1994); poor quality sequences due to ambiguity at sites were discarded, less than 1% of the overall sample. This step presumably excludes heteroduplexes as well (Thompson et al. 2002). All sequences from each PCR experiment were compiled and aligned manually using MacClade (Maddison and Maddison 2005). Similarity trees were generated by NJ algorithm using PAUP* (Swofford 2000) to tally cloned sequences as original or chimeric haplotype (Supplementary Material, available at http://www.biotechniques.com). Polymorphisms were confirmed by eye for sequences that were not 100% identical to the original haplotypes. Breakpoints were determined manually for chimeric haplotype (Supplementary Material, available at http://www.biotechniques.com), by aligning each chimera against all eight initial haplotypes in Megalign (DNASTAR 1994). A similarity tree for all encountered chimeric haplotypes was used to search for the exact same chimera in independent amplification events. Statistical analyses regarding distribution of breakpoints and One Way ANOVAs were performed in STATA/SE 9.1(StataCorp 2005).

2.3.6 Software recognition of the chimeras obtained

We used the online software Bellerophon (Huber et al. 2004) to determine efficiency of automated chimera recognition (Supplementary Material, available at

http://www.biotechniques.com). We chose to use this particular software as an exemplary system because: 1) it is widely used; 2) it is explicitly designed to recognize PCR-mediated recombination, and not historical recombination signal; 3) it does not rely on a database of confirmed sequences, which is useful to only a fraction of studies.

2.4 Results

PCR-mediated recombination between eight initial haplotypes (2.4-20.8% divergence) was investigated across a gradient of template concentrations in two different cycling conditions (Table 1.1). Extensive experiments using the Phusion polymerase were compared to a more limited dataset using $Vent_R$ polymerase (Table 1.2).

2.4.1 PCR recombination events across different treatments

We determined the number of chimeras recovered in each individual amplification/cloning event Figure 1.2 (see Supplementary Material, available at http://www.biotechniques.com, for a complete tally). For Phusion polymerase, 50 cycles amplification reaction yields significantly more recombinants then the 30 cycles amplification (One Way ANOVA: df=23, F=9.03, p=0.006), with an average of at least 65% recombinant sequences for all initial template concentrations (Table 2.2). Reduction of concentration does not significantly reduce chimera formation for the 50 cycles condition (One Way ANOVA: df=11, F=2.20, p=0.16). Chimera formation remained high (~60%) at 30 cycles with higher initial template concentration not significantly different from the same concentrations at 50 cycles (One Way ANOVAs -- concentration $1.4x10^9$: df=5, F=4.17, p=0.11; concentration $1.4x10^7$: df=5, F=3.76, p=0.12), but

decreased to 5% at a starting concentration of $1.4x10^5$ molecules/μl and no chimeras were recovered in the lowest concentration ($1.4x10^3$ molecules/μl), both significant decreases (One Way ANOVA: df=11, F=36.33, p=0.00001). For $Vent_R$ polymerase, we had fewer conditions to compare as the PCRs failed at the lower DNA concentrations. At 50 cycles $Vent_R$ yielded an average of 40-44% chimeras at all observed concentrations, and at 30 cycles 45% and 36% for higher and lower concentrations respectively (Table 1.2).

2.4.2 Recovery of original haplotypes per treatment

The number of original haplotypes recovered out of the eight initially mixed varied for different treatments (Figure 1.2), and do not differ significantly with varying number of cycles (One Way ANOVA: df=23, F=2.52, p=0.12) or starting template concentration (One Way ANOVAs – 30 cycles: df=11, F=1.85, p=0.21; 50 cycles: df=11, F=2.91, p=0.1). Nevertheless, a few trends emerge. For Phusion, at 50 cycles, all template concentrations on average recover an assortment of four haplotypes out of the eight originals. At 30 cycles, the intermediate concentrations ($1.4x10^7$ and $1.4x10^5$, molecules/μl) recovered seven haplotypes on average, with individual PCR reactions actually being able to recover all eight (Figure 1.2); the higher and lower template concentrations recovered four haplotypes on average. $Vent_R$ polymerase recovered an average of five original haplotypes for all conditions. Some sequences were more prone to be recovered: Original Haplotype 2 and Original Haplotype 5 are recovered in almost all experiments; and Original Haplotype 3 is recovered only in half the experiments (Supplementary Material, available at http://www.biotechniques.com).

2.4.3 Number of breakpoints in chimeric sequences

We determined breakpoints and participant original haplotypes for each chimeric haplotype. Both the distribution of numbers of breakpoints per sequence and the distribution of breakpoints along the sequence suggests that under PCR conditions the recombination events are random. Chimeras varied from having a single breakpoint with two clear parental sequences to having eight breakpoint and six parental sequences alternating in participation (see Supplementary Material, available at http://www.biotechniques.com, for a complete tally). The majority of chimeras (65%) have more than one breakpoint and in most cases there are more than two parental sequences for each chimera (Figure 1.3). There is no clear pattern between number of breakpoints and template concentration or number of cycles: the distribution of sequences with breakpoints follows a Poisson distribution when taken together (Poisson regression likelihood ratio chi-squared=7.83, p=0.02, df=2; Pearson goodness-of-fit chi-squared=110.67, p=0.99, df=152), and follow that distribution when partitioned by concentration (p=0.01) or cycling number (p=0.04) (Figure 1.3). Additionally, we were unable to determine a correlation between sequence features (local similarity, conservation) and susceptibility to be a breakpoint. The distributions of breakpoints along the sequence are not significantly different from the expected in a normal distribution (Shapiro-Wilk's W=0.95, p=0.04; Supplementary Material, available at http://www.biotechniques.com).

2.4.4 Haplotype participation in chimeric sequences

The frequency that a specific original haplotype was involved in chimeric events corresponds with the frequency that haplotype was recovered overall in PCRs (Supplementary Material, available at http://www.biotechniques.com). Original Haplotype 2 was recovered most times across experiments (51 clones overall), and it also was involved as a part of a chimera in 81 cases. Original Haplotype 3 was the least recovered haplotype across experiments (16 clones overall) and was also the least likely haplotype to participate in chimeras (23 counts). Such pattern in the composition of chimeras is further evidence that recombination events are mostly random. The more readily available sequences are more likely to participate in recombination events, without any bias toward a particular haplotype or group of haplotypes.

2.5 Discussion

Chimeras are more likely to be observed when both high cycle condition and large initial concentration of templates are used (Table 1.2). In contrast, no recombinants were observed in the low cycle/low concentration conditions. Although our analyses corroborate the inference that high cycle numbers induce chimera formation in PCR (Bradley and Hillis 1997; Judo et al. 1998; Kanagawa 2003; Kurata et al. 2004; Qiu et al. 2001; Wang and Wang 1997), they also highlight the importance of initial template concentration in chimera formation (Figure 1.2). The effect of template concentration revealed here is due to the wider range (seven orders of magnitude) of concentrations analyzed compared to previous research (Qiu et al. 2001). Such a range is enabled by the

ability of Phusion, a processivity-enhanced polymerase, to amplify very low concentrations of template (Table 1.1).

Rates of recombination obtained for Phusion are highly variable across conditions. Previous research has proposed that enzymes with higher processivity yield more chimeric sequences (Qiu et al. 2001). This is confirmed at high cycle/high template concentration conditions, with an average yield of 71% chimeras (Figure 1.2). However, the surprising result is that the same enzyme yields absolutely no chimeras when the initial template concentration is low enough ($1.4x10^3$ starting molecules) and when the cycle number is reduced to 30 cycles (Figure 1.2). The processivity-enhanced enzyme makes up for its high rate of chimera formation by being able to amplify initial concentrations that are four to five orders of magnitude lower than what the strict proofreading polymerase can amplify (Table 1.1), effectively reducing chimera formation to zero (Table 1.2). Non-proofreading polymerases (i. e. *Taq*) might also benefit from less concentrated initial templates but they are less desirable for molecular evolution studies. Furthermore, we attribute Phusion's ability to amplify low template concentrations to the enhanced processivity. Ordinary *Taqs* might not be able to amplify concentrations that are low enough to reduce artifact formation.

The percentages of chimeras formed per reaction are higher for certain conditions in the present survey than have been reported in much of the literature, which we believe is due to the greater complexity of our starting templates (*i.e.* 8 actin haplotypes). The rates of recombination for a proofreading enzyme ($Vent_R$) under standard conditions (30 cycles) averages around 40% recombinants in the present study, while the highest available literature reports are 32% for *Taq* polymerase after 30 cycles on 7 distinct initial

haplotypes (Wang and Wang 1997) 16% for the proofreading Expand H-F system (Boehringer, Mannheim, Germany) after 25 cycles using 8 initial haplotypes and 31% recombinants using *Taq* polymerase across multiple loci in polyploid cotton (Speksnijder et al. 2001). Since the reported chimera formation rates in other available literature ranges from 1-5% across a variety of enzymes (Acinas et al. 2005; Bradley and Hillis 1997; Brakenhoff et al. 1991; Judo et al. 1998; Kanagawa 2003; Meyerhans et al. 1990; Qiu et al. 2001), we attribute the higher rates in our experiment as well as in two others to the higher number of initial haplotypes; most studies used 2-4 initial haplotypes to test chimera formation. In contrast, up to 35% recombinants were reported for only two MHC loci (Lenz and Becker 2008), which may indicate a possible influence of the template itself. Further, we demonstrate that there is a rapid increase in chimera formation as diversity in the original sample increases. This reinforces the idea that molecular environmental studies might be plagued with a slew of artificial sequences (Hugenholtz and Huber 2003). Moreover, initial template concentrations (i. e. abundances) in environmental samples will most likely be unequally distributed, which might influence the formation of artifacts. However, differential abundance of templates probably has a larger impact on the detection of true diversity and our analyses indicate that multiple (>2) PCRs will be required to capture the true diversity of a sample, even when abundances of templates are equivalent.

We find that the majority of chimeras contain more than one breakpoint, indicating that more than two parental sequences can be involved in PCR-mediated recombination. This high rate of cross-over is independent of cycle number or initial template concentration (Figure 1.3). This observation will create problems for chimera

detecting software that base their search criteria in finding one breakpoint per sequence. For example, using the online software Bellerophon (Huber et al. 2004) to detect the chimeras in the present dataset, on average only 65±18% of chimeras are detected and even more worrisome there is a false-positive rate of 40±31% (see Supplementary Material, available at http://www.biotechniques.com, for details).

Capturing the full diversity within a sample requires a combination of multiple PCR reactions that have been performed under chimera-reducing conditions. On average, PCRs at high cycle numbers are unable to recover all diversity (average 4±1 out of 8 starting haplotypes, Figure 1.2), even if all three replicates are combined, and have the added bias of generating false haplotypes. While low cycle number improves recovery (7±1 out of 8 starting haplotypes, Figure 1.2), it is likely that a single PCR experiment will not capture all the diversity. For example in chimera-reducing conditions with low cycle number and lowest initial template concentration possible, individual PCRs detect an average of four out of the eight haplotypes, but performing three replicates will certainly describe all diversity in this eight haplotype system (Supplementary Material, available at http://www.biotechniques.com). Hence we reiterate the necessity of replicating PCRs to assess biodiversity (Acinas et al. 2005; Kanagawa 2003; Qiu et al. 2001), and we add the recommendation that these replicate PCRs be run at minimal DNA concentrations and cycle numbers, which need to be established on a sample-by-sample basis in environmental studies.

Table 2.1: Conditions examined for the formation of PCR-mediated recombinants (chimeras).

Template Concentration	**$1.4x10^9$**	**$1.4x10^7$**	**$1.4x10^5$**	**$1.4x10^3$**	**$6.8x10^2$**
Phusion – template mix 1*					
50 cycles (2x)	12, 10	23, 25	23, 22	NA	24, 24
30 cycles (1x)	27	29	24	15	failed
Phusion – template mix 2*					
50 cycles (1x)	21	17	22	NA	15
30 cycles (2x)	17, 19	20, 22	22, 19	3, 11	failed
$Vent_R$ **					
50 cycles (2x)	7, 8	8, 3	failed	failed	failed
30 cycles (2x)	8, 10	10, 12	failed	failed	failed

We have determined eight distinct experimental conditions to study, comprising of four initial template concentrations, each amplified using two different numbers of cycles.
* For Phusion polymerase, each treatment was repeated three times, two of them with the same template mixing event, and one independent.
** For $Vent_R$ Polymerase, we used one mixture, independently assayed two times.
NA - not available. Our objective was to survey the most dilute condition possible. Therefore, we skipped this condition for the 50 cycle experiments. The amplification works normally under these conditions.
failed - Amplification under the established parameters did not yield a suitable product for cloning.

Table 2.2: Total clones observed, average numbers of chimeras, original haplotypes recovered and percentage of chimera haplotypes formed under varying conditions.

Cycle Number	50 cycles				30 cycles			
Concentration	**$1.4x10^9$**	**$1.4x10^7$**	**$1.4x10^5$**	**$6.8x10^2$**	**$1.4x10^9$**	**$1.4x10^7$**	**$1.4x10^5$**	**$1.4x10^3$**
Phusion								
Total clones	43	65	67	63	63	71	65	29
Chimera Haplotypes	6	12	11	8	9	7	0	0
Original Haplotypes	3	5	6	4	5	7	7	4
Chimera %	65±9	70±11	64±3	65±2	67±9	50±10	5±8	0
$Vent_R$								
Total clones	15	11	-	-	18	22		
Chimera Haplotypes	2	2	-	-	3	2	-	-
Original Haplotypes	3	3	-	-	3	4	-	-
Chimera %	38±18	42±12	-	-	46±5	38±31	-	-

Numbers represent averages across all replicates for the particular experiment. Percentages are represented along with the standard deviation. For Phusion polymerase, number of clones for each condition is averaged out of 3 replicate experiments; for $Vent_R$ polymerase, out of 2 replicate experiments. A full detailed table is available as **Supplementary Material 1** online.

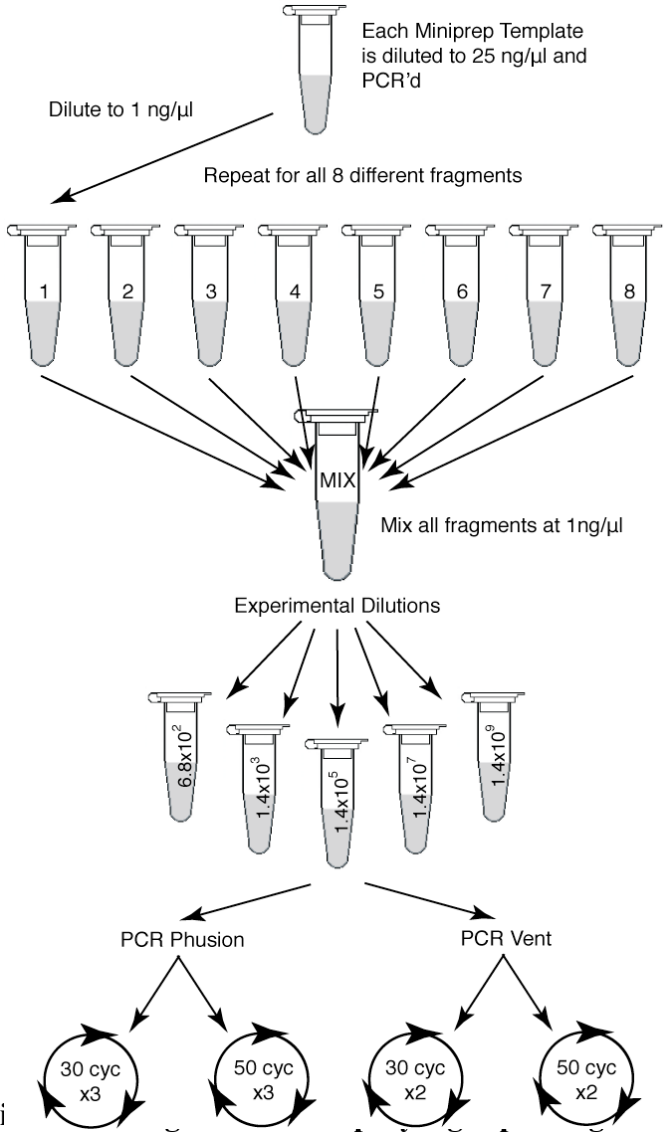

Figure 2.1: Experi ıctin haplotypes in different cycling conditions using multiple initial template concentrations.

The haplotypes, obtained from a previous study, were diluted and amplified individually. All eight were mixed to a concentration of 1 ng/µl and then diluted to five successive experimental concentrations. Each experimental concentration was amplified in triplicate using both a processivity-enhanced proofreading polymerase (Phusion) and a strict proofreading polymerase ($Vent_R$), in a low cycle number (30) and a high cycle number (50) conditions. Each amplification was subsequently cloned, sequenced and scanned for chimeras.

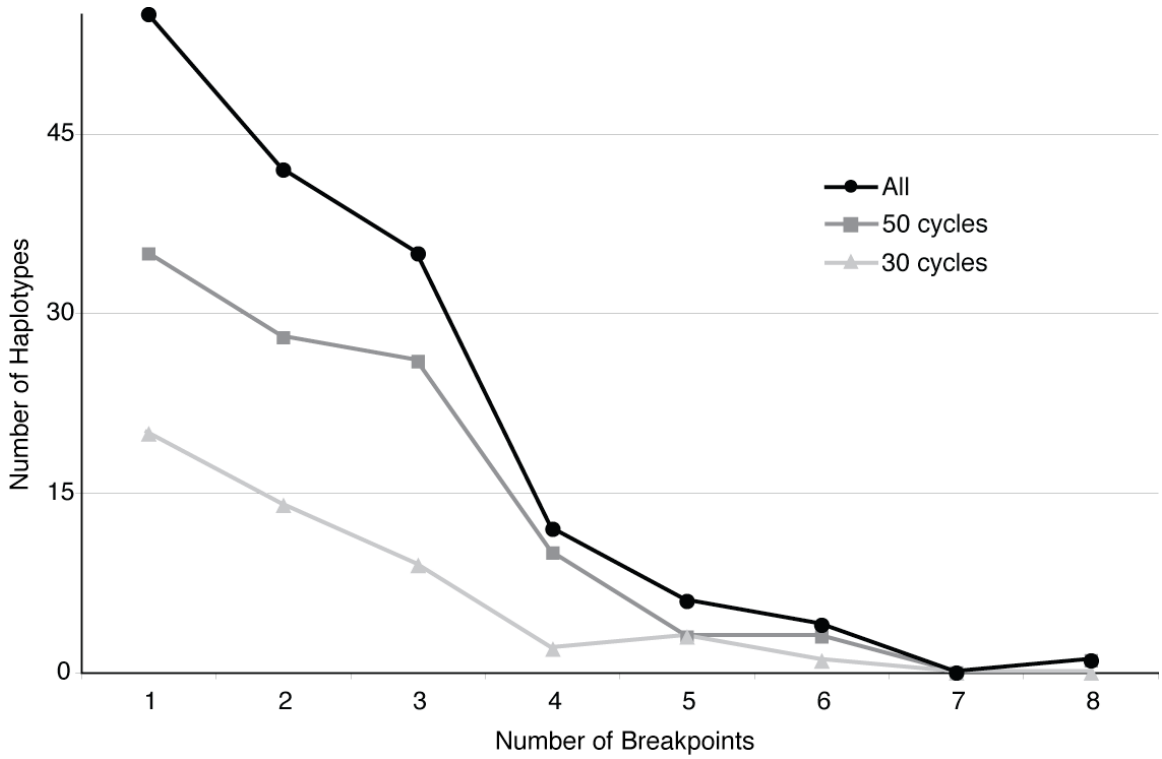

Figure 2.2: Distribution of chimeric haplotypes according to number of breakpoints. There is no significant difference in distribution between cycle numbers.

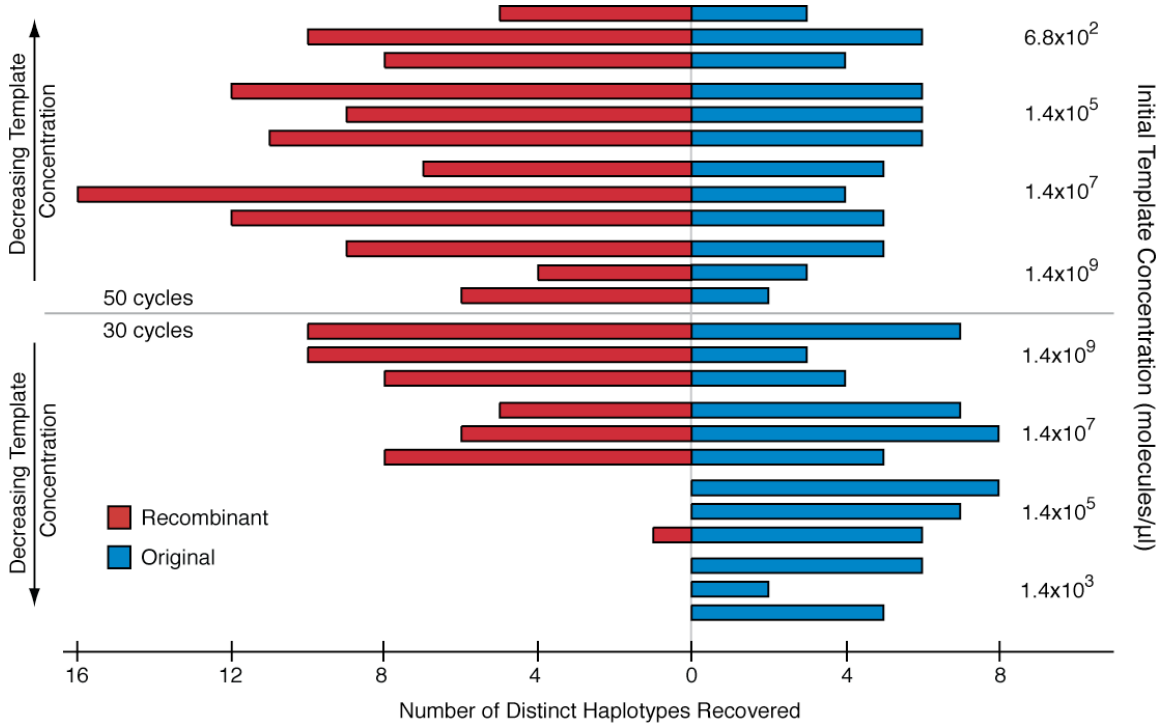

Figure 2.3: Number of chimeras and original haplotypes recovered across all concentrations and cycling conditions analyzed using a processivity-enhanced proofreading polymerase (Phusion).

The three replicates for each condition are shown separately. Numbers on x axis represent the initial concentration of molecules.

CHAPTER 3

EVOLUTION OF THE ACTIN GENE FAMILY IN TESTATE LOBOSE AMOEBAE (ARCELLINIDA) IS CHARACTERIZED BY TWO DISTINCT CLADES OF PARALOGS AND RECENT INDEPENDENT EXPANSIONS

3.1 Abstract

The evolution of actin gene families is characterized by independent expansions and contractions across the eukaryotic tree of life. Here we assess diversity of actin gene sequences within three lineages of the genus *Arcella*, a free-living testate (shelled) amoeba in the Arcellinida. We established four clonal lines of two morphospecies, *Arcella hemisphaerica* and *Arcella vulgaris*, and assessed their phylogenetic relationship within the 'Amoebozoa' using SSU-rDNA genealogy. We determined that the two lines of *A. hemisphaerica* are identical in SSU-rDNA, while the two *A. vulgaris* are independent genetic lineages. Further, we characterized multiple actin gene copieshaplotypes from all four lineages. Analyses of the resulting sequences reveal numerous diverse actin genes, which differ mostly by synonymous substitutions. We estimate that the actin gene family contains 40-50 paralogous members in each lineage. None of the three independent lineages share the same paralog with another, and divergence between actins reaches 29% in contrast to just 2% in SSU-rDNA. Analyses of effective number of codons, compositional bias, recombination signatures and genetic diversity in the context of a gene tree -genealogy indicate that there are two groups of actins evolving with distinct patterns of molecular evolution. Within these groups, there

have been multiple independent expansions of actin genes within each lineage. Together, these data suggest that the two groups are located in different regions of the *Arcella* genome. Further, we compare the *Arcella* actin gene family to the relatively well-described gene family in the slime mold *Dictyostellium discoideum* and other members of the 'Amoebozoa' clade. Overall patterns of molecular evolution are similar in *Arcella* and *Dictyostelium*. However the separation of genes in two distinct groups coupled with recent expansion is characteristic of *Arcella* and might reflect an unusual pattern of gene family evolution in the lobose testate amoebae. We provide a model to account for both the existence of two distinct groups as well as the pattern of recent independent expansion leading to a large number of actins in each lineage.

3.2 Introduction

Though actin is one of the most abundant proteins in eukaryotic cells and has been the subject of many studies, much remains to be understood about the tempo and mode of evolution of this gene family (Reisler and Egelman 2007). Actin cytoskeletal functions are well characterized (Goodson and Hawse 2002), but actins and actin related proteins (ARPs) are also implicated in nuclear processes (Chen and Shen 2007; Pederson 2008; Reisler and Egelman 2007).

There is a high level of structural and sequence conservation among actin proteins between disparate organisms, and between eukaryotic and bacterial homologues such as MreB (Erickson 2007; Goodson and Hawse 2002; Hightower and Meagher 1986; van den Ent et al. 2001). However, different lineages show different patterns of evolution in their

sets of actin paralogs across the eukaryotic tree of life (Goodson and Hawse 2002; Wade et al. 2009).

In multicellular eukaryotes, appearance of distinct actin gene duplicates and subsequent innovation is associated with tissue differentiation. In the green algal lineage (including land plants), an increase of paralogs up to 18 in the soybean (McDowell et al. 1996) is associated with an increase in morphological complexity (Bhattacharya et al. 2000). In animals, the appearance of specific muscle and cytoplasmic actin types is ancient, and subsequent duplications within each type seem independent, yielding up to 6 actin genes in vertebrates and arthropods (Hooper and Thuma 2005; Kusakabe et al. 1997).

Our understanding of actin gene family evolution in microbial eukaryotes is incomplete. Across the estimated ~70 lineages of microbial eukaryotes (Parfrey et al. 2006; Patterson 1999), the breadth of knowledge on actin diversity is largely limited to organisms with completed genomes (Reisler and Egelman 2007). In addition, diverse lineages such as dinoflagellates (Bachvaroff and Place 2008), Foraminifera (Flakowski et al. 2006), red algae (Wu et al. 2009) have been shown to contain large collections of actin gene paralogs of up to 28, 7, and 10 genes, respectively. In these three cases, paralogs are divided into two groups with different evolutionary characteristics. In all three cases there is significantly more synonymous substitutions than replacement ones, indicating purifying selection. In the Dinoflagelate *Amphidinium carterae* (Bachvaroff and Place 2008), both actin groups contain introns and are tandemly organized. There is indication that the size of introns and expression differs between the 2 groups.

The ‘Amoebozoa’ contains many familiar amoeboid organisms, including *Amoeba proteus, Dictyostelium discoideum,* and the testate lobose amoebae (Arcellinida) that are the subject of this study. Most knowledge in this group stems from studies in the model slime molds (*Dictyostelium* and *Physarum*) and some pathogenic lineages (*Entamoeba*, *Acanthamoeba*). The completed genome of *Dictyostelium discoideum* (Eichinger et al. 2005) reveals a 41-member actin gene family encompassing 17 paralogs that code for the exact same amino acid sequence (Act8-group), 7 potential pseudogenes, and 16 other paralogs ranging from canonical actins to very divergent proteins (Joseph et al. 2008). Identical paralogs for the most highly expressed type of actin (Act8-group) are spread across 4 chromosomes.

The remaining ~14 major lineages in ‘Amoebozoa’ remain largely unexplored with respect to gene family evolution (Pawlowski and Burki 2009). We have investigated the actin gene family in the lobose testate amoebae (Arcellinida). The Arcellinida are characterized by the presence of a test (shell), but despite a 750 Ma fossil record (Porter et al. 2003) and high abundance in numerous environments (Smith et al. 2008) the Arcellinida remain relatively understudied. We have isolated four clonal lines that represent two morphospecies, *Arcella hemisphaerica* and *Arcella vulgaris*. We established their relationship by analyzing small subunit ribosomal DNA (SSU-rDNA) genealogies and characterized their actin genes. Analyses of actin gene -genealogies coupled with analysis of the effective number of codons, genetic diversity indices and recombination signatures reveal that actin genes in the Arcellinida are under an intriguing mode of evolution which combines paralogy predating the divergence of these morphospecies with recent independent gene expansions.

3.3 Methods

3.3.1 Taxa studied

We isolated and cultured four lineages of *Arcella* spp. during this study. Two lineages of *A. hemisphaerica* were isolated from commercial cultures, both marketed as *A. vulgaris*. The "Blue" lineage was purchased from Connecticut Valley Biological Supply Company, Southampton, MA and is the same strain described in Tekle et al. (2008) (Table 2.1, Figure 2.1a). The "Red" lineage was purchased from Carolina Biological, Burlington, NC (Table 2.1, Figure 2.1b). The two *A. vulgaris* lineages were isolated from nature by sampling two geographically separated lake sediments in Massachusetts, USA (Table 2.1). The "SC" lineage was isolated from Lyman Lake at the Smith College Campus, Northampton, MA (Figure 2.1d) and the WP lineage was isolated at Weeks Pond in Falmouth, MA (Figure 2.1c). After starting initial mixed cultures, individual organisms were picked and washed to start clonal cultures (from a single organism) by placing cells into autoclaved pond water, and adding 0.05 volume cereal grass media (Fisher Scientific, Cat No NC9735391), as well as bacteria. Identification follows Lahr and Lopes (2009), briefly *A. hemisphaerica* are 60-80 μm wide (Figure 2.1a, b), with a markedly semi-circular lateral profile; *A. vulgaris* are 100-120 μm wide (Figure 2.1c, d), with a slightly flattened lateral profile and the presence of a rim on the border where the abapertural and apertural surfaces meet.

3.3.2 DNA extraction and amplification experiments

For DNA extraction, 100-1000 clonal individuals were harvested multiple times for each isolate. Individuals were either handpicked or harvested by spinning culture flasks into DNA extraction buffer (100mM NaCl, 10mM Tris, 25mM EDTA, 0.5% SDS, 10μg/ml Proteinase K). DNA was extracted following a standard 2 Phenol: 1 Chloroform extraction followed by cold Ethanol precipitation. Amplification of target genes was achieved by PCR in a PTC-200 Thermal Cycler (MJ Research, Waltham USA) with Phusion HotStart polymerase (Cat. Nº M0254; New England BioLabs), using concentrations of reagents per manufacturers recommendations (1X HF buffer, 1.5 mM $MgCl_2$, 0.2 mM each dNTP, and 0.01 U/μl polymerase), except for primers which were used at a final concentration 2 to 4-fold higher than recommended (1-2 μm for each) in a total reaction volume of 25 μl. Cycling conditions were: 98° for 3 min; followed by 35-60 cycles at 98° for 15 s, 56° for 15 s and 72° for 90 s; and then a final extension at 72° for 5 min. The number of cycles varied from 35-60 according to conditions inherent to different DNA extractions (see Supplementary Material available at www.mbe.oxfordjournals.org and Section 2.2.3 for details). We performed single-celled PCRs by picking an individual, washing it 3 times in autoclaved water and transferring directly into a PCR master mixture. Primers for SSU-rDNA genes are eukaryote specific from Medlin et al. (1988) and primers for actin genes are either eukaryote specific from Tekle et al. (2007) or *Arcella* specific as described in Lahr and Katz (2009) (see Supplementary Material available at www.mbe.oxfordjournals.org for details). The target amplicon for *Arcella* actins were either 669 or 795 basepairs long depending on the primer set used. Cloning experiments were performed for all amplified products using

the Invitrogen TOPO cloning kit, exactly as described in Lahr and Katz (2009). All plasmids containing inserts were purified using a PureLink kit (Invitrogen). The positive colonies were either amplified individually using a Big Dye Terminator kit (Perkin-Elmer) and run on an ABI 3100 automated sequencer at the Center for Molecular Biology (Smith College, Northampton, MA) or in a 96-well format at the Pennsylvania State University Nucleic Acid Facility (University Park, PA, USA).

3.3.3 Detection and avoidance of artifactual PCR-recombinants (chimeras)

3.3.3.1 Avoiding chimera formation

PCR amplification yields artifactual recombinants (chimeras) when multiple closely related target sequences are present in the mixture (Judo et al. 1998). We chose to rely on an empirical strategy, performing extensive PCR-artifact formation experiments in our study system to determine chimera-reducing PCR conditions (Lahr and Katz 2009). We have found that PCR recombinants result of a combination of too many amplification cycles and too much starting DNA template. Both conditions were determined on a sample-by-sample basis, given differences in genomic DNA extractions (number of individuals in the culture for example), and thus differ from experiment to experiment (see Supplementary Material available at www.mbe.oxfordjournals.org for details). Hence, most PCRs were performed across a gradient of DNA concentrations and amplification cycles, to choose the most chimera- restricting amplification possible (lowest cycle number coupled with lowest initial DNA concentration). Only those PCR products were cloned and sequenced, the others were discarded. In addition, we have performed all PCRs using three times the recommended extension time, and two to four-

fold more concentrated primers according to general guidelines for chimera reduction (Judo et al. 1998).

3.3.3.2 Determining a dataset of "real" sequences

We chose a conservative approach to distinguish "real" and "chimeric" sequences, based on observed experimental properties of chimera formation (Lahr and Katz 2009). The appearance of a specific gene sequence in two or more independent PCRs is almost certainly indicative of a "real" sequence. Hence we consider all sequences in this condition to be "real" (a total of 41 sequences for *A. hemisphaerica*, 16 for *A. vulgaris* SC and 3 for *A. vulgaris* WP). Additionally, when a particular gene sequence is found multiple times in the same PCR experiment (i.e. multiple clones) there is a reduced chance that this sequence is a recombinant, though this chance is larger than using the former criterion. This probability increases with sampling effort, and we consider that for *A. hemisphaerica*, with a total sampling effort of 440 clones and 30 PCRs, sequences that appeared or more times can be considered "real" for a total of additional sequences. For *A. vulgaris* SC, with a quarter of that effort, we also consider clones that appeared three times or more as "real," an additional four gene sequences. For *A. vulgaris* WP, with only 43 clones sequenced, we consider that clones which appeared two or more times to be "real," yielding an additional five gene sequences. Using these two criteria, we come to a final dataset that includes 45 actin genes for *A. hemisphaerica*, 20 genes for *A. vulgaris* SC and 8 genes for *A. vulgaris* WP.

3.3.4 Total number of actin genes per lineage

We estimated the total number of actin genes in each lineage using estimation tools most commonly used in ecological sampling, freely available in the package EstimateS 8.0 (Colwell 2006). We have used only actin sequences that were considered non-chimeric for these estimations. We have used two sub-partitions of the dataset to estimate diversity: 1) a conservative dataset that considers each PCR a sample, and uses only genes that were found in 2 or more independent PCRs as real paralogs; and 2) a more liberal dataset that considers a real paralog every gene that was found 3 or more times for *A. hemisphaerica* and 2 or more times for the *A. vulgaris* SC lineage, as described above. In *A. hemisphaerica*, sampling was more intense and we were able to use interpolation methods (Mao T) to estimate the total diversity. For *A. vulgaris* SC, not enough samples were taken to plateau the accumulation curve. Hence we used an extrapolation method (MMMeans) to estimate total diversity. Estimates were calculated using 1000 randomizations and sampling without replacement. The remaining lineage, *A. vulgaris* WP had too few samples to allow a consistent estimate of total diversity, but the pattern of discovery of new genes is similar to the other two lineages, and we expect results to apply to this lineage as well.

3.3.5 SSU-rDNA analysis

Sequences of representative organisms in the 'Amoebozoa' were retrieved from GenBank (Figure 2.2 lists all accession numbers). Taxon sampling reflects an effort to include representatives of all major lineages in the 'Amoebozoa' (Pawlowski 2008; Pawlowski and Burki 2009; Tekle et al. 2008). One *Arcella* sequence used in previous

reconstructions of the Arcellinida (*A. artocrea* AY848969) is likely a contaminant (Edward Mitchell, pers. com.) and was not included in this study. Alignments were constructed in SeaView (Galtier et al. 1996; Gouy et al. 2010) with alignment algorithm MAFFT (Katoh et al. 2009) using the L-INS-I setting, and adjusted manually in MacClade (Maddison and Maddison 2005). We have masked the alignment to exclude regions that had over 50% missing data, as well as ambiguously aligned sites identified by ALISCORE, using default settings (Misof and Misof 2009). The resulting alignment is 1587 sites. We generated a second, more conservative alignment, by manually excluding ambiguous regions, to a total of 1357 sites. Phylogenetic reconstruction was made using RAxML-HPC 7.0.4 (Stamatakis et al. 2008) through the online server CIPRES Portal 2.0 (Miller et al. 2009), using the GTRGAMMA model of nucleotide substitution and running 200 automatic rapid bootstraps followed by a slow search for the best-scoring ML tree. The GTRGAMMA model was selected as the most appropriate model for our dataset through a ModelTest analysis performed on HyPhy (Pond et al. 2005).

3.3.6 Actin Genealogical Analysis

Actins from other 'Amoebozoa' were obtained from GenBank accessions, curated Genome databases, and EST databases. From genome databases, we included every available actin paralog, except four variant putative actin genes (>30% divergence at nucleotide level, >30% divergence at amino acid level) in *D. discoideum* (Joseph et al. 2008). To collect actins from ESTs, we first performed a BLAST search with one described actin against the EST database of the same organism. We then constructed

contigs using SeqMan, and with a cutoff point of 1% we established putative actin paralogs. The dataset, including 73 *Arcella* actins and another 103 'Amoebozoa' actins, was aligned on SeaView (Galtier et al. 1996; Gouy et al. 2010) using the alignment algorithm MAFFT (Katoh et al. 2009) set to L-INS-I optimization. Proper codon alignment was confirmed visually. The final total alignment consists of 179 sequences, 1134 nucleotide sites or 378 amino acid sites, which we designate "Actin Alignment A". Phylogenetic analyses were performed at the nucleotide level considering all sites, as well as third positions excluded with RAxML-HPC 7.0.4 in the CIPRES Portal 2.0, using the GTRGAMMA model of nucleotide substitution and running automatic bootstrapping followed by a slow ML search. An additional analysis limiting the dataset to only the 795 homologous sites that were amplified in *Arcella*, which we designate "Actin Alignment B", was performed. The translated datasets consisting of 378 total sites and 265 homologous sites were also analysed on RAxML using the JTT model of amino acid substitution, chosen through a ProtTest analysis performed on HyPhy (Pond et al. 2005).

3.3.7 Codon usage and compositional bias

We determined the codon usage and compositional bias for collections of actin paralogs in lineages that are represented by 3 or more actin sequences, totaling 149 actin sequences that do not have internal stop codons or frame-shift deletions in "Actin Alignment B" (which comprises only the 795 basepair region that contains the largest *Arcella* amplicons) using the algorithm CodonW (Peden 1999) as implemented in the online server MOBYLE (Neron et al. 2009). We calculated the effective number of codons (ENC), total GC content and GC content at four-fold degenerate sites.

3.3.8 Genetic diversity indices

We have used tools available for nucleotide diversity calculation at the online server DPDB (Casillas et al. 2005) to calculate: average pairwise distance in the group (k=measured by averaging all pairwise uncorrect-p distances), the average number of nucleotide differences per site (π), the number of segregating sites (S) and respective variances. The alignment used for this analysis is the most restrictive in terms of length, since we only used the 669 homologous basepair region that is available for our shortest *Arcella* amplicons ("Actin Alignment C"). Each group of actin sequences representing an 'Amoebozoa' taxon was analyzed separately, and the *Arcella* dataset was divided into Groups 1 and 2 from the phylogenetic reconstruction.

3.3.9 Recombination Detection

We used the algorithm GARD (Pond et al. 2006) implemented in the online server Datamonkey (Pond and Frost 2005) to infer historical recombination between actins in each *Arcella* lineage, as well as other collections of sequences in each 'Amoebozoa' lineage separately. Analyses were run by first calculating the most appropriate model of substitution for each case, then running the genetic algorithm estimating site-to-site variation with Beta-Gamma distribution and 4 rate classes. The output is given in terms of most likely points for recombination in the dataset, and those points are further submitted to the Kishino-Hasegawa test because initial detection of points could be due to rate heterogeneity. We then used the statistically significant recombination points to divide the dataset into partitions, and independently reconstructed ML trees using

RAxML. Despite incomplete sampling in the *A. vulgaris* lineages, we can infer which genes have recombined by analyzing their relative position in different partitions. We cannot infer, in most cases, exactly which other gene they recombined with.

3.3.10 Comparison of actin gene copies across eukaryotes

We have performed a comparison of average pairwise distances between chosen eukaryotes with large numbers of actin paralogs in their genomes. We aligned GenBank deposits for actin genes in: the 'Amoebozoa' *Entamoeba histolytica* (7 actins) and *Dictyostelium discoideum* (29 actins), the dinoflagelate *Amphidinium carterae* (28 actins); the red algae *Flintiella sanguinaria* (10 actins) and *Glaucosphaera vaculolata* (7 actins); the Metazoa *Drosophila melanogaster* and *Homo sapiens* (6 actins for each); and the plants *Glycine max* (18 actins), *Zea mais* (12 actins) and *Arabdopsis thaliana* (10 actins). We calculated uncorrected pairwise distances both at the nucleotide and amino acid levels using Paup* 4.0 beta 10 (Swofford 2000), and averaged the distances for each taxon.

3.4 Results

3.4.1 SSU-rDNA analysis

Maximum likelihood analysis of the SSU-rDNA gene (Figure 2.2) including a total of 55 taxa and 1587 characters is largely concordant with other recently published reconstructions of the 'Amoebozoa' (Nikolaev et al. 2006; Parfrey et al. 2010b; Pawlowski and Burki 2009; Smirnov et al. 2005; Tekle et al. 2008). Most major lineages are recovered with high bootstrap supports (BS): the Amoebidae, Leptomyxidae,

Echinamoebidae, Archamoebae, Dictyosteliida and Acanthamoebidae are all monophyletic and receive full support (BS=100). The Flabellinea and the Hartmannellidae are recovered as paraphyletic. This analysis uses an alignment with liberal masking. When analysis are repeated with more stringent masking, yielding an alignment with 1357 sites similar to Tekle et. al (2008), the relationships remain the same and bootstrap supports increase slightly (data not shown). The Arcellinida are recovered with low support (BS<50%), and the structure of the Arcellinida sub-tree agrees with more focused reconstructions (Lara et al. 2008; Lara et al. 2007; Nikolaev et al. 2005), with the exception that the genus *Argynnia* appears related to *Arcella* in our reconstruction, as opposed to *Heleopera sphagni* in Lara et al. (2008).

The *Arcella* lineages are monophyletic within the Arcellinida (BS=100). The SSU-rDNA sequences for both lineages of *A. hemisphaerica* (Red and Blue) are identical, which may not be surprising as these were both obtained from biological supply companies. Based on this and the sharing of numerous identical actin sequences, we consider these to be two populations of the same species. However, the two *A. vulgaris* SSU-rDNAs differ by 1.7%, and do not share any actin sequences. This indicates that these two lineages are independently evolving units. Furthermore, the phylogenetic reconstruction places *A. vulgaris* WP isolate as a sister-group to *A. hemisphaerica* (BS=98), to the exclusion of the *A. vulgaris* SC isolate.

3.4.2 Actin genes identified

We identified a total of 166 distinct actin sequences from one lineage of *A. hemisphaerica* (2 populations that we interpret as the same genetical lineage) and 2

lineages of *A. vulgaris*. Given the possibility of chimera formation, and taking into account the results from a previous experimental approach to PCR in this system (Lahr and Katz 2009), we use the sole criterion of redundancy to exclude chimeric sequences. We include 45 sequences of *A. hemisphaerica* that were found in at least 2 separate PCRs or were represented by 3 or more clones in a single PCR twice, or were found at least thrice in the same experiment. For both lineages of *A. vulgaris*, we analyze 28 sequences (20 from the SC lineage and 8 from the WP lineage) that were found in at least two PCRs or were represented by two or more clones. These 73 sequences, along with representatives from other 'Amoebozoa' lineages, were used in subsequent phylogenetic, recombination, diversity and codon usage analyses.

After chimera exclusion, we have estimated the total number of actins likely to be present in each lineage, using tools for estimating total species richness commonly used by ecologists (Table 2.3). For the *A. hemisphaerica* dataset, the most appropriate statistic is the species accumulation curve calculated by the Mao T parameter. The estimate for this lineage is 45±1 actin genes. For *A. vulgaris*, only the SC lineage was sampled sufficiently enough to enable estimation of total number of actin genes. In this case, there are fewer samples than in *A. hemisphaerica*. Hence, it is more appropriate to use an extrapolation method instead of a species accumulation curve. The estimate for this lineage lies within 25-50 actin genes, consistent with the estimate for *A. hemisphaerica*. We conclude that each lineage has around 50 actin paralogs in their genome (Table 2.3).

Unique gene sequence discovery varied with intensity of sampling effort. Sampling efforts were greater in *A. hemisphaerica*, where we obtained a total of 41 non-chimeric gene sequences (Table 2.2). The two populations (Blue and Red) share 30 out

of 45 gene sequences. These 45 sequences are 1-29% divergent from each other at the nucleotide level. The large majority of polymorphisms are synonymous substitutions, thus the amino acid sequences are identical for 36 actin genes. Ahem_act19 shows major modifications (20 amino acid substitutions), Ahem_act45 shows 9 substitutions and 7 other sequences show 1-3 amino acid substitutions. Three sequences show deletions: Ahem_act33 has a frame-shifting deletion of 26 nucleotides and no amino acid modifications; Ahem_act38 shows two in-frame deletions and one frame-shifting, and 3 amino acid modifications if made to be in-frame; and Ahem_act41 has a 1-nucleotide frame-shifting deletion as well as a 6-nucleotide in-frame deletion, and 2 amino acid substitutions if made to be in-frame (Figure 2.3).

For both isolates identified as the morphospecies *A. vulgaris*, sampling was less intense than in *A. hemisphaerica*. We have obtained 20 distinct sequences for the SC isolate and 8 for the WP isolate. The levels of divergence are similar to those found in *A. hemisphaerica* with up to 27% nucleotide divergence in pairwise comparisons, and most sequences (16 in the SC lineage and 7 in the WP lineage) code for the same amino acid sequence. The most divergent sequence found is AvulSC_act09 (10 AA substitutions, including a stop codon), other 4 show 2-3 amino acid substitutions (AvulSC_act20, AvulSC_act17, AvulSC_act18 and AvulWP_act02). No nucleotide sequences are shared between isolates, but the most common coding sequence is the same across all four *Arcella* analyzed (59 out of 73 sequences).

3.4.3 Actins in the Arcellinida

Maximum likelihood analysis at the nucleotide level of the 73 actin genes described in the present study reveal that instead of yielding monophyletic clades, the genes interdigitate between the three lineages (the 2 populations of *A. hemisphaerica* that we interpret as the same genetical lineage as well as the independent lineages, *A. vulgaris* SC and *A. vulgaris* WP; Figure 2.4). These analyses were performed both including and excluding third positions. There is no correspondence between actin genealogy and morphospecies or SSU-rDNA relationship for the *Arcella*. Instead, the gene copies fall into 2 groups, with a well-supported split (BS=96): Group 1 is paraphyletic while Group 2 is monophyletic and falls within Group 1 (Figure 2.4). A phylogenetic analysis excluding third-positions to minimize the effect of saturation reveals the same pattern (Figure 2.5a). The interdigitation indicates that gene duplication predated the divergence of these strains. Yet there is also evidence of independent gene copy expansion within the 3 lineages as evidenced by the shallow clustering of paralogs from within a lineage in the actin topology.

3.4.4 Actins in the 'Amoebozoa'

The Maximum Likelihood reconstruction of the actin gene tree recovers most major lineages with moderate/high support (Figure 2.5a): the Amoebidae, Dictyosteliida, Entamoebida and Arcellinida are all monophyletic (BS>75). The reconstruction based on amino acid sequences is topologically similar regarding the placement of *Arcella* actin paralogs with the reconstruction at the nucleotide level (Figure 2.5b). Collections of paralogs within each lineage appear to have expanded independently in each species in

the ‘Amoebozoa’ for which sufficient data exist (e.g. intense PCR study, EST analyses). In almost all cases, actin paralogs of a given species group together, to the exclusion of paralogs in another species’ gene family. There are two exceptions: the *D. purpureum* set of actin genes is a monophyletic group that falls within the *D. discoideum* family; and in *Arcella* the 3 lineages interdigitate, i.e., no one isolate is monophyletic to the exclusion of others.

3.4.5 Codon usage and base composition

We compared codon usage and base composition for sets of actin genes in lineages with three or more representative sequences (Figure 2.6). In the genus *Arcella*, there are two separate groups of genes based on codon usage, and these correspond to the Groups 1 and 2 recovered in the phylogenetic reconstruction (Figure 2.4): Group 1 is moderately biased with average effective number of codons (ENC) of 34.6 and Group 2 is less constrained with average ENC of 42.3 (Figure 2.6). Group 1 also has higher GC content in four-fold degenerate sites, with an average of 65% compared to 44% in Group 2.

In contrast, sets of actin paralogs in the other ‘Amoebozoa’ lineages analyzed are restricted to a range of codon usage that is biased and low, with ENC generally less than 30 (Figure 2.6). Base compositions for the actin gene are highly variable in the ‘Amoebozoa’: *Mastigamoeba balamuthi* has an average GC composition of 65% and *Entamoeba dispar* has 35%. These organisms have biased codon usage for the actin gene, probably reflecting GC bias in the genome.

3.4.6 Recombination among actin genes

We have searched for recombination among gene sequences in each *Arcella* lineage, using the online server GARD (Table 2.4). For *A. hemisphaerica* two putative points of recombination were detected (Kishino-Hasegawa test, $p<0.01$). Analyzing trees for each partition allows determination of sequences that have likely recombined (Table 2.4). Strikingly, sequences within Group 1 only recombine among themselves; the same is true for sequences in Group 2 with one exception: EBP27 is a member of Group 2 and appears to have recombined with members of Group 1 for the first third of the sequence. For *A. vulgaris* WP, the one breakpoint inferred by GARD is not statistically significant. For *A. vulgaris* SC, no inferred recombination breakpoints were statistically significant. We are confident that these are historical and not artifactual recombination events because there are further point mutations in recombined segments. When using the same methodology for other sets of actin genes in 'Amoebozoa' lineages, only the genus *Dictyostelium* shows statistically significant recombination between paralogs (Table 2.4).

3.4.7 Genetic diversity indices

We calculated genetic diversity indices to elucidate general patterns of molecular evolution (Table 2.5). *Arcella* sequences were analyzed separately according to the two phylogenetic Groups 1 and 2 (Figure 2.4). Both *Arcella* groups show a high propensity for substitution, revealed by a high number of segregating sites per site (Group 1 $s=0.23$; Group 2 $s=0.37$). Additionally, Group 2 shows higher average nucleotide differences per site ($\pi=0.11$) than Group 1 ($\pi=0.07$). *D. discoideum* is the only other 'Amoebozoa' that

shows a comparable average pairwise distance, intermediate between Group 1 and Group 2 (π=0.08).

3.5 Discussion

The two main observations for the actin gene family in the genus *Arcella* are: 1) the gene family is organized in two distinct groups whose members share similar patterns of molecular evolution; and 2) there have been recent independent expansions within each group. To establish the pattern of molecular evolution, we first assessed the phylogenetic position of two isolates from each of the morphospecies *A. hemisphaerica* and *A. vulgaris.* Analysis of both SSU-rDNA and sharing of identical actin gene sequences indicate that the 2 *A. hemisphaerica* isolates represent the same genetical lineage, whereas the 2 *A. vulgaris* isolates are independently evolving. The genus *Arcella* forms a monophyletic clade in Maximum Likelihood genealogies of the SSU-rDNA and actin genes (Figures 2.2, 2.4). However, the 2 *A. vulgaris* morphospecies are not monophyletic, indicating that there might be more genetic divergence than seen at the phenotypical level. Though taxon sampling is limited here, these data also provide a framework for additional phylogenetic hypotheses. For example, the genus *Argynnia* previously assigned to Hyalospheniidae (Lara et al. 2008) is recovered here as a sister group to *Arcella* with moderate support (BS=67). This might be indicative of an unpredicted relationship between some testate amoebae with chitinous shells (*Arcella*) and others with biomineralized siliceous plates (*Argynnia*).

The collection of actin gene copies in the genus *Arcella* is organized in two distinct genomic groups, based on multiple lines of evidence. Phylogenetic analysis

reveals a well-supported split in a paraphyletic Group 1 and a nested monophyletic Group 2 (Figure 2.4). Group 1 has lower codon usage and higher GC content (Figure 2.6), as well as lower substitution rates (Table 2.5) than Group 2. Recombination inference indicates that members in each group recombine mostly amongst themselves (Table 2.4). Although most of the classic literature lists testate amoebae as asexual organisms, evidence for meiosis (Mignot and Raikov 1992) validates our recombination inferences. Since recombination *via* unequal crossing-over is more likely within physically close segments, this pattern is consistent with two or more groups of tandemly arrayed actin paralogs in separate parts of the genome (different chromosomes or chromosomal regions).

The two groups of actin paralogs experience strong purifying selection as they maintain a common coding sequence. The majority of *Arcella* genes (59 out of 73) encode exactly the same amino acid sequence, even though uncorrected nucleotide divergence reaches 29%. The slime mold *D. discoideum* presents a similar scenario: the core group of 17 highly expressed actins (Act8-group) have the same coding sequence and are separated in groups across four chromosomes (Joseph et al. 2008). However, there are two significant differences in *D. discoideum* compared to the pattern of actin evolution observed for *Arcella* spp. Firstly, *D. discoideum* does not show a separation in two distinct groups of actins with respect to codon usage and compositional bias (Figure 2.6). Secondly, there is no evidence of maintenance of ancient paralogous groups within different species as *D. discoideum* sequences form a single clade. In contrast, the two distinct groups of *Arcella* sp. sequences are interdigitated showing that they predate the divergence of the three lines: *A. hemisphaerica*, *A. vulgaris* SC and *A. vulgaris* WP.

The second main observation of this study is the evidence for recent and rapid duplications within each lineage, and most likely within each group. Within Group 2, there are multiple closely related copies for each of the three lineages of *Arcella* studied: *A. hemisphaerica*, *A. vulgaris SC* and *A. vulgaris* WP (Figure 2.4). Within Group 1, there are multiple closely related actin copies for the *A. hemisphaerica*, but only one copy for each of the two other lineages, which might reflect either incomplete sampling or really a large reduction in this Group of actins for the two lineages. The recency of these gene family expansions is evidenced by the presence of frame-shifting deletions in paralogs that have no additional amino acid substitutions (Figure 2.3). We consider these an indication of recent recombination, since a locus that is no longer useful should quickly accumulate mutations.

We propose a model consistent with our main observations: *Arcella* has a large collection of actin genes encoded in two distinct regions of the genome that evolve under strong purifying selection and yet are also expanding (Figure 2.7). The two groups are subject to different evolutionary pressures, as evidenced by differing levels of codon usage. These two groups may be evolving under distinct regulatory constraints, or the mutational background differs between different areas of the genome, or both. At the same time, there are multiple independent expansions at the tips of the actin tree, especially within Group 2 where we see higher rates of recombination (Figure 2.3, Table 2.4).

There are at least two mechanisms to explain the pattern of recent gene family expansion: 1) Group 2 actins are in a recombination hotspot; or 2) actins are the target of developmentally regulated genome rearrangements (DRGR). If Group 2 actins are in a

recombination hotspot, we should expect the appearance and elimination of genes at a higher than usual rate. Somatic events that alter genomes of either specific cells or at specific life-cycle stages are referred to as DRGR (Zufall et al. 2005). *Arcella* might show the kind of DRGR Zufall et al (2005) classified as genome-wide rearrangements. The actins in this scenario would be amplified many times, as in a ciliate macronucleus, and might even reside on extra-chromosomal pieces of DNA. Under this scenario the expansion pattern observed in the tips of our tree (Figure 2.4) really depicts one genomic copy and many "extra" copies. Other 'Amoebozoa' are known to have extra-chromosomal rDNA (*D. discoideum* and *E. histolytica*). Additionally, *Amoeba proteus*, which is more closely related to *Arcella*, has been shown to exhibit DNA synthesis outside of cell division (see Parfrey, Lahr and Katz 2008 for a review).

The pattern of actin evolution revealed for *Arcella* is unusual among eukaryotes (Figure 2.8). Other unicellular eukaryotes show either a limited number of actins encoding the same amino acid sequence (e.g. *Entamoeba* spp.), or a large number of actin gene copies with detectable positive selection for some members (e. g. *Amphidinium carterae*). Animals and plants have multiple actin copies, all with divergent amino acid sequences attributed to adaptive evolution concerning tissue differentiation. *Arcella* has a large collection of genes that generally maintain the same coding sequence. Yet, actins within these amoebae appear to be evolving under varying tempos of gene duplication.

Table 3.1: Lineages of *Arcella* used in this study. NA – Cultures perished before preservation.

Species	Source	Coordinates	Isolation	SSUGenBank
A. hemisphaerica Blue	CT V. Bio. Cat# L 1B	-	May 2005	EU273445
A. hemisphaerica Red	Carolina Cat# 131310	-	March 2007	-
A. vulgaris SC	Lyman Lake, MA	N42°19’ 07”; W72°38’24”	April 2007	HM853761
A. vulgaris WP	Weeks Pond, MA	N41°33’21”; W70°36’52”	Nov 2008	HM853762

The lineages A. hemisphaerica Blue and A. vulgaris SC have been deposited at the American Type Culture Collection (ATCC). The other two lineages perished before preservation.

Table 3.2: Summary of actin PCR experiments on *Arcella* lineages.

Lineage	Pop PCRs	SC PCRs	Clones	Genes	Non-chimera
A. hemisphaerica Blue	15	3	246	69	41
A. hemisphaerica Red	6	6	194	58	33
A. vulgaris SC	11	NA	132	48	20
A. vulgaris WP	8	NA	43	20	8

Pop PCR – PCR performed on DNA extracted from a clonal culture; SC PCR – PCR performed on a single-cell. A complete table is available as Table S1; Clones – number of clones sequenced; Genes – number of distinct actin gene sequences obtained; Non-chimera – number of distinct actin sequences determined to be non-chimeric.

Table 3.3: Estimates of total number of actin genes for lineages of *Arcella*.

	A. hemisphaerica		*A. vulgaris* SC	
	≥2 PCRs (n=365)	≥3 Clones (n=381)	≥2 PCRs (n=85)	≥2 Clones (n=111)
Sobs (Mao T)	41±0	45±1	16±0	27±2
MMMeans	55	61	25.1	50

The most appropriate statistic for each dataset is in bold. Only actin sequences that were deemed non-chimeric were used for estimation. Sobs – expected total number of sequenceshaplotypes sample-based statistic, using the Mao TTau calculation; MMMeans – expected total number of actin sequenceshaplotypes by functional extrapolation, based on the Michaelis-Menten richness estimator, computed analytically; n – number of actin sequenceshaplotypes used to calculate the statistic. We did not perform estimates of diversity in the lineage *A. vulgaris* WP due to low sampling effort.

Table 3.4: Number of recombinants for each actin gene family in the 'Amoebozoa'.

	# BP	#BP p<0.05*	#seqs	#Recombinants
A. hemisphaerica	2	2	15 (G1) 30 (G2)	4 4
A. vulgaris WP	1	0	1 (G1) 7 (G2)	0 2
A. vulgaris SC	0	0	1 (G1) 19 (G2)	0 0
D. discoideum	2	2	25	2
D. purpureum	1	1	11	3
M. balamuthi	0	0	12	0
E. histolytica	1	0	7	0
E. dispar	1	0	7	0
A. castellani	2	0	6	0

BP – number of inferred recombination points; G1 – *Arcella* Group 1 actins; G2 – *Arcella* Group 2 actins; * p-values are calculated by the Kishino-Hasegawa test, after breakpoint (BP) inference by GARD.

Table 3.5: Genetic diversity indices for actin gene families across the 'Amoebozoa'.

	s	π nucleotide (SD)	N	Source
Arcella Group 2	0.37	0.11 (0.015)	52	PCR
D. discoideum	0.27	0.08 (0.011)	22	Genome
Arcella Group 1	0.23	0.07 (0.011)	17	PCR
D. purpureum	0.05	0.02 (0.004)	11	EST
A. castellani	0.04	0.02 (0.005)	6	EST
M. balamuthi	0.04	0.01 (0.004)	13	EST
E. dispar	0.01	0.01 (0.003)	6	Genome
E. histolytica	0.01	0.01 (0.003)	7	Genome

s – number of segregating sites per site, π - average number of nucleotide differences per site, SD – standard deviation assuming free recombination, N – number of genes used to calculate indices, Source – indicates whether sequences were obtained from Whole Genome Projects, Expressed Sequence Tag Projects or PCR-based experiments.

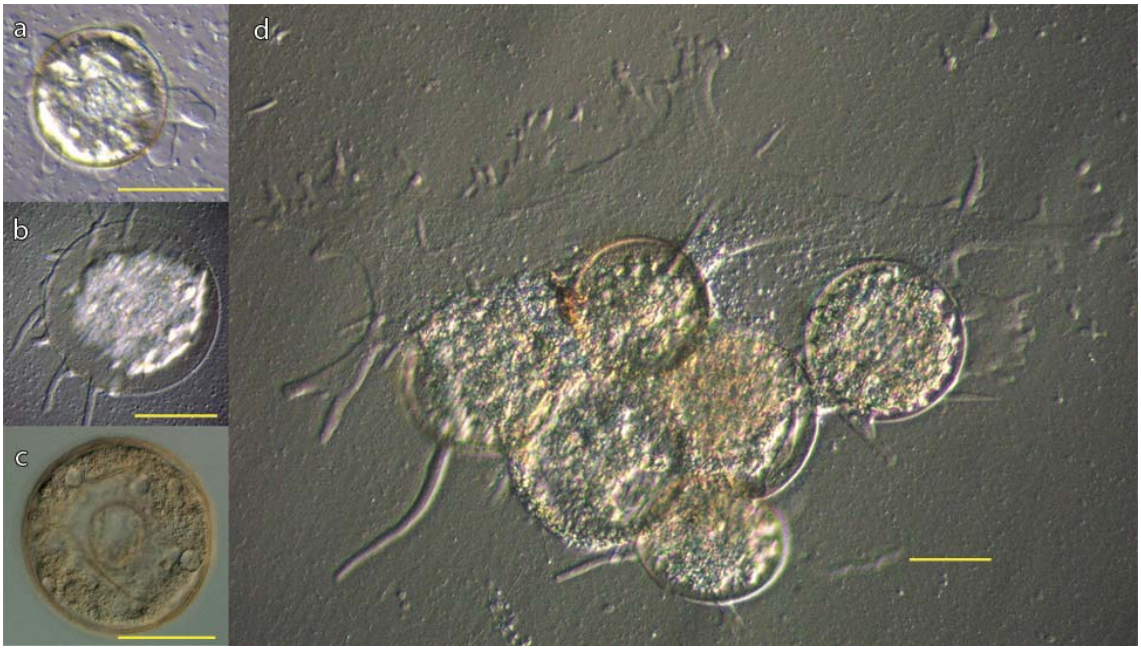

Figure 3.1: Representative images of *Arcella* lineages used in this study.

a- *A. hemisphaerica* Blue lineage. b- *A. hemisphaerica* Red lineage. c- *A. vulgaris* WP lineage. d- *A. vulgaris* SC lineage: the image shows 6 individuals undergoing plasmogamy, where multiple individuals fuse their cytoplasm. It is unknown whether nuclear fusion also occurs. Images a, b and d are Hoffman Interference Contrast, image c is Differential Interference Contrast. Scale bars are 50 um.

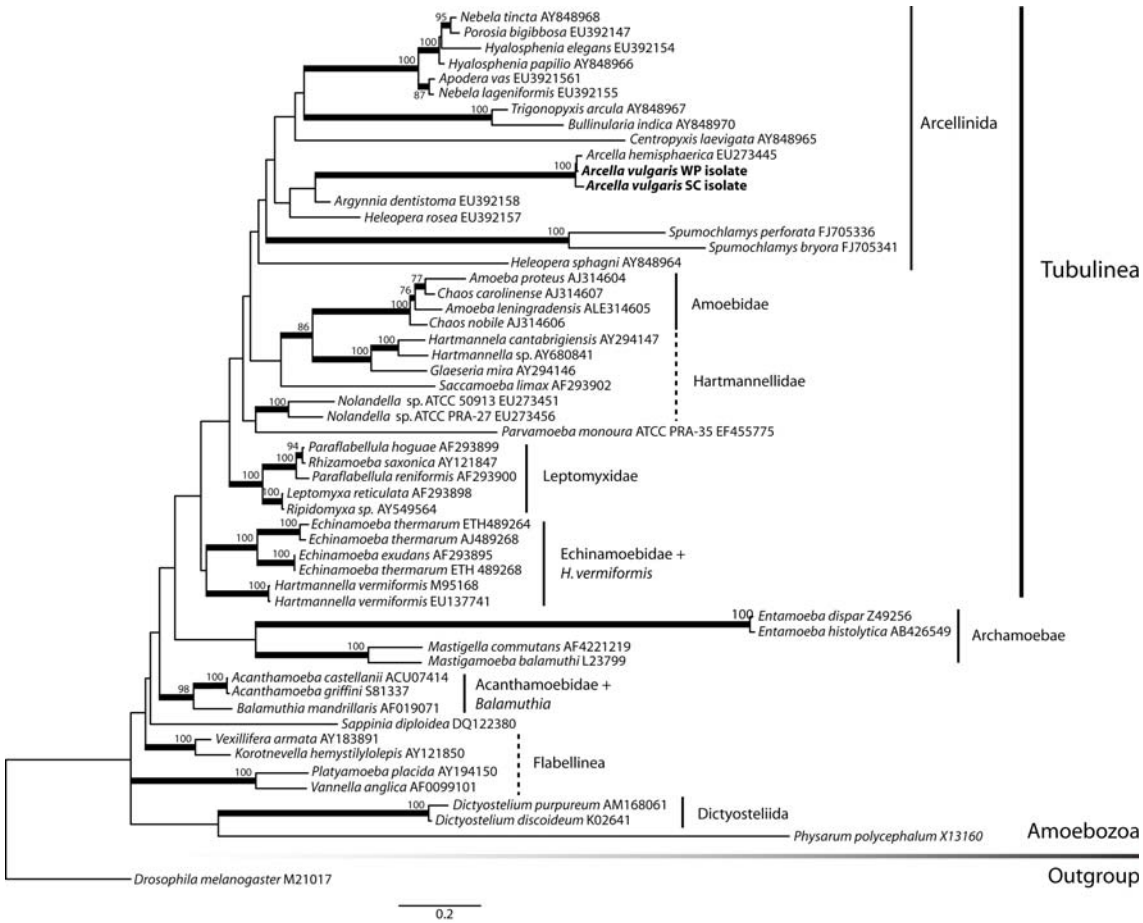

Figure 3.2: Most likely SSU-rDNA gene tree of the 'Amoebozoa'.

For this reconstruction, we used an alignment consisting of 55 taxa and 1587 characters, and performed a maximum likelihood analysis on RaxML using the GTR model of evolution and 200 bootstrap replicates (-ln*L* = 29300.662). The position of 2 new isolates of *A. vulgaris* are indicated in bold. Thicker branches represent nodes that have >75% bootstrap support. Branches are drawn to scale. Dashed lines represent paraphyletic groupings.

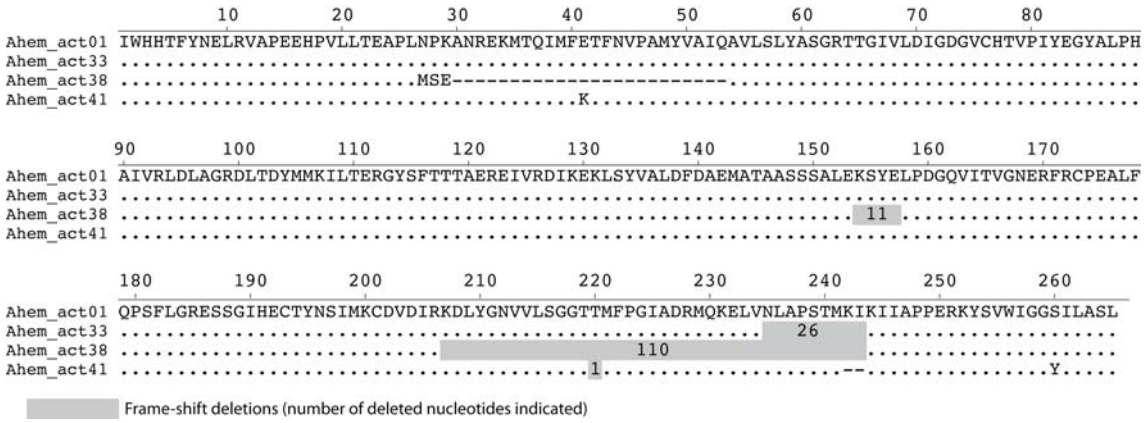

Figure 3.3: Recent frame-shifting deletions in *Arcella* actin genes.

The amino acid alignment compares three actin sequences to the most common actin found (Ahem_act01). Identities are shown as dots, and substitutions are indicated with respective amino acid symbol. Dashes show in-frame deletions and gray areas show frame-shifting deletions. Insets show the number of nucleotides deleted in each case. Notice that all three haplotypes show at least one frame-shifting deletion, but the amino acid sequence remains largely unchanged, suggesting that these deletion events are recent.

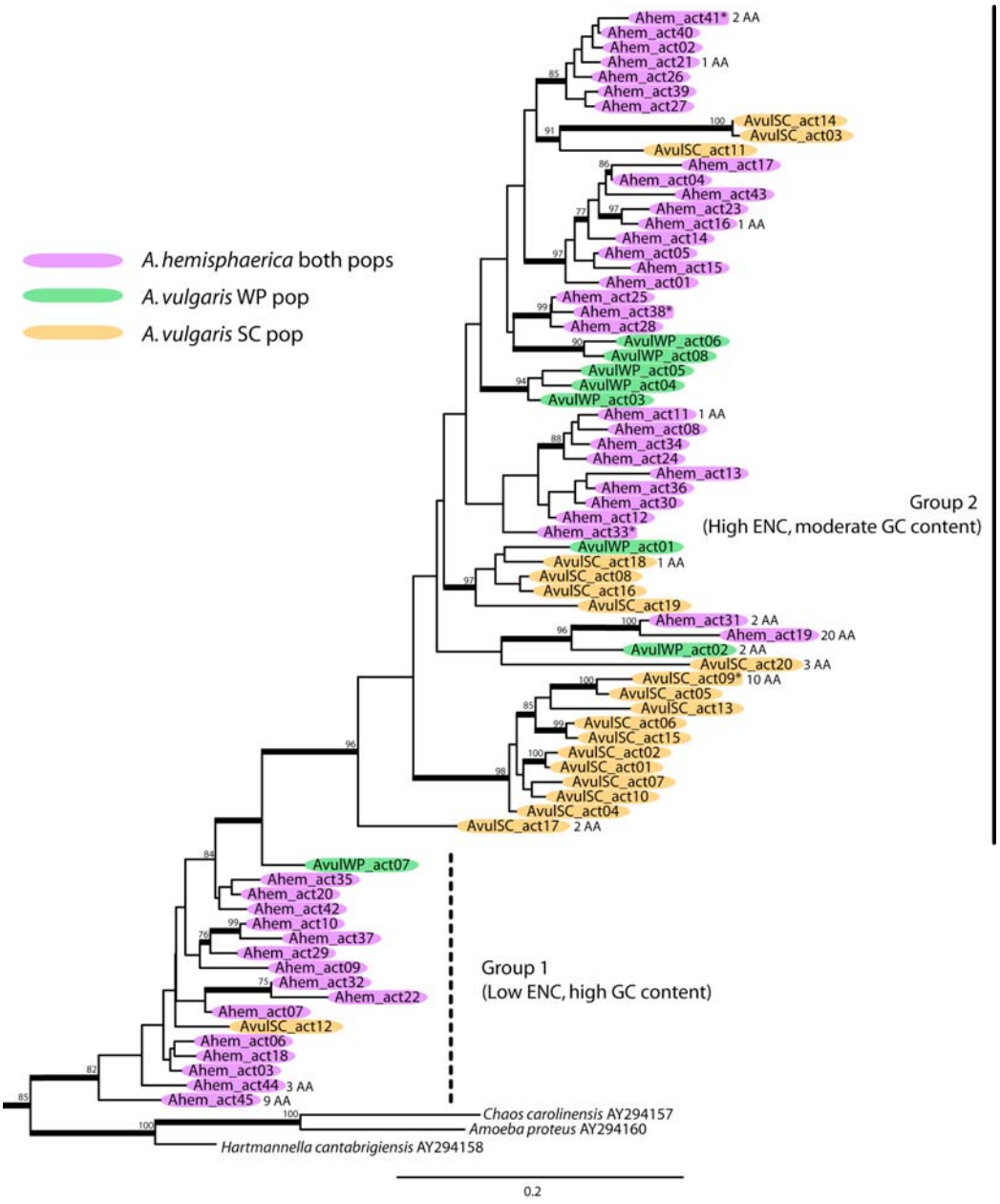

Figure 3.4: Most likely genealogy of the actin gene family in the genus *Arcella* inferred from maximum likelihood.

For this reconstruction, we performed a RaxML analysis using the GTR model of evolution with 300 bootstrap replicates on the dataset "Actin aligment B", which consists of 795 nucleotides, third positions are included. Genes are colored according to lineage. Most genes (59) code the exact same amino acid sequence. A minority of genes (14) encode divergent amino acid sequences, these are indicated by the number of AA substitutions. Sequences with an asterisk (*) represent putative pseudogenes. Thicker branches represent nodes that have >75% bootstrap support. All branches are drawn to scale. Dashed lines represent paraphyletic groupings.

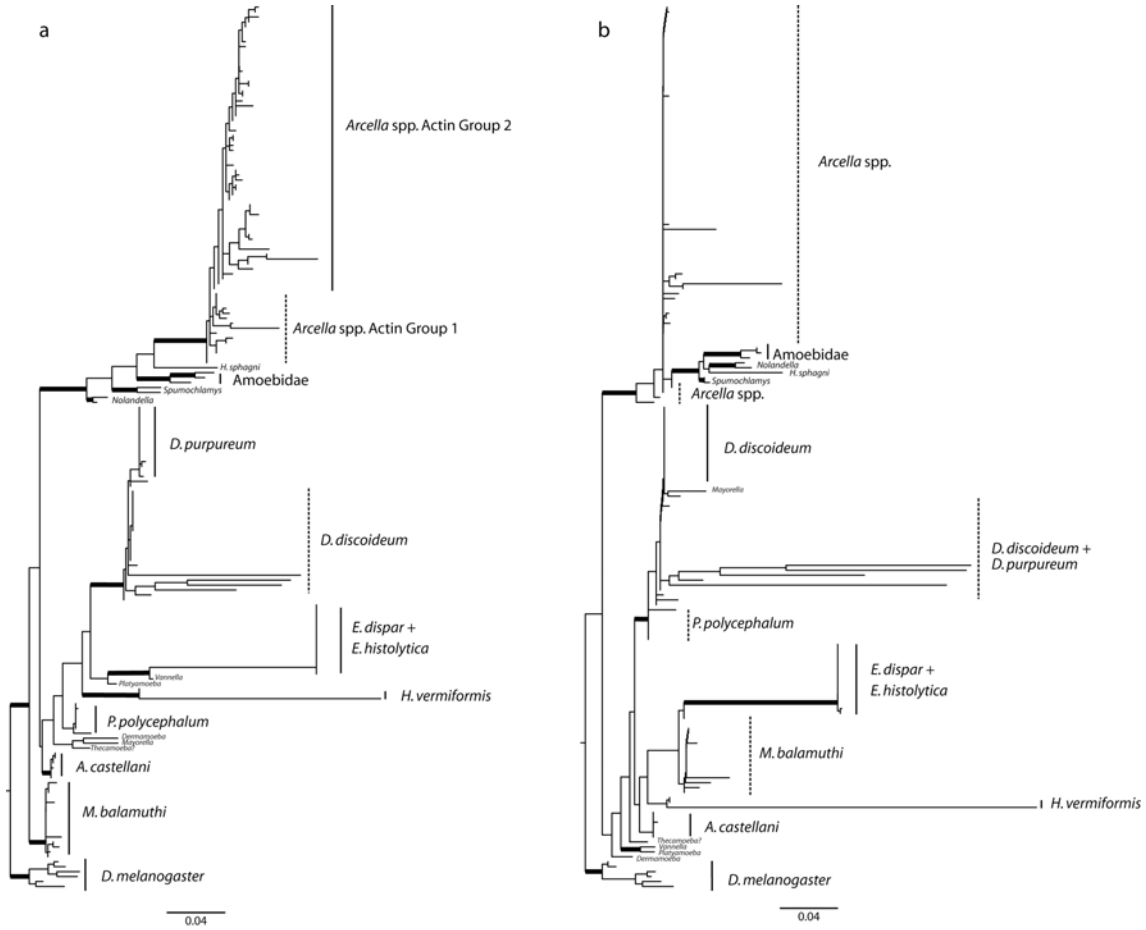

Figure 3.5: Genealogy of actin gene families across the 'Amoebozoa', showing multiple independent expansion in different lineages.

A - nucleotide tree without third positions inferred from the "Actin alignment A", which consists of 179 sequences and 756 (after excluding third positions from a total of 1134 sites) using maximum likelihood with the GTR model of substitution and running 650 bootstrap replicates (-ln*L* = 5991.952). B - amino acid tree inferred from translated "Actin alignment A", which consists of 179 sequences and 378 amino acid sites, using maximum likelihood with the JTT model of substitution and running 1000 bootstrap replicates (-ln*L* = 4392.092). Thick branches represent >75% bootstrap support for the backbone of the tree. All branches are drawn to scale. Dashed lines represent paraphyletic groupings.

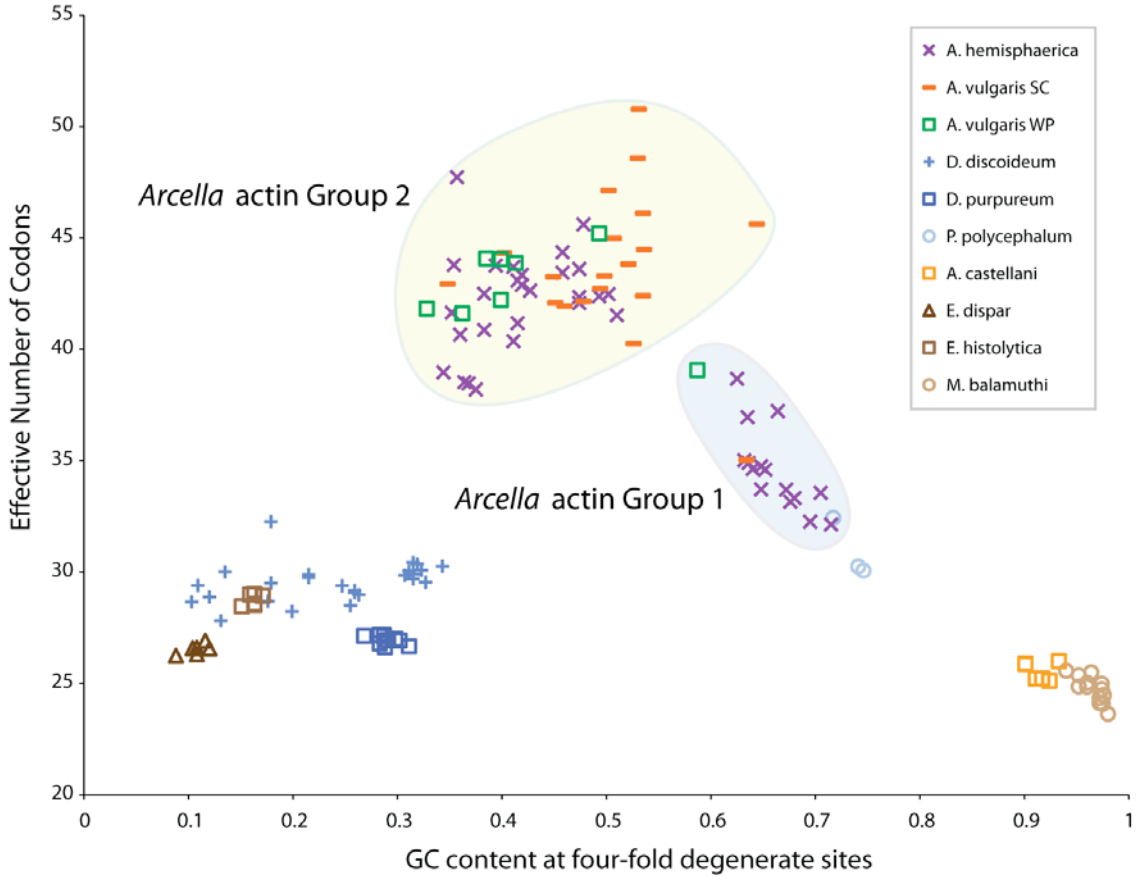

Figure 3.6: Effective number of codons (ENC) versus GC content in 4-fold degenerate sites for actin gene families in the 'Amoebozoa'.

This analysis is based on a subset of "Actin Alignment B" and comprises 149 sequences and 795 basepairs. Most gene families are restricted to an area of low effective number of codons as well as highly biased GC content, consistent with actins being highly expressed genes. The *Arcella* on the other hand are able to explore a more relaxed space regarding both ENC and GC content, to the middle-upper area of the graph.

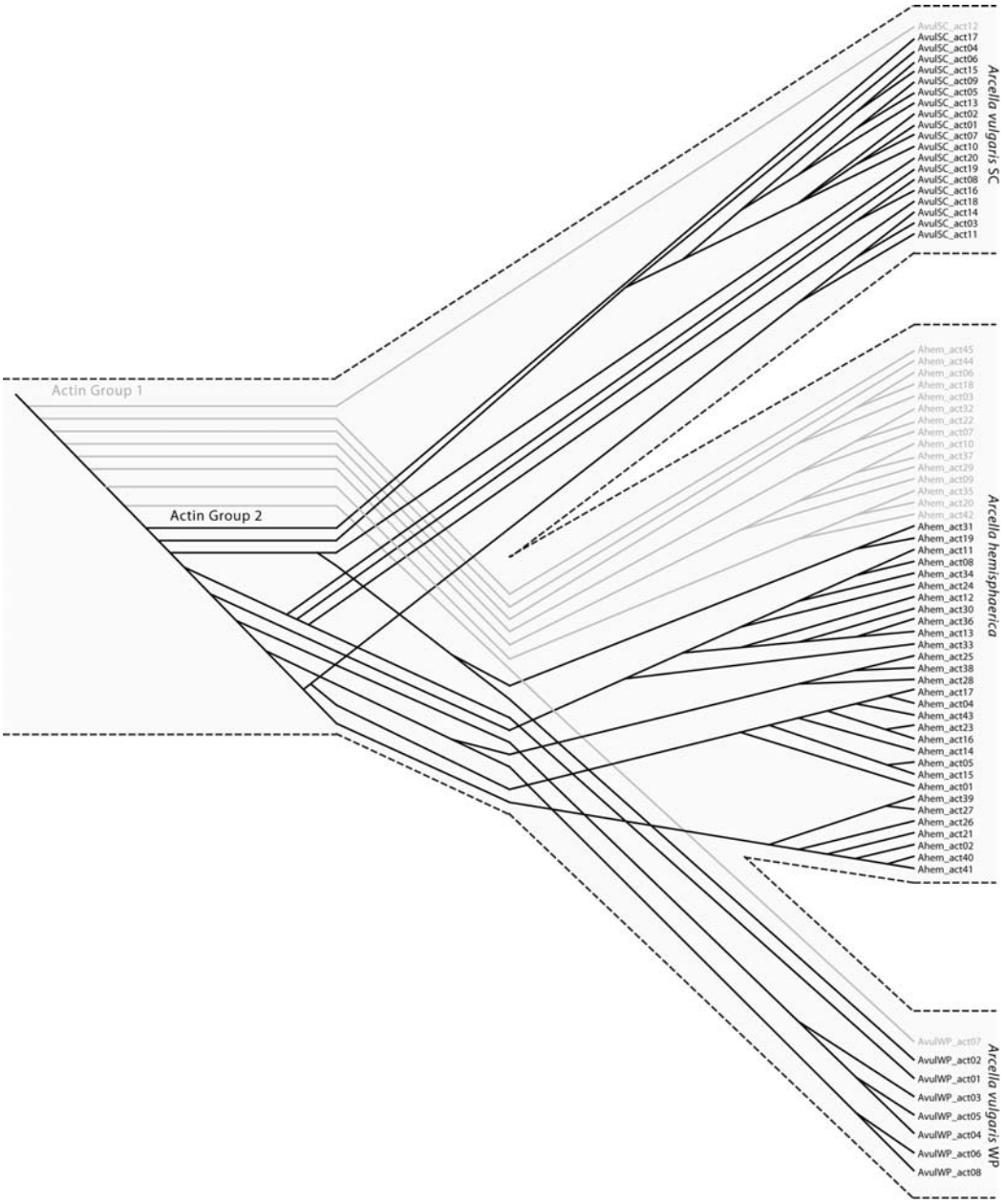

Figure 3.7: A hypothetical model for actin gene family evolution among species in the genus *Arcella*.

The branching order for species is obtained from the SSU-rDNA reconstruction and the branching order for actin paralogs is exactly as in fig. 4. The first event depicted is the separation of actins in 2 genomic groups (gray and black), which predates the divergence of lineages. Following separation, each group is under distinct regulatory constraints. Perhaps actins located in different areas are activated/deactivated following the life cycle, thus may be subject to different evolutionary pressures. Further, speciation happens, with maintenance of the 2 actin groups in all 3 lineages. Within each lineage there is a high level of independent duplications, the mechanism for which might be either a recombinational hotspot or developmentally regulated genome rearrangements (DRGR).

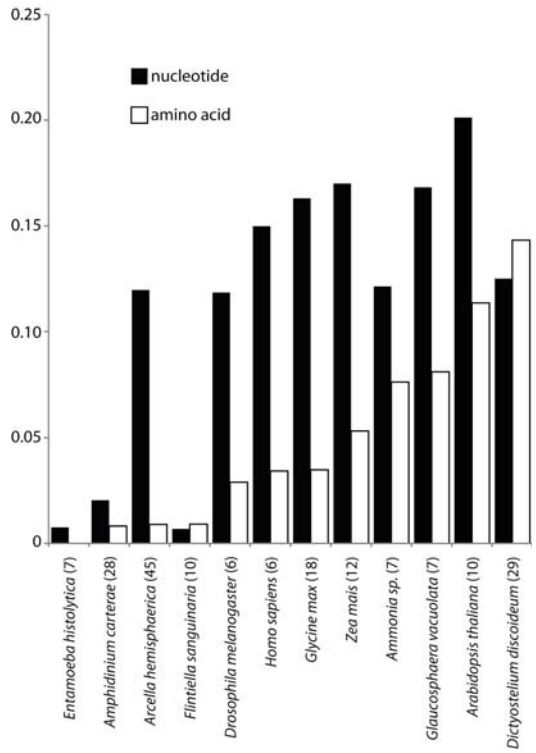

Figure 3.8: Average pairwise distances within actin gene paralogs for different eukaryotic taxa.

The number of members in each family is indicated in parenthesis after the organism name on the x-axis. The distances were calculated as uncorrected pairwise distances and then averaged over the number of actins in the taxon.

CHAPTER 4

COMPREHENSIVE PHYLOGENETIC RECONSTRUCTION OF AMOEBOZOA BASED ON CONCATENATED ANALYSIS OF SSU-RDNA AND ACTIN GENES.

4.1 Abstract

Evolutionary relationships within Amoebozoa have been the subject of controversy for two reasons: 1) paucity of morphological characters in traditional surveys and 2) haphazard taxonomic sampling in modern molecular reconstructions. These along with other factors have prevented the erection of a definitive system that resolves confidently both higher and lower-level relationships. Additionally, the recent recognition that many protostelids are in fact scattered throughout the Amoebozoa suggests that phylogenetic reconstructions have been excluding an extensive and integral group of organisms. Here we provide a comprehensive phylogenetic reconstruction based on 139 taxa using molecular information from both SSU-rDNA and actin genes. We provide molecular data for 13 of those taxa, 12 of which had not been previously characterized. We explored the dataset extensively by generating 18 alternative reconstructions that assess the effect of missing data, long-branched taxa, unstable taxa, fast evolving sites and inclusion of environmental sequences. We compared reconstructions with each other as well as against previously published phylogenies. Our analyses show that many of the morphologically established lower-level relationships (defined here as relationships roughly equivalent to Order level or below) are congruent

with molecular data. However, the data is insufficient to corroborate or reject the large majority of proposed higher-level relationships (above the Order-level), with the exception of Tubulinea, Archamoebae and Myxogastrea, which are consistently recovered. Moreover, contrary to previous expectations, the inclusion of available environmental sequences does not significantly improve the Amoebozoa reconstruction. This is probably because key amoebozoan taxa are not easily amplified by environmental sequencing methodology due to high rates of molecular evolution and regular occurrence of large indels and introns. Finally, in an effort to facilitate future sampling of key amoebozoan taxa, we provide a novel methodology for genome amplification and cDNA extraction from single or a few cells, a method that is culture-independent and allows both photodocumentation and extraction of multiple genes from natural samples.

4.2 Introduction

Reconstructing relationships between amoeboid organisms is challenging. The intrinsic paucity of morphological characters when compared to macroscopic taxa, as well as difficulties in establishing homology, made deep inferences nearly impossible for the ~200 years of studies based on microscopy. As a result, most taxa were lumped into the artificial Sarcodina (Pawlowski 2008). However, a number of well-defined morphological groups emerged from morphological information and are rarely disputed (Smirnov and Brown 2004), including lobose shelled amoebae (the Arcellinida); and the amitochondriate, parasitic amoebae with intra-nuclear mitotic spindles (the Entamoebidae). Major advances were achieved with the use of electronic microscopy techniques, but these generally helped stabilize further the lower-level relationships with

additional putative synapomorphies, rather than resolve deep relationships (*eg.* (Bovee 1985; Page 1987)).

With the advent of molecular techniques, amoeboid groups were found to be scattered across at least 30 lineages in the eukaryotic tree of life, with the amoebae producing lobose pseudopodia now included in the Amoebozoa (Cavalier-Smith 1998). It was only in the early 2000s that the promise of molecular phylogenetic reconstruction reached the fine-grained relationships within Amoebozoa, with well-sampled analysis of SSU-rDNA and actin genes (Amaral Zettler et al. 2000; Bolivar et al. 2001; Fahrni et al. 2003). The number of available amoebozoan sequences has increased steadily in the last decade, though not exponentially as occurred in other groups. A handful of medically important taxa and model organisms had their complete genomes sequenced or EST data made available (eg. *Dictyostelium discoideum* (Eichinger et al. 2005), *Entamoeba histolytica* (Stanley 2005)), but this sampling is still sparse making phylogenomic reconstructions difficult for this diverse group (Watkins and Gray 2008). Currently, there are about 150 diverse strains of Amoebozoa for which the SSU-rDNA has been characterized, followed by the actin gene for a few dozen lineages. These strains basically cover all the traditionally proposed morphological diversity (Page 1987; Smirnov and Brown 2004; Smirnov and Goodkov 1999).

The last few years provided further stabilization in purported relationships within the Amoebozoa (Figure 3.1). Two competing classifications emerged almost simultaneously: the higher-level taxonomic system of eukaryotes of Adl et al. (2005), and the Amoebozoa system of Cavalier-Smith et al. (2004). Subsequently, both systems were combined using both morphological and molecular evidence in the now standard

classification of Smirnov et al. (2005). Numerous additions have been made to the system of Smirnov et al. (2005), generally placing *incertae sedis* taxa without much modification into the higher-level proposed relationships (eg. (Corsaro et al. 2010; Dykova et al. 2008a; Fiore-Donno et al. 2005; Fiore-Donno et al. 2010; Kudryavtsev et al. 2009a; Lara et al. 2008; Romeralo et al. 2010)). Subsequent large-scale reconstructions largely corroborated the proposed relationships in the Smirnov et al. (2005) system (Kudryavtsev et al. 2005; Nikolaev et al. 2006; Pawlowski and Burki 2009; Tekle et al. 2008). Notable exceptions to this rule come from analyses of organisms traditionally considered slime molds. The Protostelia, once united by the ability to produce a unicellular fruiting body (Olive 1975), proved to be scattered in virtually every major branch of the Amoebozoa except for the Tubulinea (Shadwick et al. 2009). In addition, the sorocarpic slime mold *Fonticula alba* was shown to be more closely related to the opisthokont amoebae (Brown et al. 2009), and *Copromyxa protea* is shown to be in the Tubulinea (Brown et al. 2010). The implications and impact of these important insights are yet to be fully appreciated, either: 1) the ability to produce stalked fruiting bodies has emerged multiple times; 2) this ability has emerged once and was either lost or modified many times and; 3) many more lineages of amoebae are able to do so and the differences in the methodological traditions of typological protistology and mycology have failed to take this into account, as suggested by Shadwick et al. (2009).

Reconstructing these ancient relationships is an outstanding question difficult to resolve both due to the scattered understanding of the diversity of organisms and the highly heterogeneous rates of molecular evolution within the group (Nikolaev et al. 2006; Pawlowski and Burki 2009). The Amoebozoa may have radiated as early as 1200 Mya

(Berney and Pawlowski 2006), with the oldest unambiguous fossil being Arcellinida shells at 750 Mya (Bosak et al. 2011; Porter and Knoll 2000; Porter et al. 2003). Here, we provide a comprehensive reconstruction based on available data, concatenating SSU-rDNA and actin genes for 129 amoebozoan lineages and 10 outgroups. We introduce new molecular data for 13 lineages, 12 of which had not been previously characterized. In order to scrutinize the range of techniques used to reconstruct the Amoebozoa, we explore multiple iterations of taxa and data sampling, aiming to obtain reliable estimates of consistent groups, and to assess critically the support for proposed relationships. We perform comparative analysis using 18 different reconstruction approaches, including differential taxon sampling, removal of fast evolving sites, removal of long-branched and unstable taxa, and inclusion of environmental sequences. We test previously proposed relationships at both lower and higher-levels, and provide a summary of which groups are corroborated given the current molecular and, to a lesser extent, morphological data.

4.3 Methods

4.3.1 New taxa and morphology

Molecular sequences of SSU-rDNA and/or Actin were generated for 13 taxa (Table 3.1, Figure 3.2). The testate amoeba *Cryptodifflugia operculata* (Figure 3.2a-c) was isolated from a mixed Protozoa culture (Carolina Biological Supply Company, Cat. No. 131970). *Arcella mitrata* (Figure 3.2d-f), *Arcella gibbosa* (Figure 3.2g-i), *Arcella discoides* (Figure 3.2j-l), *Hyalosphenia papilio* (Figure 3.2s, t) and *Nebela carinata* (Figure 3.2z,a') were isolated from *Sphagnum* sp. moss in Hawley Bog, MA. *Pyxidicula operculata* (Figure 3.2m, n) was isolated from Hiddensee, Germany and kindly donated

to us by Mr. Wolfgang Bettighofer. *Gocevia fonbrunei* ATCC® 50196 (Figure 3.2o-r), *Stereomyxa ramosa* ATCC® 50982 (Figure 3.2u-y), *Stygamoeba regulata* ATCC® 50892 (Figure 3.2b'-e'), isolate CHINC-5 ATCC® 50979 (Figure 3.2f'-h'), *Thecamoeba* sp. ATCC® 50185, *Paraflabellula hoguae* ATCC® 30733 were obtained from the American Type Culture Collection (Manassas, VA).

All ATCC® species were identified following the original depositor information, and, when possible, comparison of photodocumentation provided by ATCC® to the original literature. We maintained the original depositor identification for all organisms except isolate CHINC-5 ATCC® 50979, which is certainly not a *Sexangularia* since it does not possess a shell (Figure 3.2f'-h'). This organism presents morphological characteristics similar to the dactylopodids, and will be described in detail elsewhere. The accuracy of the original identification for all other accessions will be discussed further after molecular analyses. However complete morphological characterization of these isolates is outside the scope of the current essay, and only limited morphological conclusions will be drawn.

The Arcellinida were identified by light microscopy and scanning electron microscopy where necessary (for electronic microscopy methods, see (Lahr and Lopes 2006)). We established a clonal culture of *Cryptodifflugia operculata*, whose morphological characteristics are in accordance with the original description (Page 1966), including the presence of a mucous aperture plug after encystation (operculum, Figure 3.2a). The operculum is regarded as the only distinguishing characteristic between *C. operculata* and the type species *C. oviformis* Penard 1890, and its use as a distinguishing character has been challenged as it may vary intra-specifically (Hedley et al. 1977). We

use Page's *C. operculata* definition since the operculum has indeed been observed in our isolate, and further research on non-operculum forming lineages is needed to elucidate this issue. Our clonal culture of *Pyxidicula operculata* had morphological characteristics in accordance with those described in Cash et al. (Cash et al. 1905). Some individuals presented a small funnel shaped rim attached to the inner side of the shell that is characteristic of *Pyxidicula patens* (Claparede and Lachmann 1859) indicating that the character may vary intra-specifically. The three *Arcella* isolates were identified in accordance with appropriate literature (Deflandre 1928; Lahr and Lopes 2009), *Arcella discoides* and *Arcella gibbosa* were culturable, while one *A. mitrata* individual was isolated from nature, photodocumented and genome amplified. *Hyalosphenia papilio* and *Nebela carinata* individuals were isolated from nature, photodocumented and further processed, morphological characteristics in accordance with those of Lara et al. (2008).

4.3.2 Molecular methods: DNA extractions, primers used, PCR conditions, cloning.

A combination of multiple methods was used to characterize both SSU-rDNA and actin genes from the diverse lineages. The ATCC® samples were processed as described in Tekle et al. (2008). Briefly, cultures were harvested and DNA extracted using DNA Stat60 (Tel-Test, Inc., Friendswood, Texas, Cat. No. TL-4220) following manufacturer's instructions, with the addition of a Phenol-Chloroform-Isoamyl step using Phase Lock Gel Heavy tubes (Eppendorf AG, Hamburg, Germany, Cat. No. 955154070).

We used multiple strategies for obtaining DNA from the testate amoebae species, due to their resistance to PCR methods and the difficulty in culturing some species. *Arcella gibbosa, Arcella discoides, Pyxidicula operculata* and *Cryptodifflugia operculata*

were cultured in autoclaved pond water enriched with cereal grass media extract and bacteria as described in (Lahr et al. 2010). DNA was extracted using a standard Phenol:Chloroform protocol on rapidly growing cultures as in (Lahr et al. 2010; Lahr and Katz 2009). *Arcella mitrata*, *Nebela carinata* and *Hyalosphenia papilio* were not amenable to culture, so we adopted two alternative strategies before PCR: whole genome amplification and cDNA extraction of single individuals. Briefly, for both strategies, a single or a small group of individuals were cleaned through several sterile pond water washes, left overnight to purge any remaining food/prey organisms being digested, re-washed in sterile pond water, and photo-documented in a light microscope. The individuals were then placed in either buffer DLB from a Repli-g Mini Kit (Qiagen, Cat. No. 150023) for whole genome amplification, or in Resuspension buffer with Lysis Enhancer from a SuperScript III CellsDirect cDNA synthesis kit (Invitrogen, Cat. No. 18080-200). Genome amplification and generation of complementary DNA libraries were then performed following manufacturer's instructions. PCR reactions on obtained DNAs were tested on a serial dilution (1x-1:1000 in ddH_2O), and the lowest concentration amplification was chosen to avoid formation of chimeras for further processing (Lahr and Katz 2009). Using this strategy enables a similar comparison to clonal cultures, since we have obtained the genetic material from a single individual. Primers for SSU-rDNA amplification were from (Medlin et al. 1988) with three additional primers used to generate overlapping sequences from each clone (Snoeyenbos-West et al. 2002), or shorter internal sequences for organisms where full SSU-rDNA amplification was not possible. Primers for actin were from (Tekle et al. 2007) and (Lahr and Katz 2009). Phusion Hot Start DNA polymerase (New England BioLabs, Cat. no.

F540) was used to amplify the genes of interest, and Zero Blunt TOPO cloning kits (Invitrogen, Cat. No. K280020) were used to clone PCR products. Cloned plasmid DNA was purified in a 96 well format using a PureLink Kit (Invitrogen, Cat. No. 12263018) and sequencing reactions performed using an ABI3100 sequencer (Applied Biosystems, Foster City, CA, USA) either at the Smith College Center for Molecular Biology (Northampton, MA) or at the Pennsylvania State Nucleic Acid Facility (University Park, PA, USA).

4.3.3 Multiple Sequence Alignments

4.3.3.1 SSU-rDNA datasets

Sequences for SSU-rDNA of 117 Amoebozoa and 10 Opisthokonta outgroups were retrieved from GenBank (see Supplementary Material available online at www.plosone.org for details), along with the 9 SSU-rDNA sequences generated in this study (Table 3.1) for a total of 136 SSU-rDNA sequences. Taxon sampling reflects an effort to include representatives of all available lineages in the 'Amoebozoa' (Adl et al. 2005; Pawlowski 2008; Pawlowski and Burki 2009; Shadwick et al. 2009; Tekle et al. 2008). Alignments were constructed in SeaView (Galtier et al. 1996; Gouy et al. 2010) with alignment algorithm MAFFT (Katoh et al. 2009) using the L-INS-I setting. Alignments were then curated manually to adjust ambiguous regions. This larger alignment was then subject to manual removal of ambiguous sites, to generate the dataset named M139 (Figure 3.3, Table 3.2). Independent automated removal of ambiguous sites was done using the online server GUIDANCE (Penn et al. 2010) with default parameters, to generate the dataset named A139 (Figure 3.3, Table 3.2).

Additional datasets with limited number of taxa were generated to explore the interaction between taxon sampling and missing actin sequences (Figure 3.3, Table 3.2). We removed taxa from both A139 and M139 to contain at least one representative of each major lineage, while maintaining all taxa for which actin sequences are available (43), to a total of 101 taxa, generating thus the alignments names A101 and M101 (Figure 3.3, Table 3.2, see Supplementary Material available online at www.plosone.org for a detailed list of taxon inclusion). Both datasets were subjected to further taxon removal to maintain only the 43 Amoebozoa lineages for which both actin and SSU-rDNA sequences are available as well as the 10 outgroup sequences, generating datasets A53 and M53 (Figure 3.3, Table 3.2). Datasets were then concatenated with the amino-acid actin dataset and subject to post-phylogenetic analyses treatment.

4.3.3.2 Actin datasets

Representative sequences for actin genes of Amoebozoa were retrieved from GenBank, curated Genome databases and EST databases, as detailed in (Lahr et al. 2010). The dataset, containing a total of 130 actin genes, 40 of them generated in this study (13 taxa, some with multiple paralogs, Table 3.1), was aligned at the amino-acid level in the software SeaView (Gouy et al. 2010) using the alignment algorithm MAFFT (Katoh et al. 2009) set to L-INS-I optimization, and trimmed down to retain only a central homologous region. The dataset for actin consists of 130 sequences with 265 amino acid sites. To choose sequences for concatenation, we determined the shortest branched actin genes for each group of paralogs, through a PhyML (Guindon and Gascuel 2003) analysis using a GTR model, with optimized estimation of invariable sites,

gamma variation with 6 rate categories across sites, combining the best of NNI and SPR searches, as implemented in Seaview (Gouy et al. 2010). We then trimmed the alignment to contain only the shortest branched paralog for each species, totaling 46 Amoebozoa taxa, and 265 amino-acid sites. This dataset was then concatenated to six SSU-rDNA datasets obtained in section 3.1 (A139, M139, A101, M101, A53 and M53). Additionally the alignment with all 130 paralogs was analyzed separately to determined events in the evolution of actin gene families in Amoebozoa. We performed maximum likelihood analyses on the amino acid dataset.

4.3.4 Phylogenetic analyses

4.3.4.1 Concatenated datasets

We performed maximum likelihood phylogenetic reconstruction in each of the initial six concatenated datasets using RAXML HPC 7.2.7 (Stamatakis 2006; Stamatakis et al. 2008) as implemented in the online server CIPRES (Miller et al. 2009). We ran 1000 fast bootstrap analysis using the GTRCAT approximation, and 100 independent maximum likelihood reconstructions using the GTRGAMMA model for the SSU-rDNA partition and the LG model for the protein partition. The most appropriate model for amino-acid evolution was determined using model testing implemented in the online server Datamonkey (Delport et al. 2010). Bootstrap values of the GTRCAT search were then plotted on the best tree found by maximum likelihood search for comparative analysis. Additional Mr. Bayes analyses were performed on the two largest datasets Auto139 and Manual139 to test independence of results from algorithm. We used the implementation on BioHPC cluster at Cornell University (http://cbsuapps.tc.cornell.edu/).

Using a random starting tree, the analyses did not converge after 20 million generations. Because Mr. Bayes is so computationally heavy, we had to resort to starting analyses from the best ML tree obtained in RAXML, although this may lead to exaggeration of support values in the final Bayesian tree. Hence, we started the analysis using the topology obtained in the RAxML analysis, and obtained convergence after 4 million generations. We performed two independent MCMC runs with 8 chains each, and a heating parameter of 0.05, with a burnin of 1 million generations. We applied the GTR+gamma model for the SSU-rDNA partition, and the WAG model for protein partition, since the available version of Mr.Bayes did not implement the LG model at the time of writing this report. The WAG model was the second best fit to our data according to the model selection analysis performed in the online server Datamonkey.

4.3.4.2 Removal of fast rate sites, long-branched and unstable taxa

To assess the effect of rate heterogeneity on SSU-rDNA topologies, we partitioned the Manual139 dataset into 8 rate classes using the GTR model with rate variation among sites following a discrete gamma distribution, as implemented in HyPhy v1.0beta (Pond et al. 2005). Classes 0 and 7 represent the slowest and fastest rate classes, respectively. We then proceeded to eliminate the fastest rate class (7) to generate the alignment M139-7 (Table 3.2). Similarly, we removed the two fastest rate classes (7 and 6) for the dataset M139-76, and the three fastest rate classes (7, 6 and 5) for the dataset M139-765 (Table 3.2).

To assess the effect of long-branched taxa on final topologies the root-tip branch lengths of each terminal from section 4.1 was calculated as implemented in the freely

distributed program TreeStat v1.2 (http://tree.bio.ed.ac.uk/software/treestat/). The results were then compared within reconstructions and we proceeded to remove the 10 overall longest branched taxa (*Arcyria denudata, Dydimium nigripes, Echinostelium arboreum, Lindbladia tubulina, Pelomyxa palustris, Physarum polycephalum, Polysphondylium pallidum, Protophysarum phloiogenum, Trichia persimilis* and *Tricosphaerium* sp. ATCC 40318 (a list of all branch lengths is available as Supplementary Material online at www.plosone.org), to generate the alignments M139-LB and A139-LB, with a total of 129 taxa each (Figure 3.3, Table 3.2).

To assess the effect of unstable taxa on final topologies we calculated terminal Leaf Stabilities (Thorley and Page 2000) as implemented by the script THOR (http://code.google.com/p/phylogenetics/) using as input the outgroup-rooted 1000 bootstrap trees generated from Section 4.1. After performing comparative analysis between the different datasets, we removed the 10 most unstable taxa (the three *Cochiopodium* spp., *Dermamoeba algensis, Endostelium zonatum, Gocevia fonbrunei* ATCC 50196, *Pessonella* sp., isolate CHINC-5 ATCC 50979, *Trichosphaerium* sp. and *Vexilifera minutissima*) to generate the datasets M139-us and A139-us, with a total of 129 taxa each (Figure 3.3, Table 3.2). Additionally, we generated datasets by removing both the most unstable taxa and the most long-branched taxa at the same time, to a total of 120 taxa in the dataset A139-LB-us and M139-LB-us (Figure 3.3, Table 3.2).

4.3.4.3 Sampling of environmental sequences

A next logical step for our analyses was to determine whether increased taxon sampling will enable more robust phylogenetic reconstructions. An available method

widely used to increase taxon sampling is to add environmental sequences that represent unculturable organisms or taxonomic representatives in environments that were not yet studied by specialists. The number of environmental sequences available is very large, and there is a tendency to recover closely related organisms since most environmental studies are focusing on a specific type of habitat, rather then targeting phylogenetic coverage. It is desirable then to use representatives from different parts of the tree rather than many representatives in a single branch (cherries). We performed BLAST searches querying all 129 Amoebozoa taxa in our dataset against the environmental database in GenBank. We retrieved the top 100 hits for each taxon to create a combined dataset, excluding redundant sequences of ~3,000 entries. We then eliminated all entries that are 98% similar to each other using the Rid v0.3 script (Grant, J.). This approach recovered 25 sequences that were then included in the M139 datasets, generating the dataset MEnv (Figure 3.3, Table 3.2).

4.3.5 Comparative analyses of resulting trees

We used three methods to assess the information in our reconstructions: comparison of bootstrap supports for different levels of groupings, Treeness Index and Leaf Stabilities. For comparative analysis of support for different groupings, we divided the hypothesized groupings in two categories: higher-level relationships and morphology based lower-level relationships. We then assessed bootstrap supports from the 18 reconstructions performed to compare stability of clades between analyses. We also compared data for the Treeness index, a measure of the proportion of total tree length that is taken up by internal branches, thought to be a rough assessment of how much of the

dataset's information is being used to reconstruct stem relationships as opposed to substitutions along terminal branches. We calculated Treeness values as implemented in TreeStat (http://tree.bio.ed.ac.uk/software/treestat/). Finally, we calculated the average leaf stability for each reconstruction; this is useful in informing how much overall instability is present in a particular dataset, and whether our manipulations are improving overall resolution.

4.3.6 Approximately unbiased (AU) testing of alternative hypotheses

We tested whether non-recovered hypotheses could be rejected using the Approximately Unbiased test (Shimodaira 2002). Briefly, we generated maximum likelihood reconstructions with constraints for each of 12 alternative hypotheses by running 100 independent maximum likelihood analysis in RAxML using the exact same parameters as before, and choosing the most likely tree. All trees were then compared to the best tree found on the standard analysis using RAxML to calculate per-site likelihoods. The per-site likelihoods were then analyzed in CONSEL (Shimodaira and Hasegawa 2001) with standard parameters to obtain p-values.

4.4 Results

4.4.1 General topology

The SSU-rDNA and actin genes for 13 lineages were sequenced (Figure 3.2, Table 3.1) and phylogenetic analyses were performed on a total of 139 taxa, using multiple reconstruction strategies (Figure 3.3, Table 3.2). Topologies obtained in the 18 distinct phylogenetic reconstructions of concatenated SSU-rDNA and actin genes (Table

3.3) largely agree with previous analyses regarding the monophyletic status of lower-level relationships (defined here as in roughly equivalent to Order level or below). These groups are also consistent with morphological characters, as outlined in Smirnov et al. (2005) and Shadwick et al. (2009): the Amoebidae, Dictyosteliida, dark spored myxogastrids, Hartmannellidae (excluding *Saccamoeba limax* ATCC® 30942), Leptomyxida, protosporangiids, protosteliids, schizoplasmodiids, soliformoviids and Tubulinida are always recovered with high support (Table 3.3); the Acanthamoebida, cavosteliids, Dactylopodiida, Echinamoeboidea, light spored myxogastrids, Mastigamoebida, Pelobiontida, Thecamoebida and Vannellida are recovered with moderate to high support; the Arcellinida are recovered with low support (Table 3.3).

In contrast, almost all proposed higher-level relationships (defined here as above the Order level) are not recovered in our analyses, with three exceptions (Table 3.3): 1) the Myxogastrea (=myxomycetes) are recovered with high support in virtually all analyses, and both proposed nested groups are also strongly supported (dark spored myxogastrids and light spored myxogastrids); 2) the Tubulinea is recovered with moderate to high bootstrap supports in 15 out of 18 analyses, and 3 of the 4 group members Echinamoeboidea, Leptomyxoidea and Tubulinida are consistently recovered with moderate to high bootstrap supports. The fourth group, Arcellinida is recovered with low support in 13 out of 18 analyses. A further group within the Tubulinea (Hypothesis 1 – 'Poseidonida', see below) is highly supported in all analyses (Table 3.3); 3) the Archamoebae are recovered in 8 out of 18 analyses with weak to moderate support, the two proposed groups within are also moderately supported, the Pelobiontida is

recovered with moderate to high support in 7 out of 14 analyses, and the Mastigamoebida in 8 out of 16.

Another two higher-level relationships worth noting are inconsistently recovered. The Mycetozoa *sensu* Cavalier-Smith *et al.* (2004) (Archamoebae+Dictyostellida+Protosteliidae) are only recovered in analysis with low number of taxa included (Analyzes A53, M53 in Table 3.3). The Flabellinea are only recovered in analysis where long branched taxa and/or unstable taxa were removed (Table 3.3). All other proposed higher-level relationships are never recovered in our reconstructions: Flabellinea, Conosea, Discosea, Stellamoebia, Variosea and Varipodida (Table 3.3), but these are also not rejected using an AU test (Table 3.4).

4.4.2 Placement of newly characterized lineages

4.4.2.1 Arcellinida lineages

The newly introduced Arcellinida sequences consistently group with previously available lobose testate amoebae. The *Nebela carinata* and *Hyalosphenia papilio* fall consistently with other members of the Hyalospheniidae previously sequenced (Figure 3.4). The three new lineages of the genus *Arcella* also consistently group with the other available *Arcella*, including *Arcella discoides,* only represented by actin genes (Figure 3.4). This demonstrates that at least in principle we should be able to infer relationships for the other two lineages represented only by actin genes (see below *Steromyxa ramosa* ATCC® 50982 and isolate CHINC-5 ATCC® 50979), as long as taxonomic sampling is significant. *Pyxidicula operculata* and *Cryptodifflugia operculata,* both representing previously unsampled genera, fall consistently in the Arcellinida, but with no consistent

home. The Arcellina hypothesis, which unites the testate amoebae that have secreted chitinous shells (Meisterfeld 2002), would encompass the *Arcella, Pyxidicula* and *Spumochlamys*, but was not recovered.

4.4.2.2 Other Amoebozoa lineages

The ATCC® accession 50196 identified as *Gocevia fonbrunei* is found to be strongly related to the protosteloid amoeba *Endostelium zonatum.* This relationship is moderately or highly supported in 9 out of 10 analyses where both taxa were present (Figure 3.4, Table 3.3). Further, *Gocevia fonbrunei* + *Endostelium zonatum* is monophyletic with *Cochliopodium* spp., albeit with moderate or low bootstrap supports in 9 out of 11 analyses where all taxa were present (Figure 3.4, Table 3.3). The ATCC® accession 50185, deposited as a member of the genus *Thecamoeba*, is nested within the genus *Sappinia*, with high support in all analyses (see Supplementary Material available online at www.plosone.org for details). *Sappinia* is in its turn the sister-group to the genus *Thecamoeba* (Brown et al. 2007; Dykova et al. 2010b). Analysis of the SSU-rDNA sequence performed by BLAST reveals that ATCC® 50185 is almost identical (99% similarity) to a specimen identified as *Sappinia* sp. Noaf EU881941 (Wylezich et al. 2009), presumably related to *Sappinia diploidea*. This is an indication that isolate ATCC® 50185 is in fact a novel *Sappinia* lineage, and further research into its morphology should shed light on the distinctions between the two genera. The ATCC® accession 50892 identified as *Stygamoeba regulata*, and with morphological characters consistent with the original description (Smirnov 1996) does not reliably fall into any of

the proposed groups (see Supplementary Material available online at www.plosone.org for details). Leaf stability analyses do not indicate this as a particularly rogue taxon.

4.4.2.3 Lineages represented only by actin genes

The two non-Arcellinida lineages for which we were only able to amplify the actin gene do not group reliably with any other Amoebozoa, which may either indicate their status as *incertae sedis* is granted, or that a single gene is not sufficient to reconstruct their evolutionary history. The ATCC® accession 50982 deposited as *Stereomyxa ramosa* does not reliably fall into any of the proposed groups, or lower-level morphological relationships. In most reconstructions, it falls outside of the Archamoebae, but this is not supported by bootstrap analyses. The leaf stability index for this taxon is generally one of the lowest, ranking 26 out of 29 (29 being the most unstable taxa), further confirming its status as *incertae sedis* at least for this single gene. The isolate CHINC-5 ATCC® 50979 (misidentified as *Sexangularia* sp., see Material and Methods section) is found to be related to the also *incertae sedis Pessonella* sp., albeit with low bootstrap support. Leaf stability analysis shows that both taxa are unstable, ranking 27 and 26 out of 29.

4.4.3 Comparative analyses of different types of reconstruction

The general performance of 18 different reconstruction approaches was assessed by three measures: bootstrap supports of well-established morphological groups and proposed higher-level relationships (Table 3.3); leaf stability measures (Table 3.5, Supplementary Table S2); and Treeness indices (Table 3.5). Overall, trees tend to score

higher with more taxa added; when manual removal of ambiguous sites is performed and when long branched as well as "rogue" taxa are removed (see Supplementary Material available online at www.plosone.org for a detailed discussion). Since removal of 19 long branched or unstable taxa significantly impairs interpretation of relationships (for instance, Pelobiontida and Myxomycetes are almost completely removed), we consider that both Mr. Bayes and RaxML reconstructions based on the dataset with 139 taxa and manual removal of ambiguous sites (M139, Table 3.2) best represents our results (Figure 3.4), and comparisons will be made to other reconstructions as necessary.

4.4.4 Addition of environmental sequences

The addition of 25 environmental sequences neither improves support for the groups recovered in other reconstructions, nor stabilizes rogue taxa. The added sequences group with: Arcellinida (11 sequences), Mastigamoebidae (3 sequences), Hartmannellidae (2 sequences), undetermined (2 sequences) and one sequence in each *Cochliopodium*, Echinamoebidae, Filamoebidae, Myxogastrea, Poseidonidae, protosteliids and *Saccamoeba*. The bootstrap supports for lower-level relationships remain largely unchanged when compared to other types of reconstruction (Table 3.3); the average Leaf Stability is not significantly different from reconstructions with large taxon sampling (Table 3.5) and the Treeness index decreases when environmental taxa are added, probably the result of an increase in total tree length without a concomitant increase in signal (Table 3.5).

4.4.5 Actin gene family reconstruction

A reconstruction using multiple actin gene paralogs for 46 Amoebozoa taxa largely agrees with the reconstruction in Lahr et al. (2010) (Figure 3.5). Using a reconstruction based on amino acids fails to recover monophyly of Amoebozoa, because under these conditions the Opisthokont *Amoebidium* does not fall as an outgroup. Still, many lower-level relationships are recovered (Figure 3.5): Leptomyxida, Tubulinida, Thecamoebidae, and one of the well-supported higher-level relationships is recovered: Archamoebae. However, the isolate *Hartmannella vermiformis* does not fall into the Tubulinea, another well-supported high-level relationship in the concatenated reconstruction. The Arcellinida appear as paraphyletic with the invasion of Tubulinida (Amoebidae+Hartmannellidae), indicating that some of the actin paralogs in these lineages may be ancient duplicates. Additionally, throughout the tree many taxa display recent independent expansions of the actin gene family (*Arcella*, *Cryptodifflugia*, *Dictyostelium*, *Phalansterium*, *Trichosphaerium*, *Gocevia*).

4.5 Discussion

Our analyses of available SSU-rDNA and actin genes confirm the monophyly of several previously reported lower-level relationships (defined here as roughly equivalent to Order level and below) within the Amoebozoa, and indicate an additional six uncharacterized well-supported relationships (Figure 3.4, Table 3.3). However, only three of the previously proposed higher-level relationships (defined here as deep relationships that are above the Order level) are consistently recovered: the Myxogastrea are strongly supported; the Tubulinea are moderately supported; and the Archamoebae

are weakly supported. Other proposed higher-level relationships such as Conosea along with the included Mycetozoa and Stellamoeba, as well as the Protamoebae with the included Discosea and Variosea are never recovered, but our data also does not reject these relationships (Table 3.4). The three recovered higher-level relationships are distinguished from other proposed groups in that they all have well-established morphological synapomorphies: the Tubulinea present cylindrical pseudopods with monoaxial streaming (Smirnov et al. 2005); the Archamoebae unite all amitochondriate Amoebozoa (likely a secondary loss (Patterson 1999), rather than a primitive condition as previously suggested elsewhere (Cavalier-Smith 1991)); and the Myxogastrea are characterized by a fruiting body arising from a syncytial diploid stage (Fiore-Donno et al. 2010). The remaining non-confirmed higher-level relationships (Table 3.3), which were proposed largely based on analysis single gene analyses of SSU-rDNA, are not marked by strong morphological synapomorphies.

Most of the morphologically defined lower-level relationships are reliably recovered, as well as six previously undescribed groups, referred to here as Hypothesis 1-6 (Figure 3.4, Table 3.3). Proposed names for each hypothesis are stated in single quotes, to denote their speculative nature, and a taxonomic summary is provided for each group following regulations of the International Code of Zoological Nomenclature. The Tubulinea and nested groups are consistently well supported: Echinamoebida, Leptomyxida and Tubulinida are moderately to strongly supported and the Arcellinida is consistently recovered, albeit with weak support. Additionally Hypothesis 1 (‘Poseidonida’), a monophyletic group composed of *Nollandela* spp. and ‘*Hartmannella*’ *abertawensis* is distinct from the other four groups in the Tubulinea (Figure 3.4, Table

3.3). Indications of this relationship have been shown in previous reconstructions (Brown et al. 2010; Fiore-Donno et al. 2010), and our analysis suggests that this strongly supported group (Table 3.3) is not embedded within any other Tubulinea clade. *Nolandella* spp. and '*Hartmannella' abertawensis* were isolated from near-shore marine environments in the same publication (Page 1980). Another species with similar morphological features, *Hartmannella vacuolata*, also marine, has been described with notes about the unusual feature for limax amoeba of a floating form with extended arms (Anderson et al. 1997), a character shared with *Nolandella*. Given the stable status of this clade, and the fact that the organisms share the marine environment as a habitat, we suggest the name 'Poseidonida', in reference to the Greek god of the seas, Poseidon (see taxonomic summary for a formal account). The type genus and species for the group should be *Nolandella hibernica* (Page 1980) for stability reasons, since *Hartmannella abertawensis* will likely require re-assignment to a new genus with further research.

The genus *Soliformovum*, common protosteloid amoebae found associated with dead plant material (Spiegel 1990), forms a monophyletic group with *Grellamoeba robusta*, an amoeba isolated from fish kidneys (Dykova et al. 2010a), which we designate as Hypothesis 2 ('Fractovitellida', Figure 3.4, Table 3.3). *Grellamoeba robusta* is putatively related to *Acramoeba dendroida* based on SSU-rDNA analysis (Dykova et al. 2010a), which justified inclusion in the group Acramoebidae (Smirnov et al. 2008). However Dykova et al. (Dykova et al. 2010a) emphasize that no well-supported relationships could be found in their analysis, either morphologically or phylogenetically, so they settled for an *incertae sedis* status. *Acramoeba dendroida* never groups with *G. robusta* in our analyses, suggesting *A. dendroida* is still the only representative of the

Acramoebidae. On the other hand, *G. robusta* composes a new, highly supported clade with two soliformoviids (Hypothesis 2, see taxonomic summary for details). *Soliformovum* spp. was removed from the genus *Protostelium* based on a series of gross morphology and ultrastructural characteristics (Spiegel et al. 1994). Spiegel et al. (Spiegel et al. 1994) suggests that the nucleus with an irregular, multilobed nucleolus is a putative synapomorphy of the genus *Soliformovum*, although the cavosteliid *Schizoplasmodiopsis amoeboidea* also presents a diffuse nucleolus (Lindley et al. 2006; Shadwick et al. 2009). *Grellamoeba robusta* presents oval nucleoli more similar to *Protostelium* spp. and *S. amoeboidea* as a trophozoite, but shows a lobed morphology in cyst form (Dykova et al. 2010a), which may be consistent with the *Soliformovum-Schizoplasmodiopsis* type (Lindley et al. 2006). The micrographs provided by Dykova et al. (Dykova et al. 2010a) do not indicate that *G. robusta* has a microtubular organizing center (MTOC), so this is possibly a further shared characteristic with the genus *Soliformovum* (Spiegel et al. 1994). Both amoebae are generally uninucleate, without pigmentation and exhibit multiple contractile vacuoles. They both present sharply pointed sub-pseudopodia an thus an acanthopodid morphotype (*sensu* Smirnov et al. (Smirnov et al. 2005)). However, *G. robusta* tends to be more branched and exhibit fan-shaped regions, while *Soliformovum'*s entire cell tends to be fan-shaped and less branched. No sorocarp formation was observed in *G. robusta* (Dykova et al. 2010a), making this novel relationship a suitable clade to further research the evolution of fruiting body formation in amoebae. We suggest this grouping be named 'Fractovitellida' (*fractus*-broken, *vitellum*-yolk) in reference to the diffuse aspect of the nucleoli, with type

genus and species *Soliformovum irregularis* (Olive and Stoinanovich 1969), see taxonomic summary for a formal account.

Our analyses confirms the highly supported grouping of filopodia producing Amoebozoa in the genera *Flamella* and *Arachnula* sp. ATCC® 50593, which we designate as Hypothesis 3 ('Flamellidae', Figure 3.4, Table 3.3). *Flamella* are characterized by a fan-shaped morphology, with a wide anterior hyaloplasm that produces thin sub-pseudopodia and long trailing thin filipodia. Trophozoites present a central non-diffuse nucleolus, although *F. balnearia* shows an irregularly shaped nucleolus in the cyst form (Kudryavtsev et al. 2009b). Morphological information for *Arachnula* ATCC® 50593 reveals that it is a multinucleate amoeba with branched thin filopodia (Tekle et al. 2008). This monophyletic relationship is within the moderately supported clade Hypothesis 4 ('Gracilipodida') as sister to *Filamoeba* spp., characterized by a flattened locomotive form with a thin anterior hyaloplasm and long, thin, filiform subpseudopodia (Dykova et al. 2005; Page 1967a). Hypothesis 4 has also been previously recovered, along with other environmental sequences (Kudryavtsev et al. 2009b; Nikolaev et al. 2006). However, the previously proposed relationship between *Flammella* and *Acramoeba dendroida* is not recovered (Smirnov et al. 2008). Gross morphological features characterize Hypothesis 4 as outlined in Kudryavtsev et al. (Kudryavtsev et al. 2009b), but no putative ultra-structural synapormophies can be suggested at this point. The corroboration of both hypothetical clades in our analyses justify the designation of two nested amoeboid groups: Hypothesis 3 'Flammellidae', containing *Flamella* + ATCC® 50593; and Hypothesis 4 'Gracilipodida' (*gracilis*-slender, *pedes*-foot), in reference to the filose pseudopodia present in all members of the

clade. The type genus and species for both groups is *Flammella magnifica* Schaffer 1926 according to the Principle of Priority (see taxonomic summary for a formal account).

The identification of ATCC® 50593 as *Arachnula* sp. (Tekle et al. 2008) has been the subject of some controversy (Bass et al. 2009). Bass et al. (2009) suggest that large terminal fans provided with many thin reticulating pseudopodia, a conspicuous character in Cienkowski's description of *Arachnula* (Cienkowski 1876), are not present in the available images of isolate ATCC® 50593 (Tekle et al. 2008). Bass et al. (2009) isolated an additional organism that they argue is more consistent with the original description (Cienkowski 1876). In molecular analysis of SSU-rDNA, this organism falls in the Rhizaria along with other similar forms such as *Platyreta*. Bass et al. (2009) then suggest that ATCC® 50593 is misidentified, and is instead more closely related to *Acramoeba dendroida* (Smirnov et al. 2008), but these do not group together in the current report. The isolate ATCC® 50593 instead is included in the well-supported clade of filopodia bearing Amoebozoa (Hypothesis 4 'Gracilipodida') enforcing the notion that extremely similar, convergent morphologies are present in Amoebozoa and Rhizaria (Bass et al. 2009), corroboration based on molecular data is necessary to determine relationships. The taxonomic identity of *Arachnula* is further obscured because the organism in Figure 8 of Bass et al. (2009) was unfortunately not amenable to culture (therefore cannot be studied further), and the authors themselves raise the possibility of contamination. Establishing a taxonomic identity by comparing traditional descriptions with modern techniques is a complicated affair (Lahr et al. 2008; Lahr and Lopes 2009), and is made worse in this case by the multiple uncertainties introduced by different studies. The

question of which organism is the real *Arachnula*, either ATCC® 50593 or the organism pictured in Figure 8 of Bass et al. (2009) remains an open debate.

Hypothesis 5 (‘Goceviidae’) unites the amoeba *Gocevia fonbrunei* ATCC® 50196 and the protosteloid amoeba *Endostelium zonatum*, a relationship that has been previously suggested based on ultra-structure (Bennett 1986). Although we present limited morphological data on ATCC® 50196, its morphology is generally consistent with that of *Gocevia fonbrunei* as having a lens-like locomotive morphology, few thin subpseudopodia and covered in a hyaline cuticle without foreign bodies and an unornamented cyst (Page 1976; Page 1987; Rogerson and Patterson 2002; Smirnov and Brown 2004). The protosteloid amoeba *Endostelium zonatum* is characterized by a fibrous covering, and the amoeba has numerous thin subpseudopodia (Olive et al. 1984). The taxonomic status of this organism has been a conundrum, and has evaded classification in relation to other protosteloid amoebae (Patterson 1999; Shadwick et al. 2009; Spiegel 1990), the very monophyly of the genus *Endostelium* has been called into question (Spiegel 1990). The high stability of Hypothesis 5 enables us to suggest a novel Amoebozoa group, defined morphologically by the presence of an outer cuticle of fibrose or hyaline material. We suggest this group be named ‘Goceviidae’, the type genus and species should be *Gocevia fonbrunei* Pussard 1965 following the Principle of Priority. ’Goceviidae’ is strongly supported and often recovered within a larger clade designated Hypothesis 6, along with the genus *Cochliopodium*, consistent with the ‘Himatismenida’ *sensu* Page (Page 1987), with the added inclusion of *Endostelium zonatum* (see taxonomic summary for a formal account). However, support is low for Hypothesis 6

and there is a chance that *Cochliopodium* spp. are grouping here due to a long-branch attraction artifact.

Our observation of a clade uniting the Dictyosteliida and the protosporangiids is inconsistent with previously published works. We do not recover the previously proposed Stellamoebae (protosteliids + Dictyosteliida) within the Mycetozoa (Stellamoebae+Myxogastrea) *sensu* (Cavalier-Smith et al. 2004); nor the Macromycetozoa (Dictyostelidae+Myxogastrea) *sensu* (Fiore-Donno et al. 2010), also observed in (Nikolaev et al. 2006); neither the grouping with cavosteliids (Shadwick et al. 2009). However our analyses do not allow rejection of any of these hypotheses (Table 4). Given the moderate support for this clade, availability of equally likely alternative topologies, and lack of morphological features supporting any of these hypothesis, we suggest that the Dictyosteliida should receive an *incertae sedis* status.

Additionally, numerous taxa remain unplaced in our analyses: *Parvamoeba monoura*, *Stereomyxa ramosa*, *Dermamoeba algensis, Acramoeba dendroida*, *Multicilia marina, Phalansterium solitarium, Stygamoeba regulata*, ATCC® 50979, *Pessonella* sp., *Trichosphaerium* sp., *Vermistella antarctica* and *Mayorella* spp. are taxa with highly unstable relationships (Figure 3.4, Supplementary Figure S2). Morphological features of both *Vermistella antarctica* and *Stygamoeba regulata* would suggest these are closely related (Moran et al. 2007; Sawyer 1975; Smirnov 1996), but this relationship was not recovered (Table 3.3). However, AU testing does not reject a possible relationship (Table 3.4).

We hoped that including environmental sequences would increase resolution of the tree, a strategy previously adopted by several authors (Cavalier-Smith et al. 2004;

Fiore-Donno et al. 2010; Kudryavtsev et al. 2009b; Nikolaev et al. 2006; Smirnov et al. 2008; Wylezich et al. 2009). However, the environmental sequences at most only add to already stablished morphological groups, and fail to resolve deep branches. This is corroborated by the low increase in the Treeness index, coupled with non-significant improvement in the average Leaf Stabilities (Table 3.5). We conjecture that obtaining phylogenetically meaningful SSU-rDNA sequences for amoebozoans from environmental surveys is an unreasonable expectation, given current technologies for environmental sampling of molecular sequences. SSU-rDNAs in Amoebozoa are often very divergent, exhibiting over 2,000 base pairs, and reaching 3,000-4,000 bp in some taxa (e.g. *Pelomyxa*, *Trichosphaerium*, *Lindbladia*). Additionally, many of these exhibit unusual secondary structure features (Nikolaev et al. 2006). In our experience, many amoeboid taxa are not easily amenable to routinely used molecular techniques, even the model organism *Dictyostelium discoideum* requires special techniques for reliable DNA preparation (Charette and Cosson 2004). Key amoebozoan taxa likely have divergent SSU-rDNAs and will not be detected by current environmental sequencing methodology, but rather will need to be isolated and specifically targeted until better tools are developed for environmental sequencing. We provide in this paper two new methodologies that might simplify this task, by using single cell genome amplification as well as single cell cDNA extraction, while maintaining an acceptable morphological record through photodocumentation. These methods are superior to single cell PCR because they allow extraction of multiple genes from the same organism, crucial to the reconstruction of deep phylogenies.

What course of action should be taken to resolve the deep relationships within amoeboid organisms remains an open question. Our analyses demonstrate that single or few genes are not sufficient to unravel the relationships between deep groupings. Single gene analyses may however be enough to characterize relationships within well-supported lineages such as the Myxogastrea and Tubulinea. Morphological data are useful to establish synapomorphies among lower-level lineages, but also does not (at this point) help resolve the deeper relationships. Whether phylogenomic approaches (analyzing alignments of entire genomes) hold the key to resolve these ancient relationships remain to be seen. It is not clear as yet that such analyses actually result in more signal or yield strongly supported biased answers (Delsuc et al. 2005), another option may be using a selection of well chosen genes as in Parfrey *et al.* (Parfrey et al. 2010b). An additional important factor in unraveling the phylogenetic relationships within the Amoebozoa is comprehensive taxon sampling. The recognition that protosteloid amoebae are an integral part of the Amoebozoa (Brown et al. 2010; Shadwick et al. 2009) opens up many possibilities for exploring possible taxa with key phylogenetic positions, as suggested by the stabilization of three homeless amoeboid taxa (*Gocevia fonbrunei*, *Cochliopodium* sp. and *Grellamoeba robusta*) due to inclusion of protosteloid amoebae in our analyses (Figure 3.4, Table 3.3). This integration will most likely be useful not only in phylogeny, but also allow more meaningful studies on several aspects of Amoebozoa evolution, such as the evolution of the many diverse life cycle strategies (Lahr et al. 2011b).

4.6 Taxonomic summary of proposed hypotheses.

Remarks on nomenclature: At the time of writing of this report, there is no widely agreed upon consensus on how microbial eukaryote taxa should be named and treated. Some advocate a rankless approach while others continue to propose categorical ranks along with their taxon names. The International Code of Zoological Nomenclature, International Code for Botanical Nomenclature and the Bacterial Code do not assume direct responsibility for new microbial eukaryote names, they merely suggest ways to deal with names that were originally described under their provisions. We have taken a pluralistic approach with the aim to stabilize and make the taxa we propose available under many circumstances. We suggest taxa under categorical ranks, but those who wish to create a 94ankles taxonomy are welcome to ignore the proposed ranks, and be guided by the Hypotheses in Figure 3. Names are suggested in accordance with the ICZN: we provide diagnosis, etymology and name-bearing types. Additionally, we provide putative synapomorphies (where possible), which are not required by the ICZN.

Phylum Amoebozoa Luhe, 1913

Class Tubulinea Smirnov et al. 2005

Order Poseidonida ord. nov. Lahr and Katz 2011
Diagnosis: marine limax amoebae; small (5-20 um); pseudopods with a cylindrical or semi-cylindrical cross-section and monoaxial streaming.
Type species: *Nolandella hibernica* (Page 1980)
Etymology: in reference to the Greek god of the seas, Poseidon. All organisms in this group are marine, or capable of tolerating high-levels of salinity.
Putative Synapomorphy: marine limax Tubulinea.

Family Nolandellidae fam. nov. Lahr and Katz 2011
Included taxa: *Nolandella*; *'Hartmannella' abertawensis*.
Diagnosis: with characters of the order Poseidonida.
Type species: *Nolandella hibernica* (Page 1980)

Etymology: in direct reference to the type species.

Incertae sedis Amoebozoa

Order Fractovitellida ord. nov. Lahr and Katz 2011
Diagnosis: uninucleate amoebae without coloration, irregularly triangular with sharply pointed hyaline sub-pseudopodia, lobed nucleoli, and absence of a microtubular organizing center (MTOC).
Type species: *Soliformovum irregularis* (Olive and Stoianovich 1969) Spiegel 1994
Etymology: From the Latin fractus (broken) and vitellum (yolk), in reference to the appearance of the nucleoli. Also to acknowledge the etymology of the genus *Soliformovum*, which alludes to the resemblance of the pre-spore to a fried egg "sunny-side up" (Spiegel et al. 1994).
Putative Synapomorphy: presence of lobed nucleoli in at least one stage of the life-cycle.

Family Soliformoviidae fam. nov. Lahr and Katz 2011
Included taxa: *Soliformovum, Grellamoeba*
Diagnosis: with characters of the order Fractovitellidae.
Type species: *Soliformovum irregularis* (Olive and Stoianovich 1969) Spiegel 1994
Etymology: in direct reference to the type species.

Order Gracilipodida ord. nov. Lahr and Katz 2011
Included taxa: Flamellidae fam. nov. Lahr and Katz 2011, Filamoebidae Cavalier-Smith 2004
Diagnosis: gross morphological features outlined in Kudryavtsev 2009: flattened locomotive form either with expanded fan-shaped hyaloplasm regions producing thin sub-pseudopodia, or pseudopods coming out from cell body. Pseudopods are thin, filiform. Single or multinucleated.
Type species: *Flamella magnifica* Schaeffer 1926
Etymology: from the Latin gracilis (slender) and pedes (feet), in reference to the ability shared by these organisms to produce thin pseudopodia.
Putative Synapomorphy: filiform pseudopodia.

Family Flamellidae fam. nov. Lahr and Katz 2011
Included taxa: *Flammella, Arachnula* ATCC® 50593
Diagnosis: flattened, sometimes fan-shaped amoebae that can produce digitiform sub-pseudopodia from an anterior wide hyaloplasm margin, or can produce thin pseudopods from the cell body. Central, non-diffuse nucleolus in trophozoites.
Type species: *Flamella magnifica* Schaeffer 1926
Etymology: in direct reference to the type species, and most well described genus.

Order Himatismenida Page 1987 emend.
Diagnosis: amoebae with a locomotive lens-like shape, with an organic covering that does not enclose the cell completely, and may be organized in scales.
Type species: *Cochliopodium bilimbosum* Auerbach 1856

Putative Synapomorphy: an organic outer covering which does not completely seal the amoeba.

Family Cochliopodidae Hertwig and Lesser 1874 emend.
Included taxa: *Cochliopodium*
Diagnosis: himatismenid amoebae capable of producing an organic tectum composed of structured scales.
Type species: *Cochliopodium bilimbosum* Auerbach 1856
Putative Synapomorphy: structured scales composing the outer covering.

Family Goceviidae fam. nov. Lahr and Katz 2011
Included taxa: *Gocevia*, *Endostelium*
Diagnosis: himatismenid amoebae capable of producing non-organized outer cuticle, hyaline or granular.
Type species: *Gocevia fonbrunei* Pussard 1965
Etymology: in direct reference to the type species.
Putative Synapomorphy: an outer cuticle made of non-structured organic material.

Table 4.1: Newly characterized Amoebozoa lineages.

Organism	Source	SSU-rDNA	Actin genes
Cryptodifflugia operculata	commercial culture	JF694280	JF694297-305
Pyxidicula operculata	Hiddensee Germany	JF694284	JF694316-318
Arcella mitrata	Hawley Bog, MA	JF694279	JF694293-296
Arcella discoides	Hawley Bog, MA	-	JF694287-292
Arcella gibbosa	Hawley Bog, MA	JF694278	-
Hyalosphenia papilio	Hawley Bog, MA	JF694282	JF694311
Nebela carinata	Hawley Bog, MA	JF694283	JF694312
Gocevia fonbrunei	ATCC® 50196	JF694281	JF694306-310
Stereomyxa ramosa	ATCC® 50982	-	JF694320-321
Stygamoeba regulata	ATCC® 50892	JF694285	JF694322
'Thecamoeba' sp.	ATCC® 50185	JF694286	JF694323-326
Paraflabellula hoguae	ATCC® 30733	AF293899[a]	JF694313-315
CHINC-5 isolate[b]	ATCC® 50979	-	JF694319

[a] The SSU-rDNA for *Paraflabellula hoguae* ATCC® 30733 has been published previously (Amaral Zettler et al. 2000). We have obtained an identical sequence from our independently retrieved DNA.
[b] Morphological analysis confirms this isolate is not *Sexangularia*, mislabeled in the ATCC® collection.
Source indicates origin of the organism, GenBank numbers are listed for both SSU-rDNA and actin genes. Name in single quotes indicate that identification provided by ATCC® may be incorrect.

Table 4.2: Concatenated datasets generated to perform phylogenetic analyses.

Dataset name	Taxa #	Sites SSU-rDNA	Sites Actin	Removal of amb sites
A53	53	989	265	Automated
M53	53	1270	265	Manual
A101	101	989	265	Automated
M101	101	1270	265	Manual
A139	139	989	265	Automated
M139	139	1270	265	Manual
M139-7	139	1115	265	Manual
M139-76	139	1003	265	Manual
M139-765	139	860	265	Manual
A139-LB	129	1270	265	Automated
M139-LB	129	989	265	Manual
A139-us	129	1270	265	Automated
M139-us	129	989	265	Manual
A139-LB-us	120	1270	265	Automated
M139-LB-us	120	989	265	Manual
MEnv	164	1260	265	Manual

A list detailing which taxa were included in each reconstruction is available as Supplementary Table S1. Taxa # - number of taxa included in reconstruction; Sites – number of sites included in alignment for each of SSU-rDNA and actin genes; Removal of amb sites – method used for dealing with ambiguously aligned sites: Manual indicates that sites were hand curated, and Automated indicates usage of the GUIDANCE algorithm (Penn et al. 2010).

Table 4.3: Summary of bootstrap values obtained in all 18 reconstructions for previously proposed relationships and hypothesis suggested in the current report.

	A53	M53	A101	M101	A139	A139(B)	M139	M139(B)	M139-7	M139-76	M139-765	A139-LB	M139-LB	A139-us	M139-us	A139 -LB-us	M139-LB-us	MEnv
Higher-level hypot.																		
Amoebozoa	92	89	84	90	81	1	85	1	90	49	nm	85	93	78	94	84	93	87
Myxogastrea (FD)	-	-	94	96	93	1	96	1	94	97	95	-	-	97	97	-	-	95
Tubulinea (S)	81	74	53	nm	66	0.79	60	0.84	54	59	26	66	65	72	55	67	75	4
Archamoebae (CS)	40	nm	nm	54	nm	nm	53	0.81*	55	55	nm	37*	61	nm	49	53*	45*	54
Mycetozoa (CS)	51	67	nm	nm	nm	nm	nm	nm	nm	nm	nm	-	-	nm	nm	-	-	nm
Flabellinea (S)	-	-	nm	nm	nm	nm	nm	nm	6	nm	nm	nm	nm	19	21	21	24	nm
Conosea (CS)	nm	nm	nm	nm	nm	nm	nm	nm	nm	nm	nm	nm	nm	nm	nm	nm	nm	nm
Discosea (CS)	-	-	nm	nm	nm	nm	nm	nm	nm	nm	nm	nm	nm	nm	nm	nm	nm	nm
Dermamoebida (CS)	nm	nm	nm	nm	nm	nm	nm	nm	nm	nm	nm	nm	nm	nm	nm	nm	nm	nm
Stellamoebia (CS)	-	-	nm	nm	nm	nm	nm	nm	nm	nm	nm	nm	nm	nm	nm	nm	nm	nm
Variosea (CS)	nm	nm	nm	nm	nm	nm	nm	nm	nm	nm	nm	nm	nm	nm	nm	nm	nm	nm
Varipodida (CS)	-	-	nm	nm	nm	nm	nm	nm	nm	nm	nm	nm	nm	nm	nm	nm	nm	nm
Protamoebae (CS)	nm	nm	nm	nm	nm	nm	nm	nm	nm	nm	nm	nm	nm	nm	nm	nm	nm	nm
Proposed Hypotheses																		
H1 'Poseidonida'	-	-	100	100	100	1	100	1	100	100	100	100	100	100	100	100	100	100
H2 'Fractovitellida'	-	-	100	100	100	1	100	1	100	100	100	100	100	100	100	100	100	100
H3 'Flamellidae'	-	-	57	82	nm	nm	76	0.89	nm	64	77	54	76	nm	69	52	75	67
H4 'Gracilipodida'	-	-	nm	nm	43	1	34	0.83	46	19	nm	48	41	40	41	42	50	24
H5 'Goceviidae'	-	-	-	-	90	1	83	0.88	88	76	34	97	96	-	-	-	-	80
H6 'Himatismenida+'	-	-	nm	nm	59	1	35	0.51	nm	44	18	49	30	-	-	-	-	28
Morphogroups																		
Amoebidae	100	100	100	100	100	1	100	1	100	100	100	100	100	100	100	100	100	99
Hartmannellidae**	100	100	100	100	100	1	100	1	100	100	100	100	100	100	100	100	100	79
Dictiosteliida	100	100	96	96	97	1	97	1	96	96	98	97	97	100	100	100	100	96
protosporangiids	-	-	-	-	100	1	100	1	100	100	100	100	100	100	100	100	100	100
DS Myxogastrea	-	-	96	98	95	1	97	1	96	98	98	100	100	99	98	100	100	96
soliformoviids	-	-	-	-	100	1	100	1	100	99	97	100	100	100	100	100	100	100
Leptomyxida	100	100	94	96	98	1	96	1	96	92	87	97	95	99	97	99	96	96
schizoplasmodiids	-	-	-	-	100	1	100	1	99	97	96	100	100	99	100	100	100	97
Himatismenida	-	-	100	100	100	1	100	1	100	100	100	100	100	-	-	-	-	100
protosteliids	-	-	-	-	100	1	100	1	99	98	89	91	96	100	100	91	98	100
Tubulinida (Am+Hart)	100	100	80	69	75	1	82	1	86	92	nm	81	85	79	83	97	82	69
Dactylopodiida	-	-	-	-	97	1	92	0.99	80	47	36	45	90	99	98	99	97	92
Thecamoebidae	96	90	79	51	87	1	80	1	44	81	59	88	77	86	74	86	78	64
LS Myxogastrea	-	-	-	-	84	1	84	1	70	87	83	-	-	94	91	-	-	79
Echinamoeboidea	-	-	64	nm	73	0.96	77	0.97	83	70	44	75	81	75	80	80	84	70
Vannellida	100	100	99	99	68	1	54	0.99	42	36	nm	66	59	59	71	66	68	28
Centramoebida	-	-	58	42	77	1	73	1	33	30	nm	76	64	77	71	75	73	64
Mastigamoebidae	-	-	nm	39	nm	nm	66	0.58	71	77	59	nm	28	nm	64	nm	nm	59
Pelobiontidae	27	78	nm	36	nm	nm	58	0.91	71	78	13*	-	-	nm	56	-	-	56
cavosteliids	-	-	-	-	nm	nm	60	1	54	nm	nm	nm	37	nm	50	nm		51
Arcellinida	nm	nm	30	31	31	0.95	35	nm	27	nm	nm	33	32	37	36	36	34	2
Sty + Ver	-	-	nm	nm	nm	nm	nm	nm	nm	nm	nm	nm	nm	nm	nm	nm	nm	nm
Dic + pro	-	-	nm	54	nm	nm	69	1	62	47	45	nm	64	nm	78	nm	84	70

1. denotes that the group is invaded by one incertae sedis taxon;

** excluding *Saccamoeba limax* ATCC® 30942

All reconstructions performed on RaxML, except the two indicated by (B) on Mr. Bayes. Bootstrap and posterior probability values above 95 and 0.95 respectively are in bold. Notes: nm – non-monophyletic; - not enough taxa to test the group in reconstruction; DS Myxogastrea – dark spored myxogastrids; LS Myxogastrea – light spored myxogastrids; Am+Hart – Amoebidae + Hartmannellidae; *Sty + Ver* – *Stygamoeba + Vermistella*; Dic + pro – Dictyosteliida + protosporangiids; FD – taxon as defined in Fiore-Donno et al. (Fiore-Donno et al. 2010); S – taxon as defined in Smirnov et al. (Smirnov et al. 2005); CS – taxon as defined in Cavalier-Smith et al. (Cavalier-Smith et al. 2004).

Table 4.4: Summary of values obtained from approximately unbiased test.

Hypothesis	wkh	au	wsh
Conosa (CS)	0.153	0.185	0.632
Dermamoebida (CS)	0.354	0.482	0.893
Discosea (CS)	0.127	0.144	0.480
Flabellinea (S)	0.250	0.503	0.882
Glycosteliida (CS)	0.132	0.184	0.514
Macromycetozoa (FD)	0.254	0.450	0.806
Mycetozoa (CS)	0.130	0.250	0.669
Protamoeba (CS)	0.153	0.146	0.632
Stellamoeba (CS)	0.284	0.494	0.825
Variosea *sensu* (CS)	0.068	0.062	0.318
Varipodida *sensu* (CS)	0.254	0.423	0.794
Stygamoeba + *Vermistella*	0.253	0.387	0.743

Values are comparing our best phylogeny against phylogenies where proposed relationships were constrained. None of the hypothesis can be rejected, since all p values are above the 0.05 threshold. wkh – weighted Kishino-Hasegawa test; au – approximately unbiased test; wsh – weighted Shimodaira-Hasegawa test; FD – taxon as defined in Fiore-Donno et al. (Fiore-Donno et al. 2010); S – taxon as defined in Smirnov et al. (Smirnov et al. 2005); CS – taxon as defined in Cavalier-Smith et al. (Cavalier-Smith et al. 2004).

Table 4.5: Summary of tree indices obtained for 16 RAxML reconstructions.

Analysis	Tree Length	Treeness	LStability	95% CI
A53	9.47	0.35	0.84	0.02
M53	12.00	0.30	0.84	0.01
A101	21.66	0.36	0.82	0.02
M101	26.13	0.35	0.86	0.01
A139	31.48	0.40	0.80	0.01
M139	38.83	0.35	0.84	0.01
M139-7	21.05	0.34	0.86	0.01
M139-76	11.30	0.32	0.77	0.01
M139-765	7.78	0.32	0.73	0.01
A139-LB	26.77	0.45	0.80	0.01
M139-LB	30.73	0.41	0.85	0.01
A139-us	27.56	0.41	0.80	0.01
M139-us	34.36	0.37	0.85	0.01
A139-LB-us	24.14	0.45	0.83	0.01
M139-LB-us	27.76	0.41	0.88	0.01
MEnv	49.66	0.38	0.85	0.01

Tree length is the total length of the tree. Treeness index is the ratio of tree length that is in internal branches over the total tree length. Leaf Stability values are averaged over all taxa in 1000 boostrap reconstructions. The 95% Confidence Interval refers to Leaf Stability values.

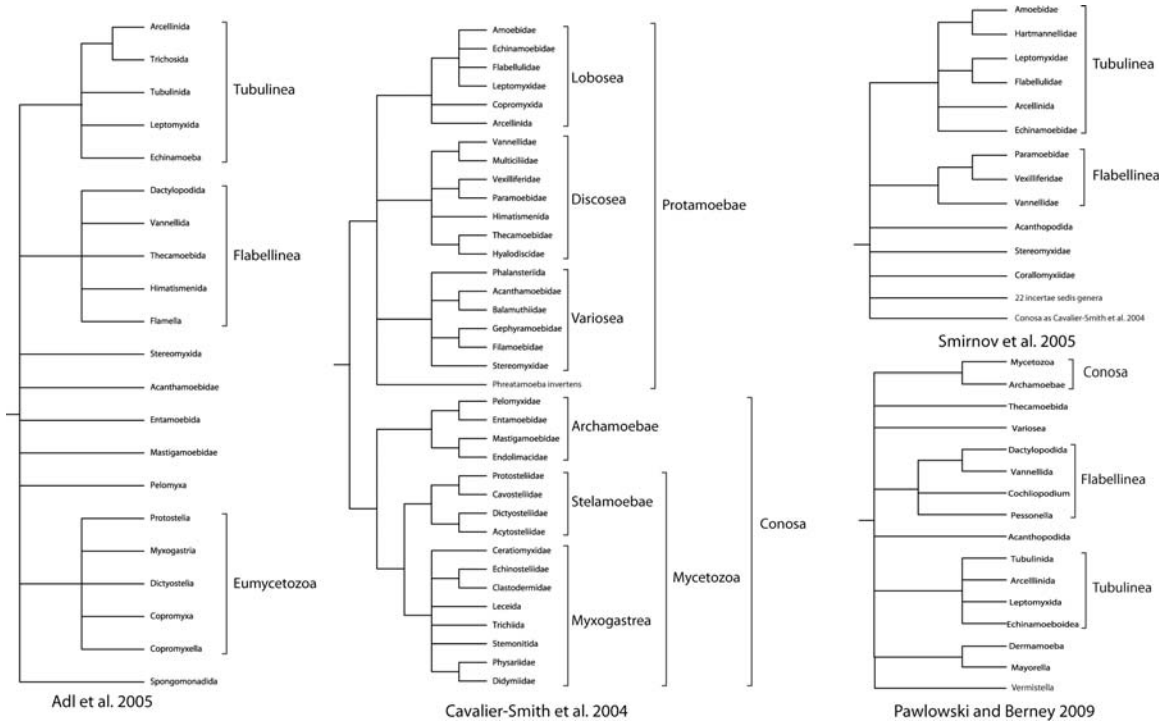

Figure 4.1: A summary of previously proposed relationships between the Amoebozoa.

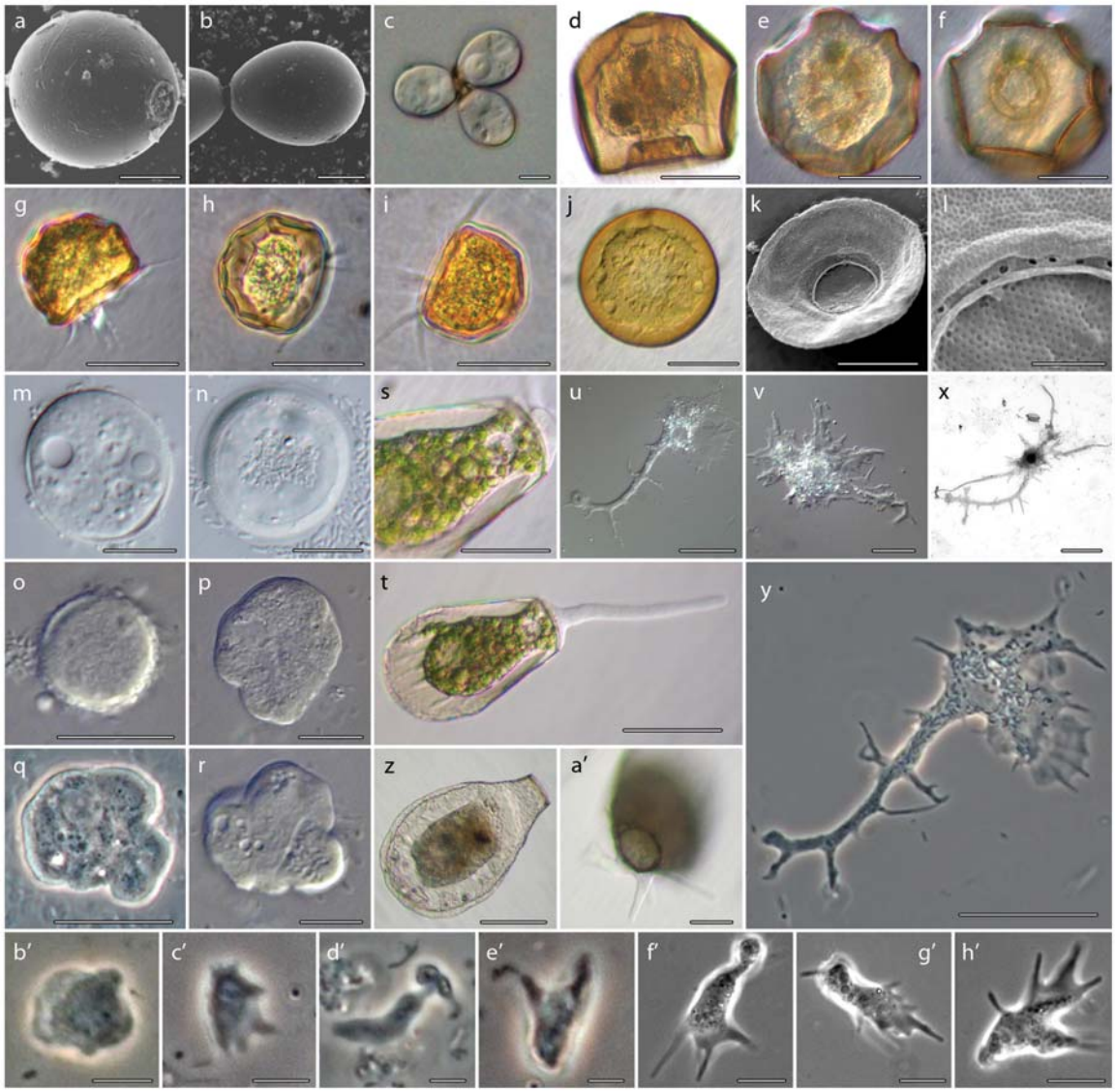

Figure 4.2: Morphology of the amoeboid lineages isolated for this study.

1a-c. *Cryptodifflugia operculata*: a) Scanning electron micrograph (SEM) of *C. operculata* in ventral view, showing the distinctive mucous operculum covering the aperture; b) Dorsal view of two *C. operculata* with a cytoplasmic connection, this state is often seen in culture; c) Differential interference contrast images (DIC) of 3 connected *C. operculata* individuals. Scale bars are 5 μm. 1d-f. Light microscopy images of the *Arcella mitrata* individual that was genome amplified to generate the sequences used in this study: d) lateral view showing the typical polygonal profile; e) top view of the same individual, focal plane at the middle of test height; f) top view of the same individual, focal plane at bottom of test height, showing the characteristic rippled apertural margin. Scale bars are 100 μm. 1g-i. Hoffman interference contrast (HIC) images of cultured individuals of *Arcella gibbosa*: g) lateral view showing hemispherical profile and pseudopods; h) another individual showing the shell's ridges and depressions; i) lateral view of a third individual. Scale bars are 60 μm. 1j-l. *Arcella discoides*: j) HIC image of

a cultured individual; k) SEM image showing the thin lateral profile; l) close-up on the apertural margin of individual in k, showing pores surrounding the aperture. Scale bars for j, k are 30 μm, for l 3 μm. 1m-n. DIC images of cultured *Pyxidicula operculata*: m) focal plane at middle of test height showing the nucleus and one contractile vacuole; n) focal plane at the bottom of a different individual, surrounded by bacteria on which it was feeding. Scale bars are 10 μm. 1o-r. DIC images of '*Govecia fonbrunei*' ATCC® 50196: o) Encysted individual; p) resting individual, note the hyaline covering visible at the top margin; q) individual shape immediately after excystation; r) initial stages of locomotion. Scale bars are 10 μm. 1s-t. HIC images of *Hyalosphenia papilio*: s) close up on one of the individuals that was genome amplified to obtain sequences in this study, scale bar 30 μm; t) a more general view of the same individual, scale bar 50 μm. 1u-y. Images of '*Stereomyxa ramosa*' ATCC® 50982: u,v) Phase contrast images of a cultured individual; x) protargol staining, showing the single nucleus; y) DIC image of a '*S. ramosa*' showing the variety of pseudopods it can produce. Scale bars are 20 μm. 1z-a'. HIC images of *Nebela carinata*: z) a lateral profile of one of the individuals used to obtain sequences in this study, this image shows the characteristic rim around the margin of the shell; a') same individual observed in the typical raised shell locomotive position. Scale bars are 20 μm. 1b'-e'. '*Stygamoeba regulata*' ATCC® 50892: b') sedentary shape; c') beginning of movement morphology; d') start of monopodial movement; e') polypodial movement. Scale bars are 5 μm. 1f'-h'. Three images of isolate CHINC-5 ATCC® 50979 (misidentified as *Sexangularia*) showing locomotive form. The absence of a shell, among other significant characters, indicates the identification as *Sexangularia* is incorrect. Note the finger-like pseudopods, similar to dactylopodids. Scale bars are 10 μm. Images of ATCC® isolates were generated by Jeffrey Cole and kindly provided by Robert Molestina, director of ATCC® collections, except for images on isolate CHINC-5 ATCC® 50979 provided by O. Roger Anderson.

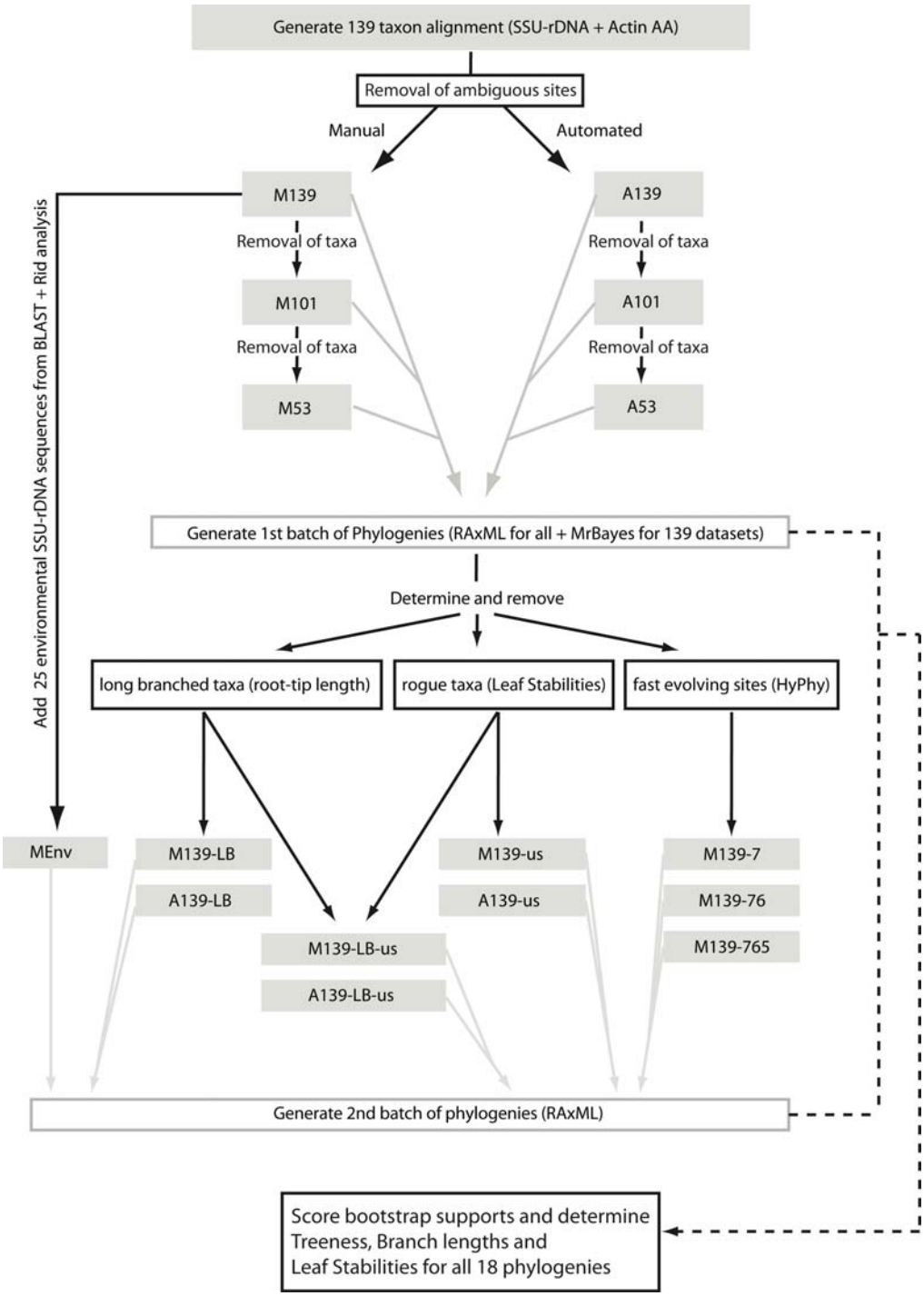

Figure 4.3: Computational pipeline implemented for phylogenetic analysis.

Grey boxes indicate a dataset, grey arrows indicate phylogenetic analyses performed on that dataset. Black arrows and boxes indicate other types of analyses performed on particular datasets, and the black dotted lines indicate the final analyses performed to obtain scores for each phylogenetic reconstruction.

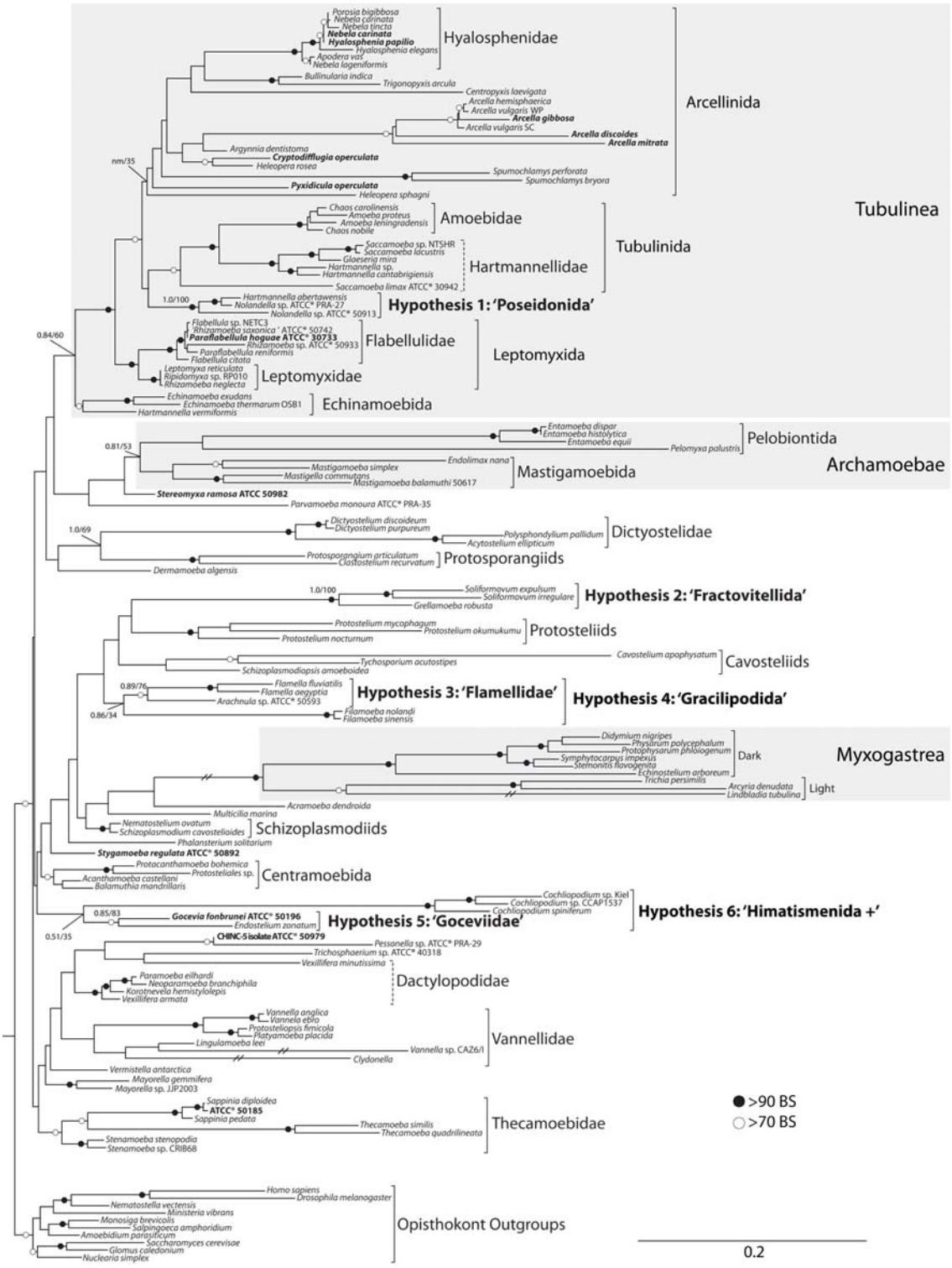

Figure 4.4: Phylogenetic reconstruction of the Amoebozoa, based on concatenated analysis of SSU-rDNA and actin genes of 139 lineages.

This reconstruction is the best maximum likelihood tree obtained from the dataset Manual139, which we consider exhibits the optimal combination of tree indices and taxonomic coverage. Both Bayesian posterior probabilities and bootstrap supports are

plotted on branches of interest. Branches without any support indication had bootstrap support of less than 70. The three well-supported higher-level groupings are shaded gray. The lower-level, morphologically consistent relationships are indicated. The novel relationships uncovered in the current study are in bold, and the suggested name for the group is shown in single quotes. Terminals in bold indicate lineages for which we are providing novel molecular information. Dashed brackets represent lower-level groups that are morphologically consistent but not recovered in this reconstruction. All branches are drawn to scale, except the branches leading to Myxomycetes, *Lindbladia*, *Vannella* CAZ6/I and *Clydonnella* which were trimmed to half-length for display purposes.

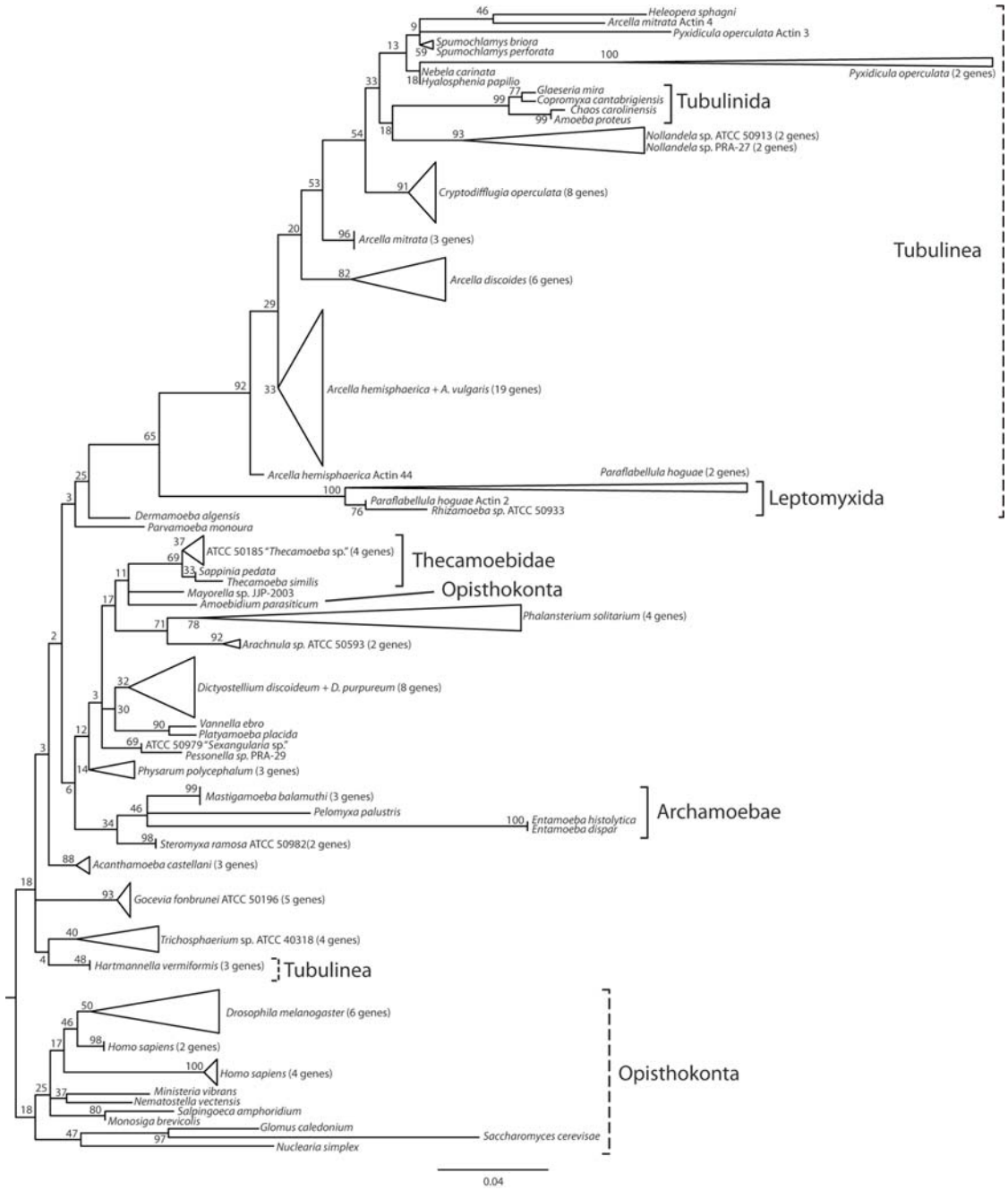

Figure 4.5: Reconstruction of actin gene family evolution in Amoebozoa, using 140 paralogs.

Triangles indicate multiple paralogs (number indicated in parenthesis), the length of triangle is equal to the length of longest branching paralog within the group.

CHAPTER 5

INTERPRETING THE EVOLUTIONARY HISTORY OF THE TUBULINEA (AMOEBOZOA), IN LIGHT OF A MULTIGENE PHYLOGENY

5.1 Abstract

The Tubulinea are a higher-level group of amoeboid organisms characterized by monoaxially streaming cylindrical pseudopods. They emerge in molecular phylogenetic reconstructions as one of the few stable, highly supported groups within Amoebozoa. However, contemporary reconstructions have largely relied on SSU-rDNA, and to a lesser extent, actin genes, to reveal the relationships among these organisms. Additionally, one of the most species rich amoebozoan groups is nested within Tubulinea, the test (shell) forming Arcellinida, still suffers from substantial undersampling of taxa. Here, we aim to increase both taxonomic and gene sampling within the Tubulinea, by characterizing novel molecular data for 21 taxa and 6 genes (SSU-rDNA, actin, α and β tubulin, elongation factor 2α and the regulatory 14-3-3). We perform concatenated phylogenetic analyzes using these genes and assess alternative hypothesis of relationships within the Tubulinea using approximately unbiased tests. We confirm the monophyly of Tubulinea and five of the six included lineages (Amoebida, Arcellinida, Echinamoeboidea, Leptomyxida, and Poseidonida). We show instances of non-monophyly for well-defined morphological groups at various hierarchical levels. Most strikingly, relationships within Arcellinida seem to be more consistent with general test and aperture (opening) shape than on test composition as previously proposed. Our

multigene analyses yield two large groups with high support: the Nebelidae/Hyalosphenidae (with inclusion of *Quadrulella*) that have an elongate and flattened shell with an ellipsoid aperture; and a novel grouping of *Arcella* and *Netzelia*, both of which have a more rounded shell with circular aperture. The group composed of silica biomineralizers (Lesquereusiidae) is non-monophyletic, indicating multiple origins of silica biomineralization within the Arcellinida. We also discuss the implications of this phylogeny for interpretations of the Precambrian fossil record of testate amoebae.

5.2 Introduction

The Tubulinea are a monophyletic lineage within the Amoebozoa (Smirnov et al. 2005). Unlike many other eukaryotic groups proposed in recent years, the clade Tubulinea has a defining morphological character, or synapomorphy: monoaxial streaming of cytoplasm within a roughly cylindric pseudopod (Smirnov et al. 2005). Some organisms in this group can produce several cylindrical pseudopods, as in the genus *Amoeba*, while others have a single semi-cylindrical pseudopodial protrusion that comprises the entire body, giving them a slug-like (limax) shape as in the genus *Saccamoeba*. Tubulinea emerged from phylogenetic reconstructions based on SSU-rDNA (Amaral Zettler et al. 2000; Bolivar et al. 2001), and has been corroborated subsequently in multiple analyses (Cavalier-Smith et al. 2004; Fahrni et al. 2003). This assemblage of organisms has been referred in some instances as Lobosea (Bolivar et al. 2001) or Gymnamoebia (Amaral Zettler et al. 2000). However, given the historical instability of these alternative names (Pawlowski 2008), we follow the classification of Smirnov et al. (2011) for clarity.

The Tubulinea are composed of six major groups with defining morphological characteristics and varying levels of support in molecular reconstructions: Echinamoeboidea, Leptomyxida, Amoebidae, Hartmannellidae, Poseidonida and Arcellinida. The Tubulinida, Echinamoebida, Leptomyxida and the recently described Poseidonida are recovered with high levels of support in most studies (see Lahr et al. 2011a). The Arcellinida are the most species-rich assemblage and are characterized by the ability to secrete or agglutinate a test (shell). In molecular reconstructions, the Arcellinida are either not monophyletic or exhibit low to moderate levels of support, with more comprehensive taxon sampling tending to decrease support, (Lahr et al. 2011a). Finally, a core group of organisms within the Hartmannellidae are often recovered with high levels of support, but with the lineage *Saccamoeba limax* ATCC 30942 often falling outside of the main group, rendering the Hartmannellidae paraphyletic (Bolivar et al. 2001; Cavalier-Smith et al. 2004; Fahrni et al. 2003; Lahr et al. 2011a; Pawlowski and Burki 2009; Smirnov et al. 2005; Tekle et al. 2008).

Taxonomic instability extends further to impact genera within the Tubulinea. The genus *Hartmannella* (and Family Hartmannellidae by consequence) is probably one of the most affected by recent molecular reconstructions. Many small (10-30um) amoebae that present a limax-like locomotive form were described as different species in the genus *Hartmannella* (Page 1987). Based on morphological evidence, several species were removed from the genus (eg. *Nolandella* (Page 1983) and *Echinamoeba* (Page 1975)). More recently, molecular studies showed that *Hartmannella vermiformis* (now transferred to *Vermamoeba* (Smirnov et al. 2011)), a common freshwater amoeba, is in fact more closely related to *Echinamoeba* than to other limax-shaped amoebae now

considered “core-hartmannellids” (*Glaeseria, Saccamoeba*; (Amaral Zettler et al. 2000; Fahrni et al. 2003). Further, marine species of *Nolandella* and *H. abertawensis* were shown to form the highly-supported Poseidonida, a distinct lineage from other hartmannellids (Lahr et al. 2011a). Even more surprisingly, the sorocarpic (fruiting body producing) slime mold *Copromyxa protea* is very closely related to *H. cantabrigiensis*, prompting transfer of the latter to the genus *Copromyxa* (Brown et al. 2010).

Taxonomic instability is also seen within the Arcellinida. These are conspicuous and abundant amoebae that build distinctive shells (tests), which have been argued to be valuable structures for both species delimitations and phylogenetic inferences (Meisterfeld 2002). In this group, molecular evidence does not corroborate morphological predictions in three significant and distinct instances: 1) some genera appear not to be monophyletic, including *Heleopera* and *Nebela* (Lara et al. 2008; Nikolaev et al. 2005); 2) relationships proposed based on shell form and composition are not recovered. Genera (e.g. *Pyxidicula*, *Arcella* and *Spumochlamys*) within the Suborder Arcellina, which is defined based on the possession of a organic membranous shell, are not monophyletic (Lahr et al. 2011a), and finally; 3) at the most inclusive level, increased taxonomic sampling results in reduction of support for the entire group, opening up a possibility of non-monophyly for the Arcellinida (Lahr et al. 2011a). However, taxonomic sampling is still far from comprehensive in this species-rich group, making difficult to evaluate these taxonomic instabilities.

A further limitation of previous work is that phylogenetic inference in the Tubulinea has relied mostly on SSU-rDNA and to a lesser extent on actin genes. The problems associated with single gene reconstructions are well known and have been

extensively dealt with elsewhere (eg. Baldauf et al. 2000; Philippe and Douady 2003). The actin gene family in Amoebozoa poses challenges as the high levels of paralogy present in many members of the group complicates phylogenetic reconstruction (Lahr et al. 2010). Here, we present a phylogenetic reconstruction of the Tubulinea that capitalizes on sampling of SSU-rDNA, actin and an additional 4 protein-coding genes (α and β tubulin, elongation factor 2α and the regulatory protein 14-3-3). We provide molecular data for 21 taxa from all six groups contained in the Tubulinea, with greatest emphasis on the diverse Arcellinida (15 taxa), adding a total of 112 gene sequences. We perform phylogenetic reconstructions including a representative sample of eukaryotes to test monophyly at higher taxonomic levels, as well as specific hypotheses of evolution within the Tubulinea.

5.3 Methods

5.3.1 Taxon sampling

Amoebae were obtained by two methods: 1) culturing of newly isolated or deposited strains and 2) isolation, photo-documentation and genome amplification or construction of cDNA libraries of individuals or small groups of freshly isolated organisms (Table 1, Fig. 1). *Arcella hemisphaerica*, *Cryptodifflugia operculata* and *Hartmannella vermiformis* were isolated and cultured as previously described (Lahr et al. 2010; Lahr and Katz 2009). *Chaos carolinensis* (Cat. no 131324), *Amoeba proteus* (Cat. No 131306) and *Lesquereusia spiralis* (Cat. no 131334, listed as *Difflugia*) cultures were obtained from Carolina Biological Supply. These amoebae were cleaned by multiple transfers of sterilized pond water and allowed to sit overnight to finish digestion of prey

organisms before being subjected to cDNA construction. *Arcella gibbosa, Difflugia bryophila, Difflugia lanceolata* and *Difflugia* sp., *Heleopera sphagni, Hyalosphenia papilio, Lesquereusia modesta, Nebela carinata, Nebela penardiana, Netzelia wailesi, Netzelia tuberculata* and *Quadrulella* were isolated from natural sources (details in Table 1), cleaned through successive transfers, photodocumented and then subjected to genome amplification and/or construction of cDNA libraries. *Saccamoeba lacustris* CCAP 1572/4, *Rhizamoeba saxonica* CCAP 1570/2 and *Nolandella hibernica* CCAP 1534/10 were obtained from the Culture Collection of Algae and Protozoa. These cultures were grown according to instructions from the repository, and large numbers of amoebae were harvested for cDNA construction.

5.3.2 DNA and cDNA isolation

Genetic material was obtained by three methods: 1) genomic extraction; 2) genomic amplification and 3) construction of cDNA. For genomic extraction (gEXT), cultures were grown either in liquid media or agar plates as previously described (Lahr et al. 2011a), amoebae were harvested and cleaned either through several washes or by filtering, and subjected to a standard phenol/chloroform protocol (Lahr et al. 2011a). For genomic amplification (gAMP), one or a small group of organisms was isolated, cleaned through washes in sterile water, left overnight to finish digestion of prey organisms, and subjected to amplification using a Repli-g Genomic amplification kit (Qiagen, Cat. No. 150023) following manufacturer's directions. The same strategy for isolation and cleaning of organisms was adopted for construction of cDNA libraries, but in the final step organisms were subjected to lysis and cDNA contruction protocol through a

SuperScript III Cells Direct kit (Invitrogen, Cat. No. 18080-200), following manufacturer's instructions.

5.3.3 Amplification of target genes, cloning and sequencing

We performed amplification of genes of interest using Phusion Hot Start DNA polymerase (New England BioLabs, Cat. no. F540), following manufacturer's instructions. We targeted 6 genes using a variety of previously described and novel primers (Table 2). In general, reactions were performed on serial dilutions of starting material (1x, 1:10x, 1:100x, 1:500x) to determine the lowest amount of starting DNA necessary for amplification, in an attempt to minimize the formation of chimeras as recommended in (Lahr and Katz 2009). Successfully amplified products were then gel isolated using the Millipore Ultra Free DA spin column, and cloned using the Zero Blunt TOPO cloning kits (Invitrogen, Cat. No. K280020) according to manufacturer's instructions. Colonies were then screened by PCR and positive colonies were sequenced in a ABI3100 sequencer (Applied Biosystems, Foster City, CA, USA) at the Smith College Center for Molecular Biology.

5.3.4 Analytical methods

With the resulting set of 112 new sequences (Table 1), we reconstructed the genealogy of each gene independently to determine possible ancient paralogy and chose the most appropriate paralogs to be concatenated. Taxon sampling for Amoebozoa is identical to the dataset used in Lahr et al. (2011), with the addition of the taxa sampled in the current study. For ougroups we used a dataset of representative eukaryotic organisms

proposed by Parfrey et al. (2010b), the dataset (named 10-16) is available for download at Treebase (www.treebase.org). For each protein coding gene, alignments were constructed in SeaView (Galtier et al. 1996; Gouy et al. 2010) with alignment algorithm MAFFT (Katoh et al. 2009) using the L-INS-I setting. Alignments were then subjected to automated removal of ambiguously aligned sites using the software GUIDANCE (Penn et al. 2010). We performed maximum likelihood phylogenetic reconstruction for each gene using RAXML HPC 7.2.7 (Stamatakis 2006; Stamatakis et al. 2008) as implemented in the online server CIPRES (Miller et al. 2009). We ran a 100-replicate bootstrap analysis using the GTRCAT approximation followed by a slow maximum likelihood search using the GTRGAMMA model for the SSU-rDNA partition and the LG model with gamma distribution of site heterogeneity for the protein partition. The most appropriate model for amino-acid evolution was determined using model testing implemented both in the software ProtTest 3.0 (Darriba et al. 2011) and the online server Datamonkey (Delport et al. 2010), which gave similar results. Bootstrap values of the GTRCAT search were then plotted on the best tree found by maximum likelihood search for comparative analysis.

Each gene genealogy was analyzed to determine which paralogs should be used for concatenation. In most cases, there was no indication of ancient paralogy so we chose the shortest branching paralog for concatenation (Table 3). In the few cases where duplication predated species divergence, we took care to choose orthologous genes (Table 3). We concatenated all six genes into one file and performed the analyses using RAXML HPC 7.2.7 with two partitions: one for the SSU-rDNA gene and one large partition with all 5 protein coding genes and LG model of substitution with gamma

distribution of site heterogeneity, as determined by the software ProtTest 3.0 (Abascal et al. 2005; Darriba et al. 2011). The analysis consisted of 250 rapid bootstraps followed by a slow maximum likelihood search. We have also performed a slower, more accurate search in RAXML HPC 7.2.7 consisting of 100 multiparametric bootstraps using GTRGAMMA (as opposed to rapid bootstraps based on the GTRCAT approximation), followed by 25 maximum likelihood searches, each starting from an independent maximum parsimony starting tree. The resulting topology from this slower, more accurate approach was identical to the topology obtained from the faster approach, with minimal increases in bootstrap values. The faster approach is at least one order of magnitude less time consuming and less computationally intensive. We present results from the faster approach, because these are more logically comparable to the 33 constrained reconstructions we generated for the approximately unbiased test (AU, see below).

With the results from the unconstrained reconstruction at hand, we designed constraints to several proposed groups as well as non-monophyletic groups to be tested by the approximately unbiased (AU) test (Shimodaira 2002; Shimodaira 2004). The AU test provides a statistical measure whether the current dataset is able to reject the monophyly of specifically constrained groups. We tested 36 relationship hypotheses that were not monophyletic in the most likely tree. For each, we generated constrained maximum likelihood reconstructions. Parameters for tree searching in RAxML HPC 7.2.7 were identical to the standard reconstruction (here the advantages of a less computationally intensive approach become critical). These trees were then compared to the best tree found on the standard analysis using RAxML to calculate per-site

likelihoods. The per-site likelihoods were then analyzed in CONSEL (Shimodaira and Hasegawa 2001) with standard parameters to obtain p-values. We use a conservative α of 0.05 for rejection of hypothesis.

5.4 Results

5.4.1 Genes characterized

We characterized 112 sequences from six genes, representing a total of 21 taxa (Table 1): 11 SSU-rDNAs for 8 taxa; 44 actin genes from 17 taxa; a total of 18 a-tubulin genes for 14 taxa; 10 b-tubulin genes for 8 taxa; 9 elongation factor 2-a genes for 9 taxa and; 20 regulatory 14-3-3 genes for 13 taxa. For the SSU-rDNA, both *Lesquereusia spiralis* and *Heleopera sphagni* yielded multiple sequences: the 2 SSU-rDNAs for *H. sphagni* are identical except that one contains a group I intron; two of the three *L. spiralis* SSU-rDNAs are very similar (0.6% divergence) and a third one is more divergent with average 2.4% divergence from the other two. As the DNA extraction for both taxa was performed from a pool of individuals, the yield of multiple divergent SSU-rDNAs suggests intra-population variation.

We found varying levels of paralogy in protein-coding genes. There is extensive paralogy of actin genes as expected based on previous work on the genus *Arcella* (Lahr et al. 2010), with 11 out of the 17 sampled taxa containing duplicated genes. For α-tubulin the taxa *Difflugia sp.* and *Quadrulella symmetrica* contained paralogs; for β-tubulin the taxa *Difflugia sp.* and *Chaos carolinensis* contained paralogs, for 14-3-3 the taxa *Difflugia* sp., *Hyalosphenia papilio*, *Nebela penardiana* and *Netzelia wailesi* had paralogs (Table 1). We found no indication of paralogy for the gene EF2α.

5.4.2 Single gene trees results

We performed phylogenetic reconstructions on each of the genes sampled, with the objective of looking for cases of ancient gene duplication events so we could choose appropriate genes for concatenation. These single-gene trees are generally not efficient in reconstructing deep relationships, and the variable taxon sampling for each gene makes comparisons difficult (Figure 2). In most cases, paralogy seems to occur independently at shallow levels, without evidence for ancient duplications. For both actin and 14-3-3, there is evidence of duplication events that predate the divergence of genera within the Nebelidae and so we used the single-gene topologies to choose putative orthologs here for concatentation. In other cases where multiple paralogs for an isolate were monophyletic, we chose the shortest branching paralog.

5.4.3 Concatenated trees results

5.4.3.1 General topology

The topology obtained from concatenated analyses (Figure 3) is largely congruent with comprehensive eukaryotic analyses (Hampl et al. 2009; Parfrey et al. 2010b; Yoon et al. 2008) and Amoebozoa specific reconstructions (Lahr et al. 2011a; Shadwick et al. 2009; Smirnov et al. 2005; Tekle et al. 2008). The Tubulinea appear monophyletic with low support (31% BS, Figure 3); four of the six major included lineages are moderately to highly supported (Figure 3b): Amoebida (100% BS), Echinamoeboidea (79% BS), Leptomyxida (98% BS), and Poseidonida (100% BS); while the Arcellinida are poorly supported (27% BS). The remaining lineage (Hartmannelidae) is non-monophyletic due to a single taxon (*Saccamoeba limax*, Figure 3) falling outside a highly supported core

group containing *Saccamoeba lacustris*, *Glaeseria mira* and *Copromyxa* spp. (100% BS, Figure 3). An approximately unbiased test (AU) cannot reject the possibility that Hartmannellidae *sensu strictu* (including *S. limax* ATCC 30942) is monophyletic (Table 4).

5.4.3.2 Topology of the major Tubulinea lineages

The internal topology of major Tubulinea lineages (Figure 3) is generally concordant with morphological observations as well as previous phylogenetic reconstructions, with a few exceptions detailed below. Within the Echinamoeboidea, the genus *Echinamoeba* is monophyletic (98% BS) and our newly isolated *Vermamoeba vermiformis* SC groups with the other available *V. vermiformis* strain (100% BS). The Echinamoeboidea is not only monophyletic (Figure 3b), but also an the AU test rejects the possibility of its sister grouping with any other major Tubulinea lineage (Table 4, except with Leptomyxida in one out of three tests).

The topology of Leptomyxa is generally concordant with previous phylogenetic reconstructions, except for positioning of the isolate *Rhizamoeba saxonica* CCAP 1570/2 characterized here. We recover two highly supported groups within the Leptomyxida (Figure 3). The isolate *Rhizamoeba saxonica* CCAP 1570/2, considered as the most morphologically accurate representative of the *Rhizamoeba* genus (Smirnov et al. 2008), falls within a highly supported group sister to representatives of *Paraflabellula*, *Flabellula* and the isolate *'Rhizamoeba'* sp. ATCC 50933 (Figure 3, BS 100%). This result contrasts with previous reconstructions using only SSU-rDNA, where *R. saxonica* falls sister to the Leptomyxida (Dykova et al. 2008a; Smirnov et al. 2008), or sister to the

group comprising *Leptomyxa reticulata*, *Rhizamoeba neglecta* and two strains identified as *Ripidomyxa* sp. (Smirnov et al. 2009). The second group within the Leptomyxida contains the isolate *Leptomyxa reticulata* ATCC 50242; one *Ripidomyxa* sp. isolate (RP-010) as well as *Rhizamoeba neglecta*, consistent with the reconstruction in (Smirnov et al. 2009).

The Hartmannellidae topology recovered here is congruent with previous reconstructions as the strain *Saccamoeba limax* ATCC 30942 falls outside a well supported group of "core hartmannelids", comprising *Copromyxa cantabrigiensis*, *Saccamoeba lacustris* CCAP 1572/4 and *Glaeseria mira* (Figure 3b). This result is consistent with the majority of previous reconstructions (Amaral Zettler et al. 2000; Bolivar et al. 2001; Brown et al. 2010; Cavalier-Smith et al. 2004; Corsaro et al. 2010; Fahrni et al. 2003; Shadwick et al. 2009; Tekle et al. 2008). However, the AU test does not reject the possibility that the Hartmannellidae *sensu strictu* (i.e. including *S. limax*) is monophyletic (Table 4).

The Poseidonida appear monophyletic and strongly supported (100% BS, Figure 3b), with the addition of a partial SSU-rDNA sequence for the isolate *Nolandella hibernica* CCAP 1534/10. This is the strain used in the original description of the species, though the original designation was *Hartmannella hibernica* in Page (1980) and then transferred to *Nolandella hibernica* in Page (1983). This result validates taxonomically the Family Nolandellidae Lahr & Katz 2011 and Order Poseidonida Lahr & Katz 2011, since the type strain is now shown to nest within the previously characterized lineages. The AU test shows Poseidonida is likely not included within any

other major Tubulinea lineage, as monophyly can be rejected with all but Arcellinida (Table 4).

The Amoebidae are monophyletic (100% BS, Figure 3). The SSU-rDNA sequence for the isolate of *Chaos carolinense* presented here is identical to the previously-characterized SSU-rDNA (GB# AJ314607). Two moderately supported groups emerge within the Amoebidae, corresponding to the genera *Chaos* (55% BS) and *Amoeba* (BS 67%), a result contradictory to previous reconstructions where lineages of *Amoeba* and *Chaos* interdigitate (Smirnov et al. 2005), but consistent with the results of Fahrni et al. (2003) where both genera also appear monophyletic.

The Arcellinida are monophyletic, albeit with a low bootstrap support (27%). Despite the low value, monophyletic Arcellinida were recovered multiple times with widely varying taxon sampling (Kudryavtsev et al. 2009a; Lahr et al. 2010; Lahr et al. 2011a; Lara et al. 2008; Nikolaev et al. 2005). Two groups within the Arcellinida show high support: a group uniting *Netzelia* and *Arcella* (80% BS, Figure 3) and a group uniting the Hyalosphenidae and Nebelidae (100% BS, Figure 3). These two highly supported groups are in disagreement with the morphologically based classification of Meisterfeld (2002), where the Hyalosphenidae and Nebelidae are independent lineages, and *Arcella* and *Netzelia* fall within distinct clades due to differences in shell composition.

Three previously proposed groups within the Arcellinida are not recovered: the Suborders Arcellina and Difflugina; and the Family Lesquereusiidae. The Arcellina comprise amoebae capable of producing organic membranous or chitinoid shells, and are represented in the current sampling by the genera *Arcella*, *Pyxidicula* and *Spumochlamys*.

The group appears polyphyletic (Figure 3): *Arcella* is in a well-supported clade with *Netzelia* (80% BS), *Spumochlamys* is in a poorly-supported clade (41% BS) with *Difflugia*, and *Pyxidicula* appears at the base of the Arcellinida clade. Monophyly of the Arcellina can be marginally rejected by AU tests (Table 4), while combinations of any two taxa within cannot be rejected with the exception of *Arcella+Pyxidicula* which can be marginally rejected. The group Difflugina comprise the majority of Arcellinida, uniting amoebae that construct the shell by agglutination and are represented in the present survey with members of 9 out of 11 putative included families (Heleoperidae, Hyalospheniidae, Difflugiidae, Nebelidae, Lesquereusiidae, Paraquadrulidae, Centropyxidae, Plagiopyxidae, Trigonopyxidae). The Suborder is non-monophyletic in our reconstruction (Figure 3), and monophyly of the group can be rejected by the AU test (Table 4). The Lesquereusiidae, defined as the Arcellinida capable of biomineralizing silica (Ogden 1979), originally included the genera *Lesquereusia*, *Quadrullela* and *Netzelia*, with the later additions of *Microquadrulla* and the marine *Pomoriella* (Meisterfeld 2002). This group is not monophyletic in our multigene reconstructions: *Quadrulella* appears within the Nebelidae, *Lesquereusia* is sister to a poorly-supported *Difflugia+Spumochlamys* clade, and *Netzelia* is in a well-supported position sister to the genus *Arcella*. Additionally, AU tests reject the possibility that Lesquereusiidae (*Lesquereusia+Quadrulella+Netzelia*) is monophyletic (Table 4). However, the monophyly of *Netzelia+Lesquereusia* is rejected the two least conservative statistical tests (Table 4). The remaining taxon, *Quadrulella* is nested within the Hyalosphenidae, a result confirmed by studies on cytochrome oxidase 1 (Kosakyan et al. Submitted), and

the fact that its monophyly with either of the other two taxa can be rejected by AU tests (Table 4).

Monophyly of genera within the Arcellinida is variable: while the genera *Arcella* and *Spumochlamys* are monophyletic (80% and 100% BS respectively, Figure 3), the other three genera represented by more than one species are non-monophyletic: *Heleopera*, *Hyalosphenia* and *Nebela* (Figure 3). However the AU test cannot reject the monophyly of any of these three genera (Table 4).

5.5 Discussion

The addition of taxa combined with larger gene sampling reveals a phylogeny that is generally consistent with hypotheses on the six principal Tubulinea lineages (Figure 3), albeit with low resolution at deep nodes. The monophyly of Echinamoebidae, Leptomyxida, Poseidonida, "Hartmannellidae" (excluding *Saccamoeba limax*) and Amoebidae that were previously recovered in numerous SSU-rDNA and actin gene reconstructions (Amaral Zettler et al. 2000; Brown et al. 2010; Cavalier-Smith et al. 2004; Corsaro et al. 2010; Dykova et al. 2008a; Dykova et al. 2008b; Fahrni et al. 2003; Kudryavtsev et al. 2009a; Lahr et al. 2010; Lahr et al. 2011a; Lara et al. 2008; Nikolaev et al. 2005; Smirnov et al. 2005; Smirnov et al. 2009; Tekle et al. 2008) are confirmed here with the addition of sequences for four genes (α and β tubulins, EF2α and 14-3-3). The Arcellinida appear in our most likely tree as monophyletic with low support (27% BS, Figure 3). With the exception of a strongly supported relationship between the Amoebidae and the "Hartmannelidae" (83% BS), together comprising the taxon

Tubulinida (Smirnov et al. 2005), the relationships between the six main lineages remain uncertain as evidenced by low support for the backbone of the tree (Figure 3).

The more comprehensive sampling presented here also enables scrutiny of more fine-grained hypotheses within each of the six major lineages. Higher level relationships within the Arcellinida are currently defined according to shell composition, though this classification was proposed as explicitly provisional (Meisterfeld 2002). The three more inclusive groups are: 1) Difflugina, characterized by an agglutinated shell composed of either collected particles (xenosomes, e.g. *Difflugia*) or biomineralized particles (idiosomes, e.g. *Lesquereusia*); 2) Arcellina, characterized by a secreted organic membranous (e.g. *Microchlamys*) or chitinoid shell (e.g. *Arcella)*; and 3) Phryganelina, which are classified separately by their distinctive pseudopodial morphology rather than by features of the shell (*Cryptodifflugia* and *Phryganella*) (Meisterfeld 2002). The Arcellina (represented here by the genera *Arcella, Pyxidicula* and *Spumochlamys*) do not appear monophyletic, though monophyly is not rejected by the AU test (Figure 3, Table 4). The monophyly of Difflugina can be rejected by the AU test (Figure 3, Table 4), indicating that agglutination is either an ancestral character state in the group or evolved several times convergently. The current topology indicates that agglutination is the ancestral state because this would require fewer transitions (Figure 4).

The monophyly of the Lesquereusiidae can also be rejected by AU tests (Table 4), indicating at least two origins of silica biomineralization within the Arcellinida (the monophyly of *Lesquereusia* and *Netzelia* cannot be rejected, but *Quadrulella* falls within the Nebelidae, separate from the other two). This contrasts with the hypothesized single origin of biomineralization in a clade of testate amoebae in Rhizaria: the Euglyphida

(Heger et al. 2010; Lara et al. 2007). The results presented here and elsewhere regarding multiple independent origins of silica biomineralization will make interpretation of fossil Arcellinida more difficult, as biomineralization is one of the few characters that can be unambiguously determined in poorly preserved fossil tests (Bosak et al. 2011).

The current reconstruction, as well as other recent phylogenies based on SSU-rDNA, actin and Cox1 (Gomaa et al. Submitted; Kosakyan et al. Submitted; Lahr et al. 2011a), reveals that shell shape might be more indicative of relationships than shell composition (Figure 4). Organisms with similar shell shape group together: *Arcella* and *Netzelia* both have shells that are round in cross-section with a round aperture. *Quadrulella* and other Nebelidae (*Hyalosphenia*, *Nebela*, *Apodera, Porosia*) have vase-shaped shells that are flattened in cross-section and ellipsoid apertures (Figure 4). Of further evolutionary interest, there are a number of "intermediate" taxa, that is taxa displaying shell shape of a group and shell composition of another, which have yet to be sampled for molecular data. *Lesquereusia mimetica*'s shell has the typical *Lesquereusia* shell, with the neck bent over the body of the test. However *L. mimetica*'s shell is built with roughly agglutinated material, in a manner more similar to *Difflugia* (figures 21-28 in (Lahr and Lopes 2007)). Similarly, *Difflugia gramen* and *Difflugia achlora* have shells similar in shape to *Netzelia* (round shell with lobed aperture) but agglutination is more like *Difflugia*, (ie., with the absence of idiosomes; figures 11-15 in (Lahr and Lopes 2006)). *Pseudonebela africana* and *Nebela nebeloides* are both shaped like pyriform *Difflugia*, i.e., vase-shaped shells, with round cross-section and round apertures, respectively figures 1b-m in Lahr and Souza (2011) and figures 6-11 in Todorov et al.

(2010)) but the shell composition is more akin to that of *Nebela*, with agglutinated biomineralized plates.

Certain assumptions about test construction in the Arcellinida may need to be revised in light of the current results. The siliceous plates in *Nebela* are assumed to be collected either from the environment or from prey organisms, rather than autogenously produced (Meisterfeld 2002). However, the current results placing *Quadrulella* amidst the Nebelidae prompts a re-evaluation of this assumption, as it is possible that at least some members of the Nebelidae are actually able to synthesize silica. If so, a case of parallel evolution can be drawn by comparing the order of events in the two well supported clades shown here (Figure 4): *Netzelia* biomineralizes silica and the sister group *Arcella* secretes an organic shell; in the Nebelidae/ Hyalosphenidae, the *Nebela*, *Quadrulella*, *Porosia* and *Apodera* biomineralize silica while *Hyalosphenia* produces an organic shell. Assuming agglutination is the ancestral character state in the group, the evolution of the ability to biomineralize silica in both of these clades may have been followed by loss of this character independently in the *Arcella* and *Hyalosphenia.*

Within the Arcellinida there is extensive non-monophyly of well-established genera: *Nebela, Heleopera* and *Hyalosphenia* all appear non-monophyletic in the current and previous reconstructions (Lahr et al. 2011a; Lara et al. 2008), though the AU test does not allow rejection of monophyly for *Hyalosphenia* and *Heleopera*, while for *Nebela* the two out of three tests are able to reject the monophyly (Table 4). All three genera are well defined morphologically, and it comes as a surprise that multiple isolates end up in disparate portions of the tree (Figure 4). One possibility is that some of the isolates are contaminants or misidentifications. This hypothesis is unlikely, for in all

cases at least two independent laboratories have generated the sequences (e.g. both ssu-rDNA and actin genes for *Heleopera sphagni* were generated independently by the Edward Mitchell lab in Switzerland and the Laura Katz lab in the USA). One key aspect to keep in mind is that adding representatives of the 44 unsampled genera Meisterfeld (2002) will most certainly resolve/change the topology of the Arcellinida, as the majority of the tree is currently unsupported.

Non-monophyly of less inclusive lineages (e.g. genera and species) runs rampant in the Tubulinea beyond the Arcellinida. The genus *Hartmannella* is probably the most striking example, with taxa scattered in three of the five major Tubulinea lineages. Taking into account that the type strain *Hartmannella hyalina* is lost (Brown et al. 2010; Page 1967b), it is going to be extremely difficult to determine which of the many lineages should retain the taxon name and *Hartmannella* may qualify as *nomen nudum*. Perhaps the best solution will be to invalidate the genus, by transferring or proposing novel genera for each of the three major lineages. *Hartmannella abertawensis* stands out as an immediate candidate to be transferred to *Nolandella*, given its stable position in the current reconstruction as well as morphological characteristics and ecology—such transfer has also been recently suggested by Smirnov et al. (2011). Another case of non-monophyletic genus is *Rhizamoeba*, at least given the current taxon and gene sampling. The type strain *Rhizamoeba saxonica* CCAP 1570/2, which previously did not group with *Rhizamoeba* sp. ATCC 50742 (Dykova et al. 2008a; Smirnov et al. 2009; Smith et al. 2008), does so in the current reconstruction (Fig. 4). Morphologically, the CCAP isolate is distinct from the ATCC isolate, hence further taxonomic sampling will be necessary clarify the extent of morphological convergence within this group.

Observations of diverse testate amoebae in the Precambrian combined with the phylogeny of Arcellinida presented here generate several hypotheses on the early evolution of this taxon. Arcellinida fossils in marine sediments from 750 million years ago represent some of the most ancient and unambiguous records of eukaryotic life (Bosak et al. 2011; Porter and Knoll 2000; Porter et al. 2003). There is considerable taxonomic diversity in these marine sediments, including putative representatives morphologically similar to the modern genera *Arcella*, *Difflugia*, *Heleopera*, *Lesquereusia*, *Nebela* and *Trigonopyxis* (Bosak et al. 2011; Porter et al. 2003). These fossil marine morphologies interdigitate with the freshwater species characterized for this study, yet very few extant marine representatives have been described (e.g. *Pomoriella valkanovi* (Golemansky 1970)). Further, the monophyly of Arcellinida and the marine Poseidonida cannot be rejected (Table 4). Together, these observations lead to two related hypotheses: 1) Arcellinida evolved in a marine environment, perhaps from a common ancestor with the Poseidonida, and after extensive diversification each independent lineage switched to freshwater environments; 2) there is considerable diversity of extant marine Arcellinida yet to be discovered.

Table 5.1: Distribution of the 112 sequences characterized from 21 taxa.

Taxon	Source	SSU-rDNA	Actin	αtub	βtub	EF2α	14-3-3
Arcella gibbosa	Bear Swamp	JF694278	2	1	-	1	-
Arcella hemisphaerica	ATCC	EU273445	HM853688	1	1	1	1
Cryptodifflugia operculata	comm. Cult.	JF694279	JF694297	1	1	-	1
Difflugia bryophila	Hawley Bog	-	1	1	-	1	-
Difflugia lanceolata	Hawley Bog	-	3	-	-	-	-
Difflugia sp.	Hawley Bog	-	6	3	2	-	3
Heleopera sphagni	Hawley Bog	2	3	1	-	-	1
Hyalosphenia papilio	Hawley Bog	JF694282	1	1	-	-	2
Lesquereusia modesta	Bear Swamp	-	7	1	1	1	1
Lesquereusia spiralis	CB 131334	3	2	-	-	-	-
Nebela penardiana	Hawley Bog	1	3	1	-	-	4
Netzelia wailesi	Hawley Bog	1	2	-	-	-	2
Netzelia tuberculata	Hawley Bog	1	3	-	-	1	-
Quadrulella symmetrica	Hawley Bog	1	1	3	1	-	1
Hartmannella vermiformis	Smith Coll.	1	5	1	1	1	1
Saccamoeba lacustris	CCAP 1572/4	GQ221845	2	1	1	1	1
Rhizamoeba saxonica	CCAP 1570/2	EU719197	1	-	-	-	1
Nolandella hibernica	CCAP 1534/10	1	-	-	-	-	-
Chaos carolinense	CB 131324	AJ314607	1	1	2	1	1
Amoeba proteus	CB 131306	-	1	1	-	-	-
Nolandella sp.	ATCC 50913	EU273451	EU273446	EU273448	EU273450	1	-

Table 5.2: List of primers used to amplify genes in the current study.

Gene	Primer name	Primer sequence	Reference
SSU	SSU5'	ACC TGG TTG ATC CTG CCA GT	(Medlin et al. 1988)
	SSU3'	GAT CCT TCT GCA GGT TCA CCT AC	
	SSU Int -2	TTY YCC GTG TTG ART CAR ATT RAG	(Snoeyenbos-West et al. 2002)
	SSU Int +1	YGG AGA RDS RGC YTG AKA RAY GGC	
Btub	318F	TGGGCTAAGGGTCAYTAYACNGARGG	(Tekle et al. 2007)
	734F	CTCCGTTTCCCNGGNCARYTNAA	
	792R	GAAGAAGTGNAGNCKNGGRAANGG	
	1191R	GGTGTACCAGTGNARRAARGCYTT	
atub	AtubF94	GGC AAG GAG GAC GCN GCN AAY AAY TWY GC	(Tekle et al. 2007)
	AtubR341	TTG AAG CCT GTC GGR CAC CAR TCN ACR AAY TG	
	AtubR400	ACC TTC GCC GAC RTA CCA RTG NAC RAA NGC	
Actin	Actin245F	AAC TGG GAY GAY ATG GAR AAG AT	(Tekle et al. 2007)
	Actin1080R	ATC CAC ATY TGY TGG AAN GT	
14-3-3	34	CTG AGC AAG CTG ARM GNT AYG ANG ARA TGG	(Yoon et al. 2008)
	130	GTT GCC TAC AAR AAY GTY RTY GGN GC	
	455	AGT GCA AGA CCN ARN CGG ATN GGG TG	
	541	GCG ATG GCA TCA TCG AAN GCN TGR TTN GC	
ef2	351	GAA GTC ACT GCT GCN CTN CGN GTN ACN GA	(Yoon et al. 2008)
	411	GGT GTT TGC GTC CAA ACN GAR ACN GTN CT	
	1285	CGC CCG AAG GCA TAG AAN CGN CCY TTR TC	
	1756	AAA TCT CCA GGT GNA GYT CNC CNG CNC C	

Table 5.3: List of genes chosen for concatenation.

Taxon	actin	α-tubulin	β-tubulin	ef2-α	14-3-3
A. proteus	JG11.44_f4	SSU_JG11.60_abcd4	-	-	-
A. gibbosa	DL.133_JG10.86_b8_f8	JG10.89_a1_b1_c1_d1_e1_f1	-	JG10.150_b3	-
A. hemisphaerica	HM853688	JG8.93_a11_b11_c11_e1 1_a1	JG8.116_4clones	JG8.114	JG10.126_a1b1c1d1
C. carolinesis	Act_JG10.98_a1_b1	A6_JG10.105_a5_e5	JG11.60_b1_d1	JG11.46_a4_b4_c4	JG11.66_abcd6
C. operculata	JF694279	Crypto_atub	JHL_119_ABCD	-	JG11.5_7clones
Difflugia (comb)	DL121_JG40_DL131	JG10.25	JG8.126_btub2	JG11.38	JG11.66_b3
H.vermiformis	DL3.133_JG10.89_a2_c2	JG10.132_b5_c5_g5_h5	Hverm_btub	JG10.150_g4	JG10.126_a3c3b3
H. sphagni	JG11.28_a4_b5	JG10.105_a10_b10_c10_	-	-	JG11.5_f6g6h6
H. papilio	DL3.121.2_JG10.25_f11	JG10.25_a3	-	-	JG11.34_e7_f7_h7
Lesquereusia (comb)	DL3.139_JG10.98_b5	JG19.68_a3_b3_c3_d3_g2_	JG10.68_b3	JG10.68_c2e3	JG11_34
N. penardiana	JG11.33_e1_b1_d1	A6_JG11.33	-	-	JG11.33_f3_d3
Netzelia (comb)	DL3.99_JG8.138_h4	-	-	JG11.58_a3_e3	JG11.28_b4_f4
Q. symmetrica	JG8.127_b4_b5	DL137_JG10.86_b4__h4	JG10.86_c6f6h6	-	JG11.39_a10
R. saxonica	DL3.145_JG10.147_g3	-	-	-	DL_Rd3
S. lacustris	DL3.141_JG10.115_a9	JG11.5_a3_b3_c3_e3	JG11.5_e2f2g2h2	JG11.38_c6_e6_b7_d7_e	JG10.147_a7a8
Nolandella sp. 50913	EU273446	EU273448	EU273450	JG8.29	-

Table 5.4: Results from the approximately unbiased test.

Constraint tested	wkh	au	wsh
Amoebidae+Arcellinida	**0.00**	**0.00**	**0.00**
Amoebidae+Echinamoeboidea	**0.00**	**0.00**	**0.00**
Amoebidae+Hartmannelidae *core*	0.46	0.58	0.99
Amoebidae+Hartmannellidae *s.s.*	0.36	0.50	0.98
Amoebidae+Leptomyxidae	**0.00**	**0.00**	**0.00**
Arcellinida+Echinamoeboidea	**0.00**	**0.00**	**0.00**
Arcellinida+Leptomyxida	**0.00**	**0.00**	**0.00**
Leptomyxida+Echinamoeboidea	**0.01**	**0.01**	0.08
Hartmannellidae *s.s.*	0.36	0.45	0.98
Hartmannellidae *core*+Echinamoeboidea	**0.00**	**0.00**	**0.00**
Hartmannellidae s.s.+Echinamoeboidea	**0.00**	**0.00**	**0.00**
Hartmannellidae *core*+Leptomyxida	**0.00**	**0.00**	**0.00**
Hartmannellidae s.s.+Leptomyxida	**0.00**	**0.00**	**0.00**
Hartmannellidae *core*+Arcellinida	**0.00**	**0.00**	**0.00**
Hartmannellidae *s.s.*+Arcellinida	**0.00**	**0.00**	**0.00**
Poseidonida+Amoebidae	**0.00**	**0.00**	**0.01**
Poseidonida+Echinamoeboidea	**0.00**	**0.00**	**0.00**
Poseidonida+Hartmannellida *core*	**0.00**	**0.00**	**0.01**
Poseidonida+Hartmannellida *s.s.*	**0.00**	**0.00**	**0.00**
Poseidonida+Leptomyxida	**0.00**	**0.00**	**0.00**
Poseidonida+Arcellinida	0.35	0.43	0.98
Lesquereusiidae	**0.00**	**0.00**	**0.00**
Lesquereusia+Quadrulella	**0.00**	**0.00**	**0.00**
Netzelia+Quadrulella	**0.00**	**0.00**	**0.00**
Lesquereusia+Netzelia	**0.02**	**0.01**	0.19
Difflugina	**0.00**	**0.00**	**0.00**
Arcellina	**0.02**	**0.02**	0.25
Arcella+Pyxidicula	**0.04**	**0.04**	0.33
Arcella+Spumochlamys	0.21	0.22	0.83
Spumochlamys+Pyxidicula	0.19	0.19	0.83
Hyalosphenia	0.44	0.51	0.99
Heleopera	0.29	0.35	0.96
Nebela	**0.03**	**0.04**	0.33

The constraints tested column lists the taxa that were tested for monophyly, if $p<0.05$ then the monophyly of the constrained group can be rejected. wkh – weighted Kishino-Hasegawa test; au – Approximatelly unbiased test; wsh – weighted Shimodaira-Hasegawa test. The tests are listed in increasing order of conservativeness, that is, the wkh test is the least conservative, most prone to type I error. The wsh is the most conservative, most prone to type II error. The au test is the most balanced test. In bold are p values smaller than 0.05, indicating that monophyly of the group can be rejected.

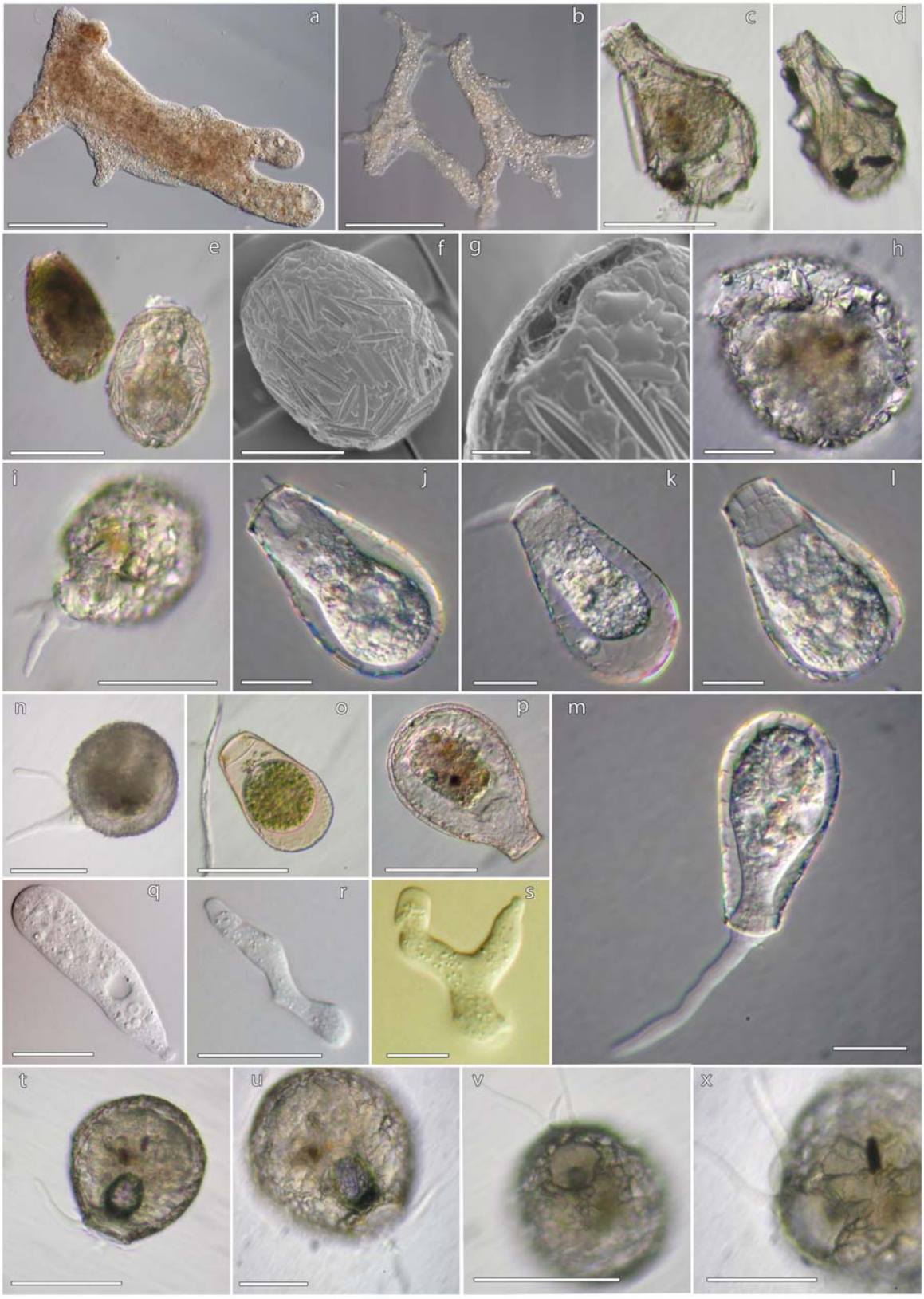

Figure 5.1: Images of organisms used in this study.

a) *Chaos carolinensis*, stack of eight images under differential interference contrast (DIC); b) two *Amoeba proteus* individuals, DIC; c, d) Individuals of *Difflugia bryophila* that were genome amplified, Hoffman Modulation Contrast (HMC); e) individuals of *Heleopera sphagni* that were genome amplified, HMC image; f, g) details of *Heleopera sphagni* shell under scanning electron microscopy; h, i) individuals of *Lesquereusia modesta* that were genome amplified, HMC images; j, k, l, m) individuals of *Quadrulella symmetrica* that were genome amplified (j, k) and had their cDNA libraries constructed (l, m); n) *Lesquereusia spiralis* individual that was genome amplified (HMC); o) *Hyalosphenia papilio* that was genome amplified (HMC); p) *Nebela carinata*; q) representative individual from culture of *Saccamoeba lacustris* CCAP 1572/4 (DIC); r, s) representative individuals from *Rhizamoeba saxonica* CCAP 1570/2 (DIC); t, u) *Netzelia wailesi* individual that was genome amplified; v, x) *Netzelia tuberculata* individual that was genome amplified, although images don't quite show the characteristic protuberances of the shell, these were prominent while observing the living individual. Scale bars are 100μm for a, b, c, d, e, i, n, o, p, t; 50μm for h, j, k, l, m, u, v, x; 30μm for f; 25μm for q, r; and 10μm for g.

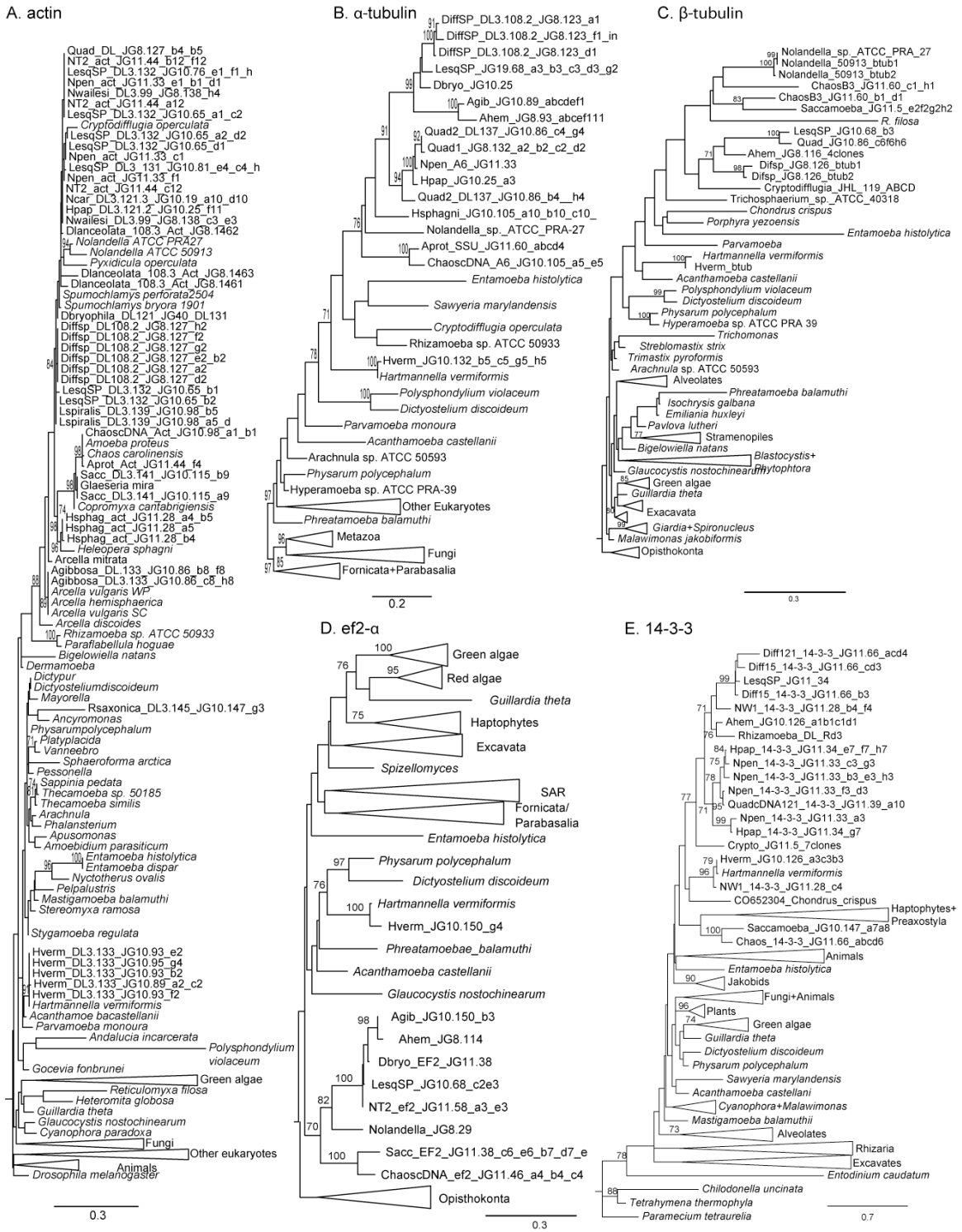

Figure 5.2: Gene-genealogies for each of the protein coding genes surveyed in the present study, including all characterized paralogs.

A) actin; B) α-tubulin; C) β-tubulin; D) elongation factor 2α; E) 14-3-3. Scale bar for each genealogy is indicated underneath the respective tree. Note: GenBank numberswill be substituted for paralog names as soon as available. Only bootstrap supports above 70% are shown.

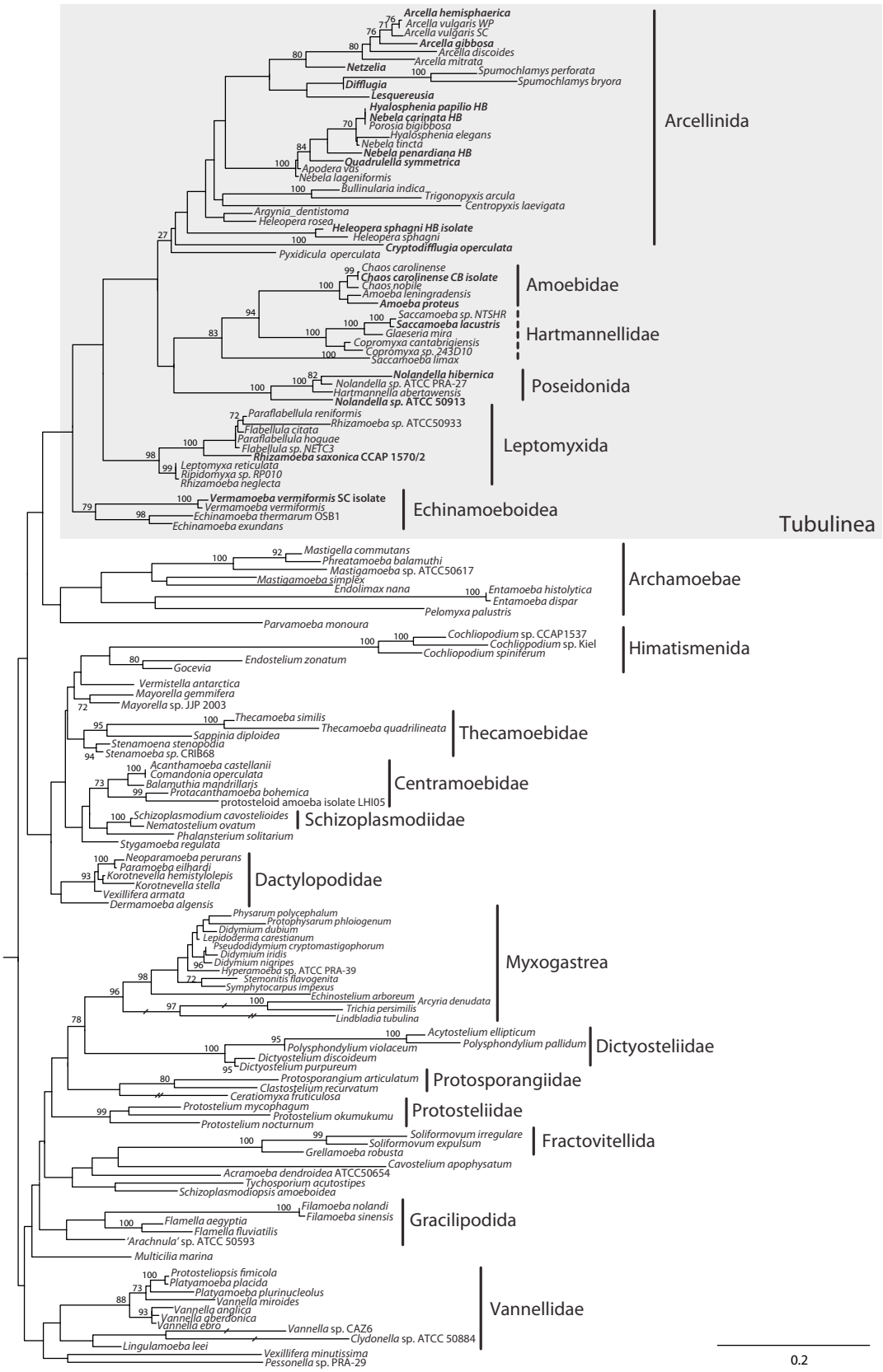
Arcella hemisphaerica
Arcella vulgaris WP
Arcella vulgaris SC
Arcella gibbosa
Arcella discoides
Arcella mitrata
Netzelia
Spumochlamys perforata
Spumochlamys bryora
Difflugia
Lesquereusia
Hyalosphenia papilio HB
Nebela carinata HB
Porosia bigibbosa
Hyalosphenia elegans
Nebela tincta
Nebela penardiana HB
Quadrulella symmetrica
Apodera vas
Nebela lageniformis
Bullinularia indica
Trigonopyxis arcula
Centropyxis laevigata
Argynia_dentistoma
Heleopera rosea
Heleopera sphagni HB isolate
Heleopera sphagni
Cryptodifflugia operculata
Pyxidicula operculata
Arcellinida
Chaos carolinense
Chaos carolinense CB isolate
Chaos nobile
Amoeba leningradensis
Amoeba proteus
Amoebidae
Saccamoeba sp. NTSHR
Saccamoeba lacustris
Glaeseria mira
Copromyxa cantabrigiensis
Copromyxa sp. 243D10
Saccamoeba limax
Hartmannellidae
Nolandella hibernica
Nolandella sp. ATCC PRA-27
Hartmannella abertawensis
Nolandella sp. ATCC 50913
Poseidonida
Paraflabellula reniformis
Rhizamoeba sp. ATCC50933
Flabellula citata
Paraflabellula hoguae
Flabellula sp. NETC3
Rhizamoeba saxonica CCAP 1570/2
Leptomyxa reticulata
Ripidomyxa sp. RP010
Rhizamoeba neglecta
Leptomyxida
Vermamoeba vermiformis SC isolate
Vermamoeba vermiformis
Echinamoeba thermarum OSB1
Echinamoeba exundans
Echinamoeboidea
Tubulinea
Mastigella commutans
Phreatamoeba balamuthi
Mastigamoeba sp. ATCC50617
Mastigamoeba simplex
Endolimax nana
Entamoeba histolytica
Entamoeba dispar
Pelomyxa palustris
Archamoebae
Parvamoeba monoura
Cochliopodium sp. CCAP1537
Cochliopodium sp. Kiel
Cochliopodium spiniferum
Himatismenida
Endostelium zonatum
Gocevia
Vermistella antarctica
Mayorella gemmifera
Mayorella sp. JJP 2003
Thecamoeba similis
Thecamoeba quadrilineata
Sappinia diploidea
Stenamoena stenopodia
Stenamoeba sp. CRIB68
Thecamoebidae
Acanthamoeba castellanii
Comandonia operculata
Balamuthia mandrillaris
Protacanthamoeba bohemica
protosteloid amoeba isolate LHI05
Centramoebidae
Schizoplasmodium cavostelioides
Nematostelium ovatum
Phalansterium solitarium
Stygamoeba regulata
Schizoplasmodiidae
Neoparamoeba perurans
Paramoeba eilhardi
Korotnevella hemistylolepis
Korotnevella stella
Vexillifera armata
Dermamoeba algensis
Dactylopodidae
Physarum polycephalum
Protophysarum phloiogenum
Didymium dubium
Lepidoderma carestianum
Pseudodidymium cryptomastigophorum
Didymium iridis
Didymium nigripes
Hyperamoeba sp. ATCC PRA-39
Stemonitis flavogenita
Symphytocarpus impexus
Echinostelium arboreum
Arcyria denudata
Trichia persimilis
Lindbladia tubulina
Myxogastrea
Acytostelium ellipticum
Polysphondylium pallidum
Polysphondylium violaceum
Dictyostelium discoideum
Dictyostelium purpureum
Dictyosteliidae
Protosporangium articulatum
Clastostelium recurvatum
Ceratiomyxa fruticulosa
Protosporangiidae
Protostelium mycophagum
Protostelium okumukumu
Protostelium nocturnum
Protosteliidae
Soliformovum irregulare
Soliformovum expulsum
Grellamoeba robusta
Cavostelium apophysatum
Acramoeba dendroidea ATCC50654
Tychosporium acutostipes
Schizoplasmodiopsis amoeboidea
Fractovitellida
Filamoeba nolandi
Filamoeba sinensis
Flamella aegyptia
Flamella fluviatilis
'Arachnula' sp. ATCC 50593
Gracilipodida
Multicilia marina
Protosteliopsis fimicola
Platyamoeba placida
Platyamoeba plurinucleolus
Vannella miroides
Vannella anglica
Vannella aberdonica
Vannella ebro
Vannella sp. CAZ6
Clydonella sp. ATCC 50884
Vannellidae
Lingulamoeba leei
Vexillifera minutissima
Pessonella sp. PRA-29
0.2

Figure 5.3: Most likely reconstruction of the Amoebozoa phylogeny.

The Tubulinea, focus of the current research, are highlighted by a gray box. Taxa in bold are taxa for which the present work has contributed novel data. Tree was rooted with eukaryotic outgroups (not shown). Branches are drawn to scale, except in cases indicated by a dash, where branches where cut in half, or two dashes, where branches were cut to a quarter of original length. Dashed lines indicate non-monophyletic groupings. Only bootstraps supports above 70% are shown.

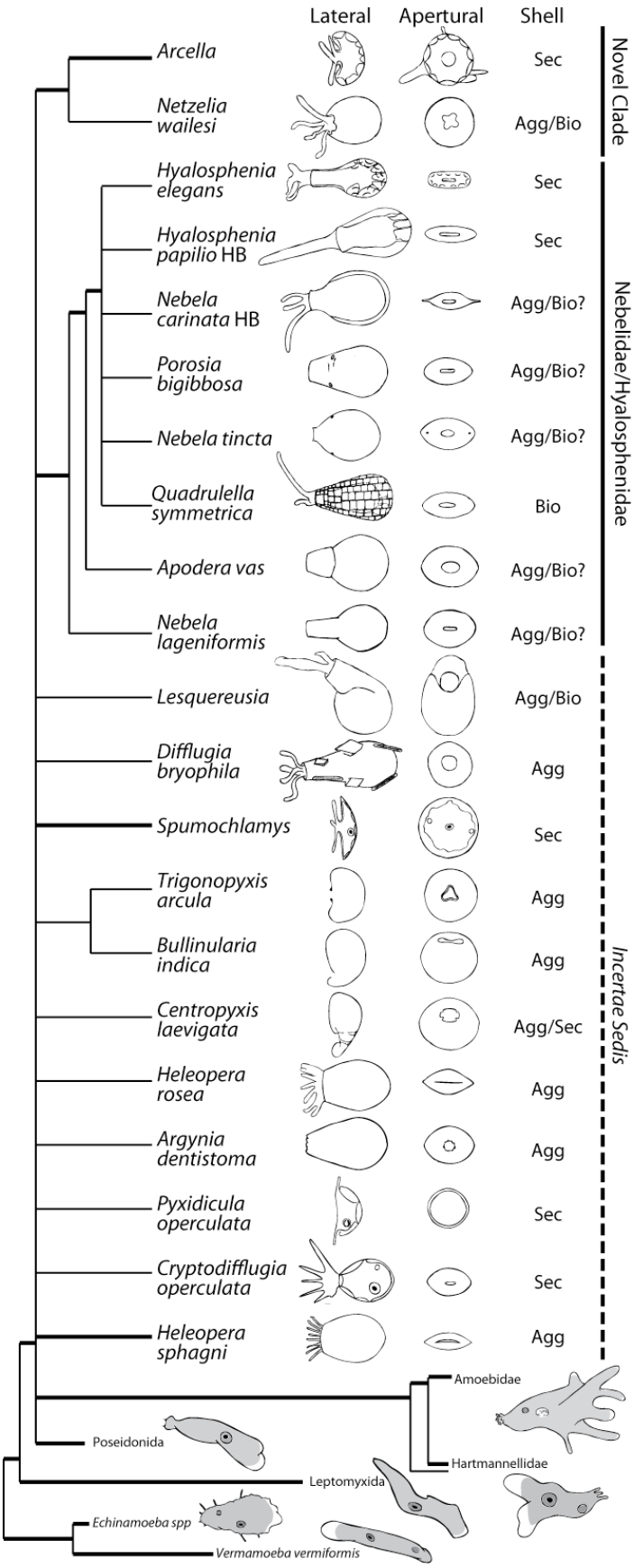
Lateral
Apertural
Shell
Arcella
Sec
Netzelia wailesi
Agg/Bio
Hyalosphenia elegans
Sec
Hyalosphenia papilio HB
Sec
Nebela carinata HB
Agg/Bio?
Porosia bigibbosa
Agg/Bio?
Nebela tincta
Agg/Bio?
Quadrulella symmetrica
Bio
Apodera vas
Agg/Bio?
Nebela lageniformis
Agg/Bio?
Lesquereusia
Agg/Bio
Difflugia bryophila
Agg
Spumochlamys
Sec
Trigonopyxis arcula
Agg
Bullinularia indica
Agg
Centropyxis laevigata
Agg/Sec
Heleopera rosea
Agg
Argynia dentistoma
Agg
Pyxidicula operculata
Sec
Cryptodifflugia operculata
Sec
Heleopera sphagni
Agg
Novel Clade
Nebelidae/Hyalosphenidae
Incertae Sedis
Amoebidae
Hartmannellidae
Poseidonida
Leptomyxida
Echinamoeba spp
Vermamoeba vermiformis

Figure 5.4: Relationships among the Tubulinea, illustrating morphological traits.

Morphological aspect for testate amoebae is illustrated from both the lateral and apertural views. Method used for shell construction is indicated in the third column: Agg – agglutinated, Bio – biomineralized, Sec – secreted. Branches were collapsed to polytomies where support is less than 70%. Thus, all resolved relationships shown have higher than 70% bootstrap support. Branches are not drawn to scale.

CHAPTER 6

THE CHASTITY OF AMOEBAE: RE-EVALUATING EVIDENCE FOR SEX IN AMOEBOID ORGANISMS

6.1 Abstract

Amoebae are generally assumed to be asexual. We argue that this view is relictual of early classification schemes that lumped all amoebae together inside the 'lower' protozoa separated from the 'higher' plants, animals, and fungi. This artificial classification allowed microbial eukaryotes, including amoebae, to be dismissed as primitive, and implied that the biological rules and theories developed for macroorganisms need not apply to microbes. Eukaryotic diversity is made up of 70+ lineages, most of which are microbial. Plants, animals and fungi are nested among these microbial lineages. Theories that apply to macroorganisms should in fact apply to microbial eukaryotes, though the theories may need to be refined and generalized (e.g. to account for the variation in sexual strategies and prevalence of facultative sex in natural populations of many microbial eukaryotes). We use a revised phylogenetic framework to assess evidence for sex in several amoeboid lineages that are traditionally considered asexual, and we interpret this evidence in light of theories on the evolution of sex developed for macroorganisms. We emphasize that the limited data available for many lineages coupled with natural variation in microbial life cycles have led to overestimate the extent of asexuality. Mapping sexuality onto the eukaryotic tree of life demonstrates that the majority of amoeboid lineages are, contrary to popular belief, anciently sexual

and that most asexual groups have probably arisen recently and independently. Additionally, several unusual genomic traits are prevalent in amoeboid lineages, including cyclic polyploidy, that may serve as alternative mechanisms to minimize the deleterious effects of asexuality.

> "Let us consider for a moment, a single *Ameba* … not as a cause of disease, but as a unit mass of protoplasm which … performs all of the fundamental vital activities common to living things … there is no reason to doubt that [the chemical composition of these unit masses] agrees with that of other living substances, since the accompanying properties of protoplasm—metabolism, growth and reproduction—are obviously performed in the same way."
>
> Gary N. Calkins, 1916

6.2 Introduction

Microbial eukaryotes were historically classified as primitive plants and animals (Haeckel 1866) or separated into their own kingdom (Corliss 1984; Margulis and Schwartz 1988; Whittaker 1969). This view received wide support with Whittaker's five kingdom classification system (Whittaker 1969) and continues to be popular in many circles. One consequence of lumping microbial eukaryotes into an artificial taxonomic unit (variously called Protista, Protoctista, or Protozoa) is the implicit view that microbes are fundamentally different entities than plants, animals, and fungi. As a result, microbial eukaryotes have been either dismissed as primitive or ignored in much of the theoretical work on eukaryotes, such as speciation theory (Mayr 1964) and theories on the evolution of sex (Maynard Smith 1978), with the notable exception of Bell (1988). However, given the current classification of eukaryotes, this dismissal is no longer acceptable. In recent

analyses, the eukaryotic tree of life is divided into a number of high-level lineages in which macroorganisms nest within predominantly microbial clades, demonstrating that the evolution of multicellularity has occurred multiple times (Adl et al. 2005; Baldauf 2003; Bonner 1998; Cavalier-Smith 1998; Keeling et al. 2005; Parfrey et al. 2010a). Hence, there is no evidence to suggest that unicellularity represents a "primitive" condition in eukaryotes.

The realization that there is no fundamental distinction between macro- and micro- eukaryotes calls for reassessment of the applicability of theories on the evolution of sex that were developed in macroorganisms to their microbial relatives. Differences between macroorganisms and microbial eukaryotes must be understood, as suggested by Calkins (1916), in terms of cell characteristics, habit and life cycle rather than an artificial and outdated taxonomic split. Current evidence suggests that sex has a single evolutionary origin and was present in the last common ancestor of eukaryotes (Dacks and Roger 1999). Hence, sex is a synapomorphy for extant eukaryotes and, where sex is absent, it must have been secondarily lost. The patchy distribution of sexual and asexual amoeboid lineages in current phylogenetic reconstructions requires many independent losses of sex (Figure 5.1), or may indicate that sex is present but not reproted in many lineages. We argue here that the amoeboid lineages are ideal candidates to investigate whether asexuality has been lost many times, because amoebae have generally been assumed to be asexual and are widespread in the tree of eukaryotes.

The body of theory developed from macroorganismal observations holds that sexuality should be pervasive and that asexuality should be limited to recent twigs on the tree of life (Schwander and Crespi 2009). We define sex as the presence of a meiotic

reduction of the genome complement followed eventually by karyogamy (nuclear fusion) in an organism's life cycle. In contrast to amphymyxis (Kondrashov 1997), our definition allows autogamy to be considered sex. Sex is argued to be advantageous because it generates variability by allowing independent assortment of genetic material through recombination (the advantage of sex, (Muller 1932; Weismann 1889)). Conversely, asexual lineages are argued to be subject to the accumulation of deleterious mutations through a process described as Muller's ratchet (Muller 1932; Muller 1964), leading to the prediction that asexual lineages should be short-lived and hence ancient asexuals will be rare (Judson and Normark 1996; Maynard Smith 1978; Muller 1964). On the other hand, sex is not beneficial for the individual in the short term, because only half of its genetic material is transmitted to the next generation (the cost of meiosis, (Maynard Smith 1978)). Recent efforts in modeling the evolution of sex show that incorporating genetic drift is essential to understand the dynamics of populations with finite size: when both drift and selection are taken into account, sex and recombination bring together alleles with higher selection coefficients that tend to be found in different individuals, outcompeting asexual lineages (Otto 2009). Thus, there are two main situations where asexuality is expected: 1) in relatively young lineages such as several species of scale insects with obligate apomictic thelytoky (Ross et al. 2010); and 2) in systems with very large population sizes, which rely on strategies for rapid reprodution (cell/organism replication) (Judson and Normark 1996).

We posit that the purported advantages and disadvantages of sex observed in multicellular macroorganisms should also apply to microbial eukaryotes. However, some caveats must be taken into account when comparing them. Firstly, life cycles are much

more varied and complex in microbial eukaryotes (Parfrey et al. 2008). For instance, in most plants and animals sex and growth are tightly linked, i.e., they cannot complete development without sex (Dacks and Roger 1999). Conversely, many microbial eukaryotes are only facultatively sexual, i.e., they may turn sex on or off depending on environmental conditions.

Knowledge about the natural history of microbial eukaryotes is deeply hindered by the difficulties of observation, when compared to macroorganisms. In most cases, organisms are assumed to be asexual because no sex has been observed; the gold standard for establishing sexuality remains direct observation of sexual phases of the life cycle. Proving that sex occurs in microbial eukaryotes is further hindered as there are often no sexually dimorphic forms and sexual life cycle stages may not occur readily in laboratory conditions, or they may be cryptic (Dunthorn and Katz 2010). Further, many amoebae are not culturable (e.g. polycystine radiolaria (Anderson 1981)). Despite these difficulties, sex has been observed in several microbial and non-microbial taxa long considered asexual when culturing conditions were modified or appropriate mating types were made available, including Darwinullid ostracods (Smith et al. 2006), arbuscular mycorrhizal fungi (Croll and Sanders 2009) and the filamentous mold *Aspergillus* (O'Gorman et al. 2009), and *Dictyostelium* (see below). Thus, it may not be prudent to rely on the absence of evidence as evidence for the absence of sex (Dunthorn and Katz 2010; Judson and Normark 1996).

Given the long history of study and diversity of methods used, evidence for sex in amoeboid lineages comes in a wide range of forms. We divide the continuum of evidence for sex into three categories: 1) confirmed sexual life cycle, 2) direct evidence

for sex, and 3) indirect evidence that suggest a sexual life cycle but is inconclusive. A confirmed sexual life cycle is the irrefutable compilation of both meiosis and karyogamy (nuclear fusion). Direct evidence for sex is provided by microscopic observations of either meiosis or karyogamy without confirmation of the other, or the presence of meiosis specific genes. We realize that for many biologists documenting meiosis alone is enough to confirm sexuality. However we feel that observation of both parts of the cycle are necessary given the variation in sexual mechanisms found in microbial eukaryotes. We are defending a more logical stance: if we define a phenomenon by the union of two elements, then we must expect to see the two elements for confirmation of said phenomenon. Conversely, the confirmation of karyogamy alone may indicate a parasexual system (one where subsequent haploidization occurs by some other means than meiosis (Pontecorvo 1956), also see the case of *Giardia* (Birky 2010)) but more strongly indicates the possibility of sex. Finally, many characteristics provide indirect evidence for the hypothesis that an organism is sexual, but fall short of conclusively demonstrating sex. These include molecular evidence of recombination, cytoplasmic fusion, evidence for complex life cycles with more than one trophic stage and production of putative reproductive cells (e.g. swarmer cells that can be interpreted as gametes).

6.3 Amoeboid lineages

The broad distribution of amoeboid organisms across the eukaryotic tree of life make them an ideal system for assessing the applicability of theories on sex to microbial lineages. Amoeboid organisms are defined by the ability to produce pseudopodia for locomotion or feeding. They were historically lumped into a single group, named

Sarcodina or Rhizopoda depending on the classification system (Pawlowski and Burki 2009). However, recent work demonstrates that amoebae are found in at least 30 distinct lineages (i.e. close to half of all described eukaryotic lineages) that are scattered throughout the tree of eukaryotes (Patterson 1999; Pawlowski and Burki 2009) (Figure 5.1). The majority of these lineages are clustered in the Amoebozoa and Rhizaria (Pawlowski and Burki 2009), with the remaining lineages scattered across the tree (Figure 5.1). The term "amoeba" is used here descriptively as a morphological category and has no phylogenetic meaning. Here, we reexamine the sexuality of amoebae in the context of the current phylogenetic framework of eukaryotes. We review evidence for sex in lineages traditionally considered asexual, and discuss reports of sexual life cycles that were originally considered exceptions or misinterpretations.

6.3.1 Amoebozoa

The Amoebozoa are a higher-level grouping encompassing over 5,000 species and are currently divided in ~14 lineages (Figure 5.2a). These lineages include familiar amoebae, such as the star of high school biology classes *Amoeba proteus* and the human enteric parasite *Entamoeba histolytica*. The majority of organisms shown to belong within Amoebozoa have amoeboid characteristics (Pawlowski and Burki 2009), although these encompass a wide range of morphologies, such as slime molds, lobose testate amoebae (Arcellinida), and amoeboflagellates. Asexuality in this group is thought to be either a defining characteristic (Hurst et al. 1992) or unknown (Cavalier-Smith 2002). However, deep inspection of the literature reveals evidence for sex in several Amoebozoa lineages: the dictyostelid sorocarpic slime molds and myxogastrid plasmodial slime

molds, Thecamoebida, Arcellinida, Leptomyxida, the genera *Entamoeba*, *Pelomyxa*, *Mastigamoeba*, *Trichosphaerium*, the sorocarpic slime mold *Copromyxa* and a number of protosteloid amoebae (Figure 5.2a, Table 5.1). We will briefly review the evidence for each of these groups.

Confirmed sexual life cycles are described for two lineages: the dictyostelid sorocarpic slime molds and the myxogastrid plasmodial slime molds. The dictyostelids illustrate the difficulty of observing sex in the laboratory. Known for their asexual life cycles (Bonner 1944; Bonner 1947), it was only in the 1970s that appropriate mating types of *Dictyostellium* were brought into culture and the sexual life cycle was fully documented (Erdos et al. 1973; Erdos et al. 1975; Macinnes and Francis 1974). The Myxogastria go through meiosis and fuse to form diploid plasmodia (Fiore-Donno et al. 2005; Martin and Alexopoulos 1969). Myxogastria have complex mating systems, with up to 13 mating types (roughly equivalent to sexes) described (Collins and Tang 1977).

Three lineages within Amoebozoa have direct evidence of sexual life cycles: the free-living thecamoebids, the sorocarpic slime mold *Copromyxa*, and the testate lobose amoebae (Arcellinida). The thecamoebid *Sappinia diploidea* makes a bicellular cyst where zygote formation is thought to occur (Goodfellow et al. 1974; Michel et al. 2006; Wenrich 1954), similar cysts have been reported in the related *Sappinia pedata* (Brown et al. 2007). The slime mold *Copromyxa*, has a life cycle that is consistent with sex although no secondary confirmation of meiosis has been described (Brown et al. 2007). *Copromyxa* was initially thought to be related to the acrasids *sensu lattu*, which in their turn were recognized as polyphyletic (Spiegel and Olive 1978) and are currently limited to the species placed in Excavata: Heterolobosea (Adl et al. 2005). However, molecular

studies demonstrate that *Copromyxa* is closely related to the Tubulinea (Brown et al. 2007). Multiple lines of evidence indicate that the Arcellinida are sexual: *Arcella vulgaris* shows microscopic evidence of synaptonemal complexes (Mignot and Raikov 1992), a typical structure that forms only during meiosis (Moses 1969). Molecular data from both *A. hemispherica* and *A. vulgaris* also demonstrate recombination in the actin gene (Lahr et al. 2010). *Paraquadrulla* and *Heleopera* go through nuclear division and subsequent fusion (Lüftenegger and Foissner 1991; Meisterfeld 2002). Finally, cell fusion (which we consider indirect evidence for sex, see below) has been reported for many genera of Arcellinida, though it is unclear whether karyogamy also occurs when cells fuse, or whether gamete formation occurs at other time points, (reviewed in (Wenrich 1954)). The most complete report of karyogamy following cytoplasmic fusion is for *Difflugia lobostoma* (Dangeard 1937), though Rhumbler (1898) reports not observing fusion during long-term culturing of this species. This apparent contradiction may indicate that these were different strains, a probable situation given the prevalence of cryptic species and other uncertainty in the taxonomy of Arcellinida (Heger et al. 2009; Lahr and Lopes 2009). Different life cycle observations may also be due to different culturing conditions.

A number of lineages have described complex life cycles, with the formation of multiple types of trophic cells that are consistent with sex, these are: the polyphyletic protosteloid amoebae *Clastostelium recurvatum*, *Protosporangium* spp., *Cavostelium apophysatum* and *Ceratiomyxella tahitiensis* (Shadwick et al. 2009) and the archamoebae *Pelomyxa palustris* (Whatley and Chapman-Andresen 1990)

Finally, three taxa have direct, but controversial evidence for sex: cell fusion reports in the free-living naked amoebae Leptomyxida, a complement of meiotic gene in the human pathogen *Entamoeba histolytica*, and a life cycle consistent with sex in *Trichosphaerium*. Cell fusion is widely reported for Amoebozoa (Seravin and Goodkov 1984a; Seravin and Goodkov 1984b), among the Leptomyxids, *Leptomyxa reticulata* (Seravin and Goodkov 1984a), *Flabellula baltica* (Smirnov and Goodkov 1999), and multiple strains of flabellulids (Dykova et al. 2008a) are observed to fuse. Subsequently the cells separate or persist as multinucleate stages. It is unclear whether this fusion facilitates genetic exchange or serves another purpose (Cavalier-Smith 2002), hence we consider this as only supporting evidence for sex.

Entamoeba histolytica has long been considered asexual despite numerous pieces of evidence pointing to the contrary, such as appearance of putative heterozygote populations after mixing of homozygotic populations for certain isozyme classes (Blanc et al. 1989; Sargeaunt et al. 1988). The availability of the whole genome (Loftus et al. 2005) shows that *E. histolytica* has the full complement of genes required for meiosis (Ramesh et al. 2005; Stanley 2005), which should have decayed if *E. histolytica* abandoned a sexual life cycle. The enigmatic genus of marine amoebae *Trichosphaerium* is reported to have an alternation of generations with gamont (sexual, including karyogamy) and schizont (asexual) stages (Angell 1976). Since meiosis has not been properly documented (Schaudinn 1899; Schuster 1976), we consider there is only direct evidence for sex in *Trichosphaerium*.

6.3.2 Rhizaria

The Rhizaria are a heterogeneous assemblage encompassing lineages such as Foraminifera, radiolarians and euglyphid testate amoebae, chlorarachniophytes, parasitic groups (Phytomyxea, Haplosporidia) as well as a multitude of other lesser-known flagellates (Figure 5.2b) that emerge as having fundamental ecological roles (Bass et al. 2009; Ekelund and Patterson 1997; Foissner 1991; Parfrey et al. 2010a). Filamentous pseudopodia are a recurrent morphological feature among amoeboid members of Rhizaria, in contrast to the lobose or broad pseudopodia of many Amoebozoa. Complete sexual life cycles are documented for two lineages: Foraminifera and *Gromia*; karyogamy or meiosis (direct evidence) was observed in five lineages: Euglyphida, Thecofilosea, Chlorarachniophyta, Plasmodiophorida and Phaeodaria; and indirect evidence such as cell fusion or formation of putative gametes in five lineages: Acantharia, Polycistinea, *Cercomonas*, *Helkesimastix* and *Lateromyxa*.

There are at least two lineages in the Rhizaria with confirmed sexual life cycles. Foraminifera are marine amoebae defined by a dynamic network of anastomosing pseudopodia (Bowser and Travis 2002), and well-known for producing intricate shells. They exhibit complex sexual life cycles with meiosis and gamete production occurring at separate stages (Goldstein 1999). The Gromiidae also have confirmed sexual life cycles (Arnold 1972). These large protists (up to several centimeters) have been observed in shallow and deep-sea sediments (Matz et al. 2008), where they are capable of denitrification in anoxic environments (Pina-Ochoa et al. 2010). *Gromia* was originally classified as a genus of Foraminifera based on gross morphology, but lacks the distinctive anastomosing pseudopods of Foraminifera and branches separately in molecular

phylogenies (Bass et al. 2009). The life cycle of *Gromia* resembles that of Foraminifera, with meiosis and gamete fusion occurring at different stages.

The Euglyphid testate amoebae and the Thecofilosa have many reports of cytoplasmic fusion, which we consider indirect evidence, and also reports of karyogamy, a form of direct evidence. Euglyphid testate amoebae have primarily been studied from a faunistic perspective, as bioindicators of past and present environmental conditions (Mitchell et al. 2008; Tolonen et al. 1992), and recently from a molecular phylogenetic perspective (Heger et al. 2010; Lara et al. 2007; Wylezich et al. 2002). In the family Euglyphidae, *Euglypha alveolata* (Reukauf 1912), *Euglypha scutigera* (Penard 1902) and *Euglypha* sp. (Awerintzew 1906) combine their cellular contents to form a cyst, or in one case a third larger shell (*E. alveolata* (Blochmann 1887)). Similar processes have been observed in other closely related families: Assulinidae (Awerintzew 1906), Trinematidae (Cash et al. 1915; Penard 1902), Cyphoderiidae (Cash et al. 1915; Rhumbler 1898); and in the unclassified *Tracheleuglypha dentata* (Chardez 1965). The formation of a third, larger cell has been reported only in Assulinidae and Euglyphidae (Schonborn and Peschke 1990; Valkanov 1962a), and not in Trinematidae and Cyphoderiidae, where cell fusion occurs within one of the copulating cells.

In some Euglyphids, cytoplasmic fusion is followed by karyogamy, providing direct evidence for sex. In *Trinema lineare*, *Valkanovia delicatula* (Valkanov 1962b), *Assulina muscorum* and *Valkanovia elegans* (Schonborn and Peschke 1990), karyogamy was documented but the ultimate fate of the synkaryon (fused nuclei) remains unknown. In *Corythion delamarei* (familiy Trinematidae) the synkaryon divides into four nuclei, interpreted as the result of meiosis (Iudina and Sukhanova 2000). The cytoplasm is then

distributed around the four nuclei, and four naked daughter cells leave the mother shell, which is left empty. These naked cells eventually secrete a test. If the interpretation is correct and *C. delamarei* indeed goes through meiosis after cytoplasmic and karyogamy, these organisms spend most of their life cycle in a haploid stage, being diploid only when karyogamy occurs. In contrast, *Trinema lineare* (Trinematidae) performs "conventional" binary divisions in addition to a sexual life cycle similar to *Corythion delamarei* (Sukhanova and Cheban 1990). Binary divisions were not observed in *Corythion delamarei*, or its sister species *Corythion dubium* (Iudina and Sukhanova 2000). This suggests that *Corythion* is a genus of obligate sexual organisms. In sum, there is direct evidence for sex in four families out of the five that compose Euglyphida.

The other lineage of filose testate amoebae, Thecofilosea (sensu (Cavalier-Smith and Chao 2003)) present direct evidence for sex. Recent phylogenetic analyses show they are not sister to the Euglyphida (Bass et al. 2009; Parfrey et al. 2010a). These amoebae may have proteinaceous or agglutinated tests and are often overlooked in environmental samples due to their small size. Cytoplasmic fusion followed by karyogamy has been observed in both *Pseudodifflugia gracilis* and *P. fascicularis*. The fate of the synkaryon is unknown (Valkanov 1962b).

Chlorarachniophytes, a group known for their ancient secondary endosymbiosis (Archibald 2009), go through an elaborate alternation of flagellate and amoeboid life stages and show indirect evidence for sex. In *Chlorarachnion reptans* flagellate cells fuse with coccoid cells, these are interpreted as "male" and "female" gametes (Grell 1990). In *Cryptochlora perforans*, two morphologically identical amoeboid cells fuse and produce a cyst where meiosis is thought to occur in a manner similar to euglyphids.

The DNA content of the cyst is double that of the amoeboid stages, suggesting karyogamy (Beutlich and Schnetter 1993). As meiosis has not been confirmed, we consider this direct evidence as opposed to confirmed sexual life cycle. The Plasmodiophorida are obligate intracellular parasites of plants, characterized by a specific type of mitotic division named cruciform nuclear division (Braselton 2002). They have a complex life cycle with a plasmodial amoeboid phase, and meiosis has been confirmed in the group. However, karyogamy has not yet been observed (Braselton 2002).

The organisms collectively designated "Radiolaria", a non-monophyletic assemblage containing Phaeodarea, Acantharia and Polycistinea, are large pelagic cells ubiquitous in the oceans. These organisms are extremely difficult to maintain in laboratory conditions, and their full life cycle has never been documented, but observations reveal evidence that suggests sex. All three groups of radiolarians generally produce small bi-flagellated cells, whose fate remains unclear (Anderson 1981; Raikov 1982), but may be gametes that are released into the water column.

The strongest evidence for sex within the 'Radiolaria' is found in Phaeodaria, specifically in the well-studied species *Aulacantha scolymantha*, which falls in the Cercozoa (Bass et al 2009). Synaptonemal complexes have been documented between the numerous (1000+) composite chromosomes. Each of these composite chomosomes subsequently segregates into developing bi-flagellated swarmer cells (Grell and Ruthmann 1964) and divides into eight chromosomes. However, complete evidence for sex is still lacking for this group, as cellular fusion and karyogamy have not been documented. Production of small bi-flagellated swarmer cells has also been observed in Polycystinea and Acantharea, which are closely related to Foraminifera (Parfrey et al.

2010). These have been interpreted as "isogametes" in the case of Acantharea (Febvre et al. 2002), but cell fusion has not been observed for either lineage (Anderson et al. 2002).

Evidence for sex becomes scarce as organisms get smaller and more difficult to observe. For the small amoeboflagellate forms there are reports of cell fusions with subsequent encystment: *Helkesimastix faecicola* (Woodcock and Lapage 1915) and *Cercomonas longicauda* (Woodcock 1916). In *Cercomonas*, cells can aggregate and fuse in some species, thus forming plasmodia containing up to 100 nuclei (Karpov 1997; Shirkina 1987). Such plasmodia have also been documented in the vampyrellid *Lateromyxa gallica* (Hulsmann 1993; Ropstorf et al. 1993), though the fate of these nuclei is unknown.

6.3.3 Other amoeboid lineages: Heliozoa, Heterolobosea, Stramenopila and Opisthokonta

There are other amoeboid lineages scattered in the tree of eukaryotes, most with limited information on sex. The Heliozoa have been split in four morphological lineages (Patterson 1999), three of which have been confirmed in molecular reconstructions (Nikolaev et al. 2004). One lineage, the Actinophryida nested within the Stramenopila, is reported to go through autogamy in the cyst (Mikrjukov and Patterson 2001). The life cycles of all three remaining "heliozoan" lineages, the Desmothoracida, Centrohelida and Gymnosphaerida remain poorly documented.

The Heterolobosea are a lineage of amoeboflagellates nested within the Excavata (Simpson 2003). *Heteramoeba clara* is reported to have a sexual life cycle consisting of a two mating-type system (Droop 1961), although there is a certain amount of doubt to

these experiments. The genome of *Naegleria gruberi* was recently sequenced, and reveals the presence of meiosis specific genes, supporting the presence of sex in this clade (Fritz-Laylin et al. 2010). The acrasid cellular slime molds have been shown to fall within the Heterolobosea rather than with other sorocarpic slime molds in Amoebozoa (Adl et al. 2005). Complete life cycles have been documented for acrasids, but these contain no evidence for meiosis or karyogamy. Hence, we consider there is no evidence pointing to sex in this group.

The Labirynthulidae and Thraustochytriidae are amoeboid organisms currently placed within the Stramenopiles (or Heterokonta), which also includes the diatoms, brown algae and water molds. A complete sexual cycle is described, with well documented meiosis (Moens and Perkins 1969; Perkins and Amon 1969).

A number of orphan amoeboid lineages have recently been placed amidst the Opisthokonts (which also includes the Fungi and Metazoa). *Amoebidium parasiticum*, originally thought to be a fungus, has a multi-stage life cycle, but no sex has been reported (Sumbali 2005). Similarly, the nucleariid amoebae and *Fonticula alba* have shown no evidence of sex (Brown et al. 2009). However, only a limited number of studies have focused on these taxa.

6.4 Conclusion

Evolutionary theory predicts that long-lived lineages should be sexual (Maynard Smith 1978), and that asexual lineages derived from sexual ancestors will be short-lived due to the negative effects of Muller's ratchet on the genome (Felsenstein 1974; Hamilton 2001). The two major clades that are dominated by amoebae, the Rhizaria and

Amoebozoa, (Figure 5.2) are certainly very ancient. Fossil Arcellinida, a clade of testate amoebae within the Amoebozoa, have been found in 750 Million year old rocks (Porter and Knoll 2000); Foraminifera and Polycystinea, two clades within Rhizaria, have fossil records that extend back at least to the Cambrian, i.e. 488-542 Mya (Anderson 1981; Sen Gupta 1999). Sex is a complex character and it is unlikely to have evolved independently in multiple lineages, or lost and regained multiple times (Dunthorn and Katz 2010). Thus, the presence of sexual lineages scattered across Amoebozoa and Rhizaria suggest that these clades were ancestrally sexual. As in other branches of the eukaryotic tree sex may then have been lost independently in derived lineages.

Some amoeboid lineages may be genuinely asexual. One candidate for asexuality is *Amoeba proteus*, which is the textbook example of binary fission in eukaryotes. A multitude of research groups have been culturing *Amoeba proteus* and its relatives for more than a century without uncovering evidence supporting the existence of sex in this group. Yet, assuming asexuality may be precarious given the uncertainties regarding culturing conditions. Although the ultimate proof for sex, as defined here, is the observation of meiosis and subsequent karyogamy, genomic data from populations of *A. proteus* could reveal evidence of recombination. Such data is yet lacking for these and the majority of amoeboid protists.

The logical equation “lack of evidence=asexual” is precarious, but the opposite stance is perhaps equally dangerous. Assuming that all lineages in Amoebozoa are sexual may mean discarding the possibility that alternative means to deal with Muller’s ratchet have independently arisen. Microbial eukaryote lineages may well have different strategies, such as lateral gene transfer and cyclic polyploidy. Bdelloid rotifers, a clade

of asexual microscopic animals, provide the most famous example of an alternative mechanism to avoid the ratchet: during rehydration following anhydrobiosis (a suspended animation state that allows the organism to survive dehydration), these organisms acquire foreign DNA and reorganize genomic regions (Gladyshev and Arkhipova 2010). This may well be a remarkable example of an evolutionary approach to reap the benefits of recombination, and could represent one of many strategies that eukaryotes have explored to avoid the deleterious effects of Muller's ratchet. If such an unusual mechanism appeared in Metazoa, comparably non-canonical mechanisms may have probably have evolved among 30+ amoeboid lineages.

Cyclic polyploidy may be another evasion method for avoiding the impact of Muller's ratchet. Ploidy cycles may reduce the mutational load usually associated with high ploidy, and maintain the selective advantages of haploid genetic transmission (Kondrashov 1997). Many microbial eukaryotes (amoeboid and others) experiment with ploidy changes that go far beyond the metazoan n-2n fluctuation (Parfrey et al. 2008). For instance, *Amoeba proteus* shows up to 3n variation during interphase, suggesting a cycle of polyploidization and return to haploidy before mitosis; and *Entamoeba histolytica* shows heterogeneity in nuclear ploidy due to varying levels of endomitosis: within a population, individual trophozoites exhibit continuous variation from 4n to 40n (Lohia 2003). The consequences of these phenomena are still poorly understood, as implications about the dynamics of eukaryotic genomes are only beginning to be explored (Parfrey et al. 2008).

An open question is whether lateral gene transfer (LGT) through endosymbiotic organisms may supply genetic variability to populations of amoebae. Diverse amoebae

(e.g. *Acanthamoeba* spp., *Hartmannella* spp., *Arcella* spp., *Amoeba* spp.) harbor a wide variety of bacterial endosymbionts and viruses during their life cycle (Corsaro and Venditti 2009; Greub and Raoult 2004; Jeon 2004). The possibility of genetic recombination between the amoeba and their multiple cytoplasmic inhabitants has just begun to be studied, as is the case of the giant amoeba-infecting Marseillevirus and Mimivirus that show evidence of chimeric genomes, with fragments of DNA acquired from multiple sources (Boyer et al. 2009).

Well-resolved phylogenetic trees may be used as a framework to investigate possible sexual and identify truly asexual lineages. *Amoeba proteus* is a member of the Amoebidae clade, for which no evidence for sex has been uncovered. The closely related Arcellinida are most likely sexual. Hence, the Amoebidae make an ideal group for deeply searching for signs of sex/asexuality. Documentation of the complete life cycle is difficult, but suitable alternative methods to identify the presence of sex include intense culturing and/or surveying of natural populations to document recombination (as predicted by meiosis) and genetic studies to identify a set of meiosis genes. In this case, there are three possible outcomes: 1) the Amoebidae are indeed sexual and we failed to document sex so far; 2) the Arcellinida-Amoebidae ancestral was sexual and the Amoebidae became truly asexual independently; or 3) the Amoebidae use a distinct strategy for evading Muller's ratchet, which might involve extensive LGT and/or ploidy cycles.

We conclude that the generalization about asexuality of amoeboid organisms is a superficial one and a product of two main forces: 1) an intrinsic practical difficulty in studying microbial organisms, and 2) the long held belief that amoeboid organisms are a

single unit of evolution, as opposed to an evolutionary strategy that was adopted by a wide variety of independent lineages. Amoebae are not fundamentally chaste. The timing and flow of events that lead each independent lineage to adopt an asexual or sexual life cycle must be evaluated separately. A multiple evidence approach, using a phylogenetic framework, gathering evidence on life cycles, genetic information on recombination and/or suits of meiotic genes will be more efficient in reconstructing the history of eukaryotic sexual life cycles. In line with Calkin's reasoning almost a century ago about the chemical constitution of amoebae (Moreira and Brochier-Armanet 2008), there is no reason to doubt that the rules of evolution governing sex in amoeboid organisms agree with that of other living beings. We predict that thorough and careful study of amoeboid organisms will reveal even more unusual ways of performing sex or otherwise exchanging genetic information. When discussing the sex of amoeboid protists, the existing evidence does not evoke chastity but rather Kama Sutra.

Table 6.1: Summary of evidence for sex in amoeboid organisms. Quoted names represent paraphyletic lineages.

Clade Organisms		Evidence	References
Amoebozoa			
Dictyosteliida	*Dictyostelium*	Full cycle	Erdos et al. 1973, Erdos et al. 1975, Macinnes and Francis 1974
Myxogastriida	several	Full cycle	Fiore-Donno et al. 2005, Martin and Alexopoulos 1969, Collins and Tang 1977
Thecamoebida	*Sappinia diploidea, S. pedata*	Nuclear fusion	Goodfellow et al. 1974, Michel et al. 2006, Wenrich 1964, Brown et al. 2007
"Hartmannellida"	*Copromyxa protea*	Life cycle with multiple trophic stages	Brown et al. 2011
Arcellinida	*Arcella*	Meiosis, actin gene recombination	Mignot and Raikov 1992, Lahr et al. 2011a
	Paraquadrulla, Heleopera	Nuclear fusion	Luftenegger and Foissner 1991
	Difflugia lobostoma, several others	Cellular fusion	Wenrich 1954, Dangeard 1937, Rhumbler 1898, Meisterfeld 2002
"Protosteloids"	several	Life cycle with multiple trophic stages	Shadwick et al. 2009
Archamoebae	*Pelomyxa palustris*	Life cycle with multiple trophic stages	Whatley and Chapman-Andresen 1990
	Entamoeba histolytica	Heterozygosity, full complement of meiotic genes	Blanc et al. 1989, Sargeaunt et al. 1988, Loftus et al. 2005, Ramesh et al. 2005, Stanley 2005
Leptomyxida	*Leptomyxa reticulata, Flabelulla baltica* + others	Cell fusion	Seravin and Goodkov 1984a, b, Smirnov and Goodkov 199, Dykova et al. 2008
Insertae Sedis	*Trichosphaerium*	Life cycle with multiple trophic stages, karyogamy	Schaudinn 1899, Schuster 1976
Rhizaria			
Foraminifera	several	Full cycle	Goldstein 1999
Gromiidae	*Gromia*	Full cycle	Arnold 1972
Euglyphida	*Euglypha, Trinema, Tracheleuglypha, Cyphoderia*	Cytoplasmic fusion	Reukauf 1912, Penard 1902, Awerintzew 1906, blochmann 1887, Cash et al. 1915, Rhumbler 1898, Chardez 1965, Shconborn and Peschke 1990, Valkanov 1962a
	Trinema, Valkanovia, Corythion	Nuclear fusion	Valkanov 1962b, Schonborn and Peschke 1990, Iudina and Sukhanova 2000, Sukhanova and Cheban 1990
Thecofilosea	*Pseudodifflugia*	Nuclear fusion	Valkanov 1962b

Chlorarachniophyta	*Chlorarachnion reptans, Cryptochlora perforans*	Life cycle with multiple trophic stages, cell fusion	Grell 1990, Beutlich and Schnetter 1993
Plasmodiophorida	several	Life cycle with multiple trophic stages, meiosis	Braselton 2002
Acantharia, Polycistinea	several	Putative gametes	Anderson 1981, Raikov 1982, Anderson et al. 2002, Febvre et al. 2002
Phaeodaria	*Aulacantha scolymantha*	Meiosis	Grell and Ruthmann 1964
Vampyrellids	Lateromyxa gallica	Cell fusion	Hulsmann 1993, Ropstorf et al. 1993
Insertae Sedis	*Helkesimastix faecicola, Cercomonas longicauda*	Cell fusion	Woodcock and Lapage 1915, Woodcock 1916, Karpov 1997, Shirkina 1987
Actinophryiida	several	Nuclear fusion	Mikrjukov and Patterson 2001
Labirynthulidae, Thraustochytriidae	several	Full cycle	Moens and Perkins 1969, Perkins and Amon 1969
Excavata			
Heterolobosea	*Heteramoeba clara*	Life cycle with multiple trophic stages	Droop 1961
	Naegleria gruberi	Full complement of meiosis genes	Fritz-Laylin et al 2010
Acrasida	several	Life cycle with multiple trophic stages	Adl et al. 2005
Opisthokonta			
Insertae Sedis	*Amoebidium parasiticum*	Life cycle with multiple trophic stages	Sumbali 2005
"nucleariids"	several, including *Fonticula*	Life cycle with multiple trophic stages	Brown et al. 2009

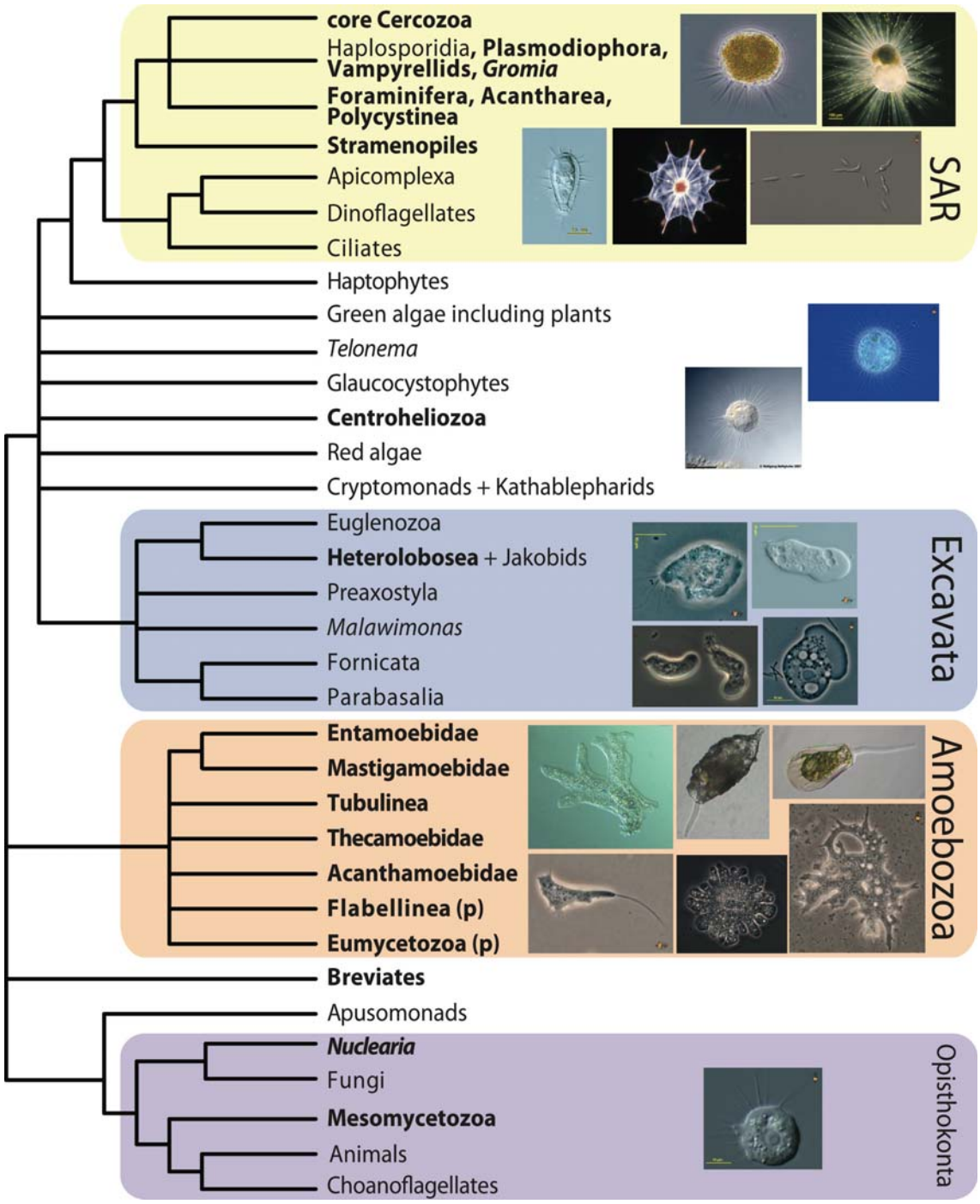

Figure 6.1: Distribution of amoeboid lineages in the eukaryotic tree of life.

This phylogenetic hypothesis of eukaryotic evolution is adapted from Parfrey et al. (2010), and depicts the well-supported higher-level groupings of eukaryotes. The lineages that have members with amoeboid morphology are in **bold**. Images depict exemplary amoeboid organisms and were retrieved from micro*scope.

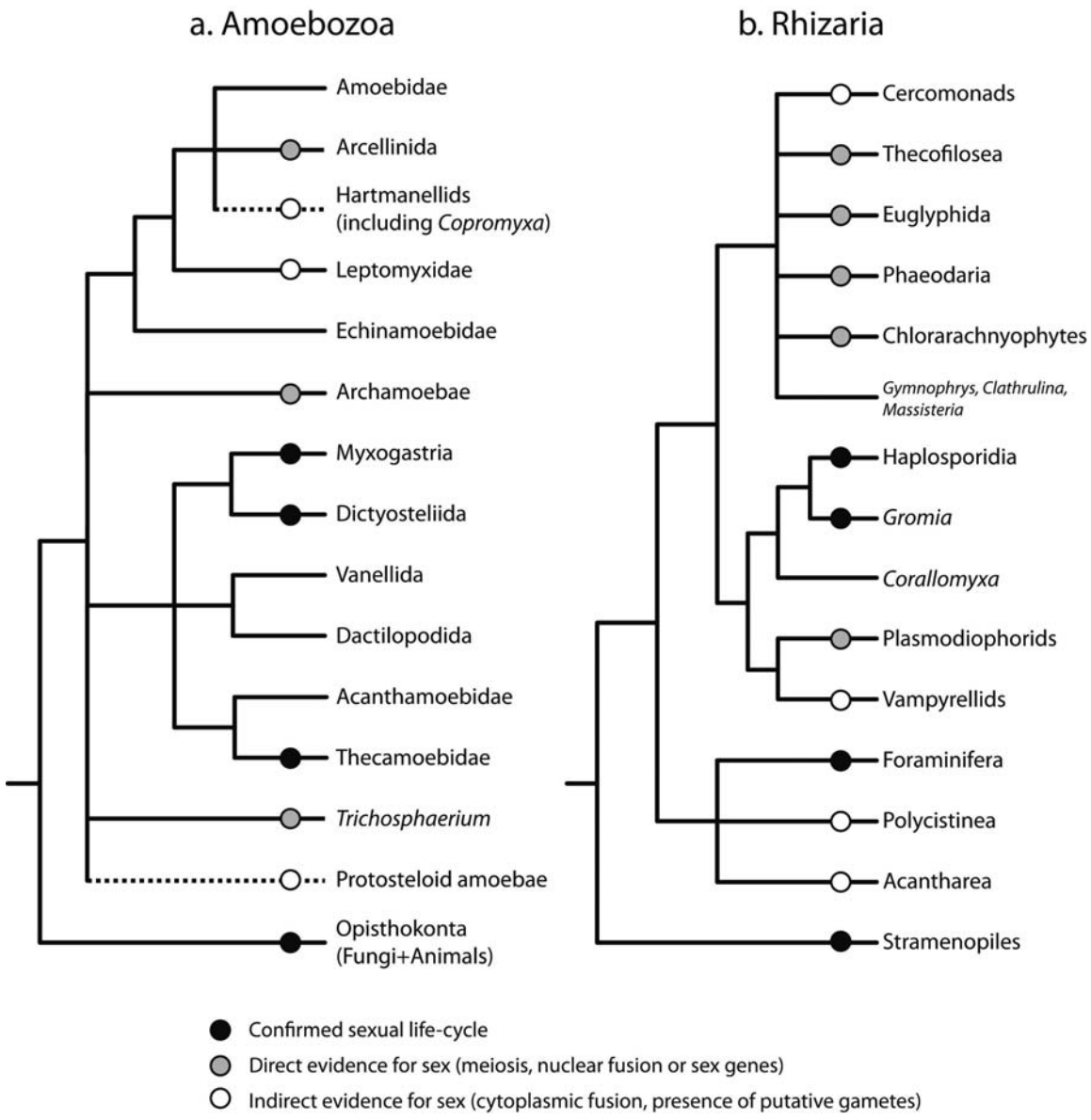

Figure 6.2: Distribution and types of evidence for sex in the main lineages of the two largest amoeboid groups: a) Amoebozoa and b) Rhizaria.

The topology of these illustrative trees are a consensus of well-supported lineages derived from Tekle et al. 2008; Burki and Pawlowski 2009; Shadwick et al. 2009 and Parfrey et al. 2010. Dashed lines represent non-monophyletic taxa.

BIBLIOGRAPHY

Abascal, F., R. Zardoya, and D. Posada. 2005. ProtTest: selection of best-fit models of protein evolution. Bioinformatics 21:2104-2105.

Acinas, S. G., R. Sarma-Rupavtarm, V. Klepac-Ceraj, and M. F. Polz. 2005. PCR-induced sequence artifacts and bias: insights from comparison of two 16S rRNA clone libraries constructed from the same sample. Appl Environ Microbiol 71:8966-8969.

Adl, S. M., A. G. B. Simpson, M. A. Farmer, R. A. Andersen, O. R. Anderson, J. R. Barta, S. S. Bowser, G. Brugerolle, R. A. Fensome, S. Fredericq, T. Y. James, S. Karpov, P. Kugrens, J. Krug, C. E. Lane, L. A. Lewis, J. Lodge, D. H. Lynn, D. G. Mann, R. M. McCourt, L. Mendoza, O. Moestrup, S. E. Mozley-Standridge, T. A. Nerad, C. A. Shearer, A. V. Smirnov, F. W. Spiegel, and M. Taylor. 2005. The new higher level classification of eukaryotes with emphasis on the taxonomy of protists. J Euk Mic 52:399-451.

Allen, E. E., and J. F. Banfield. 2005. Community genomics in microbial ecology and evolution. Nat Rev Microbiol 3:489-98.

Amaral Zettler, L. A., T. A. Nerad, C. J. O'Kelly, M. T. Peglar, M. T. Peglar, P. M. Gillevet, J. D. Silberman, and M. L. Sogin. 2000. A molecular reassessment of the leptomyxid amoebae. Protist 151:275.

Anderson, O. R. 1981. Radiolaria. Springer Verlag. Heidelberg, Germany.

Anderson, O. R., C. Nigrini, D. Boltovskoy, K. Takahashi, and N. R. Swanberg. 2002. Class Polycistina. Pages 994-1022 *in* The illustrated guide to the Protozoa. Second Edition. (J. J. Lee, G. F. Leedale, and P. C. Bradbury, eds.). Society of Protozoologists, Lawrence, KS.

Anderson, O. R., A. Rogerson, and F. Hannah. 1997. Three new limax amoebae isolated from marine surface sediments: *Vahlkampfia caledonica* n. sp., *Saccamoeba marina* n. sp., and *Hartmannella vacuolata* n. sp. J Euk Mic 44:33-42.

Angell, R. W. 1976. Observations on *Trichosphaerium platyxyrum* sp. n. J Protozool 23:357-364.

Archibald, J. M. 2009. The puzzle of plastid evolution. Current biology : CB 19:R81-R88.

Arnold, Z. M. 1972. Observations on the biology of the protozoan *Gromia oviformis* Dujardin. University of California Press, Berkley.

Awerintzew, S. 1906. Die Süsswasser Rhizopoden. I and II. Trudui S-Peter Oscch 36:351.

Bachvaroff, T. R., and A. R. Place. 2008. From stop to start: tandem gene arrangement, copy number and *trans*-splicing sites in the dinoflagellate *Amphidinium carterae*. PLoS ONE 3:e2929.

Baldauf, S. L. 2003. The deep roots of eukaryotes. Science 300:1703-1706.

Baldauf, S. L., A. J. Roger, I. Wenk-Siefert, and W. F. Doolittle. 2000. A kingdom-level phylogeny of eukaryotes based on combined protein data. Science 290:972-977.

Bapteste, E., H. Brinkmann, J. A. Lee, D. V. Moore, C. W. Sensen, P. Gordon, L. Durufle, T. Gasterlan, P. Lopez, M. Muller, and H. Philippe. 2002. 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.

Barns, S. M., C. F. Delwiche, J. D. Palmer, and N. R. Pace. 1996. Perspectives on archaeal diversity, thermophily and monophyly from environmental rRNA sequences. Proc Natl Acad Sci USA 93:9188 - 9193.

Bass, D., E. E. Y. Chao, S. Nikolaev, A. Yabuki, K. I. Ishida, C. Berney, U. Pakzad, C. Wylezich, and T. Cavalier-Smith. 2009. Phylogeny of novel naked filose and reticulose Cercozoa: Granofilosea cl. n. and Proteomyxidea revised. Protist 160:75-109.

Bell, G. 1988. Sex and death in protozoa: the history of an obsession. Cambridge University Press, Cambridge.

Bennett, W. E. 1986. Fine-structure of the trophic stage of *Endostelium* Olive, Bennet and Deasey, 1984 (Eumycetozoa, Protostelia). Protistologica 22:205-212.

Berney, C., J. Fahrni, and J. Pawlowski. 2004. How many novel eukaryotic 'Kingdoms'? Pitfalls and limitations of environmental DNA surveys. BMC Biology 2:13.

Berney, C., and J. Pawlowski. 2006. A molecular time-scale for eukaryote evolution recalibrated with the continuous microfossil record. Proc Roy Soc B 273:1867-1872.

Beutlich, A., and R. Schnetter. 1993. The life-cycle of *Cryptochlora perforans* (Chlorarachniophyta). Acta Bot 106:441-447.

Bhattacharya, D., J. Aubry, E. C. Twait, and S. Jurk. 2000. Actin gene duplication and the evolution of morphological complexity in land plants. J Phycol 36:813-820.

Birky, C. W. 2010. Giardia Sex? Yes, but how and how much? Tr Parasitol 26:70-74.

Blanc, D., R. Nicholls, and P. G. Sargeaunt. 1989. Experimental production of new zymodemes of *Entamoeba histolytica* supports the hypothesis of genetic exchange. Trans R Soc Trop Med Hyg 83:787-90.

Blochmann, F. 1887. Zur Kenntniss der Fortpflanzung von *Euglypha alveolata* Dujardin. Morphologische Jahrbuch 13:173-183.

Bolivar, I., J. F. Fahrni, A. Smirnov, and J. Pawlowski. 2001. SSU rRNA-based phylogenetic position of the genera *Amoeba* and *Chaos* (Lobosea, Gymnamoebia): The origin of Gymnamoebae revisited. Mol Biol Evol 18:2306-2314.

Bonner, J. T. 1944. A descriptive study of the development of the slime mold *Dictyostelium discoideum*. Amer J Bot 31:175-182.

Bonner, J. T. 1947. Evidence for the formation of cell aggregates by chemotaxis in the development of the slime mold *Dictyostelium discoideum*. J Exp Zool 106:1-26.

Bonner, J. T. 1998. The origins of multicellularity. Integrative Biology: Issues, News, and Reviews 1:27-36.

Bosak, T., D. J. G. Lahr, S. B. Pruss, and F. A. Macdonald. 2011. Agglutinated tests in post-Sturtian cap carbonates of Namibia and Mongolia. Earth Planet Sci Lett 308:29-40.

Bovee, E. C. 1985. Class Lobosea, Carpenter 1861. Pages 158-211 *in* An illustrated guide to the Protozoa. (J. J. Lee, S. H. Hutner, and E. C. Bovee, eds.). Society of Protozoologists., Lawrence, Kansas.

Bowser, S. S., and J. L. Travis. 2002. Reticulopodia: structural and behavioral basis for the suprageneric placement of Granuloreticulosan protists. J Foram Res 32:440-447.

Boyer, M., N. Yutin, I. Pagnier, L. Barrassi, G. Fournous, L. Espinosa, C. Robert, S. Azza, S. Sun, M. G. Rossmann, M. Suzan-Monti, B. La Scola, E. V. Koonin, and D. Raoult. 2009. Giant Marseillevirus highlights the role of amoebae as a melting pot in emergence of chimeric microorganisms. Proc Natl Acad Sci USA 106:21848-53.

Bradley, R. D., and D. M. Hillis. 1997. Recombinant DNA sequences generated by PCR amplification. Mol Biol Evol 14:592-593.

Brakenhoff, R. H., J. G. G. Schoenmakers, and N. H. Lubsen. 1991. Chimeric cDNA clones: a novel PCR artifact. Nucl Acids Res 19:1949.

Braselton, J. P. 2002. Phylum Plasmodiophora (Plasmodiophoromycota) *in* The illustrated guide to the Protozoa, 2nd. Ed. (J. J. Lee, G. F. Leedale, and P. C. Bradbury, eds.). Society of Protozoologists, Lawrence, KS.

Brown, M. W., J. D. Silberman, and F. W. Spiegel. 2010. "Slime molds" among the Tubulinea: Molecular systematics and taxonomy of Copromyxa. Protist 162:277-287.

Brown, M. W., F. W. Spiegel, and J. D. Silberman. 2007. Amoeba at attention: phylogenetic affinity of Sappinia pedata. J Euk Mic 54:511-519.

Brown, M. W., F. W. Spiegel, and J. D. Silberman. 2009. Phylogeny of the "forgotten" cellular slime mold, *Fonticula alba*, reveals a key evolutionary branch within Opisthokonta. Mol Biol Evol 26:2699-2709.

Calkins, G. N. 1916. General biology of the protozoan life cycle. Am Nat 50:257-270.

Cash, J., J. Hopkinson, and G. H. Wailes. 1905. The British freshwater Rhizopoda and Heliozoa. Ray Scociety, London.

Cash, J., G. H. Wailes, and J. Hopkinson. 1915. The British freshwater Rhizopoda and Heliozoa: Rhizopoda part 3. Ray Society, London.

Casillas, S., N. Petit, and A. Barbadilla. 2005. DPDB: a database for the storage, representation and analysis of polymorphism in the *Drosophila* genus. Bioinformatics 21:26-30.

Cavalier-Smith, T. 1991. Archamoebae - the ancestral eukaryotes. Biosystems 25:25-38.

Cavalier-Smith, T. 1998. A revised six-kingdom system of life. Biol Rev Cambridge Philosophic Soc 73:203-266.

Cavalier-Smith, T. 2002. The phagotrophic origin of eukaryotes and phylogenetic classification of Protozoa. Int J Syst Evol Microbiol 52:297-354.

Cavalier-Smith, T., and E. E. Y. Chao. 2003. Phylogeny and classification of phylum Cercozoa (Protozoa). Protist 154:341-358.

Cavalier-Smith, T., E. E. Y. Chao, and B. Oates. 2004. Molecular phylogeny of Amoebozoa and the evolutionary significance of the unikont Phalansterium. Eur J Protistol 40:21-48.

Chardez, D. 1965. On the particular mode and little understood reproduction in aquatic thecamoebae (Rhizopoda; Testacea.). Bull Inst Agron Sta Rech Gembloux 33:26-34.

Charette, S. J., and P. Cosson. 2004. Preparation of genomic DNA from *Dictyostelium discoideum* for PCR analysis. Biotechniques 36:574-575.

Chen, M. M., and X. T. Shen. 2007. Nuclear actin and actin-related proteins in chromatin dynamics. Curr Opin Cell Biol 19:326-330.

Cienkowski, L. 1876. Ueber einige Rhizopoden und vermandte Organismen. Arch Mikr Anat 12:15-50.

Collins, O. R., and H. C. Tang. 1977. New mating types in *Physarum polycephalum*. Mycologia 69:421-423.

Colwell, R. K. 2006. EstimateS: Statistical estimation of species richness and shared species from samples. Version 8.

Corliss, J. O. 1984. The kingdom Protista and its 45 phyla. Biosystems 17:87-126.

Corsaro, D., R. Michel, J. Walochnik, K. D. Muller, and G. Greub. 2010. *Saccamoeba lacustris*, sp. nov. (Amoebozoa: Lobosea: Hartmannellidae), a new lobose amoeba, parasitized by the novel chlamydia 'Candidatus *Metachlamydia lacustris*' (Chlamydiae: Parachlamydiaceae). Eur J Protistol 46:86-95.

Corsaro, D., and D. Venditti. 2009. Detection of Chlamydiae from freshwater environments by PCR, amoeba coculture and mixed coculture. Res Microbiol 160:547-552.

Costas, B. A., G. McManus, M. Doherty, and L. A. Katz. 2007. Use of species-specific primers and PCR to measure the distributions of planktonic ciliates in coastal waters. Limnol Oceanogr Methods 5:163-173.

Croll, D., and I. R. Sanders. 2009. Recombination in *Glomus intraradices*, a supposed ancient asexual arbuscular mycorrhizal fungus. BMC Evol Biol 9:13.

Dacks, J., and A. J. Roger. 1999. The first sexual lineage and the relevance of facultative sex. J Mol Evol 48:779-783.

Dangeard, L. 1937. Memoirs sur le *Difflugia globulosa* Dujardin. Le Botaniste 28:229-274.

Darriba, D., G. L. Taboada, R. Doallo, and D. Posada. 2011. ProtTest 3: fast selection of best-fit models of protein evolution. Bioinformatics 27:1164-1165.

Dawson, S. C., and N. R. Pace. 2002. Novel kingdom-level eukaryotic diversity in anoxic environments. Proc Natl Acad Sci USA 99:8324 - 8329.

Deflandre, G. 1928. Le genre *Arcella* Ehrenberg. Morphologie-Biologie. Essai phylogénétique et systématic. Arch Protistenk 64:152-287.

Delport, W., A. F. Y. Poon, S. D. W. Frost, and S. L. K. Pond. 2010. Datamonkey 2010: a suite of phylogenetic analysis tools for evolutionary biology. Bioinformatics 26:2455-2457.

Delsuc, F., H. Brinkmann, and H. Philippe. 2005. Phylogenomics and the reconstruction of the tree of life. Nat Rev Genet 6:361-375.

DNASTAR. 1994. Lasergene. DNASTAR Inc.

Doherty, M., B. A. Costas, G. B. McManus, and L. A. Katz. 2007. Culture-independent assessment of planktonic ciliate diversity in coastal northwest Atlantic waters. Aquat Mic Ecol 48:141-154.

Droop, M. R. 1961. Evidence for a sexual phase in life-cycle of an amoeba. Nature 192:281-&.

Dunthorn, M., and L. A. Katz. 2010. Secretive ciliates and putative asexuality in microbial eukaryotes. Trends Microbiol 18:183-188.

Dykova, I., I. Fiala, H. Peckova, and H. Dvorakova. 2008a. Phylogeny of Flabellulidae (Amoebozoa: Leptomyxida) inferred from SSU rDNA sequences of the type strain of *Flabellula citata* Schaeffer, 1926 and newly isolated strains of marine amoebae. Folia Parasitol 55:256-264.

Dykova, I., M. Kostka, and H. Peckova. 2008b. Morphology and SSU rDNA-based phylogeny of a new strain of *Saccamoeba* sp. (*Saccamoeba* Frenzel, 1892, Amoebozoa). Acta Protozool 47:397-405.

Dykova, I., M. Kostka, and H. Peckova. 2010a. *Grellamoeba robusta* gen. n., sp. n., a possible member of the family Acramoebidae Smirnov, Nassonova et Cavalier-Smith, 2008. Eur J Protistol 46:77-85.

Dykova, I., M. Kostka, and H. Peckova. 2010b. Two new species of the genus *Stenamoeba* Smirnov, Nassonova, Chao et Cavalier-Smith, 2007. Acta Protozool 49:245-251.

Dykova, I., H. Peckova, I. Fiala, and H. Dvorakova. 2005. *Filamoeba sinensis* sp n., a second species of the genus *Filamoeba* Page, 1967, isolated from gills of *Carassius gibelio* (Bloch, 1782). Acta Protozool 44:75-80.

Edgcomb, V. P., D. T. Kysela, A. Teske, A. de Vera Gomez, and M. L. Sogin. 2002. Benthic eukaryotic diversity in the Guaymas Basin hydrothermal vent environment. Proc Natl Acad Sci USA 99:7658-7662.

Eichinger, L., J. A. Pachebat, G. Glockner, M. A. Rajandream, R. Sucgang, M. Berriman, J. Song, R. Olsen, K. Szafranski, Q. Xu, B. Tunggal, S. Kummerfeld, M. Madera, B. A. Konfortov, F. Rivero, A. T. Bankier, R. Lehmann, N. Hamlin, R. Davies, P. Gaudet, P. Fey, K. Pilcher, G. Chen, D. Saunders, E. Sodergren, P. Davis, A. Kerhornou, X. Nie, N. Hall, C. Anjard, L. Hemphill, N. Bason, P. Farbrother, B. Desany, E. Just, T. Morio, R. Rost, C. Churcher, J. Cooper, S. Haydock, N. van Driessche, A. Cronin, I. Goodhead, D. Muzny, T. Mourier, A. Pain, M. Lu, D. Harper, R. Lindsay, H. Hauser, K. James, M. Quiles, M. M. Babu, T. Saito, C. Buchrieser, A. Wardroper, M. Felder, M. Thangavelu, D. Johnson, A. Knights, H. Loulseged, K. Mungall, K. Oliver, C. Price, M. A. Quail, H. Urushihara, J. Hernandez, E. Rabbinowitsch, D. Steffen, M. Sanders, J. Ma, Y. Kohara, S. Sharp, M. Simmonds, S. Spiegler, A. Tivey, S. Sugano, B. White, D. Walker, J. Woodward, T. Winckler, Y. Tanaka, G. Shaulsky, M. Schleicher, G. Weinstock, A. Rosenthal, E. C. Cox, R. L. Chisholm, R. Gibbs, W. F. Loomis, M. Platzer, R. R. Kay, J. Williams, P. H. Dear, A. A. Noegel, B. Barrell, and A. Kuspa. 2005. The genome of the social amoeba *Dictyostelium discoideum*. Nature 435:43-57.

Ekelund, F., and D. J. Patterson. 1997. Some heterotrophic flagellates from a cultivated soil in Australia. Arch Protistenk 148:461-478.

Erdos, G. W., K. B. Raper, and L. K. Vogen. 1973. Mating types and macrocyst formation in *Dictyostelium discoideum*. Proc Natl Acad Sci USA 70:1828-1830.

Erdos, G. W., K. B. Raper, and L. K. Vogen. 1975. Sexuality In cellular slime mold *Dictyostelium giganteum*. Proc Natl Acad Sci USA 72:970-973.

Erickson, H. P. 2007. Evolution of the cytoskeleton. Bioessays 29:668-677.

Fahrni, J. F., I. Bolivar, U. Berney, E. Nassonova, A. Smirnov, and J. Pawlowski. 2003. Phylogeny of lobose amoebae based on actin and small-subunit ribosomal RNA genes. Mol Biol Evol 20:1881-1886.

Febvre, C., J. Febvre, and A. F. Michaels. 2002. Class Acantharea. Pages 1055-1084 *in* The illustrated guide to the protozoa (J. J. Lee, G. F. Leedale, and P. C. Bradbury, eds.). Society of Protozoologists, Lawrence, Kansas, USA.

Felsenstein, J. 1974. The evolutionary advantage of recombination. Genetics 78:737-756.

Fiore-Donno, A. M., C. Berney, J. Pawlowski, and S. A. Baldauf. 2005. Higher-order phylogeny of plasmodial slime molds (Myxogastria) based on elongation factor 1-A and small subunit rRNA gene sequences. J Euk Microbiol 52:201-210.

Fiore-Donno, A. M., S. I. Nikolaev, M. Nelson, J. Pawlowski, T. Cavalier-Smith, and S. L. Baldauf. 2010. Deep phylogeny and evolution of slime moulds (Mycetozoa). Protist 161:55-70.

Flakowski, J., I. Bolivar, J. Fahrni, and J. Pawlowski. 2006. Tempo and mode of spliceosomal intron evolution in actin of foraminifera. J Mol Evol 63:30-41.

Foissner, W. 1991. Diversity and ecology of soil flagellates. Pages 93-112 *in* The biology of free-living heterotrophic flagellates. (D. J. Patterson, and J. Larsen, eds.). Clarendon Press, Oxford.

Fritz-Laylin, L. K., S. E. Prochnik, M. L. Ginger, J. B. Dacks, M. L. Carpenter, M. C. Field, A. Kuo, A. Paredez, J. Chapman, J. Pham, S. Q. Shu, R. Neupane, M. Cipriano, J. Mancuso, H. Tu, A. Salamov, E. Lindquist, H. Shapiro, S. Lucas, I. V. Grigoriev, W. Z. Cande, C. Fulton, D. S. Rokhsar, and S. C. Dawson. 2010. The genome of *Naegleria gruberi* illuminates early eukaryotic versatility. Cell 140:631-642.

Galtier, N., M. Gouy, and C. Gautier. 1996. SEAVIEW and PHYLO_WIN: Two graphic tools for sequence alignment and molecular phylogeny. Comp Applic Bioscienc 12:543-548.

Gladyshev, E. A., and I. R. Arkhipova. 2010. Genome structure of bdelloid rotifers: shaped by asexuality or desiccation? J Hered 101:S85-93.

Goldstein, S. T. 1999. Foraminifera: A biological overview. Pages 37-56 *in* Modern Foraminifera (B. K. Sen Gupta, ed.) Kluwer, Dordrecht.

Golemansky, V. 1970. Rhizopodes nouveaus du psammon littoral de la Mer Noire (Note préliminaire). Protistologica 6:365-371.

Gomaa, F., M. Todorov, T. J. Heger, E. A. D. Mitchell, and E. Lara. Submitted. SSU-rDNA phylogeny questions de monophyly of Arcellinida (Amoebozoa) and reveals that *Difflugia*, the largest Arcellinida genus, is paraphyletic. Protist.

Goodfellow, L. P., J. H. Belcher, and F. C. Page. 1974. A light- and electron-microscopical study of *Sappinia diploidea*, a sexual amoeba. Protistologica 2:207-216.

Goodson, H. V., and W. F. Hawse. 2002. Molecular evolution of the actin family. J Cell Sci 115:2619-2622.

Gouy, M., S. Guindon, and O. Gascuel. 2010. SeaView Version 4: a multiplatform graphical user interface for sequence alignment and phylogenetic tree building. Mol Biol Evol 27:221-224.

Grant, J., Y. I. Tekle, O. R. Anderson, D. J. Patterson, and L. A. Katz. 2009. Multigene evidence for the placement of a heterotrophic amoeboid lineage *Leukarachnion* sp. among photosynthetic stramenopiles. Protist 160:376-385.

Grell, K. G. 1990. Indications Of Sexual Reproduction In The Plasmodial Protist Chlorarachnion-Reptans Geitler. Zeitschrift Fur Naturforschung C-A Journal Of Biosciences 45:112-114.

Grell, K. G., and A. Ruthmann. 1964. Uber die Karyologie des Radiolars *Aulacantha scolymantha* und die Feinstruktur seiner Chromosomen. Chromosoma 15:185.

Greub, G., and D. Raoult. 2004. Microorganisms resistant to free-living amoebae. Clin Microbiol Rev 17:413-33.

Guindon, S., and O. Gascuel. 2003. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. Syst Biol 52:696-704.

Haeckel, E. H. P. A. 1866. Generelle morphologie der organismen. Allgemeine grundzüge der organischen formen-wissenschaft, mechanisch begründet durch die von Charles Darwin reformirte descendenztheorie. G. Reimer, Berlin.

Hamilton, W. D. 2001. Narrow roads of gene land, Vol. 2: the evolution of sex. Oxford University Press, Oxford.

Hampl, V., L. Hug, J. W. Leigh, J. B. Dacks, B. F. Lang, A. G. B. Simpson, and A. J. Roger. 2009. Phylogenomic analyses support the monophyly of Excavata and resolve relationships among eukaryotic "supergroups". Proc Natl Acad Sci USA 106:3859-3864.

Hedley, R. H., C. G. Ogden, and N. J. Mordan. 1977. Biology and fine structure of *Cryptodifflugia oviformis* (Rhizopodea: Protozoa). Bull Brit M Nat Hist 30:311-328.

Heger, T. J., E. A. D. Mitchell, P. Ledeganck, S. Vincke, B. Van de Vijver, and L. Beyens. 2009. The curse of taxonomic uncertainty in biogeographical studies of free-living terrestrial protists: a case study of testate amoebae from Amsterdam Island. Journal Of Biogeography 36:1551-1560.

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

Hightower, R. C., and R. B. Meagher. 1986. The molecular evolution of actin. Genetics 114:315-332.

Holt, R. A., and S. J. Jones. 2008. The new paradigm of flow cell sequencing. Genome Res 18:839-46.

Hooper, S. L., and J. B. Thuma. 2005. Invertebrate muscles: muscle specific genes and proteins. Physiol Rev 85:1001-1060.

Huber, T., G. Faulkner, and P. Hugenholtz. 2004. Bellerophon: a program to detect chimeric sequences in multiple sequence alignments. Bioinformatics 20:2317-2319.

Hugenholtz, P., and T. Huber. 2003. Chimeric 16S rDNA sequences of diverse origin are accumulating in the public databases. Int J Syst Evol Microbiol 53:289-293.

Hulsmann, N. 1993. *Lateromyxa gallica* n. g., n. sp. (Vampyrellidae) - A filopodial ameboid protist with a novel life-cycle and conspicuous ultrastructural characters. J Eukaryot Microbiol 40:141-149.

Hurst, L. D., W. D. Hamilton, and R. J. Ladle. 1992. Covert sex. Trends Ecol Evol 7:144-145.

Iudina, T. A., and K. M. Sukhanova. 2000. The testate amoeba *Corythion delamarei*: cell biology and life cycle. Tsitologiya 42:613-623.

Jeon, K. W. 2004. Genetic and physiological interactions in the amoeba-bacteria symbiosis. J Eukaryot Microbiol 51:502-8.

Joseph, J. M., P. Fey, N. Ramalingam, X. I. Liu, M. Rohlfs, A. A. Noegel, A. Muller-Taubenberger, G. Glockner, and M. Schleicher. 2008. The actinome of *Dictyostelium discoideum* in comparison to actins and actin-related proteins from other organisms. PLoS ONE 3:e2654.

Judo, M. S. B., A. B. Wedel, and C. Wilson. 1998. Stimulation and suppression of PCR-mediated recombination. Nucl Acids Res 26:1819-1825.

Judson, O. P., and B. B. Normark. 1996. Ancient asexual scandals. Trends Ecol Evol 11:A41-A46.

Kanagawa, T. 2003. Bias and artifacts in multitemplate polymerase chain reactions (PCR). J Bioscienc Bioengin 96:317.

Karpov, S. A. 1997. Cercomonads and their relationship to the myxomycetes. Arch Protistenk 148:297-307.

Katoh, K., G. Asimenos, and H. Toh. 2009. Multiple alignment of DNA sequences with MAFFT. Methods Mol Biol 537:39-64.

Keeling, P. J., G. Burger, D. G. Durnford, B. F. Lang, R. W. Lee, R. E. Pearlman, A. J. Roger, and M. W. Gray. 2005. The tree of eukaryotes. Trends Ecol Evol 20:670-676.

Keller, M., and K. Zengler. 2004. Tapping into microbial diversity. Nat Rev Micro 2:141.

Kondrashov, A. S. 1997. Evolutionary genetics of life cycles. Annu Rev Ecol Syst 28:391-435.

Kosakyan, A., T. J. Heger, B. S. Leander, M. Todorov, E. A. D. Mitchell, and E. Lara. Submitted. COI barcoding of Nebelid Testate Amoebae (Amoebozoa:Arcellinida): extensive pseudocryptic diversity and redefinition of Family Nebelidae. Protist.

Kudryavtsev, A., D. Bernhard, M. Schlegel, E. E-Y Chao, and T. Cavalier-Smith. 2005. 18S ribosomal RNA gene sequences of *Cochliopodium* (Himatismenida) and the phylogeny of Amoebozoa. Protist 156:215.

Kudryavtsev, A., J. Pawlowski, and K. Hausmann. 2009a. Description and phylogenetic relationships of *Spumochlamys perforata* n. sp and *Spumochlamys bryora* n. sp (Amoebozoa, Arcellinida). J Eukaryot Microbiol 56:495-503.

Kudryavtsev, A., C. Wylezich, M. Schlegel, J. Walochnik, and R. Michel. 2009b. Ultrastructure, SSU rrna gene sequences and phylogenetic relationships of *Flamella* Schaeffer, 1926 (Amoebozoa), with description of three new species. Protist 160:21-40.

Kurata, S., T. Kanagawa, Y. Magariyama, K. Takatsu, K. Yamada, T. Yokomaku, and Y. Kamagata. 2004. Reevaluation and reduction of a PCR bias caused by reannealing of templates. Appl Environ Microbiol 70:7545-7549.

Kusakabe, T., I. Araki, N. Satoh, and W. R. Jeffery. 1997. Evolution of chordate actin genes: evidence from genomic organization and amino acid sequences. J Mol Evol 44:289-98.

Lahr, D. J., T. B. Nguyen, E. Barbero, and L. A. Katz. 2010. Evolution of the actin gene family in testate lobose amoebae (Arcellinida) is characterized by 2 distinct clades of paralogs and recent independent expansions. Mol Biol Evol 28:223-236.

Lahr, D. J. G., P. J. Bergmann, and S. G. B. C. Lopes. 2008. Taxonomic Identity in Microbial Eukaryotes: A Practical Approach Using the Testate Amoeba *Centropyxis* to Resolve Conflicts Between Old and New Taxonomic Descriptions. Eukar Microbiol 55:409-416.

Lahr, D. J. G., J. Grant, T. B. Nguyen, J. H. Lin, and L. A. Katz. 2011a. Comprehensive phylogenetic reconstruction of Amoebozoa based on concatenated analyses of SSU-rDNA and actin genes. PLoS ONE 6:e22780.

Lahr, D. J. G., and L. A. Katz. 2009. Reducing the impact of PCR-mediated recombination in molecular evolution and environmental studies using a new-generation high-fidelity DNA polymerase. Biotechniques 47:857-66.

Lahr, D. J. G., and S. G. B. C. Lopes. 2006. Morphology, biometry, ecology and biogeography of five species of *Difflugia* Leclerc, 1815 (Arcellinida : Difflugiidae), from Tiete River, Brazil. Acta Protozoologica 45:77-90.

Lahr, D. J. G., and S. G. B. C. Lopes. 2007. Ultra-structure and biometry of three lobose testate amoebae of the family Lesquereusiidae (Tubulinea : Arcellinida) based on specimens from Sao Paulo, Brazil. Acta Protozoologica 46:339-348.

Lahr, D. J. G., and S. G. B. C. Lopes. 2009. Evaluating the Taxonomic Identity in Four Species of the Lobose Testate Amoebae Genus Arcella Ehrenberg, 1832. Acta Protozoologica 48:127-142.

Lahr, D. J. G., L. W. Parfrey, E. A. D. Mitchell, L. A. Katz, and E. Lara. 2011b. The chastity of amoebae: re-evaluating evidence for sex in amoeboid organisms. Proc Roy Soc B 278:2081-2090.

Lahr, D. J. G., and M. B. G. e Souza. 2011. Occurrence of the lobose testate amoeba Pseudonebela africana (Amoebozoa, Arcellinida) in the Brazilian "cerrado". Eur J Protistol 47:231-234.

Lara, E., T. J. Heger, F. Ekelund, M. Lamentowicz, and E. A. D. Mitchell. 2008. Ribosomal RNA genes challenge the monophyly of the Hyalospheniidae (Amoebozoa : Arcellinida). Protist 159:165-176.

Lara, E., T. J. Heger, E. A. D. Mitchell, R. Meisterfeld, and F. Ekelund. 2007. SSU rRNA reveals a sequential increase in shell complexity among the euglyphid testate amoebae (Rhizaria : Euglyphida). Protist 158:229-237.

Lenz, T. L., and S. Becker. 2008. Simple approach to reduce PCR artefact formation leads to reliable genotyping of MHC and other highly polymorphic loci -- Implications for evolutionary analysis. Gene 427:117.

Liesack, W., H. Weyland, and E. Stackebrandt. 1991. Potential risks of gene amplification by PCR as determined by 16S rDNA analysis of a mixed-culture of strict barophilic bacteria. Microb Ecol 21:191.

Lindley, L. A., S. M. Edwards, and F. W. Spiegel. 2006. Variations in nucleolar morphology in *Eumycetozoans*. Rev Mexican Micol 23:75-81.

Loftus, B., I. Anderson, R. Davies, U. C. M. Alsmark, J. Samuelson, P. Amedeo, P. Roncaglia, M. Berriman, R. P. Hirt, B. J. Mann, T. Nozaki, B. Suh, M. Pop, M. Duchene, J. Ackers, E. Tannich, M. Leippe, M. Hofer, I. Bruchhaus, U. Willhoeft, A. Bhattacharya, T. Chillingworth, C. Churcher, Z. Hance, B. Harris, D. Harris, K. Jagels, S. Moule, K. Mungall, D. Ormond, R. Squares, S. Whitehead, M. A. Quail, E. Rabbinowitsch, H. Norbertczak, C. Price, Z. Wang, N. Guillen, C. Gilchrist, S. E. Stroup, S. Bhattacharya, A. Lohia, P. G. Foster, T. Sicheritz-Ponten, C. Weber, U. Singh, C. Mukherjee, N. M. El-Sayed, W. A. Petri, C. G. Clark, T. M. Embley, B. Barrell, C. M. Fraser, and N. Hall. 2005. The genome of the protist parasite *Entamoeba histolytica*. Nature 433:865-868.

Lohia, A. 2003. The cell cycle of *Entamoeba histolytica*. Mol Cell Biochem 253:217-22.

Lüftenegger, G., and W. Foissner. 1991. Morphology and biometry of twelve soil testate amoebae (Protozoa, Rhizopoda) from Australia, Africa, and Austria. Bull Brit M Nat Hist (Zool) 57:1-16.

Macinnes, M. A., and D. Francis. 1974. Meiosis in *Dictyostelium mucoroides*. Nature 251:321.

Maddison, W. P., and D. R. Maddison. 2005. MacClade v. 4.0.8. Sinauer Assoc.

Margulis, L., and K. Schwartz. 1988. Five Kingdoms: An illustrated guide to the phyla of life on earth (2nd edition). W. H. Freeman and Company, New York.

Martin, G. W., and C. J. Alexopoulos. 1969. The Myxomycetes. University of Iowa Press, Iowa City.

Matz, M. V., T. M. Frank, N. J. Marshall, E. A. Widder, and S. Johnsen. 2008. Giant deep-sea protist produces bilaterian-like traces. Current Biol 18:1849-1854.

Maynard Smith, J. 1978. The evolution of sex. Cambridge University Press, Cambridge.

Mayr, E. 1964. Systematics and the origin of species: From the viewpoint of a zoologist (With a new introduction by the author.). Dover Publications, Inc.

McDowell, J. M., S. R. Huang, E. C. McKinney, Y. Q. An, and R. B. Meagher. 1996. Structure and evolution of the actin gene family in *Arabidopsis thaliana*. Genetics 142:587-602.

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

Meisterfeld, R. 2002. Order Arcellinida Kent, 1880. Pages 827-860 *in* The Illustrated Guide to the Protozoa. (J. J. Lee, L. G. F., and B. P., eds.). Allen Press, Lawrence, Kansas, USA.

Meyerhans, A., J.P. Vartanian, and S. Wain-Hobson. 1990. DNA recombination during PCR. Nucl Acids Res 18:1687-1691.

Michel, R., C. Wylezich, B. Hauroder, and A. Smirnov. 2006. Phylogenetic position and notes on the ultrastructure of *Sappinia diploidea* (Thecamoebidae). Protistology 4:319-325.

Mignot, J.P., and I. B. Raikov. 1992. Evidence for meiosis in the testate amoeba *Arcella*. J Euk Microbiol 39:287-289.

Mikrjukov, K. A., and D. J. Patterson. 2001. Taxonomy and phylogeny of Heliozoa. III. Actinophryids. Acta Protozool 40:3-25.

Miller, M., M. Holder, R. Vos, P. Midford, T. Leiebowitz, L. Chan, P. Hoover, and T. Warnow. 2009. The CIPRES portals.

Misof, B., and K. Misof. 2009. A Monte Carlo Approach Successfully Identifies Randomness in Multiple Sequence Alignments : A More Objective Means of Data Exclusion. Syst Biol 58:21-34.

Mitchell, E. A. D., D. J. Charman, and B. G. Warner. 2008. Testate amoebae analysis in ecological and paleoecological studies of wetlands: past, present and future. Biodivers Conserv 17:2115-2137.

Moens, P. B., and F. O. Perkins. 1969. Chromosome number of a small protist - accurate determination. Science 166:1289.

Moran, D. M., O. R. Anderson, M. R. Dennett, D. A. Caron, and R. J. Gast. 2007. A description of seven Antarctic marine Gymnamoebae including a new subspecies, two new species and a new genus: *Neoparamoeba aestuarina antarctica* n. subsp., *Platyamoeba oblongata* n. sp., *Platyamoeba contorta* n. sp and *Vermistella antarctica* n. gen. n. sp. J Euk Microbiol 54:169-183.

Moreira, D., and C. Brochier-Armanet. 2008. Giant viruses, giant chimeras: the multiple evolutionary histories of mimivirus genes. BMC Evol Biol 8.

Moses, M. J. 1969. Structure and function of the synaptonemal complex. Genetics 61:Suppl:41-51.

Muller, H. J. 1932. Some genetic aspects of sex. Am Nat 66:118-138.

Muller, H. J. 1964. The relation of recombination to mutational advance. Mut Res 1:2-9.

Neron, B., H. Menager, C. Maufrais, N. Joly, J. Maupetit, S. Letort, S. Carrere, P. Tuffery, and C. Letondal. 2009. Mobyle: a new full web bioinformatics framework. Bioinformatics 25:3005-3011.

Nikolaev, S. I., C. Berney, J. F. Fahrni, I. Bolivar, S. Polet, A. P. Mylnikov, V. V. Aleshin, N. B. Petrov, and J. Pawlowski. 2004. The twilight of Heliozoa and rise of Rhizaria, an emerging supergroup of amoeboid eukaryotes. Proc Natl Acad Sci USA 101:8066-8071.

Nikolaev, S. I., C. Berney, N. B. Petrov, A. P. Mylnikov, J. F. Fahrni, and J. Pawlowski. 2006. Phylogenetic position of *Multicilia marina* and the evolution of Amoebozoa. Int J Syst Evol Microbiol 56:1449-1458.

Nikolaev, S. I., E. A. D. Mitchell, N. B. Petrov, C. Berney, J. Fahrni, and J. Pawlowski. 2005. The testate lobose amoebae (Order Arcellinida Kent, 1880) finally find their home within Amoebozoa. Protist 156:191.

O'Gorman, C. M., H. T. Fuller, and P. S. Dyer. 2009. Discovery of a sexual cycle in the opportunistic fungal pathogen *Aspergillus fumigatus*. Nature 457:471-474.

Ogden, C. G. 1979. Siliceous structures secreted by members of the subclass Lobosia (Rhizopodea:Protozoa). Bull Brit M Nat Hist (Zool) 36:203-207.

Olive, L. S. 1975. The mycetozoans. Academic Press, New York,.

Olive, L. S., W. E. Bennett, and M. C. Deasey. 1984. The new protostelid genus *Endostelium*. Mycologia 76:884-891.

Otto, S. P. 2009. The evolutionary enigma of sex. Am Nat 174 Suppl 1:S1-S14.

Page, F. C. 1966. *Cryptodifflugia operculata* n. sp. (Rhizopodea - Arcellinida Cryptodifflugiidae) and status of genus *Cryptodifflugia*. Trans Am Microscop Soc 85:506.

Page, F. C. 1967a. *Filamoeba nolandi* n. g., n. sp. a filose amoeba. Trans Am Microscop Soc 86:405.

Page, F. C. 1967b. Taxonomic criteria for limax amoebae, with descriptions of 3 new species of *Hartmannella* and 3 of *Vahlkampfia*. J Euk Microbiol 14:499-521.

Page, F. C. 1975. New family of amebas with fine pseudopodia. Zool J Linn Soc 56:73-89.

Page, F. C. 1976. An illustrated key to freshwater and soil amoebae : with notes on cultivation and ecology. Freshwater Biological Association, Ambleside.

Page, F. C. 1980. A light-microscopical and electron-microscopical comparison of limax and flabellate marine amebas belonging to 4 genera. Protistologica 16:57-78.

Page, F. C. 1983. Marine gymnamoebae. Institute of Terrestrial Ecology, Cambridge.

Page, F. C. 1987. The classification of naked amebas (Phylum Rhizopoda). Arch Protistenk 133:199-217.

Parfrey, L. W., E. Barbero, E. Lasser, M. Dunthorn, D. Bhattacharya, D. J. Patterson, and L. A. Katz. 2006. Evaluating support for the current classification of eukaryotic diversity. PLoS 2:2062-2073.

Parfrey, L. W., J. Grant, Y. I. Tekle, E. Lasek-Nesselquist, H. G. Morrison, M. L. Sogin, D. J. Patterson, and L. A. Katz. 2010. Broadly sampled multigene analyses yield a well-resolved eukaryotic tree of life. Syst Biol 59:518-533

Parfrey, L. W., D. J. G. Lahr, and L. A. Katz. 2008. The dynamic nature of eukaryotic genomes. Mol Biol Evol 25:787-794.

Patterson, D. J. 1999. The diversity of eukaryotes. Am Nat 154:S96-S124.

Pawlowski, J. 2008. The twilight of Sarcodina: a molecular perspective on the polyphyletic origin of amoeboid protists. Protistology 5:22.

Pawlowski, J., and F. Burki. 2009. Untangling the phylogeny of amoeboid protists. J Euk Microbiol 56:16-25.

Peden, J. F. 1999. Analysis of codon usage. University of Nottingham.

Pederson, T. 2008. As functional nuclear actin comes into view, is it globular, filamentous, or both? J Cell Biol 180:1061-1064.

Penard, E. 1902. Faune rhizopodique du bassin du Léman. Henry Kündig, Geneva.

Penn, O., E. Privman, H. Ashkenazy, G. Landan, D. Graur, and T. Pupko. 2010. GUIDANCE: a web server for assessing alignment confidence scores. Nuc Acid Res 38:W23-W28.

Perkins, F. O., and J. P. Amon. 1969. Zoosporulation in *Labyrinthula* sp - an electron microscope study. J Protozool 16:235.

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

Pina-Ochoa, E., S. Hogslund, E. Geslin, T. Cedhagen, N. P. Revsbech, L. P. Nielsen, M. Schweizer, F. Jorissen, S. Rysgaard, and N. Risgaard-Petersen. 2010. Widespread occurrence of nitrate storage and denitrification among Foraminifera and Gromiida. Proc Natl Acad Sci USA 107:1148-1153.

Pond, S. L. K., and S. D. W. Frost. 2005. Datamonkey: rapid detection of selective pressure on individual sites of codon alignments. Bioinformatics 21:2531-2533.

Pond, S. L. K., S. D. W. Frost, and S. V. Muse. 2005. HyPhy: hypothesis testing using phylogenies. Bioinformatics 21:676-679.

Pond, S. L. K., D. Posada, M. B. Gravenor, C. H. Woelk, and S. D. W. Frost. 2006. Automated phylogenetic detection of recombination using a genetic algorithm. Mol Biol Evol 23:1891-1901.

Pontecorvo, G. 1956. The parasexual cycle in fungi. Annu Rev Microbiol 10:393-400.

Porter, S. M., and A. H. Knoll. 2000. Testate amoebae in the Neoproterozoic Era: evidence from vase-shaped microfossils in the Chuar Group, Grand Canyon. Paleobiology 26:360-385.

Porter, S. M., R. Meisterfeld, and A. H. Knoll. 2003. Vase-shaped microfossils from the Neoproterozoic Chuar Group, Grand Canyon: A classification guided by modern testate amoebae. Journal of Paleontology 77:409-429.

Qiu, X., L. Wu, H. Huang, P. E. McDonel, A. V. Palumbo, J. M. Tiedje, and J. Zhou. 2001. Evaluation of PCR-generated chimeras, mutations, and heteroduplexes with 16S rRNA gene-based cloning. Appl Environ Microbiol 67:880-887.

Raikov, I. B. 1982. The protozoan nucleus: morphology and evolution. Springer-Verlag, Wien.

Ramesh, M. A., S.-B. Malik, and J. M. Longsdon. 2005. A phylogenomic inventory of meiotic genes: evidence for sex in *Giardia* and an early eukaryotic origin of meiosis. Curr Biol 15:185-191.

Reisler, E., and E. H. Egelman. 2007. Actin structure and function: what we still do not understand. J Biol Chem 282:36133-36137.

Reukauf, E. 1912. Zur Encystierung von *Euglypha alveolata*. Zoologischer Anzeiger 39:372-375.

Rhumbler, H. 1898. Zelleib-, Schalen-, und Kern-verschmelzungen bei den Rhizopoden und deren wahrscheinliche Beziehungen zu phylogenetischen Vorstufen der Metazoenbefruchtung. 2. Biologisches Centralblatt 18:33-38.

Rogerson, A., and D. J. Patterson. 2002. The naked ramicristate amoebae (Gymnamoebae). Pages 1023-1053 *in* An Illustrated Guide to the Protozoa, 2nd Edition (J. J. Lee, G. F. Leedale, and P. Bradbury, eds.). Society of Protozoologists, Lawrence, KS.

Romeralo, M., F. W. Spiegel, and S. L. Baldauf. 2010. a fully resolved phylogeny of the social amoebas (Dictyostelia) based on combined SSU and ITS rDNA sequences. Protist 161:539-548.

Ropstorf, P., N. Hulsmann, and K. Hausmann. 1993. Karyological investigations on the vampyrellid filose ameba *Lateromyxa gallica* Hulsmann 1993. Eur J Protistol 29:302-310.

Ross, L., I. Pen, and D. M. Shuker. 2010. Genomic conflict in scale insects: the causes and consequences of bizarre genetic systems. Biol Rev Cambridge Philosophic Soc.

Sargeaunt, P. G., T. F. Jackson, S. R. Wiffen, and R. Bhojnani. 1988. Biological evidence of genetic exchange in Entamoeba histolytica. Trans R Soc Trop Med Hyg 82:862-7.

Sawyer, T. K. 1975. Marine amebas from surface waters of Chincoteague Bay, Virginia - 2 new genera and 9 new species within Families Mayorellidae, Flabellulidae, and Stereomyxidae. Trans Am Microscop Soc 94:71-92.

Schaudinn, F. R. 1899. Untersuchungen uber den generationswechsel von Trichosphaerium sieboldi. Sch Abh Konigl Preuss Akd Wiss Berlin Suppl.:1-93.

Schonborn, W., and T. Peschke. 1990. Evolutionary studies on the *Assulina-Valkanovia* Complex (Rhizopoda, Testaceafilosia) in *Sphagnum* and soil. Biol Fert Soils 9:95-100.

Schuster, F. L. 1976. Fine structure of the schizont stage of the testate marine ameba, *Trichosphaerium* sp. J Euk Microbiol 23:86-93.

Schwander, T., and B. J. Crespi. 2009. Twigs on the tree of life? Neutral and selective models for intergrating macroevolutionary patterns and microevolutionary processes in the analysis of asexuality. Molec Ecol 18:28-42.

Sen Gupta, B. K. 1999. Introduction to Modern Foraminifera. Pages 1-6 *in* Modern Foraminfera (B. K. Sen Gupta, ed.) Kluwer, Dordrecht.

Seravin, L. N., and A. V. Goodkov. 1984a. Possible forms of agamic genetic interactions in protists and ways of establishment of the sexual process. Tsitologiya 26:1224-1236.

Seravin, L. N., and A. V. Goodkov. 1984b. The main types and forms of agamic cell fusion in protozoa. Tsitologiya 26:123-131.

Shadwick, L. L., F. W. Spiegel, J. D. L. Shadwick, M. W. Brown, and J. D. Silberman. 2009. Eumycetozoa = Amoebozoa?: SSUrDNA phylogeny of protosteloid slime molds and its significance for the Amoebozoan supergroup. PLoS ONE 4:e6754.

Shimodaira, H. 2002. An approximately unbiased test of phylogenetic tree selection. Syst Biol 51:492-508.

Shimodaira, H. 2004. Approximately unbiased tests of regions using multistep-multiscale bootstrap resampling. Ann Statist 32:2616-2641.

Shimodaira, H., and M. Hasegawa. 2001. CONSEL: for assessing the confidence of phylogenetic tree selection. Bioinformatics 17:1246-1247.

Shirkina, N. I. 1987. Biological features in some flagellates of the genus *Cercomonas* (Rhizomastigida, Cercomonadidae). Zoologichesky Zhurnal 66:485-489.

Simpson, A. G. B. 2003. Cytoskeletal organization, phylogenetic affinities and systematics in the contentious taxon Excavata (Eukaryota). Int J Syst Evol Microbiol 53:1759-1777.

Smirnov, A., and S. Brown. 2004. Guide to the methods of study and identification of soil gymnamoebae. Protistology 3:148-190.

Smirnov, A., E. Nassonova, C. Berney, J. Fahrni, I. Bolivar, and J. Pawlowski. 2005. Molecular Phylogeny and Classification of the Lobose Amoebae. Protist 156:129.

Smirnov, A., E. Nassonova, J. Fahrni, and J. Pawlowski. 2009. *Rhizamoeba neglecta* n. sp (Amoebozoa, Tubulinea) from the bottom sediments of freshwater Lake Leshevoe (Valamo Island, North-Western Russia), with notes on the phylogeny of the order Leptomyxida. Eur J Protistol 45:251-259.

Smirnov, A. V. 1996. *Stygamoeba regulata* n sp (Rhizopoda) - A marine amoeba with an unusual combination of light-microscopical and ultrastructural features. Arch Protistenk 146:299-307.

Smirnov, A. V., E. Chao, E. S. Nassonova, and T. Cavalier-Smith. 2011. A revised classification of naked lobose amoebae (Amoebozoa: Lobosa). Protist 162:545-570.

Smirnov, A. V., and A. V. Goodkov. 1999. An Illustrated list of basic morphotypes of Gymnamoebia (Rhizopoda, Lobosea). Protistology 1:20-29.

Smirnov, A. V., E. S. Nassonova, and T. Cavalier-Smith. 2008. Correct identification of species makes the amoebozoan rRNA tree congruent with morphology for the order Leptomyxida Page 1987; with description of *Acramoeba dendroida* n. g., n. sp., originally misidentified as *Gephyramoeba* sp. Eur J Protistol 44:35.

Smith, H. G., A. Bobrov, and E. Lara. 2008. Diversity and biogeography of testate amoebae. Biodivers Conserv 17:329-343.

Smith, R. J., T. Kamiya, and D. J. Horne. 2006. Living males of the 'ancient asexual' Darinulidae (Ostracoda: Crustacea). Proc R Soc Lond B 273:1569-1578.

Snoeyenbos-West, O. L. O., T. Salcedo, G. B. McManus, and L. A. Katz. 2002. Insights into the diversity of choreotrich and oligotrich ciliates (Class : Spirotrichea) based on genealogical analyses of multiple loci. Int J Syst Evol Microbiol 52:1901-1913.

Speksnijder, A. G. C. L., G. A. Kowalchuk, S. De Jong, E. Kline, J. R. Stephen, and H. J. Laanbroek. 2001. Microvariation artifacts introduced by pcr and cloning of closely related 16S rRNA gene sequences. Appl Environ Microbiol 67:469-472.

Spiegel, F. W. 1990. Phylum plasmodial slime molds: Class Protostelids. Pages 484-497 *in* Handbook of Protoctista. (L. Margulis, J. O. Corliss, M. J. Melkonian, and D. Chapman, eds.). Jones and Bartlett Publishers, Inc., Boston, MA.

Spiegel, F. W., S. C. Gecks, and J. Feldman. 1994. Revision of the genus *Protostelium* (Eumycetozoa) 1. The *Protostelium mycophaga* group and the *P. irregularis* group. J Euk Microbiol 41:511-518.

Spiegel, F. W., and L. S. Olive. 1978. New evidence for validity of *Copromyxa protea*. Mycologia 70:843-847.

Stamatakis, A. 2006. RAxML-VI-HPC: Maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. Bioinformatics 22:2688-2690.

Stamatakis, A., P. Hoover, and J. Rougemont. 2008. A rapid bootstrap algorithm for the RAxML web servers. Syst Biol 57:758-771.

Stanley, J. S. L. 2005. The *Entamoeba histolytica* genome: something old, something new, something borrowed and sex too? Trends Parasitol 21:451.

StataCorp. 2005. Stata Statistical Software: Release 9. StataCorp LP.

Sukhanova, I. N., and E. A. Cheban. 1990. Heterotrophic phytoplankton of the black-sea in the early spring period of the phytocen development. Okeanologiya 30:979-986.

Sumbali, G. 2005. The Fungi. Alpha Science International, Middlesex, UK.

Suzuki, M., M. S. Rappe, and S. J. Giovannoni. 1998. Kinetic bias in estimates of coastal picoplankton community structure obtained by measurements of small-subunit rRNA gene PCR amplicon length heterogeneity. Appl Environ Microbiol 64:4522-4529.

Suzuki, M. T., and S. J. Giovannoni. 1996. Bias caused by template annealing in the amplification of mixtures of 16S rRNA genes by PCR. Appl Environ Microbiol 62:625-630.

Swofford, D. L. 2000. PAUP*: Phylogenetic Analysis Using Parsimony and Other Methods. Sinauer Associates, Sunderland, MA.

Tekle, Y. I., J. Grant, O. R. Anderson, T. A. Nerad, J. C. Cole, D. J. Patterson, and L. A. Katz. 2008. Phylogenetic placement of diverse amoebae inferred from multigene analyses and assessment of clade stability within 'Amoebozoa' upon removal of varying rate classes of SSU-rDNA. Mol Phylogenet Evol 47:339-352.

Tekle, Y. I., J. Grant, J. C. Cole, T. A. Nerad, O. R. Anderson, D. J. Patterson, and L. A. Katz. 2007. A multigene analysis of *Corallomyxa tenera* sp. nov. suggests its membership in a clade that includes Gromia, Haplosporidia and Foraminifera. Protist 158:457-472.

Thompson, J. R., L. A. Marcelino, and M. F. Polz. 2002. Heteroduplexes in mixed-template amplifications: formation, consequence and elimination by 'reconditioning PCR'. Nucl Acids Res 30:2083-2088.

Thorley, J. L., and R. D. Page. 2000. RadCon: phylogenetic tree comparison and consensus. Bioinformatics 16:486-7.

Todorov, M., V. Golemansky, and R. Meisterfeld. 2010. Is *Difflugia nebeloides* (Amoebozoa: Arcellinida) really a *Difflugia*? Re-description and new combination. Acta Zool Bulgar 62:13-20.

Tolonen, K., B. G. Warner, and H. Vasander. 1992. Ecology of testaceans (Protozoa, Rhizopoda) in mires in Southern Finland .1. Autecology. Arch Protistenk 142:119-138.

Valkanov, A. 1962a. *Euglyphella delicatula* n. g., n. sp. (Rhizopoda Testacea) und ihre Kopulation. Doklad Bolgarkoi Akademii Nauk 15:207-209.

Valkanov, A. 1962b. Über die Kopulation der Testaceen (Rhizopoda Testacea). Doklad Bolgarskoi Akademii Nauk 15:305-308.

van den Ent, F., L. A. Amos, and J. Lowe. 2001. Prokaryotic origin of the actin cytoskeleton. Nature 413:39-44.

von Wintzingerode, F., U. B. Gobel, and E. Stackebrandt. 1997. Determination of microbial diversity in environmental samples: pitfalls of PCR-based rRNA analysis. FEMS Microbiol Rev 21:213.

Wade, R. H., I. Garcia-Saez, and F. Kozielski. 2009. structural variations in protein superfamilies: actin and tubulin. Mol Biotechnol 42:49-60.

Wang, G. C., and Y. Wang. 1997. Frequency of formation of chimeric molecules as a consequence of PCR coamplification of 16S rRNA genes from mixed bacterial genomes. Appl Environ Microbiol 63:4645-4650.

Wang, Y., D. E. Prosen, L. Mei, J. C. Sullivan, M. Finney, and P. B. Vander Horn. 2004. A novel strategy to engineer DNA polymerases for enhanced processivity and improved performance in vitro. Nucl Acids Res 32:1197-1207.

Watkins, R. F., and M. W. Gray. 2008. Sampling gene diversity across the supergroup Amoebozoa: Large EST data sets from *Acanthamoeba castellanii, Hartmannelia vermiformis, Physarum polycephalum, Hyperamoeba dachnaya* and *Hyperamoeba* sp. Protist 159:269-281.

Weismann, A. 1889. Essays on heredity and kindred biological subjects. Oxford University Press, Oxford, UK.

Wenrich, D. H. 1954. Sex in Protozoa: a comparative review. Page 362 *in* Sex in microorganisms. A symposium. (D. H. Wenrich, ed.) AAAS, Washington.

Whatley, J., and C. Chapman-Andresen. 1990. Phylum Karyoblastea. Pages 167-185 *in* Handbook of Protoctista (L. Margulis, J. O. Corliss, M. Melkonian, and D. Chapman, eds.). Jones and Bartlett Publishers, Boston, MA.

Whittaker, R. H. 1969. New concepts of kingdoms of organisms. Science 163:150-160.

Woodcock, H. M. 1916. Observations on coprozoic flagellates. Phil Trans R Soc B 207:375-412.

Woodcock, H. M., and G. Lapage. 1915. Observations on the life-cycle of a new flagellate, *Helkesimastix faecicola*, n. g., n. sp.: together with remarks on the question of syngamy in the Trypanosomes. Proc R Soc B 88:353-370.

Wu, M., J. M. Comeron, H. S. Yoon, and D. Bhattacharya. 2009. Unexpected dynamic gene family evolution in algal actins. Mol Biol Evol 26:249-253.

Wylezich, C., R. Meisterfeld, S. Meisterfeld, and M. Schlegel. 2002. Phylogenetic analyses of small subunit ribosomal RNA coding regions reveal a monophyletic lineage of euglyphid testate amoebae (order Euglyphida). J Euk Microbiol 49:108-118.

Wylezich, C., J. Walochnik, and R. Michel. 2009. High genetic diversity of *Sappinia*-like strains (Amoebozoa, Thecamoebidae) revealed by SSU rRNA investigations. Parasitol Res 105:869-873.

Yoon, H. S., J. Grant, Y. Tekle, M. Wu, B. Chaon, J. Cole, J. Logsdon, D. Patterson, D. Bhattacharya, and L. Katz. 2008. Broadly sampled multigene trees of eukaryotes. BMC Evol Biol 8:14.

Yu, W., K. J. Rusterholtz, A. T. Krummel, and N. Lehman. 2006. Detection of high levels of recombination generated during PCR amplification of RNA templates. Biotechniques 40:499-507.

Zaphiropoulos, P. G. 1998. Non-homologous recombination mediated by *Thermus aquaticus* DNA polymerase I. Evidence supporting a copy choice mechanism. Nucl Acids Res 26:2843-2848.

Zufall, R. A., T. Robinson, and L. A. Katz. 2005. Evolution of developmentally regulated genome rearrangements in eukaryotes. J Exp Zool B 304B:448-455.