Molecular Phylogenetics and Evolution 99 (2016) 144–154

Contents lists available at ScienceDirect

Molecular Phylogenetics and Evolution

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

Phylogenomics of 'Discosea': A new molecular phylogenetic perspective on Amoebozoa with flat body forms

Yonas I. Tekle [a,*], O. Roger Anderson [b], Laura A. Katz [c,d], Xyrus X. Maurer-Alcalá [c,d], Mario Alberto Cerón Romero [c,d], Robert Molestina [e]

[a] Spelman College, 350 Spelman Lane Southwest, Atlanta, GA 30314, USA
[b] Biology, Lamont-Doherty Earth Observatory of Columbia University, Palisades, NY, USA
[c] Department of Biological Sciences, Smith College, Northampton, MA, USA
[d] Graduate Program in Organismic and Evolutionary Biology, University of Massachusetts, Amherst, MA, USA
[e] American Type Culture Collection, Manassas, VA, USA

ARTICLE INFO

Article history:
Received 27 November 2015
Revised 18 March 2016
Accepted 21 March 2016
Available online 22 March 2016

Keywords:
RNA-Seq
Transcriptome
Amoeba
Lobosea
Flabellinea
Tubulinea

ABSTRACT

The majority of amoeboid lineages with flattened body forms are placed under a taxonomic hypothetical class 'Discosea' *sensu* Smirnov et al. (2011), which encompasses some of the most diverse morphs within Amoebozoa. However, its taxonomy and phylogeny is poorly understood. This is partly due to lack of support in studies that are based on limited gene sampling. In this study we use a phylogenomic approach including newly-generated RNA-Seq data and comprehensive taxon sampling to resolve the phylogeny of 'Discosea'. Our analysis included representatives from all orders of 'Discosea' and up to 550 genes, the largest gene sampling in Amoebozoa to date. We conducted extensive analyses to assess the robustness of our resulting phylogenies to effects of missing data and outgroup choice using probabilistic methods. All of our analyses, which explore the impact of varying amounts of missing data, consistently recover well-resolved and supported groups of Amoebozoa. Our results neither support the monophyly nor dichotomy of 'Discosea' as defined by Smirnov et al. (2011). Rather, we recover a robust well-resolved clade referred to as Eudiscosea encompassing the majority of discosean orders (seven of the nine studied here), while the Dactylopodida, Thecamoebida and Himatismenida, previously included in 'Discosea,' are non-monophyletic. We also recover novel relationships within the Eudiscosea that are largely congruent with morphology. Our analyses enabled us to place some *incertae sedis* lineages and previously unstable lineages such as *Vermistella*, *Mayorella*, *Gocevia*, and *Stereomyxa*. We recommend some phylogeny-based taxonomic amendments highlighting the new findings of this study and discuss the evolution of the group based on our current understanding.

1. Introduction

The taxonomy and phylogeny of amoeboid lineages currently classified within the eukaryotic clade Amoebozoa have been challenging to study (Adl et al., 2012; Amaral Zettler et al., 2000; Cavalier-Smith et al., 2004; Lahr et al., 2011a; Smirnov et al., 2005, 2011; Tekle et al., 2008). While one of the main hurdles is scarcity of diagnosable morphological characters in this group (Lee et al., 1985; Page, 1987; Rogerson and Patterson, 2002; Schaeffer, 1926), other challenges are the under sampling of taxa, and particularly insufficient gene sampling in molecular phylogenetic studies (Amaral Zettler et al., 2000; Lahr et al., 2011a; Tekle et al., 2008). While making significant advances, studies based on a single or few molecular gene sequences resolve mainly shallow rather than deep nodes (Lahr et al., 2011a; Tekle et al., 2008). Moreover, they support only some of the well-established relationships based on morphology (Lahr et al., 2011a; Smirnov et al., 2011).

Molecular phylogenetic studies in Amoebozoa have been steadily growing albeit with limited gene sampling (Lahr et al., 2011a, 2013; Tekle et al., 2008). These studies have contributed to our knowledge of the taxonomic breadth of the Amoebozoa, which is far more diverse than originally conceived. Amoebozoa now includes reticulate/filose and flagellate amoeboid forms (Berney et al., 2015; Lahr et al., 2012; Nikolaev et al., 2006; Tekle et al., 2008), in addition to the Mycetozoa (i.e. slime molds) and diverse naked and testate lobose amoebae. Despite these major

* Corresponding author.
E-mail address: ytekle@spelman.edu (Y.I. Tekle).

http://dx.doi.org/10.1016/j.ympev.2016.03.029
1055-7903/

advances, the monophyly and deep relationships of most of the amoebozoan subclades, hypothesized largely from morphological data, remain unresolved (Kudryavtsev et al., 2014; Lahr et al., 2011a; Tekle et al., 2008). Some confounding factors contributing to the failure of the commonly used genes (e.g. SSU-rDNA and actin) in reconstructing phylogenies in amoebozoans is related to long-branch attraction (Felsenstein, 1978) and complicated gene history (Lahr et al., 2011b). The SSU-rDNA genes of some amoebozoans are amongst the fastest evolving in eukaryotic lineages (Tekle et al., 2008), which is an impediment to phylogenetic reconstruction methods (Lartillot et al., 2007; Pisani, 2004). Similarly, presence of multiple copies of actin (i.e. paralogs) in amoebozoans hinders accurate species tree inferences (Lahr et al., 2011b).

Resolving deep relationships in Amoebozoa requires increased gene sampling in an effort to amass phylogenetic signal over noise, which might help circumvent some of the above confounding factors. Large scale data from Amoebozoa such as whole genome, transcriptome and expressed sequence tag (EST) are sorely lacking, and limited only to a few lineages (Eichinger et al., 2005; Grant and Katz, 2014; Stanley, 2005); although this will very likely change in the foreseeable future as EST projects focusing on diverse taxonomic groups are steadily emerging (Cavalier-Smith et al., 2015; Grant and Katz, 2014). A recent multigene study of amoebozoans using transcriptome data, with limited taxonomic sampling, showed promise in resolving deep relationships within the group (Cavalier-Smith et al., 2015). In the study reported here, we used a large-scale analysis approach to resolve the phylogeny of one of the most problematic Amoebozoa subclades ('Discosea') encompassing the majority of flat amoebae.

Lobose amoebae are grouped into two large clusters: 'Discosea' and Tubulinea, primarily based on morphology plus limited molecular data (Smirnov et al., 2011). While the Tubulinea has consistently received some support in molecular phylogenetic analyses (Lahr et al., 2011a, 2013; Tekle et al., 2008), the 'Discosea' as it is currently defined is not recovered in molecular studies (hence we place it in quotes to indicate uncertainty). 'Discosea' was originally erected based on limited molecular data sampling and loosely defined morphological data (Cavalier-Smith et al., 2004). Even though subsequent studies with increased taxon sampling and refined morphological data have improved our understanding of the 'Discosea' (Kudryavtsev et al., 2014; Lahr et al., 2011a; Nikolaev et al., 2006; Smirnov et al., 2005; Tekle et al., 2008), it remains one of the most contentious groups within the Amoebozoa (Smirnov et al., 2011). 'Discosea' *sensu* Smirnov et al. (2011) is broadly defined as diverse groups of flattened naked amoebae with polyaxial cytoplasmic flow. The group encompasses nine morphotypes of lobose amoebae, compared to three that are found within Tubulinea (Smirnov et al., 2011). As the result of this huge diversity, 'Discosea' is defined largely by the absence of characters that are used to distinguish it from the Tubulinea, rather than by a unifying morphological synapomorphy. Thus, it is unclear whether 'Discosea' is a natural clade, or a paraphyletic collection of lineages.

'Discosea' has historically been revised repeatedly and its members redefined, as evidence has emerged from analyses with different taxon sampling and reevaluation of morphological characters (Cavalier-Smith et al., 2004; Smirnov et al., 2005, 2011). Smirnov et al. (2011) recognized two discosean subclades, Longamoebia and Flabellinia, based on morphology. Longamoebia includes flat amoebae with pointed subpseudopodia and elongated cell shape, while members of Flabellinia are discoid to fan-shaped without pointed subpseudopodia (Smirnov et al., 2011). Moreover, some members of Longamoebia possess centrosomes, while this feature is not reported in Flabellinia (Smirnov et al., 2011). The dichotomy of 'Discosea' into these two subclades, Longamoebia and Flabellinia, is rarely recovered in molecular analyses involving SSU-rDNA or actin (Lahr et al., 2011a; Tekle et al., 2008). Similarly, the phylogenetic affinity of several well-characterized taxa within the discosean subclades is inconsistent (Kudryavtsev et al., 2014; Lahr et al., 2011a; Tekle et al., 2008). 'Discosea' also has the highest number of *incertae sedis* (lineages of unknown taxonomic placement), 14 of 24, in one recent taxonomy of Amoebozoa (Adl et al., 2012). This list does not include several putative members of 'Discosea' that never or rarely form a group, including members of Dermamoebida (e.g. *Mayorella*, *Dermamoeba*), Vannellida (e.g. *Pessonella*), Himatismenida (e.g. *Cochliopodium*, *Parvamoeba*, *Gocevia*), Trichosida (e.g. *Trichosphaerium*), and Stygamoebida (e.g. *Vermistella*, *Stygamoeba*) (Berney et al., 2015; Kudryavtsev et al., 2014; Lahr et al., 2011a; Tekle et al., 2008).

In this study we seek to resolve the phylogeny of the 'Discosea' including its monophyly, dichotomy into Longamoebia and Flabellinia, and placement of enigmatic taxa using large-scale analysis. We analyze a total of 40 amoebozoans, including 24 putative discoseans, from newly characterized RNA-Seq data and existing EST and genomic data using probabilistic methods. Our findings provide new and deeper insights into the taxonomic composition and evolution of the 'Discosea'. We also report evidence for several novel relationships and find a taxonomic home for some putative discosean *incertae sedis*. Finally, we make taxonomic amendments based on our current understanding of the group.

2. Materials and methods

2.1. Cultures

A list of the six amoebae strains characterized for this study, including culture conditions and description of bacteria used as food sources during cultivation, are provided (Table S1). *Clydonella* sp. ATCC® 50884™, *Gocevia fonbrunei* ATCC® 50196™, *Vermistella antarctica* ATCC® PRA-216™, and *Parvamoeba monoura* ATCC® PRA-35™ were grown in a minimum of eight 75 cm^2 plastic culture flasks to obtain enough starting material for RNA isolation. *Thecamoeba quadrilineata* ATCC® PRA-259™ and *Unda schaefferi* ATCC® 50810™ were grown in a minimum of four 10 cm diameter agar plates. Cultures were incubated for 4–8 days. Adherent amoebae were removed from plastic culture flasks or agar plates and centrifuged at 300*g*. Pellets were resuspended in 600 µl of buffer RLT (Qiagen®) and frozen at −80 °C. The minimum cell counts used in RNA extraction were $\geqslant 5 \times 10^6$ cells for ATCC® PRA-259™, ATCC® PRA-216™, and ATCC® 50196™, and $\geqslant 2 \times 10^7$ cells for ATCC® PRA-35™, ATCC® 50884™, and ATCC® 50810™.

2.2. RNA isolation

Total RNA from lysates above was isolated using the Qiagen® RNeasy® Mini Kit (QIAGEN group, USA) according to the manufacturer's instructions. The concentration of the RNA samples was measured using a Qubit® RNA HS Assay Kit and Qubit® 2.0 fluorometer (Life Technologies, Carlsbad, CA). Integrity of RNA was evaluated using the Agilent RNA 6000 Pico Assay kit and an Agilent 2100 Bioanalyzer (Agilent Technologies, USA). Electropherograms obtained from the Agilent 2100 Bioanalyzer were inspected visually. RNA concentration, 28S:18S ratio, and RNA integrity number (RIN) were obtained with the Agilent 2100 Expert Software.

2.3. Preparation of libraries and sequencing

A total RNA input of between 0.5 and 4 µg was used in library preparation. Libraries were prepared using the TruSeq Stranded mRNA Library Prep kit (Illumina, San Diego, CA). This kit includes a poly(A) + RNA selection step using oligo (dT) magnetic beads, fragmentation of RNA, followed by first- and second-round cDNA