The complete mitochondrial genome from an unidentified Phalansterium species.

Abstract
We describe the complete sequence and organization of the mitochondrial genome from an unidentified species of Phalansterium. This is the first sequenced mitochondrial genome of a member of Variosea clade (Amoebozoa, Conosa). The sequence was assembled from shotgun reads of DNA from a mixed culture containing the euglenid Monomorphae aenigmatica and an amoebozoan that we demonstrate here is closely related to Phalansterium (in nuclear SSU rRNA phylogenies, it branches between two sequences from described species of Phalansterium). Sequence assembly resulted in two distinct mitochondrial genome types, one fragmented and euglenid-like, and the second a single circular-mapping contig of 53,614 bp with an amoebozoan-like set of genes. The Phalansterium sp. mitochondrial genome is gene-rich and densely packed, with a large number of tRNAs and an unusually low ratio of identifiable protein-coding genes to unidentified ORFs. These ORFs potentially encode ribosomal proteins exhibiting a divergent character at the sequence level, and whose identification may be hindered by the presence of RNA editing in Phalansterium mitochondria, as inferred from numerous acceptor stem mis-matches typical of amoebozoan tRNA 5’ editing.

Keywords
Phalansterium • Amoebozoa • Variosea • Mitochondrion • Genome

Introduction
The genus Phalansterium was established by Cienkowsky [1] for two species of colonial uniciliate flagellates: Monas consociatum Frenesius 1858, renamed Phalansterium consociatum, and newly described species P. intestinum, transferred eight years later to the genus Spongomonas because it was shown to have two closely apposed flagella instead of one [2]. Cienkowsky observed that, together, Phalansterium species fashion their living place together almost like a society in ‘heaps of jelly’, a palace-like building capable of housing four hundred families imagined in the early 19th century by French utopian socialist Charles Fourier as part of his ‘Théorie Sociale’ [3]. The genus Phalansterium currently unifies four species - Phalansterium digitatum [2], P. solitarium [4], P. filosum (Cavalier-Smith and Chao, in [5]) and the type species P. consociatum [1]. Of those, the first three species are relatively well studied (Cavalier-Smith and Chao in [5], [6,7]) whereas P. consociatum on the other hand was never reliably re-isolated and Hibberd [6] suggested that it may be co-specific with P. digitatum.

Phalansterium cells are uni-flagellate, with the flagellum partially surrounded by a collar-like structure and associated with a single basal body that is situated at the apex of a radiating cone of microtubules [6,7]. This relatively simple structure led to the proposal that Phalansterium might represent a primitive, early-branching lineage [8]. However, molecular studies show that Phalansterium branches within Amoebozoa, in the clade named Variosea within the subphylum Protamoebae [9]. The phylogenetic position of Variosea was not clear for some time but members of this clade tend to group with archamoebae and mycetozoans rather than with typical lobose amoebae [10–13]. Cavalier-Smith [14] suggests transfer of the class Variosea to the subphylum Conosa Cavalier-Smith 1998, unifying them with two major amoebozoan groups, mycetozoans (dictyostelids, physarids and some of protostelids) and archamoebae (pelobions and Entamoeba). Close relationships of Variosea with other Conosa were confirmed in further studies [15–20] and also are recognised in the systems of amoebae by Smirnov and Cavalier-Smith [5].

The mitochondria of Variosea are of some interest because members of one of the major conosean groups, Mycetozoa, have relatively normal mitochondria, while members of the other group, Archamoebae, are anaerobes possessing highly reduced mitosomes [21]. However few mitochondrial genomes of Amoebozoa are sequenced and most are from mycetozoans while other groups are virtually not covered [22–25]. Here we describe the complete mitochondrial genome from an unidentified species.
species of *Phalansterium* (the first sequenced mtDNA from a variosean representative) found as a contaminating species in a culture of a photosynthetic euglenid.

**Materials and methods**

**Genome sequencing.**

UTEX 1284 culture containing *Monomorpha aenigmatica* (and subsequently also found to contain the *Phalansterium* species) was obtained directly from the University of Texas Culture Collection, cultured, and purified DNA prepared for sequencing as described [26]. 320,000,000 filtered reads corresponding to 16,000,000 pairs of raw sequence were assembled in parallel with Ray [27] version 2.0.0 rc4, as described [26]. From the 1,849,947 contigs generated (546,420,088 bp total), 130,292 contigs larger than 500 bp (241,607,848 bp total) were searched for sequences similar to genes expected to be encoded in a mitochondrial genome. Two general types of contig resulted: small linear mapping contigs with genes sharing high similarity to *Euglena gracilis* mitochondrial genes, and a single, 53,614 bp circular mapping contig (71X coverage) with sequences sharing closest similarity to amoebozoan mitochondria.

**Genome annotation.**

Open reading frames were located on the circular-mapping *Phalansterium* sp. mitochondrial genome sequence with Artemis 14 [28] and known genes identified by BLAST homology searches [29] against the NCBI non-redundant databases. ORFs showing low or no similarity with known sequences were further queried against the NCBI databases using iterative PSI-BLAST searches [30]. Genes coding for tRNAs were located by BLAST homology searches. Artemis annotations were converted to TBL format using the built-in tools, the TBL file further manually curated, and the accession number generated with NCBI’s TBL2ASN. The physical map was generated from the accession number using OGDRAW 1.2 [32]. The *Phalansterium* sp. mitochondrial genome has been deposited in GenBank under accession number KC121006.

**Phylogenetic analysis.**

The number of species present in what was thought to be a uni-eukaryotic culture was inferred by examining the phylogeny of all sampled nuclear small subunit ribosomal RNA (SSU rRNA) genes. From 546,420,088 bp of sequence, three types were found: a 10.8 kb contig (99,073 reads; 910X coverage) corresponding to the euglenid *Monomorphe aenigmatica*, a 5.8 kb contig (3,021 reads, 52X coverage) closely related to *Drosophila melanogaster*, and a 8.5 kb contig (1,573 reads; 19X coverage) with the closest match to *Phalansterium* (deposited under GenBank accession number KFS539978).

*Phalansterium*-like sequences were aligned to an extensive alignment covering all major groups of eukaryotes using SeaView 4 [33]; the alignment was manually polished and Maximum Likelihood phylogenies were inferred with PhyML under the LG+I+Γ model of amino acid substitution.

**Results and discussion**

A complete mitochondrial genome from an amoebozoan related to *Phalansterium*.

We recently reported the complete plastid genome from shotgun sequencing of *Monomorpha aenigmatica* strain UTEX 1284 [26]. This culture is not axenic, so we used the shotgun sequence data to assess its species diversity by characterizing all nuclear SSU rRNA-encoding contigs. From 320,000,000 filtered reads assembled into 130,292 contigs over 500 bp in length (corresponding to 241,607,848 bp of sequence), only three distinct nuclear SSU rRNA types could be found: that of *M. aenigmatica*, a 5.8 kb fragment related to *Drosophila melanogaster*, and a third type corresponding to a novel sequence closely related to *Phalansterium solitarium*. Based on this deep sequencing survey, we conclude the culture was dominated by *M. aenigmatica* and that the other two organisms were low-level contaminants, with the novel strain of *Phalansterium* present at about 2% of the dominant species (the few *Drosophila* sequences are interpreted as a low level sequencing contaminant as is common with such volumes of sequence).

Phylogenetic analyses on a representative alignment containing all major groups of eukaryotes (tree not shown)
The complete mitochondrial genome from an unidentified Phalansterium species.

indicating that the newly identified Phalansterium SSU sequence robustly branches within the Amoebozoa, where it always clusters (with 100% support by all methods used) with two named Phalansterium species. This position is genuine, and is confirmed by analyses performed with an expanded number of amoebozoan sequences (Figure 1). In the latter tree, Phalansterium species appear as a group inside the Variosea clade, with Filamoeba species and a number of environmental sequences as nearest neighbors. This position is not supported statistically, however, and the Phalansterium clade sometimes appears basal to the entire Variosea. Overall in our analysis, Variosea, together with dicystostelids, mixogastrians and archamoebians

Figure 1. 18S rDNA phylogenetic tree showing the position of Phalansterium sp. (present strain) amongst Amoebozoa. PP/Bootstraps over 0.5/50% are indicated (even if only one of the two values exceeds the thresholds); dashes indicate branches not reproduced in the corresponding Bayesian or ML analyses.
form a monophyletic group corresponding to Conosa while Lobosa appears paraphyletic. All major groups of Amoebozoa are represented in our trees in the same configuration as in Smirnov et al. [5] and the class Discosea remains paraphyletic. The new sequence branched within the Phalansterium lineage, specifically related to Phalansterium solitarium to the exclusion of Phalansterium filosum with complete statistical support, but showing clear differences with both species (the new sequence differs from existing ones at 1.8% and 5.5%, whereas they differ from one another by 6.3%). The phylogeny therefore supports the conclusion that the novel strain represents a new species of Phalansterium, and we henceforth refer to it (in the absence of morphological data allowing description and naming) as Phalansterium sp.

Searching the assembled sequence (filtered for Drosophila contamination) for contigs derived from mitochondrial genomes yielded not one, but two distinct mitochondrial types, as expected. The first is a collection of small linear contigs encoding genes with high similarity to mitochondrial genes from Euglena gracilis. We infer that this collection is derived from the M. aenigmatica mitochondrial genome (not shown). The second type formed a single, gene-rich 53,614 bp circular-mapping contig (Figure 2). Phylogenetic inferences based on the analysis of mitochondrial cytochrome b (Figure 3) supported a close relationship between this protein and amoebozoan homologues, and we accordingly infer the genome encoding this protein to be that of Phalansterium sp.

The Phalansterium sp. mitochondrion has an ancestral, gene-rich genome.

The Phalansterium sp. mitochondrial genome is in some ways similar to that of other amoebozoans, but in others is unique (Figure 2, Table 1). It falls within the size range of other sequenced amoebozoan mitochondrial genomes, is slightly less AT-rich, and contains no identifiable introns (Table 1). It is gene-rich, with a similar coding density as other amoebozoan mtDNAs, but exhibits an unusually low ratio of identifiable protein-coding genes to ORFs. Only 19 protein-coding genes could be identified as
The complete mitochondrial genome from an unidentified Phalansterium species.

Figure 3. Phylogenetic position of the Phalansterium sp. cytochrome oxidase b mitochondrial protein (Cob). The figure shows a zoom-in of the best ML tree computed with PhyML under the LG+F+F model of amino acid substitution (the full tree is available in Nexus format in Data S1). The dashed line indicates the junction with the rest of the tree (not shown). Bootstrap support over 50% is shown above the major nodes, with asterisks for 100%. 
null
structure modeling of tRNA gene sequences has proven to be a powerful predictor of mitochondrial tRNA 5’ editing [48]. Of the 24 tRNAs encoded by Phalansterium mtDNA, we infer on this basis that at least 16 require post-transcriptional 5’ editing to generate the functional species. The almost certain existence of tRNA 5’ editing in Phalansterium mitochondria argues that this system was already established in a last common amoebozoan ancestor.

Overall, the Phalansterium mitochondrial genome might be summed up as having an ‘ancestral’ architecture (according to the classification of mitochondrial genome architecture types proposed in [49]), but populated with highly derived genes.

Acknowledgements

This work was supported by a grant (227301) from the Natural Science and Engineering Research Council of Canada (NSERC) to PJK. PJK and MWG are Fellows of the Canadian Institute for Advanced Research.

References

content and genome organization, J. Mol. Biol., 1995, 245, 522–537


