Molecular Phylogenetics and Evolution 70 (2014) 314–322

Contents lists available at ScienceDirect

Molecular Phylogenetics and Evolution

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

Dinoflagellate phylogeny revisited: Using ribosomal proteins to resolve deep branching dinoflagellate clades

Tsvetan R. Bachvaroff [a,*], Sebastian G. Gornik [b], Gregory T. Concepcion [c], Ross F. Waller [b], Gregory S. Mendez [c], J. Casey Lippmeier [d], Charles F. Delwiche [c]

[a] Smithsonian Environmental Research Center, 647 Contees Wharf Rd., Edgewater, MD 21037, United States
[b] School of Botany, University of Melbourne, Victoria 3010, Australia
[c] University of Maryland College Park, Maryland Agricultural Extension, College Park, MD 20742, United States
[d] DSM, 6480 Dobbin Rd., Columbia, MD 21045, United States

ARTICLE INFO

Article history:
Received 6 December 2012
Revised 24 September 2013
Accepted 7 October 2013
Available online 14 October 2013

Keywords:
Dinoflagellate
Alveolate
Heterokont
Apicomplexan
Ribosomal protein

ABSTRACT

The alveolates are composed of three major lineages, the ciliates, dinoflagellates, and apicomplexans. Together these 'protist' taxa play key roles in primary production and ecology, as well as in illness of humans and other animals. The interface between the dinoflagellate and apicomplexan clades has been an area of recent discovery, blurring the distinction between these two clades. Moreover, phylogenetic analysis has yet to determine the position of basal dinoflagellate clades hence the deepest branches of the dinoflagellate tree currently remain unresolved. Large-scale mRNA sequencing was applied to 11 species of dinoflagellates, including strains of the syndinean genera *Hematodinium* and *Amoebophrya*, parasites of crustaceans and dinoflagellates, respectively, to optimize and update the dinoflagellate tree. From the transcriptome-scale data a total of 73 ribosomal protein-coding genes were selected for phylogeny. After individual gene orthology assessment, the genes were concatenated into a >15,000 amino acid alignment with 76 taxa from dinoflagellates, apicomplexans, ciliates, and the outgroup heterokonts. Overall the tree was well resolved and supported, when the data was subsampled with gblocks or constraint trees were tested with the approximately unbiased test. The deepest branches of the dinoflagellate tree can now be resolved with strong support, and provides a clearer view of the evolution of the distinctive traits of dinoflagellates.

1. Introduction

Alveolates include three major lineages, the ciliates, dinoflagellates and apicomplexans (Gajadhar et al., 1991; Taylor, 1999; Burki et al., 2007). The dinoflagellates are notable primary producers, especially in marine environments, and the apicomplexans are known as parasites, particularly the malaria-causing genus *Plasmodium*, which infects 149 million to 274 million and kills 537,000–907,000 individuals annually (WHO, 2010). The third alveolate group, the ciliates, is most notable for the diversity of their habitats and unusual cell biology including dual nuclei, one germinal and the other somatic. The alveolates, in turn, are most closely related to the heterokonts, a very diverse group ranging from members of the human gut flora, plant pathogens, to the photosynthetic diatoms and the giant kelps (Baldauf, 2003; Burki et al., 2007; Parfrey et al., 2010).

Within the alveolates, the dinoflagellates and apicomplexans are more closely related, and the area between them has been one of recent discovery which confounds simple interpretations of the evolution of these groups (Okamoto et al., 2012; Leander et al., 2003). For example, while parasitic apicomplexans are known to harbor a non-photosynthetic plastid, the discovery of *Chromera velia* demonstrates that photosynthetic members of the apicomplexan lineage are extant (Moore et al., 2008; Keeling, 2013). Meanwhile, in the dinoflagellates, best known as free-living photosynthetic autotrophs, the non-photosynthetic oyster parasite *Perkinsus marinus* diverges from the base of the dinoflagellate lineage (Bachvaroff et al., 2011; Reece et al., 1997; Saldarriaga et al., 2003). Clearly both *C. velia* and *P. marinus* have the potential to independently lose or gain characteristics, but at the simplest level the lifestyles of the deepest branching members of the apicomplexan and dinoflagellate clades strongly contrast with the more familiar members of these lineages.

Simultaneously with the description of new species between apicomplexans and dinoflagellates has been the discovery of an astonishing breadth and abundance of sequences attributable to 'marine alveolates' from marine environmental clone libraries. At a first approximation many of these sequences are placed with known syndinean dinoflagellates in phylogenies, although the raw abundance of such sequences (>1000 in GenBank) dwarfs

* Corresponding author. Present address: Institute for Marine and Environmental Technology, 701 E. Pratt St., Baltimore, MD 21202, United States.
E-mail address: bachvarofft@gmail.com (T.R. Bachvaroff).

1055-7903/$ - see front matter © 2013 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.ympev.2013.10.007

the tens of sequences attributed to described syndinean species or genera (Bachvaroff et al., 2012). The relationships between Marine Alveolate clades I–VIII are not resolved. Indeed, all Marine Alveolate clades may not be syndinean dinoflagellates or parasites, although certainly clades I, II and IV contain syndinean taxa (Bachvaroff et al., 2012; Coats and Bachvaroff, 2012; Harada et al., 2007; Skovgaard et al., 2005, 2009). In the present study we define two major lineages of dinoflagellates, the syndineans and the core dinoflagellates (Hoppenrath and Leander, 2010; Okamoto et al., 2012). The term core dinoflagellate is used in preference to what were formally called the dinokaryotes, since recent studies have cast doubt on the synapomorphies of the dinokaryon (Gornik et al., 2012; Sano and Kato, 2009).

The dinokaryotic state lacks a strict definition, but could be defined as a nucleus with chromosomes condensed throughout the cell cycle, a very low basic protein:DNA ratio, and lack of bulk DNA packaging into nucleosomes. Together these characteristics produce an 'arched fibrillar' appearance of the DNA in transmission electron micrographs of dinokaryote chromosomes (Taylor, 1989). Also, these features appear to be correlated with a high degree of gene duplication (Bachvaroff and Place, 2008; Bachvaroff et al., 2009; Shoguchi et al., 2013).

The outlying species *Oxyrrhis marina* has features reminiscent of dinokaryotes including high DNA content, 'conspicuously banded' chromosomes, and multiple gene copies (Sano and Kato, 2009). In recent reviews on the evolution of the dinokaryon *O. marina* is placed just outside of the core dinoflagellates (Saldarriaga et al., 2004; Wisecaver and Hackett, 2011). Such a placement, however, disagrees with other taxonomic treatments that place *O. marina* outside of both the syndineans and core dinoflagellates based on cell morphology and flagellar arrangement (Adl et al., 2005; Fensome et al., 1993). Clearly independent phylogenetic assessment of *O. marina* is warranted to resolve this discordance.

Well-defined relationships between *P. marinus*, *O. marina*, syndineans and core dinoflagellates are essential to interpreting the evolution of distinctive characters found in dinoflagellates, notably the state of the dinokaryon. In pursuit of this used 'next-generation' sequencing data acquired from two syndineans and their dinoflagellate hosts and another 8 novel datasets from core dinoflagellate taxa, in combination with existing data from *P. marinus* and *O. marina*, to develop basic taxon sampling for the dinoflagellate clade.

Here a specific category of protein-coding genes, the ribosomal proteins, was used to create a phylogeny of the dinoflagellate lineage, other alveolates and heterokonts. These proteins contribute the protein fraction of the ribosome (Nakao et al., 2004). Given that the rRNA may be the most commonly used sequence for nuclear molecular phylogeny, a logical progression would be to use ribosome associated proteins where orthology assignment and horizontal gene transfer may be less of a problem. There are approximately 70+ ribosomal genes in eukaryotes, with some diversification into gene families (Nakao et al., 2004). Many of these genes are quite small, conserved, and highly expressed, providing an easily obtainable fraction of orthologous genes for phylogeny, particularly from EST datasets already available in GenBank. For the core dinoflagellates many genes for ribosomal proteins exist as multiple duplicated gene copies, however most of the differences between gene copies are synonymous, and thus amino acid translations were used (Bachvaroff et al., 2009; Bachvaroff and Place, 2008; Kim et al., 2011).

2. Materials and methods

The photosynthetic dinoflagellates were cultured under autotrophic growth conditions (Table 1). The two parasitic dinoflagellates from the genus *Amoebophrya* used free-living photosynthetic hosts. One was grown on *Karlodinium veneficum*, and the second on *Akashiwo sanguinea* and so are referred to here as *Amoebophrya* sp. ex. *Karlodinium veneficum* and *Amoebophrya* sp. ex *Akashiwo sanguinea* (Gunderson et al., 1999, 2002). Host cultures of ~10,000 hosts ml^{-1} were inoculated with ~100,000 parasite dinospores ml^{-1}. After incubation for 48–72 h, parasite dinospores were isolated from remaining hosts using nucleopore (Whatman, Piscataway, NJ) filters (5 μm for dinospores produced from *K. veneficum* host, and 8 μm for dinospores from *A. sanguinea* host) (Coats and Park, 2002; Park et al., 2002). Parasite cells were pelleted by centrifugation at 10,000*g* for 10 min. Total RNA was isolated using the RNAqueous kit (Ambion, Grand Island, NY) with LiCl precipitation as recommended by the manufacturer. The RNA quality was assessed on the Experion system (BioRad, Hercules, CA). Illumina (San Diego, CA) sequencing was performed by Macrogen with paired end reads of 76 or 100 bases (Table 1). The sequence data were assembled using Trinity for most datasets (Grabherr et al., 2011) or Abyss (Simpson et al., 2009). The choice of assembly program was arbitrary although Trinity required larger memory space computers and longer run times than Abyss. *Hematodinium* sp. ex *Nephops norvegicus* was cultured, its RNA extracted, sequenced and assembled as described in Jackson et al. 2012.

2.1. Assembling orthologous genes

A non-composite strategy was used in this study. Data from individual studies, strains and species were treated as individual taxa. Sequences were downloaded from GenBank using the species-specific taxonomic identifier from refseq, nr, or db_est as appropriate (Supplemental Table T1) and formatted into blast databases, with one database for each species. Similarly, in-house assembled datasets were formatted into blast databases. All species within the heterokonts and apicomplexans with >1000 EST sequences in db_est or a comparable sized nucleotide dataset in the nr database were used.

Sequences for *Nannochloropsis gaditana* were downloaded from http://nannochloropsis.genomeprojectsolutions-databases.com. Sequences from recent publications based on 454 or Illumina sequencing of RNA from dinoflagellates were also downloaded and formatted into local databases for *Symbiodinium* spp. (Bayer et al., 2012), *Alexandrium tamarense* (Moustafa et al., 2010) and *Lingulodinium polyedrum* (Roy and Morse, 2012).

A manually curated set of *Perkinsus marinus* ribosomal proteins was used as a reference query against the individual species' blast databases. Sequences were gleaned from the individual taxa by combining blast search (using the ncbi blast+ suite) with sequence retrieval and translation as necessary using perl scripts. For the sequences from the two *Amoebophrya* host–parasite cultures a total of 10 sequences (if available) from each host–parasite system were harvested. The increased depth for host–parasite datasets was used to ensure both host and parasite versions of each gene were collected. For nucleotide sequences from nr or db_est, or the autotrophic dinoflagellates three sequences were collected (if available) followed by translation into amino acids using the blast hit reading frame and the appropriate genetic code. Amongst the study organisms the ciliates use an alternate translation of the codons TAA and TAG which are translated as Glutamine (Caron and Meyer, 1985). Thus for *Ichthyopthiriius multifiliis*, *Anophryoides haemophila*, *Entodinium caudatum* and *Miamiensis avidis* the ciliate genetic code was used. For *Tetrahymena thermophila* and *Paramecium tetraurelia* the NCBI ref_seq protein database was used so translation was not necessary (see Supplemental Table T1). For amino acid sequences from ref_seq only the best hit was retained. The resulting amino acid sequences were then aligned with clustalo version 1.0.3 (Sievers et al., 2011) using the full iteration