Molecular Phylogenetics and Evolution 70 (2014) 240–243

Contents lists available at ScienceDirect

Molecular Phylogenetics and Evolution

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

Short Communication

Phylogenomic analyses support the bifurcation of ciliates into two major clades that differ in properties of nuclear division

Feng Gao [a,b], Laura A. Katz [b,c,*]

[a] Laboratory of Protozoology, Institute of Evolution & Marine Biodiversity, Ocean University of China, Qingdao 266003, China
[b] Department of Biological Sciences, Smith College, Northampton, MA 01063, USA
[c] Program in Organismic and Evolutionary Biology, UMass-Amherst, Amherst, MA 01003, USA

ARTICLE INFO

Article history:
Received 25 June 2013
Revised 27 September 2013
Accepted 1 October 2013
Available online 9 October 2013

Keywords:
Phylogenomic analysis
Macronucleus
Ciliophora
Postciliodesmatophora
Intramacronucleata

ABSTRACT

Ciliates are a diverse assemblage of eukaryotes that have been the source of many discoveries including self-splicing RNAs, telomeres and trans-splicing. While analyses of ciliate morphology have given rise to robust hypotheses on relatively shallow level relationships, the deeper evolutionary history of ciliates is largely unknown. This is in part because studies to date have focused on only a single locus, small subunit ribosomal DNA (SSU-rDNA). In the present study, we use a taxon-rich strategy based on multiple loci from GenBank and recently completed transcriptomes to assess deep phylogenetic relationships among ciliates. Our phylogenomic data set includes up to 537 taxa, all of which have been sampled for SSU-rDNA and a subset of which have LSU-rDNA and up to 7 protein-coding sequences. Analyses of these data support the bifurcation of ciliates as suggested by SSU-rDNA, with one major clade defined by having somatic macronuclei that divide with intranuclear microtubules (Intramacronucleata) and the other clade containing lineages that either divide their macronuclei with microtubules external to the macronucleus or are unable to divide their macronuclei (Postciliodesmatophora). These multigene phylogenies provide a robust framework for interpreting the evolution of innovations across the ciliate tree of life.

1. Introduction

Ciliates are characterized by the presence of cilia in at least one of their life stages and by their nuclear dimorphism (i.e. the presence of a somatic macronucleus and a germline micronucleus in each cell). Following conjugation, macronuclei derive from zygotic nuclei through chromosomal rearrangements that include fragmentation, elimination of internal sequences and amplification of the processed chromosomes; in all but one class of ciliates (Karyorelictea, see below) the resulting macronuclei divide by amitosis during asexual division (Prescott, 1994). Given their age of approximately one billion years (Parfrey et al., 2011), estimation of deep relationships within this clade are difficult.

The class Karyorelictea Corliss, 1974 had been argued to be sister to all other ciliates based on the relatively simple morphologies and the presence of nearly-diploid, non-dividing macronuclei within this clade (Raikov, 2006). Subsequent phylogenetic analyses of ciliates based only on SSU-rDNA divides ciliates into two clades, Intramacronucleata Lynn, 1996 and Postciliodesmatophora Gerassimova and Seravin, 1976 (Lynn, 1996). The Intramacronucleata includes the bulk of the ciliate classes such as Oligohymenophorea (e.g. *Paramecium* and *Tetrahymena*), and are united by the feature of division of the macronucleus involving intramacronuclear microtubules (Hirt et al., 1995; Lukashenko, 2009; Lynn, 1996). In contrast, ciliates in Postciliodesmatophora either have macronuclei that cannot divide (i.e. Karyorelictea) or macronuclei that divide with microtubules external to the macronucleus (i.e. Heterotrichea).

Given the limitation of single gene trees, we use a taxon-rich strategy based on multiple loci to assess the relationships within ciliates. We expand the taxonomic sampling of SSU-rDNA to 537 species representing all major ciliate lineages and combine these data with large subunit-rDNA and up to seven protein genes from a subset of taxa. We analyzed the full data matrix as well as six submatrices to assess the impact of taxon sampling and missing data.

2. Methods

2.1. Dataset assembly

We collected small subunit ribosomal DNA (SSU-rDNA) and large subunit ribosomal DNA (LSU-rDNA) sequences for all the ciliates from GenBank using a custom Python script. The SSU-rDNA of *Philasterides armatalis* (FJ848877) and LSU-rDNA of *Stylonychia*

* Corresponding author. Address: Department of Biological Sciences, Smith College, 44 College Lane, Northampton, MA 01063, USA. Fax: +1 413 585 3786.
E-mail address: lkatz@smith.edu (L.A. Katz).

1055-7903/$ - see front matter © 2013 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.ympev.2013.10.001

lemnae (AF508773) were used as queries in a Blast analysis against the GenBank nr database and one sequence ⩾ 1000 bp per taxon ID was kept. The taxon IDs of the ciliates from GenBank were downloaded in February 2012. Environmental and uncultured sequences were removed. As our preliminary analyses showed that the sequences of *Mesodidium* and *Myrionecta* formed a long, unstable branch as discussed elsewhere (Strüder-Kypke et al., 2006), we excluded these taxa in our final analyses. The sequences for at most two species per genus and for all the species that have available protein sequences used in the analyses (see below) were kept, resulting in 537 and 111 sequences for SSU-rDNA and LSU-rDNA, respectively. Sequences were aligned in GUIDANCE (Penn et al., 2010b) and ambiguous columns in the alignment were removed with default parameters using GUIDANCE web server (Penn et al., 2010a).

Assembly of the protein-coding gene dataset relied on a custom built pipeline that uses Python scripts to collect homologs from one of three sources: directly downloaded from GenBank, translated from EST data, or translated from transcriptome data. First, in January 2012, we downloaded all 1935 amino acid sequences from Ciliophora, excluding those from *Paramecium*, *Tetrahymena*, and *Ichthyophthirius* as these taxa have complete genome data. We then used Proteinortho4 (Lechner et al., 2011) to bin proteins into orthologous groups. We chose seven proteins that had sequences available from the largest number of species (i.e. Actin, α-tubulin, β-tubulin, cytochrome oxidase subunit 1, elongation

Table 1
The division of ciliates into Intramacronucleata and Postciliodesmatophora is supported by all but one analyses using subsamples of the full data set.

	All: 9	All: 7	2P: 9	2P: 7	3P: 9	3P: 7	SSU-rDNA
Intramacronucleata	89	nm	84	49	85	64	17
Armophorea	75	76	100	99	100	100	73
Litostomatea	100	43	100	100	100	100	100
Spirotrichea	nm	nm	100	88	100	100	nm
Colpodea	43	nm	–	–	–	–	nm
Oligohymenophorea	98	nm	93	nm	93	nm	88
Nassophorea	nm	–	–	–	–	–	nm
Phyllopharyngea	100	57	–	–	–	–	100
Plagiopylea	87	nm	–	–	–	–	54
Prostomatea	nm	nm	–	–	–	–	nm
Postciliodesmatophora	100	nm	100	91	100	89	92
Heterotrichea	100	34	100	100	100	98	65
Karyorelictea	100	nm	–	–	–	–	100

Note: nm = non-monophyletic; All:9 = our full data matrix consisted of 9 genes (7 protein-coding genes plus SSU-rDNA and LSU-rDNA) and 537 taxa; all: 7 = data matrix consisted of 7 protein-coding genes and 537 taxa; 2P:9 = data matrix consisted of the 9 genes and the taxa that had at least 2 proteins; 2P: 7 = data matrix consisted of 7 protein-coding genes and the taxa that had at least 2 proteins; 3P:9 = data matrix consisted of the 9 genes and the taxa that had at least 3 proteins; 3P:7 = data matrix consisted of 7 protein-coding genes and the taxa that had at least 3 proteins; SSU-rDNA = small subunit ribosomal DNA.

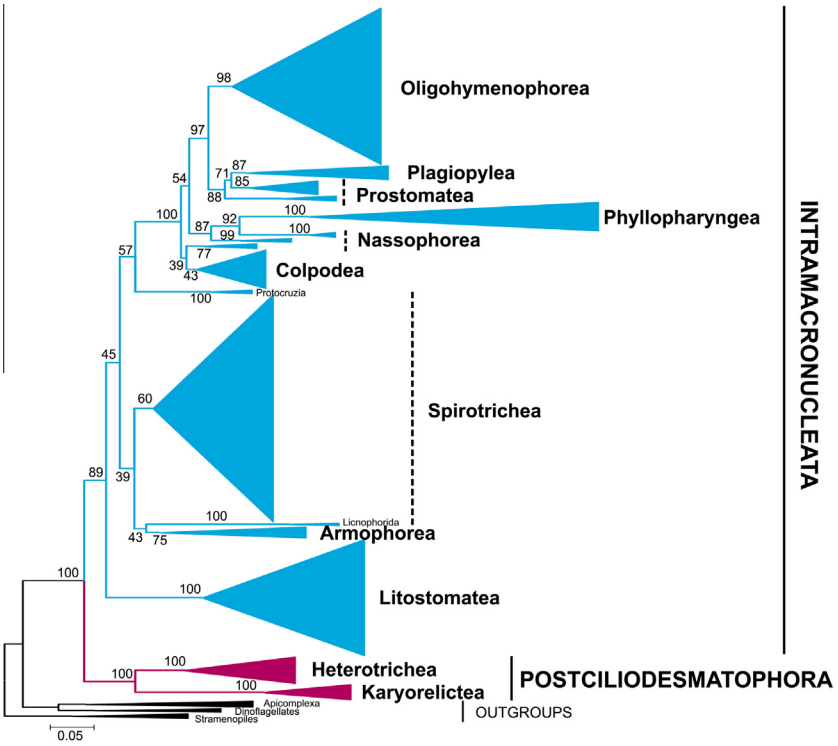

Fig. 1. Maximum likelihood (ML) tree reconstructed using 537 ciliates and all 9 genes (all:9; SSU-rDNA, LSU-rDNA plus 7 protein genes) yields a well-resolved bifurcation of ciliates phylogeny, Intramacronucleata (in blue) and Postciliodesmatophora (in maroon). Numbers at the nodes represent ML support values. The scale bar corresponds to 0.05 expected substitutions per site. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)