Molecular Phylogenetics and Evolution xxx (2014) xxx–xxx

Contents lists available at ScienceDirect

Molecular Phylogenetics and Evolution

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

Large-scale phylogenomic analysis reveals the phylogenetic position of the problematic taxon *Protocruzia* and unravels the deep phylogenetic affinities of the ciliate lineages

E. Gentekaki [a,*], M. Kolisko [a,b], V. Boscaro [c], K.J. Bright [d], F. Dini [c], G. Di Giuseppe [c], Y. Gong [e], C. Miceli [f], L. Modeo [c], R.E. Molestina [g], G. Petroni [c], S. Pucciarelli [f], A.J. Roger [a], S.L. Strom [d], D.H. Lynn [h,*]

[a] Department of Biochemistry and Molecular Biology, Dalhousie University, Halifax, Nova Scotia B3H 4R2, Canada
[b] Department of Botany, University of British Columbia, Vancouver, British Columbia V6T 1Z4, Canada
[c] Department of Biology, University of Pisa, Pisa 56126, Italy
[d] Shannon Point Marine Center, Western Washington University, Anacortes, WA 98221, United States
[e] The Key Laboratory of Aquatic Biodiversity and Conservation of Chinese Academy of Sciences, Institute of Hydrobiology, Chinese Academy of Sciences, Wuhan 430072, PR China
[f] School of Biosciences and Biotechnology, University of Camerino, Camerino, Italy
[g] Protistology Department, American Type Culture Collection, Manassas, VA 20110, United States
[h] Department of Integrative Biology, University of Guelph, Guelph, Ontario N1G 2W1, Canada

ARTICLE INFO

Article history:
Received 10 November 2013
Revised 18 March 2014
Accepted 18 April 2014
Available online xxxx

Keywords:
Ciliophora
Phylogenomics
CONthreeP
Protocruzia
SAL

ABSTRACT

The Ciliophora is one of the most studied protist lineages because of its important ecological role in the microbial loop. While there is an abundance of molecular data for many ciliate groups, it is commonly limited to the 18S ribosomal RNA locus. There is a paucity of data when it comes to availability of protein-coding genes especially for taxa that do not belong to the class Oligohymenophorea. To address this gap, we have sequenced EST libraries for 11 ciliate species. A supermatrix was constructed for phylogenomic analysis based on 158 genes and 42,158 characters and included 16 ciliates, four dinoflagellates and nine apicomplexans. This is the first multigene-based analysis focusing on the phylum Ciliophora. Our analyses reveal two robust superclades within the Intramacronucleata; one composed of the classes Spirotrichea, Armophorea and Litostomatea (SAL) and another with Colpodea and Oligohymenophorea. Furthermore, we provide corroborative evidence for removing the ambiguous taxon *Protocruzia* from the class Spirotrichea and placing it as *incertae sedis* in the phylum Ciliophora.

© 2014 Elsevier Inc. All rights reserved.

1. Introduction

Over the past few years, phylogenomic approaches have been employed to untangle deep relationships among major microbial eukaryotic lineages and place divergent taxa of evolutionary significance (Brown et al., 2012, in press; Burki et al., 2007, 2009, 2012; Hampl et al., 2009; Parfrey et al., 2010; Zhao et al., 2012). Large-scale phylogenomic analyses are now being utilized to resolve questions associated with the shallower nodes of the eukaryotic tree of life (Bachvaroff et al., 2011; Burki et al., 2010). One protistan group where such analyses have never been performed is the phylum Ciliophora. This is mainly because sufficient data exist for only a limited number of taxa (Abernathy et al., 2007; Aury et al., 2006; Eisen et al., 2006; Ricard et al., 2008; Swart et al., 2013). Consequently, phylogenetic inference has been based largely on the 18S ribosomal RNA; however, a single locus is insufficient to infer robust phylogenetic relationships (Gribaldo and Philippe, 2000). Thus, several evolution and taxonomy related questions remain unresolved.

The ciliate tree is divided in two deep lineages – the Postciliodesmatophora and Intramacronucleata, a split that is supported by both molecular and morphological lines of evidence (Baroin-Tourancheau et al., 1998; Embley et al., 1995; Gao et al., 2010; Hirt et al., 1995; Lynn, 1996, 2003). Beyond this deep division, it is generally agreed that there are 11 major ciliate lineages or classes and a twelfth single species clade of Cariacotrichea (Adl et al., 2012; Lynn, 2008; Orsi et al., 2012; Stoeck et al., 2003). Most of these classes are strongly supported by morphology, with the exception of the "riboclasses" – the Armophorea and Plagiopylea, which are identified only by sequences of the 18S rRNA genes, as included taxa lack any morphological synapomorphies (Bernhard et al., 1995, 2001; Cameron et al., 2001; Embley et al., 1995;

* Corresponding authors. Present address: Department of Zoology, University of British Columbia, Vancouver, British Columbia V6T 1Z4, Canada (D.H. Lynn).
E-mail addresses: eleni@dal.ca (E. Gentekaki), lynn@zoology.ubc.ca (D.H. Lynn).

http://dx.doi.org/10.1016/j.ympev.2014.04.020
1055-7903/© 2014 Elsevier Inc. All rights reserved.

Please cite this article in press as: Gentekaki, E., et al. Large-scale phylogenomic analysis reveals the phylogenetic position of the problematic taxon *Protocruzia* and unravels the deep phylogenetic affinities of the ciliate lineages. Mol. Phylogenet. Evol. (2014), http://dx.doi.org/10.1016/j.ympev.2014.04.020

UNCORRECTED PROOF

Greenwood et al., 1991; Leipe et al., 1994; Lynn, 2008; Lynn et al., 1999; Lynn and Strüder-Kypke, 2002; Snoeyenbos-West et al., 2004; Stechmann et al., 1998). Nevertheless, the monophyly of the class Spirotrichea has been challenged in several studies, notably those that include *Protocruzia* spp. While *Protocruzia* has been formally assigned to the Spirotrichea (Lynn, 2008), its phylogenetic position remains as one of the most ambiguous since it is rarely recovered with the Spirotrichea (Bernhard and Schlegel, 1998; Shin et al., 2000; Song and Wilbert, 1997). Some researchers have suggested that *Protocruzia* be assigned its own independent lineage status (Li et al., 2010).

The phylogenetic relationships between the classes of the phylum generally are uncertain, although there is a robust clustering of Colpodea, Oligohymenophorea, Nassophorea, Prostomatea, Plagiopylea and Phylopharyngea, named CONthreeP (Adl et al., 2012; Lynn, 2008). The class Spirotrichea is of uncertain affiliation making it an orphan lineage in the ciliate tree, though some studies do recover a moderately supported cluster with the Litostomatea and Armophorea (da Silva Paiva et al., 2013; Riley and Katz, 2001; Vd'ačný et al., 2010).

To examine whether the relationships between some of the classes differ to those recovered by 18S rRNA phylogenies and to elucidate the phylogenetic position of *Protocruzia*, we increased both taxon and character sampling by obtaining RNAseq data from 11 ciliate taxa. Some of these taxa belong to ciliate lineages for which only limited data are available, namely the Colpodea, Litostomatea and Heterotrichea. Phylogenomic analyses of 158 genes show maximally supported groupings of Colpodea + Oligohymenophorea and Spirotrichea + Armophorea + Litostomatea. Furthermore, our study illustrates that *Protocruzia* is not a spirotrich, though it remains unclear, if it is an independent lineage or a member of the Heterotrichea.

2. Materials and methods

2.1. Identification, isolation and culturing

Details regarding identification, isolation and culturing of individual organisms are provided in Supplementary Information section, Appendix A.

2.2. RNA extraction, cDNA libraries, Illumina sequencing, EST clustering and annotation

Details regarding RNA extraction from individual organisms are provided in Supplementary Information section, Appendix A.

Poly(A)+ RNA was isolated and cDNA libraries with an insert size of ~200 bp were constructed according to the standard protocol of the National Center for Genome Resources (NCGR, http://ncgr.org). The cDNA libraries were then sequenced on an Illumina Hi-Seq 2000 using paired-end Illumina sequencing at NCGR (New Mexico, USA). Raw sequence reads were assembled into clusters and those were subsequently annotated using the standard protocols of NCGR. Further details are provided in Supplementary Information section, Appendix A.

2.3. Basic phylogenomic dataset construction

The 158-gene set used in this study was derived from Brown et al. (2012), (see also Appendix B, Tables S1 and S2 of this manuscript for details on gene and taxon sampling). All genes used in the analyses are encoded in the nucleus. Briefly, protein sequences of *Arabidopsis thaliana* were used as the seed reference dataset. For each gene, a "raw" dataset was assembled as follows: (1) for each taxon up to five sequences per gene were recovered by BLASTP or TBLASTN using the reference dataset as query and with an e-value cut-off of e-10; and (2) well-characterized paralogues (e.g. HSP70) were identified by reciprocal BLAST against the reference dataset.

Subsequently, each "raw" gene dataset was aligned using MAFFT v7.045b (Katoh et al., 2002; Katoh and Toh, 2010) and the ambiguously aligned positions were masked with the Block Mapping and Gathering with Entropy (BMGE v1.1) software (parameters: -g 0.3 -b 5 -m BLOSUM62) (Criscuolo and Gribaldo, 2010). Single gene trees were constructed with FastTree (Price et al., 2009) and each tree was examined by eye to identify and remove paralogues and contaminants. Deep branching paralogues that span the eukaryotic tree (e.g. HSP70, where cytosolic, mitochondrial and endoplasmic reticulum versions exist) were identified by visual inspection and supplemented by BLAST searches. In case of paralogy resulting from gene duplication within a specific group (i.e. gene duplication in metazoans), the clade with the largest taxon sampling was retained. If multiple in-paralogues per species (i.e. multiple sequences from one species forming a clade) were present, the longest sequence or the shortest branching sequence was retained. A single orthologue was kept per taxon.

2.4. Adding ciliate taxa

The assembled contigs of the 11 newly sequenced ciliates were screened for orthologues using TBLASTN (e-value cut-off of e-10) with the reference dataset as query and translated using an in-house script. As ciliates are known to have large numbers of paralogues, the top five hits for each ciliate taxon were retained. Single gene datasets were then aligned using MAFFT v7.045b, ambiguously aligned positions were automatically masked with the BMGE v1.1 software program, and single gene trees were constructed using FastTree. The single gene trees were then inspected visually and paralogous/contaminating sequences were removed. Contaminating sequences were identified using BLAST searches against the GenBank database. In case of ciliate-specific gene duplications, the clade with the largest number of ciliate sequences was retained. As in the construction of the base line dataset, if multiple in-paralogues per species were present (i.e. multiple sequences from one species forming a clade), the longest sequence or the shortest branching sequence was kept.

To additionally test each single gene tree for presence of contaminants and/or paralogues, we generated 100 rapid bootstraps for each single gene alignment using RAxML v7.2.6 (model setting PROTGAMMALGF). Subsequently, we extracted highly supported bipartitions (BP > 70%) and compared them to a multi-furcating eukaryotic consensus tree containing all widely accepted eukaryotic clades (Brown et al., 2012). Highly supported bipartitions that were conflicting with the consensus tree were examined by eye and corrected by additional removing of contaminants or hidden paralogues. This step was repeated until no obvious conflicts remained.

The final single gene datasets were then aligned using MAFFT v7.045b (algorithm linsi) and ambiguously aligned positions were automatically masked with the BMGE v1.1 software program. The masked alignments were then concatenated into a final 42,158 amino acid supermatrix using ALVERT from the BARREL-o-MONKEYS software suite (http://rogerlab.biochemistryandmolecularbiology.dal.ca/Software/Software.htm# Monkeybarrel). From this supermatrix two datasets were constructed: (1) medium size dataset, containing only alveolates (30 species); and (2) large size dataset, containing members from all major eukaryotic super-groups.

2.5. Phylogenomic analyses

Maximum Likelihood (ML) analyses were performed using RAxML v7.2.6 under the LG model of amino acid substitution + Γ

UNCORRECTED PROOF

Please cite this article in press as: Gentekaki, E., et al. Large-scale phylogenomic analysis reveals the phylogenetic position of the problematic taxon *Protocruzia* and unravels the deep phylogenetic affinities of the ciliate lineages. Mol. Phylogenet. Evol. (2014), http://dx.doi.org/10.1016/j.ympev.2014.04.020

distribution (four rate categories) + F (Le and Gascuel, 2008; Stamatakis, 2006). The model was selected using the program ProtTest v2.4 (Abascal et al., 2005). The ML tree topologies were generated via 50 random tree searches in RAxML (as implemented in RAxML). To obtain statistical support, 500 bootstrap replicates were analyzed for each tree (Fig. 1, Appendix C, Fig. S1).

The Bayesian Inference (BI) trees were generated for the mid-size dataset using the program PhyloBayes-MPI 1.4e with the CAT-GTR model + Γ distribution (Lartillot et al., 2009). Four independent chains were run for 20,000 generations (convergence Maxdiff = 0.00996335 with 10% burn-in). Posterior probabilities were computed using the program bpcom as part of the PhyloBayes package (Fig. 2).

2.6. Removal of fast evolving sites

To test for potential phylogenetic artifacts, the fast evolving sites removal analysis was performed. Rates per site were computed using the ML tree in the program Dist_Est (Susko et al., 2003). Sites were then sorted from fastest to slowest evolving. The fastest evolving sites were then sequentially removed in blocks of 1000, until 42,000 sites were removed. This resulted in 42 datasets. These datasets were analyzed by rapid bootstrapping in RAxML v7.2.6 (model setting PROTCATLG) and bootstrap support for nodes of interest was plotted (Appendix B, Table S3).

2.6.1. Access

All transcriptomic data are publicly available through the CAMERA portal (https://portal.camera.calit2.net/gridsphere/gridsphere). The accession numbers are as follows: *Aristerostoma* sp. MMETSP0125-20120918; *Condylostoma magnum* MMETSP0210-20121227; *Euplotes focardii* MMETSP0205-20121125; *Euplotes harpa* MMETSP0213-20121227; *Litonotus* sp. MMETSP0209-20121228; *Platyophrya macrostoma* MMETSP0127-20121128; *Protocruzia adherens* MMETS P0216-20120918; *Pseudokeronopsis riccii* MMETSP0211-20121228; *Schmidingerella arcuata* [*Favella ehrenbergii*] MMETSP0123-2013 0129; *Strombidinopsis acuminatum* MMETSP0126-20121128; *Strombidium inclinatum* MMETSP0208-20121228.

3. Results and discussion

In the present study, we increased the taxonomic breadth of the ciliate clade by integrating data from 11 ciliates for which EST data were not available previously with five pre-existing genomic datasets. Our taxon sampling covered six major ciliate groups – Heterotrichea, Colpodea, Oligohymenophorea, Litostomatea, Armophorea and Spirotrichea. We assembled a 158-gene dataset containing 42,158 amino acid positions. This is the largest "ciliate-based" phylogenetic dataset assembled to date in terms of number of bases included. All previous studies were based either on single or a couple of genes or contained extremely limited sampling of ciliates (Bapteste et al., 2002; Brown et al., 2012; Budin and Philippe, 1998; Burki et al., 2009, 2013; Greenwood et al., 1991;Hampl et al., 2009; Hammerschmidt et al., 1992; Katz et al., 2004; Lynn and Sogin, 1988; Sogin and Elwood, 1986).

All of the analyzed datasets recover maximal statistical support for the monophyly of ciliates. This is in agreement with numerous studies on morphological characters as well as with the results of several previously published works based on single gene phylogenies (Baldauf and Doolittle, 1997; Barroin et al., 1988; Baroin-Tourancheau et al., 1998; Bernhard et al., 1995; Budin and Philippe, 1998; Bütschli, 1887–1889; Chatton and Lwoff, 1935a, 1935b; Elwood et al., 1985; Fauré-Fremiet, 1950; von Gelei, 1932, 1934; Greenwood et al., 1991; Hammerschmidt et al., 1992; Hirt et al., 1995; Israel et al., 2002; Jankowski, 1967, 1973; Katz et al., 2004; Klein, 1928, 1929; Leander and Keeling, 2003; Leipe et al., 1994; Lynn and Sogin, 1988; Philippe and Adoutte, 1998; Sogin and Elwood, 1986).

3.1. Is Protocruzia a spirotrich?

Protocruzia is a marine benthic ciliate with a highly ambiguous taxonomic history. In the first molecular study of its histone H4 and H3 genes, Bernhard and Schlegel (1998) showed that

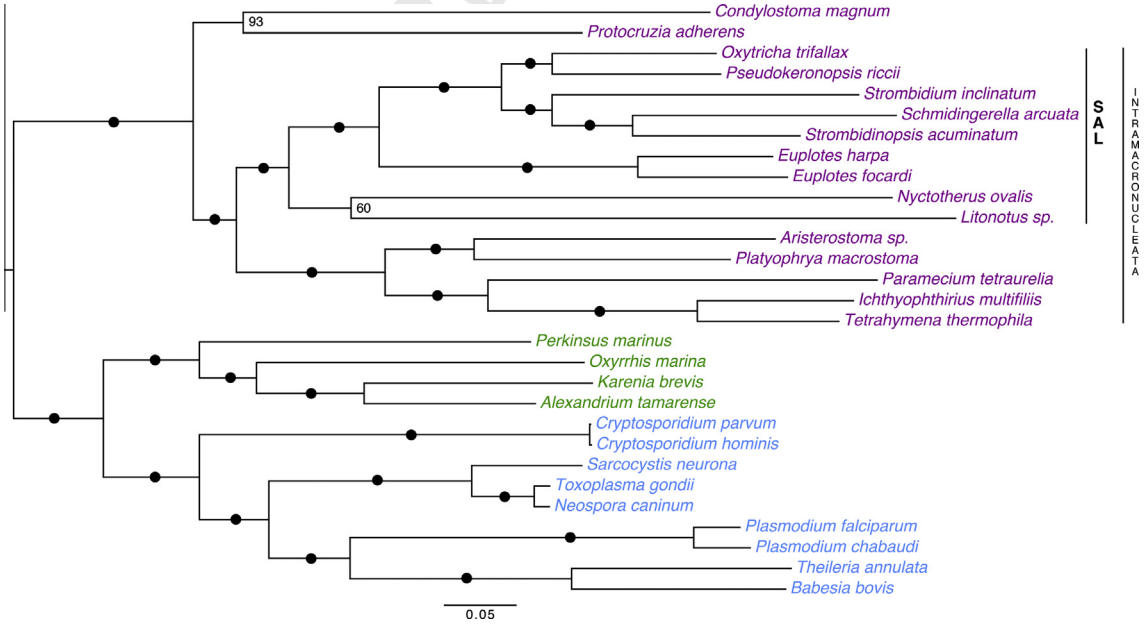

Fig. 1. Phylogenetic tree estimated from a 158 gene dataset inferred by RAxML under the LG + Γ model. The numbers at nodes indicate bootstrap support (BS) values. Solid black circles indicate BS of 100%. The long black line indicates the subphylum Intramacronucleata. The shorter black line marks the newly identified assemblage of Spirotrichea + Armophorea + Litostomatea (SAL).

Please cite this article in press as: Gentekaki, E., et al. Large-scale phylogenomic analysis reveals the phylogenetic position of the problematic taxon *Protocruzia* and unravels the deep phylogenetic affinities of the ciliate lineages. Mol. Phylogenet. Evol. (2014), http://dx.doi.org/10.1016/j.ympev.2014.04.020

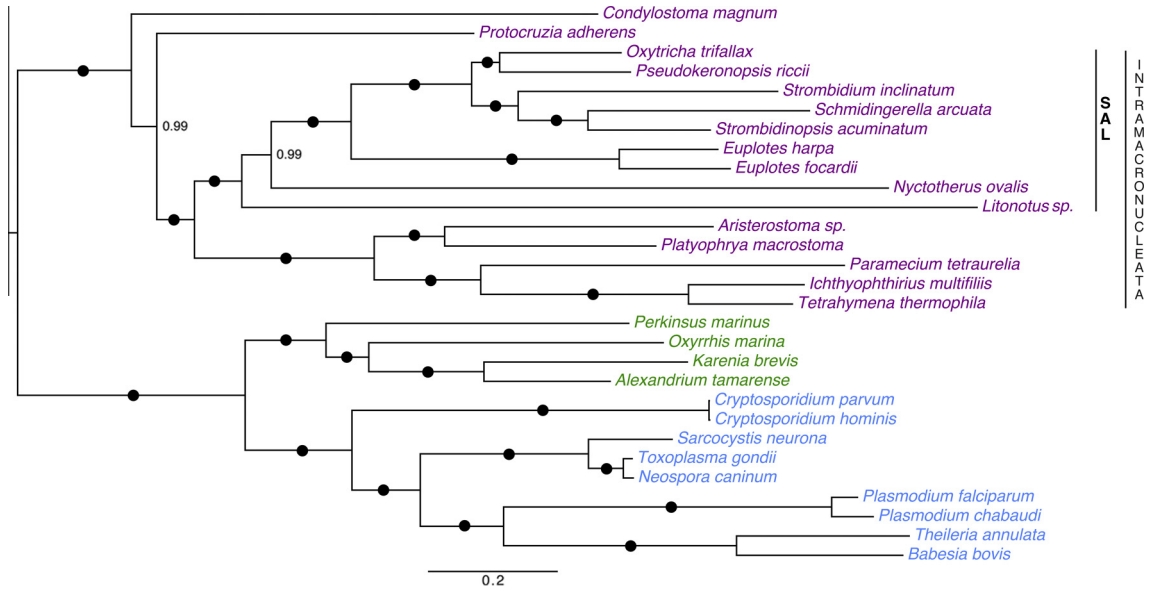

Fig. 2. Phylogenetic tree estimated from a 158 gene dataset inferred by PhyloBayes under the CAT-GTR + Γ model. The numbers at nodes indicate Bayesian posterior probabilities (PPs). Solid black circles indicate PPs of 1.0. The long black line indicates the subphylum Intramacronucleata. The shorter black line marks the newly identified assemblage of Spirotrichea + Armophorea + Litostomatea (SAL).

Protocruzia had an ambiguous position dependent upon whether nucleotide or protein sequences were used. Subsequent studies using 18S rRNA gene sequences showed the genus to be more closely related to Spirotrichea (in a basal position) than Postciliodesmatophora (Hammerschmidt et al., 1996; Shin et al., 2000).

Multiple analyses have since been employed using a limited number of genes but the exact phylogenetic position of *Protocruzia* remains unresolved; it has been considered a karyorelictid, heterotrich and spirotrich (Bernhard and Schlegel, 1998; De Puytorac, 1994; Grolière et al., 1980; Lynn, 1981, 1991; Lynn, 2008; Shin et al., 2000; Small and Lynn, 1981; Song and Wilbert, 1997). Some investigators have proposed that *Protocruzia* be given its own lineage status (Li et al., 2010). Currently, the formal taxonomic placement of the taxon is within the class Spirotrichea as the only species of the subclass Protocruziidia (Lynn, 2003, 2008).

The present study shows that *Protocruzia* is not a member of the class Spirotrichea (Figs. 1 and 2, Appendix C, Fig. S1). Instead, both our analyses place *Protocruzia* in a deeper and earlier diverging position in the ciliate tree. In the ML analysis *Protocruzia* is sister to the heterotrich *Condylostoma* and this relationship is strongly supported (Fig. 1). In the BI analysis the taxon branches after *Condylostoma* indicating an independent lineage separate from both Heterotrichea and Spirotrichea, supporting the conclusion of Li et al., 2010 (Fig. 2). The discrepancy between the ML and BI analyses may suggest that they are affected by long-branch attraction (LBA) (Philippe et al., 2000). To test whether the ML topology might be the result of LBA, we performed removal of fast evolving sites, one of the most common ways for suppressing such artifacts (Brown et al., in press; Hampl et al., 2009; Philippe et al., 2000). Nevertheless, even after doing so, the results of our analyses remained unchanged. Therefore, it is more likely that the differences in the two topologies are due to the phylogenetic models used in our ML (LG model) and BI (CAT model) analyses. Several recent studies that involve large datasets have shown that the CAT model is better fitting and more biologically realistic than LG (Brown et al., in press; Burki et al., 2013; Lartillot et al., 2009). Thus, the topology derived from the BI analysis is in all likelihood the most accurate. It is very likely that improved sampling, especially addition of deep-branching karyorelictid taxa may resolve this part of the ciliate tree and therefore the placement of *Protocruzia*. Until these gene data are available, we place *Protocruzia incertae sedis* in the Phylum Ciliophora as our analyses very strongly show that it is not a spirotrich.

The exclusion of *Protocruzia* from Spirotrichea is further supported by a combination of both morphological and ultrastructural features: the spirotrich-specific S-phase band that passes through the macronucleus during DNA synthesis is absent in *Protocruzia* (Lynn, 2008; Ruthmann and Hauser, 1974). Furthermore, division of the macronucleus in *Protocruzia* exhibits some mitosis-like features, a characteristic that has never been observed in any other spirotrich, and which is in fact unique within the phylum (Ammermann, 1968; Lynn, 2008; Ruthmann and Hauser, 1974). Additional information on its cortical ultrastructure would also be informative. Grolière et al. (1980) clearly showed the presence of overlapping postciliary ribbons. However, the critical feature for systematics is the manner in which the postciliary ribbons overlap, and this information is not provided in the micrographs of Grolière et al. (1980). To be precise, the postciliodesmatophorans have postciliodesmata with either a "2 + ribbon + 1" structure as in the Class Karyorelictea or a "ribbon + 1" structure as in the Class Heterotrichea (Lynn, 2008). To our knowledge, all other overlapping postciliary ribbons of ciliates are not separated by singlet or doublet microtubules. Thus, research on the details of the cortical ultrastructure of *Protocruzia* would provide significant phylogenetic information. Together, these additional morphological data along with a broader taxon sampling of gene sequences would enable resolution of the phylogenetic position of this unusual genus but also shed light on the early evolution of ciliates.

3.2. Relationships between ciliate lineages

All classes that have more than one representative – Colpodea, Oligohymenophorea and Spirotrichea – are recovered as monophyletic and the relationships are strongly supported in all methods of analyses (Figs. 1 and 2). The Colpodea + Oligohymenophorea clade is very strongly supported confirming previous studies. This clade is part of a bigger assemblage that comprises six ciliate lineages: Colpodea + Oligohymenophorea + Nassophorea + Plagiopylea + Prostomatea + Phyllopharyngea (CONthreeP). CONthreeP is consistently recovered on 18S rRNA phylogenies but there is no associated morphological synapomorphy (Adl et al., 2012; Cavalier-Smith, 2004; Lynn, 2008). Regrettably, at the time of the

Please cite this article in press as: Gentekaki, E., et al. Large-scale phylogenomic analysis reveals the phylogenetic position of the problematic taxon *Protocruzia* and unravels the deep phylogenetic affinities of the ciliate lineages. Mol. Phylogenet. Evol. (2014), http://dx.doi.org/10.1016/j.ympev.2014.04.020

analyses, we did not have multigene data for the other four classes to conclusively determine the monophyly of CONthreeP.

Both the ML and BI analyses have supported maximally an assemblage formed by Litostomatea + Armophorea + Spirotrichea (SAL) in agreement with a previous studies, though the support was very weak (Riley and Katz, 2001; Vd'ačný et al., 2010). The Litostomatea + Armophorea assemblage has been recovered frequently in phylogenetic studies. However, the support for this relationship has never been strong (Embley and Finlay, 1994; Gong et al., 2009; Hammerschmidt et al., 1996; Hirt et al., 1995; Vd'ačný et al., 2010). Nevertheless, the two classes do share some ontogenetic features (Foissner and Agatha, 1999). Based on these findings, Vd'ačný et al. (2010) proposed that Litostomatea and Armophorea be united into the Lamellicorticata. In our ML analysis, the two classes do indeed have a sister relationship, although the support is weak (Fig. 1). In the BI analysis, litostomes and armophoreans do not have a sister relationship, instead the armophoreans are sister to spirotrichs (Fig. 2). Spirotrichea and Armophorea undergo extensive chromosomal fragmentation resulting in gene-sized chromosomes, whereas the Litostomatea possess macronuclear chromosomes of larger size, a character shared by the CONthreeP cluster (Lipscomb et al., 2012; McGrath et al., 2007; Riley and Katz, 2001; Swart et al., 2013). This suggests that gene-sized chromosomes arose only twice within ciliate evolution. Regrettably, at the time of the analyses there were no available data from Phyllopharyngea, the only other group of ciliates known to have gene-sized macronuclear chromosomes.

4. Conclusions

Our aims in this first phylogenomic analysis of major clades of the phylum Ciliophora were to confirm the monophyly of the group, to resolve the phylogenetic position of the cytologically unusual ciliate *Protocruzia*, and to explore the deeper relationships within the phylum. We have vigorously confirmed the monophyly of the Ciliophora in agreement with ultrastructural, rRNA gene studies and some protein gene sequences. We postulate that *Protocruzia* is not a spirotrich, but its exact position remains unclear, at least until a representative of the Karyorelictea is included in the analyses. Although we do not have a complete sampling of all classes assigned to CONthreeP, we have representatives from all other classes in the Intramacronucleata. In this regard, our analyses have confirmed the "super" clade SAL, which is strongly supported. It will be intriguing to see if this "super" clade remains stable as future studies complete the sampling of the classes.

5. Uncited reference

Katz (2001).

Acknowledgments

This research was funded in part by the Gordon and Betty Moore Foundation through Grant #2637 to the National Center for Genome Resources. Samples MMETSP0125-20120918 (*Aristerostoma* sp.), MMETSP0210-20121227 (*Condylostoma magnum*), MMETSP0205-20121125 (*Euplotes focardii*), MMETSP0213-20121227 (*Euplotes harpa*), MMETSP0209-20121228 (*Litonotus* sp.), MMETSP0127-20121128 (Platyophrya macrostoma), MMETSP0216-20120918 (*Protocruzia adherens*), MMETSP0211-20121228 (*Pseudokeronopsis riccii*), MMETSP0123-20130129 (*Schmidingerella arcuata* [*Favella ehrenbergii*]), MMETSP0126-20121128 (*Strombidinopsis acuminatum*) and MMETSP0208-20121228 (*Strombidium inclinatum*) were sequenced, assembled and annotated at the National Center for Genome Resources.

Part of this research, was supported by an NSERC Discovery grant awarded to A.J.R. and D.H.L. A.J.R. acknowledges support from the Canada Research Chair Program and the Canadian Institute for Advanced Research Program in Integrated Microbial Diversity. E.G. was supported by a postdoctoral fellowship from the TULA Foundation. M.K. was supported by a NSERC Discovery grant awarded to A.J.R. Y.G. is funded by the Chinese Academy of Sciences. S.L.S. and K.J.B. are supported by the National Science Foundation Grant 1021189 to S.L.S. C.M. and S.P. are supported by the Italian "Programma Nazionale di Ricerche in Antartide" (PNRA), the Italian "Ministero dell'Istruzione, dell'Università e della Ricerca" (MIUR) (PRIN 2008) and the COST (action BM1102). L.M., V.B. and G.P. were supported by the European Commission FP7-PEOPLE-2009-IRSES project CINAR PATHOBACTER (project number 247658), by the COST action BM1102 to G.P. The authors G.D.G. and F.D. would like to thank Francesco Paolo Frontini for curating the *Protocruzia* cultures and Claudia Vannini for providing the *Euplotes harpa* culture.

Appendix A. Supplementary material

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.ympev.2014.04.020.

References

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

Abernathy, J.W., Xu, P., Li, P., Xu, D.H., Kucuktas, H., Klesius, P., Arias, C., Liu, Z.J., 2007. Genration and analysis of expressed sequence tags from the ciliate parasite *Ichthyophthirius multifiliis*. BMC Genomics 8, 176.

Adl, S.M., Simpson, A.G.B., Lane, C.E., Lukes, J., Bass, D., Bowser, S.S., Brown, M.W., Burki, F., Dunthorn, M., Hampl, V., et al., 2012. The revised classification of eukaryotes. J. Eukaryot. Microbiol. 59, 429–493.

Ammermann, D., 1968. Die Kernverhältnisse des Ciliaten Protocrucia depressa n. sp. Arch. Protistenkd. 110, 434–438.

Aury, J.M., Jaillon, O., Duret, L., Noel, B., Jubin, C., Porcel, B.M., Segurens, B., Daubin, V., Anthouard, V., Aiach, N., et al., 2006. Global trends of whole-genome duplications revealed by the ciliate *Paramecium tetraurelia*. Nature 444, 171–178.

Bachvaroff, T.R., Handy, S.M., Place, A.R., Delwiche, C.F., 2011. Alveolate phylogeny inferred using concatenated ribosomal proteins. J. Eukaryot. Microbiol. 58, 223–233.

Baldauf, S.L., Doolittle, W.F., 1997. Origin and evolution of the slime molds (Mycetozoa). Proc. Natl. Acad. Sci. USA 94, 12007–12012.

Bapteste, E., Brinkmann, H., Lee, J.A., Moore, D.V., Sensen, C.W., Gordon, P., Duruflé, L., Gaasterland, T., Lopez, P., Müller, M., Philippe, H., 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.

Baroin-Tourancheau, A., Villalobo, E., Tsao, N., Torres, A., Pearlman, R.E., 1998. Protein coding gene trees in ciliates: comparison with rRNA-based phylogenies. Mol. Phylogenet. Evol. 10, 299–309.

Barroin, A., Perasso, R., Qu, L.H., Brugerolle, G., Bachellerie, J.P., Adoutte, A., 1988. Partial phylogeny of the unicellular eukaryotes based on rapid sequencing of a portion of 28S ribosomal RNA. Proc. Natl. Acad. Sci. USA 85, 9764–9768.

Bernhard, D., Schlegel, M., 1998. Evolution of histone H4 and H3 genes in different ciliate lineages. J. Mol. Evol. 46, 344–354.

Bernhard, D., Leipe, D.D., Sogin, M.L., Schlegel, K.M., 1995. Phylogenetic relationships of the Nassulida within the phylum Ciliophora inferred from the complete small subunit rRNA gene sequences of *Furgasonia blochmanni*, *Obertrumia Georgiana* and *Pseudomicrothorax dubius*. J. Eukaryot. Microbiol. 42, 126–131.

Bernhard, D., Stechmann, A., Foissner, W., Ammermann, D., Hehn, M., Schlegel, M., 2001. Phylogenetic relationships within the class Spirotrichea (Ciliophora) inferred from small subunit rRNA gene sequences. Mol. Phylogenet. Evol. 21, 86–92.

Brown, M.W., Kolisko, M., Silberman, J.D., Roger, A.J., 2012. Aggregative multicellularity evolved independently in the eukaryotic supergroup Rhizaria. Curr. Biol. 22, 1123–1127.

Brown, M.W., Sharpe, S.C., Silberman, J.D., Heiss, A.A., Lang, B.F., Simpson, A.G.B., Roger, A.J., in press. Phylogenomics demonstrates that breviate flagellates are related to opisthokonts and apusomonads. Proc. Biol. Sci. http://dx.doi.org/10.1098/rspb.2013.1755 (20131755).

Budin, K., Philippe, H., 1998. New insights into the phylogeny of eukaryotes based on ciliate Hsp70 sequences. Mol. Biol. Evol. 15, 943–956.

UNCORRECTED PROOF

Burki, F., Shalchian-Tabrizi, K., Minge, M., Skjaeveland, A., Nikolaev, S.I., Jakobsen, K.S., Pawlowski, J., 2007. Phylogenomics reshuffles the eukaryotic supergroups. PLoS ONE 2, e790. http://dx.doi.org/10.1371/journal.pone.0000790.

Burki, F., Inagaki, Y., Bråte, J., Archibald, J.M., Keeling, P.J., Cavalier-Smith, T., Sakaguchi, M., Hashimoto, T., Horak, A., Kumar, S., Klaveness, D., Jakobsen, K.S., Pawlowski, J., Shalchian-Tabrizi, K., 2009. Large-scale phylogenomic analyses reveal that two enigmatic protist lineages, Telonemia and Centroheliozoa, are related to photosynthetic Chromalveolates. Genome Biol. Evol. 9, 231–238.

Burki, F., Kudryavtsev, A., Matz, M., Aglyamova, G., Bulman, S., Fiers, M., Keeling, P.J., Pawlowski, J., 2010. Evolution of Rhizaria: new insights from phylogenomic analysis of uncultivated protists. BMC Evol. Biol. 10, 377.

Burki, F., Okamoto, N., Pombert, J.F., Keeling, P.J., 2012. The evolutionary history of haptophytes and cryptophytes: phylogenomic evidence for separate origins. Proc. Biol. Sci. 279, 2246–2254.

Burki, F., Corradi, N., Sierra, R., Pawlowski, J., Meyer, G.R., Abbott, C.L., Keeling, P.J., 2013. Phylogenomics of the intracellular parasite *Mikrocytos mackini* reveals evidence for a mitosome in Rhizaria. Curr. Biol. 23, 1541–1547.

Bütschli, O., 1887–1889. Protozoa. Abt. III Infusoria und System der Radiolaria. In: Bronn, H.G. (Ed.), Klassen und Ordnung des Thiers-Reichs. Leipzig, vol. 1, pp. 1098–2035.

Cameron, S.L., Adlard, R.D., O'Donoghue, P.J., 2001. Evidence for an independent radiation of endosymbiotic litostome ciliates within Australian marsupial herbivores. Mol. Phylogenet. Evol. 20, 302–310.

Cavalier-Smith, T., 2004. Chromalveolate diversity and cell megaevolution: interplay of membranes, genomes and cytoskeleton. In: Hirt, P.R., Horner, D.S. (Eds.), Organelles, Genomes and Eukaryote Phylogeny. Boca Raton, London, New York, Washington, pp. 75–108.

Chatton, E., Lwoff, A., 1935a. Les ciliés apostomes. Morphologie, cytology, éthologie, evolution, systématique. Première partie. Aperçu historique et general. Etude monographique des genres et des espéces. Arch. Zool. Exp. Générale. 77, 1–453.

Chatton, E., Lwoff, A., 1935b. La constitution primitive de la strie ciliaire des infusoires. La desmodexie. Comptes Rendus des Séances de la Société de Biologieet de ses Filiales 118, 1068–1072.

Criscuolo, A., Gribaldo, S., 2010. BMGE (Block Mapping and Gathering with Entropy): a new software for selection of phylogenetic informative regions from multiple sequence alignments. BMC Evol. Biol. 10, 210–231.

Da Silva Paiva, T., do Nascimento Borges, B., da Silva-Neto, I.D., 2013. Phylogentic study of Class Armophorea (Alveolata, Ciliophora) based on 18S-rDNA data. Genet. Mol. Biol. 36, 571–585.

De Puytorac, P., 1994. Phylum Ciliophora Doflein, 1901. In: de Puytorac, P. (Ed.), Traité de zoologie, Infusoires Ciliés. Paris, vol. 2, pp. 1–15.

Elwood, H.K., Olsen, G.I., Sogin, M.L., 1985. The small subunit ribosomal RNA gene sequences from the hypotrichous ciliates *Oxytricha nova* and *Stylonychia pustulata*. Mol. Biol. Evol. 2, 399–410.

Embley, T.M., Finlay, B.J., 1994. The use of small subunit rRNA sequences to unravel the relationships between anaerobic ciliates and their methanogen endosymbionts. Microbiology 140, 225–235.

Embley, T.M., Finlay, B.J., Dyal, P.L., Hirt, R.P., Wilkinson, M., Williams, A.G., 1995. Multiple origins of anaerobic ciliates with hydrogenosomes within the radiation of aerobic ciliates. Proc. Biol. Sci. 262, 87–93.

Fauré-Fremiet, E., 1950. Morphologie compare et systématique des ciliés. Bulletin Société de Zoologique de France 75, 109–122.

Foissner, W., Agatha, S., 1999. Morphology and morphogenesis of *Metopus hasei* Sondheim, 1929 and *M. inversus* (Jankowski, 1964) nov. comb. (Ciliophora, Metopida). J. Eukaryot. Microbiol. 46, 174–193.

Gao, S., Strüder-Kypke, M.C., Al-Rasheid, K.A.S., Lin, X., Song, W., 2010. Molecular phylogeny of three ambiguous ciliate genera: *Kentrophoros*, *Trachelophos* and *Trachelotractus*. Zool. Scr. 39, 305–313.

Gong, J., Stoeck, T., Yi, Z., Miao, M., Zhang, Q., Roberts, D.M.C.L., Warren, A., Song, W., 2009. Small subunit rRNA phylogenies show that the class Nassophorea is not monophyletic (phylum Ciliophora). J. Eukaryot. Microbiol. 56, 339–347.

Greenwood, S.J., Schlegel, M., Sogin, M.L., Lynn, D.H., 1991. Phylogenetic relationships of *Blepharisma americanum* and *Colpoda inflata* within the phylum Ciliophora inferred from the complete small subunit rRNA gene sequences. J. Protozool. 38, 1–6.

Gribaldo, S., Philippe, H., 2000. Pitfalls in tree reconstruction and the phylogeny of eukaryotes. In: Hirt, P.R., Horner, D.S. (Eds.), Organelles, Genomes and Eukaryote Phylogeny. Boca Raton, London, New York, Washington.

Grolière, C.A., de Puytorac, P., Detcheva, R., 1980. À prpopos d' observations sur la stomatogenèse et l' ultrastructure du cilié *Protocruzia tuzeti* Villeneuve-Brachon, 1940. Protistologica 16, 453–466.

Hammerschmidt, B., Schlegel, M., Lynn, D.H., Leipe, D.D., Sogin, M.L., Raikov, I.B., 1996. Insights into the evolution of nuclear dualism in the ciliates revealed by phylogenetic analysis of rRNA sequences. J. Eukaryot. Microbiol. 43, 225–230.

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

Hirt, P.R., Dyal, P.L., Wilkinson, M., Finlay, B.J., Roberts, D.M., Embley, T.M., 1995. Phylogenetic relationships among karyorelictids and heterotrichs inferred from small subunit rRNA sequences: resolution at the base of the ciliate tree. Mol. Phylogenet. Evol. 4, 77–87.

Israel, R.L., Kosakovsky Pond, S.L., Muse, S.V., Katz, L.A., 2002. Evolution of duplicated alpha-tubulin genes in ciliates. Evolution 56, 1110–1122.

Jankowski, A.W., 1967. A new system of ciliate Protozoa (Ciliophora). Akademii Nauk SSSR, Trudy Zoologicheskogo Instituta. 43, 3–54.

Jankowski, A.W., 1973. Taxonomic revision of subphylum Ciliophora Doflein, 1901. Zoologichesky Zhurnal 52, 165–175.

Katoh, K., Toh, H., 2010. Parallelization of the MAFFT multiple sequence alignment program. Bioinformatics 26, 1899–1900.

Katoh, K., Misawa, K., Kuma, K.I., Miyata, T., 2002. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. Nucleic Acids Res. 30, 3059–3066.

Katz, L.A., 2001. Evolution of nuclear dualism in ciliates: a reanalysis in light of recent molecular data. Int. J. Syst. Evol. Microbiol. 51, 1587–1592.

Katz, L.A., Bornstein, J.G., Lasek-Nesselquist, E., Muse, S.V., 2004. Dramatic diversity of ciliate histone H4 genes revealed by comparisons of patterns of substitutions and paralogs divergences among eukaryotes. Mol. Biol. Evol. 21, 555–562.

Klein, B.M., 1928. Die Silberliniensysteme der Ciliaten. Weitere Resultate. Arch. Protistenkd. 62, 177–260.

Klein, B.M., 1929. Weitere Beiträge zur Kenntnis des Silberliniensystems der Ciliaten. Arch. Protistenkd. 65, 183–257.

Lartillot, N., Lepage, T., Blanquart, S., 2009. PhyloBayes 3: A Bayesian software package for phylogenetic reconstruction and molecular dating. Bioinformatics 25, 2286–2288.

Le, S.Q., Gascuel, O., 2008. An improved general amino acid replacement matrix. Mol. Biol. Evol. 25, 1307–1320.

Leander, B.S., Keeling, P.J., 2003. Morphostasis in alveolates evolution. Trends Ecol. Evol. 18, 395–402.

Leipe, D.D., Bernhard, D., Schlegel, M., Sogin, M.L., 1994. Evolution of 16S-like ribosomal RNA genes in the ciliophoran taxa Litostomatea and Phyllopharyngea. Eur. J. Protistol. 30, 354–361.

Li, L., Stoeck, T., Kyoon, S.M., Al-Rasheid, K.A.S., Al-Khedhairy, A.B., Song, W., 2010. *Protocruzia*, a highly ambiguous ciliate (Protozoa; Ciliophora): very likely an ancestral form for Heterotrichea, Colpodea or Spirotrichea? With reevaluation of its evolutionary position based on multigene analyses. Sci. China Life Sci. 53, 131–138.

Lipscomb, D.L., Bowditch, B.M., Riordan, G.P., 2012. A molecular and ultrastructural description of *Spathidiopsis buddenbrocki* and the phylogenetic position of the family Placidae (Ciliophora). J. Eukaryot. Microbiol. 59, 67–79.

Lynn, D.H., 1981. The organization and evolution of microtubular organelles in ciliated protozoa. Biol. Rev. Camb. Philos. Soc. 56, 243–292.

Lynn, D.H., 1991. The implications of recent descriptions of kinetid structure to the systematics of the ciliated protists. *Protoplasma*. 164, 123–142.

Lynn, D.H., 1996. My journey in ciliate systematics. J. Eukaryot. Microbiol. 43, 253–260.

Lynn, D.H., 2003. Morphology or molecules: How do we identify the major lineages of ciliates (Phylum Ciliophora)? Eur. J. Protistol. 39, 356–364.

Lynn, D.H., 2008. The Ciliated Protozoa: Characterization, classification and guide to the literature. New York.

Lynn, D.H., Sogin, M.L., 1988. Assessment of phylogenetic relationships among ciliated protists using partial ribosomal RNA sequences derived from reverse transcripts. Biosystems 21, 249–254.

Lynn, D.H., Strüder-Kypke, M., 2002. Phylogenetic position of *Licnophora*, *Lechriopyla* and *Schizocaryum*, three unusual ciliates (Phylum Ciliophora) endosymbiotic in echinoderms (Phylum Echinodermata). J. Eukaryot. Microbiol. 49, 460–468.

Lynn, D.H., Wright, A.D.G., Schlegel, M., Foissner, W., 1999. Phylogenetic relationships of orders within the class Colpodea (Phylum Ciliophora) inferred from small subunit rRNA gene sequences. J. Mol. Evol. 48, 605–614.

McGrath, C.L., Zufall, R.A., Katz, L.A., 2007. Variation in macronuclear genome content of three ciliates with extensive chromosomal fragmentation: a preliminary analysis. J. Eukaryot. Microbiol. 54, 242–246.

Orsi, W., Edgcomb, V., Faria, J., Foissner, W., Fowle, W.H., Hohmann, E., Suarez, P., Taylor, C., Taylor, G.T., Vd'ačný, P., Epstein, S.S., 2012. Class Cariacotrichea, a novel ciliate taxon from the anoxic Cariaco Basin, Venezuela. ISME J. 62, 1425–1433.

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

Philippe, H., Adoutte, A., 1998. The molecular phylogeny of Eukaryota: Solid facts and uncertainties. In: Coombs, G.H., Vickerman, K., Sleigh, M.A., Warren, A. (Eds.), Evolutionary relationships among protozoa. London. pp 25–56.

Philippe, H., López, P., Brinkmann, H., Budin, K., Germot, A., Laurent, J., Moreira, D., Le Guyader, H., 2000. Early-branching or fast-evolving eukaryotes? an answer based on slowly evolving positions. Proc. Biol. Sci. 267, 1213–1221.

Price, M.N., Dehal, P.S., Arkin, A.P., 2009. FastTree: computing large minimum-evolution trees with profiles instead of a distance matrix. Mol. Biol. Evol. 27, 1641–1650.

Ricard, G., de Graaf, R.M., Dutilh, B.E., Duarte, I., van Alen, T.A., van Hoek, A.H.A.M., Boxma, B., van der Staay, G.W.M., et al., 2008. Macronuclear genome structure of the ciliate *Nyctotherus ovalis*: single-gene chromosomes and tiny introns. BMC Genomics 9, 587.

Riley, J.L., Katz, L.A., 2001. Widespread distribution of extensive chromosomal fragmentation in ciliates. Mol. Biol. Evol. 18, 1372–1377.

Ruthmann, A., Hauser, M., 1974. Mitosis-like macronuclear division in a ciliate. Chromosoma 45, 261–272.

Shin, M.K., Hwang, U.W., Kim, W., Wright, A.D.G., Krawczyk, C., Lynn, D.H., 2000. Phylogenetic position of the ciliates *Phacodinium* (Order Phacodiniida) and *Protocruzia* (Subclass Protocruziidia) and systematics of the spirotrich ciliates examined by small subunit ribosomal RNA gene sequences. Eur. J. Protistol. 36, 293–302.

UNCORRECTED PROOF

Please cite this article in press as: Gentekaki, E., et al. Large-scale phylogenomic analysis reveals the phylogenetic position of the problematic taxon *Protocruzia* and unravels the deep phylogenetic affinities of the ciliate lineages. Mol. Phylogenet. Evol. (2014), http://dx.doi.org/10.1016/j.ympev.2014.04.020

Small, E.B., Lynn, D.H., 1981. A new macrosystem for the phylum Ciliophora Doflein, 1901. Biosystems 14, 387–401.

Snoeyenbos-West, O.L.O., Cole, J., Campbell, A., Coats, W.D., Katz, L.A., 2004. Molecular phylogeny of phyllopharyngean ciliates and their Group I introns. J. Eukaryot. Microbiol. 51, 441–450.

Sogin, M.L., Elwood, H.J., 1986. Primary structure of the *Paramecium tetraurelia* small subunit RNA coding region: phylogenetic relationships within the Ciliophora. J. Mol. Evol. 23, 53–60.

Song, W., Wilbert, N., 1997. Morphological investigations on some free living ciliates (Protozoa, Ciliophora) from China seas with description of a new hypotrichous genus, Hemigastrostyla nov. gen. Arch. Protistenkd. 148, 413–444.

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

Stechmann, A., Schlegel, M., Lynn, D.H., 1998. Phylogenetic relationships between prostome and colpodean ciliates tested by small subunit rRNA sequences. Mol. Phylogenet. Evol. 9, 48–54.

Stoeck, T., Taylor, G.T., Epstein, S.S., 2003. Novel eukaryotes from the permanently anoxic Cariaco Basin (Caribbean Sea). Appl. Environ. Microbiol. 69, 5656–5663.

Susko, E., Field, C., Blouin, C., Roger, A.J., 2003. Estimation of rates-across-sites distributions in phylogenetic substitutions models. Syst. Biol. 52, 594–603.

Swart, E.C., Bracht, J.R., Magrini, V., Minx, P., Chen, X., Zhou, Y., Khurana, J.S., Goldman, A.D., Nowacki, M., Schotanus, K., et al., 2013. The *Oxytricha trifallax* macronuclear genome: a complex eukaryotic genome with 16,000 tiny chromosomes. PLoS Biol. 11, e1001473. http://dx.doi.org/10.1371/journal.pbio.1001473.

Vd'ačný, P., Orsi, W., Foissner, W., 2010. Molecular and morphological evidence for a sister group relationship of the classes Armophorea and Litostomatea (Ciliophora, Intramacronucleata, Lamellicorticata infraphyl. nov.), with an account on basal litostomateans. Eur. J. Protistol. 46, 298–309.

von Gelei, J., 1932. Die reizleitenden Elemente der Ciliaten in nass hergestellen Silberbzw. Goldpräparaten. Arch. Protistenkd. 77, 152–174.

von Gelei, J., 1934. Eine mikrotechnische Studie über die Färbung der subpelliculären Elemente der Ciliaten. Zeitschrift für Wissenschaftliche Mikroskopie und für Mikroscopische Technik. 51, 103–178.

Zhao, S., Burki, F., Bråte, J., Keeling, P.J., Klaveness, D., Shalchian-Tabrizi, K., 2012. *Collodictyon* – an ancient lineage in the tree of eukaryotes. Mol. Biol. Evol. 29, 1557–1568.

UNCORRECTED PROOF