Phylogenomic analysis of *Balantidium ctenopharyngodoni* (Ciliophora, Litostomatea) based on single-cell transcriptome sequencing

Zongyi Sun¹, Chuanqi Jiang², Jinmei Feng³, Wentao Yang², Ming Li¹,²,* and Wei Miao²,*

¹ Hubei Key Laboratory of Animal Nutrition and Feed Science, Wuhan Polytechnic University, Wuhan 430023, PR China
² Institute of Hydrobiology, Chinese Academy of Sciences, No. 7 Donghu South Road, Wuchang District, Wuhan 430072, Hubei Province, PR China
³ Department of Pathogenic Biology, School of Medicine, Jianghan University, Wuhan 430056, PR China

Received 22 April 2017, Accepted 12 October 2017, Published online 14 November 2017

Abstract – In this paper, we present transcriptome data for *Balantidium ctenopharyngodoni* Chen, 1955 collected from the hindgut of grass carp (*Ctenopharyngodon idella*). We evaluated sequence quality and *de novo* assembled a preliminary transcriptome, including 43.3 megabits and 119,141 transcripts. Then we obtained a final transcriptome, including 17.7 megabits and 35,560 transcripts, by removing contaminative and redundant sequences. Phylogenomic analysis based on a supermatrix with 132 genes comprising 53,873 amino acid residues and phylogenetic analysis based on SSU rDNA of 27 species were carried out herein to reveal the evolutionary relationships among six ciliate groups: Colpodea, Oligohymenophorea, Litostomatea, Spirotrichea, Heterotrichea and Protocruziida. The topologies of both phylogenomic and phylogenetic trees are discussed in this paper. In addition, our results suggest that single-cell sequencing is a sound method of obtaining sufficient omics data for phylogenomic analysis, which is a good choice for uncultivable ciliates. The transcriptome data for *Balantidium ctenopharyngodoni* are the first omics data within the subclass Trichostomatia, and provide a good basis for ciliate phylogenomic analysis, as well as related omics analysis.

Keywords: *Balantidium ctenopharyngodoni*, Trichostomatia, single-cell transcriptome sequencing, SSU rDNA, phylogenomic analysis

Résumé – Analyse phylogénomique de *Balantidium ctenopharyngodoni* (Ciliophora, Litostomatia) basée sur le séquençage du transcriptome d’une seule cellule. Les données du transcriptome de *Balantidium ctenopharyngodoni* Chen, 1955, recueillies à partir de l’intestin de la carpe herbivore (*Ctenopharyngodon idella*), sont présentées dans cet article. Nous avons évalué la qualité de sa séquence et avons assemblé de novo un transcriptome préliminaire, incluant 43,3 mégabits et 119 144 transcriptions. Ensuite, nous avons obtenu un transcriptome final, incluant 17,7 mégabits et 35 560 transcriptions, en supprimant les séquences contaminatrices et redondantes. Une analyse phylogénomique basée sur une supermatrice avec 132 gènes comprenant 53 873 résidus d’acides aminés et une analyse phylogénétique basée sur l’ADNr SSU de 27 espèces a été réalisée pour révéler les relations évolutives entre six groupes de ciliés: Colpodea, Oligohymenophorea, Litostomatea, Spirotrichea, Heterotrichea et Protocruziida. Les topologies des arbres phylogénomiques et phylogénétiques sont discutées dans ce document. En outre, nos résultats suggèrent que le séquençage d’une seule cellule est une méthode solide pour obtenir suffisamment de données omiques pour l’analyse phylogénomique, ce qui est un bon choix pour les ciliés incultivables. Les données du transcriptome de *Balantidium ctenopharyngodoni* sont les premières données omiques dans la sous-classe Trichostomatia et fournissent une base pour l’analyse phylogénomique des ciliés ainsi que l’analyse omique associée.

*Corresponding authors: liming@ihb.ac.cn; miaowei@ihb.ac.cn

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.
Introduction

The genus Balantidium Claparède & Lachmann, 1858 belongs to the order Vestibiuliferida (Trichostomatia, Ciliophora). It was established by Claparède and Lachmann when they separated Bursaria entozoon Ehrenberg, 1838 from the genus Bursaria Ehrenberg, 1838 and assigned it to Balantidium entozoon [4]. A large number of Balantidium species, thereafter, have been reported as endocommensals in the digestive tracts of both vertebrate and invertebrate hosts, especially fish and amphibians. At present, 19 fish-isolated species (13 freshwater species and 6 marine species) and 31 amphibian-isolated species (26 species from anura amphibians and 5 species from urode amphibians) have been found [21].

Balantidium ctenopharyngodoni Chen, 1955 is a native species of China and a common obligatory intestinal parasite of grass carp (Ctenopharyngodon idella). It was first discovered and simply documented by Chen [2], then, its morphology was redescribed [20], pathogenicity [28], ultrastructural anatomy [22] and related molecular phylogeny [22, 24] were studied, respectively. As for phylogenetic analysis, the related studies reconstructed the phylogenetic trees only based on the small subunit ribosomal DNA (SSU rDNA) gene. The SSU rDNA gene is the most common gene marker for phylogenetic analysis [42], which was first used by Lynn et al. [23] to demonstrate divergences among ciliates over 25 years ago. For example, it was applied to place the hymenostomes and peritrichs as sister lineages [26, 27]. However, with the development of high-throughput sequencing in recent decades, enormous amounts of omics data have become available for phylogenetic analysis, potentially including hundreds of thousands of loci, which provide much more information compared to single or multi-gene analysis [7, 18, 25]. To take full advantage of this abundant data, phylogenomics—research into evolutionary relationships based on comparative analysis of omics data—was proposed to replace single-gene phylogenetics [30, 33], through the use of genomics rather than single-gene information [29]. This makes it possible to avoid problems related to insufficient sample size by addition of more sites from multiple genes or to compensate for biased base compositions. Although some of these common problems will also affect phylogenies reconstructed from large datasets, others will be substantially reduced and/or easily diagnosed [5]. To date, several ciliates have been sequenced and analyzed, such as Tetrahymena thermophila [10], Paramecium tetraurelia [1], Ichthyophthirius multifiliis [6], Pseudocohnilembus persalinus [43] and Oxytreca trifallax [39], and the results obtained were used for phylogenomic analysis [3, 11, 12]. Along with the steady decline in costs as new technologies have been developed and/or refined, sequencing projects have provided a possible method, single-cell sequencing, for many uncultivable non-model organisms to fill in the gaps in phylogenomics, such as for Balantidium. Here we obtained RNA-Seq data for B. ctenopharyngodoni, which are the first omics data within the subclass Trichostoma-
Alveolata, the transcripts were reserved for further analysis; otherwise they were excluded from the transcriptome. Then, both the transcript fragments with hits belonging to Alveolata and transcript fragments without hits were put together and removed the redundancy of remnant transcriptome (transcripts of no-hits and top hit matching Alveolata) by using Corset [8] with default parameters. Finally, we obtained a contamination-exclusion and non-redundant transcriptome.

We used the contamination-exclusion and non-redundant transcriptome to carry out further analysis. Potential rDNA sequences were extracted from the transcripts, and species identification were verified by BLAST searching against the GenBank database by using the rDNA sequences as queries. Predicted protein sequences were obtained by translating the assembled transcripts using the ciliates’ codon table (only TGA as stop codon, such as Tetrahymena) by the getorf program of the EMBASS site [32] (table 6) and we picked the longest protein sequences from every translated sequence. Subsequently, we used interproscan-5.19-58.0 [16] with default parameters to implement functional annotation about these protein sequences and then classified the annotated result with WEGO (http://wego.genomics.org.cn/cgi-bin/wego/index.pl) [44]. For phylogenomic analysis, translated protein sequences, which are more than 200 amino acid residues, were selected to find orthologous genes.

**Data sources**

Omics data and SSU rDNA gene sequences used in this study are listed in Table S1. Genomic data for four Tetrahymena spp. (T. borealis, T. elliotti, T. malaccensis and T. thermophila) were downloaded from the Tetrahymena Genome Database (TGDb) (http://ich.ciliate.org/index.php/home/welcome); Ichthyophthirius multifilis, Oxytricha trifallax, Paramecium tetraurelia and Stylochlamys lemnæ were obtained from the Ichthyophthirius multifilis Genome Database (IchDB) (http://ich.ciliate.org/index.php/home/welcome) the Oxytricha Genome Database (OxyDB) (http://oxy.ciliate.org/index.php/home/welcome), the ParameciumDB (http://paramecium.cgm.cnrs-gif.fr/) and the Stylochlamys Genome Database (StyloDB) (http://styro.ciliate.org/index.php/home/welcome), respectively. Genomic data for two Plasmodium spp. (P. falciparum and P. yoelii) were downloaded from the Plasmodb (http://plasmodb.org/plasmo/). Omics data for Pseudocohnilembus persalinus, Parabodo digitiformis, Colpoda aspera, Campanella umbellaria, and Carchesium polygynum were downloaded from the National Center for Biotechnology Information (NCBI) with accession numbers SRR1768438, SRR1768439, SRR1768440, SRR1768423, and SRR1768437, respectively. All transcriptomic data of the eleven ciliate species were downloaded from iMicrobe (http://imicrobe.us/) under project numbers MMETSP0125: Aristerostoma spp., MMETSP0210: Condylostoma magnum, MMETSP0265: Euplotes fociardi, MMETSP0213: Euplotes harpa, MMETSP0209: Litonotus sp., MMETSP0127: Platypo- phyla macrostoma, MMETSP0216: Protocraziæ adherens, MMETSP0211: Pseudokeronopsis ricci, MMETSP0123: Schmidingerella arcuata, MMETSP0126: Strombidinop- sis acuminatum, and MMETSP0208: Strombidium inclinatum. We obtained orthologous genes between two species by using the Reciprocal Best Hits (RBH) approach in BLAST with E-value no more than $10^{-20}$. If the score of the second-best hit of either gene in the other genome was less than half of the score of the best hit, the RBH-pair was retained. Only gene sets including more than 70% (at least 19 different species of our 27 species) of all species were retained for further analysis [11].

The SSU rDNA gene sequence for B. ctenopharyngodon was obtained from an assembled 1500-bp fragment (Figure S2) in this work. The SSU rDNA gene sequences of the other 27 species were downloaded from GenBank and iMicrobe.

**Sequence alignment and phylogenetic analysis**

For phylogenomic analysis, sequence alignments were implemented by MUSCLE version 3.6 [9], with the default parameters. Sequence alignments were re-optimized using Gblock 0.91b ($-t = \text{d}, -b2 = 0.65, -b3 = 10, -b4 = 5, -b5 = a$) to detect and trim the ambiguously aligned regions [40]. The 132 individual alignments of the 27 taxa were concatenated into a supermatrix by ScaFaSo software version 4.42 using the data set assembling panel [35]. The maximum-likelihood (ML) analysis and Bayesian inference (BI) analysis followed Feng et al. [11]. For ML analysis, the concatenated data set was conducted in RAxML version 7.2.6 [38], with the LG model of amino acid substitution + F + Γ4 distribution (Γ four rate categories). The ML analysis was evaluated with 100 replicates. For BI analysis, the concatenated data set was implemented by the software PhyloBayes 3.3 [17] under CAT-POI model + Γ4 (a discrete gamma distribution of rate variation with four rate categories) with two independent Markov chain Monte Carlo (MCMC) runs. In order to evaluate the convergence of the two independent MCMC runs, the bcomp program was used to compare the discrepancy of bipartition frequencies between the two runs and output a consensus tree and the maxdiff equalled 0.068, which was less than 0.1.

For the SSU rDNA gene, sequence alignments were implemented by ClustaW algorithm in MEGA6, with the default parameters (Gap Extension Penalty: 6.66; Transition Weight: 0.5) [41]. In order to decrease the number of characters presenting ambiguous alignment, we trimmed the ambiguously aligned positions. Our final data set comprised 1,695 nucleotide sites, which were used for subsequent analysis. A phylogenetic tree based on the SSU rDNA gene was constructed by the maximum-likelihood (ML) method with the MEGA6 and Bayesian Inference (BI) method with MrBayes 3.2 [34]; respectively. For the ML analysis, the Tamura 3-parameter model of nucleotide substitution with a proportion of invariable sites (I) and a Gamma distribution (G) ($T92 + G + I$) as the best model of
nucleotide evolution based on the Akaike Information Criterion [15] were used to calculate in MEGA6. For the BI analysis, a GTR + G + I model (n = 6, rates = invgamma) was selected to perform for 1,000,000 generations (or until the standard deviations of split frequencies were below 0.01) with trees sampled every 1,000 generations and an initial relative burn-in of 25% of the trees by MrBayes 3.2 and other parameters were default.

Results

Morphological description

*B. ctenopharyngodoni* specimens are mainly found in the luminal contents of the hindgut, especially near the anal opening. The organism is spindle-like or oval in shape (Figure 1A). Body size 76.2 – 160.7 μm (X = 98.0 μm; n = 30) × 40.5 – 70.1 μm (X = 59.2 μm; n = 30). Its body is highly flexible to permit the ciliate to move through narrow intestinal folds. The vestibulum is situated anteromedially, extending directly posterior to about one sixth of the body length (Figure 1B). All ciliary rows started from the border of the vestibulum, arranged in closely set lines parallel to the longitudinal axis and packed over the body (Figures 1C and D). The macronucleus is kidney-shaped, usually with a tiny spherical micronucleus embedded in the middle concavity (Figure 1E).

Single-cell RNA-sequencing and transcriptome analysis

The Illumina HiSeq4000 sequencer produced a total of 19,440,851 paired-end reads (averaged 150 bp for each read) for *B. ctenopharyngodoni* (including 5,832,255,300 bases with 34% average GC content), and 96.88% reads were retained after filtering low quality reads. The remaining high-quality reads were used to assemble the transcriptome and a 43.3 megabit (Mb) preliminary transcriptome was obtained, including 119,141 transcripts. To exclude contamination, 119,141 transcripts were searched against nt database by using BLAST. In the result, 40,411 transcripts were hits and 78,730 transcripts had no hits. Moreover, the top hit of 40,236 transcripts belonged to non-Alveolata and the top hit of 175 transcripts belonged to Alveolata among the 40,411 transcripts. As a result, we discarded 40,236 transcripts from the preliminary redundant transcriptome to gain redundant transcriptome, including 78,905 transcripts. To remove redundancy from the new transcriptome, 43,345 transcripts were abandoned by using Corset with default. In the end, we obtained a final transcriptome including 35,560 transcripts (Figure S1). We implemented functional annotation of 35,560 transcripts with WEGO, and 1,427 transcripts were annotated and classified (Figure S3).

Two fragments, the 698 bp fragment and the 1277 bp fragment, of SSU rDNA sequences were extracted from the transcriptome assembled here by using two rDNA sequences (GU480804 and KU170972) of *B. ctenopharyngodoni* from GenBank as queries. Sequence alignments (Figure S2) showed that there is a 25-bp overlap between the two fragments, which leads to a 1950 bp potential rDNA. The 1950 bp rDNA sequence shows 99.69% identity to the GU480804 with the alignment region from 19 to 1368 bp. The 1950 bp rDNA sequence shows 100% identity to the KU170972 with the alignment region from 1 to 330 bp.

Phylogenomic analysis

We presented the taxonomic sampling of 27 species by integrating data from one newly obtained transcriptome (*B. ctenopharyngodoni*) and other omics data from NCBI, TGD, IchDB, OxyDB, ParameciumDB, StyleDB and iMicrobe. Altogether, omics data for 25 ciliates were used in our analysis and two species of *Plasmodium* were used as the outgroups. The 25 ciliate samples included six ciliate groups: Colpodea, Oligohymenophorea, Litostomatea, Spirotrichea, Heterotrichae and Protocruziida. We assembled a 132-gene dataset comprising 53,873 amino acid residues of the 27 species by analyzing their genome and transcriptome data. Both the ML tree and BI tree show a consistent topological tree. In addition, the ML tree topology is presented with support values of both ML and BI analysis indicated on branches (Figure 2). In the phylogenomic tree, Colpodea is sister to Oligohymenophorea and Litostomatea is sister to Spirotrichea. The phylogenomic analysis provides full statistical support that *B. ctenopharyngodoni* is sister to *Litonotus* sp., and both of them are Litostomatea. *Protocruzia adherens* and *Condylostoma magnum* are placed in the earliest diverging position in the ciliate tree.

The topology of the phylogenetic tree derived from the SSU rDNA genes reconstructed by ML and BI is very similar to that of the phylogenomic tree. However, the phylogenetic tree derived from the SSU rDNA genes (Figure 3) showed that Peritrichia (*Campanella umbellaria* and *Carchesium polypinum*) is sister to the Hymenostomatia (*Tetrahymena* spp. and *Ichthyophthirius multifiliis*), and Peniculida (*Paramecium tetraurelia*) is sister to Scuticociliatia (*Pseudocohnilembus persalinus* and *Paralembus digitiformis*). Moreover, *Protocruzia adherens* is sister to Colpodea and Oligohymenophorea, but not in the deepest branch.

Discussion

The important morphological characteristics of *B. ctenopharyngodoni* observed herein are consistent with those described by Li et al. [20]. Combined with evidence from the SSU rDNA sequences, we confirmed that the ciliate isolated in the present study is *B. ctenopharyngodoni*.

For single-cell sequencing, there is also a problem in transcriptome assembly of ciliates which are uncultivable, such as *B. ctenopharyngodoni*. When we collected the
experimental sample, some contaminations were inevi-
tably included. It was therefore very important concern-
ing uncultivable ciliates to exclude contamination before
implementing bioinformatics analysis. In this work,
contaminants were excluded in two steps: (1) eradict-
ing surface impurities of ciliates; (2) eliminating se-
quences of residual contaminants. In the first step, a simple
and effective method is to clean ciliates by washing with
distilled water. However, since many ciliates are sensitive
to osmotic pressure, cleaning with distilled water may
result in cell lysis. We found that B. ctenopharyngodoni
remained alive for about 10 minutes in distilled water and
most of them could be kept alive for more than 30 minutes
in 0.65% NaCl, while others survived more than 2 hours.
As a result, we starved B. ctenopharyngodoni for about
2 hours in 0.65% NaCl before collecting samples. This was
done because keeping the microorganisms alive for a long
time to digest the intracellular contamination (food
residue) is a very effective method of reducing contami-
nation. Washing several times with 0.65% saline and dis-
tilled water was done to remove epiphytic contamination. In
the second step, excluding contamination by using bioinfor-
matics analysis, we de novo assembled transcriptome
directly and then searched against the nt database to
discard the contaminating transcripts and redundant ones
(see above). Furthermore, guanine-cytosine (GC)-content
is a useful feature of ciliate sequences to distinguish them
from one another [43]; thus, we used the GC-content
distribution to evaluate the quality of the transcriptome
assembled here. The GC-content distribution of the
preliminary redundant transcriptome (Figure S4) of the
sequencing sample B. ctenopharyngodoni shows two
peaks, a main peak (32%) and a sub-peak (50%),
indicating that some other sequences were present in the
preliminary redundant transcriptome. After excluding
contamination and reducing redundancy by bioinfor-
matics analysis, there was only one peak (32%) in each of
the two new transcriptomes, indicating that the method of
bioinformatics analysis to exclude contamination was very
efficient. Therefore, we suggest that integrated methods
to reduce contamination using sterile isosmotic solution
to wash ciliates is a good way to exclude contamination.
Moreover, poly-A is a specific characteristic of eukaryote
mRNA and is a target for ciliates to obtain transcripts

Figure 1. Light microscopy images of Balantidium ctenopharyngodoni. A. Living specimens, showing numerous B. ctenopharyngodoni specimens in the luminal contents of the hindgut. Scale bar = 50 μm. B. Living specimens, showing the general form and vestibulum (arrow). Scale bar = 50 μm. C-D-E. Specimens stained with protargol, showing its somatic kineties (C.) peripheral fibres (pf), (D.) macronucleus (mac) and micronucleus (mic). (E.). Scale bar = 50 μm.
before sequencing, which can in theory filter bacterial contamination. Sequencing transcriptome is therefore a relatively sound method for uncultivable ciliates to be applied in phylogenomic analysis based on single-cell sequencing.

Phylogenomic trees show similar topological structures to the tree topology derived from the SSU rDNA gene. However, there is some debate on the phylogenetic position of Oligohymenophorea and *Protocravia adherens* by comparing with the tree from these two methods. For Oligohymenophorea, the phylogenetic tree derived from the SSU rDNA gene shows that Peritrichia (*Campanella umbellaria* and *Carchesium polypinum*) is close to Hymenostomatia (*Tetrahymena* spp. and *Ichthyophthirius multifiliis*). By contrast, the phylogenomic tree derived from the supermatrix of the 132 gene containing 53,873 unambiguously aligned amino acid residues of 27 taxa supports the hypothesis that Scuticociliattia (*Pseudocohnilembus persalinus* and *Paralembus digitiformis*) is closer to Hymenostomatia (*Tetrahymena* spp. and *Ichthyophthirius multifiliis*).
The taxonomic relationships among Peritrichia, Scuticociliatia and Hymenostomatia in Oligohymenophorea were previously controversial [11], and phylogenomic analysis has revealed that Scuticociliatia is sister to Hymenostomatia, but not Peritrichia. Concerning Protocruzia adherens, the marine benthic ciliate genus Protocruzia of has a long history of ambiguous taxonomy. Some molecular studies have shown that Protocruzia could be more closely related to Spirotrichea [14,37]. Similarly, some researchers have proposed that Protocruzia be given its own lineage status, although they failed to define the taxon [19]. Recently, Gentekaki et al. [12] discussed the taxon Protocruzia in detail based on phylogenomic analysis and postulated that Protocruzia is not a spirotrich. This incongruence may be caused by insufficient data, long branch attraction artifact or species imbalance. High-throughput sequencing can provide plentiful nucleotide sequence information to be used for phylogenomic analysis, which can provide more representative genes to construct phylogenetic trees and is

![Figure 3. Phylogenetic tree based on the SSU rDNA sequences of ciliates. The ML tree was implemented by MEGA6 using the Tamura 3-parameter model, and Bayesian Inference by Mrbayes 3.2 with the GTR + G + I model. Two Plasmodium spp. were used as outgroups. The scale bar corresponds to 0.02 expected substitutions per site. The sequence extracted from the present transcriptome is in red script with yellow shading.](image-url)
conducive for us to reveal the true evolutionary relationships among species. Therefore, phylogenomic analysis is superior to phylogenetic analysis derived from single genes to discuss taxonomic placement when sufficient omics data are available. As for the uncultivable species, the sample preparation for high-throughput sequencing is a great challenge. Fortunately, the rapidly developing single-cell sequencing technology has helped to resolve this problem.

No interesting gene families or metabolic pathways were found by Gene Ontology (GO) annotation, KEGG pathway analysis (data not shown) and searching against the non-redundant protein (nr) database (data not shown), which probably results from the fact that the present transcriptome is incomplete, and complete annotation for analyzing the function of genes is unavailable. Therefore, further research should aim to sequence the genome of B. ctenopharyngodoni and resquence its transcriptome.

Thus far, more than 8,000 ciliates have been discovered. Unfortunately, most of them are uncultivable, such as rumen ciliates, Trichodina spp., Balantidium spp. and so on, making it difficult to ascertain their taxonomy. Our present results suggest that the methods in this paper are a sound way of obtaining sufficient omics data for phylogenomics analysis to solve this problem.

Supplementary materials

Table S1 List of the omics data and SSU rDNA sequences used in the present study. Figure S1 The flowchart for obtaining transcriptome without contaminations and redundancy. Figure S2 Sequence alignments of SSU rDNA of B. ctenopharyngodoni. Black script, the length of fragment; blue script, the start site and termination site of alignment; red script, the identity of alignment; gray bar, second transcripts extracted from assembled transcriptome; green bar, GU48080 (B. ctenopharyngodoni small subunit ribosomal RNA gene, partial sequence); yellow bar, KU170972 (B. ctenopharyngodoni internal transcribed spacer 1, partial sequence); 5.8S ribosomal RNA gene, complete sequence; and internal transcribed spacer 2, partial sequence). Figure S3 Gene Ontology (GO) annotation of the final transcriptome. Figure S4 The guanine-cytosine (GC)-content distribution of B. ctenopharyngodoni’s transcriptome. Yellow line, the GC content of preliminary redundant transcriptome; green line, the GC content of redundant transcriptome; red line, the GC content of final transcriptome. The Supplementary Material is available at http://www.parasite-journal.org/10.1051/parasite/2017043/olm

Acknowledgements. This work was supported by the Natural Science Foundation of China (Project Nos. 31372168, 31471978, and 31501854) to WM, ML and JF.

References

1. Aury JM, Jaillon O, Duret L, Noel B, Jubitin C, Porcell BM, Seguerra B, Daubin V, Anthouard V, Aiach N, Arraiz O, Billaut A, Beisson J, Blanc I, Bouhouche K, Camara F, Duharcourt S, Guigo R, Gogendeau D, Katinka M, Keller AM, Kissmehl R, Klotz C, Kolf F, Le Monel A, Lepere G, Malinsky S, Nowacki M, Nowak JK, Plattner H, Poulain J, Ruiz F, Serrano V, Zagulski M, Dessen P, Betermier M, Weissenbach J, Scarpelli C, Schachter V, Sperling L, Meyer E, Cohen J, Wineker P. 2006. Global trends of whole-genome duplications revealed by the ciliate Paramecium tetraurelia. Nature, 444(7116), 171-178.

2. Chen CL. 1955. The protozoan parasites from four species of Chinese pond fishes: Ctenopharyngodon idellus, Mylopharyngodon ptenothecus, Aristichthys nobilis and Hypophthalmichthys molitrix, I. The protozoan parasites of Ctenopharyngodon idellus. Acta Hydrobiologica Sinica, 2, 123-164.

3. Chen X, Zhao X, Liu X, Warren A, Zhao F, Miao M. 2015. Phylogenomics of non-model ciliates based on transcriptomic analyses. Protein Cell, 6(5), 373-85.

4. Claparède E, J Lachmann. 1858. Études sur les infusoires et les rhizopodes. Mémoires de l’Institut Genevois, 5, 1-260.

5. Comas I, Moya A, González-Candelas F. 2007. From Phylogenetics to Phylogenomics. The evolutionary relationships of insect endosymbiotic γ-Proteobacteria as a test case. Systematic Biology, 56(1), 1-16.

6. Coyne RS, Hannick L, Shanmugam D, Hostetler JB, Bram D, Joardar VS, Johnson J, Radune D, Singh I, Badger JH, Kumar U, Saier M, Wang YF, Cai H, Gu JY, Mather MW, Vaidya AB, Wilkes DE, Rajagopalan V, Asai DJ, Pearson CG, Findly RC, Dickerson HW, Wu M, Martens C, Van de Peer Y, Roos DS, Cassidy-Hanley DM, Clark TG. 2011. Comparative genomics of the pathogenic ciliate Ichthyophthirius multifiliis, its free-living relatives and a host species provide insights into adoption of a parasitic lifestyle and prospects for disease control. Genome Biology, 12(10), 1-26.

7. Crawford NG, Faircloth BC, McCormack JE, Brumfield RT, Winker K, Glenn TC. 2012. More than 1000 ultra-conserved elements provide evidence that turtles are the sister group of archosaurs. Biology Letters, 8(5), 783-786.

8. Davidson NM, Oshlack A. 2014. Corset: enabling differential gene expression analysis for de novo assembled transcriptomes. Genome Biology, 410, 1-14.

9. Edgar RC. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Research, 32(5), 1792-1797.

10. Eisen JA, Coyne RS, Wu M, Wu D, Thiagarajan M, Worman JR, Badger JR, Ren Q, Amedeo P, Jones KM, Tallon LJ, Delcher AL, Salzberg SL, Silva JC, Haas BJ, Majoros WH, Farzad M, Carlton JM, Smith RK, Jr., Garg J, Pearlman RE, Karrer KM, Sun L, Manning G, Elde NC, Turkewitz AP, Asai DJ, Wilkes DE, Wang Y, Cai H, Collins K, Stewart BA, Lee SR, Wlanoowska K, Weinberg R, Ruzzo WL, Wloga D, Gaertig J, Frankel JT, Tsao CC, Gorovsky MA, Keeling PJ, Winker K, Patron NJ, Cherry JM, Stover NA, Krieger CJ, del Toro C, Ryder HF, Williamson SC, Barbeau RA, Hamilton EP, Orias E. 2006. Macronuclear genome sequence of the ciliate Tetrahymena thermophila, a model eukaryote. PLoS Biology, 4(9), 1620-1642.

11. Feng JM, Jiang CQ, Warren A, Tian M, Cheng J, Liu GL, Xiong J, Miao W. 2015. Phylogenomics analyses reveal subclass Scuticociliatia as the sister group of subclass Hymenostomatia within class Oligohymenophorea. Molecular Phylogenetics and Evolution, 90, 104-111.
12. Gentekaki E, Kolisko M, Boscaro V, Bright KJ, Dini F, Di Giuseppe G, Gong Y, Miceli C, Mdeo L, Molesta RE, Petroni G, Pucciarelli S, Roger AJ, Strom SL, Lynn DH. 2014. Large-scale phylogenetic analysis reveals the phylogenetic position of the problematic taxon Protocruzia and unravels the deep phylogenetic affinities of the ciliate lineages. Molecular Phylogenetics and Evolution, 78, 36-42.

13. Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, Adiconis X, Fan L, Raychowdhury R, Zeng Q, Chen Z, Mauceli E, Haconen N, G nirke A, Rhind N, di Palma F, Birren BW, Nusbaum C, Lindblad-Toh K, Friedman N, Regew A. 2011. Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nature Biotechnology, 29(7), 644-652.

14. Hammerschmidt B, Schlegel M, Lynn DH, Leipe DD, Sogin ML, Raikov IB. 1996. Insights into the evolution of nuclear dualism in the ciliates revealed by phylogenetic analysis of rRNA sequences. Journal of Eukaryotic Microbiology, 43, 225-230.

15. Hirotugun Akaile M, Ieee. 1974. A new look at the statistical model identification. IEEE Transactions on Automatic Control, Ac-19(6), 716-722.

16. Jones P, Binns D, Chang HY, Fraser M, Li W, McAnulla C, Micheel M, Mitchell A, Nuka G, Pesse S, Quinn AF, Sangrador-Vegas A, Scheremetjew M, Yong SY, Lopez R, Hunter S. 2014. InterProScan 5: genome-scale protein function classification. Bioinformatics, 30(9), 1236-40.

17. Lartillot N, Leache AD, Chavez AS, Jones LN, Grummer JA, Gottscho AD, Linkem CW. 2015. Phylogenomics of phrynosomatid lizards: conflicting signals from sequence capture versus restriction site associated DNA sequencing. Genome Biology, 25(17), 2286-2288.

18. Leache AD, Chavez AS, Jones LN, Grummer JA, Gotschot AD, Linkem CW. 2015. Phylogenomics of phrynosomatid lizards: conflicting signals from sequence capture versus restriction site associated DNA sequencing. Genome Biology and Evolution, 7(3), 706-719.

19. Li L, Thorsten S, Kyoon SM, Al-Rasheid Khaled AS, Al-Khedhairy Belulaziz A, Song W. 2010. Protocruzia, a highly ambiguous ciliate (Prototaxa; Ciliophora): very likely an ancestral form for Heterotricha, Colpodoea or Spirotrichae? With reevaluation of its evolutionary position based on multigene analyses. Science China Life Sciences, 53(1), 131-8.

20. Li M, Li D, Wang J, Zhang J, Gu Z, Gong X. 2007. Light and scanning electron microscopic study of Balantidium ctenopharyngodonii from small intestine of the grass carp (Chen, 1955) (Class: Litostomatea) from Honghu Lake, China. Korean Journal of Parasitology, 51(4), 427-431.

21. Li M, Li W, Zhang L, Wang C. 2013. Balantidium honghuensis n. sp. (Ciliophora: Trichostomatidae) from the rectum of Rana nigromaculata and R. limnocharis from Honghu Lake, China. Korean Journal of Parasitology, 51(4), 427-431.

22. Li M, Wang C, Grim JN, Wang J, Gong X. 2014. Ultrastructural study of Balantidium ctenopharyngodonii Chen, 1955 (Class: Litotomata) from China with an emphasis on its vestibulum. Acta Parasitologica, 59(3), 413-419.

23. Lynn D, Sogin ML. 1988. Assessment of phylogenetic relationships among ciliated protists using partial ribosomal RNA sequences derived from reverse transcripts. BioSystems, 21, 249-254.

24. Marton S, Banyai K, Forro B, Lengyel G, Szekely C, Varga A, Molnar K. 2016. Molecular genetic investigations on Balantidium ctenopharyngodonii Chen, 1955, a parasite of the grass carp (Chen, 1955) (Class: Litostomatea) from Acta Veterinaria Hungarica, 64(2), 213-221.

25. Mccormack JE, Faircloth BC. 2013. Next-generation phylogenetics takes root. Molecular Ecology, 22, 19-21.

26. Miao W, Fan W, Yu Y, XiuyuanZhang, Shen Y. 2004. Phylogenetic relationships of the subclass Peritrichia (Oligokaryophorea, Ciliophora) inferred from small subunit rRNA gene sequences. Journal of Eukaryotic Microbiology, 51(2), 180-186.

27. Miao W, Yu Y, Shen Y. 2001. Phylogenetic relationships of the subclass Peritrichia (Oligokaryophorea, Ciliophora) with emphasis on the genus Epistylistis, inferred from small subunit rRNA gene sequences. Journal of Eukaryotic Microbiology, 48(5), 583-587.

28. Molnar K, Reinhardt M. 1978. Intestinal lesions in grass carp (Chen, 1955) (Valenciencie) infected with Balantidium ctenopharyngodonii Chen. Journal of Fish Diseases, 1 (151-156).

29. Negrisolo E, Kuhl H, Forcato C, Vitulo N, Reinhardt R, Patarnello T, Bargelloni L. 2010. Different phylogenomic approaches to resolve the evolutionary relationships among model fish species. Molecular Biology and Evolution, 27(12), 2757-2774.

30. Philippe H, Desuc F, Brinkmann H, Lartillot N. 2005. Phylogenomics. Annual Review of Ecology, Evolution, and Systematics, 36(1), 541-562.

31. Picelli S, Bjorklund AK, Faridani OR, Sagasser S, Winberg G, Sandberg R. 2013. Smart-seq2 for sensitive full-length transcriptome profiling in single cells. Nature Methods, 10(11), 1096-1098.

32. Rice PM, Bleasby AJ, Ison J. 2011. EMBOSS User's Guide: Practical Bioinformatics. Cambridge University Press.

33. Rokas A, Williams BL, King N, Carroll SB. 2003. Genome-scale approaches to resolving incongruence in molecular phylogenies. Nature, 425, 798-804.

34. Ronquist F, Teslenko M, van der Mark P, Ayres DL, Darling A, Holma S, Larget B, Liu L, Suchard MA, Huelsenbeck JP. 2012. MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. Systematic Biology, 61(3), 539-542.

35. Roure B, Rodriguez-Ezepeleta N, Philippe H. 2007. SCAFoS: a tool for selection, concatenation and fusion of sequences for phylogenomics. BMC Evolutionary Biology, 7 Suppl 1, S2.

36. Shi X, Frankel J. 1990. Morphology and development of mirror image doubles of Styloynchia mytilus. Journal of Protozoology, 37(1), 1-13.

37. Shin MK, Hwang UW, Kim W, Wright A-DG, Krawczyk C, Lynn DH. 2000. Phylogenetic position of the ciliates Phacodinium (Order Phacodiniida) and Protocruzia (Subclass Protocruziaida) and systematics of the spirotrich ciliates examined by small subunit ribosomal RNA gene sequences. European Journal of Protistology, 36(3), 293-302.

38. Stamatakis A. 2006. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. Bioinformatics, 22(21), 2688-2690.

39. Swart EC, Bracht JR, Magrini V, Minx P, Chen X, Zhou Y, Khurana JS, Goldman AD, Nowacki M, Schotanus K, Aravind L, Hahn M, Knight R, Normark M, Stetter K, Harmsen W, Stueben J, Stenstrom A, Wilson RK, Mardis ER, Landweber LF. 2013. The Oxytricha trifallax eukaryotic genome with 16,000 tiny chromosomes. PLoS Biology, 11(1), e1001473.
40. Talavera G, Castresana J. 2007. Improvement of phylogenies after removing divergent and ambiguously aligned blocks from protein sequence alignments. Systematic Biology, 56(4), 564-577.
41. Tamura K, Stecher G, Peterson D, Filipski A, Kumar S. 2013. MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. Molecular Biology and Evolution, 30(12), 2725-2729.
42. Woese CR. 1987. Bacterial Evolution. Microbiological Research, 51(2), 221-271.
43. Xiong J, Wang G, Cheng J, Tian M, Pan X, Warren A, Jiang C, Yuan D, Miao W. 2015. Genome of the facultative scuticociliatosis pathogen Pseudocohnilembus persalinus provides insight into its virulence through horizontal gene transfer. Scientific Reports, 5, 15470.
44. Ye J, Fang L, Zheng H, Zhang Y, Chen J, Zhang Z, Wang J, Li S, Li R, Bolund L, Wang J. 2006. WEGO: a web tool for plotting GO annotations. Nucleic Acids Research, 34(Web Server issue), W293-7.

Cite this article as: Sun Z, Jiang C, Feng J, Yang W, Li M, Miao W. 2017. Phylogenomic analysis of Balantidium ctenopharyngodoni (Ciliophora, Litostomatea) based on single-cell transcriptome sequencing. Parasite 24, 43