Received: 31 October 2018 | Revised: 15 June 2019 | Accepted: 6 July 2019

DOI: 10.1111/zsc.12380

ORIGINAL ARTICLE

Zoologica Scripta WILEY

Further analyses on the evolutionary "key-protist" *Halteria* (Protista, Ciliophora) based on transcriptomic data

Chundi Wang[1,2] | Ying Yan[1,2] | Xiao Chen[1,2,3] | Saleh A. Al-Farraj[4] | Hamed A. El-Serehy[4] | Feng Gao[1,2]

[1]Institute of Evolution & Marine Biodiversity, Ocean University of China, Qingdao, China

[2]Key Laboratory of Mariculture (Ocean University of China), Ministry of Education, Qingdao, China

[3]Department of Genetics and Development, Columbia University Medical Center, New York, NY, USA

[4]Zoology Department, College of Science, King Saud University, Riyadh, Saudi Arabia

Correspondence
Feng Gao, Institute of Evolution & Marine Biodiversity, Ocean University of China, Qingdao, China.
Email: gaof@ouc.edu.cn

Funding information
Marine S&T Fund of Shandong Province for Pilot National Laboratory for Marine Science and Technology (Qingdao), Grant/Award Number: 2018SDKJ0406-1; Young Elite Scientists Sponsorship Program by CAST, Grant/Award Number: 2017QNRC001; National Natural Science Foundation of China, Grant/Award Number: 31772428 and 31702009; Fundamental Research Funds for the Central Universities, Grant/Award Number: 201841013 and 201762017; Deanship of Scientific Research at King Saud University, Grant/Award Number: RG-1435-018

Abstract

Considerable discordance between morphologies and molecules, which can be caused by convergent morphologies, cryptic species or plastic phenotypes, has been frequently detected, especially in microbes. One of the best examples could be the well-known halteriid ciliates, whose systematic position remains highly unresolved, because it resembles oligotrichs in terms of morphology while associates with hypotrichs in phylogeny based on maker genes. In the present work, we report the deep sequencing and analyses of RNA-seq data from the representative halteriid, *Halteria grandinella.* The results indicate that: 1) the transcriptome includes 92,114 genes (aa ≥ 50) with the N50 of 957 bp, which is much better than the single-cell transcriptome; 2) *H. grandinella* shares more homologous genes in higher sequence identity with hypotrichous oxytrichids than with oligotrichs; 3) the codon usage bias of *H. grandinella* is much more similar with that of oligotrich *Strombidium sulcatum*, and UAA and UAG are reassigned to encode glutamine, which is a common feature of oligotrichs and hypotrichs; and 4) phylogenomic analyses based on 132 orthologs and 47,263 amino acid sites place *H. grandinella* as a sister to hypotrichs while other oligotrichs cluster together. Based on all the information so far available, we thus suggest that *H. grandinella* could be an extremely specialized "oligotrich-like" hypotrich. The similar morphology between halteriids and oligotrichs results from convergent evolution to adapt to the planktonic lifestyle. We hypothesize the evolutionary relationship of the core spirotrichs that halteriids and oxytrichids share the most common ancestor, followed successively by other hypotrichs, oligotrichs/choreotrichs, and euplotids.

KEYWORDS
ciliated protist, *Halteria grandinella*, phylogenomics, stop codon reassignment, transcriptome

1 | INTRODUCTION

Free-living ciliated protists constitute a large group of single-celled eukaryotes and play important roles in microbial food webs and energy flow processes (Kathol, Norf, Arndt, & Weitere, 2009; Liu et al., 2017; Xu, Zhang, & Jiang, 2014). Their highly diverse morphology, unique nuclear dimorphism, distinctive chromosomal fragmentation and special sexual reproduction process (conjugation), have attracted many researchers' attention in the field of taxonomy, ecology, evolution, genomics, epigenetics, etc. (Chen et al., 2016; Gao et al., 2013; Huang et al., 2018; Wang, Sheng, et al.,

Chundi Wang and Ying Yan are contributed equally to this work.

 © 2019 Royal Swedish Academy of Sciences

2017; Wang, Chen, Sheng, Liu, & Gao, 2017; Wang, Zhang, et al., 2017; Xiong et al., 2016; Yan et al., 2018; Yan, Rogers, Gao, & Katz, 2017; Zhang, Wang, Katz, & Gao, 2018; Zhao, Wang, Wang, Liu, & Gao, 2017; Zhao, Yi, Warren, & Song, 2018). Although these topics have been studied for hundreds of years, there are still many unresolved issues, such as the considerable discordance between morphologies and molecules, which can be caused by convergent morphologies, cryptic species, plastic phenotypes or life-cycle stage variation (reviewed in Lahr, Laughinghouse, Oliverio, Gao, & Katz, 2014). One of these examples could be halteriids, a unique group in the phylum Ciliophora, whose systematic position has puzzled taxonomists for decades (Agatha, 2004; Foissner, Muller, & Agatha, 2007; Gao et al., 2016; Hu, Hu, Al-Rasheid, Al-Farraj, & Song, 2011).

Represented by the well-known species, *Halteria grandinella*, halteriids are a cosmopolitan group that lives predominantly in freshwater habitats and is typically planktonic (Foissner et al., 2004; Šimek, Jürgens, Nedoma, Comerma, & Armengol, 2000). They are characterized mainly by the globular body shape, reduced somatic ciliature (cirrus-like "bristles" used for performing conspicuous jumps), apical adoral structure (with the "collar" as an "open" circle) and enantiotropic cell division mode (the cell axes of proter and opisthe not oriented in the same direction, but in opposite orientation) (Figure 1) (Fauré-Fremiet, 1953; Foissner et al., 2007; Petz & Foissner, 1992; Song, 1992). All these characteristics are shared with high similarity with these in the oligotrichs (mostly globular, reduced somatic ciliature and enantiotropic cell division) (Agatha, 2004; Liu et al., 2016; Petz & Foissner, 1992; Song et al., 2018), but differ from these in the hypotrichs (mostly dorsoventrally compressed, ventral cirri and synclastic cell division) (Figure 1) (Chen, Zhao, Shao, Miao, & Clamp, 2017; Lu, Huang, Chen, & Berger, 2018; Luo et al., 2017, 2018). However, phylogenetic analyses based on small subunit ribosomal DNA (SSU rDNA), internal transcribed spacer (ITS), 5.8S rDNA, large subunit ribosomal DNA (LSU rDNA), α-tubulin and actin I genes have suggested that halteriids are closely related to oxytrichids, a highly specialized group of hypotrichs (Gao et al., 2016; Hu et al., 2011; Lynn & Sogin, 1988; Paiva, Borges, Harada, & Silva-Neto, 2009). However, these phylogenetic analyses are based on single or a few genes/loci, which is not convincing enough to address conclusively any controversy.

With the rapid development of the high-throughput sequencing techniques, genomic or transcriptomic analyses can provide comprehensive insights into the evolution of many organisms (Farlow et al., 2015; Kapusta, Suh, & Feschotte, 2017; Shubha et al., 2016). Besides, phylogenomic investigation has become an effective approach to study the evolutionary relationships of ciliates (eg: Chen, Wang, Sheng, Warren, & Gao, 2018; Chen et al., 2015; Feng et al., 2015; Gentekaki et al., 2014; Gentekaki, Kolisko, Gong, & Lynn, 2017; Sheng, He, Zhao, Shao, & Miao, 2018). In order to investigate the systematic position of halteriids, Lynn and Kolisko (2017) sequenced the transcriptome of *H. grandinella* based on single-cell sequencing techniques and found *H. grandinella* clustered with hypotrichs in multigene trees. Here, we sequenced and analysed the culture-based transcriptome of *H. grandinella*, including assembly, annotation, and comparison with the single-cell transcriptome data downloaded from National Center for Biotechnology Information (NCBI, Lynn & Kolisko, 2017). The amino acid codon usage and stop codon reassignment of *H. grandinella* were also analysed and compared with other ciliates. Phylogenomic analysis based on 47,263 amino acid sites from 132 orthologous proteins of 43 taxa was performed to infer the evolutionary relationships among halteriids, hypotrichs and oligotrichs.

2 | MATERIAL AND METHODS

2.1 | Ciliate culture

Halteria grandinella, collected from a pond in Baihuayuan Park (36°04'N, 120°22'E) in Qingdao, China, was identified based on the morphological traits observed both in vivo and via protargol staining (Song, 1992). Its SSU rDNA sequence was amplified and characterized according to Wang, Zhang, et al. (2017), which is 100% identical to the published SSU rDNA sequence of *H. grandinella* (acc. no. MF002423). Cells were isolated, cleaned (i.e. diluted the surroundings multiple times) and cultured in filtered and autoclaved in situ water (from the pond where we found *H. grandinella*) with the monoclonal bacteria *Escherichia coli* as the food source. Cultures were maintained in 600 ml cell culture flasks at 25°C in the incubator.

2.2 | RNA extraction and Illumina sequencing

When the cultures reaching ca. 10^5 cells, the cultures were starved and treated with ampicillin (Antibiotic-Antimycotic, ThermoFisher Scientific, Cat No. 15,240,062, 1ml for 1L) for 24 hr before RNA extraction. Cells were collected by centrifugation at 300g for 5 min, and total RNA was extracted using RNeasy Plus Mini Kit (Qiagen, Hiden, Germany; Lot No. 74,136) according to the kit's instructions.

Poly-A mRNAs were isolated using Sera-Mag Magnetic Oligo (dT) beads from Illumina, and then fragmented with divalent cations. Double-stranded cDNA was synthesized using the SuperScript Double-Stranded cDNA Synthesis kit (Invitrogen, Camarillo, USA; Lot No. 11,917,020) with random hexamer primers in Illumina. Thereafter, the cDNA was subjected to end-repair, phosphorylation, 3' adenylation and adapter ligation. After these steps, cDNA fragments with an insert size of approximately 250 bp were purified and

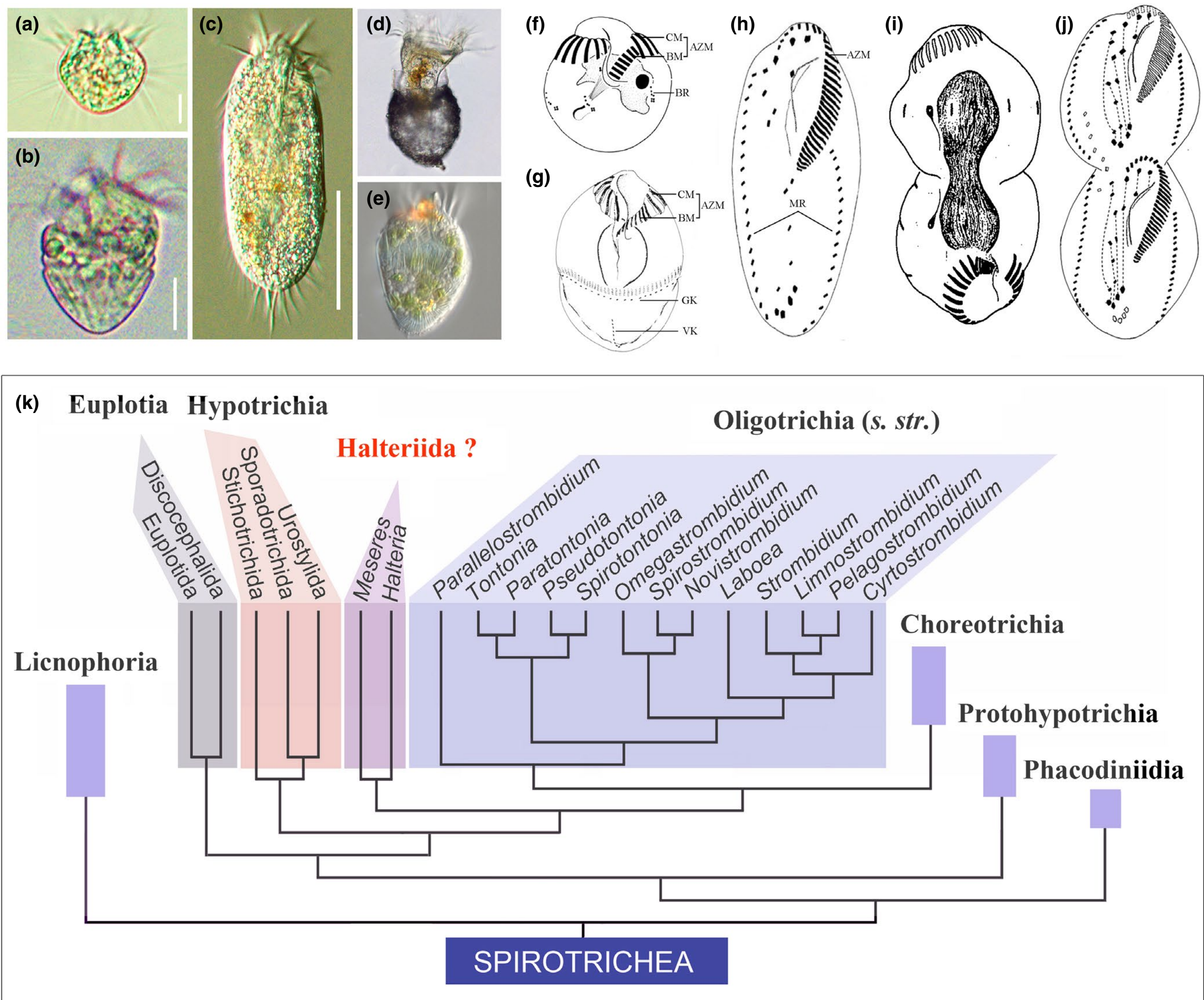


FIGURE 1 Morphology and ontogenesis of oligotrichs (sensu lato) and hypotrichs (a–j) and the relationships within the class Spirotrichea based on morphological data (k). (a, f, i), *Halteria grandinella*; b, (g), *Strombidium sulcatum*; (c, h, j), *Oxytricha trifallax*; (d), *Codonellopsis mobilis*, (e), *Strombidium* sp. (f) is from Agatha and Foissner (2009); (g) is from Song, Wang, and Warren (2000); (h and j) are from Jiang, Ma, and Shao (2013); (i) is from Song (1992). (k) is modified from Agatha (2004) and Foissner et al. (2007). The scale bar in (a), (b) and (c) represent 10μm, 20μm and 50μm, respectively [Colour figure can be viewed at wileyonlinelibrary.com]

enriched to construct the library, which was sequenced for 125 bp at both ends (paired-end) using an Illumina HiSeq 2,500 sequencer (Novogene, Beijing, China).

2.3 | Assembly and gene annotation

The quality of reads was assessed using FastQC v.0.11.4 (Andrews, 2010), and the low-quality reads were removed with the FASTX-Toolkit for both the mass-culture and single-cell RNA-seq reads (acc. no. SRX2788015 and SRX2788016) with the parameters of -q 20 -p 80 (Gordon & Hannon, 2010). The rRNA reads were removed by mapping to *H. grandinella's* rRNA gene sequences (acc. no. AF508759) using Bowtie v.1.0.0 with the parameters of –a, –n 2 and –p 8 (Langmead, Trapnell, Pop, & Salzberg, 2009). The transcriptome was assembled de novo using Trinity v.2.2.0 (Grabherr et al., 2011) with default parameters. Redundancy of contigs was eliminated by CD-HIT v.4.6.1 (Fu, Niu, Zhu, Wu, & Li, 2012) with a 98% sequence identity threshold (-c 0.98, -n 10). The mitochondrial genomes of *Oxytricha trifallax* (acc. no. JN383843), *Tetrahymena pyriformis* (acc. no. AF160864), *Paramecium aurelia* (acc. no. NC001324), *Euplotes minuta* (acc. no. GQ903130) and *Nyctotherus ovalis* (acc. no. GU057832) as well as the bacterial nucleotide database were downloaded from GenBank as mapping references to remove contaminations of mitochondria or bacteria using blastn with the e-value of 1e-2. The regions of coding sequences (CDS) of *H. grandinella* were predicted using Augustus v.3.2.3 with *Tetrahymena thermophila* as model

species (Stanke, Diekhans, Baertsch, & Haussler, 2008) and predicted peptides longer than 50 amino acids were retained. The predicted peptides were annotated using InterProScan v.5.2–45.0 with the parameters of -iprlookup and -goterms (Jones et al., 2014). The Gene Ontology (GO) analyses were carried out with Web Gene Ontology Annotation Plotting v.2.0 (WEGO, http://wego.genomics.org.cn/). The gene clusters were performed by MCL v.14–137 with --abc and -I 1.5 (Enright, Van Dongen, & Ouzounis, 2002). Subsequently, the predicted peptides were blasted with the protein sequences of ciliates downloaded from NCBI and The Marine Microbial Eukaryote Transcriptome Sequencing Project (MMETSP) to investigate the sequence similarity between *H. grandinella* and other ciliates (e-value: 1e-5). The orthologs among *H. grandinella*, *O. trifallax* and *Strombidium sulcatum* were detected via OrthoFinder v.2.2.7 with default parameters (Emms & Kelly, 2015), and the Venn diagram was generated by R v.3.4.0.

2.4 | Codon usage analysis

To analyse the amino acid codon usage of *H. grandinella*, the contigs from the mass-culture and single-cell-combined transcriptome data of *H. grandinella* were blasted against all the ciliate protein sequences in NCBI and MMETSP and the contigs with the e-value ≤ 1e-10, sequence identity ≥ 50% and alignment length ≥ 300 amino acids were retained for analyses. The codon and its corresponding amino acid were extracted by a custom Perl script. The codon usage of *S. sulcatum* was evaluated with the same method as *H. grandinella*. The codon usages of *O. trifallax*, *Stylonychia lemnae*, *Euplotes octocarinatus*, *Tetrahymena thermophila* and *Paramecium tetraurelia* were assessed using their coding sequences published in NCBI.

The coding regions of the ATP-dependent 26S proteasome regulatory subunit and actin II genes were amplified and sequenced according to Wang, Zhang, et al. (2017). The primers are 26S_regulator_F: 5’-ATG CCG TAA TTT TTG GTT AAT-3’, 26S_regulator_R: 5’-CCC TTA CTG ACA AGA GAG TTG-3’ and Actin_II-F: 5’-ATG CGT ATC AGT GAT AAA GC-3’, Actin_II-R: 5’-TCA AAT ATG CAG TTA GAA TAA TT-3’, positioned outside the coding regions.

2.5 | Phylogenomic analyses

For the phylogenomic analyses, except for the transcriptome of *H. grandinella*, 10 ciliate genomes, 30 ciliate transcriptomes and two *Plasmodium* genomes as outgroups were included. The sources of the omics data used in the present work are listed in Table S1.

Orthologs were detected employing the reciprocal BLAST hit (RBH) approach according to Chen et al. (2015). The predicted peptides of *H. grandinella* were blasted to the peptide dataset of other species by blastp with the parameters of e-value ≤ 1e-10, identity ≥ 30%, length ≥ 50 aa, and vice versa. Only the proteins present in more than 90% of the taxa (at least 39 different species of our 43 species) were retained for further analyses. The protein sequences were aligned using ClustalW v.2.1 with the bootstrap value of 1,000 (Sievers et al., 2011). The ambiguous alignment sites were masked using Gblocks v.0.91b (-b2 = 0.65, -b3 = 10, -b4 = 5, -b5 = a) (Castresana, 2000). A final dataset containing 47,263 amino acid sites from 132 proteins was used for maximum-likelihood (ML) and Bayesian inference (BI) analyses. The ML analysis was conducted in CIPRES Science Gateway with RAxML-HPC v.8 on XSEDE (Stamatakis, 2014) with the model of LG + F + Γ4 and 1,000 replicates. BI analysis was implemented by PhyloBayes v.4.1 (Lartillot, Lepage, & Blanquart, 2009) under a mixture model of CAT + POI + Γ4. Four Markov chain Monte Carlo (MCMC) chains were run for 1,000 generations. The bpcomp programme within PhyloBayes was applied to compare the discrepancy of bipartition frequencies and to yield a consensus tree (convergence Maxdiff = 0.00722892). The ML tree based on single gene was built in CIPRES Science Gateway using RAxML-HPC v.8 on XSEDE with the same parameters of the concatenated tree. MEGA v.6.06 (Tamura, Stecher, Peterson, Filipski, & Kumar, 2013) was utilized to visualize tree topologies.

As a comparison, phylogenetic analyses based on SSU rDNA sequences (accession numbers as in Figure 4b) were also performed according to Wang, Zhang, et al. (2017). Briefly, the 44 SSU rDNA sequences were aligned with the GUIDANCE Server (http://guidance.tau.ac.il/ver2/) with default parameters (Sela, Ashkenazy, Katoh, & Pupko, 2015) and then manually modified using BioEdit v.7.0.1 (Hall, 1999) to generate a matrix of 44 taxa with 1,669 nucleotide sites. The ML analysis was performed in CIPRES Science Gateway using RAxML-HPC2 on XSEDE v.8.1.24 with the model of GTR + I + G (Stamatakis, Hoover, & Rougemont, 2008). The reliability of internal branches was assessed with a non-parametric bootstrap method with 1,000 replicates. The BI tree was constructed using MrBayes on XSEDE v.3.2.6 tool in CIPRES Science Gateway with the model of GTR + I + G, which was selected by MrModeltest v.2.0 (Nylander, 2004). Markov chain Monte Carlo simulations were run with two sets of four chains for 6,000,000 generations with a sampling frequency of 100 generations, and the first 10% of generations were discarded as burn-in.

3 | RESULTS

3.1 | Comparisons between single-cell and culture-based datasets of *Halteria grandinella* transcriptome

The high-throughput RNA-seq data of *H. grandinella* from mass-culture (present study) and single-cell downloaded

from NCBI were assembled via Trinity (Table 1). For the RNA-seq data from mass-culture of *H. grandinella* in the current work, a total of 31,412,761 paired-end reads (125 bp for each read) were produced (7,853,190,250 bases, GC: 49%). After removing low-quality reads, 90.9% of reads were retained and assembled into 129,114 contigs (maximum length: 14,811 bp, N50: 1,159 bp) using Trinity. After removing the redundancy and contamination of bacteria, rRNA and mitochondria, 117,978 contigs were retained (Table 1).

Two datasets from single cell (single-cell 1 and single-cell 2) were analysed both separately and after combination. Compared with the data from the mass-culture, the single-cell transcriptome recovered much less contigs, and even the combined data did not exhibit much improvement (117,978 vs. 71,498 contigs). The N50 and maximum length of the mass-culture transcriptome are higher than those of single-cell transcriptomes. The number of predicted genes longer than 50 amino acids in the mass-culture is about two-fold higher than that in the single-cell-combined transcriptome (92,114 vs. 44,219). After CDS prediction, 21,782 and 12,007 peptide sequences were annotated in the mass-culture and single-cell transcriptomes, respectively (Figure 2d). In the annotation results, single-cell transcriptome has higher percentage of genes in cellular component and biological process, while mass-culture transcriptome data have higher proportion of genes in molecular function.

To determine whether the number of reads affects the assembly quality and size, we used the same number of reads (7,433,129 pair-end reads, which is all of the combined two single-cell RNA-seq reads and part of the mass-culture RNA-seq reads) with the same length (125 bp), and assembled them via Trinity (Table S2). Similarly, the N50 and maximum length of the mass-culture transcriptome are still higher than those of single-cell transcriptomes (1,234 vs. 833 bp). The number of predicted genes longer than 50 amino acids in the mass-culture is much higher than that in the single-cell-combined transcriptome (55,798 vs. 32,805).

The predicted protein sequences of the mass-culture transcriptome were binned into 47,158 protein clusters, more than that of the combined single-cell transcriptome (14,313). The largest protein cluster of the mass-culture transcriptome contained 1,790 proteins with a protein kinase function. For the peptides from both the mass-culture and single-cell-combined transcriptomes, 54,752 protein clusters were produced.

3.2 | Analysis of the mass-culture transcriptome

The culture-based dataset was analysed further to assess the content of the *H. grandinella* transcriptome. The Web Gene Ontology Annotation (WEGO) analyses of the annotated proteins mainly focused on cell and intracellular in cellular component, binding in molecular function and metabolic process in biological process.

The predicted peptides were first blasted with the protein sequences of eight other ciliate genomes to investigate

TABLE 1 The assembly comparison of *Halteria grandinella* between mass-culture and single-cell transcriptome data

	Mass-culture	Single-cell-1	Single-cell-2	Single-cell-combined
Assembly size (M)	97.3	33.4	24.5	41.8
Contigs (*n*)	129,114	53,874	50,021	75,270
N50 (bp)	1,159	746	558	658
Maximum length (bp)	14,811	5,524	3,349	5,710
Remove redundancy	119,845	51,631	48,664	72,325
Remove bacteria	118,041	51,149	48,120	71,532
Remove rDNA	117,993	51,136	48,105	71,512
Remove mitochondria	117,978	51,123	48,090	71,498
Predicted gene number (aa ≥ 50)	92,114	33,768	29,926	44,219
Predicted gene size (M)	61.0	19.8	14.3	23.6
GC (%)	48.4	44.5	45.6	44.7

Note: Mass-culture, the transcriptome data from mass cells obtained in the present work; Single-cell-1 and Single-cell-2, the transcriptome data from two cells using the single cell sequencing technology from Lynn and Kolisko, (2017). Single-cell-combined, merge the data of Single-cell-1 and Single-cell-2.

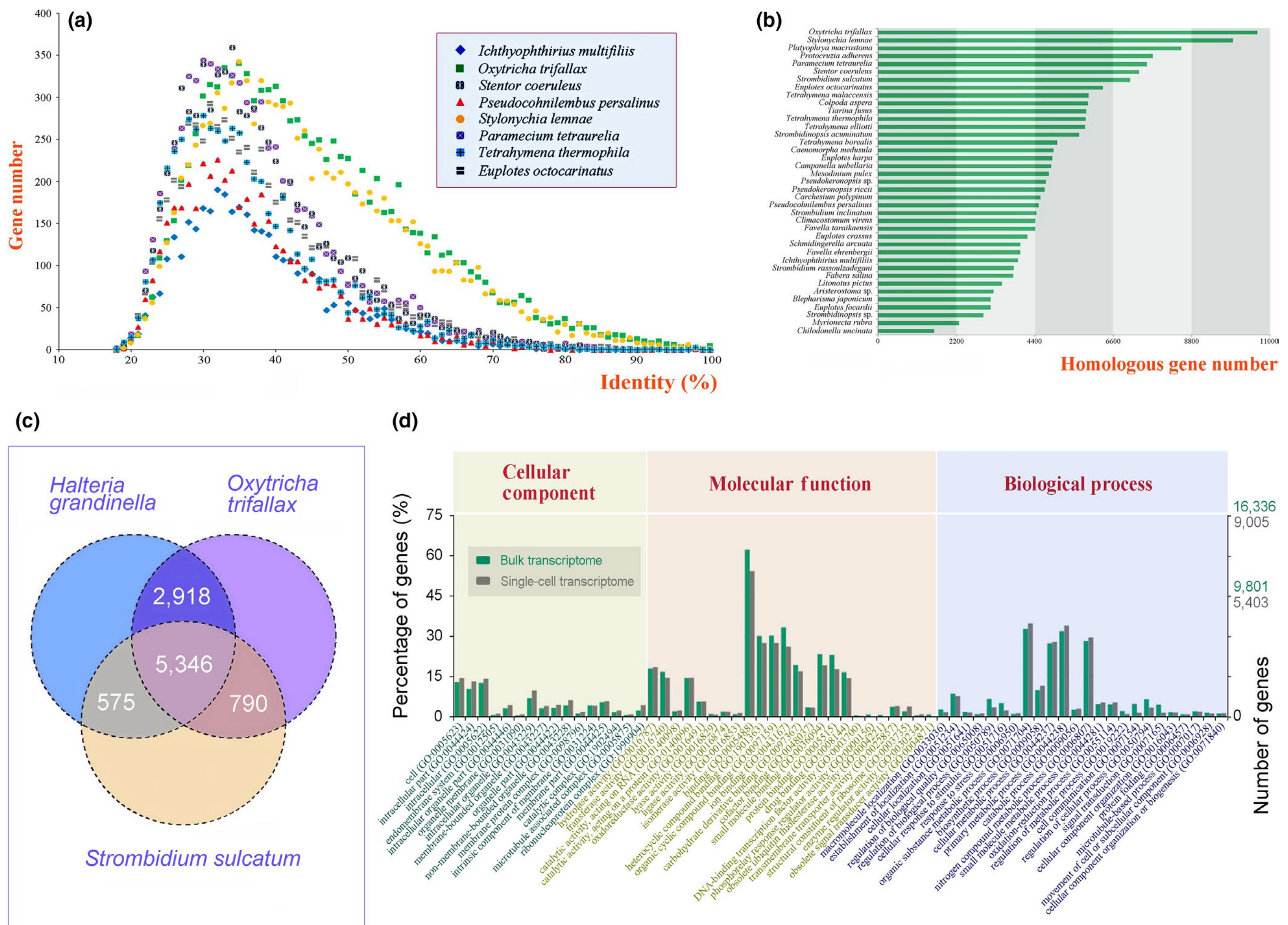


FIGURE 2 The analyses of *Halteria grandinella* transcriptome. (a), the distribution of the protein sequence identity between *H. grandinella* and eight ciliate genomes; (b), the homologous gene numbers between *H. grandinella* and other ciliates; (c), Venn diagram of orthologous gene groups among *H. grandinella*, *Strombidium sulcatum* and *Oxytricha trifallax*; (d), GO term enrichment analysis of the predicted genes in mass-culture and single-cell transcriptomes of *H. grandinella* [Colour figure can be viewed at wileyonlinelibrary.com]

sequence similarity (Figure 2a). The sequence identities between *H. grandinella* and the two hypotrichs *O. trifallax* and *S. lemnae* were higher than that with other genomes. The predicted peptides of *H. grandinella* were then blasted with the protein sequences of ciliates that are available for both genomes and transcriptomes downloaded from NCBI and MMETSP (Figure 2b). *O. trifallax* and *S. lemnae* are the foremost two species sharing homologous genes with *H. grandinella* in the amount of 10,645 and 9,964, respectively.

To gain further insights into the relationships among *H. grandinella*, oligotrichs and hypotrichs, the proteins predicted from the *H. grandinella* transcriptome were grouped with the transcriptome or genome of the oligotrich *S. sulcatum* and hypotrich *O. trifallax* using OrthoFinder (Figure 2c). *H. grandinella*, *O. trifallax* and *S. sulcatum* shared 5,346 ortholog groups. Except the orthologs shared by these three species, *H. grandinella* and *O. trifallax* shared 2,918 specific ortholog groups, far more than that shared specifically by *H. grandinella* and *S. sulcatum* (575).

3.3 | Codon usage in *Halteria grandinella*

To evaluate the amino acid codon usage of *H. grandinella*, the mass-culture transcriptome data of *H. grandinella* were blasted against all the ciliate protein sequences in NCBI and MMETSP, and the 6,141 contigs with sequence identity ≥ 50% and alignment length ≥ 300 amino acids were retained for analyses (Figure 3a, Table S3). Among the 3,105,110 amino acid codons of the 6,141 contigs, AAG and GAG were the two most common amino acid codons at a percentage of 5.55% and 4.66%, respectively. Leucine (L) had the highest proportion at 9.41% (292,248) among the 20 amino acids, and the least frequent amino acid was Tryptophan (W, 0.99%, 30,681). Compared with A/U-ending codons, G/C-ending codons were more frequently used for the 6,141 contigs (G/C-ending: 1,866,739 vs. A/U-ending: 1,238,251).

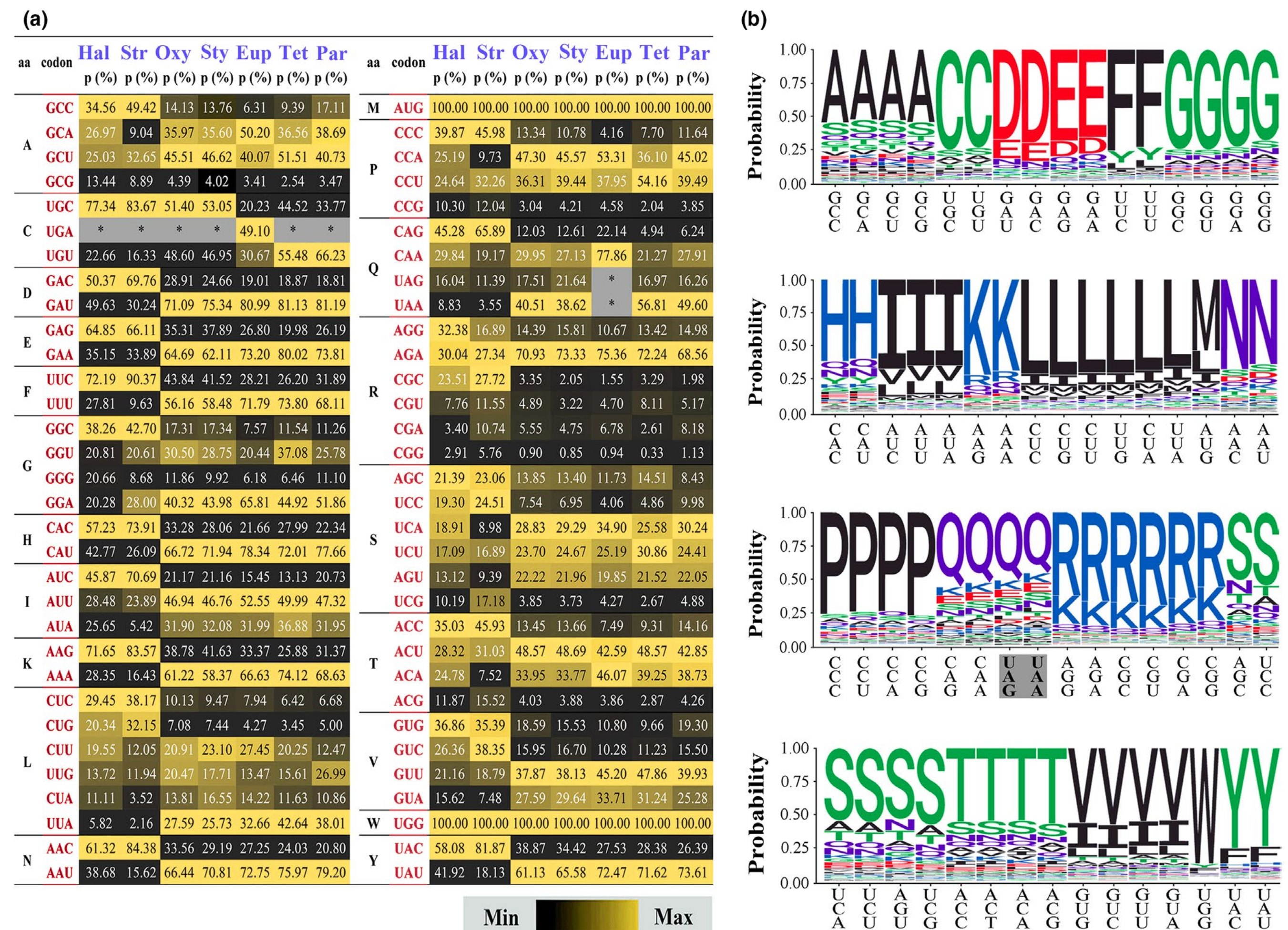
Among the amino acid codons in the coding regions, 14,447 UAA, 26,228 UAG and 120 UGA were detected

(Figure 3a). For both UAA and UAG, the highest frequency amino acid is Glutamine (Q) with at proportions of 45.55% and 50.93%, respectively (Figure 3b). The 120 UGA and the very low frequency (<5%) amino acids (except for Glutamine) encoded by UAA and UAG may be subject to noise. Besides, we sequenced the macronuclear sequences of ATP-dependent 26S proteasome regulatory subunit gene and actin II gene. In these two genes, there are UAA and UAG positioned within the coding regions (mostly corresponding to Glutamine), while UGA is used as a stop codon (Figure S1).

There is an obvious usage bias among the synonymous codons in our results (Figure 3a). For example, for the four amino acid codons encoding for Glutamine, UAA and UAG have lower percentage compared with CAG and CAA, especially UAA that encodes only 8.83% with Glutamine. The codon usage of single-cell-combined transcriptome has similar tendency with that of mass-culture data, although has a few differences (Table S3). Amino acid codon usage of *S. sulcatum*, *O. trifallax*, *S. lemnae*, *E. octocarinatus*, *T. thermophila* and *P. tetraurelia* was also analysed and compared. Interestingly, for the 18 amino acids that have synonymous codons, the codon usage bias of *H. grandinella* is much more similar to that of *S. sulcatum*, while the bias of the other five species are more akin to each other, despite not being phylogenetically related to each other.

3.4 | Phylogenomic analyses

We used the RBH approach to identify orthologs between *H. grandinella* and the other 42 species (40 ciliates and two *Plasmodium* species as outgroups). A final dataset containing 132 proteins with 47,263 amino acids from 43 taxa was used to conduct phylogenomic analyses with both algorithms of maximum likelihood (ML) and Bayesian inference (BI). The concatenated trees of ML and BI algorithms shared a similar topology; therefore, only the ML tree was shown in Figure 4a with the support values from both ML and BI methods.



(a)

aa	codon	Hal p (%)	Str p (%)	Oxy p (%)	Sty p (%)	Eup p (%)	Tet p (%)	Par p (%)
A	GCC	34.56	49.42	14.13	13.76	6.31	9.39	17.11
	GCA	26.97	9.04	35.97	35.60	50.20	36.56	38.69
	GCU	25.03	32.65	45.51	46.62	40.07	51.51	40.73
	GCG	13.44	8.89	4.39	4.02	3.41	2.54	3.47
C	UGC	77.34	83.67	51.40	53.05	20.23	44.52	33.77
	UGA	*	*	*	*	49.10	*	*
	UGU	22.66	16.33	48.60	46.95	30.67	55.48	66.23
D	GAC	50.37	69.76	28.91	24.66	19.01	18.87	18.81
	GAU	49.63	30.24	71.09	75.34	80.99	81.13	81.19
E	GAG	64.85	66.11	35.31	37.89	26.80	19.98	26.19
	GAA	35.15	33.89	64.69	62.11	73.20	80.02	73.81
F	UUC	72.19	90.37	43.84	41.52	28.21	26.20	31.89
	UUU	27.81	9.63	56.16	58.48	71.79	73.80	68.11
G	GGC	38.26	42.70	17.31	17.34	7.57	11.54	11.26
	GGU	20.81	20.61	30.50	28.75	20.44	37.08	25.78
	GGG	20.66	8.68	11.86	9.92	6.18	6.46	11.10
	GGA	20.28	28.00	40.32	43.98	65.81	44.92	51.86
H	CAC	57.23	73.91	33.28	28.06	21.66	27.99	22.34
	CAU	42.77	26.09	66.72	71.94	78.34	72.01	77.66
I	AUC	45.87	70.69	21.17	21.16	15.45	13.13	20.73
	AUU	28.48	23.89	46.94	46.76	52.55	49.99	47.32
	AUA	25.65	5.42	31.90	32.08	31.99	36.88	31.95
K	AAG	71.65	83.57	38.78	41.63	33.37	25.88	31.37
	AAA	28.35	16.43	61.22	58.37	66.63	74.12	68.63
L	CUC	29.45	38.17	10.13	9.47	7.94	6.42	6.68
	CUG	20.34	32.15	7.08	7.44	4.27	3.45	5.00
	CUU	19.55	12.05	20.91	23.10	27.45	20.25	12.47
	UUG	13.72	11.94	20.47	17.71	13.47	15.61	26.99
	CUA	11.11	3.52	13.81	16.55	14.22	11.63	10.86
	UUA	5.82	2.16	27.59	25.73	32.66	42.64	38.01
N	AAC	61.32	84.38	33.56	29.19	27.25	24.03	20.80
	AAU	38.68	15.62	66.44	70.81	72.75	75.97	79.20

aa	codon	Hal p (%)	Str p (%)	Oxy p (%)	Sty p (%)	Eup p (%)	Tet p (%)	Par p (%)
M	AUG	100.00	100.00	100.00	100.00	100.00	100.00	100.00
P	CCC	39.87	45.98	13.34	10.78	4.16	7.70	11.64
	CCA	25.19	9.73	47.30	45.57	53.31	36.10	45.02
	CCU	24.64	32.26	36.31	39.44	37.95	54.16	39.49
	CCG	10.30	12.04	3.04	4.21	4.58	2.04	3.85
Q	CAG	45.28	65.89	12.03	12.61	22.14	4.94	6.24
	CAA	29.84	19.17	29.95	27.13	77.86	21.27	27.91
	UAG	16.04	11.39	17.51	21.64	*	16.97	16.26
	UAA	8.83	3.55	40.51	38.62	*	56.81	49.60
R	AGG	32.38	16.89	14.39	15.81	10.67	13.42	14.98
	AGA	30.04	27.34	70.93	73.33	75.36	72.24	68.56
	CGC	23.51	27.72	3.35	2.05	1.55	3.29	1.98
	CGU	7.76	11.55	4.89	3.22	4.70	8.11	5.17
	CGA	3.40	10.74	5.55	4.75	6.78	2.61	8.18
	CGG	2.91	5.76	0.90	0.85	0.94	0.33	1.13
S	AGC	21.39	23.06	13.85	13.40	11.73	14.51	8.43
	UCC	19.30	24.51	7.54	6.95	4.06	4.86	9.98
	UCA	18.91	8.98	28.83	29.29	34.90	25.58	30.24
	UCU	17.09	16.89	23.70	24.67	25.19	30.86	24.41
	AGU	13.12	9.39	22.22	21.96	19.85	21.52	22.05
	UCG	10.19	17.18	3.85	3.73	4.27	2.67	4.88
T	ACC	35.03	45.93	13.45	13.66	7.49	9.31	14.16
	ACU	28.32	31.03	48.57	48.69	42.59	48.57	42.85
	ACA	24.78	7.52	33.95	33.77	46.07	39.25	38.73
	ACG	11.87	15.52	4.03	3.88	3.86	2.87	4.26
V	GUG	36.86	35.39	18.59	15.53	10.80	9.66	19.30
	GUC	26.36	38.35	15.95	16.70	10.28	11.23	15.50
	GUU	21.16	18.79	37.87	38.13	45.20	47.86	39.93
	GUA	15.62	7.48	27.59	29.64	33.71	31.24	25.28
W	UGG	100.00	100.00	100.00	100.00	100.00	100.00	100.00
Y	UAC	58.08	81.87	38.87	34.42	27.53	28.38	26.39
	UAU	41.92	18.13	61.13	65.58	72.47	71.62	73.61

FIGURE 3 The amino acid and codon usage of *Halteria grandinella*. (a), the codon usage bias of *H. grandinella*. The percentage of synonymous codons is indicated by gradient colours filled from black (minimum) to yellow (maximum). (b), the percentage of the predicted amino acids encoded by UAA and UAG. Hal, *Halteria grandinella*; Str. *Strombidium sulcatum*; Oxy, *Oxytricha trifallax*; Sty, *Stylonychia lemnae*; Eup, *Euplotes octocarinatus;* Tet, *Tetrahymena thermophila*; Par, *Paramecium tetraurelia* [Colour figure can be viewed at wileyonlinelibrary.com]

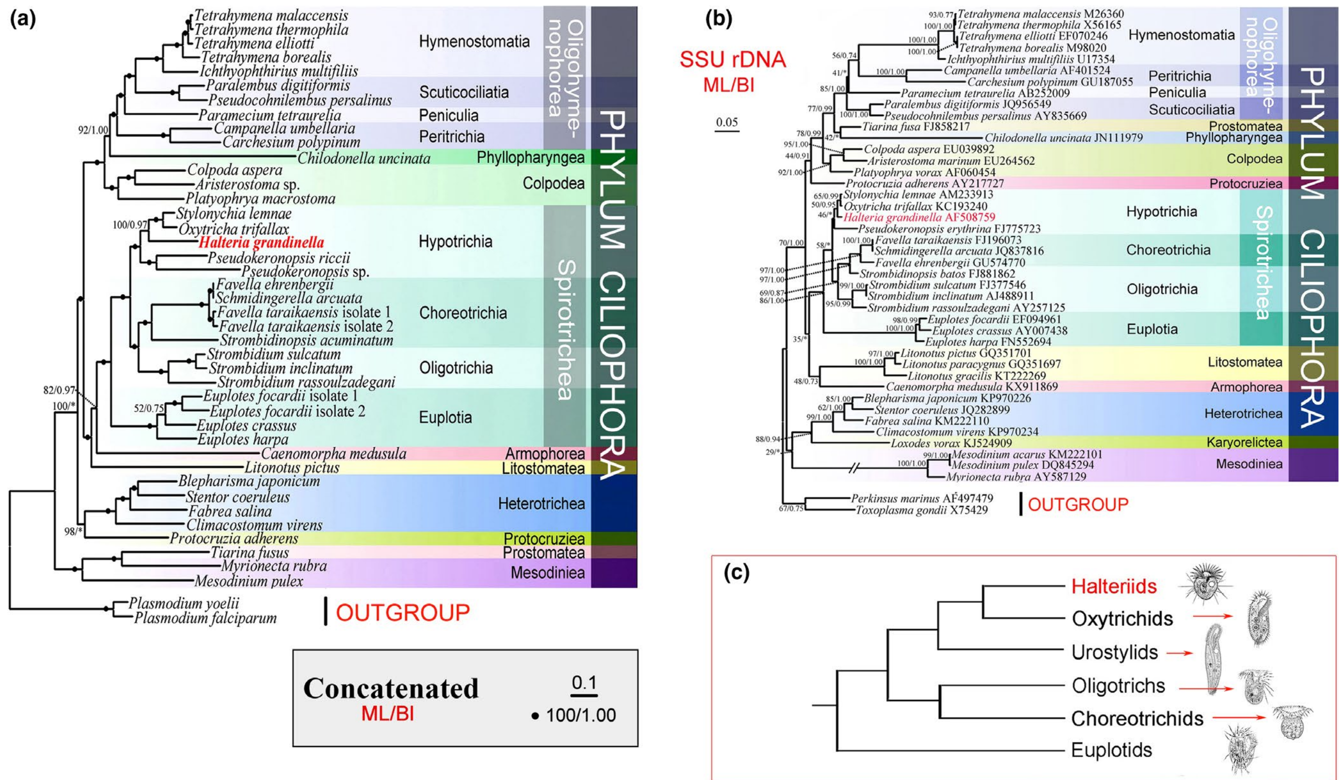


FIGURE 4 Maximum-likelihood trees based on a concatenated alignment of 132 proteins (a) and SSU rDNA sequences (b). Numbers at each node are ML bootstrap support and Bayesian posterior probability, respectively. The black dots indicate the full bootstrap support in both ML and BI trees. The scale bar corresponds to ten and five substitutions per one hundred nucleotide positions in concatenated and SSU rDNA trees, respectively. (c), the hypothesized evolution relationships of the core groups in the class Spirotrichea based on the current phylogenetic reconstruction [Colour figure can be viewed at wileyonlinelibrary.com]

In the concatenated trees, the class Spirotrichea is monophyletic and clusters with the classes Armophorea and Litostomatea. Within Spirotrichea, the subclass Euplotia is placed as an early branching lineage. The subclasses Choreotrichia and Oligotrichia form a fully supported clade, which then groups with Hypotrichia with full support. *H. grandinella* is positioned within Hypotrichia by clustering with *O. trifallax* and *S. lemnae* with strong support (100% ML, 0.97 BI), which is consistent with the SSU rDNA trees (Figure 4b). Among the 132 single-gene trees, *H. grandinella* clusters within Hypotrichia in 74 trees while it groups with oligotrichs in only six trees (data not shown).

4 | DISCUSSION

4.1 | Comparison between mass-culture and single-cell transcriptomes

The single-cell transcriptome recovered much less genes compared with the data from the mass-culture. The previously reported transcriptomes of *H. grandinella* are sequenced from two cells using single-cell sequencing techniques (Lynn & Kolisko, 2017). The assembly size, contig numbers, N50 and predicted genes are much less than those of the mass-culture transcriptome. The first explanation is that the genes expressed in single cell at one time could be less than those expressed in mass-culture. The second explanation might be the PCR bias during cDNA synthesis and amplification (Gawad, Koh, & Quake, 2016; Zhang et al., 2006; Zong, Lu, Chapman, & Xie, 2012). Amplification bias can result in gene loss with the coverage from about 30% to 90% (Gawad et al., 2016; Hou et al., 2015; Huang, Ma, Chapman, Lu, & Xie, 2015; Kolisko, Boscaro, Burki, Lynn, & Keeling, 2014; Lao, Xu, & Straus, 2008; Zhang et al., 2006, 1992; Zong et al., 2012). Besides, the PCR amplification can cause fragmentation of the assembled contigs (X. Chen, Wang, et al., 2018; Zong et al., 2012). However, despite the fact that gene recovery of single-cell sequencing techniques is less efficient than those of the mass-culture method, it is obviously much more efficient than traditional cloning and sequencing to recover transcripts in isolated cells (De Bourcy et al., 2014; Tang et al., 2009; Wu et al., 2014). Moreover, the transcriptome assembled by combining the reads of single-cell-1 and single-cell-2 can recover more genes and has longer N50, which indicates that we can increase the number of parallel samples to obtain more data when using single-cell sequencing techniques.

For GO annotation, mass-cultured and single-cell transcriptomes have some differences, for example single-cell transcriptome recovered higher percentage of genes in cellular component and biological process while mass-culture transcriptome data recovered higher proportion of genes in molecular function (Figure 2d). This might be because that transcriptomes from mass-culture represent the average stage of the whole population, while single-cell datasets reflect living conditions of individual cells, which might be in different living conditions or cell process. Besides, it is possible that only the majority of highly expressed genes are amplified by single-cell transcriptomes, while some lowly expressed genes are not represented in the data due to the limitation of single-cell sequencing techniques.

4.2 | The codon bias and stop codon reassignment of *Halteria grandinella*

Synonymous codons are used with different frequencies in many organisms, known as codon bias, which can reflect a balance between selection, mutation and genetic drift (Bulmer, 1991; Shah & Gilchrist, 2011). The codon bias in *H. grandinella* is more similar with *S. sulcatum* compared with *O. trifallax*, *S. lemnae*, *E. octocarinatus*, *P. tetraurelia* and *T. thermophila* (Figure 3a). Previous studies reported that amino acid codon usage was driven by GC content (Knight, Freeland, & Landweber, 2001). The GC content of the coding region in *H. grandinella*, *S. sulcatum*, *O. trifallax*, *S. lemnae*, *E. octocarinatus*, *P. tetraurelia* and *T. thermophila* is 48%, 52%, 31%, 33%, 32%, 28% and 22%, respectively. The similarity of GC content between *H. grandinella* and *S. sulcatum* might explain why their codon bias resembles each other more faithfully.

The standard stop codons in *H. grandinella* are reassigned (UGA as stop codon while UAA and UAG encoding Glutamine), which is consistent with other hypotrichs and oligotrichs (Swart, Serra, Petroni, & Nowacki, 2016). In most eukaryotes, UAA, UAG and UGA are standard stop codons and do not encode amino acid (Knight et al., 2001; Ling, Patrick, & Dieter, 2015). However, in ciliates, the reassignments of standard stop codons UAA, UAG and UGA to amino acids are detected frequently, which show high divergence and evolve independently in various ciliate lineages (Eliseev, Kryuchkova, Alkalaeva, & Frolova, 2010; Lozupone, Knight, & Landweber, 2001; Swart et al., 2016). For example, in *O. trifallax*, *Strombidium inclinatum* (Cl: Spirotrichea) and *T. thermophila* (Cl: Oligohymenophorea), UAA and UAG encode Glutamine whereas UGA is stop codon (Horowitz & Gorovsky, 1985; Swart et al., 2016); in *Euplotes* (Cl: Spirotrichea), UAA and UAG are stop codons whereas UGA encode Cysteine (Meyer & Heckmann, 1991); in *Blepharisma japonicum* (Cl: Heterotrichea), UAA and UAG are stop codon and UGA encodes Tryptophan (Lozupone et al., 2001; Swart et al., 2016); in *Condylostoma magnum*, UAA, UAG and UGA can either encode amino acid (UAA and UAG encode Glutamine, UGA encode Tryptophan) or be used as stop codons in a context-dependent manner (Swart et al., 2016). The stop codons for UAA, UAG and UGA are decoded by the N-terminal domain of eukaryote class-1 polypeptide release factor (eRF1) (Eliseev et al., 2010; H. Song, Mugnier, et al., 2000; Swart et al., 2016). In some ciliates, the features of the eRF1 protein molecule are changed and can recognize only one or two stop codons (Eliseev et al., 2010; Lekomtsev et al., 2007; Seit-Nebi et al., 2002).

4.3 | The systematic position of *Halteria grandinella*

In the concatenated phylogenomic trees, *H. grandinella* falls within the subclass Hypotrichia by clustering with oxytrichids *O. trifallax* and *S. lemnae* (Figure 4a), which is consistent with the topologies inferred from SSU rDNA, internal transcribed spacers (ITS), 5.8S rDNA, LSU rDNA, α-tubulin and actin I genes (Gao et al., 2016; Hu et al., 2011; Lynn & Sogin, 1988; Paiva et al., 2009). The multigene tree in Lynn and Kolisko (2017) based on the single-cell transcriptome of *H. grandinella* revealed similar topology to our concatenated trees. Moreover, *H. grandinella* has more orthologous genes with the oxytrichid *O. trifallax* than the oligotrich *S. sulcatum* (Figure 2c). All these analyses support that *H. grandinella* has a closer relationship with hypotrichs than oligotrichs.

As discussed in Lynn and Kolisko (2017), it is reasonable to hypothesize that the similar morphology (e.g. globular body shape, apical adoral membranelles and reduced somatic ciliature) of halteriids and oligotrichs might be evolution-adaptation to the planktonic lifestyle, and their similar enantiotropic cell division is evolved to provide enough locomotive force to separate the dividing cells using the developing oral ciliature instead of reduced somatic ciliature (Foissner et al., 2007; Lynn & Kolisko, 2017). Besides, we concur that the long cilia of halteriids, also called jumping bristles, might be specialized cirri or homologous to the cirri in hypotrichs. Notably, all the species in the hypotrich genus *Hypotrichidium* Ilowaisky, 1921 have several spiral cirral rows, which can be considered as an intermediate stage between halteriids and traditional hypotrichs (Chen, Liu, Liu, Al-Rasheid, & Shao, 2013). Moreover, some characteristics (e.g. both paroral and endoral membranes present, stomatogenesis epiapokinetal, somatic ciliature originates de novo) can also indicate a closer relationship between halteriids and hypotrichs (Chen, Wu, El-Serehy, Hu, & Clamp, 2018; Lu et al., 2018; Luo et al., 2017; Petz & Foissner, 1992; Song, 1992).

Hence, our results support the opinion that *H. grandinella* might be an "oligotrich-like" hypotrich. Based on the morphological and phylogenetic data, the evolutionary relationship of the core spirotrichs is hypothesized as follows: the

ancestor of euplotids was firstly isolated from the ancestor of other spirotrichs, which then divided into two groups: one was the ancestor of oligotrichs and choreotrichs, and the other was the ancestor of hypotrichs; within hypotrichs, the ancestor of urostylids and/or stichotrichids separated firstly, while halteriids and oxytrichids shared the most common ancestor (Figure 4c). Nonetheless, the hypothesis here needs to be tested in the future, especially that many taxa are still missing. We believe that with the genomic/transcriptomic data of more and more taxa being available, incorporated analyses of phylogenomic and morphological/morphogenetic results will depict a clearer evolutionary story.

ACKNOWLEDGMENTS

We would like to thank Dr. Weibo Zheng for his help with data analyses. Many thanks are due to Prof. Weibo Song, Dr. Xiaotian Luo, Mr. Bo Pan and Mr. Song Li (Ocean University of China) for their help in manuscript preparation, sample collection and species identification.

ORCID

Feng Gao https://orcid.org/0000-0001-9395-0125

REFERENCES

Agatha, S. (2004). A cladistic approach for the classification of oligotrichid ciliates (Ciliophora: Spirotricha). *Acta Protozoologica*, *43*(3), 201–217.

Agatha, S., & Foissner, W. (2009). Conjugation in the spirotrich ciliate *Halteria grandinella* (Müller, 1773) Dujardin, 1841 (Protozoa, Ciliophora) and its phylogenetic implications. *European Journal of Protistology*, *45*(1), 51–63. https://doi.org/10.1016/j.ejop.2008.07.004

Andrews, S. (2010). FastQC: a quality control tool for high throughput sequence data.

Bulmer, M. (1991). The selection-mutation-drift theory of synonymous codon usage. *Genetics*, *129*(3), 897–907.

Castresana, J. (2000). Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Molecular Biology and Evolution*, *17*(4), 540–552. https://doi.org/10.1093/oxfordjournals.molbev.a026334

Chen, L., Liu, W., Liu, A., Al-Rasheid, K. A., & Shao, C. (2013). Morphology and molecular phylogeny of a new marine hypotrichous ciliate, *Hypotrichidium paraconicum* n. sp.(Ciliophora, Hypotrichia).. *Journal of Eukaryotic Microbiology*, *60*(6), 588–600. https://doi.org/10.1111/jeu.12064

Chen, L., Wu, W., El-Serehy, H. A., Hu, X., & Clamp, J. C. (2018). Morphology, morphogenesis, and phylogeny of an *Anteholosticha intermedia* (Ciliophora, Urostylida) population from the United States. *European Journal of Protistology*, *65*, 1–15. https://doi.org/10.1016/j.ejop.2018.04.003

Chen, L., Zhao, X., Shao, C., Miao, M., & Clamp, J. C. (2017). Morphology and phylogeny of two new ciliates, *Sterkiella sinica* sp. nov. and *Rubrioxytricha tsinlingensis* sp. nov. (Protozoa, Ciliophora, Hypotrichia) from north-west China. *Systematics and Biodiversity*, *15*(2), 131–142. https://doi.org/10.1080/14772000.2016.1219426

Chen, X., Gao, S., Liu, Y., Wang, Y., Wang, Y., & Song, W. (2016). Enzymatic and chemical mapping of nucleosome distribution in purified micro- and macronuclei of the ciliated model organism, *Tetrahymena thermophila*. *Science China Life Sciences*, *59*(9), 909–919. https://doi.org/10.1007/s11427-016-5102-x

Chen, X., Wang, Y. R., Sheng, Y., Warren, A., & Gao, S. (2018). GPSit: An automated method for evolutionary analysis of nonculturable ciliated microeukaryotes. *Molecular Ecology Resources*, *18*, 700–713. https://doi.org/10.1111/1755-0998.12750

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), 373–385. https://doi.org/10.1007/s13238-015-0147-3

De Bourcy, C. F., De Vlaminck, I., Kanbar, J. N., Wang, J., Gawad, C., & Quake, S. R. (2014). A quantitative comparison of single-cell whole genome amplification methods. *PLoS ONE*, *9*(8), e105585. https://doi.org/10.1371/journal.pone.0105585

Eliseev, B., Kryuchkova, P., Alkalaeva, E., & Frolova, L. (2010). A single amino acid change of translation termination factor eRF1 switches between bipotent and omnipotent stop-codon specificity. *Nucleic Acids Research*, *39*(2), 599–608. https://doi.org/10.1093/nar/gkq759

Emms, D. M., & Kelly, S. (2015). OrthoFinder: Solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. *Genome Biology*, *16*(1), 157. https://doi.org/10.1186/s13059-015-0721-2

Enright, A. J., Van Dongen, S., & Ouzounis, C. A. (2002). An efficient algorithm for large-scale detection of protein families. *Nucleic Acids Research*, *30*(7), 1575–1584. https://doi.org/10.1093/nar/30.7.1575

Farlow, A., Long, H., Arnoux, S., Sung, W., Doak, T. G., Nordborg, M., & Lynch, M. (2015). The spontaneous mutation rate in the fission yeast *Schizosaccharomyces pombe*. *Genetics*, *201*(2), 737–744. https://doi.org/10.1534/genetics.115.177329

Fauré-Fremiet, E. (1953). La bipartition énantiotrope chez les ciliés oligotriches. *Achieves Anatomy Microscopy Morphine Experimental*, *42*, 209–225.

Feng, J.-M., Jiang, C.-Q., Warren, A., Tian, M., Cheng, J., Liu, G.-L., … Miao, W. (2015). Phylogenomic analyses reveal subclass Scuticociliatia as the sister group of subclass Hymenostomatia within class Oligohymenophorea. *Molecular Phylogenetics and Evolution*, *90*, 104–111. https://doi.org/10.1016/j.ympev.2015.05.007

Foissner, W., Moon-van der Staay, S. Y., van der Staay, G. W. M., Hackstein, J. H. P., Krautgartner, W.-D., & Berger, H. (2004). Reconciling classical and molecular phylogenies in the stichotrichines (Ciliophora, Spirotrichea), including new sequences from some rare species. *European Journal of Protistology*, *4*, 265–281. https://doi.org/10.1016/j.ejop.2004.05.004

Foissner, W., Muller, H., & Agatha, S. (2007). A comparative fine structural and phylogenetic analysis of resting cysts in oligotrich and hypotrich Spirotrichea (Ciliophora). *European Journal of Protistology*, *43*(4), 295–314. https://doi.org/10.1016/j.ejop.2007.06.001

Fu, L., Niu, B., Zhu, Z., Wu, S., & Li, W. (2012). CD-HIT: Accelerated for clustering the next-generation sequencing data. *Bioinformatics*, *28*(23), 3150–3152. https://doi.org/10.1093/bioinformatics/bts565

Gao, F., Warren, A., Zhang, Q., Gong, J., Miao, M., Sun, P., … Song, W. (2016). The all-data-based evolutionary hypothesis of ciliated protists with a revised classification of the phylum Ciliophora (Eukaryota, Alveolata). *Scientific Reports*, *6*, 24874. https://doi.org/10.1038/srep24874

Gao, S., Xiong, J., Zhang, C., Berquist, B. R., Yang, R., Zhao, M., … Liu, Y. (2013). Impaired replication elongation in *Tetrahymena* mutants deficient in histone H3 Lys 27 monomethylation. *Genes and Development*, *27*(15), 1662–1679. https://doi.org/10.1101/gad.218966.113

Gawad, C., Koh, W., & Quake, S. R. (2016). Single-cell genome sequencing: Current state of the science. *Nature Reviews Genetics*, *17*(3), 175–188. https://doi.org/10.1038/nrg.2015.16

Gentekaki, E., Kolisko, M., Boscaro, V., Bright, K. J., Dini, F., Di Giuseppe, G., … Lynn, D. H. (2014). Large-scale phylogenomic 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. https://doi.org/10.1016/j.ympev.2014.04.020

Gentekaki, E., Kolisko, M., Gong, Y., & Lynn, D. (2017). Phylogenomics solves a long-standing evolutionary puzzle in the ciliate world: The subclass Peritrichia is monophyletic. *Molecular Phylogenetics and Evolution*, *106*, 1–5. https://doi.org/10.1016/j.ympev.2016.09.016

Gordon, A., & Hannon, G. (2010). Fastx-toolkit. FASTQ/A short-reads Preprocessing Tools, (unpublished).

Grabherr, M. G., Haas, B. J., Yassour, M., Levin, J. Z., Thompson, D. A., Amit, I., … Regev, A. (2011). Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nature Biotechnology*, *29*(7), 644–652. https://doi.org/10.1038/nbt.1883

Hall, T. A. (1999). BioEdit: A user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symposium Series*, *41*, 95–98.

Horowitz, S., & Gorovsky, M. A. (1985). An unusual genetic code in nuclear genes of *Tetrahymena*. *Proceedings of the National Academy of Sciences of the United States of America*, *82*(8), 2452–2455. https://doi.org/10.1073/pnas.82.8.2452

Hou, Y., Wu, K., Shi, X., Li, F., Song, L., Wu, H., … Wang, J. (2015). Comparison of variations detection between whole-genome amplification methods used in single-cell resequencing. *GigaScience*, *4*(1), 37. https://doi.org/10.1186/s13742-015-0068-3

Hu, X. Y., Hu, X. Z., Al-Rasheid, K. A., Al-Farraj, S. A., & Song, W. B. (2011). Insights into the phylogeny of sporadotrichid ciliates (Protozoa, Ciliophora: Hypotricha) based on genealogical analyses of multiple molecular markers. *Chinese Journal of Oceanology and Limnology*, *29*(1), 96–102. https://doi.org/10.1007/s00343-011-9008-y

Huang, J. B., Zhang, T., Zhang, Q., Li, Y., Warren, A., Pan, H., & Yan, Y. (2018). Further insights into the highly derived haptorids (Ciliophora, Litostomatea): Phylogeny based on multigene data. *Zoologica Scripta*, *47*(2), 231–242. https://doi.org/10.1111/zsc.12269

Huang, L., Ma, F., Chapman, A., Lu, S., & Xie, X. S. (2015). Single-cell whole-genome amplification and sequencing: Methodology and applications. *Annual Review of Genomics and Human Genetics*, *16*, 79–102. https://doi.org/10.1146/annurev-genom-090413-025352

Jiang, J., Ma, H., & Shao, C. (2013). Morphology and morphogenesis of *Sterkiella histriomuscorum* (Ciliophora, Hypotricha). *Acta Hydrobiologica Sinica*, *37*(2), 227–234. https://doi.org/10.7541/2013.9

Jones, P., Binns, D., Chang, H.-Y., Fraser, M., Li, W., McAnulla, C., … Hunter, S. (2014). InterProScan 5: Genome-scale protein function classification. *Bioinformatics*, *30*(9), 1236–1240. https://doi.org/10.1093/bioinformatics/btu031

Kapusta, A., Suh, A., & Feschotte, C. (2017). Dynamics of genome size evolution in birds and mammals. *Proceedings of the National Academy of Sciences of the United States of America*, *114*(8), E1460. https://doi.org/10.1073/pnas.1616702114

Kathol, M., Norf, H., Arndt, H., & Weitere, M. (2009). Effects of temperature increase on the grazing of planktonic bacteria by biofilm-dwelling consumers. *Aquatic Microbial Ecology*, *55*, 65–79. https://doi.org/10.3354/ame01286

Knight, R. D., Freeland, S. J., & Landweber, L. F. (2001). A simple model based on mutation and selection explains trends in codon and amino-acid usage and GC composition within and across genomes. *Genome Biology*, *2(4), research0010*, 0011. https://doi.org/10.1186/gb-2001-2-4-research0010

Kolisko, M., Boscaro, V., Burki, F., Lynn, D. H., & Keeling, P. J. (2014). Single-cell transcriptomics for microbial eukaryotes. *Current Biology*, *24*(22), R1081–R1082. https://doi.org/10.1016/j.cub.2014.10.026

Lahr, D., Laughinghouse, H., Oliverio, A., Gao, F., & Katz, L. (2014). How discordant morphological and molecular evolution among microorganisms can revise our notions of biodiversity on earth. *BioEssays*, *36*(10), 950–959. https://doi.org/10.1002/bies.201400056

Langmead, B., Trapnell, C., Pop, M., & Salzberg, S. L. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biology*, *10*(3), R25. https://doi.org/10.1186/gb-2009-10-3-r25

Lao, K., Xu, N. L., & Straus, N. A. (2008). Whole genome amplification using single-primer PCR. *Biotechnology Journal*, *3*(3), 378–382. https://doi.org/10.1002/biot.200700253

Lartillot, N., Lepage, T., & Blanquart, S. (2009). PhyloBayes 3: A Bayesian software package for phylogenetic reconstruction and molecular dating. *Bioinformatics*, *25*(17), 2286–2288. https://doi.org/10.1093/bioinformatics/btp368

Lekomtsev, S., Kolosov, P., Bidou, L., Frolova, L., Rousset, J.-P., & Kisselev, L. (2007). Different modes of stop codon restriction by the *Stylonychia* and *Paramecium* eRF1 translation termination factors. *Proceedings of the National Academy of Sciences*, *104*(26), 10824–10829. https://doi.org/10.1073/pnas.0703887104

Ling, J., Patrick, O., & Dieter, S. (2015). Genetic code flexibility in microorganisms: Novel mechanisms and impact on physiology. *Nature Reviews Microbiology*, *13*, 707–721. https://doi.org/10.1038/nrmicro3568

Liu, W., Jiang, J., Xu, Y., Pan, X., Qu, Z., Luo, X., … Pan, H. (2017). Diversity of free-living marine ciliates (Alveolata, Ciliophora): Faunal studies in coastal waters of China during the years 2011–2016. *European Journal of Protistology*, *61*, 424–438. https://doi.org/10.1016/j.ejop.2017.04.007

Liu, W., Xu, D., Ma, H., Al-Farraj, S. A., Warren, A., & Yi, Z. (2016). Taxonomy and molecular systematics of three oligotrich (s.l.) ciliates including descriptions of two new species, *Strombidium guangdongense* sp. nov. and *Strombidinopsis sinicum* sp. nov. (Protozoa, Ciliophora). *Systematics and Biodiversity*, *14*(5), 452–465. https://doi.org/10.1080/14772000.2016.1162872

Lozupone, C. A., Knight, R. D., & Landweber, L. F. (2001). The molecular basis of nuclear genetic code change in ciliates. *Current Biology*, *11*(2), 65–74. https://doi.org/10.1016/S0960-9822(01)00028-8

Lu, X., Huang, J., Chen, S., & Berger, H. (2018). Morphology, cell-division, and phylogeny of *Schmidingerothrix elongata* spec. nov. (Ciliophora, Hypotricha), and brief guide to hypotrichs with *Gonostomum*-like oral apparatus. *European Journal of Protistology*, *62*, 24–42. https://doi.org/10.1016/j.ejop.2017.11.001

Luo, X., Li, L., Wang, C., Bourland, W., Lin, X., & Hu, X. (2017). Morphologic and phylogenetic studies of two hypotrichous ciliates, with notes on morphogenesis in *Gastrostyla steinii* Engelmann, 1862 (Ciliophora, Hypotrichia). *European Journal of Protistology*, *60*, 119–133. https://doi.org/10.1016/j.ejop.2017.05.001

Luo, X., Yan, Y., Shao, C., Al-Farraj, S. A., Bourland, W. A., & Song, W. (2018). Morphological, ontogenetic and molecular data support strongylidiids as being closely related to Dorsomarginalia (Protozoa, Ciliophora) and reactivation of the family Strongylidiidae Fauré-Fremiet, 1961. *Zoological Journal of the Linnean Society*, *184*(2), 237–254. https://doi.org/10.1093/zoolinnean/zly001.

Lynn, D. H., & Kolisko, M. (2017). Molecules illuminate morphology: Phylogenomics confirms convergent evolution among 'oligotrichous' ciliates. *International Journal of Systematic and Evolutionary Microbiology*, *67*(9), 3676–3682. https://doi.org/10.1099/ijsem.0.002060

Lynn, D. H., & Sogin, M. L. (1988). Assessment of phylogenetic relationships among ciliated protists using partial ribosomal RNA sequences derived from reverse transcripts. *Bio Systems*, *21*(3–4), 249. https://doi.org/10.1016/0303-2647(88)90020-2

Meyer, F., Schmidt, H. J., Plumper, E., Hasilik, A., Mersmann, G., Meyer, H. E., … Heckmann, K. (1991). UGA is translated as cysteine in pheromone 3 of *Euplotes octocarinatus*. *Proceedings of the National Academy of Sciences of the United States of America*, *88*(9), 3758–3761. https://doi.org/10.1073/pnas.88.9.3758

Nylander, J. (2004). *MrModeltest v2. Program distributed by the author*. Uppsala, Sweden: University Press.

Paiva, T. S., Borges, B. N., Harada, M. L., & Silva-Neto, I. D. (2009). Comparative phylogenetic study of Stichotrichia (Alveolata: Ciliophora: Spirotrichea) based on 18S-rDNA sequences. *Genetics and Molecular Research*, *8*(1), 223–246. https://doi.org/10.4238/vol8-1gmr529

Petz, W., & Foissner, W. (1992). Morphology and morphogenesis of *Strobilidium caudatum* (Fromentel), *Meseres corlissi* n. sp., *Halteria grandinella* (Müller), and *Strombidium rehwaldi* n. sp., and a proposed phylogenetic system for oligotrich ciliates (Protozoa, Ciliophora). *Journal of Eukaryotic Microbiology*, *39*(1), 159–176. https://doi.org/10.1111/j.1550-7408.1992.tb01296.x

Seit-Nebi, A., Frolova, L., & Kisselev, L. (2002). Conversion of omnipotent translation termination factor eRF1 into ciliate-like UGA-only unipotent eRF1. *EMBO Reports*, *3*(9), 881–886. https://doi.org/10.1093/embo-reports/kvf178

Sela, I., Ashkenazy, H., Katoh, K., & Pupko, T. (2015). GUIDANCE2: Accurate detection of unreliable alignment regions accounting for the uncertainty of multiple parameters. *Nucleic Acids Research*, *43*(W1), 7–14. https://doi.org/10.1093/nar/gkv318

Shah, P., & Gilchrist, M. A. (2011). Explaining complex codon usage patterns with selection for translational efficiency, mutation bias, and genetic drift. *Proceedings of the National Academy of Sciences*, *108*(25), 10231–10236. https://doi.org/10.1073/pnas.1016719108

Sheng, Y., He, M., Zhao, F., Shao, C., & Miao, M. (2018). Phylogenetic relationship analyses of complicated class Spirotrichea based on transcriptomes from three diverse microbial eukaryotes: *Uroleptopsis citrina*, *Euplotes vannus* and *Protocruzia tuzeti*. *Molecular Phylogenetics and Evolution*, *129*, 338–345. https://doi.org/10.1016/j.ympev.2018.06.025

Sievers, F., Wilm, A., Dineen, D., Gibson, T. J., Karplus, K., Li, W., … Higgins, D. G. (2011). Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Molecular Systems Biology*, *7*(1), 539. https://doi.org/10.1038/msb.2011.75

Šimek, K., Jürgens, K., Nedoma, J., Comerma, M., & Armengol, J. (2000). Ecological role and bacterial grazing of *Halteria* spp.: Small freshwater Oligotrichs as dominant pelagic ciliate bacterivores. *Aquatic Microbial Ecology*, *22*(1), 43–56. https://doi.org/10.3354/ame022043

Song, H., Mugnier, P., Das, A. K., Webb, H. M., Evans, D. R., Tuite, M. F., … Barford, D. (2000). The crystal structure of human eukaryotic release factor eRF1-mechanism of stop codon recognition and peptidyl-tRNA hydrolysis. *Cell*, *100*(3), 311–321. https://doi.org/10.1016/S0092-8674(00)80667-4

Song, W. B. (1992). Studies on the cortical morphogenesis during cell division in *Halteria grandinella* (Müller, 1773) (Ciliophora, Oligotrichida). *Chinese Journal of Oceanology and Limnology*, *11*(2), 122–129. https://doi.org/10.1007/BF02850818

Song, W., Wang, L., Li, L., Al-Farraj, S. A., Aleidan, A., Smith, S., & Hu, X. (2018). Morphological characterizations of four species of *Parallelostrombidium* (Ciliophora, Oligotrichia), with a note on the phylogeny of the genus. *Journal of Eukaryotic Microbiology*, 1–15. https://doi.org/10.1111/jeu.12513

Song, W. B., Wang, M., & Warren, A. (2000). Redescriptions of three marine ciliates, *Strombidium elegans*, Florentin, 1901, *Strombidium sulcatum*, Claparède & Lachmann, 1859 and *Heterostrombidium paracalkinsi*, Lei, Xu & Song, 1999 (Ciliophora, Oligotrichida). *European Journal of Protistology*, *36*(3), 327–342. https://doi.org/10.1016/S0932-4739(00)80010-3

Stamatakis, A. (2014). RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics*, *30*(9), 1312–1313. https://doi.org/10.1093/bioinformatics/btu033

Stamatakis, A., Hoover, P., & Rougemont, J. (2008). A rapid bootstrap algorithm for the RAxML web servers. *Systematic Biology*, *57*(5), 758–771. https://doi.org/10.1080/10635150802429642

Stanke, M., Diekhans, M., Baertsch, R., & Haussler, D. (2008). Using native and syntenically mapped cDNA alignments to improve de novo gene finding. *Bioinformatics*, *24*(5), 637–644. https://doi.org/10.1093/bioinformatics/btn013

Swart, E. C., Serra, V., Petroni, G., & Nowacki, M. (2016). Genetic codes with no dedicated stop codon: Context-dependent translation termination. *Cell*, *166*(3), 691–702. https://doi.org/10.1016/j.cell.2016.06.020

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. https://doi.org/10.1093/molbev/mst197

Tang, F., Barbacioru, C., Wang, Y., Nordman, E., Lee, C., Xu, N., … Surani, M. A. (2009). mRNA-Seq whole-transcriptome analysis of a single cell. *Nature Methods*, *6*(5), 377. https://doi.org/10.1038/nmeth.1315

Vij, S., Kuhl, H., Kuznetsova, I. S., Komissarov, A., Yurchenko, A. A., Van Heusden, P., … Orbán, L. (2016). Chromosomal-level assembly of the asian seabass genome using long sequence reads and multi-layered scaffolding. *PLoS Genetics*, *12*(4), e1005954. https://doi.org/10.1371/journal.pgen.1005954

Wang, C. D., Zhang, T., Wang, Y. R., Katz, L. A., Gao, F., & Song, W. (2017). Disentangling sources of variation in SSU rDNA sequences from single cell analyses of ciliates: Impact of copy number variation and experimental error. *Proceedings of the Royal Society B: Biological Sciences*, *284*, 20170425. https://doi.org/10.1098/rspb.2017.0425

Wang, Y. Y., Chen, X., Sheng, Y., Liu, Y., & Gao, S. (2017). N6-adenine DNA methylation is associated with the linker DNA of H2A. Z-containing well-positioned nucleosomes in Pol II-transcribed genes in *Tetrahymena*. *Nucleic Acids Research*, *45*(20), 11594–11606. https://doi.org/10.1093/nar/gkx883

Wang, Y. Y., Sheng, Y., Liu, Y., Pan, B., Huang, J., Warren, A., & Gao, S. (2017). N6-methyladenine DNA modification in the unicellular eukaryotic organism *Tetrahymena thermophila*. *European Journal of Protistology*, *58*, 94–102. https://doi.org/10.1016/j.ejop.2016.12.003

Wu, A. R., Neff, N. F., Kalisky, T., Dalerba, P., Treutlein, B., Rothenberg, M. E., … Quake, S. R. (2014). Quantitative assessment of single-cell RNA-sequencing methods. *Nature Methods*, *11*(1), 41. https://doi.org/10.1038/nmeth.2694

Xiong, J., Gao, S., Dui, W., Yang, W., Chen, X., Taverna, S. D., … Liu, Y. (2016). Dissecting relative contributions of cis-and trans-determinants to nucleosome distribution by comparing *Tetrahymena* macronuclear and micronuclear chromatin. *Nucleic Acids Research*, *44*(21), 10091–10105. https://doi.org/10.1093/nar/gkw684

Xu, H., Zhang, W., & Jiang, Y. (2014). Do early colonization patterns of periphytic ciliate fauna reveal environmental quality status in coastal waters? *Environmental Science and Pollution Research International*, *21*(11), 7097–7112. https://doi.org/10.1007/s11356-014-2615-3

Yan, Y., Fan, Y., Luo, X., El-Serehy, H. A., Bourland, W., & Chen, X. (2018). New contribution to the species-rich genus *Euplotes* : Morphology, ontogeny and systematic position of two species (Ciliophora; Euplotia). *European Journal of Protistology*, *64*, 20–39. https://doi.org/10.1016/j.ejop.2018.03.003

Yan, Y., Rogers, A. J., Gao, F., & Katz, L. A. (2017). Unusual features of non-dividing somatic macronuclei in the ciliate class Karyorelictea. *European Journal of Protistology*, *61*, 399–408. https://doi.org/10.1016/j.ejop.2017.05.002

Zhang, K., Martiny, A. C., Reppas, N. B., Barry, K. W., Malek, J., Chisholm, S. W., & Church, G. M. (2006). Sequencing genomes from single cells by polymerase cloning. *Nature Biotechnology*, *24*(6), 680–686. https://doi.org/10.1038/nbt1214

Zhang, L., Cui, X., Schmitt, K., Hubert, R., Navidi, W., & Arnheim, N. (1992). Whole genome amplification from a single cell: Implications for genetic analysis. *Proceedings of the National Academy of Sciences*, *89*(13), 5847–5851. https://doi.org/10.1073/pnas.89.13.5847

Zhang, T., Wang, C., Katz, L. A., & Gao, F. (2018). A paradox: Rapid evolution rates of germline-limited sequences are associated with conserved patterns of rearrangements in cryptic species of *Chilodonella uncinata* (Protista, Ciliophora). *Science China Life Sciences*, *61*(9), 1071–1078. https://doi.org/10.1007/s11427-018-9333-1

Zhao, X., Wang, Y., Wang, Y., Liu, Y., & Gao, S. (2017). Histone methyltransferase TXR1 is required for both H3 and H3. 3 lysine 27 methylation in the well-known ciliated protist *Tetrahymena thermophila*. *Science China Life Sciences*, *60*(3), 264–270. https://doi.org/10.1007/s11427-016-0183-1

Zhao, Y., Yi, Z., Warren, A., & Song, W. B. (2018). Species delimitation for the molecular taxonomy and ecology of the widely distributed microbial eukaryote genus *Euplotes* (Alveolata, Ciliophora). *Proceedings of the Royal Society B: Biological Sciences*, *285*(1871), 20172159. https://doi.org/10.1098/rspb.2017.2159

Zong, C., Lu, S., Chapman, A. R., & Xie, X. S. (2012). Genome-wide detection of single-nucleotide and copy-number variations of a single human cell. *Science*, *338*(6114), 1622–1626. https://doi.org/10.1126/science.1229164

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of the article.

How to cite this article: Wang C, Yan Y, Chen X, Al-Farraj SA, El-Serehy HA, Gao F. Further analyses on the evolutionary "key-protist" *Halteria* (Protista, Ciliophora) based on transcriptomic data. *Zool Scr*. 2019;48:813–825. https://doi.org/10.1111/zsc.12380