Comparative Genome Analysis Reveals Cis-Regulatory Elements on Gene-Sized Chromosomes of Ciliated Protists

Weibo Zheng1,2,3†, Huan Dou1†, Chao Li3†, Saleh A. Al-Farraj4, Adam Byerly5, Naomi A. Stover6, Weibo Song1,3, Xiao Chen1* and Lifang Li1*†

1 Laboratory of Marine Protozoan Biodiversity and Evolution, Marine College, Shandong University, Weihai, China, 2 School of Life Sciences, Ludong University, Yantai, China, 3 Institute of Evolution and Marine Biodiversity, Ocean University of China, Qingdao, China, 4 Zoology Department, College of Science, King Saud University, Riyadh, Saudi Arabia, 5 Department of Computer Science, Bradley University, Peoria, IL, United States, 6 Department of Biology, Bradley University, Peoria, IL, United States

Gene-sized chromosomes are a distinct feature of the macronuclear genome in ciliated protists known as spirotrichs. These nanochromosomes are often only several kilobase pairs long and contain a coding region for a single gene. However, the ways in which transcription is regulated on nanochromosomes is still largely unknown. Here, we generated macronuclear genome assemblies for two species of Pseudokeronopsis ciliates to better understand transcription regulation on gene-sized chromosomes. We searched within the short subtelomeric regions for potential cis-regulatory elements and identified distinct AT-rich sequences conserved in both species, at both the 5' and 3' end of each gene. We further acquired transcriptomic data for these species, which showed the 5' cis-regulatory element is associated with active gene expression. Gene family evolution analysis suggests nanochromosomes in spirotrichs may originated approximately 900 million years ago. Together our comparative genomic analyses reveal novel insights into the biological roles of cis-regulatory elements on gene-sized chromosomes.

Keywords: ciliates, Pseudokeronopsis carnea, Pseudokeronopsis flava, phylogenomics, nanochromosome

INTRODUCTION

Gene-sized chromosomes (nanochromosomes) are an intriguing genetic architecture found in a subset of ciliates, one of the most diverse clades of unicellular eukaryotes. Ciliates contain two distinct types of nuclei, called the micro- and macronucleus, which are both present in the cell throughout its vegetative life cycle (Yan et al., 2017; Jiang et al., 2019; Sheng et al., 2021). During sexual reproduction (conjugation), zygotic micronuclear (MIC) chromosomes are fragmented into somatic macronuclear (MAC) chromosomes through a series of genome wide rearrangements (Chalker and Yao, 2011; Chen et al., 2014; Zhao et al., 2019, 2020; Sheng et al., 2020). For spirotrichous ciliates including Oxytricha, Stylonychia, Uroleptopsis, Euplotes and Strombidium, this fragmentation is extensive, resulting in a MAC genome composed of more than 10,000 telomere-capped chromosomes (Steinbrück et al., 1981; Prescott, 2000; Swart et al., 2013; Zheng et al., 2018; Chen et al., 2019; Li et al., 2021). These linear nanochromosomes are usually smaller than
mitochondrial chromosomes and can carry one or few genes and can be few hundred base pairs up to several kilobases in size (Swart et al., 2013; Zhang et al., 2021). These general features make spirotrich ciliates excellent models to study chromosomal architecture and functions (Zhao et al., 2021; Zheng et al., 2021).

The cis-regulatory elements (CREs, e.g., promoter, enhancer, and insulator) needed for transcription regulation in eukaryotes are often located in intergenic regions (Levine and Tjian, 2003). The existence and nature of cis-regulatory elements, and the methods of gene regulation in ciliates in general, are poorly understood. Furthermore, since nanochromosomes contain only very short regions outside of the coding sequence, the space to harbor CREs is limited. Surprisingly, studies of other spirotrich genomes have not immediately revealed the presence of any conserved, recognizable sequences or patterns that could be tied to the regulation of gene expression (Swart et al., 2013; Chen et al., 2019; Li et al., 2021; Zheng et al., 2021).

To help elucidate the mechanism of transcription regulation on gene-sized chromosomes, we carried out deep genomic sequencing and assembly for the MAC genomes of two new spirotrich ciliates, Pseudokeronopsis carnea and Pseudokeronopsis flava. Pseudokeronopsis species have long been recognized and studied for their distinct cell shape and fascinating pigment colors (Song et al., 2004; Baek et al., 2011) (Figure 1 and Supplementary Table 1). Like other spirotrichs, Pseudokeronopsis cells are large and can provide abundant DNA, making them ideal prospects for in-depth genetic studies (Dong et al., 2020; Luo et al., 2021). To date these types of studies have been limited by a lack of genomic data, and many features of Pseudokeronopsis genomes, including the presence of nanochromosomes, have been unknown until now. Combining both genomic and transcriptomic data, we searched the subtelomeric regions of their compact chromosomes for potential conserved CREs. Using a variety of genome evolution analyses, we reveal the origin of gene-sized chromosomes in spirotrichs and the regulatory elements they harbor.

MATERIALS AND METHODS

Cell Culture and Sample Preparation

Pseudokeronopsis carnea and P. flava cells were isolated from a freshwater pond in Baihuayuan Park (36°04′N, 120°22′E), Qingdao, China. Species were initially determined by morphological features and later confirmed by sequencing their SSU-rrRNA genes. A single cell was picked, washed, and cultured in flasks using filtered and autoclaved pond water. Cells were morphological features and later confirmed by sequencing their

The RNA libraries were generated using NEBNext Ultra RNA Library Prep Kit for Illumina (NEB, United States) following the manufacturer’s instructions.

Illumina Sequencing and Genome Assembly

Pair-end 150 bp sequencing reads were performed on the Illumina Hiseq 2500 platform, producing 20 Gb and 10 Gb of clean data for the DNA and RNA libraries, respectively. Genomes were assembled using SPAdes v3.12 (Bankevich et al., 2012) (parameters: -k 21,33,55,77 –careful). Contigs with low coverage (< 2 ×) or small size (< 200 bp) were removed. QUAST v5.0.2 (Gurevich et al., 2013) was used to measure genomic statistics including GC content and N50. RSEM v1.3.3 (Li and Dewey, 2011) was used to calculate the sequencing depth of each contig. For homologous gene annotation, genomic contigs were aligned with protein sequences from the SWISS-PROT database using BLASTX v2.3.0 (Camacho et al., 2009) (parameters: evalue = 1e-5, querygenecode = 6). Gene model annotation information was extracted using a custom Perl script and used to train AUGUSTUS v2.3.0 using custom Perl script and used to train AUGUSTUS v2.5.5 (Stanke et al., 2006) (parameters: –species = pseudokeronopsis –min_intron_len 15,39) for gene model prediction. RNA-seq reads were assembled into transcripts using rnaSPAdes v3.11.1 (Bushmanova et al., 2019) and aligned with the genome assembly by BLAT v3.6 (Kent, 2002) to optimize the gene models. Predicted genes without start and stop codons were filtered out using a custom Perl script. The RNA-seq reads were mapped to genome contigs using TopHat2 v2.0.10 (Kim et al., 2013). The mapped read count of each gene was measured by featureCounts v1.6.1 (Liao et al., 2013). Potential cis-regulatory sequence motifs were searched within subtelomeric regions using MEME v5.3.3 (Bailey et al., 2015). The sequence motifs identified were visualized using WebLogo 3 (Crooks et al., 2004). Frequency of stop codon usage (TAA, TGA, and TAG) was measured from the homolog sequence alignment between the CDS or transcript sequences of each species and the ciliate protein library using BLASTX v2.3.0 (parameter: evalue = 1e-5), as previously described (Pan et al., 2019).

Phylogenomic Analysis and Genome Evolution

A total of 238 orthologous were identified among 31 ciliates (see Supplementary Table 2) using OrthoFinder and were aligned using mafft (Emms and Kelly, 2019). The concatenated ortholog sequence alignment dataset was used for phylogenomic analysis on CIPRES Science Gateway server v3.3 (Miller et al., 2010). RAxML-HPC2 v8.2.9 (Stamatakis, 2014) under LG model of amino acid substitution (Γ distribution + F, four rate categories, 1,000 bootstrap replicates) was used to perform maximum likelihood (ML) analysis. PhyloBayes MPI 1.5a (Lartillot et al., 2009) (CAT-GTR model + Γ distribution, four independent chains, 4,000 generations with 10% burn-in, convergence Maxdifs < 0.3) was used to perform Bayesian inference (BI) analysis. The phylogenetic tree was visualized using MEGA v7.0.20 (Kumar et al., 2016). The time of speciation was estimated using r8s (Sanderson, 2003) and corrected using calibration times obtained from the TimeTree database (Hedges et al., 2006).
Zheng et al. Cis-Regulatory Element on Nanochromosomes

FIGURE 1 | Morphology and genome size of model organisms including ciliates. From left to right, bacteria (Staphylococcus aureus and Escherichia coli), cyanobacteria (Gloeobacter violaceus), yeast (Saccharomyces cerevisiae), amoeboid protists (Entamoeba moshkovskii), ciliated protists (Ichthyophthirius multifiliis, Stylonychia lemmae, Oxytricha trifallax, Pseudokeronopsis flava, Halteria grandinella, Paramecium tetraurelia, Pseudokeronopsis carnea, Stentor coeruleus, Euplotes octocarinatus, Tetrahymena thermophila), nematode (Caenorhabditis elegans), thale cress (Arabidopsis thaliana), fruit fly (Drosophila melanogaster), and zebrafish (Danio rerio).

RESULTS AND DISCUSSION

Compact Genome Architecture and Reassigned Stop Codons
Using high-throughput sequencing data, we assembled the MAC genomes of two Pseudokeronopsis species (Table 1). The genome assemblies of P. carnea and P. flava are 76.8 Mb and 52.5 Mb in size, respectively, in line with other known ciliate genomes (Figure 1). Most of the contigs bear telomeric repeats (C4A4) on at least one end (P. carnea, 81.1%; P. flava, 82.8%). The average size of contigs capped with telomeric repeats on both ends is 1.7 kb (P. carnea) and 1.1 kb (P. flava), and the vast majority (P. carnea, 97%; P. flava, 96%) appear to be gene-sized chromosomes (Figures 2A,B), consistent with the nanochromosome architecture found in other spirotrich genomes. Predicted gene numbers in P. carnea and P. flava are 12,734 and 9,520, respectively. 84% of P. carnea and 78%

| TABLE 1 | MAC genome assembly information for two Pseudokeronopsis species. |
|-------|-----------------|-----------------|
|             | P. carnea       | P. flava       |
| Genome size (Mb) | 76.8            | 52.5            |
| % GC        | 41.7%           | 40.4%           |
| Contig     | 37,909          | 37,545          |
| % contig with telomere* | 81.1%          | 82.8%           |
| % scaffold (two telomeres)** | 38.6%          | 32.7%           |
| N50 (scaffold) | 2,120           | 1,712           |
| Gene       | 12,734          | 9,520           |
| Exon       | 31,869          | 21,757          |

*Percentage of contigs capped with telomere repeats on at least one of two ends.  
**Percentage of scaffolds capped with telomere repeats on both ends.
of *P. flava* genes are annotated using SWISS-PROT or NR databases. Introns in these species are tiny (only 36 nucleotides on average for both *P. carnea* and *P. flava*), a feature also observed in several other ciliate genomes (Figure 2C). The subtelomeric regions between the transcription start site (TSS) or transcription end site (TES) and the adjacent 5’ telomere repeat are also short (Figure 2D), similar to previous observations in the nanochromosomes of *Oxytricha*, *Euplotes* and *Strombidium* (Kim et al., 2013; Swart et al., 2013; Chen et al., 2019). Overall, the combined genomic features of *Pseudokeronopsis* represent an extremely compact eukaryotic genome architecture, with single genes containing minimal introns, nested between short subtelomeric regions (Figures 2D,E).

To perform the phylogenomic analysis, we collected public genomic/transcriptomic datasets available for 31 ciliates (Supplementary Table 2), and identified 238 orthologs among *P. carnea*, *P. flava*, and these species. The system assignment of species we describe here based on maximum likelihood (ML) and Bayesian inference (BI) methods generally agrees with previous studies (Gentekaki et al., 2017; Chen et al., 2018). Phylogenomic analysis of the *P. carnea* and *P. flava* sequenced in the current study shows full support for their cluster with two previously reported *Pseudokeronopsis* species (Figure 2F). The analysis also supports a larger cluster that includes the spirotrichs *Oxytricha trifallax* and *Stylonychia lemnæ*. Standard stop codons in ciliates are frequently reassigned to code for
Conserved AT-Rich Sequences Identified as Potential Cis-Regulatory Elements

Although *Pseudokeronopsis* nanochromosomes have limited space, we found an AT-rich region exists in the 5’ subtelomeric regions of most complete nanochromosomes that carry single genes (92.7 and 98.5% for *P. carnea* and *P. flava*, respectively) (Figure 3A). Further analysis 5’ subtelomeric regions reveals a 15 nt putative CRE in the sequence motif W_nAWTW_n which is positioned between the transcription start site (TSS) and translation start codon in both species (Figure 3B). Similar CREs were identified in the 5’ subtelomeric regions on the nanochromosomes of *Strombidium* (Li et al., 2021) and *Nyctotherus* (McGrath et al., 2007). We identified an additional AT-rich element in the *Pseudokeronopsis* species, this time in the 3’ subtelomeric regions, with the sequence motif “TTNATTTCNTTAA.” This sequence is found between the translation stop codon and the transcription end site (TES) (Supplementary Figure 1A), and is distinct from the 5’ subtelomeric sequence at the other end.

To demonstrate that these potential CREs are conserved evolutionarily, we investigated the base substitution rate of nucleotides in the 5’ subtelomeric regions. The base substitution rate in the putative CRE is 50% lower than that in the entire subtelomeric region (48.2 and 55.2% for *P. carnea* and *P. flava*, respectively), indicating that nucleotides in the CRE are under amino acids (Swart et al., 2016). Similar to these two species, the stop codons “TAA” and “TAG” are consistently reassigned in all *Pseudokeronopsis* species, leaving “TGA” as the only stop codon (Figure 2F and Supplementary Table 2), reflecting the close evolutionary relationship between these groups. Interestingly, compared to previous phylogenomic analysis result in which the litostomatous *Entodinium* is not included (Chen et al., 2018), we find that the clustering of class Litostomatea and class Spirotrichea is poorly supported in the current BI tree, and not supported at all by the ML tree (Litostomatea clusters with Colpodea first after including *Entodinium*). This opens the possibility that the assignment of litostomatans could require further review in the future by expanding sampling.
stronger selection. The substitution pattern also shows a distinct difference between the 5′ CRE and surrounding nucleotides (Figure 3C and Supplementary Figure 1B). Compared with the entire subtelomeric region, the substitution rate of A-to-G and T-to-C is greatly reduced in the CRE, but dramatically increased in C-to-A substitutions, indicating that G/C nucleotides in the CRE are consistently being replaced by A/T nucleotides.

Although not positioned upstream of the TSS, as is seen for TATA-box-like elements in other eukaryotic organisms like yeast (Lin et al., 2010), these sequences may still act as non-canonical regulatory CREs that bind transcriptional trans-activating factors. To determine whether this motif is related to transcription initiation, we compared the expression of genes on nanochromosomes that either possess or lack this CRE in their 5′ subtelomeric regions. We observed that genes with this CRE have significantly higher transcription levels in both species (Figure 3D). The majority of genes without the 5′ CRE are silent and the putative CRE is more associated with active genes, which could be the source of this transcription activity difference (Figure 3E). On the contrary, most of the genes with the adjacent CRE are actively transcribed (94.2 and 90.4% for *P. carnea* and *P. flava*, respectively). These observations suggest that transcription initiation on Pseudokeronopsis nanochromosomes depends on this CRE near the transcription start site, though the nature of this CRE is not clear. The sequence may act as a promoter by binding directly to a transcription factor, or it may contribute a necessary structural feature to the DNA in this region. Future studies should further test whether chromatin accessibility is greater at this location, and if active chromatin marks are enriched (Sheng et al., 2021). A similar association between the 3′ CRE and gene transcription was also identified (Supplementary Figures 1C,D).

Together with the conserved base substitution patterns, our results reveal a strong evolutionary selection pressure upon the AT-rich CRE, and suggest it plays a functionally important role in transcription regulation. Considering the compact nanochromosome architecture, the inclusion of these sequences in the primary transcript UTRs, and reassignment of stop codons in these species, it is also possible that these CREs assist in translation initiation/termination at the 5′ and 3′ ends, respectively.

### The Evolution History of Spirotrich Nanochromosomes

To help understand the origins of spirotrich nanochromosomes, we investigated the expansion and contraction of gene families in 31 ciliates based on 238 orthogroups. Nanochromosome architecture has been reported in species of several disparate ciliate clades (classes Spirotrichae, Armophorea, and
Litostomatea), suggesting multiple origins of extensive fragmentation within ciliates (Riley and Katz, 2001; McGrath et al., 2007; Huang and Katz, 2014; Špaková et al., 2014; Park et al., 2021). The spirotrich clade, which features gene-sized chromosomes, originated approximately 900 million years ago, accompanied by a dramatic expansion/contraction of several gene families (Figure 4). Within spirotrichs, Pseudokeronopsis species originated 220 million years ago, which in turn separated from the clade containing Oxytricha and Stylonychia 430 million years ago. As gene family expansion reflects phenotypic diversity and genetic adaptations during evolution (Harris and Hofmann, 2015; Yan et al., 2019), we identified 1079 and 768 expanded gene families (\( p < 0.05 \)) in P. carnea and P. flava, respectively. These expanded gene families are enriched in a variety of pathways (Supplementary Figure 2). Compared with three representative species that do not carry nanochromosomes (Supplementary Figure 3), the expanded gene families in both Pseudokeronopsis species contribute to transcription factor binding and sequence-specific DNA binding. Although given the incomplete nature of the current ciliate genomic datasets as a limitation, our analyses provide a baseline about the transcription regulation pathways rewiring in species with nanochromosomal organization for the future studies.

CONCLUDING REMARKS

In summary, we report the first macronuclear genome assemblies of two Pseudokeronopsis ciliates, which consist of compact, gene-sized nanochromosomes. Similar to other spirotrichs, Pseudokeronopsis nanochromosomes have tiny introns and small subtelomeric regions. We identified AT-rich sequences conserved within the 5′ and 3′ subtelomeric regions in both species and observed that these potential CREs are associated with active gene expression, suggesting a role in transcription regulation. Both P. carnea and P. flava have expanded their complement of genes related to nucleotide binding and gene expression regulation since the origin of gene-sized chromosomes in spirotrichs approximately 900 million years ago. Together, these findings suggest that ciliates may have developed a unique mechanism to regulate transcription from gene-sized chromosomes during evolution.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found below: https://www.ncbi.nlm.nih.gov/, PRJNA507672, PRJNA534036, https://pse.ciliate.org, Pseudokeronopsis DB.

AUTHOR CONTRIBUTIONS

WZ, CL, XC, and LL conceived the study. WZ provided the biological materials. WZ, WS, and XC designed the experiments. WZ and HD performed the experiments. WZ, CL, HD, and XC performed computational and experimental analysis for all figures and tables. WZ, CL, SAA, WS, and XC interpreted the data. NAS and AB constructed the genome database website. WZ, CL, XC, and LL wrote the manuscript with contribution from all authors. All authors read and approved the final manuscript.

FUNDING

This work was supported by the National Natural Science Foundation of China (31772431) to LL, the National Science Foundation of Shandong Province (ZR2021QC187) to WZ, and the Program of Qilu Young Scholars of Shandong University to XC. Research reported in this publication was also supported by the Researchers Supporting Project (RSP-2022R7) of the King Saud University, Saudi Arabia to SAA and National Institutes of Health grant No. P40OD010964 subaward to NAS. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

ACKNOWLEDGMENTS

We thank to Ying Yan (Ocean University of China, China) for her advice on preparing the manuscript. We acknowledge the computing resources provided on IEMB-1, a high-performance computing cluster operated by the Institute of Evolution and Marine Biodiversity, Ocean University of China.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2022.775646/full#supplementary-material

Supplementary Figure 1 | A potential cis-regulatory element (CRE) in 3′ subtelomeric regions of two Pseudokeronopsis ciliate genomes. (A) Density plot showing the intermediate position of the CRE in a motif “TTNATTCTTAA” between the transcription end site (TES) and translation stop codon in 3′ subtelomeric regions of P. carnea (left) and P. flava (right) genomes. (B) Frequency of base substitutions in 5′ CRE and entire subtelomeric region in P. carnea. “AC” indicates substitutions where adenine is replaced by cytosine, for example. (C) The presence of a putative 3′ CRE is associated with the transcription level of its adjacent genes. TPM, transcripts per million. ***/\( p < 0.0001 \). (D) Pie plots showing the percentage of active and silent genes correlated with the presence or absence of the putative 3′ CRE on the nanochromosome.

Supplementary Figure 2 | Pathway annotation by Gene Ontology (GO) of expanded gene families in spirotrichs: (A) Pseudokeronopsis carnea, (B) P. flava, (C) Oxytricha trifallax, (D) Stylonychia lemnata, (E) Strombidium sculcatum, and (F) Euplotes vanus. Q values (FDR) are indicated by color scale and number of expanded genes in each pathway is indicated by the dot size. Rich factor (as indicated in the Y-axis) is the ratio of the number of expanded genes in a pathway to the number of all annotated genes in this pathway.

Supplementary Figure 3 | Pathway annotation by Gene Ontology (GO) of expanded gene families in Tetrahymena thermophila (A), Paramecium tetraurelia (B), and Stentor coeruleus (C). Q values (FDR) are indicated by color scale and number of expanded genes in each pathway is indicated by the dot size. Rich factor (as indicated in the Y-axis) is the ratio of the number of expanded genes in a pathway to the number of all annotated genes in this pathway.
REFERENCES

Baek, Y.-S., Jung, J.-H., and Min, G.-S. (2011). Redescription of two marine ciliates (Ciliophora: Urostylida: Pseudokororopodidae), Pseudokororopopsis arnea and Uroleptopus citrina, from Korea. Anton. Syst. Evol. Diversity 27, 220–227. doi: 10.5365/KSIS.2011.27.3.220

Bailey, T. L., Johnson, J., Grant, C. E., and Noble, W. S. (2015). The MEME suite.

Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., et al. (2009). BLAST+: architecture and applications. BMC Bioinformatics 10:421. doi: 10.1186/1471-2105-10-421

Chalker, D. L., and Yao, M.-C. (2011). DNA elimination in ciliates: transposon domestication and genome surveillance. Annu. Rev. Genet. 45, 227–246. doi: 10.1146/annurev-genet-110410-112432

Chen, X., Jiang, Y., Gao, F., Zheng, W., Krock, T. J., Stover, N. A., et al. (2019). Genome analyses of the new model protist Euplotes vannius focusing on genome rearrangement and resistance to environmental stressors. Mol. Ecol. Resour. 19, 1292–1308. doi: 10.1111/1755-0998.13023

Chen, X., Wang, Y., Sheng, Y., Warren, A., and Gao, S. (2018). GPSt: an automated method for evolutionary analysis of nonculturable ciliated microeukaryotes. Mol. Ecol. Resour. 18, 700–713. doi: 10.1111/1755-0998.12750

Crooks, G. E., Hon, G., Chandonia, J. M., and Brenner, S. E. (2004). WebLogo: a sequence logo generator. Genome Res. 14, 1188–1190. doi: 10.1101/gr.49004

De Bie, T., Cristianini, N., Demuth, J. P., and Hahn, M. W. (2006). CAFE: a computational tool for the study of gene family evolution. Bioinformatics 22, 1269–1271. doi: 10.1093/bioinformatics/btl097

Dong, J., Li, L., Fan, X., Ma, H., and Warren, A. (2020). Two Urosoma species (Ciliophora, Hypotrichia): a multidisciplinary approach provides new insights into their ultrastructure and systematics. Eur. J. Protistol. 72:125661. doi: 10.1016/j.ejop.2019.125661

Emms, D. M., and Kelly, S. (2019). OrthoFinder: phylogenetic orthology inference for comparative genomes. Bioinformatics 20:238. doi: 10.1186/s13059-019-1832-y

Gurevich, A., Saveliev, V., Vyahhi, N., and Tesler, G. (2013). QUAST: quality assessment of genome assemblies. Bioinformatics 29, 1072–1075. doi: 10.1093/bioinformatics/btt086

Harris, R. M., and Hofmann, H. A. (2015). Seeing is believing: dynamic evolution of gene families. Proc. Natl. Acad. Sci. U. S. A. 112, 1252–1253. doi: 10.1073/pnas.1423683112

Hedges, S. B., Dudley, J., and Kumar, S. (2006). TimeTree: a public knowledge-base of divergence times among organisms. Bioinformatics 22, 2971–2972. doi: 10.1093/bioinformatics/btl055

Huang, J., and Katz, L. A. (2021). Taxonomy of three oxytrichids (Protozoa, Ciliophora, Hypotrichia), with establishment of the new species Rubrioxytricha guangzhouensis spec. nov. Front. Mar. Sci. 7:1228. doi: 10.3389/fmars.2020.625436

McGrath, C. L., Zufall, R. A., and Katz, L. A. (2007). Variation in macronuclear genome content of three ciliates with extensive chromosomal fragmentation: a preliminary analysis. J. Eukaryot. Microbiol. 54, 242–246. doi: 10.1111/j.1550-7407.2008.00257.x

Miller, M. A., Pfeiffer, W., and Schwartz, T. (2010). “Creating the CIPRES Science Gateway for inference of large phylogenetic trees,” in 2010 Gateway Computing Environments Workshop (GCE) (New Orleans: IEEE), 1–8. doi: 10.1109/GCE.2010.5676129

Pan, B., Chen, X., Hou, L., Zhang, Q., Qu, Z., Warren, A., et al. (2019). Comparative genomics analysis of ciliates provides insights on the evolutionary history within “Nassophorea–Synhymenia–Phyllopharyngea” assemblage. Front. Microbiol. 10:2819. doi: 10.3389/fmicb.2019.02819

Park, T., Wijeratne, S., Meulia, T., Firkins, J. L., and Yu, Z. (2021). The topology and divergence times in the absence of a molecular clock. Genomics 113, 4146–4127. doi: 10.1016/j.ygeno.2021.03.014

Prescott, D. M. (2000). Genome gymnastics: unique modes of DNA evolution and processing in ciliates. Nat. Rev. Genet. 1, 191–198. doi: 10.1038/35042057

Riley, J. L., and Katz, L. A. (2001). Widespread distribution of extensive chromosomal fragmentation in ciliates. Mol. Biol. Evol. 18, 1372–1377. doi: 10.1093/oxfordjournals.molbev.a003921

Sanderson, M. J. (2003). eLRT: inferring absolute rates of molecular evolution and divergence times in the absence of a molecular clock. Bioinformatics 19, 301–302. doi: 10.1093/bioinformatics/btf2019

Sheng, Y., Duan, L., Cheng, T., Qiao, Y., Stover, N. A., and Gao, S. (2020). The completed macronuclear genome of a model ciliate Tetrahymena thermophila and its application in genome scrambling and copy number analyses. Sci. China Life Sci. 63:1534. doi: 10.1007/s11427-020-1689-4

Song, W., Sun, P., and Ji, D. (2004). Redefinition of the yellow hypotrichous ciliate, Pseudokororopopsis flava (Hypotrichia: Ciliophora). J. Mar. Biol. Assoc. U. K. 84, 1137–1142. doi: 10.1017/S0025315404010574

Kent, W. J. (2002). BLAT—the BLAST-like alignment tool. Genome Res. 12, 656–664. doi: 10.1101/gr.229202

Kim, D., Pertea, G., Trapnell, C., Pimentel, H., Kelley, R., and Salzberg, S. L. (2013). TopHat2: accurate alignment of transcripts in the presence of insertions, deletions and gene fusions. Genome Biol. 14, R36. doi: 10.1186/1475-2164-14-36

Kumar, S., Stecher, G., and Tamura, K. (2016). MEGAH: molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. Mol. Biol. Evol. 33, 1870–1874. doi: 10.1093/molbev/msw054

Lartillot, N., Lepage, T., and Blaquart, S. (2009). PhyloBayes 3: a Bayesian software package for phylogenetic reconstruction and molecular dating. Bioinformatics 25, 2286–2288. doi: 10.1093/bioinformatics/btp368

Levine, M., and Tjian, R. (2003). Transcription regulation and animal diversity. Nature 424, 147–151. doi: 10.1038/nature01763

Li, B., and Dewey, C. N. (2011). RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. BMC Bioinformatics 12:323. doi: 10.1186/1471-2105-12-323

Li, C., Chen, X., Zheng, W., Doak, T. G., Fan, G., Song, W., et al. (2021). Chromosome organization and gene expansion in the highly fragmented genome of the ciliate Strombidium stylifer. J. Genet. Genomics 48, 908–916. doi: 10.1016/j.jgg.2021.05.014

Liao, Y., Smyth, G. K., and Shi, W. (2013). featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. Bioinformatics 30, 923–930. doi: 10.1093/bioinformatics/btt556

Lin, Z., Wu, W.-S., Liang, H., Woo, Y., and Li, W.-H. (2010). The spatial distribution of cis regulatory elements in yeast promoters and its implications for transcriptional regulation. BMC Genomics 11:581. doi: 10.1186/1471-2164-11-581

Luo, X., Huang, J., Bourland, W. A., El-Serehy, H. A., Al-Farraj, S. A., Chen, X., et al. (2021). Taxonomy of three oxytrichids (Protozoa, Ciliophora, Hypotrichia), with establishment of the new species Rubrioxytricha guangzhouensis spec. nov. Front. Mar. Sci. 7:1228. doi: 10.3389/fmars.2020.625436
Zheng et al.

Cis-Regulatory Element on Nanochromosomes

Špaková, T., Pristaš, P., and Jávoršký, P. (2014). Telomere repeats and macronuclear DNA organization in the soil ciliate Kahliella matisi (Ciliophora, Hypotricha). Eur. J. Protistol. 50, 231–235. doi: 10.1016/j.ejop.2014.03.002

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

Stanke, M., Zvětková, A., and Morgenstern, B. (2006). AUGUSTUS at EGASP: using EST, protein and genomic alignments for improved gene prediction in the human genome. Genome Biol. 7:S11. doi: 10.1186/gb-2006-7-s1-s11

Steinbrück, G., Haas, L., Hellmer, K.-H., and Ammermann, D. (1981). Characterization of macronuclear DNA in five species of ciliates. Chromosoma 83, 199–208. doi: 10.1007/BF00286789

Swart, E. C., Bracht, J. R., Magrini, V., Minx, P., Chen, X., Zhou, Y., et al. (2013). The Oxytricha trifallax macronuclear genome: a complex eukaryotic genome with 16,000 tiny chromosomes. PLoS Biol. 11:29. doi: 10.1371/journal.pbio.1001473

Zhao, L., Gao, F., Gao, S., Liang, Y., Long, H., Is, Z., et al. (2021). Biodiversity-based development and evolution: the emerging research systems in model and non-model organisms. Sci. China Life Sci. 64, 1236–1280. doi: 10.1007/s11427-020-1915-y

Zhao, X., Li, Y., Duan, L., Chen, X., Mao, F., Juma, M., et al. (2020). Functional analysis of the methyltransferase SMYD in the single-cell model organism Tetrahymena thermophila. Mar. Life Sci. Technol. 2, 109–122. doi: 10.1007/s42995-019-00025-y

Zhao, X., Xiong, J., Mao, F., Sheng, Y., Chen, X., Feng, L., et al. (2019). RNAi-dependent Polycomb repression controls transposable elements in Tetrahymena. Gene Dev. 33, 348–364. doi: 10.1101/gad.320796.118

Zheng, W., Wang, C., Lynch, M., and Gao, S. (2021). The compact macronuclear genome of the ciliate Halteria grandinella: a transcriptome-like genome with 23,000 nanochromosomes. mBio 12, e01964–20. doi: 10.1128/mBio.01964-20

Zheng, W., Wang, C., Yan, Y., Gao, F., Doak, T. G., and Song, W. (2018). Insights into an extensively fragmented eukaryotic genome: de novo genome sequencing of the multinuclear ciliate Uroleptopsis citrina. Genome Biol. Evol. 10, 883–894. doi: 10.1093/gbe/evy055

Author Disclaimer: The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher’s Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Zheng, Dou, Li, Al-Farraj, Byerly, Stover, Song, Chen and Li. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.