Mitogenomic architecture and evolution of the soil ciliates *Colpoda*

Yuanyuan Zhang,[1,2] Haichao Li,[1] Yaohai Wang,[1] Mu Nie,[1] Kexin Zhang,[1] Jiao Pan,[1] Yu Zhang,[1,3] Zhiqiang Ye,[4] Rebecca A. Zufall,[5] Michael Lynch,[6] Hongan Long[1,2]

AUTHOR AFFILIATIONS See affiliation list on p. 14.

ABSTRACT *Colpoda* are cosmopolitan unicellular eukaryotes primarily inhabiting soil and benefiting plant growth, but they remain one of the least understood taxa in genetics and genomics within the realm of ciliated protozoa. Here, we investigate the architecture of *de novo* assembled mitogenomes of six *Colpoda* species, using long-read sequencing and involving 36 newly isolated natural strains in total. The mitogenome sizes span from 43 to 63 kbp and typically contain 28–33 protein-coding genes. They possess a linear structure with variable telomeres and central repeats, with one *Colpoda elliotti* strain isolated from Tibet harboring the longest telomeres among all studied ciliates. Phylogenomic analyses reveal that *Colpoda* species started to diverge more than 326 million years ago, eventually evolving into two distinct groups. Collinearity analyses also reveal significant genomic divergences and a lack of long collinear blocks. One of the most notable features is the exceptionally high level of gene rearrangements between mitochondrial genomes of different *Colpoda* species, dominated by gene loss events. Population-level mitogenomic analysis on natural strains also demonstrates high sequence divergence, regardless of geographic distance, but the gene order remains highly conserved within species, offering a new species identification criterion for *Colpoda* species. Furthermore, we identified underlying heteroplasmic sites in the majority of strains of three *Colpoda* species, albeit without a discernible recombination signal to account for this heteroplasmy. This comprehensive study systematically unveils the mitogenomic structure and evolution of these ancient and ecologically significant *Colpoda* ciliates, thus laying the groundwork for a deeper understanding of the evolution of unicellular eukaryotes.

IMPORTANCE *Colpoda*, one of the most widespread ciliated protozoa in soil, are poorly understood in regard to their genetics and evolution. Our research revealed extreme mitochondrial gene rearrangements dominated by gene loss events, potentially leading to the streamlining of *Colpoda* mitogenomes. Surprisingly, while interspecific rearrangements abound, our population-level mitogenomic study revealed a conserved gene order within species, offering a potential new identification criterion. Phylogenomic analysis traced their lineage over 326 million years, revealing two distinct groups. Substantial genomic divergence might be associated with the lack of extended collinear blocks and relaxed purifying selection. This study systematically reveals *Colpoda* ciliate mitogenome structures and evolution, providing insights into the survival and evolution of these vital soil microorganisms.

KEYWORDS ciliated protozoa, mitochondria, evolutionary genomics

A s a semi-autonomous organelle, the mitochondrion possesses its own multicopy genome, i.e., mitogenome, which contains important genes related to the electron transfer chain, ribosomal protein synthesis, and other functions (1, 2). Due to their high

genetic diversity (3, 4) and rapid evolutionary rate (5), mitogenomes have high precision in identifying highly similar species (6, 7). Multiple copies of mitogenomes in a single cell also make it relatively easy for them to be amplified for sequencing (1, 8–11). Many studies suggest that in some taxa, high spontaneous mutation rates of mitogenomes are one main driver of their high diversity (12, 13). Gene rearrangements of mitogenomes also provide essential clues regarding evolution and ancient relationships (14–16). Although mitochondria have traditionally been thought to be maternally inherited (17), reports from organisms such as *Caenorhabditis elegans* (18) and *Daphnia* (19) have confirmed the presence of considerable heteroplasmy within individual mitogenomes, which has spurred investigations into genome recombination and biparental inheritance in mitochondria (20). Up to now, mitogenomes have been widely studied in biogeography (21), ecology (22), epidemiology (23), and other fields, playing an increasing role in the study of interspecific evolutionary relationships (24–27), population structure, and dynamics (20).

Compared with mitochondrial studies on human diseases and the breeding of animals and plants (28–32), attention to mitochondria shows a much slower growth in ciliates. Since the first reports on mitochondrial DNA of *Tetrahymena thermophila*, *Paramecium primaurelia,* and *Paramecium tetraurelia* in the 1980s (33, 34), mitogenomes of approximately 30 species from the Oligohymenophorea (35), Spirotrichea (36), Heterotrichea (37), and Armophorea (38) classes have been documented. However, this constitutes only a small fraction when considering the tens of thousands of ciliate species, and population-level investigation remains nearly unstudied (39). Unlike the usual circular structure, mitogenomes of ciliates are mostly linear molecules with high AT content, generally possessing telomeric sequences at both ends (1, 34), with gene rearrangements known to occur within the phylum (23). Numerous mitogenes (mitochondrial genes) and the existence of long non-coding sequences including telomeres and central repeats make these mitogenomes of ciliates quite long, generally tens of kilobases. The existence of *ymf* genes, a kind of ciliate-specific mitogenes whose structure and function remain to be explored, also suggests that mitogenomes of ciliates may show different evolutionary characteristics and paths from those of other organisms (4, 40).

Unicellular eukaryotes represent the vast majority of eukaryotic groups (41), and globally distributed ciliated protozoa (42) are considered to be crown eukaryotes that can date back to about 1,980–2,200 million years ago (Mya) in the Paleoproterozoic. Colpodea is a substantial class of ciliates that is about 900 million years old (41, 43, 44). *Colpoda*, a group of ciliates widespread all over the world especially in soil, is a classical genus of Colpodea with kidney-like morphology (45) and unique life history (46, 47). Species in this class show great ecological significance in promoting plant growth and are heavily involved in the biogeochemical cycle (48, 49). As an ancient genus, *Colpoda* has a long evolutionary history. However, the phylogenetic relationships within the genus remain controversial (24, 45, 50). The exploration of *Colpoda* mitogenomes will provide a new perspective on the origin and evolutionary mechanisms of the genus. However, except for several phylogenetic molecular markers such as *cox1*, β-tubulin (24), and partial mito-SSU rDNA (50), mitogenome-level research in the class Colpodea has rarely been reported, which is essential to explore their evolutionary history.

In this study, we isolated 36 natural strains of six *Colpoda* species from topsoil samples collected from China (35 strains; Fig. 1A) and one from Italy. After microscopy, mass culture, omics sequencing of these isolates, and bioinformatics analyses, we *de novo* assembled mitogenomes of six species with 28–33 protein-coding genes (PCGs). By comparing mitogenomic architecture both within and between species of *Colpoda*, we revealed the phylogenomic relationships and divergence time of the *Colpoda* species, identified gene rearrangement events, estimated selection pressure on the mitogenes, analyzed polymorphisms and heteroplasmies in natural strains, and detected linkage disequilibrium signals. Our study deepens the understanding of the mitogenome
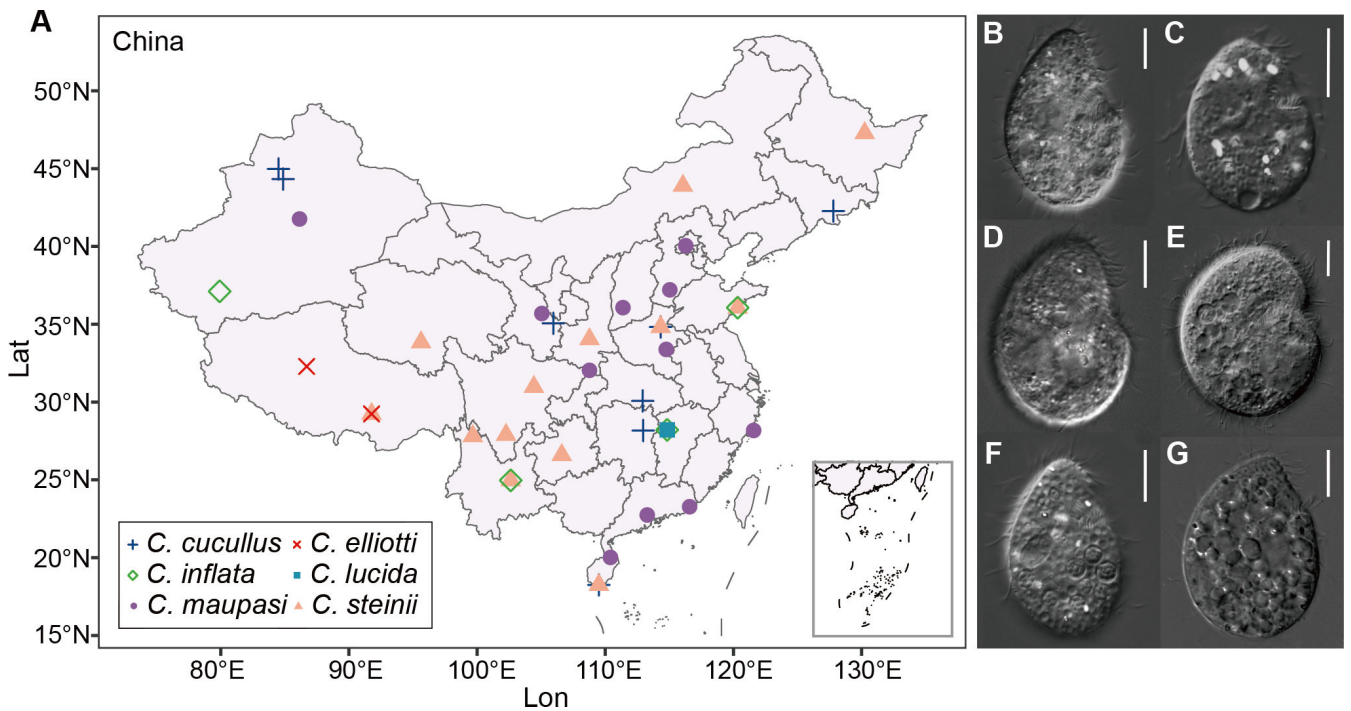
FIG 1 Sampling sites and photomicrographs of the six *Colpoda* species. (A) Geographical distribution of collected *Colpoda* strains in China. (B–G) Ventral side of *Colpoda cucullus* (MLY2A28), *C. elliotti* (LHA5931), *C. inflata* (RL4B), *C. lucida* (DZG2A41), *C. maupasi* (MW2A28), and *C. steinii* (RZ4A), respectively. The size of the scale bars is 10 µm.

macroevolution of eukaryotes and provides insights into the speciation of soil microorganisms.

RESULTS

Colpoda isolation, mitogenome assembly and annotation

Ciliates were isolated from topsoil samples collected from China and Italy. Thirty-six strains of six *Colpoda* species were obtained—*C. cucullus* (11 strains), *C. maupasi* (10), *C. elliotti* (1), *C. lucida* (1), *C. inflata* (1), and *C. steinii* (12). The geographical distribution of these species and their morphology are shown in Fig. 1. Additional details are in Table S1.

After long-read sequencing and *de novo* assembly, linear mitogenomes with lengths of 43,691–63,340 bp (mean length of 51,006 bp) and guanine-cytosine (GC) content of 18.11%–21.37% were obtained for each reference strain of the six species (Table 1). The length and GC content of mitogenomic regions with different functional contexts are shown in Fig. 2A and B, respectively. No introns were found. The mitogenomes of *C. cucullus*, *C. maupasi*, *C. elliotti,* and *C. lucida* contain telomeric sequences at both ends. The length of the telomeric repeat units ranges from 32 to 142 bp, with the repeat number ranging from 17 to 273. Contrary to the high telomeric identity observed in mitogenomes of other ciliated genera, no homology was observed between these

TABLE 1 Mitogenomic features of six *Colpoda* species

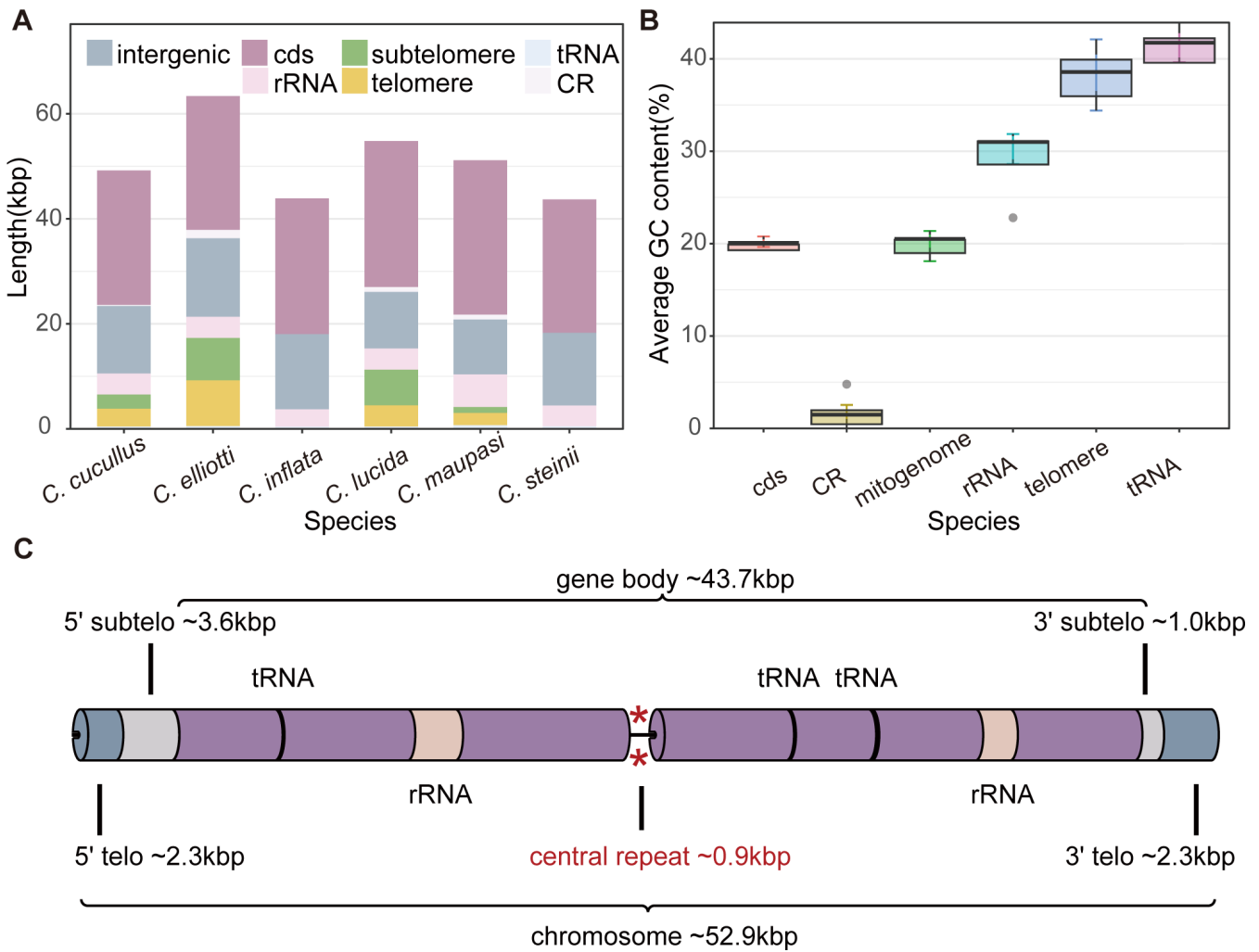| Species	Strain name	Mitogenome length (bp)	GC content	PCG numbers	tRNA numbers	Sequencing depth
C. cucullus	MLY2A28	49,189	21.04%	30	6	137
C. elliotti	LHA5931	63,340	20.22%	28	7	220
C. inflata	RL4B	43,881	19.32%	30	6	128
C. lucida	DZG2A41	54,804	21.37%	30	5	78
C. maupasi	MW2A28	51,136	20.80%	33	9	94
C. steinii	RZ4A	43,691	18.11%	30	7	80

FIG 2 Length, GC content, and blueprint of *Colpoda* mitogenomes. (A) Size distribution of different mitogenomic regions in six *Colpoda* mitogenomes. (B) Boxplots show the mean GC content of different regions; solid gray circles represent outliers. (C) Blueprint of *Colpoda* mitogenomes, visualized by EVenn (52); protein-coding genes are in purple, and their details are shown in Fig. 6.

repeat sequences, suggesting that these telomere sequences have undergone diversifying evolutionary processes or have been independently acquired by different species. For these four species, repeat sequences with high adenine-thymine (AT) content (94.65%–100%) were also found in the mitogenomes and likely constitute the central repetition (CR) regions associated with DNA replication initiation and gene transcription regulation (51). Accordingly, our analysis showed that near CR, the direction of gene transcription usually changes from the forward strand to the reverse one, implying the potential significance of this region in certain yet-to-be-explored transcription regulation mechanisms. Details of telomeres and CRs are given in Table S2. Above all, the structure of the four complete *Colpoda* mitogenomes were elucidated, although we did not acquire telomeric and central repeats in *C. inflata* and *C. steinii*, possibly due to different sequencing strategies or insufficient sequencing depth.

As shown in Fig. 2A, the total length of protein-coding sequences in the six mitogenomes varies from 25,434 to 29,379 bp, accounting for 46.59%–74.39% of the total length (excluding telomeres), which is a relatively low fraction compared to the published ciliate mitogenomes. 28–33 PCGs have been annotated and are assigned to six gene families: complex I: NADH dehydrogenase gene (*nad*), complex III: ubiquinol cytochrome c reductase (*cob*), complex IV: cytochrome C oxidase (*cox*), complex V: ATP
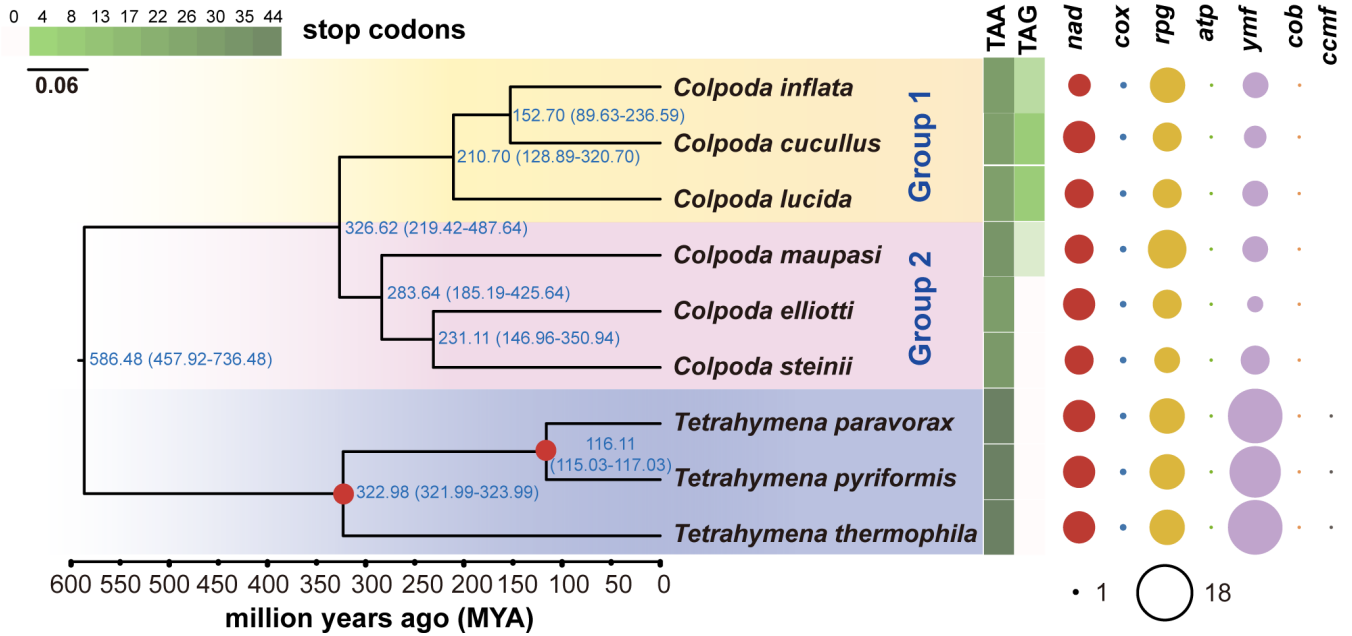
FIG 3 Maximum likelihood (ML) phylogenomic tree based on 22 core genes of the *Colpoda* mitogenomes. The yellow and pink background marks the *Colpoda* species; the blue background marks the outgroup *Tetrahymena*. Numbers beside each node show the estimated divergence time and its confidence intervals in parentheses. The red dots on the nodes indicate the time calibration points based on the divergence time estimation by Xiong et al. (53). The heat map in green gradients represents stop codon usage in all PCGs of each mitogenome, and the bubble map on the right panel displays the types and numbers of mitogenes.

synthetase (*atp*), ribosome protein genes (*rpg*), and *ymf*, ciliates-specific genes with high diversity and evolutionary rate (40). *nad1* is split into two parts (*nad1_a*, *nad1_b*) in all *Colpoda* species, as in other ciliates. However, a highly conserved mitogene related to the cytochrome c maturation in ciliates, *ccmf* (51), is not found in *Colpoda* mitogenomes. There are two stop codons in the mitochondrial genomes of *Colpoda*: TAA and TAG, with TAA being the main stop codon of *Colpoda* and the only one in the *Tetrahymena* mitogenome, while stop codon TAG is used less but exists widely in four *Colpoda* species
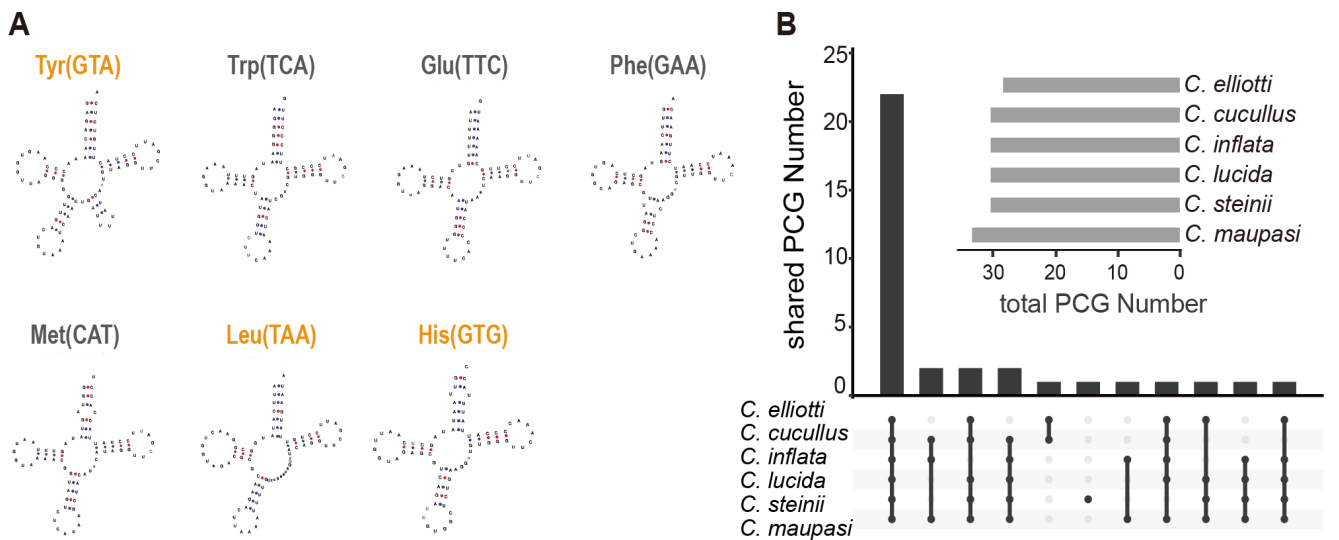

FIG 4 tRNAs and protein-coding gene numbers of *Colpoda* mitogenomes. (A) Secondary structures and loss events of *Colpoda* mito-tRNA. Yellow texts indicate the tRNAs that have been lost in the mitogenomes of group 1 (*C. cucullus*, *C. inflata*, and *C. lucida*). (B) The horizontal bars in the inset show the total gene number of the mitogenomes; vertical bars show the number of shared PCGs by different *Colpoda* species.

(Fig. 3). As in other ciliates, two mito-rRNA genes were annotated in all six *Colpoda* species, but the number of tRNAs varies from 6 to 9 among different species. All of the latter possess the typical four-armed cloverleaf secondary structure, as shown in Fig. 4A.

Phylogenomic relationships and evolutionary history of the *Colpoda* genus

Homologous gene analysis by Orthofinder v2.5.4 revealed 35 orthogroups (Fig. 4B), of which 22 were core genes that existed in all six *Colpoda* mitogenomes: *nad1, nad3, nad4, nad4L, nad6, nad7, nad10, cox1, cox2, cob, atp9, rpl2, rpl16, rps12, rps13, rps14, rps19, ymf57, ymf65, ymf66, ymf67,* and *ymf68*. Additionally, the homology analysis showed that certain *ymf* genes were homologous to some ribosomal protein genes, i.e., *ymf60* was homologous to *rpl6, ymf59* to *rps10*, and *ymf76* to *rps4*.

As shown in Fig. 3, the phylogenomic tree based on concatenated core gene sequences showed that *Colpoda* species form two groups, with *C. cucullus, C. inflata,* and *C. lucida* as the first group and another three species as the second. Although our results were slightly different from those of previous studies, such as the closer relationship of *C. inflata* and *C. cucullus* in this study vs *C. lucida* and *C. cucullus* in Foissner et al. (45), this may be caused by different input genes and length and tree construction methods. High bootstrap values, which are shown in Fig. S1, indicate the reliability of core mito-PCGs in reconstructing the evolutionary relationship of *Colpoda* species.

The molecular clock is a dependable technique for gauging the rate of organism evolution, which relies on the premise that genomes accumulate neutral mutations at a consistent rate throughout time. The extrapolation of the divergence time based on phylogenomics calibrated by the molecular clock indicates that *Colpoda* diverged from the lineage leading to *Tetrahymena* about 586 Mya during the Proterozoic period (Fig. 3). At least 326 Mya, i.e., the Carboniferous period, when nearly 50% of the animals on land experienced extinction (54), *Colpoda* diverged into two lineages. The second lineage began to diverge within the next 43 million years or so, and it fully diverged into three species: *C. maupasi, C. steinii,* and *C. elliotti* within the next 52 million years or so. The ancestors of the first lineage took 116 million years to diverge into *C. lucida*, and finally, *C. inflata* and *C. cucullus* originated about 152 Mya. We verified the reliability of the inferences by two independent runs for the divergence time ($R^2 = 1.00$, Fig. S2).

High mitogenomic divergence within the *Colpoda* genus

Using collinearity analysis, we identified many syntenic blocks among mitogenomes within the genus. As shown in Fig. 5, these collinear blocks avoid the regions where the telomeres and CRs are located; there are also some gaps in the remaining regions, primarily consisting of intergenic sequences and a few genes with low sequence identity. The total length of collinearity blocks between species pairs varies largely from 8.50 kbp (*C. elliotti* and *C. inflata*) to 33.50 kbp (*C. cucullus* and *C. inflata*). Overall, the interspecies collinearity block in group 1 is the largest, with the mean length of 30.25 kbp, followed by 17.79 kbp of group 2. The synteny degree between the two groups is the lowest, which is just 14.0 kbp in length. For single collinear blocks, most of them are less than 2.50 kbp in length, suggesting that long continuous collinear sequences may have been interrupted by frequent evolutionary events such as gene rearrangements. All seven blocks longer than 2.50 kbp are from species within the first group, possibly due to closer relatedness between species in this clade.

Unusual gene rearrangement patterns in *Colpoda* mitogenomes

Generally, the mitogenome structure is unlikely to be totally identical in class-level phylogenetic comparison, although gene orders of mammals, insects, and fish are relatively conservative (14). However, ciliate mitogenomes show a unique organization compared with other organisms. Usually, they show high agreement in gene order at the genus level but lack a uniform pattern within the whole phylum. In striking contrast, *Colpoda* mitogenomes show a nearly scrambled gene order between congeners, leading
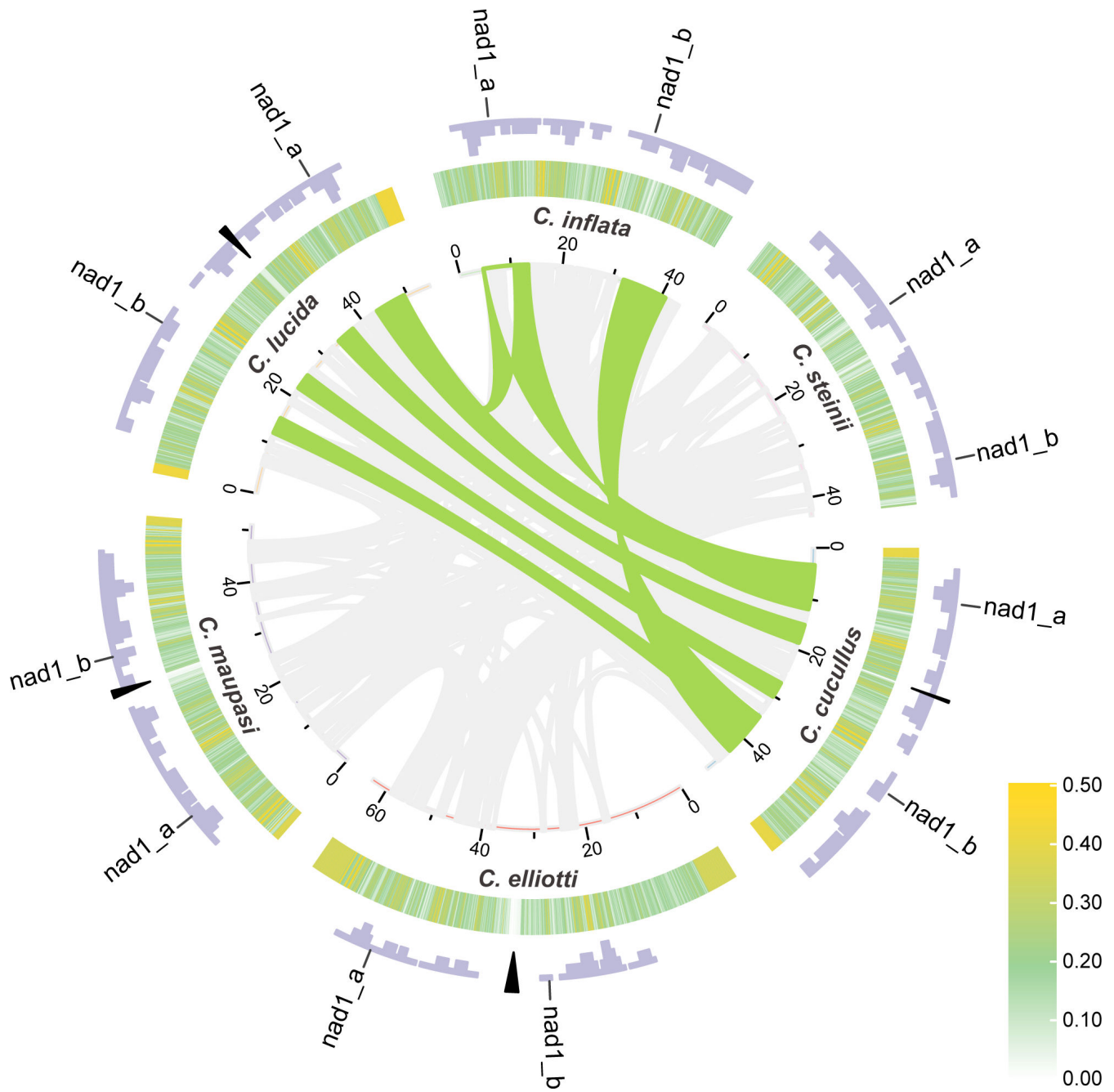
FIG 5 Collinearity of *Colpoda* mitogenomes. The lines connecting different species in the central circle represent collinearity, gray lines show synteny blocks less than 2.5 kbp, and green lines show synteny blocks greater than 2.5 kbp. The tick marks represent the coordinates of the mitogenomes in kilobase pair. The heat map in the middle track shows the GC content (window size = 0.1 kbp, step size = 0.05 kbp). The outmost purple bars show gene density (window size = 1 kbp, step size = 0.5 kbp), the black triangles show the locations of the CR regions, and nad1_a and nad1_b indicate two parts of the split gene *nad1*.

to a no consensus mitochondrial gene order for the genus (Fig. 6). Given the resemblance of mitochondrial gene order between *C. maupasi* and most non-*Colpoda* species in the class Oligoheymenophorea, we established the gene order of *C. maupasi* as the reference standard and conducted comparisons with other *Colpoda* species to quantify the extent of rearrangements among congeners.

We used qMGR (56) to quantify the rearrangement events that cause mitochondrial gene order variation within *Colpoda*, which assigns equal weight to different types of rearrangement events (i.e., gene deletion, inversion, and translocation). The results are
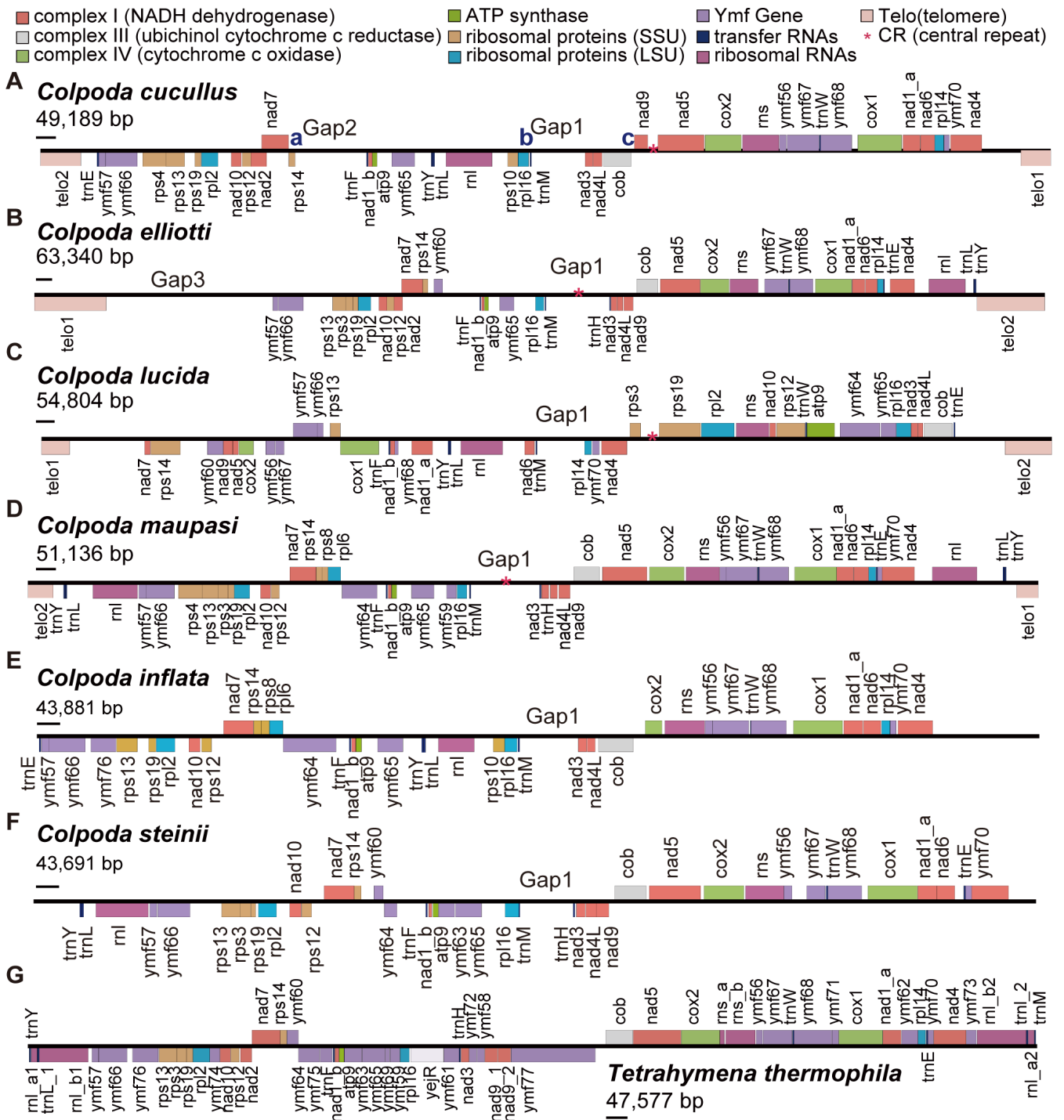
FIG 6 Gene orders of *Colpoda* spp. vs *Tetrahymena thermophila* mitogenomes. (A–F) The gene contents of six *Colpoda* mitogenomes. Blue letters indicate three rearrangement hotspots in *C. cucullus*. (G) Organization of the *Tetrahymena thermophila* mitogenome (55). The short horizontal bars beneath the genome sizes are scale bars of 1 kbp.

shown in Fig. 7A. Among all genes, the mean rearrangement frequency (RF) of tRNA is 47.91, and that of PCGs is 37.14, indicating that tRNA genes are more often involved in gene rearrangements. *trnY-trnL-rnl* is the most frequently rearranged hotspot, with a mean RF of 69.40. The most conserved gene block is *nad1_a-cox1-ymf68-trnW-ymf67*. Overall, regardless of the greater overall sequence divergence, the three species in group
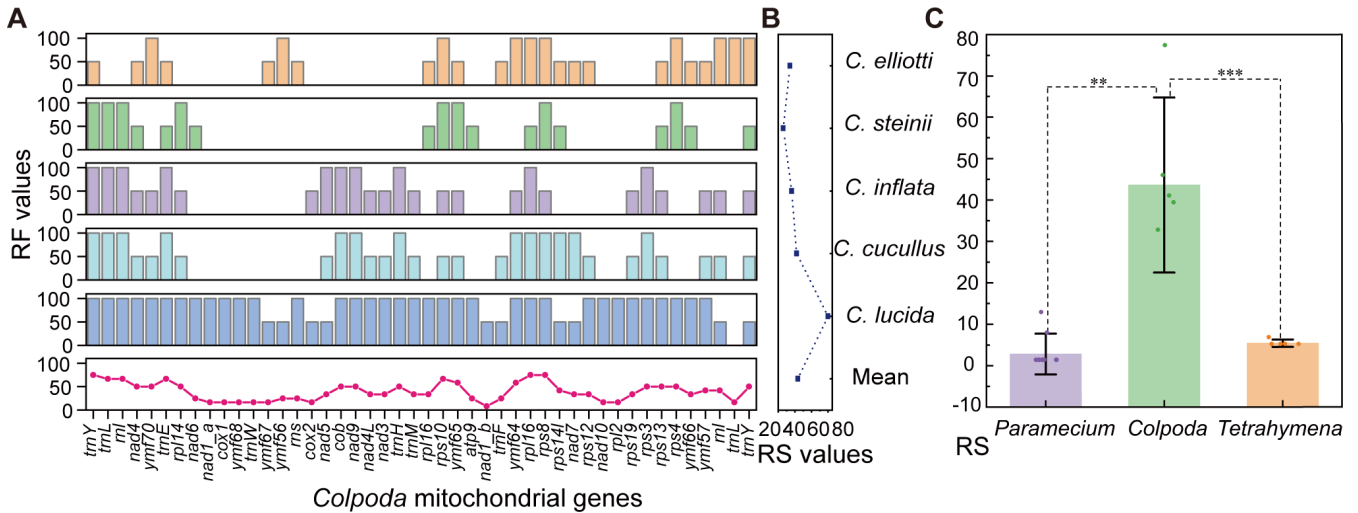
FIG 7 Rearrangement estimation of *Colpoda* mitogenomes. (A) Upper five histograms show distribution of RF of mitogenes in different species. The line plot at the bottom shows the mean RF value per gene across the five *Colpoda* species. The gene order of *C. maupasi* was used as the baseline reference. (B) RS of six *Colpoda* mitogenomes. (C) *t*-tests of RS values for *Colpoda*, *Paramecium*, and *Tetrahymena*. ** indicates a *P*-value <0.01; *** means a *P*-value <0.001.

2 that diverged earlier have experienced fewer rearrangement events [mean rearrangement score (RS) = 20]; that is, they may retain more ancestral gene order patterns. Mitogenomes of *C. inflata* and *C. cucullus* of group 1 rearranged more actively, with RS of 36 and 42 (Fig. 7B), respectively. As an exception, in the *C. lucida* mitogenome, all genes except *trnL* experienced at least one rearrangement event, elevating the total RS value to 80, which was the highest among all six species.

According to the above analyses, the gene orders of mitogenomes in different *Colpoda* species are extremely scrambled by rearrangement events. Gene losses are the most dominant rearrangement events in the mitogenomes of *Colpoda,* and the degree of rearrangements is not necessarily associated with species relatedness (Fig. 7A). In contrast, after filtering out low-quality *de novo* mitogenome assemblies of strains belonging to three species (*C. maupasi, C. cucullus,* and *C. steinii*) short-read-sequenced, we observed that the mitogene orders within nine *C. steinii* strains remained highly conserved, as shown in Fig. S3.

We also investigated gene rearrangement events in other ciliate mitogenomes, as shown in Fig. 7C and Table 2. The results of *t*-tests revealed an extreme increase in the RS of *Colpoda* mitogenomes (mean = 43.60, *P* < 0.05), compared to the conserved pattern observed in *Tetrahymena* (5.40) and *Paramecium* (2.80). After taking into account of estimates from most previous reports, we also found that *Colpoda* mitogenomes are one of the highest in rearrangement scores among mitogenomes of all studied eukaryotes (Table 2), although the selection of benchmark genome might somewhat affect the RS score. Future efforts on normalization and consideration of lineage depths may also be needed for more accurate estimation.

Widespread relaxed purifying selection

To explore the natural selection on mitogenomes of the *Colpoda* genus, we conducted an analysis of selection pressure on 35 orthogroups in six *Colpoda* mitogenomes by ParaAT v2.0 and KaKs_Calculator v2.0. As shown in Fig. 8A, the Ka/Ks values of all genes are less than 1, with a mean value of 0.37, indicative of universal purifying selection. Among all PCGs, *nad7* was the most conserved, with the lowest Ka/Ks value (0.08), while *rps3* had the highest Ka/Ks value (0.86). We also found that to a large extent, the genes directly involved in the electron transport chain of aerobic respiration are highly preserved by purifying selection (Fig. 8A). For example, those in complex V (*atp*) were the most conserved with a Ka/Ks value of 0.11, followed by complex III (*cob*), complex

TABLE 2 Rearrangement scores of mitogenomes in different organisms

Genus or lineage	Species number	RS (mean)	RS (SE)	Reference
Fishes	2,558	0.13	~0.91	(56)
Tetrahymena[a]	6	5.40	2.14	(35, 40, 51, 55)
Mammals	1,055	0.24	~1.58	(56)
Amphibians	242	1.52	~2.83	(56)
Paramecium[a]	11	2.80	4.52	(20, 51)
Reptiles	305	3.22	~4.41	(56)
Birds	614	5.85	~0.60	(56)
Fungi[a]	38	32.95	2.13	(57)
Colpoda[a]	6	43.60	23.70	This study
Acari[a]	5	69.40	3.88	(58)

[a]Indicates the RS values calculated in this study.

IV (*cox*), and complex I (*nad*) genes with values of 0.14, 0.17, and 0.20, respectively. As two groups of genes that account for more than half of the *Colpoda* mitogenomes, *rpg* genes, which are involved in ribosomal protein synthesis, and the *ymf* genes, which have the highest variability, their mean Ka/Ks values are as high as 0.49 and 0.51, which are about three times that of Ka/Ks values of the aforementioned four genes in the electron transport chain, thus experiencing less selective constraint and contributing greatly to the high overall mitogenome divergence. This is also consistent with the conclusion proposed before that *ymf* genes of *Tetrahymena* are mutation hotspots (40). Figure 8B shows the Ka/Ks values among 15 species pairs. The mean Ka/Ks is 0.33. The lowest value is 0.13 of *C. cucullus* and *C. inflata*, followed by *C. cucullus–C. lucida* and *C. lucida–C. inflata* pairs, which confirmed the lower degree of differentiation among group 1 species. The highest Ka/Ks value of 0.42 is found in the *C. lucida–C. steinii* pair. A sliding-window
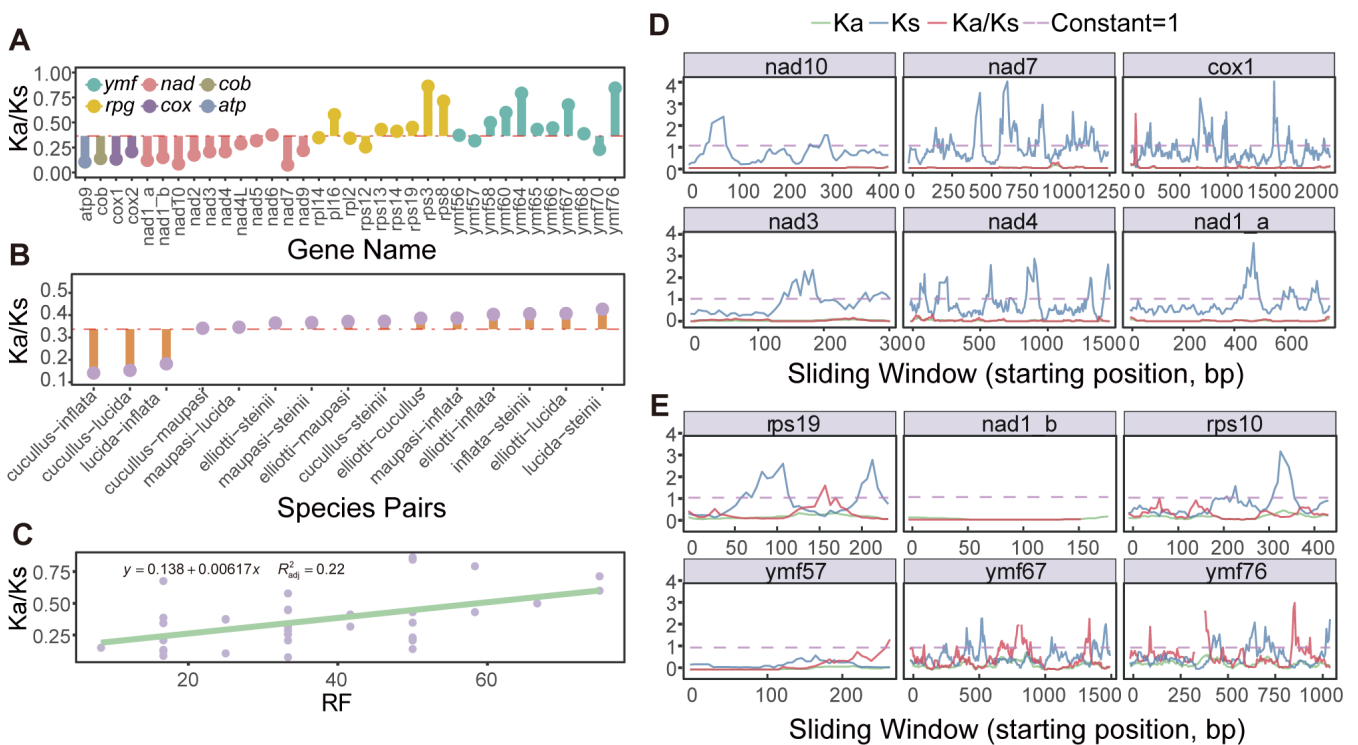

FIG 8 Selective pressure on *Colpoda* mitogenes. (A) Lollipop chart shows mean Ka/Ks values of each homologous gene. The red dotted line represents the mean value. (B) Mean Ka/Ks values of each species pair. (C) Non-significant correlation between RF and selection pressure (Ka/Ks). The Ka/Ks values increase from left to right. (D–E) Changes in Ka, Ks, and Ka/Ks values of six gene pairs with minimum and maximum Ka/Ks values of *C. cucullus* and *C. inflata*, corresponding to green, blue, and red, respectively. *X* coordinates represent the starting site of each sliding window; the purple dashed lines represent constant 1.

method was also used to show the distribution of Ka/Ks values of each gene pair in more details (Fig. 8D and E). There was not any significant correlation between selection strength and the frequency of rearrangement events (Fig. 8C; $R^2 = 0.22$, $P = 0.0028$).

Polymorphisms and heteroplasmies inferred from population genomics of *Colpoda*

The median of the sequencing depth for the 33 natural strains of *C. cucullus*, *C. maupasi*, and *C. steinii* is 68×, and the median of coverage breadth is 79.20% (Table S1). The mito-SSU rRNA single-gene tree shows the expected clustering of strains of the same species, demonstrating high species delimitation, and the phylogenetic distance of *C. maupasi* and *C. steinii* is closer than that with *C. cucullus*, which is consistent with the multi-gene phylogenetic inference and highlights the potential of this gene for species identification (Fig. 3 and 9A). After SNP calling and filtering, we identified in total 3,262, 3,752, and 4,408 SNP sites within the mitogenomes of *C. cucullus*, *C. maupasi*, and *C. steinii*, respectively. This corresponds to densities of 66, 73, and 101 SNP sites per kilobase pair (Fig. S4). Potential heteroplasmy was then examined within each strain of *Colpoda*. Among the strains of *C. cucullus* (11), *C. maupasi* (10), and *C. steinii* (12), seven, five, and six strains, respectively, were identified with potential heteroplasmic loci. The cumulative number of heteroplasmic sites for each species was 37, 14, and 20, respectively (Fig. 9). However, after the LD decay test, no significant correlation was observed between the value of r^2 and the distance separating the sites, although our estimates of r^2 values may be on the high side due to the influence of, e.g., mutation hotspots. This lack of correlation implies that recombination is rare within *Colpoda* mitogenomes (Fig. 9B).

Despite the presence of heteroplasmic sites, a single consensus sequence was obtained for each strain based on the major alleles. Using these consensus sequences, we calculated the nucleotide diversity of the mitogenomes between individuals of each species, resulting in mean π values of 0.04, 0.05, and 0.05, respectively (Fig. 9C; Fig. S4, track d). Notably, certain genes deviating from the norm are labeled in Fig.
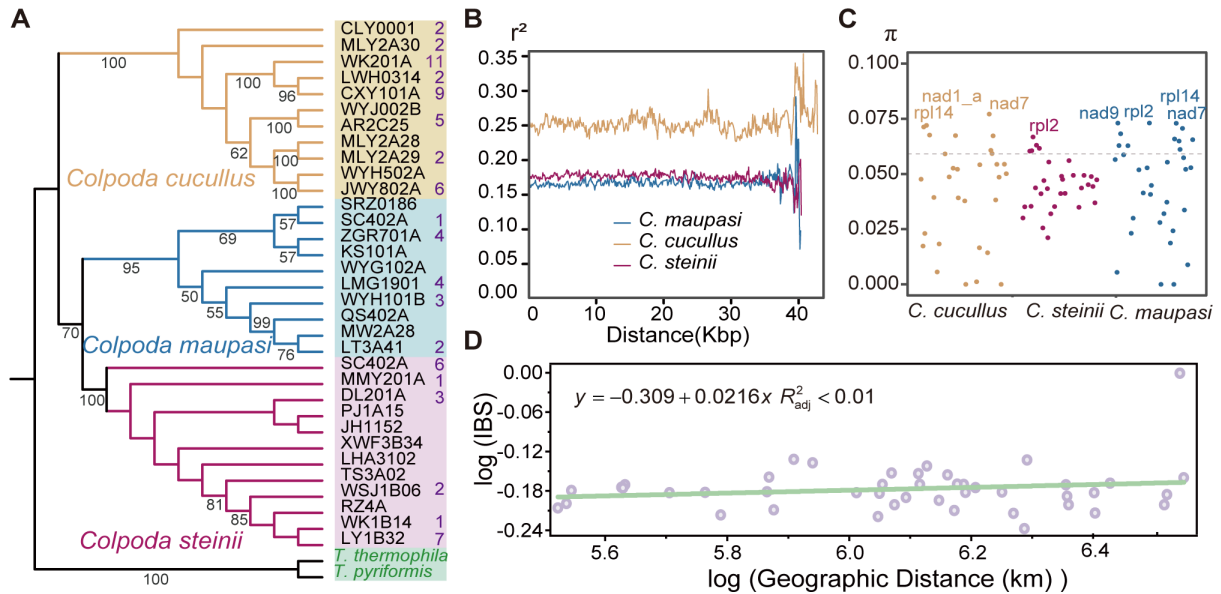

FIG 9 Population genetic analyses on natural strains of *C. cucullus*, *C. maupasi*, and *C. steinii*. (A) An ML phylogenetic tree was constructed based on 1,588 nucleotide sites of the mitochondrial small subunit rRNA gene of 33 *Colpoda* strains. Species are represented by strain names in black; the outgroups are *T. pyriformis* (AF160864.1) and *T. thermophila* (AF396436.1). Only bootstrap values >50 are shown. The rightmost numbers in purple are numbers of heteroplasmic sites. (B) Relationships between the linkage disequilibrium index (measured by r^2) and physical distance between sites. X-axis numbers are mitogenomic coordinates of the genes. (C) Nucleotide diversity of the PCGs of three *Colpoda* species. The dotted line represents the third quartile of π value 0.059. (D) Scatter plot of isolation by state (IBS) versus geographic distances (in kilometer) among 10 *C. maupasi* strains. Mantel test for matrix correlation between log (IBS) and log (geographic distance): $R^2 < 0.01$.

9C. In order to explore the role of geographical distance in the formation of genetic diversity, a correlation analysis between IBS and individual geographic distance was conducted, and the results are shown in Fig. 9D. Strains within the genus do not show an obvious geographic–genetic pattern on the whole, and geographical factors do not appear to cause universal restriction of gene flow in *Colpoda* mitogenomes. Figure 9D shows the association analysis between the 10 *C. maupasi* strains, and the same geographic–genetic independence could also be found in *C. cucullus* and *C. steinii* (Fig. S5).

DISCUSSION

In this study, we conducted a systematic analysis of the mitogenomes of six *Colpoda* species, including 36 natural strains isolated from over a large geographical range, taking advantage of intensive sample collection, microscopy, genome/transcriptome sequencing, and omics and evolutionary analyses. After complete mitogenome assembly and annotation, structures and evolution of mitogenes were elucidated. Phylogenomic analysis revealed the ancient origin of the genus, and the low degree of collinearity indicated substantial divergence between *Colpoda* mitogenomes. During evolution, mitogenomes of different *Colpoda* species have experienced extremely active rearrangement events, while the gene order between strains of the same species still exhibited high stability. This might suggest a high fidelity in mitochondrial inheritance within species and the presence of robust genetic boundaries between different species. Selection analysis demonstrated that although the coding regions of the *Colpoda* mitogenome experienced purifying selection, the selection pressure was weaker in over half of the genes (*rpg, ymf*) than those related to the electron transport chain of aerobic respiration (*nad, atp, cox,* and *cob*). Population-level analyses detected variation between (polymorphisms) and within (heteroplasmies) strains, as well as revealed no recombination and non-association between geographic and genetic distances. Nonetheless, it is crucial to acknowledge that the accuracy of our sequenced clones in reflecting the actual levels of intraspecific diversity may be influenced by the limited number of strains we studied and the geographic restrictions in our sampling.

The long-read sequencing technology enabled us to obtain near-complete mitochondrial genomes at a telomere-to-telomere scale for four of the six *Colpoda* species in this study. After investigating more than 20 published ciliate mitogenomes, we found that most of the repeat units of telomeres are shorter than 55 bp in length. Among them, those in *Tetrahymena thermophila* and *Miamiensis avidus* (Oligoheymenophorea) are 53 bp (33) and 30 bp (59), respectively. In the ciliate class Spirotrichea, such as *Strombidium* sp., *Sphaerium* cf. *sulcatum*, *Laurentiella strenua*, *Oxytricha trifallax,* and *Halteria grandinella,* they are 18, 34, 36, 35, and 31 bp, respectively (4, 36). By contrast, *Colpoda* have significantly longer telomeric repeat units without homology between species and higher GC content. Among them, *C. maupasi* has an extremely long telomere repeat unit with a length of 142 bp, followed by *C. lucida* and *C. cucullus,* with 76 and 65 bp, respectively, and finally *C. elliotti* with 32 bp.

Studies have shown that telomere structures can shield the exposed chromosome ends of telomeric DNA from DNA damage (60). In this study, except for *C. maupasi* (21.78%), the telomere sequences of the other three *Colpoda* species showed higher GC content (mean = 38.82%, SE = 0.03) than those of other ciliates (Table S2). This may correspond to a more stable nucleotide primary structure, which could play a role in protecting genes in the mitogenome. Surprisingly, the total length of *C. elliotti*'s telomeres reached 8.75 kbp, accounting for 13.70% of the total mitogenomic length, which is the longest telomere sequence ever reported in ciliates. Because this strain was collected from Tibet with high altitude and low oxygen content, we speculate that the extremely long telomeres are associated with the extra protection of the mitogenomes from especially harsh environmental conditions, although more empirical evidence on this is needed.

Based on the phylogenomic analysis, we estimated that *Colpoda* originated more than 326 Mya, which is highly similar to that of *Tetrahymena* (about 323 Mya) (53). Collinearity and other analyses further showed that *Colpoda* mitogenomes had higher genome diversity, particularly in the group 2 species, suggesting that the mitogenomes of *Colpoda* may have undergone extraordinarily rapid evolution.

We found that gene loss dominated the rearrangement events in *Colpoda* mitogenomes, with the fragments of these lost genes remaining in the mitogenome, resulting in long non-coding regions between certain genes. For example, as shown in Fig. 6A through D, Gap1 was caused by the loss of *ccmf* gene, and Gap2 in *C. cucullus* was caused by the loss of *ymf64* and *ymf75*. In diploid nuclear genomes, gene deletion events generally occur after whole-genome duplication in an organism, leading to the loss of some functional genes (61). To explain the gene order change in mitogenomes, Boore and Brown (62) proposed the tandem duplication–random loss model, which attaches great importance to the duplication of mitogenes followed by gene loss. They indicated that by pseudogenization and degeneration of the redundant genes, gene orders were scrambled in different patterns. While this model was generally used to explain similar gene orders, more mechanisms are needed to explain the extremely scrambled genes in *Colpoda* mitogenomes. Furthermore, the highly conserved intraspecific and extremely variable interspecific gene orders of the mitogenomes may contribute to reliable *Colpoda* species identification, by adding more evidence besides morphological traits and conservative gene barcoding. To determine the applicability of this concept to a wider range of taxa, further population-level investigations across diverse ciliate species are required.

Gene rearrangement hotspots may also accelerate gene order scrambling. For example, Montoya et al. (63) found that near mito-*SSU rRNA* in the human mitogenome, genes experienced more rearrangements. Similarly, the mean RF of *LSU rRNA-Leu-Tyr* block of *Colpoda* reached as high as 52.80, and this might be caused by a similar mechanism associated with large subunit rRNA (*rnl*). Additionally, Boore and Brown (62) proposed that near the origins of replication, gene rearrangements are extremely active, which is highly consistent with the high RS value of the sub-CR block *nad9-cob* (RS = 50). Gene orders could also be shaped by selection (64), though in *Colpoda* mitogenomes we only found a weak correlation between RS and Ka/Ks values (Fig. 8D).

Besides the extremely scrambled gene order, some unknown genes were also found in the *Colpoda* mitogenomes. For example, near the 5′ end of the *C. elliotti* mitogenome (Fig. 6B, Gap3), a sub-telomeric sequence of 10.45 kbp was identified, but the lost *trnY-trnL-rnl* block here was just about 1.9 kbp in length, which was uncommon in the tightly packed ciliate mitogenomes. Although this sequence does not align with any known gene, about 50 open reading frames greater than 75 nt in length were identified by ORFfinder (65), and RNAseq data also supported that four of them have transcripts (Fig. S6). Among them, the longest is 825 aa (+ strand; hypothetical protein 1), followed by 440 aa (− strand; hypothetical protein 2), 132 aa (+; hypothetical protein 3), and 90 aa (+; hypothetical protein 4), and start at positions of 4,554, 8,400, 10,132, and 11,816 bases from the 5′ end, all with TAA as the stop codon (Fig. S6). Further empirical investigations are required to determine their specific functions.

In addition, some known mitogenes of other ciliates are absent from *Colpoda* mitogenomes, such as the cytochrome C maturation protein gene (*ccmf*). *ccmf* is essential for the aerobic mitochondria of eukaryotes as it is associated with the maturation of cytochrome c, which binds to heme and transfers electrons between respiratory chains (66). Deletion of the *ccmf* gene may be a consequence of its relocation into the nuclei, but a partial remnant of this gene is present in *Colpoda* mitogenomes, which suggests that it may have gradually lost its function under relaxed selection pressure or the function was replaced by nuclear genes. Using *ccmf* and *ymf* genes as examples, it remains unclear whether the frequent gene loss events in *Colpoda* mitogenomes imply the elimination of redundant genes and if the aerobic respiration

process of *Colpoda,* which inhabit the soil, is altered compared to that of typical aquatic ciliates.

To conclude, we have identified extremely high gene rearrangement events in the mitogenomes of *Colpoda*, one of the most widely distributed ciliates. Considering their unique habitat (almost exclusively in the soil) and the scrambled gene orders compared to other ciliates, these rearrangements may shed light on their capacity to adapt to the harsh and ever-changing soil environments they inhabit. Moreover, the streamlined mitogenomes, possibly stemming from these extensive rearrangements, suggest natural selection for energy efficiency, achieved by shedding unnecessary genetic baggage, while simultaneously facilitating broader dispersal. However, the intricacies of evolutionary strategies may extend beyond the organelle level, prompting us to embark on an ongoing exploration of *Colpoda*'s macronuclear genomes. This endeavor promises to provide deeper insights into the mysteries surrounding these highly successful and prolific denizens of the soil and possibly even give clues to the evolution of other single-celled eukaryotic organisms.

MATERIALS AND METHODS

We newly isolated six *Colpoda* species, primarily sourced from the top 15 cm of mostly mainland China's soils (Table S1), resulting in the acquisition of 36 natural strains. These species were identified through 18S rRNA sequencing, with the cutoff value set at >99% sequence identity. Nucleic acids were extracted using the MasterPure Complete DNA and RNA Purification kit. To obtain comprehensive mitogenomic data, we employed long-read sequencing technologies, such as the PacBio HiFi and Nanopore MinION platforms, on reference strains of the six species. Additionally, we conducted Illumina NovaSeq6000 PE150 deep sequencing for all 36 strains.

To further investigate transcription patterns, single-cell RNA sequencing was performed on *C. elliotti*. Detailed procedures for mitogenome assembly and annotation, collinearity and phylogenomics analyses, gene rearrangements, and population genetics can be found in the Supplementary Materials and Methods section.

AUTHOR AFFILIATIONS

[1]Key Laboratory of Evolution and Marine Biodiversity (Ministry of Education), Institute of Evolution and Marine Biodiversity, KLMME, Ocean University of China, Qingdao, Shandong Province, China

[2]Laboratory for Marine Biology and Biotechnology, Laoshan Laboratory, Qingdao, Shandong Province, China

[3]School of Mathematics Science, Ocean University of China, Qingdao, Shandong Province, China

[4]School of Life Sciences, Central China Normal University, Wuhan, Hubei Province, China

[5]Department of Biology and Biochemistry, University of Houston, Houston, Texas, USA

[6]Biodesign Center for Mechanisms of Evolution, Arizona State University, Tempe, Arizona, USA

AUTHOR ORCIDs

Yu Zhang (iD) http://orcid.org/0000-0001-9179-0922
Hongan Long (iD) http://orcid.org/0000-0001-9767-8173

AUTHOR CONTRIBUTIONS

Yuanyuan Zhang, Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Validation, Visualization, Writing – original draft, Writing – review and editing | Haichao Li, Investigation, Methodology | Yaohai Wang, Investigation, Visualization | Mu Nie, Investigation | Kexin Zhang, Investigation | Jiao Pan, Investigation | Yu Zhang, Writing – review and editing | Zhiqiang Ye, Data curation, Formal analysis, Writing – review and editing | Rebecca A. Zufall, Writing – review and editing | Michael Lynch, Funding acquisition, Writing – review and editing | Hongan Long, Conceptualization, Funding acquisition, Project administration, Resources, Supervision, Writing – original draft, Writing – review and editing

DATA AVAILABILITY

Assembled mitogenomes and annotations are available for access via the National Genomics Data Center of China (https://ngdc.cncb.ac.cn/). The corresponding accession numbers are as follows: C_AA048790.1, C_AA048789.1, C_AA048788.1, C_AA048787.1, C_AA048786.1, C_AA048785.1. Furthermore, the raw sequencing data are under BioProject PRJNA1010524 on NCBI (https://www.ncbi.nlm.nih.gov/), encompassing PacBio HiFi sequencing FASTQ files for reference strains, Illumina sequencing FASTQ files for population strains (see details in Table S1), and FASTQ files for low-input RNA-seq data. The Nanopore sequencing data for C. inflata (RL4B) and C. steinii (RZ4A) can be accessed through NCBI BioProject study numbers SRR23588191 and SRR23588197.

ADDITIONAL FILES

The following material is available online.

Supplemental Material

Supplemental Figures (mSystems01161-23-s0001.pdf). Figures S1-S6.
Supplemental text (mSystems01161-23-s0002.pdf). Details of the materials and methods.
Supplemental Tables (mSystems01161-23-s0003.pdf). Tables S1 and S2.

REFERENCES

1. Suyama Y, Miura K. 1968. Size and structural variations of mitochondrial DNA. Proc Natl Acad Sci U S A 60:235–242. https://doi.org/10.1073/pnas.60.1.235
2. May-Panloup P, Chretien MF, Malthiery Y, Reynier P. 2007. Mitochondrial DNA in the oocyte and the developing embryo. Curr Top Dev Biol 77:51–83. https://doi.org/10.1016/S0070-2153(06)77003-X
3. Kersten B, Rellstab C, Schroeder H, Brodbeck S, Fladung M, Krutovsky KV, Gugerli F. 2022. The mitochondrial genome sequence of *Abies alba* Mill. reveals a high structural and combinatorial variation. BMC Genomics 23:776. https://doi.org/10.1186/s12864-022-08993-9
4. Zhang T, Li C, Zhang X, Wang C, Roger AJ, Gao F. 2021. Characterization and comparative analyses of mitochondrial genomes in single-celled eukaryotes to shed light on the diversity and evolution of linear molecular architecture. Int J Mol Sci 22:2546. https://doi.org/10.3390/ijms22052546
5. Mikhaylova AG, Mikhailova AA, Ushakova K, Tretiakov EO, Shamansky V, Yurchenko A, Zazhytska M, Zdobnov E, Makeev V, Yurov V, Tanaka M, Gostimskaya I, Fleischmann Z, Annis S, Franco M, Wasko K, Kunz WS, Knorre DA, Mazunin I, Nikolaev S, Fellay J, Reymond A, Khrapko K, Gunbin K, Popadin K. 2019. Mammalian mitochondrial mutational spectrum as a hallmark of cellular and organismal aging. bioRxiv. https://doi.org/10.1101/589168
6. Zhang HL, Liu BB, Wang XY, Han ZP, Zhang DX, Su CN. 2016. Comparative mitogenomic analysis of species representing six subfamilies in the

family Tenebrionidae. Int J Mol Sci 17:841. https://doi.org/10.3390/ijms17060841

7. Gray MW, Burger G, Lang BF. 1999. Mitochondrial evolution. Science 283:1476–1481. https://doi.org/10.1126/science.283.5407.1476

8. Ma J, Li XQ. 2015. Organellar genome copy number variation and integrity during moderate maturation of roots and leaves of maize seedlings. Curr Genet 61:591–600. https://doi.org/10.1007/s00294-015-0482-1

9. Leuthner TC, Hartman JH, Ryde IT, Meyer JN. 2021. PCR-based determination of mitochondrial DNA copy number in multiple species. Methods Mol Biol 2310:91–111. https://doi.org/10.1007/978-1-0716-1433-4_8

10. Ekstrand MI, Falkenberg M, Rantanen A, Park CB, Gaspari M, Hultenby K, Rustin P, Gustafsson CM, Larsson NG. 2004. Mitochondrial transcription factor A regulates mtDNA copy number in mammals. Hum Mol Genet 13:935–944. https://doi.org/10.1093/hmg/ddh109

11. Wang Y, Jiang Y, Liu Y, Li Y, Katz LA, Gao F, Yan Y. 2020. Comparative studies on the polymorphism and copy number variation of mtSSU rDNA in ciliates (Protista, Ciliophora): implications for phylogenetic, environmental, and ecological research. Microorganisms 8:316. https://doi.org/10.3390/microorganisms8030316

12. Neiman M, Taylor DR. 2009. The causes of mutation accumulation in mitochondrial genomes. Proc Biol Sci 276:1201–1209. https://doi.org/10.1098/rspb.2008.1758

13. Lynch M, Blanchard JL. 1998. Deleterious mutation accumulation in organelle genomes. Genetica 102–103:29–39. https://doi.org/10.1023/A:1017022522486

14. Dan ZC, Guan DL, Jiang T, Wang H, Zhao L, Xu SQ. 2022. Evolution of gene arrangements in the mitogenomes of Ensifera and characterization of the complete mitogenome of *Schizodactylus jimo*. Int J Mol Sci 23:12094. https://doi.org/10.3390/ijms232012094

15. Dowton M, Cameron SL, Dowavic JI, Austin AD, Whiting MF. 2009. Characterization of 67 mitochondrial tRNA gene rearrangements in the Hymenoptera suggests that mitochondrial tRNA gene position is selectively neutral. Mol Biol Evol 26:1607–1617. https://doi.org/10.1093/molbev/msp072

16. Magadum S, Banerjee U, Murugan P, Gangapur D, Ravikesavan R. 2013. Gene duplication as a major force in evolution. J Genet 92:155–161. https://doi.org/10.1007/s12041-013-0212-8

17. Sato M, Sato K. 2013. Maternal inheritance of mitochondrial DNA by diverse mechanisms to eliminate paternal mitochondrial DNA. Biochim Biophys Acta 1833:1979–1984. https://doi.org/10.1016/j.bbamcr.2013.03.010

18. Konrad A, Thompson O, Waterston RH, Moerman DG, Keightley PD, Bergthorsson U, Katju V. 2017. Mitochondrial mutation rate, spectrum and heteroplasmy in *Caenorhabditis elegans* spontaneous mutation accumulation lines of differing population size. Mol Biol Evol 34:1319–1334. https://doi.org/10.1093/molbev/msx051

19. Ye Z, Zhao C, Raborn RT, Lin M, Wei W, Hao Y, Lynch M. 2022. Genetic diversity, heteroplasmy, and recombination in mitochondrial genomes of *Daphnia pulex*, *Daphnia pulicaria*, and *Daphnia obtusa*. Mol Biol Evol 39:msac059. https://doi.org/10.1093/molbev/msac059

20. Johri P, Marinov GK, Doak TG, Lynch M. 2019. Population genetics of *Paramecium* mitochondrial genomes: recombination, mutation spectrum, and efficacy of selection. Genome Biol Evol 11:1398–1416. https://doi.org/10.1093/gbe/evz081

21. Di Giuseppe G, Barbieri M, Vallesi A, Luporini P, Dini F. 2013. Phylogeographical pattern of *Euplotes nobilii*, a protist ciliate with a bipolar biogeographical distribution. Mol Ecol 22:4029–4037. https://doi.org/10.1111/mec.12363

22. Weisse T, Montagnes DJS. 2022. Ecology of planktonic ciliates in a changing world: concepts, methods, and challenges. J Eukaryot Microbiol 69:e12879. https://doi.org/10.1111/jeu.12879

23. Huang YX, Wang S, Gao YQ, Chen JH, Wang XL, Li RJ. 2021. Comparison of mitochondrial genome and development of specific PCR primers for identifying two scuticociliates, *Pseudocohnilembus persalinus* and *Uronema marinum*. Parasit Vectors 14:318. https://doi.org/10.1186/s13071-021-04821-3

24. Li B, Song Y, Hao T, Wang L, Zheng W, Lyu Z, Chen Y, Pan X. 2022. Insights into the phylogeny of the ciliate of class Colpodea based on multigene data. Ecol Evol 12:e9380. https://doi.org/10.1002/ece3.9380

25. Dunthorn M, Hall MS, Foissner W, Stoeck T, Katz LA. 2014. Broad taxon sampling of ciliates using mitochondrial small subunit ribosomal DNA. Acta Protozool 53:207–213. https://doi.org/10.4467/16890027AP.14.017.1598

26. Gao F, Warren A, Zhang Q, Gong J, Miao M, Sun P, Xu D, Huang J, Yi Z, Song W. 2016. The all-data-based evolutionary hypothesis of ciliated protists with a revised classification of the phylum Ciliophora (Eukaryota, Alveolata). Sci Rep 6:24874. https://doi.org/10.1038/srep24874

27. Wang Z, Zhang J, Xu X, Witt C, Deng Y, Chen G, Meng G, Feng S, Xu L, Szekely T, Zhang G, Zhou Q. 2022. Phylogeny and sex chromosome evolution of Palaeognathae. J Genet Genomics 49:109–119. https://doi.org/10.1016/j.jgg.2021.06.013

28. Møller IM, Rasmusson AG, Van Aken O. 2021. Plant mitochondria - past, present and future. Plant J 108:912–959. https://doi.org/10.1111/tpj.15495

29. Babayev E, Seli E. 2015. Oocyte mitochondrial function and reproduction. Curr Opin Obstet Gynecol 27:175–181. https://doi.org/10.1097/GCO.0000000000000164

30. Soltész B, Pös O, Wlachovska Z, Budis J, Hekel R, Strieskova L, Liptak JB, Krampl W, Styk J, Németh N, Keserű JS, Jenei A, Buglyó G, Klekner Á, Nagy B, Szemes T. 2022. Mitochondrial DNA copy number changes, heteroplasmy, and mutations in plasma-derived exosomes and brain tissue of glioblastoma patients. Mol Cell Probes 66:101875. https://doi.org/10.1016/j.mcp.2022.101875

31. Vani S, Balasubramanyam D, Karthickeyan SMK, Parthiban M, Sesh PSL. 2022. Elucidation of genetic divergence among cattle breeds of Tamil Nadu in mitochondrial genome. Indian J Anim Sci 56:1448–1453. https://doi.org/10.18805/IJAR.B-4974

32. Liao N, Hu Z, Miao J, Hu X, Lyu X, Fang H, Zhou YM, Mahmoud A, Deng G, Meng YQ, Zhang K, Ma YY, Xia Y, Zhao M, Yang H, Zhao Y, Kang L, Wang Y, Yang JH, Zhou YH, Zhang MF, Yu JQ. 2022. Chromosome-level genome assembly of bunching onion illuminates genome evolution and flavor formation in *Allium* crops. Nat Commun 13:6690. https://doi.org/10.1038/s41467-022-34491-3

33. Morin GB, Cech TR. 1986. The telomeres of the linear mitochondrial DNA of *Tetrahymena thermophila* consist of 53 bp tandem repeats. Cell 46:873–883. https://doi.org/10.1016/0092-8674(86)90069-3

34. Pritchard AE, Cummings DJ. 1981. Replication of linear mitochondrial DNA from *Paramecium*: sequence and structure of the initiation-end crosslink. Proc Natl Acad Sci U S A 78:7341–7345. https://doi.org/10.1073/pnas.78.12.7341

35. Watt A, Haites R, Young N, Billman-Jacobe H. 2019. The mitochondrial genome of *Tetrahymena rostrata*. Mitochondrial DNA B Resour 5:53–54. https://doi.org/10.1080/23802359.2019.1693919

36. Swart EC, Nowacki M, Shum J, Stiles H, Higgins BP, Doak TG, Schotanus K, Magrini VJ, Minx P, Mardis ER, Landweber LF. 2012. The *Oxytricha trifallax* mitochondrial genome. Genome Biol Evol 4:136–154. https://doi.org/10.1093/gbe/evr136

37. Park MH, Min GS. 2019. The complete mitochondrial genome of *Gruberia lanceolata* (Gruber, 1884) Kahl, 1932 (Ciliophora: Heterotrichea). Mitochondrial DNA B Resour 4:3443–3445. https://doi.org/10.1080/23802359.2019.1674199

38. de Graaf RM, Ricard G, van Alen TA, Duarte I, Dutilh BE, Burgtorf C, Kuiper JWP, van der Staay GWM, Tielens AGM, Huynen MA, Hackstein JHP. 2011. The organellar genome and metabolic potential of the hydrogen-producing mitochondrion of *Nyctotherus ovalis*. Mol Biol Evol 28:2379–2391. https://doi.org/10.1093/molbev/msr059

39. Omar A, Choi JH, Jung J-H. 2023. New record of ten ciliate species (Protozoa, Ciliophora) from South Korea. J Species Res 12:95–108. https://doi.org/10.12651/JSR.2023.12.1.095

40. Moradian MM, Beglaryan D, Skozylas JM, Kerikorian V. 2007. Complete mitochondrial genome sequence of three *Tetrahymena* species reveals mutation hot spots and accelerated nonsynonymous substitutions in *Ymf* genes. PLoS One 2:e650. https://doi.org/10.1371/journal.pone.0000650

41. Burki F, Roger AJ, Brown MW, Simpson AGB. 2020. The new tree of eukaryotes. Trends Ecol Evol 35:43–55. https://doi.org/10.1016/j.tree.2019.08.008

42. Campello-Nunes PH, Woelfl S, da Silva-Neto ID, da S Paiva T, Fernández LD. 2022. Checklist, diversity and biogeography of ciliates (Ciliophora)

from Chile. Eur J Protistol 84:125892. https://doi.org/10.1016/j.ejop.2022.125892

43. Dale B. 2023. Paleontological evidence for dinoflagellates and ciliates as early eukaryotes. J Mar Sci Eng 11:533. https://doi.org/10.3390/jmse11030533

44. Wright A-D, Lynn DH. 1997. Maximum ages of ciliate lineages estimated using a small subunit rRNA molecular clock: crown eukaryotes date back to the Paleoproterozoic. Arch Protistenkd 148:329–341. https://doi.org/10.1016/S0003-9365(97)80013-9

45. Foissner W, Stoeck T, Agatha S, Dunthorn M. 2011. Intraclass evolution and classification of the Colpodea (Ciliophora). J Eukaryot Microbiol 58:397–415. https://doi.org/10.1111/j.1550-7408.2011.00566.x

46. Matsuoka T. 2021. Early signaling pathways mediating dormant cyst formation in terrestrial unicellular eukaryote *Colpoda*. FEMS Microbiol Lett 368:fnab019. https://doi.org/10.1093/femsle/fnab019

47. Fierer N. 2017. Embracing the unknown: disentangling the complexities of the soil microbiome. Nat Rev Microbiol 15:579–590. https://doi.org/10.1038/nrmicro.2017.87

48. Darbyshire JF. 1972. Nitrogen fixation by *Azotobacter chroococcum* in the presence of *Colpoda steini*—II. The influence of agitation. Soil Biol Biochem 4:371–376. https://doi.org/10.1016/0038-0717(72)90033-8

49. Zhang W, Lin Q, Li G, Zhao X. 2022. The ciliate protozoan *Colpoda cucullus* can improve maize growth by transporting soil phosphates. J Integr Agric 21:855–861. https://doi.org/10.1016/S2095-3119(21)63628-6

50. Dunthorn M, Foissner W, Katz LA. 2011. Expanding character sampling for ciliate phylogenetic inference using mitochondrial SSU-rDNA as a molecular marker. Protist 162:85–99. https://doi.org/10.1016/j.protis.2010.06.003

51. Burger G, Zhu Y, Littlejohn TG, Greenwood SJ, Schnare MN, Lang BF, Gray MW. 2000. Complete sequence of the mitochondrial genome of *Tetrahymena pyriformis* and comparison with *Paramecium aurelia* mitochondrial DNA. J Mol Biol 297:365–380. https://doi.org/10.1006/jmbi.2000.3529

52. Chen T, Zhang H, Liu Y, Liu Y-X, Huang L. 2021. EVenn: easy to create repeatable and editable Venn diagrams and Venn networks online. J Genet Genomics 48:863–866. https://doi.org/10.1016/j.jgg.2021.07.007

53. Xiong J, Yang W, Chen K, Jiang C, Ma Y, Chai X, Yan G, Wang G, Yuan D, Liu Y, Bidwell SL, Zafar N, Hadjithomas M, Krishnakumar V, Coyne RS, Orias E, Miao W, Malik HS. 2019. Hidden genomic evolution in a morphospecies-the landscape of rapidly evolving genes in *Tetrahymena*. PLoS Biol 17:e3000294. https://doi.org/10.1371/journal.pbio.3000294

54. Raymond A, Kelley PH, Lutken CB. 1990. Dead by degrees: articulate brachiopods, paleoclimate and the mid-carboniferous extinction event. PALAIOS 5:111. https://doi.org/10.2307/3514808

55. Brunk CF, Lee LC, Tran AB, Li J. 2003. Complete sequence of the mitochondrial genome of *Tetrahymena thermophila* and comparative

56. Zhang J, Kan X, Miao G, Hu S, Sun Q, Tian W. 2020. qMGR: a new approach for quantifying mitochondrial genome rearrangement. Mitochondrion 52:20–23. https://doi.org/10.1016/j.mito.2020.02.004

57. Groth C, Petersen RF, Piskur J. 2000. Diversity in organization and the origin of gene orders in the mitochondrial DNA molecules of the genus *Saccharomyces*. Mol Biol Evol 17:1833–1841. https://doi.org/10.1093/oxfordjournals.molbev.a026284

58. Zhang B, Havird JC, Wang E, Lv J, Xu X. 2021. Massive gene rearrangement in mitogenomes of *Phytoseiid mites*. Int J Biol Macromol 186:33–39. https://doi.org/10.1016/j.ijbiomac.2021.07.011

59. Kim NY, Ahn SJ, Seo JS, Jeon EJ, Cho MY, Choi HS. 2022. Characterization of the complete mitochondrial genome of *Miamiensis avidus* causing flatfish scuticociliatosis. Genetica 150:407–420. https://doi.org/10.1007/s10709-022-00167-5

60. Shay JW, Wright WE. 2019. Telomeres and telomerase: three decades of progress. Nat Rev Genet 20:299–309. https://doi.org/10.1038/s41576-019-0099-1

61. Gout JF, Hao Y, Johri P, Arnaiz O, Doak TG, Bhullar S, Couloux A, Guérin F, Malinsky S, Potekhin A, Sawka N, Sperling L, Labadie K, Meyer E, Duharcourt S, Lynch M. 2023. Dynamics of gene loss following ancient whole-genome duplication in the cryptic *Paramecium* complex. Mol Biol Evol 40:msad107. https://doi.org/10.1093/molbev/msad107

62. Boore JL, Brown WM. 1998. Big trees from little genomes: mitochondrial gene order as a phylogenetic tool. Curr Opin Genet Dev 8:668–674. https://doi.org/10.1016/s0959-437x(98)80035-x

63. Montoya J, Gaines GL, Attardi G. 1983. The pattern of transcription of the human mitochondrial rRNA genes reveals two overlapping transcription units. Cell 34:151–159. https://doi.org/10.1016/0092-8674(83)90145-9

64. Shtolz N, Mishmar D. 2023. The metazoan landscape of mitochondrial DNA gene order and content is shaped by selection and affects mitochondrial transcription. Commun Biol 6:93. https://doi.org/10.1038/s42003-023-04471-4

65. Sayers EW, Bolton EE, Brister JR, Canese K, Chan J, Comeau DC, Connor R, Funk K, Kelly C, Kim S, Madej T, Marchler-Bauer A, Lanczycki C, Lathrop S, Lu Z, Thibaud-Nissen F, Murphy T, Phan L, Skripchenko Y, Tse T, Wang J, Williams R, Trawick BW, Pruitt KD, Sherry ST. 2022. Database resources of the national center for biotechnology information. Nucleic Acids Res 50:D20–D26. https://doi.org/10.1093/nar/gkab1112

66. Nishimura Y, Tanifuji G, Kamikawa R, Yabuki A, Hashimoto T, Inagaki Y. 2016. Mitochondrial genome of *Palpitomonas bilix*: derived genome structure and ancestral system for cytochrome c maturation. Genome Biol Evol 8:3090–3098. https://doi.org/10.1093/gbe/evw217

methods for identifying highly divergent genes. Nucleic Acids Res 31:1673–1682. https://doi.org/10.1093/nar/gkg270