Integrative analyses on the ciliates *Colpoda* illuminate the life history evolution of soil microorganisms

Haichao Li,[1,2] Kun Wu,[1] Yuan Feng,[1] Chao Gao,[1] Yaohai Wang,[1] Yuanyuan Zhang,[1] Jiao Pan,[1] Xiaopeng Shen,[3] Rebecca A. Zufall,[4] Yu Zhang,[5] Weipeng Zhang,[1] Jin Sun,[1] Zhiqiang Ye,[6] Weiyi Li,[7] Michael Lynch,[8] Hongan Long[1,2]

AUTHOR AFFILIATIONS See affiliation list on p. 14.

ABSTRACT Microorganisms play a central role in sustaining soil ecosystems and agriculture, and these functions are usually associated with their complex life history. Yet, the regulation and evolution of life history have remained enigmatic and poorly understood, especially in protozoa, the third most abundant group of organisms in the soil. Here, we explore the life history of a cosmopolitan species—*Colpoda steinii*. Our analysis has yielded a high-quality macronuclear genome for *C. steinii*, with size of 155 Mbp and 37,123 protein-coding genes, as well as mean intron length of ~93 bp, longer than most other studied ciliates. Notably, we identify two possible whole-genome duplication events in *C. steinii*, which may account for its genome being about twice the size of *C. inflata's*, another co-existing species. We further resolve the gene expression profiles in diverse life stages of *C. steinii*, which are also corroborated in *C. inflata*. During the resting cyst stage, genes associated with cell death and vacuole formation are upregulated, and translation-related genes are downregulated. While the translation-related genes are upregulated during the excystment of resting cysts. Reproductive cysts exhibit a significant reduction in cell adhesion. We also demonstrate that most genes expressed in specific life stages are under strong purifying selection. This study offers a deeper understanding of the life history evolution that underpins the extraordinary success and ecological functions of microorganisms in soil ecosystems.

IMPORTANCE *Colpoda* species, as a prominent group among the most widely distributed and abundant soil microorganisms, play a crucial role in sustaining soil ecosystems and promoting plant growth. This investigation reveals their exceptional macronuclear genomic features, including significantly large genome size, long introns, and numerous gene duplications. The gene expression profiles and the specific biological functions associated with the transitions between various life stages are also elucidated. The vast majority of genes linked to life stage transitions are subject to strong purifying selection, as inferred from multiple natural strains newly isolated and deeply sequenced. This substantiates the enduring and conservative nature of *Colpoda*'s life history, which has persisted throughout the extensive evolutionary history of these highly successful protozoa in soil. These findings shed light on the evolutionary dynamics of microbial eukaryotes in the ever-fluctuating soil environments. This integrative research represents a significant advancement in understanding the life histories of these understudied single-celled eukaryotes.

KEYWORDS life history, microbial evolution, protists, population genomics, functional genomics

The life history of an organism refers to its survival and reproductive patterns, including traits that directly impact survival and the timing or quantity of reproduction (1). The optimization of trade-offs linked to growth, reproduction, and survival

through natural selection has shaped life history traits. Consequently, organisms sharing the same phylogeny tend to have similar traits (2). The life history strategies of soil microorganisms are highly diverse, endowing them with the exceptional capacity to thrive in a variety of environments, such as terrestrial soils, freshwater, and oceans. However, despite these, the associated regulatory mechanisms and evolutionary processes have received little research attention. *Colpoda* organismsare ciliates prevalent in diverse habitats worldwide, with a particularly notable presence in soil ecosystems. *Colpoda* species share a common feature typical of all ciliates—the presence of a dimorphic nuclear apparatus, comprising both micronucleus and macronucleus sequestered in the same cytoplasm. To date, a total of 32 nominal species of *Colpoda* have been documented, and their cosmopolitan distribution has been empirically substantiated through an extensive body of global investigations (3–6). Moreover, they show highly similar life histories, characterized by distinct life stages, including trophonts—cells in a state of vegetative growth when food bacteria are abundant, reproductive cysts—a mode of asexual reproduction within a cyst wall that distinguishes them from most other ciliates (7), and resting cysts—a dormant stage that helps endure harsh and highly fluctuating soil environments (7, 8). Importantly, these resting cysts are transformable with trophonts. The encystment process is widely acknowledged as a key mechanism contributing to the near-global distribution of *Colpoda* (8).

Colpoda species are well-known for their ease of collection and significant agricultural contributions, such as promoting the growth of crops like corn (9), cucumber (10), and rice (11). They also serve as valuable indicator species for monitoring soil quality and environmental pollution (12). The morphological features of *Colpoda* have been extensively documented in previous studies (13–15). For example, the transformation of trophonts into resting cysts includes oral absorption, shedding of cilia, filling of small vesicles within the cyst, and the formation of the resisting cyst wall (16, 17). Despite resting cysts are crucial in the life history of *Colpoda*, there are only a few studies focusing on their molecular mechanisms. For example, Sogame et al. (18, 19) proposed the involvement of elongation factor 1 (EF-1α) in inhibiting resting cyst formation. Furthermore, an increase in cAMP concentration, achieved through adenylate cyclase activation, has been shown to promote encystment (20). Additionally, Jiang et al. (21) applied transcriptomic analyses to investigate gene expression involved in the formation of resting cysts. Their findings revealed a reduction in biosynthesis and energy metabolism, coupled with a significant upregulation of autophagy during this critical stage. A systematic exploration encompassing various life stage transitions using omics tools remains an area yet to be investigated.

Cell divisions through the idiosyncratic formation of reproductive cysts in *Colpoda* (Fig. 1; Movie S1) have received limited attention, primarily investigated at the morphological level, dating back more than half a century (22). The molecular regulation of these cell divisions, albeit underexplored, may offer a valuable context for investigating non-canonical eukaryotic cell cycles involving amitosis and polyploidy (as the macronuclear chromosomes lack centromeres and are polyploid) and cellular differentiation. Similarly, while there are numerous studies on multiple whole-genome duplications, sexual processes involving micronucleus-macronucleus development and mating types, genomics, etc., in ciliates (23–26), corresponding investigations on *Colpoda* are mostly absent. It is worth noting that most of the previous efforts on *Colpoda* were conducted in the last century, and the current number of active researchers dedicated to studying *Colpoda* may be even fewer than the known *Colpoda* species.

Despite the rapid advancement of multi-omics technology, investigations into the life histories of ciliates have primarily focused on a handful of species. These studies have often relied on comparative transcriptomics but typically lacked high-quality *de novo* macronuclear genome assemblies. Such genome assemblies are essential for understanding the genetic basis and enhancing the precision and quantification of differential gene expression analyses, akin to reference-based transcriptome analyses (27–30). *Colpoda* organisms, despite their near-global distribution and vital ecological
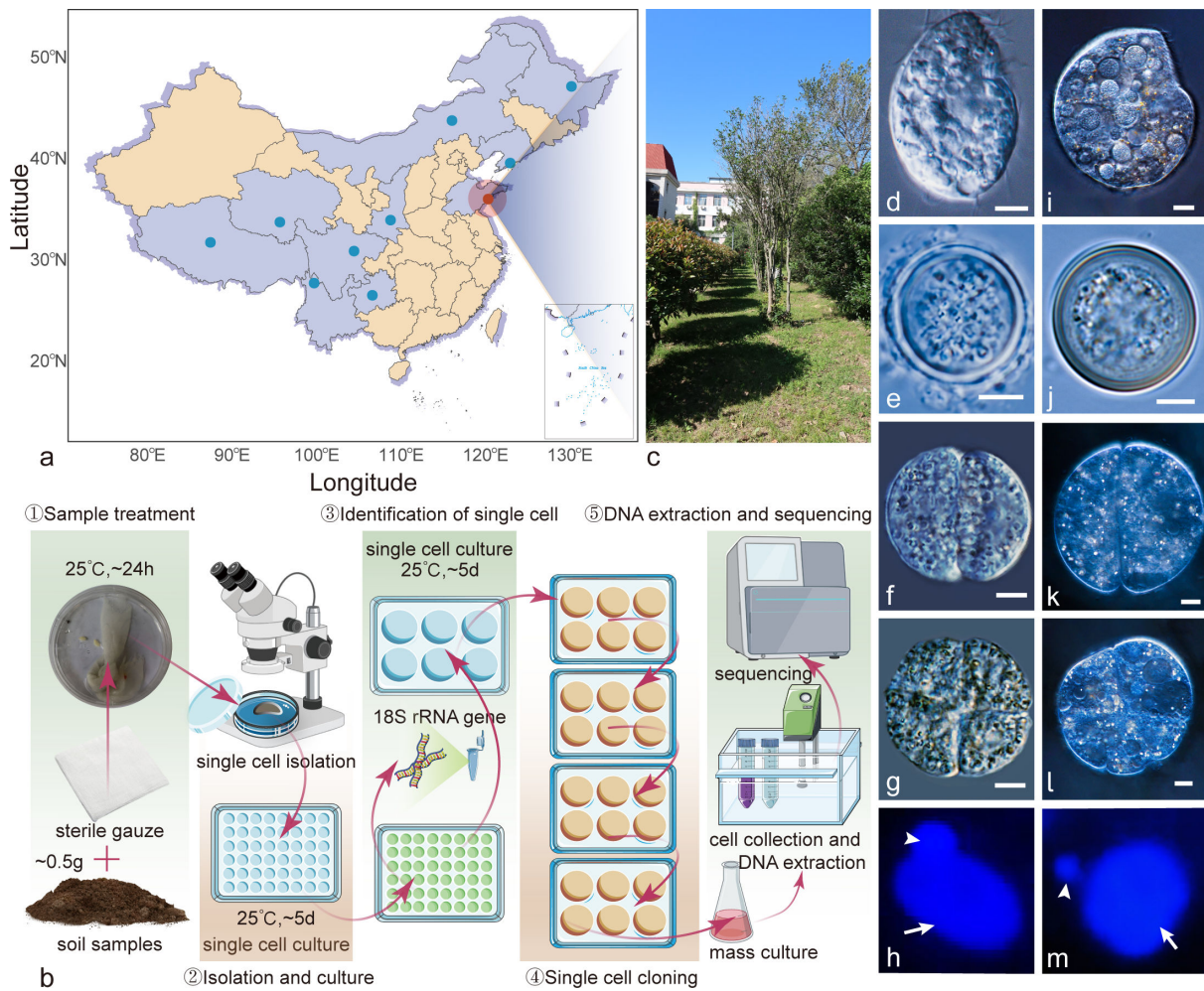
FIG 1 Life history, sample collection and preparation, and morphology of *Colpoda* spp. (a) Sampling sites of the 10 natural strains of *C. steinii*. Blue dots represent nine isolates used for population genomic analysis, and the orange dot shows the reference strain for the *de novo* assembly. (Map from https://datav.aliyun.com/portal/school/atlas/area_selector#&lat=30.332329214580188&lng=106.72278672066881&zoom=3.5.) (b) Strain isolation and preparation for genomics. (c) The sampling site of *C. steinii* and *C. inflata*. (d–g) Photomicrographs of *C. steinii* RZ4A at different life stages: d, trophonts in vegetative growth; e, resting cysts; f, g, reproductive cysts with two and four cells. (i–l) Photomicrographs of *C. inflata* RL4B at different life stages: i, trophonts; j, resting cysts; k, l, reproductive cysts. Scale bar represents 5 µm. (h and m) DAPI staining of the nuclei of *Colpoda*: h, *C. steinii* RZ4A; m, *C. inflata* RL4B. Arrow shows the macronucleus, and arrowhead marks the micronucleus.

roles, have remained among the understudied majority, in stark contrast to the genomic resources of model ciliates like *Tetrahymena* and *Paramecium* (23, 31–34).

In this study, we explored the life history of *Colpoda steinii* at both phenotypic and molecular levels. Our approach involved uncovering its molecular basis through *de novo* macronuclear assembly and gene annotation, using Nanopore long-read sequencing. We also revealed differential gene expression by performing low-input RNA library preparation across various life stages, conducted comparative genomics with the *de novo* assembled macronuclear genomes of another co-existing congener *C. inflata*, as well as published genomes of distantly-related ciliates, investigated population genomics using strains newly collected from diverse regions in China, and examined the proteomics of resting cysts. This study will provide new data to uncover the strategies and evolution of soil ciliates in resisting adverse environments. It also lays the necessary foundation for a deeper understanding of the role of protozoa in maintaining the functioning and long-term evolution of soil ecosystems.

RESULTS

De novo assembly and gene annotation of the macronuclear genome of *Colpoda steinii* RZ4A

We collected soil samples, isolated single *Colpoda* cells, and identified species by live microscopy and 18S rDNA Sanger sequencing (Fig. 1a through c). To ensure the purification of the cell lines, we executed five rounds of single-cell passaging (Fig. 1b). As a result, we successfully established a clonal cell line, denoted as *C. steinii* RZ4A, which has a typical life history akin to most previously reported *Colpoda* species.

Subsequently, we performed genome sequencing on this strain using Illumina Novaseq 6000 and Oxford Nanopore MinION platforms. Following the exclusion of reads stemming from food bacteria contamination and the removal of low-quality bases or adaptors, we generated a total of 23.89 Gbp Nanopore and 31.54 Gbp Illumina PE150 high-quality sequences (Table 1). We initiated the assembly process by first compiling the Nanopore clean reads into contigs, subsequently merging them into scaffolds. The draft genome was then polished with high-quality Illumina short reads. Based on the distribution of GC content within the scaffolds, we further filtered out scaffolds that potentially originated from bacteria, as well as contaminating scaffolds of micronucleus sources according to sequencing depth. To prevent the inadvertent exclusion of genuine genome components, we conducted a thorough examination of scaffolds that were flagged for removal (see details in Supplemental Materials and Methods).

Ultimately, our efforts yielded a high-quality macronuclear genome for *C. steinii* RZ4A, with 155.39 Mbp in length, N50 of 40,553 bp (most ciliates are known to have short chromosomes in the macronucleus due to chromosome fragmentation), a genome completeness of 90.05% according to the BUSCO assessment (Fig. S1). The genome assembly comprises 7,989 scaffolds, of which 2,717 (~34% of all scaffolds) possess 2 telomeres, while 3,838 (~48%) have 1 telomere (Table 1).

We conducted gene structural annotation through a combination of *de novo* gene prediction and RNAseq evidence. To capture the transcriptome encompassing as many genes as possible, we performed low-input RNAseq on three distinct life stages:

TABLE 1 Macronuclear genomic features of *C. steinii* RZ4A and *C. inflata* RL4B

Features	*C. steinii* RZ4A	*C. inflata* RL4B
Illumina PE150 sequences (Gbp)	31.54	34.67
Nanopore sequences (Gbp)	23.89	19.23
Genome size (bp)	155,392,961	81,751,086
Number of scaffolds	7,989	2,481
Number of scaffolds (≥1,000 bp)	7,961	2,480
Number of scaffolds (≥10,000 bp)	3,738	2,041
Number of scaffolds (≥50,000 bp)	710	385
Largest scaffolds	501,730	1,462,395
N50 (bp)	40,553	46,401
GC content (%)	33.55	35.09
Telomere	TT(T/G)GGG	TT(T/G)GGG
Scaffolds with telomere (%)	82.05	58.52
2-Telomeres scaffolds	2,717	307
1-Telomere scaffolds	3,838	1,145
BUSCO (%)	90.1	76.6
Number of genes	37,123	22,668
Mean gene size (bp)	2,034	1,876
CDS size (bp)	440	456
Number of introns	110,669	59,429
Genes with introns	22,652	12,430
Mean intron size (bp)	93.2	85.01
Mean no. of introns	2.98	2.62

trophonts, reproductive cysts, and resting cysts (Fig. 1d through g). This effort resulted in the annotation of 37,123 protein-coding genes, with a mean size of 2,034 bp, featuring 2.98 introns (mean intron size ~93 bp, much longer than those in other ciliates) and 3.98 exons per gene (Table 1).

In terms of gene functional annotation, an exceptional 97.06% of gene models show homologous matches against the NCBI NR database. In addition, 63.03% of genes can be successfully annotated in Gene Ontology (GO). These results collectively demonstrate the high quality of the functional annotation of the genome.

Transcription and translation profiles associated with the life history

To uncover the transcriptional profiles associated with the life stages of *C. steinii* RZ4A, we first isolated ~20 cells from each distinct life stage on ice with a high-power Olympus dissection microscope. Following isolation, we promptly proceeded with low-input RNA library construction. We conducted an intensive analysis of differential gene expressions across various life stages, using the low-input RNAseq data obtained from *C. steinii* RZ4A, including trophonts in vegetative growth, reproductive cysts, resting cysts, and revived trophonts, with each stage comprising a minimum of at least three replicates. To ensure the reliability of our RNAseq data sets, we performed rigorous quality control assessments by principal component and clustering analyses. The identification of significantly differentially expressed genes (DEGs) between two life stages was executed using DESeq2, with a defined threshold of |fold-change| > 2 and $P_{adj} < 0.05$.

Resting cysts vs trophonts

We compared resting cysts with trophonts, with their replicated samples clustered together (Fig. 2a). There were significant distinctions in overall gene expression between the two life stages (Fig. 2b), and we found that 69 genes were upregulated and 2,313 downregulated in resting cysts (Fig. 2c; Table S1). To gain deeper insights into the biological processes, molecular functions, and cell components involved in the life stage transition of *C. steinii* RZ4A, we performed GO enrichment analysis focusing on the DEGs between these life stages.

Consistent with previous observations indicating that resting cysts represent a dormant stage, we detected a significant downregulation in their major cellular activities, particularly genes associated with translation (Fig. 2d). As expected, resting cysts still maintain essential metabolic functions, as evidenced by the slight background expression observed in most genes. This was further corroborated by our proteome sequencing on resting cysts, which identified 1,483 peptide hits corresponding to genes associated with translation, protein folding, microtubule cytoskeleton organization, and other processes (Tables S1 and S2; Fig. S2). Notably, the abundance of LamG-domain-containing proteins was found to be the highest in resting cysts (Table S2). These proteins are usually Ca^{2+}-mediated receptors. We hypothesize that these proteins may be the primary receptors for the molecular process of the formation of resting cysts because the regulation of Ca^{2+} concentration can induce the formation of resting cysts in *Colpoda* (20).

Colpoda species are also known for their propensity to generate numerous small vesicles and vacuoles within the cytoplasm during the formation of resting cysts, a phenomenon well-documented. In the case of *C. steinii* RZ4A, we have observed that upregulated genes are enriched in lipid metabolism (GO:0006629) and the vacuole cellular component (GO:0005773) (Fig. 2d). According to previous studies, the extrusion of macronuclear chromatin is a distinctive feature in the process of encystment in a manner similar to the apoptosis-like nuclear death that occurs during conjugation in other ciliates (13). In line with this, we have also identified two additional upregulated genes (g20033.t1 and g7555.t1) enriched in cell death (GO:0008219; Fig. S3). Our results align nicely with the aforementioned conjectures suggesting that events similar to cell death may occur during encystment.
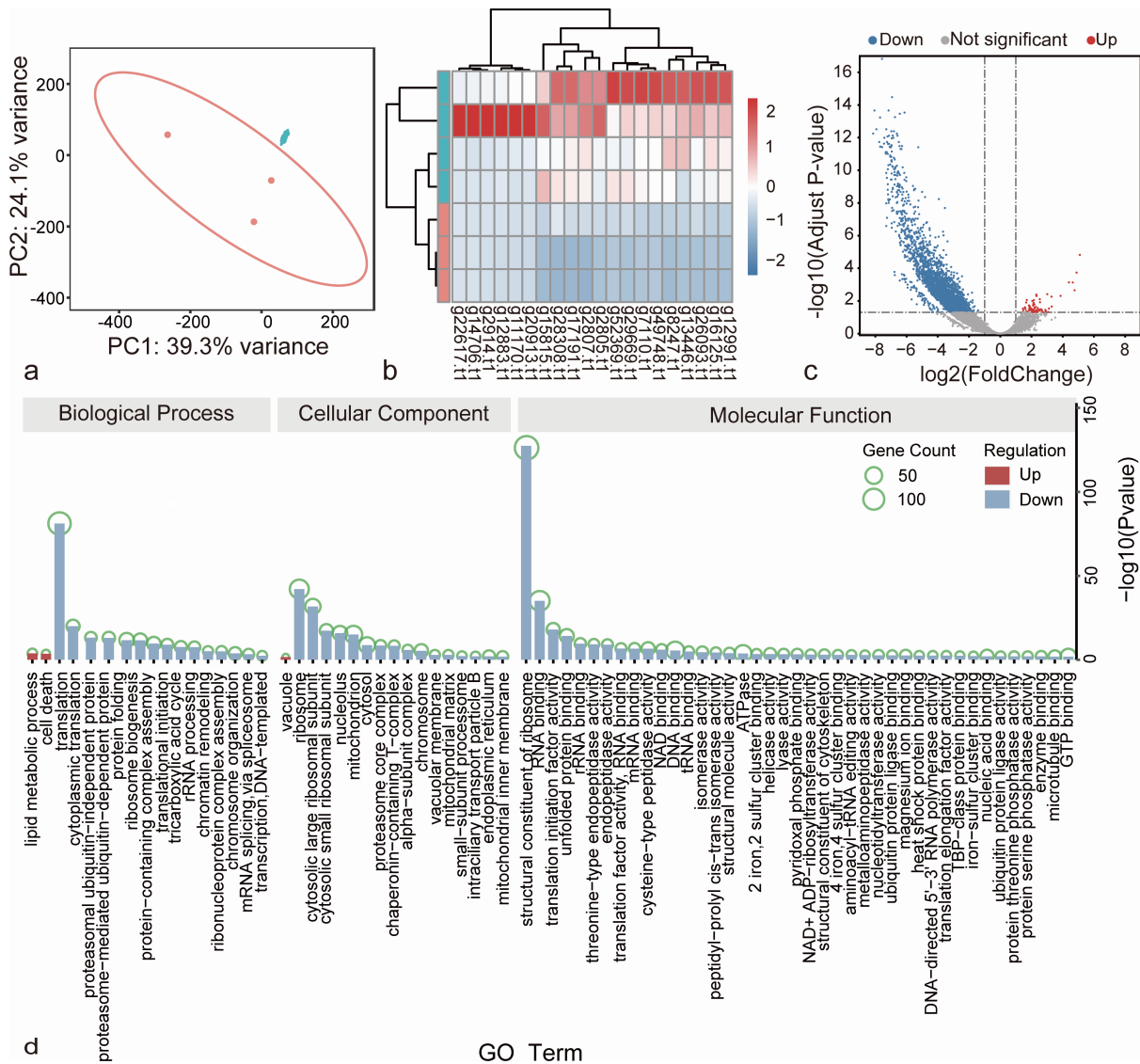
FIG 2 Differential gene expression and gene enrichment analyses of resting cysts vs trophonts in *C. steinii* RZ4A. (a) Principal component analysis of mRNAs of trophonts and resting cysts. Each dot represents one sample. Blue dots are resting cyst samples, and red ones are trophont samples. (b) The heatmap of top 20 differentially expressed genes in resting cysts (vs trophonts). The blue rectangles at the far left of the figure are resting cyst, and the red are trophont. (c) DEGs in resting cysts (vs trophonts). Gray, red, and blue dots represent not significantly, significantly up- and downregulated DEGs. (d) Results from GO enrichment analyses, using up- or downregulated DEGs.

These findings collectively reinforce the reliability of our differential gene expression analyses based on RNAseq. Furthermore, they provide valuable molecular insights into the process of encystment—a critical survival strategy used by *Colpoda* to contend with the challenging and ever-changing conditions of their soil habitats.

Revived trophonts vs resting cysts

When an ample supply of food bacteria becomes available, resting cysts undergo excystment and transition into trophonts. In our experimental setup, we initiated this process by re-introducing food bacteria to the resting cysts of *C. steinii* RZ4A. Remarkably, within a span of 12 h following this re-introduction, the resting cysts commenced excystment, giving rise to trophonts. The resting cysts before re-feeding and newly formed trophonts were promptly isolated and used for the construction of low-input RNAseq libraries, which were subsequently subject to Illumina PE150 sequencing.

Through an intensive analysis of differential gene expression, we identified a total of 2,803 upregulated and 2,486 downregulated genes in the revived trophonts when compared to the resting cysts (Fig. 3a; Table S3). We detected active expression of translation-related genes in the revived trophonts (Fig. 3e), and they are downregulated in the resting cysts (vs trophonts). By comparing the upregulated genes in the revived trophonts with those downregulated in the resting cysts, we found that more than one-third of the genes were shared between them, and these genes may play a crucial role in the mutual transformation between trophonts and resting cysts.

Reproductive cysts vs trophonts

Similar to other *Colpoda* congeners, *C. steinii* RZ4A reproduces by forming reproductive cysts, which typically give rise to two or four trophonts upon bursting (Fig. 1f and g; Movie S1). Using differential gene expression analyses, we have identified a total of 2,297 genes that are upregulated in reproductive cysts when compared to trophonts (Fig. 3b; Table S4). These genes are enriched in molecular functions commonly associated with eukaryotic cell division, encompassing processes such as mitosis, protein phosphorylation, MAPK cascade activation, lipid metabolism, and so on. Furthermore, they are associated with critical cell components including cilia, microtubules, spindles, centromeres, and others (Fig. 3f).

Among these molecular functions, those related to cilia are particularly prominent (Fig. 3f), indicating the heightened activity in the generation and movement of cilia of progeny cells enclosed within the reproductive cyst wall. By contrast, there is a noticeable reduction in genes associated with the transport of xenobiotic and communication with the extracellular environment (Fig. 3f). This observation further underscores that the cell-division process within the reproductive cysts is an independent process, reducing the interference of the external environment as much as possible to ensure high-fidelity cell divisions.

Technical validation of the DEGs analyses using RT-qPCR

To corroborate the accuracy of our differential gene expression analyses, we first conducted RT-qPCR experiments using newly prepared total RNA of trophonts and resting cysts as templates. Then, we did reverse transcription and qPCR, targeting six randomly chosen genes of *C. steinii* RZ4A (upregulated: g10370.t1, g17316.t1, g33357.t1; downregulated: g2805.t1, g2807.t1, g28308.t1; Table S5). These genes were chosen based on their observed up- or downregulation during the resting cyst stage compared to trophonts, as indicated by our earlier analyses using low-input RNAseq. Encouragingly, the expression patterns of all these genes were found to be consistent with those derived from the low-input RNAseq data sets (Fig. 3c and d). This alignment serves as compelling evidence affirming the reliability of the above differential gene expression analyses.

Validation of the gene ontology analyses

In addition, we made efforts to perform RNAi targeting the aforementioned six genes, using the L4440 plasmid system. However, we have not yet achieved success in these endeavors. Fortunately, given the high degree of conservation in the life histories of *Colpoda* species, it is reasonable to anticipate that the molecular mechanisms revealed through our GO analyses in *C. steinii* RZ4A are also likely to be present in congeners.

We then proceeded to isolate *C. inflata* RL4B from the same sampling site as *C. steinii* RZ4A (Fig. 1c, i through l). Following similar procedures, we first *de novo* assembled and annotated its macronuclear genome. The genome of *C. inflata* RL4B is approximately 81.75 Mbp in size, harboring a total of 22,668 genes, and shares identical telomere sequences [5′-TT (T/G)GGG-3′] with *C. steinii* RZ4A (Table 1). Significantly, there are notable differences in the scaffold count between the macronuclear genomes of the two species, with *C. inflata* RL4B featuring 2,481 scaffolds (with N50 ~46.40 kbp) compared to
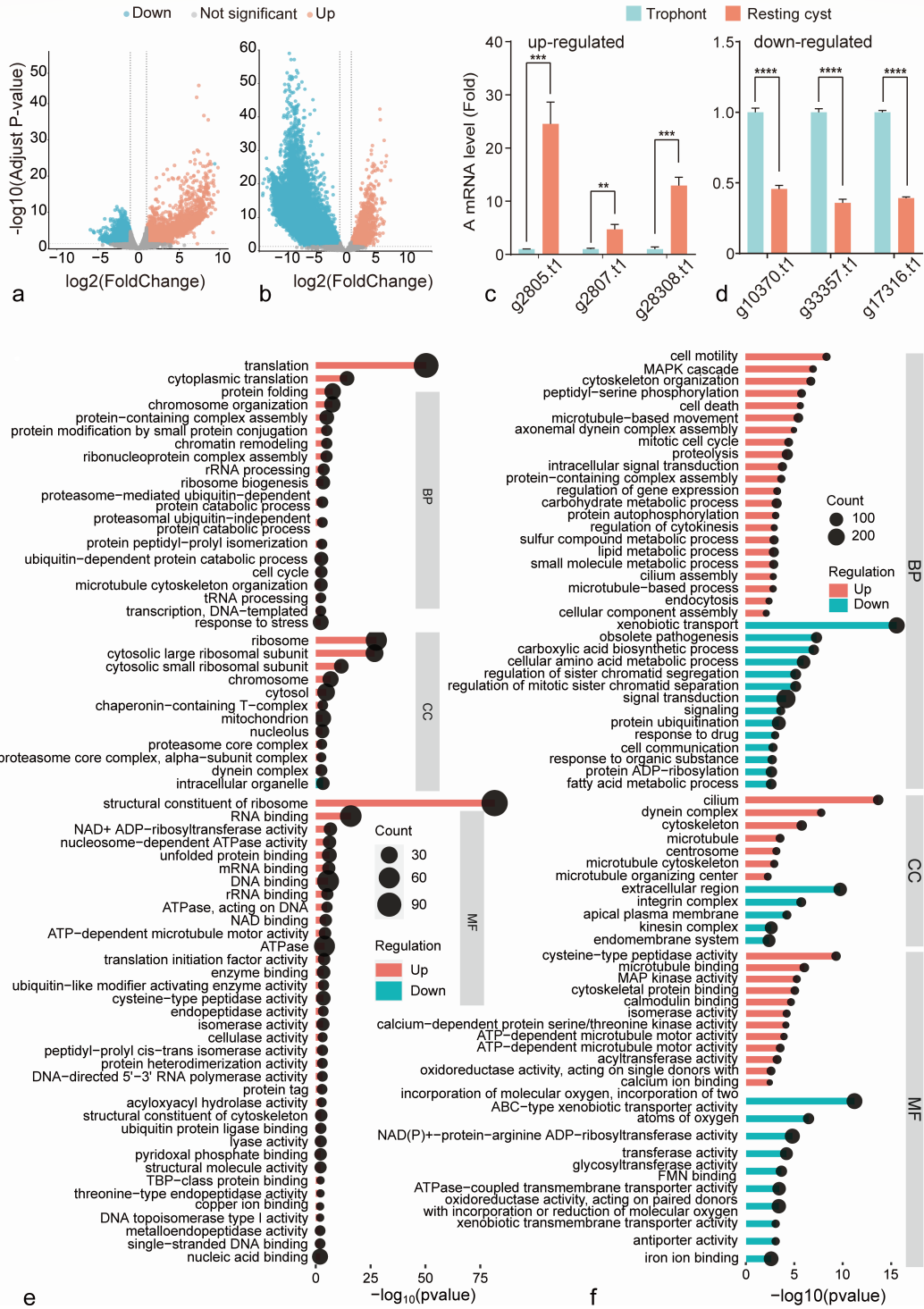
FIG 3 Differential gene expression of revived trophonts and reproductive cysts in *C. steinii* RZ4A. (a) Differential expression profiles of revived trophonts (vs resting cysts) and (b) reproductive cysts (vs trophonts in vegetative growth). Gray, red, and blue dots represent not significantly, significantly up- and downregulated DEGs, respectively. (c and d) RT-qPCR verification of DEGs in resting cysts (vs trophonts). **, ***, and **** represent *P* value < 0.01, 0.001, and 0.0001. (e) GO functional enrichment of DEGs of revived trophonts (vs resting cysts) and (f) reproductive cysts (vs trophonts). The size of the black dots represents the number of enriched genes in each pathway.

C. steinii RZ4A's 7,989 scaffolds (with N50 ~40.55 kbp). This discrepancy is consistent with the observation that the genome size of the latter (155.39 Mbp) is roughly twice that of

the former (81.75 Mbp) (Table 1). The increase in genome size may be driven by multiple processes, including whole-genome duplication (WGD), transposable elements (TE) proliferation, intron amplification, and tandem gene duplication (35). Using a synteny analysis of homologous gene pairs, we have discerned that two WGD events might have occurred in *C. steinii* RZ4A, whereas only one ancient WGD event in *C. inflata* RL4B (Fig. 4a). This fits nicely with the approximately twofold difference in genome size between the two species.

Using the genome of *C. inflata* RL4B as a reference and harnessing low-input RNAseq data from various life stages, we identified DEGs and conducted GO enrichment analyses. Impressively, numerous DEGs in *C. inflata* RL4B overlapped with those identified in *C. steinii* RZ4A and were enriched in the same molecular functions, biological processes, or cell components as those of *C. steinii* RZ4A. These included the upregulation of genes associated with cell death and vacuole formation, as well as the downregulation of translation in resting cysts (compared to trophonts in vegetative growth) (Table S6). Additionally, we observed significant enrichment of translation-related genes in revived trophonts (in contrast to resting cysts) (Table S7). Moreover, active biological processes, such as microtubule-based movement, cilium assembly, and cell motility, along with a noteworthy downregulation of cell adhesion regulated by the integrin complex, were evident in reproductive cysts (vs trophonts) (Table S8). These findings supported the reliability of our DEGs and GO analyses, as well as the robustness of gene expression patterns during the transformation between life stages.

Molecular evolution of DEGs associated with different life stages

To elucidate the evolutionary patterns underlying the above DEGs in the transcriptional profiles of different life stages, we first performed gene family expansion and contraction analyses. Since gene family expansion can facilitate adaptation, the expansion of gene families containing DEGs may serve as an indicator of complex life history conferring selective advantages for *Colpoda*. Using 14 previously published high-quality ciliate macronuclear genomes (including 2 *Colpoda* genomes from this study, as well as genomes from 4 *Tetrahymena* and 8 *Paramecium* species; Table S9), a total of 406,576 genes were clustered into 41,111 gene families. Of these, 102,897 genes were allocated into 3,584 shared gene families found in all species, while 1,876 gene families were exclusive to *Colpoda* (Fig. 4b). Our analyses also revealed a total of 19,711 gene
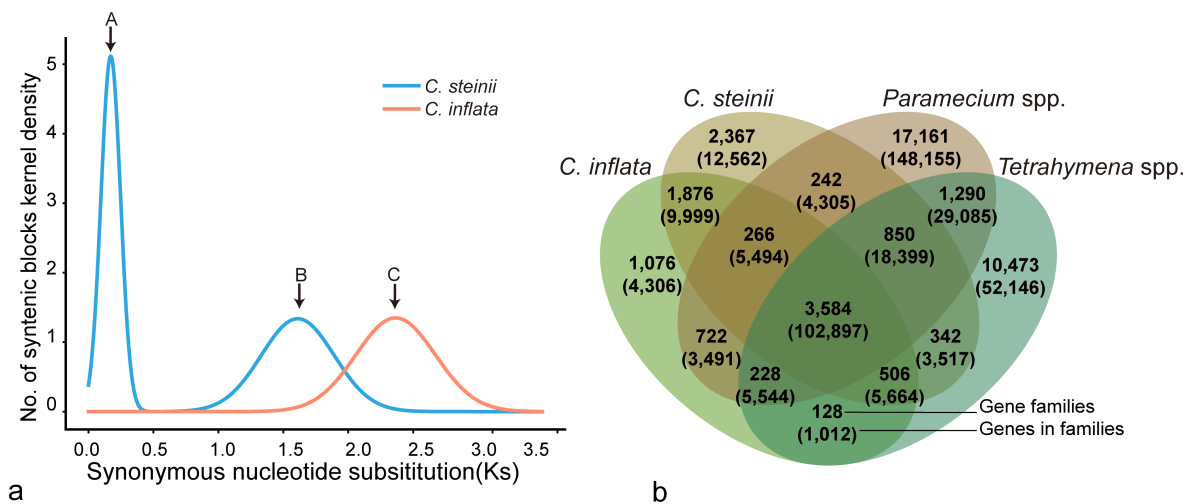
FIG 4 Genome evolution of *Colpoda*. (a) The Ks distribution of *C. steinii* RZ4A and *C. inflata* RL4B macronuclear genomes. Blue and orange represent *C. steinii* RZ4A and *C. inflata* RL4B information, respectively. After WGD occurred, paralogs doubled and accumulated mutations in parallel, which increased the synonymous mutations (Ks) and led to Ks peaks. Peaks A and B represent two WGD events in *C. steinii* RZ4A. Peak C refers to the only WGD event in *C. inflata* RL4B. (b) Homologous gene cluster analysis. Venn diagram illustrates the shared and unique gene families of *Colpoda*, compared with *Tetrahymena* and *Paramecium*.

duplication events in *C. steinii* RZ4A and 7,587 in *C. inflata* RL4B (Fig. S4), a finding that aligns consistently with the aforementioned WGD analyses.

Among the DEGs identified in resting cysts (in comparison to trophonts in vegetative growth) and in revived trophonts (vs resting cysts) of *C. steinii* RZ4A, 28.72% and 29.32% of them are in expanding gene families, while only 6.05% and 3.37% are in contracting ones (Tables S10 and S11). Significantly, many of the DEGs implicated in encystment and excystment processes are found within these expanding gene families, implying the possibility of these genes conferring selective advantages.

Similarly, for the DEGs identified in reproductive cysts (vs trophonts) of *C. steinii* RZ4A, a substantial 26.67% are associated with expanding gene families, with only 0.99% in contracting ones. It is noteworthy that the majority of these DEGs are involved in the cell adhesion processes (Table S12), suggesting that this molecular function has also been preserved by expanding gene families during the course of evolution.

The gene expansion and contraction analyses presented above provide evidence of DEGs that are subject to selection pressures. To further quantify the selective strength acting on the life-history-associated DEGs in *C. steinii*, we conducted Illumina-PE150 whole-genome sequencing on nine natural strains of *C. steinii* isolated from soils collected in China between 2020 and 2022 (Fig. 1a; Table S13). The clean Illumina reads were aligned to the genome of *C. steinii* RZ4A, and SNPs were called. In total, we identified 1,157,466 SNPs for *C. steinii*. These SNPs yielded nucleotide diversity at all genomic sites (π), nucleotide diversity at silent sites (π_s), and Tajima's D values of 0.04, 0.06, and 0.18, respectively.

Tajima's D test was used to distinguish between random neutral mutations and non-neutral mutations under selective pressure. A Tajima's D value greater than 0 suggests a low frequency of rare alleles, which may be indicative of equilibrium selection pressure or a population undergoing shrinkage/bottleneck. Because mRNA expression levels of the genes specifically expressed during one life stage, regardless of whether they are downregulated or upregulated, do not necessarily correlate with the regulation of associated biological functions. For instance, inhibitors' expression can be elevated but may result in downstream functional downregulation. We, thus, compiled the DEGs specifically expressed in each life stage as "regulation" DEGs and performed Tajima's D tests on them. Our analyses revealed that the mean Tajima's D for DEGs in reproductive cysts, resting cysts, and revived trophonts are 0.15, 0.03, and 0.09, respectively (Fig. 5a through c). We also used the McDonald-Kreitman (MK) test (a simple and widely used selection test), which is usually done by calculating the mean cross-gene neutrality index (NI) values, to summarize the selection patterns within *C. steinii* (36, 37). The majority of the NI values for the DEGs in reproductive cysts, resting cysts, and revived trophonts are larger than 1, and their distributions are highly similar (Fig. 5d through f). Taken the above together, purifying selection is, thus, dominant in genes associated with life history regulation. This further demonstrates the functional constraint on the complex life history with multiple life stages, as these stages are crucial for survival and reproduction in the highly challenging soil environments.

Possible sexual processes in *Colpoda*

The presence of sexual processes in *Colpoda* remains a subject of ongoing investigation, as there have been no direct and confirmed reports of conjugative nuclear changes in *Colpoda* to date, a fact also reflected in our own observations since 2018. In a study conducted by Dunthorn et al. (38), they explored 51 meiosis genes (11 meiosis-specific and 40 meiosis-related genes) in *Colpoda magna* and found that most of these genes align with those found in other ciliates known to undergo sexual processes.

In order to explore this, we followed the study of Dunthorn et al. (38) and reconstructed the database of these 51 meiosis genes, with a total of 391 homologs (Table S14). Subsequently, we conducted a thorough search for these meiosis genes across 16 ciliate macronuclear genomes, including the two *Colpoda* genomes assembled in this study. Our investigations revealed that a substantial portion of meiosis genes are present with
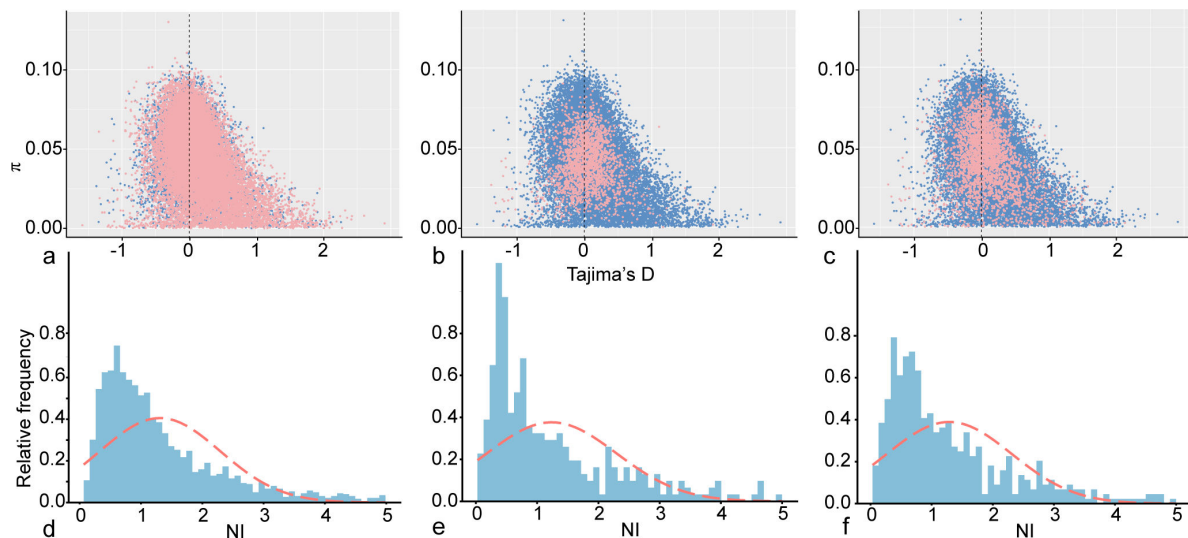
FIG 5 Selective pressures on DEGs of different life stages in *C. steinii* RZ4A. (a, b, c) π vs Tajima's *D* distribution of DEGs in reproductive cysts (vs trophonts, a), resting cysts (vs trophonts, b), and revived trophonts (vs resting cysts, c) of *C. steinii* RZ4A. The blue dots are all the genes in the genome, and the pink dots are DEGs (pink dots show up when they overlap with the blue ones). (d, e, f) The distribution of DEGs' NI values in reproductive cysts (d), resting cysts (e), and revived trophonts (f) of *C. steinii* RZ4A.

intact open reading frames in the macronuclear genome of *C. steinii* RZ4A, including all the 11 meiosis-specific genes that are also commonly found in most *Paramecium* species (indicated by underlined genes in Fig. 6). Among these genes are critical components such as *SPO11*, *DMC1*, *HOP2*, *MER3*, *MND1*, *MSH4*, and *MSH5*, all of which play pivotal roles in recombination processes in eukaryotes (39–42).

Recombination can reduce linkage equilibrium (LD). In order to investigate potential recombination events, we calculated linkage disequilibrium, measured as the squared correlation between allelic states (r^2). Our results revealed a rapid decline of LD within 1 Kbp region, and declining slowly with longer distance (Fig. S5), indicating possible recombination in *C. steinii*. We further evaluated the heterozygosity of the nine *C. steinii* natural strains by calculating method-of-moments *F* coefficient estimates, and found that their *F* values were between 0.56 and 0.87, indicating that homozygosity was more likely. This again raises the possibility of inbreeding or low mutation rate in *Colpoda*. Based on these findings, our results support the notion that all *Colpoda* congeners might undergo cryptic sexual processes, which may rarely occur or not manifest in the form of conjugation. Instead, it is possible that the cryptic sexual process is triggered by as-yet-unknown factors within the natural soil environment.

DISCUSSION

In this thorough study, our investigation spanned multiple domains of technology, including genomics (encompassing *de novo* macronuclear genome assembly and annotation—first effort using long-read sequencing for the Colpodea class), proteomics, and transcriptomics across four distinct life stages (namely, resting cysts, trophonts in vegetative growth, reproductive cysts, and revived trophonts from resting cysts) using low-input RNAseq, as well as comparative genomics and population-genomics analyses.

We successfully identified a substantial number of DEGs associated with various life stages and confirmed their occurrence during life history of the co-existing congener *C. inflata*; however, it is imperative to further determine their causal effects. Specifically, identifying the trigger genes responsible for transitioning between life stages remains to be done. Functional validation using efficient RNAi or gene-editing tools is the next logical step in our research trajectory despite our unsuccessful preliminary trials. The

FIG 6 Meiosis genes in 16 ciliate genomes. The red and blue dots represent the presence and absence of genes. The genes underlined are meiosis-specific genes. Species with "*" represent data from Dunthorn et al. (38).

resolution of these critical aspects holds the promise of providing deeper insights into the evolutionary mechanisms of cell differentiation in these unicellular eukaryotes.

As one of the most successful eukaryotic genera in soil, *Colpoda* has diverse life history strategies that are conserved across taxa (6). These strategies, including the extreme tolerance of resting cysts to harsh environmental conditions and fission

within reproductive cysts, likely contribute significantly to their broad distribution and abundant presence (3). Moreover, research has demonstrated that *Colpoda* species play a pivotal role in enhancing nutrient absorption by plants, while the presence of plants, in turn, increases the density of *Colpoda* (9, 43). This mutualistic relationship fosters both the robust growth of plants and the extensive proliferation of *Colpoda* populations.

In addition, the mean guanine-cytosine (GC) content of *Colpoda* genomes exceeds that of most other ciliates (Fig. S6). The high GC composition contributes to genome stability and resilience in extreme environments. Various mechanisms underpin this stability, including the presence of an extra hydrogen bond in G:C base pairs compared to A:T base pairs, and the fact that codons with higher GC encode more hydrophobic amino acids. These hydrophobic amino acids may enhance the stability and resistance to denaturation of proteins within *Colpoda*, reinforcing their adaptability to challenging conditions (44). Numerous predators and highly variable physicochemical factors such as moisture, pH, and temperature might exert much stronger natural selection pressure on *Colpoda* in soil habitats compared to freshwater environments where *Paramecium* and *Tetrahymena* live. These extra selection pressures might have likely contributed to the evolution of higher GC levels in *Colpoda* genomes (45). While *Colpoda* populations sharing the same ecological niches in soil may have undergone co-evolution, the specific interactions between them, including aspects such as competition and mutualism, remain largely unexplored. Future research in this domain promises to reveal additional strategies within *Colpoda*'s toolkit.

While *C. inflata* RL4B experienced a single whole-genome duplication (WGD) event, *C. steinii* RZ4A, by contrast, underwent two such events. This difference in WGD history likely contributes significantly to the observed doubling in genome size between the two *Colpoda* species. The molecular functions of both the expanded gene families and genes retained after WGD in *C. steinii* RZ4A are associated with ATP-binding, ion-binding, and protein-binding proteins (Fig. S4, S7a and b), which are the basic cell functions to maintain vital activities. We speculate that *Colpoda* require intensive energy supply due to the harsh habitats in soil, which may be related to ATP binding. Ion-binding function may be related to morphological changes during the formation of resting and reproductive cysts in *Colpoda*.

The macronuclear genome size of *C. steinii* RZ4A is larger than that of most other ciliates (Fig. S4). Genome size evolution is a multifaceted process influenced by factors such as natural selection, genetic drift, and mutations. These factors collectively contribute to significant changes in genome size (46). In the case of *C. steinii* RZ4A, its genome has experienced the most gene duplications among all the 14 ciliate genomes analyzed in this study, while this could not solely account for its large genome size. For example, in *Paramecium*, there are large genomes with few gene duplications or small genomes with numerous gene duplications (Fig. S4). These findings underscore the intricate nature of genome size evolution, wherein gene duplications can, indeed, lead to genome size expansion, but the roles of natural selection and genetic drift should not be overlooked. Further investigations are needed to disentangle the complex interplay of these genetic factors in shaping genome sizes.

Microbial eukaryotes, comprising a substantial portion of Earth's organisms, play active roles in the global biogeochemical processes that govern our soils. Their indispensability in the microbial loop, where they facilitate the transfer of materials and energy from bacteria to higher trophic levels, underscores their profound ecological significance. Despite ongoing research efforts, there remains an imperative need for further investigations to unravel the intricate survival strategies embedded in their life histories, alongside the genetic and evolutionary mechanisms that underpin their success. These diminutive yet exceptionally resilient and ancient denizens hold the key to unlocking the mysteries hidden beneath the Earth's surface at the molecular level, offering fresh insights that extend beyond the limitations of conventional methodologies. In summary, this research represents a significant advancement in the study of the life histories of these understudied single-celled eukaryotes. It does so by using

state-of-the-art "-omics" tools, coupled with extensive collections, isolations, microbiological manipulations, and advanced microscopy techniques. This study, thus, paves the way for a more profound understanding of the life history evolution of soil microorganisms.

MATERIALS AND METHODS

We collected two *Colpoda* species (*C. steinii* reference strain RZ4A and the *C. inflata* RL4B) from the Yushan campus of Ocean University of China (36°3′42″N, 120°19′56″E; Fig. 1c), Qingdao, China, on 22 August 2018. Their DNA and RNA were extracted using the MasterPure Complete DNA and RNA Purification kit. Subsequently, we used MinION (in-lab) and NovaSeq6000 (Berry Genomics, Inc., Beijing) sequencers for Nanopore and Illumina sequencing on the reference strains. We then assembled and annotated the macronuclear genomes of both species. In order to reveal transcription patterns of different life stages of each species, low-input RNAseq was performed, followed by differential gene expressions analyses. In addition, we also conducted comparative genomics and population genomics analyses to elucidate the evolutionary patterns of life-history-associated genes. Details can be found in the Supplemental Materials and Methods.

AUTHOR AFFILIATIONS

[1]Key Laboratory of Evolution and Marine Biodiversity (Ministry of Education), Institute of Evolution and Marine Biodiversity, Ocean University of China, Qingdao, Shandong Province, China
[2]Laboratory for Marine Biology and Biotechnology, Qingdao Marine Science and Technology Center, Qingdao, Shandong Province, China
[3]College of Life Sciences, Anhui Normal University, Wuhu, Anhui Province, China
[4]Department of Biology and Biochemistry, University of Houston, Houston, Texas, USA
[5]School of Mathematics Science, Ocean University of China, Qingdao, Shandong Province, China
[6]School of Life Sciences, Central China Normal University, Wuhan, Hubei Province, China
[7]Department of Genetics, Stanford University School of Medicine, Stanford, California, USA
[8]Biodesign Center for Mechanisms of Evolution, Arizona State University, Tempe, Arizona, USA

AUTHOR ORCIDs

Haichao Li 🔟 http://orcid.org/0000-0001-5782-2433

Yu Zhang [ORCID] http://orcid.org/0000-0001-9179-0922
Weipeng Zhang [ORCID] http://orcid.org/0000-0002-1231-8927
Hongan Long [ORCID] http://orcid.org/0000-0001-9767-8173

DATA AVAILABILITY

All FASTQ sequences are available at NCBI BioProject PRJNA937028. The Illumina reads and Nanopore reads of *C. steinii* RZ4A and *C. inflata* RL4B are available with the Biosample IDs of SAMN33386881 and SAMN33386882. The nine NCBI SRA samples (accession numbers SAMN36899794 to SAMN36899802) link FASTQ files containing sequences from natural strains for population analyses. The genome assemblies and annotations of *C. steinii* RZ4A and *C. inflata* RL4B are available at the CNCB (China National Center for Bioinformation) BioProject PRJCA018943 with Biosample IDs of SAMC2987422 and SAMC2989089.

ADDITIONAL FILES

The following material is available online.

Supplemental Material

Supplemental Figures (mSystems01379-23-s0001.pdf). Figures S1-S8.
Supplemental Text (mSystems01379-23-s0002.docx). Supplemental Materials and Methods.
Legend (mSystems01379-23-s0003.docx). Legend for Movie S1.
Supplemental Tables (mSystems01379-23-s0004.xlsx). Tables S1 and S2.
Additional Supplemental Tables (mSystems01379-23-s0005.xlsx). Tables S4 to S16.
Table S3 (mSystems01379-23-s0006.xlsx). Differential gene expression analysis.
Movie S1 (mSystems01379-23-s0007.mpeg). Division process of the reproductive cysts of *C. steinii*.

REFERENCES

1. Stearns SC. 1976. Life-history tactics: a review of the ideas. Q Rev Biol 51:3–47. https://doi.org/10.1086/409052

2. Petrik CM. 2019. Life history of marine fishes and their implications for the future oceans, p 165–172. In Cisneros-Montemayor AM, Cheung WWL, Ota Y (ed), Predicting future oceans. Elsevier.

3. Bowers N, Kroll TT, Pratt JR. 1998. Diversity and geographic distribution of riboprints from three cosmopolitan species of *Colpoda* Müller (Ciliophora: Colpodea). Eur J Protistol 34:341–347. https://doi.org/10.1016/S0932-4739(98)80001-1

4. Pérez-Uz B, Galfione VC, Ochoa-Hueso R, Martín-Cereceda M. 2023. Protist diversity responses to experimental N deposition in biological crusts of a semiarid mediterranean ecosystem. Protist 174:125929. https://doi.org/10.1016/j.protis.2022.125929

5. Thompson AR. 2021. Phagotrophic protists (protozoa) in Antarctic terrestrial ecosystems: diversity, distribution, ecology, and best research practices. Polar Biol 44:1467–1484. https://doi.org/10.1007/s00300-021-02896-3

6. Foissner W. 1993. *Colpodea* (Ciliophora). Inillustrated (ed)Protozoen-fauna. Vol. 4:1. Gustav Fischer Verlag, New York.

7. Padnos M, Jakowska S, Nigrelli RF. 1954. Morphology and life history of *Colpoda maupasi*, Bensonhurst strain. J Protozool 1:131–139. https://doi.org/10.1111/j.1550-7408.1954.tb00805.x

8. Müller H, Achilles-Day UEM, Day JG. 2010. Tolerance of the resting cysts of *Colpoda inflata* (Ciliophora, Colpodea) and *Meseres corlissi* (Ciliophora, Spirotrichea) to desiccation and freezing. Eur J Protistol 46:133–142. https://doi.org/10.1016/j.ejop.2009.12.004

9. 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

10. Guo S, Xiong W, Hang X, Gao Z, Jiao Z, Liu H, Mo Y, Zhang N, Kowalchuk GA, Li R, Shen Q, Geisen S. 2021. Protists as main indicators and determinants of plant performance. Microbiome 9:64. https://doi.org/10.1186/s40168-021-01025-w

11. Asiloglu R, Shiroishi K, Suzuki K, Turgay O, Murase J, Harada N. 2020. Protist-enhanced survival of a plant growth promoting rhizobacteria, *Azospirillum* sp. B510, and the growth of rice (*Oryza sativa* L.) plants. Appl Soil Ecol 154:103599. https://doi.org/10.1016/j.apsoil.2020.103599

12. Campbell CD, Warren A, Cameron CM, Hope SJ. 1997. Direct toxicity assessment of two soils amended with sewage sludge contaminated with heavy metals using a protozoan (*Colpoda steinii*) bioassay. Chemosphere 34:501–514. https://doi.org/10.1016/S0045-6535(96)00389-X

13. Akematsu T, Matsuoka T. 2008. Chromatin extrusion in resting encystment of *Colpoda cucullus*: a possible involvement of apoptosis-like nuclear death. Cell Biol Int 32:31–38. https://doi.org/10.1016/j.cellbi.2007.08.012

14. Martin-gonzalez A, Benitez L, Gutierrez JC. 1991. Cortical and nuclear events during cell division and resting cyst formation in *Colpoda inflata*. J Protozool 38:338–344. https://doi.org/10.1111/j.1550-7408.1991.tb01370.x

15. Tibbs J. 1968. Fine structure of *Colpoda steinii* during encystment and excystment. J Protozool 15:725–732. https://doi.org/10.1111/j.1550-7408.1968.tb02202.x

16. Ruthmann A, Kuck A. 1985. Formation of the cyst wall of the ciliate *Colpoda steinii*. J Protozool 32:677–682. https://doi.org/10.1111/j.1550-7408.1985.tb03101.x

17. 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

18. Sogame Y, Kojima K, Takeshita T, Kinoshita E, Matsuoka T. 2012. EF-1α and mitochondrial ATP synthase β chain: alteration of their expression in encystment-induced *Colpoda cucullus*. J Eukaryot Microbiol 59:401–406. https://doi.org/10.1111/j.1550-7408.2012.00628.x

19. Sogame Y, Hori M, Matsuoka T. 2016. EF-1α silencing by feeding RNAi suppresses resting cyst formation in *Colpoda cucullus* Nag-1 strain. Invertebr Surviv J 13:89–93. https://doi.org/10.25431/1824-307X/isj.v13i1.89-93

20. Tatsuomi M, Asuka K, Kunihisa S, Nobuaki N, Takahiko A, Akemi K, Ryota I. 2009. Role of Ca^{2+} and cAMP in a cell signaling pathway for resting cyst formation of ciliated protozoan *Colpoda cucullus*. J Protistol 6:103–110.

21. Jiang C, Wei W, Yan G, Shi T, Miao W. 2019. Transcriptome analysis reveals the molecular mechanism of resting cyst formation in *Colpoda aspera*. J Eukaryot Microbiol 66:212–220. https://doi.org/10.1111/jeu.12643

22. Padnos M. 1962. Cytology of cold induced transformation of octogenic reproductive cysts to resting cysts in *Colpoda maupasi*. J Protozool 9:13–20. https://doi.org/10.1111/j.1550-7408.1962.tb02573.x

23. Aury J-M, Jaillon O, Duret L, Noel B, Jubin C, Porcel BM, Ségurens B, Daubin V, Anthouard V, Aiach N, et al. 2006. Global trends of whole-genome duplications revealed by the ciliate *Paramecium tetraurelia*. Nature 444:171–178. https://doi.org/10.1038/nature05230

24. McGrath CL, Gout J-F, Doak TG, Yanagi A, Lynch M. 2014. Insights into three whole-genome duplications gleaned from the *Paramecium caudatum* genome sequence. Genetics 197:1417–1428. https://doi.org/10.1534/genetics.114.163287

25. Long H, Johri P, Gout J-F, Ni J, Hao Y, Licknack T, Wang Y, Pan J, Jiménez-Marín B, Lynch M. 2023. Paramecium genetics, genomics, and evolution. Annu Rev Genet 57:391–410. https://doi.org/10.1146/annurev-genet-071819-104035

26. Ahsan R, Blanche W, Katz LA. 2022. Macronuclear development in ciliates, with a focus on nuclear architecture. J Eukaryot Microbiol 69:e12898. https://doi.org/10.1111/jeu.12898

27. Rotterová J, Salomaki E, Pánek T, Bourland W, Žihala D, Táborský P, Edgcomb VP, Beinart RA, Kolísko M, Čepička I. 2020. Genomics of new ciliate lineages provides insight into the evolution of obligate anaerobiosis. Curr Biol 30:2037–2050. https://doi.org/10.1016/j.cub.2020.03.064

28. Marshall WF. 2021. Regeneration in *Stentor coeruleus*. Front Cell Dev Biol 9:753625. https://doi.org/10.3389/fcell.2021.753625

29. Lokanathan Y, Mohd-Adnan A, Wan K-L, Nathan S. 2010. Transcriptome analysis of the *Cryptocaryon irritans* tomont stage identifies potential genes for the detection and control of cryptocaryonosis. BMC Genomics 11:1–15. https://doi.org/10.1186/1471-2164-11-76

30. Lee SG, Na D, Park C. 2021. Comparability of reference-based and reference-free Transcriptome analysis approaches at the gene expression level. BMC Bioinformatics 22:1–9. https://doi.org/10.1186/s12859-021-04226-0

31. Eisen JA, Coyne RS, Wu M, Wu D, Thiagarajan M, Wortman JR, Badger JH, Ren Q, Amedeo P, Jones KM, et al. 2006. Macronuclear genome sequence of the ciliate *Tetrahymena thermophila*, a model eukaryote. PLoS Biol 4:e286. https://doi.org/10.1371/journal.pbio.0040286

32. Ackerman MS, Johri P, Spitze K, Xu S, Doak TG, Young K, Lynch M. 2017. Estimating seven coefficients of pairwise relatedness using population-genomic data. Genetics 206:105–118. https://doi.org/10.1534/genetics.116.190660

33. Cervantes MD, Hamilton EP, Xiong J, Lawson MJ, Yuan D, Hadjithomas M, Miao W, Orias E. 2013. Selecting one of several mating types through gene segment joining and deletion in *Tetrahymena thermophila*. PLoS Biol 11:e1001518. https://doi.org/10.1371/journal.pbio.1001518

34. Orias E, Singh DP, Meyer E. 2017. Genetics and epigenetics of mating type determination in *Paramecium* and *Tetrahymena*. Annu Rev Microbiol 71:133–156. https://doi.org/10.1146/annurev-micro-090816-093342

35. Gregory TR. 2001. The bigger the C-value, the larger the cell: genome size and red blood cell size in vertebrates. Blood Cells Mol Dis 27:830–843. https://doi.org/10.1006/bcmd.2001.0457

36. Rand DM, Kann LM. 1996. Excess amino acid polymorphism in mitochondrial DNA: contrasts among genes from *Drosophila*, mice, and humans. Mol Biol Evol 13:735–748. https://doi.org/10.1093/oxfordjournals.molbev.a025634

37. McDonald JH, Kreitman M. 1991. Adaptive protein evolution at the Adh locus in *Drosophila*. Nature 351:652–654. https://doi.org/10.1038/351652a0

38. Dunthorn M, Zufall RA, Chi J, Paszkiewicz K, Moore K, Mahé F. 2017. Meiotic genes in colpodean ciliates support secretive sexuality. Genome Biol Evol 9:1781–1787. https://doi.org/10.1093/gbe/evx125

39. Dray E, Dunlop MH, Kauppi L, San Filippo J, Wiese C, Tsai M-S, Begovic S, Schild D, Jasin M, Keeney S, Sung P. 2011. Molecular basis for enhancement of the meiotic DMC1 recombinase by RAD51 associated protein 1 (RAD51AP1). Proc Natl Acad Sci USA 108:3560–3565. https://doi.org/10.1073/pnas.1016454108

40. Chen Y-K, Leng C-H, Olivares H, Lee M-H, Chang Y-C, Kung W-M, Ti S-C, Lo Y-H, Wang A-J, Chang C-S, Bishop DK, Hsueh Y-P, Wang T-F. 2004. Heterodimeric complexes of Hop2 and Mnd1 function with Dmc1 to promote meiotic homolog juxtaposition and strand assimilation. Proc

Natl Acad Sci U S A 101:10572–10577. https://doi.org/10.1073/pnas.0404195101

41. Snowden T, Acharya S, Butz C, Berardini M, Fishel R. 2004. hMSH4-hMSH5 recognizes Holliday junctions and forms a meiosis-specific sliding clamp that embraces homologous chromosomes. Mol Cell 15:437–451. https://doi.org/10.1016/j.molcel.2004.06.040

42. Keeney S. 2001. Mechanism and control of meiotic recombination initiation. Curr Topics Dev Biol 52:1–53. https://doi.org/10.1016/s0070-2153(01)52008-6

43. Bamforth SS. 1971. The numbers and proportions of testacea and ciliates in litters and soils. J Protozool 18:24–28. https://doi.org/10.1111/j.1550-7408.1971.tb03274.x

44. Epstein RJ, Lin K, Tan TW. 2000. A functional significance for codon third bases. Gene 245:291–298. https://doi.org/10.1016/s0378-1119(00)00042-1

45. De Coster W, D'Hert S, Schultz DT, Cruts M, Van Broeckhoven C. 2018. NanoPack: visualizing and processing long-read sequencing data. Bioinformatics 34:2666–2669. https://doi.org/10.1093/bioinformatics/bty149

46. Lynch M. 2007. The origins of genome architecture. 1st ed. Vol. 98. Sinauer associates Sunderland, Massachusetts.