bioRxiv preprint doi: https://doi.org/10.64898/2026.06.24.734286; this version posted June 25, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license.

Evaluating the potential role and contribution of transposable elements to the evolution of microbial multicellularity across the tree of eukaryotes

Angela Correa[1], Matthew W. Brown[2], Idan Banson[2], Robert E. Jones[1], Phigy Kalulu [1], Charles Thompson[2], Alexander K. Tice[1], and David A. Ray[1]

[1] Department of Biological Sciences, Texas Tech University

[2] Department of Biological Sciences, Mississippi State University

Abstract

Multicellularity has evolved multiple times across the eukaryotic tree of life, including among protist lineages. Because transposable elements (TEs) strongly influence genome architecture and gene regulation, understanding their potential impact on genome structure and their relationship with gene expression may provide insight into the evolution of multicellularity.

Here, we generated a new genome assembly for the facultatively multicellular amoeba *Acrasis kona* and performed comparative analyses of TE composition, TE diversity, and TE-density organization across diverse protist lineages. Comparative analyses included unicellular and multicellular representatives from across the tree of eukaryotes, (Heterolobosea, Filasterea, Cristidiscoidea, and Chlorophyceae), including *Naegleria* spp., *Tetramitus jugosus*, *Capsaspora owczarzaki*, *Pigoraptor* spp., *Fonticula alba*, *Parvularia atlantis*, *Volvox carteri*, and *Chlamydomonas reinhardtii*.

To examine relationships between TEs and gene regulation, we integrated transcriptomic datasets from *A. kona*, *Capsaspora owczarzaki*, and *Volvox carteri* with genome-wide TE-density analyses of differentially expressed genes. TE abundance and composition varied substantially among lineages, with species that exhibit more complex developmental or cellular organization generally containing higher TE proportions than closely related unicellular taxa. Patterns of TE-density organization near up-regulated, down-regulated, and non-differentially expressed genes also differed among systems, ranging from strong TE depletion in *A. kona* to weaker or cell-type-specific patterns in *Capsaspora* and *Volvox*.

bioRxiv preprint doi: https://doi.org/10.64898/2026.06.24.734286; this version posted June 25, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license.

Together, these findings suggest that transposable elements are associated with multicellularity across diverse protist lineages, although the specific roles they play appear to be complex, lineage-specific, and not yet fully understood.

Keywords: transposable elements; multicellularity; protists; *Acrasis kona*; *Capsaspora owczarzaki*; *Volvox carteri*; TE density; gene regulation; comparative genomics; evolutionary genomics

Introduction

Transposable elements (TEs) are widely distributed genetic entities distinguished by their ability to move within and among genomes (Bourque et al., 2018). Through their capacity to relocate, TEs exert a substantial influence on genome composition, often accumulating over evolutionary time and playing a crucial role in shaping genome structure and function (Watson, 2014). Transposition itself is a powerful mutagenic process, but the presence of TEs can also modify genome structure by serving as substrates for non-homologous recombination and as sources of new genetic material (Wessler, 2006). Owing to their capacity to enhance genomic plasticity and variability, the study of TEs has become central to understanding how these elements contribute to species adaptation and evolutionary innovation (Casacuberta & González, 2013; Feschotte, 2008).

Because TEs can alter genome architecture, regulatory networks, and chromatin organization, their evolutionary impact becomes particularly relevant when investigating major transitions in biological complexity, including the evolution of multicellular lifestyles. The emergence of multicellular systems requires coordinated regulation of signaling, adhesion, communication, and differentiation pathways, processes that depend heavily on the organization and modulation of gene regulatory networks (King, 2004; Rokas, 2008). Increasing evidence indicates that TEs contribute to regulatory innovation by introducing novel cis-regulatory elements, modifying local chromatin environments, and influencing developmental gene expression patterns (Chuong et al., 2016; Sundaram & Wysocka, 2020). Consequently, studying TE landscapes may provide important insight into how genomes evolve to support increasingly complex developmental and cellular organizations.

bioRxiv preprint doi: https://doi.org/10.64898/2026.06.24.734286; this version posted June 25, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license.

Multicellularity is a complex trait that has independently emerged multiple times across the tree of life, including in fungi, diverse lineages of colonial and aggregatively multicellular protist lineages, charophyte green algae, the group from which land plants evolved, as well as other green, red, and brown algal lineages (Ruiz-Trillo et al., 2007). Over the past two decades, major advances in understanding the origins of multicellularity have been achieved through genomic comparisons among early-branching metazoans—such as cnidarians, ctenophores, and sponges (Putnam et al., 2007; Srivastava et al., 2008, 2010a; Moroz et al., 2014; Fortunato et al., 2014)—and their closest unicellular relatives within the Holozoa, including the choanoflagellates *Monosiga brevicollis* and *Salpingoeca rosetta* (King et al., 2008; Fairclough et al., 2013) and the filasterean *Capsaspora owczarzaki* (Suga et al., 2013; Sebé-Pedrós et al., 2017; Ocaña-Pallarès et al., 2022). Collectively, these studies have shown that multicellular traits frequently emerge through the modification and co-option of ancestral genes and regulatory pathways rather than through the origin of entirely novel gene repertoires, emphasizing the importance of comparative analyses among protists and other early-diverging eukaryotic lineages for understanding the evolution of biological complexity (Sebé-Pedrós et al., 2013, 2017; Ocaña-Pallarès et al., 2022).

Protists provide an especially valuable framework for investigating the evolution of multicellularity because they encompass a broad diversity of developmental and cellular organizations, ranging from strictly unicellular organisms to lineages capable of aggregation, cellular cooperation, or differentiated multicellularity (Sebé-Pedrós et al., 2017; Parfrey & Lahr, 2013). Comparisons among species sharing recent evolutionary histories but exhibiting contrasting developmental strategies therefore provide an opportunity to examine how changes in genome architecture, including TE composition and genomic distribution, may contribute to evolutionary diversification and increasing biological complexity. One important comparative system occurs within the volvocine green algae, which includes the unicellular species *Chlamydomonas reinhardtii* and the multicellular species *Volvox carteri*. Comparative genomic studies have demonstrated that *Volvox* retains most of the ancestral genes present in *Chlamydomonas*, while exhibiting regulatory modifications and expansions involving genes associated with extracellular matrix formation, developmental signaling, and cell-cycle regulation (Merchant et al., 2007; Prochnik et al., 2010; Herron, 2016). Transcriptomic analyses have additionally revealed extensive compartmentalization of gene

bioRxiv preprint doi: https://doi.org/10.64898/2026.06.24.734286; this version posted June 25, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license.

expression between reproductive and somatic cell types in *Volvox carteri* (Klein et al., 2017), emphasizing the value of this system for investigating how TEs may relate to developmental gene regulation within differentiated multicellular organisms (Hanschen et al., 2017; Matt & Umen, 2018).

Comparable evolutionary systems are also found among heterolobosean and holozoan protists. The facultatively aggregative amoeba *Acrasis kona* alternates between unicellular growth and multicellular aggregation under specific environmental conditions, forming fruiting structures through coordinated morphogenesis (Brown, Silberman, & Spiegel, 2012). Recent genomic and transcriptomic analyses revealed that aggregation-associated genes in *A. kona* include homologs involved in signaling, adhesion, and differentiation pathways shared with other multicellular eukaryotes, suggesting that components of multicellular regulatory toolkits predate the emergence of complex multicellular organisms (Sheikh et al., 2024). Within Holozoa, the filasterean *Capsaspora owczarzaki* has become an important model for studying the origins of multicellular complexity because it exhibits distinct life stages and possesses genes associated with adhesion, signaling, and transcriptional regulation (Suga et al., 2013). More recently, comparative genomic analyses including *Pigoraptor chileana* and *Pigoraptor vietnamica* have highlighted substantial genomic and regulatory diversification among unicellular holozoans prior to the emergence of animals, emphasizing the importance of these lineages for understanding the evolutionary foundations of cellular and developmental complexity (Ocaña-Pallarès et al., 2022). Similar comparative frameworks are also represented within Cristidiscoidea by the aggregative multicellular species *Fonticula alba* and the unicellular relative *Parvularia atlantis*, which provide additional opportunities to investigate the evolutionary origins of multicellular organization across diverse protist lineages (Tice et al., 2022; López-Escardó et al., 2017).

Although previous studies have examined TE abundance and genome composition across diverse eukaryotic lineages, most comparative analyses have primarily focused on identifying TE classes and quantifying their genomic proportions using annotation approaches such as RepeatMasker. However, little is known about how TE density in protists differs across contrasting developmental strategies relative to up-regulated, down-regulated, and non-differentially expressed genes. This study addresses two overarching questions. First, do protist species exhibiting multicellular or aggregative lifestyles differ in TE

bioRxiv preprint doi: https://doi.org/10.64898/2026.06.24.734286; this version posted June 25, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license.

abundance and diversity relative to unicellular representatives within the same major evolutionary lineages? Second, are up-regulated, down-regulated, and non-differentially expressed genes associated with distinct genomic TE-density environments across different protist systems? Based on current evidence regarding the influence of TEs on genome organization and regulatory evolution (Bourque et al., 2018; Chuong et al., 2016), we hypothesized that species exhibiting multicellular or aggregative lifestyles may exhibit increased TE abundance and distinct TE-density patterns surrounding up-regulated, down-regulated, and non-differentially expressed genes relative to related unicellular taxa.

To address these questions, we generated a new long-read genome assembly for *Acrasis kona* and performed comparative analyses of TE composition, diversity, and TE-density organization across representative species from Heterolobosea, Filasterea, Cristidiscoidea, and Chlorophyceae. Additionally, we integrated existing transcriptomic datasets from *A. kona*, *Capsaspora owczarzaki*, and *Volvox carteri* to evaluate whether developmentally regulated genes are associated with distinct genomic TE environments across multiple protist systems. These analyses provide a comparative framework for investigating how TE landscapes and TE-associated regulatory organization vary across lineages with distinct forms of developmental and cellular complexity.

Methods

Genome sequencing, assembly, and decontamination

High-molecular-weight genomic DNA from *Acrasis kona*, grown the yeast *Rhodotorula mucilaginosa* as a food source was sequenced using the Oxford Nanopore Technologies (ONT) PromethION P2 Solo platform. The data (~14 Gb; read N50 ≈ 38 kb) was assembled de novo with Flye v2.9 (Kolmogorov et al., 2019), an assembler optimized for long-read ONT data, and subsequently polished with Medaka v1.7.3 to improve base-level accuracy. To detect putative contaminants, BLASTn v2.14 searches were performed against the previously published *A. kona* genome (Sheikh et al., 2024), the NCBI RefSeq bacterial database, and the *R. mucilaginosa* genome (GCA_034406715.1). Contigs matching bacterial or yeast sequences were removed using BEDTools v2.31.1, and sequences shorter than 10 kb were filtered with SeqKit v2.3.1.

bioRxiv preprint doi: https://doi.org/10.64898/2026.06.24.734286; this version posted June 25, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license.

Following decontamination, the filtered reads were reassembled with Flye to maximize contiguity and reduce fragmentation. Assembly quality was assessed with QUAST v5.2 (Gurevich et al., 2013), and completeness was evaluated using BUSCO v5.5.0 (Manni et al., 2021). The resulting polished and decontaminated assembly served as the reference genome for TE annotation, gene prediction, and TE-density analyses in this taxon. The *Acrasis kona* genome assembly has been deposited at NCBI under BioProject accession PRJNA1447523, BioSample accession SAMN57028165, and Whole Genome Shotgun accession JBWXZU000000000.

Transposable element identification and TE library construction

Additional genome assemblies for comparative analyses were obtained from publicly available genomic repositories, including the NCBI Genome database and the European Nucleotide Archive (ENA) (Table 1). The dataset includes species exhibiting facultative or obligate multicellularity and their closely related unicellular relatives from the eukaryotic lineages Heterolobosea, Filasterea, Cristidiscoidea, and Chlorophyceae. This dataset allows for direct comparisons of TE proportion and diversity because these species share a relatively recent common ancestry but exhibit contrasting developmental or cellular organizations.

To characterize transposable elements across genomes, a multi-tool detection strategy was used to increase sensitivity and classification accuracy. Five independent pipelines were executed: RepeatModeler2 (Flynn et al., 2020), MegaLTR (Mokhtar & El Allali, 2023), Inpactor2 (Orozco-Arias et al., 2018), HELIANO (Li et al., 2024), and HiTE (Hu et al., 2024). These tools combine structural and homology-based detection approaches and differ in their strengths, allowing comprehensive identification of both autonomous and non-autonomous elements. RepeatModeler2 provided broad de novo repeat discovery, while HiTE refined element boundaries. MegaLTR, Inpactor2, and HELIANO contributed additional class-specific resolution for LTR retrotransposons and Helitrons.

For each assembly the TE libraries generated by the five pipelines were merged and cross-compared. Redundant consensus sequences were removed by clustering with USEARCH (Edgar, 2010), resulting in a non-redundant, curated TE library. This library was used as the reference database for subsequent comparative analyses. RepeatMasker v4.1.5 (Smit, Hubley, and Green, 2013–2024) was then used to annotate each genome assembly

bioRxiv preprint doi: https://doi.org/10.64898/2026.06.24.734286; this version posted June 25, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license.

using the curated TE library. Annotation outputs were parsed to quantify the proportion of masked genomic bases and to calculate the relative representation of major TE classes, including LTR retrotransposons, DNA transposons, Helitrons, LINEs, and SINEs.

Table 1. Genome assemblies analyzed in this study: Genome assemblies representing unicellular, multicellular, and facultatively multicellular protists were used to compare TE diversity and proportion across lineages. The *Acrasis kona* genome was newly assembled in this study from ONT PromethION data, while all other assemblies were retrieved from the NCBI Genome database. Accession numbers for each genome are provided.

Taxonomic Association	Species	Lifestyle	Genome Accession
Chlorophyceae	*Chlamydomonas reinhardtii*	Unicellular	GCA_000002595.3
	Volvox carteri	Multicellular	GCA_000143455.1
Cristidiscoidea	*Fonticula alba*	Multicellular facultative	GCA_000388065.2
	Parvularia atlantis	Unicellular	GCA_943704415.1
Filasterea	*Capsaspora owczarzaki*	Multicellular facultative	GCA_000151315.2
	Pigoraptor chileana	Unicellular	PRJEB52884
	Pigoraptor vietnamica	Unicellular	PRJEB52884
Heterolobosea	*Acrasis kona*	Multicellular facultative	This study (ONT PromethION assembly)
	Tetramitus jugosus	Unicellular	PRJEB50460
	Naegleria gruberi	Unicellular	GCA_000004985.1
	Naegleria fowleri	Unicellular	GCA_014843625.1
	Naegleria lovaniensis	Unicellular	GCA_022530875.1

Gene prediction and functional annotation

Gene models for *Acrasis kona* were predicted using BRAKER3 (Gabriel et al., 2023) with soft-masked genome assemblies as input. Homology-guided predictions were generated using protein evidence from the UniProtKB Heterolobosea dataset, which integrates GeneMark-EP+ and AUGUSTUS. This gene set was used as the reference for all downstream analyses.

Functional annotation of *A. kona* genes was performed using a custom pipeline that combined sequence-similarity searches, domain prediction, and non-coding RNA detection. Predicted protein sequences from BRAKER3/AUGUSTUS were searched with BLASTp against a Heterolobosea UniProtKB protein database. BLASTp hits were filtered to retain matches with ≥60% identity, alignment lengths ≥50 amino acids, and e-values ≤1e-5. Conserved domains and Gene Ontology (GO) terms were further identified using InterProScan (Jones et al., 2014). Genes lacking significant BLASTp matches were

bioRxiv preprint doi: https://doi.org/10.64898/2026.06.24.734286; this version posted June 25, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license.

additionally screened for non-coding RNA features using Infernal against the Rfam covariance model database using Rfam gathering thresholds for significant covariance-model matches (Nawrocki & Eddy, 2013). Annotation results were merged into downstream annotation tables and the BRAKER3 GFF3 file using custom Python and awk scripts rather than a packaged annotation merger. BLASTp- and InterProScan-derived annotations were linked back to BRAKER gene IDs after removing transcript suffixes, allowing transcript-level annotations to be reassigned at the gene level. For GFF3 integration, a custom Python script compared BRAKER3 gene coordinates with genes from the reference GFF3, normalized scaffold names, required same-strand overlap, and retained matches with at least 100 bp overlap and ≥50% reciprocal coverage between features. Functional fields, including gene name, locus tag, product description, and database cross-references, were then added as attributes to the corresponding BRAKER3 gene and transcript entries. This annotated gene set provided the basis for subsequent TE-density analyses.

Transposable Element Density Analysis

The density of transposable elements in upstream regulatory regions was quantified using TE-Density v1.1.0 (Teresi, Teresi, & Edger, 2022). For each annotated gene, we calculated the proportion of bases annotated as TEs within fixed upstream windows (1 kb, 2 kb, and 5 kb) relative to the transcription start site. These distances were selected because they encompass the proximal promoter regions commonly associated with cis-regulatory activity in diverse eukaryotes.

TE-density analyses were conducted for *Acrasis kona*, *Capsaspora owczarzaki*, and *Volvox carteri*, the three species for which transcriptomic datasets representing distinct developmental or cellular conditions were available. TE-density estimates were subsequently integrated with differential expression analyses to evaluate whether developmentally regulated genes occur within distinct genomic TE environments. Genes were classified as upregulated, downregulated, or non-differentially expressed according to their corresponding transcriptomic comparisons, and TE-density distributions were compared among regulatory categories within each species.

bioRxiv preprint doi: https://doi.org/10.64898/2026.06.24.734286; this version posted June 25, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license.

Differential Gene Expression Analysis

Differential gene expression analyses were conducted for *Acrasis kona*, *Capsaspora owczarzaki*, and *Volvox carteri* using publicly available RNA-seq datasets representing distinct developmental or cellular conditions. Raw sequencing reads were downloaded from the NCBI Sequence Read Archive (SRA) and reprocessed for alignment to their corresponding genome assemblies. Details regarding transcriptomic datasets and associated accession numbers are provided in Supplementary Table S1.

Read quality was assessed using FastQC v0.11.9 (Andrews, 2010), and adapter sequences and low-quality bases were removed using Trimmomatic v0.39 (Bolger, Lohse, & Usadel, 2014). Clean reads were aligned to their respective genomes using minimap2 v2.26 (Li, 2018) with splice-aware alignment settings to accommodate intron–exon structure. Gene-level read counts were obtained using featureCounts (Liao et al., 2014).

For *Acrasis kona*, differential gene expression during the transition between aggregation, growth, and germination conditions was analyzed using the RNA-seq datasets generated by Sheikh et al. (2024). Because these datasets lacked biological replicates, gene expression differences were estimated using GFOLD (Feng et al., 2012), consistent with the analytical approach implemented in Sheikh et al. (2024). Two pairwise comparisons were performed: aggregation versus growth and aggregation versus germination. Genes with positive GFOLD values were classified as upregulated, whereas genes with negative GFOLD values were classified as downregulated.

For *Capsaspora owczarzaki* and *Volvox carteri*, transcriptomic datasets with biological replicates were analyzed using DESeq2 v1.38.3 (Love, Huber, & Anders, 2014). In *C. owczarzaki*, differential expression analyses were performed across developmental life stages using RNA-seq datasets generated by Sebé-Pedrós et al. (2013). In *V. carteri*, transcriptomic comparisons between reproductive and somatic cell types were performed using RNA-seq datasets generated by Hanschen et al. (2017). Genes with adjusted p-values ($padj < 0.05$) were considered significantly differentially expressed and were subsequently classified as upregulated or downregulated according to log2 fold-change direction.

bioRxiv preprint doi: https://doi.org/10.64898/2026.06.24.734286; this version posted June 25, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license.

The resulting expression profiles for all three species were integrated with TE-density estimates to evaluate whether genes exhibiting distinct regulatory behaviors occur within different upstream TE-density environments relative to non-differentially expressed genes.

Statistical Analysis of TE Density and Composition

Genome-wide TE proportion and composition were quantified using RepeatMasker annotations generated from the curated TE library produced by RepeatModeler2, MegaLTR, Inpactor2, HELIANO, and HiTE (Flynn et al., 2020; Orozco-Arias et al., 2018; Mokhtar & El Allali, 2023; Li et al., 2024; Hu et al., 2024). For each species, the total proportion of genomic bases masked as TEs and the relative contribution of major TE orders were calculated from RepeatMasker outputs, enabling comparative analyses across protist species exhibiting distinct developmental and cellular organizations.

To evaluate whether species exhibiting greater developmental or cellular complexity differ in TE proportion relative to unicellular representatives within the same evolutionary lineages, nonparametric analyses appropriate for small sample sizes and non-normal distributions were performed. Eight pairwise comparisons were conducted across four major protist lineages: Heterolobosea (*Acrasis kona* versus *Naegleria gruberi*, *N. fowleri*, *N. lovaniensis*, and *Tetramitus jugosus*), Filasterea (*Capsaspora owczarzaki* versus *Pigoraptor chileana* and *P. vietnamica*), Chlorophyceae (*Volvox carteri* versus *Chlamydomonas reinhardtii*), and Cristidiscoidea (*Fonticula alba* versus *Parvularia atlantis*). Paired differences in TE proportions were evaluated using a one-tailed Wilcoxon signed-rank test (Wilcoxon, 1945), testing the hypothesis that species exhibiting greater developmental complexity harbor higher overall TE proportions. Statistical robustness was additionally assessed using 10,000 paired sign-flip permutation replicates.

To quantify TE diversity within each genome, Shannon (H′) and Simpson (1–D) diversity indices were calculated from normalized TE-order proportions, summarizing both the richness and evenness of TE categories while accounting for differences in genome size. Principal component analysis (PCA) was performed using normalized TE-order proportions to evaluate similarities and differences in TE composition among protist genomes.

bioRxiv preprint doi: https://doi.org/10.64898/2026.06.24.734286; this version posted June 25, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license.

To evaluate relationships between TE density and gene regulation, TE-density estimates generated by TE-Density were integrated with differential expression analyses for *Acrasis kona*, *Capsaspora owczarzaki*, and *Volvox carteri*. Genes were classified as upregulated, downregulated, or non-differentially expressed according to their corresponding transcriptomic comparisons. For *A. kona*, analyses were conducted separately for aggregation versus growth and aggregation versus germination conditions. For *C. owczarzaki*, TE-density analyses were performed independently for aggregative versus adherent and filopodial versus aggregative developmental comparisons. For *V. carteri*, analyses compared reproductive and somatic cell-associated expression profiles.

Comparisons of upstream TE-density distributions (1 kb, 2 kb, and 5 kb windows) among regulatory categories were conducted using Mann–Whitney U tests and Welch's t-tests. To further evaluate whether observed differences exceeded random expectations, permutation tests using 10,000 iterations were performed for all pairwise comparisons. Analyses were conducted both across all genes and across TE-positive genes only to evaluate whether the inclusion of genes lacking nearby TEs influenced observed regulatory patterns.

All statistical analyses were conducted in Python 3.10 using SciPy (Virtanen et al., 2020), Pandas (McKinney, 2010), NumPy (Harris et al., 2020), and scikit-learn (Pedregosa et al., 2011).

Results

Improved Genome Assembly of *Acrasis kona*

A new long-read genome assembly was generated for *Acrasis kona* using ONT PromethION sequencing, followed by a decontamination pipeline to remove bacterial, mitochondrial, and prey organism sequences identified during quality control. The resulting curated assembly shows substantial improvements in contiguity relative to the previously published version (Sheikh et al., 2024) (Table 2). Assembly contiguity increased markedly: N50 rose from 42.6 kb to 262.7 kb, and L50 decreased from 262 contigs to 47. The total number of contigs declined from 2,039 to 359, indicating a significantly less fragmented assembly. The total assembly size increased from 44.0 Mb to approximately 47.8 Mb, reflecting improved recovery of genomic regions while excluding contaminant bacterial

bioRxiv preprint doi: https://doi.org/10.64898/2026.06.24.734286; this version posted June 25, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license.

contigs and separating mitochondrial and prey organism sequences. Ambiguous bases were nearly eliminated (7.56 to 0.21 per 100 kbp), further indicating improved assembly quality. BUSCO completeness remained stable after curation (85.5% vs. 85.1%), indicating that core gene content was preserved despite removal of contaminant sequences. However, the proportion of fragmented BUSCOs decreased (7.5% to 5.1%), suggesting improved gene contiguity, while duplicated BUSCOs increased slightly (10.6% to 12.2%), likely reflecting improved resolution of repetitive or paralogous regions. A small increase in missing BUSCOs (7.0% to 9.8%) is consistent with the removal of contaminant sequences that may have previously contributed spurious matches. Overall, these results indicate that assembly refinement substantially improved structural quality without compromising biological completeness.

Gene prediction and annotation for the curated *A. kona* assembly were performed using BRAKER3, which integrates RNA-seq evidence and protein homology to generate high-confidence gene models (Gabriel et al., 2023). The final annotation contained 25,067 predicted protein-coding genes after refinement and integration of transferred annotations from previous *A. kona* genome resources. Functional characterization was further improved using InterProScan, BLASTp searches against UniProtKB, and Infernal identification of non-coding RNAs, generating a consolidated annotation set for downstream comparative and TE-density analyses.

Transposable Element Identification and Comparative Analysis

Using the curated, non-redundant TE library generated from RepeatModeler2, MegaLTR, Inpactor2, HELIANO, and HiTE, each protist genome assembly was screened with RepeatMasker to quantify the genomic proportion and class composition of transposable elements. Details regarding TE consensus sequences, classification, and identification methods are provided in Supplementary Table S2, and the complete curated TE library used for RepeatMasker analyses is provided as Supplementary File S1. This unified annotation strategy enabled consistent comparisons across representative protist species exhibiting contrasting developmental and cellular organizations.

bioRxiv preprint doi: https://doi.org/10.64898/2026.06.24.734286; this version posted June 25, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license.

Table 2. Assembly quality metrics for *Acrasis kona:* The original assembly from Sheikh et al. (2024) is compared with the new long-read assembly generated in this study. Metrics include contiguity statistics (N50, L50), total assembly length, GC content, ambiguous base frequency, and BUSCO completeness scores. The new assembly shows major improvements in contiguity and completeness, reflecting more accurate recovery of genomic structure.

Category	Metric	*A. kona* (Sheikh et al., 2024)	*A. kona* (this study)
QUAST	N50	42,575	262,724
	N90	7,953	70,789
	auN	61,431	390,397
	L50	262	47
	L90	1,274	190
	# contigs ≥ 0 bp	2,039	363
	# contigs ≥ 1 kb	2,039	351
	# contigs ≥ 5 kb	1,874	336
	# contigs ≥ 10 kb	1,021	314
	# contigs ≥ 25 kb	467	268
	# contigs ≥ 50 kb	220	224
	Total contigs	2,039	359
	Largest contig	259,475 bp	1,572,856 bp
	Total length (≥0 bp)	44,020,442 bp	47,827,436 bp
	GC (%)	38.89%	38.80%
	N's per 100 kbp	7.56	0.21
BUSCO	Complete BUSCOs (C)	85.5% (218)	85.1% (217)
	Single-copy BUSCOs (S)	74.9% (191)	72.9% (186)
	Duplicated BUSCOs (D)	10.6% (27)	12.2% (31)
	Fragmented BUSCOs (F)	7.5% (19)	5.1% (13)
	Missing BUSCOs (M)	7.0% (18)	9.8 % (25)
	Total BUSCOs (n)	255	255

Across the surveyed lineages, substantial variation in total TE proportion and TE composition was observed among species (Table 3; Figure 1). Among heteroloboseans, *Acrasis kona* contained 5.52% repetitive DNA, substantially exceeding the TE proportions detected in *Naegleria gruberi* (0.28%), *Naegleria fowleri* (0.28%), *Naegleria lovaniensis* (0.13%), and *Tetramitus jugosus* (0.14%). The TE landscape of *A. kona* was dominated by LTR retrotransposons (2.55%) and DNA transposons (2.00%), with additional contributions from rolling-circle Helitron elements (0.89%). In contrast, the compared heterolobosean unicellular species exhibited markedly reduced TE content across all TE categories. Within Filasterea, *Capsaspora owczarzaki* exhibited a total TE proportion of 3.86%, driven primarily by rolling-circle elements (1.94%) and DNA transposons (1.20%). Interestingly, the two *Pigoraptor* species displayed similarly elevated TE proportions, with *Pigoraptor chileana* and *Pigoraptor vietnamica* containing 3.18% and 3.24% repetitive content,

bioRxiv preprint doi: https://doi.org/10.64898/2026.06.24.734286; this version posted June 25, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license.

respectively. In both *Pigoraptor* species, DNA transposons represented the dominant TE category, exceeding 2% of the genome. These results indicate that elevated TE abundance within Filasterea is not restricted to *Capsaspora* and may instead characterize multiple holozoan lineages exhibiting distinct developmental or cooperative cellular behaviors.

Table 3. Proportion of transposable elements by class across unicellular and multicellular protist genomes: Proportional genomic content of major transposable element (TE) classes—including DNA transposons, LINEs, LTR retrotransposons, rolling-circle Helitrons, and SINEs—is shown for each unicellular and multicellular (or aggregatively multicellular) protist species analyzed. Values represent the percentage of the genome masked by each TE class based on RepeatMasker screening using the curated TE library generated in this study. The table highlights substantial lineage-specific variation in both total TE content and class composition across protist genomes.

Species genome assemblies	Taxonomic Group	Type	LTR (%)	DNA (%)	LINE (%)	RC (%)	SINE (%)	Total (%)
Acrasis kona	Heterolobosea	Multicellular (fac.)	2.55	2.00	0.06	0.89	0.02	5.52
Naegleria gruberi	Heterolobosea	Unicellular	0.27	0.01	0.00	0.00	0.00	0.28
Naegleria fowleri	Heterolobosea	Unicellular	0.26	0.01	0.01	0.00	0.00	0.28
Naegleria lovaniensis	Heterolobosea	Unicellular	0.10	0.01	0.00	0.00	0.00	0.13
Tetramitus jugosus	Heterolobosea	Unicellular	0.08	0.03	0.02	0.01	0.00	0.14
Capsaspora owczarzaki	Filasterea	Multicellular (fac.)	0.57	1.20	0.14	1.94	0.00	3.86
Pigoraptor chileana	Filasterea	Unicellular	0.59	2.40	0.19	0.00	0.00	3.18
Pigoraptor vietnamica	Filasterea	Unicellular	0.24	2.89	0.10	0.00	0.00	3.24
Volvox carteri	Chlorophyceae	Multicellular	8.67	2.58	1.29	0.22	0.19	12.96
Chlamydomonas reinhardtii	Chlorophyceae	Unicellular	2.75	1.77	2.76	0.37	0.12	7.76
Fonticula alba	Cristidiscoidea	Multicellular (fac.)	0.13	0.19	0.04	0.00	0.00	0.36
Parvularia atlantis	Cristidiscoidea	Unicellular	0.06	0.02	0.03	0.00	0.00	0.11

Within Chlorophyceae, *Volvox carteri* exhibited the highest TE proportion among all analyzed species, with 12.96% of the genome annotated as repetitive DNA. This increase was largely driven by extensive accumulation of LTR retrotransposons (8.67%), accompanied by substantial DNA transposon (2.58%) and LINE (1.29%) content. In comparison, *Chlamydomonas reinhardtii* contained 7.76% repetitive DNA, with a more balanced TE composition distributed among LTR, LINE, and DNA transposon categories. Although both examined volvocine algae contained relatively TE-rich genomes, *V. carteri* exhibited substantially greater expansion of LTR retrotransposons. Within Cristidiscoidea, *Fonticula alba* contained 0.36% repetitive DNA, compared with 0.11% in *Parvularia*

bioRxiv preprint doi: https://doi.org/10.64898/2026.06.24.734286; this version posted June 25, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license.

399 *atlantis*. Although both genomes exhibited comparatively low TE abundance, *F. alba*
400 consistently displayed higher proportions across most TE categories.

401 Overall, these analyses revealed strong lineage-specific variation in TE composition
402 and abundance across protist genomes. Multicellular or aggregative taxa generally tended to
403 exhibit higher overall TE content, whereas unicellular representatives often displayed
404 reduced and compositionally simpler TE landscapes. However, this pattern was not universal
405 across all examined taxa, particularly within Filasterea, where elevated TE abundance was
406 also observed in the *Pigoraptor* genomes, which are primarily unicellular but exhibit
407 transient aggregative and multicellular-like behaviors under certain conditions (Hehenberger
408 et al., 2017; Hehenberger et al., 2020).

409 To statistically evaluate whether species exhibiting greater developmental or cellular
410 complexity consistently harbor increased TE proportions, paired nonparametric analyses
411 were conducted across the eight phylogenetic comparisons representing Heterolobosea,
412 Filasterea, Chlorophyceae, and Cristidiscoidea. The paired Wilcoxon signed-rank test
413 revealed a significant enrichment of TE content among the species exhibiting greater
414 developmental complexity ($W = 36$, one-tailed $p = 0.0039$). Similarly, paired sign-flip
415 permutation analyses supported the robustness of this pattern (observed mean difference =
416 3.50; permutation $p = 0.0053$), indicating that the observed differences in TE proportion are
417 unlikely to result from stochastic variation alone.

418 These results are summarized in Figure 1, which illustrates substantial differences in
419 both total TE proportion and relative TE-class representation among the surveyed protist
420 species. The stacked barplot highlights clear lineage-specific variation in the contribution of
421 major TE categories, including LTR retrotransposons, DNA transposons, LINEs, rolling-
422 circle elements, and SINEs.

423 TE diversity quantified using Shannon (H′) and Simpson (1–D) indices, revealed
424 lineage-specific rather than universal patterns (Figure 2). Within Heterolobosea, *Acrasis kona*
425 showed high TE diversity ($H' = 1.09$; $1–D = 0.63$), markedly exceeding the three *Naegleria*
426 species, which exhibited low diversity values consistent with their reduced TE content.
427 However, *Tetramitus jugosus* also showed relatively high diversity ($H' = 1.12$; $1–D = 0.61$),
428 indicating that TE diversity within Heterolobosea does not strictly follow the same trend as

bioRxiv preprint doi: https://doi.org/10.64898/2026.06.24.734286; this version posted June 25, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license.

total TE abundance. This pattern may partially reflect the influence of low overall TE content on diversity estimates, where relatively few TE families distributed across multiple categories can yield elevated diversity indices despite limited total TE abundance.

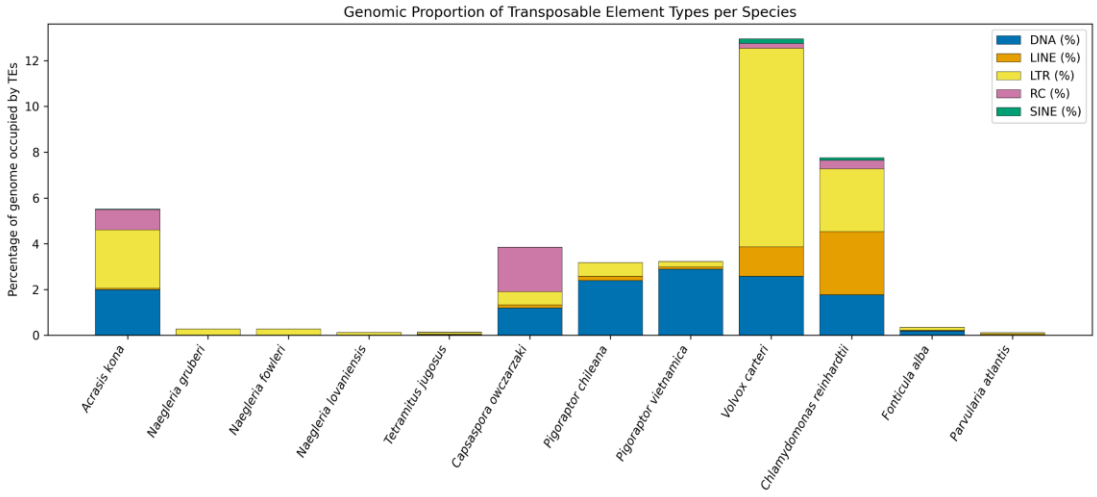


Figure 1. Genomic proportions of major transposable element types across paired protist species:

The stacked barplot shows the percentage of each genome occupied by five major transposable element (TE) categories: DNA transposons, LINEs, LTR retrotransposons, rolling-circle (RC/Helitron) elements, and SINEs. Species are arranged along the x-axis according to major evolutionary lineages, including Heterolobosea, Filasterea, Chlorophyceae, and Cristidiscoidea. Considerable variation in both total TE proportion and TE composition was observed among species. In several lineages, including Heterolobosea and Chlorophyceae, aggregative or multicellular representatives exhibited substantially expanded TE landscapes relative to related unicellular species, particularly through the accumulation of LTR retrotransposons and DNA transposons. Within Filasterea, *Capsaspora owczarzaki* exhibited slightly higher total TE content than both *Pigoraptor* species, although all three genomes contained comparatively TE-rich landscapes dominated primarily by DNA transposons and rolling-circle elements.

Within Filasterea, *Capsaspora owczarzaki* exhibited higher TE diversity ($H' = 1.11$; $1–D = 0.63$) than both *Pigoraptor chileana* ($H' = 0.70$; $1–D = 0.39$) and *Pigoraptor vietnamica* ($H' = 0.41$; $1–D = 0.20$). In Chlorophyceae, *Chlamydomonas reinhardtii* had the highest overall TE diversity among all surveyed genomes ($H' = 1.28$; $1–D = 0.69$), despite having lower total TE abundance than *Volvox carteri*. In Cristidiscoidea, *Fonticula alba* and *Parvularia atlantis* showed nearly identical diversity values despite differing in total TE proportion.

A principal component analysis (PCA) based on proportional representation of TE classes (Figure 3) revealed substantial lineage-specific differences in TE composition among

bioRxiv preprint doi: https://doi.org/10.64898/2026.06.24.734286; this version posted June 25, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license.

454 the surveyed protist genomes. PC1 explained 38.4% of the total variance, whereas PC2
455 accounted for an additional 30.9%, together capturing the majority of variation in TE
456 composition across species. Separation along both axes reflected differences in the relative
457 contribution of major TE classes, particularly LTR retrotransposons, DNA transposons,
458 rolling-circle elements, and LINEs.

459 Several multicellular or aggregative representatives occupied distinct regions of the
460 ordination space relative to related unicellular taxa. *Acrasis kona* was clearly separated from
461 the three *Naegleria* species along PC1, consistent with its expanded contribution of LTR
462 retrotransposons and rolling-circle elements. Similarly, *Capsaspora owczarzaki* was
463 differentiated from both *Pigoraptor* species, reflecting its greater relative abundance of
464 rolling-circle elements. Within Chlorophyceae, *Volvox carteri* and *Chlamydomonas*
465 *reinhardtii* were separated primarily along PC2, driven by the strong LTR retrotransposon
466 enrichment observed in *V. carteri* and the comparatively balanced TE composition of *C.*
467 *reinhardtii*.

468 Alternatively, several taxa occupied relatively similar regions of PCA space despite
469 differences in total TE abundance. For example, *Fonticula alba* and *Parvularia atlantis*
470 clustered relatively closely, suggesting similarities in proportional TE composition despite
471 differing total TE proportions. In contrast, *Tetramitus jugosus* occupied a distinct position
472 from the three *Naegleria* species despite belonging to the same major lineage and exhibiting
473 comparatively low overall TE abundance. Although the PCA did not reveal a consistent
474 clustering pattern associated with multicellular or aggregative lifestyles across the examined
475 taxa, it did indicate that TE composition varies strongly among protist lineages and that
476 increases in total TE abundance are often accompanied by shifts in the relative representation
477 of specific TE categories rather than by uniform expansion across all TE classes.

478

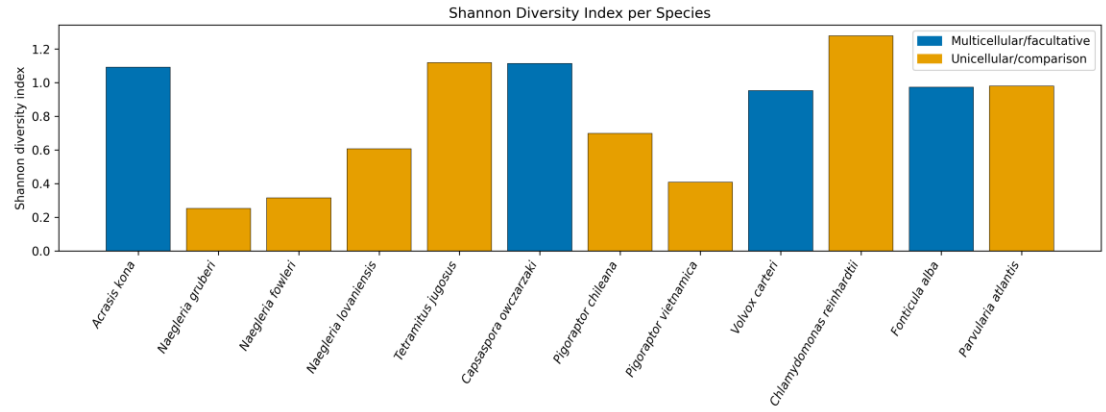
479

480

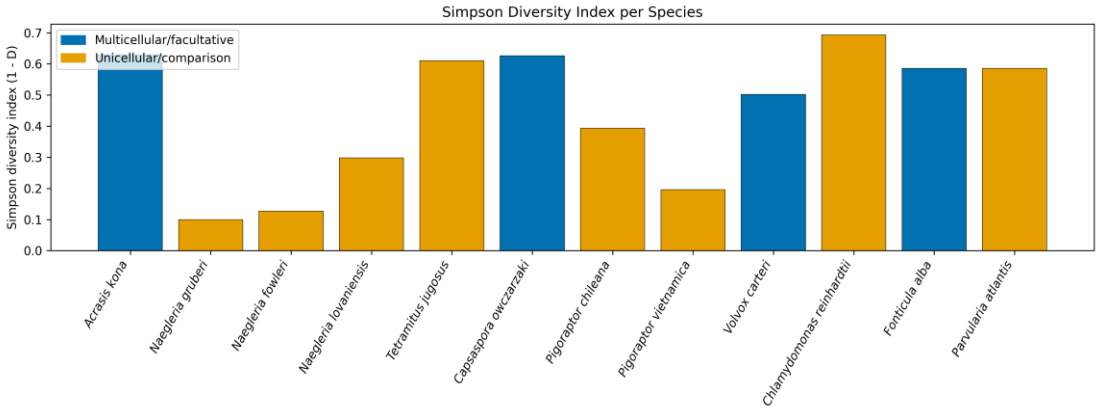
481

482

bioRxiv preprint doi: https://doi.org/10.64898/2026.06.24.734286; this version posted June 25, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license.



A)



B)

Figure 2. Transposable element diversity across multicellular and unicellular protist species: (A) Shannon diversity index (H′). (B) Simpson diversity index (1–D). Species are arranged according to major evolutionary lineages, with blue bars representing species exhibiting greater developmental or cellular complexity and orange bars representing unicellular comparison taxa. TE diversity exhibited strong lineage-specific variation rather than a universal relationship with multicellularity or aggregative behavior. *Acrasis kona* displayed substantially higher TE diversity than the compared *Naegleria* species, whereas *Tetramitus jugosus* also exhibited relatively high diversity despite its low overall TE proportion. Within Filasterea, *Capsaspora owczarzaki* exhibited greater TE diversity than both *Pigoraptor* species. In contrast, the unicellular alga *Chlamydomonas reinhardtii* exhibited the highest overall TE diversity among all analyzed genomes, despite containing lower total TE abundance than *Volvox carteri.* These results indicate that increased TE abundance does not necessarily correspond to increased TE diversity across protist lineages

bioRxiv preprint doi: https://doi.org/10.64898/2026.06.24.734286; this version posted June 25, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license.

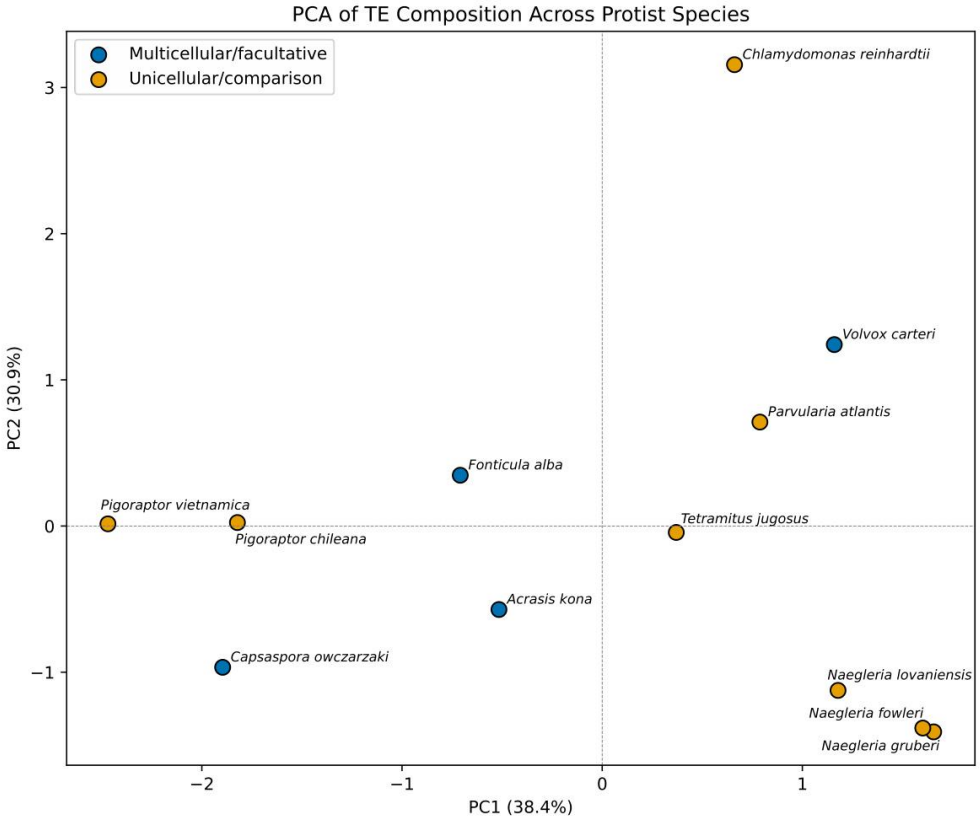


Figure 3. Principal component analysis (PCA) of transposable element composition across representative protist genomes. The PCA characterizes genome-wide TE composition using the proportional representation of major TE classes, including LTR retrotransposons, DNA transposons, LINEs, rolling-circle (RC/Helitron) elements, and SINEs. Blue points represent species exhibiting greater developmental or cellular complexity, whereas orange points represent unicellular comparison taxa. PC1 explains 38.4% of the total variance and PC2 explains 30.9%. Species are distributed according to lineage-specific differences in TE composition and relative TE-class abundance. *Acrasis kona* is separated from the *Naegleria* species primarily by its increased contribution of LTR retrotransposons and rolling-circle elements, whereas *Capsaspora owczarzaki* differs from the *Pigoraptor* species through greater RC/Helitron representation. Within Chlorophyceae, *Volvox carteri* and *Chlamydomonas reinhardtii* occupy distinct positions consistent with major differences in LTR retrotransposon abundance. Overall, the ordination highlights substantial variation in TE composition across protist lineages rather than a single conserved TE landscape associated with multicellularity or aggregative behavior.

TE-density patterns across developmental regulatory groups

Comparisons of upstream TE density across developmental regulatory categories revealed substantial variation among the analyzed protist systems (Table 4). Patterns differed

bioRxiv preprint doi: https://doi.org/10.64898/2026.06.24.734286; this version posted June 25, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license.

both in magnitude and direction, indicating that relationships between TE density and developmental gene regulation are highly lineage-specific rather than universally conserved.

In *Acrasis kona*, aggregation versus growth comparisons exhibited minimal differences in upstream TE density among regulatory categories. At the 5 kb window, upregulated (0.046), downregulated (0.044), and non-differentially expressed (NDE) genes (0.045) displayed nearly identical mean TE-density values, and statistical analyses did not detect significant differences among groups. Similar patterns were observed across the 1 kb and 2 kb upstream windows, where TE-density values also remained highly comparable among regulatory categories and lacked significant statistical support (Supplementary Table S3).

In comparison, aggregation versus germination comparisons revealed a pronounced depletion of TE density surrounding genes important during developmental transitions. Upregulated and downregulated genes exhibited substantially lower TE-density values (0.032 and 0.026, respectively) than NDE loci (0.053), and these differences were consistently supported by parametric, nonparametric, and permutation-based statistical analyses ($p < 1 \times 10^{-5}$). Similar depletion patterns were observed across all evaluated upstream windows (1, 2, and 5 kb), although the strongest and most consistent statistical support was detected at the 5 kb window (Supplementary Table S3).

In *Capsaspora owczarzaki*, comparisons involving the aggregative state produced comparatively weak and inconsistent TE-density patterns. During aggregative versus adherent transitions, upregulated genes exhibited moderately elevated TE density (0.040) relative to NDE genes (0.032), whereas downregulated genes displayed lower values (0.029). Aggregative versus filopodial comparisons showed only minor differences among regulatory categories, with mean TE-density values remaining relatively similar across groups. None of the *Capsaspora* comparisons reached statistical significance, suggesting comparatively weak associations between TE density and developmental regulation in this lineage.

In *Volvox carteri*, reproductive versus somatic comparisons revealed a contrasting pattern relative to *A. kona*. Analyses including all genes showed that somatic-upregulated genes exhibited higher upstream TE density (0.100) than both reproductive-upregulated genes (0.067) and NDE loci (0.092) at the 5 kb window. These differences were supported

bioRxiv preprint doi: https://doi.org/10.64898/2026.06.24.734286; this version posted June 25, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license.

by multiple statistical tests, including Mann–Whitney U, Welch's t-tests, and permutation analyses ($p < 0.05$). However, analyses restricted to TE-positive genes (Supplementary Table S4) revealed substantially reduced effect sizes and weaker statistical support among regulatory categories, indicating that the observed signal is driven primarily by the presence or absence of nearby TEs rather than by large differences in TE density among already TE-associated loci.

Table 4. Upstream TE-density patterns across developmental regulatory comparisons in protist systems. Mean upstream TE-density values calculated within 5 kb windows surrounding genes belonging to distinct developmental regulatory categories in *Acrasis kona*, *Capsaspora owczarzaki*, and *Volvox carteri*. All analyses were performed using the complete set of annotated genes ("All genes" dataset). Comparisons include aggregation-associated transitions in *A. kona* and *C. owczarzaki* as well as reproductive versus somatic differentiation in *V. carteri*. For each comparison, the table reports the number of genes assigned to each regulatory category, mean TE-density values, overall TE-density pattern, and statistical significance. The results reveal lineage-specific relationships between TE density and developmental gene regulation, including TE depletion near regulated genes in *A. kona*, relatively weak and variable patterns in *C. owczarzaki*, and increased TE density near somatic-associated genes in *V. carteri*.

Species	Developmental comparison	Dataset	Regulated gene category	Mean TE density	Main pattern	Significance
Acrasis kona	Aggregation vs Growth	All genes	Upregulated	0.046	weak differences	NS
			Downregulated	0.044		
			NDE	0.045		
Acrasis kona	Aggregation vs Germination	All genes	Upregulated	0.032	NDE > regulated	p < 1e-5
			Downregulated	0.026		
			NDE	0.053		
Capsaspora owczarzaki	Aggregative vs Adherent	All genes	Upregulated	0.04	Upregulated > NDE	NS
			Downregulated	0.029		
			NDE	0.032		
Capsaspora owczarzaki	Aggregative vs Filopodial	All genes	Upregulated	0.031	weak differences	NS
			Downregulated	0.033		
			NDE	0.032		
Volvox carteri	Reproductive vs Somatic	All genes	Reproductive-upregulated	0.067	Somatic > NDE > Reproductive	p < 0.05
			Somatic-upregulated	0.1		
			NDE	0.092		

Although several comparisons were statistically significant, the absolute differences in TE density among regulatory categories were relatively modest, generally differing by only a few percent of upstream sequence composition across the analyzed 5 kb

bioRxiv preprint doi: https://doi.org/10.64898/2026.06.24.734286; this version posted June 25, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license.

588 windows. These results therefore suggest modest but consistent differences in genomic TE
589 environments across developmental regulatory groups.

590 **Discussion**

591 **TE proportion and diversity across protist genomes**

592 Across the examined phylogenetic comparisons, species exhibiting aggregative,
593 cooperative, or differentiated developmental states generally contained higher overall
594 proportions of transposable elements than representative unicellular taxa. This pattern was
595 especially pronounced in *Acrasis kona* and *Volvox carteri*, which exhibited substantially
596 expanded TE landscapes relative to closely related heterolobosean and volvocine unicellular
597 representatives. Statistical analyses across the examined lineage pairs further supported a
598 significant overall enrichment of TEs among species exhibiting greater developmental
599 complexity. These findings are broadly consistent with observations from plants and animals,
600 where TE accumulation has been associated with increased genome plasticity, regulatory
601 diversification, and long-term expansion of gene regulatory networks (Bourque et al., 2018;
602 Elbarbary et al., 2016).

603 However, incorporation of the newly available holozoan genomes revealed a more
604 nuanced evolutionary pattern than initially anticipated. Although *Capsaspora owczarzaki*
605 exhibited elevated TE abundance relative to many protist genomes, both *Pigoraptor* species
606 also contained comparatively TE-rich genomes despite lacking multicellularity. Importantly,
607 *Pigoraptor* species have been reported to exhibit transient multicellular aggregations,
608 coordinated feeding behaviors, and syncytium-like cellular stages associated with
609 cooperative interactions among cells (Hehenberger et al., 2017). These observations suggest
610 that some genomic features associated with increased TE accumulation may emerge prior to
611 stable obligate multicellularity and may instead characterize lineages capable of facultative
612 multicellular or cooperative behaviors, such as those observed in *Pigoraptor* and
613 *Capsaspora*. Consequently, TE expansion may reflect broader increases in developmental,
614 behavioral, or regulatory complexity across lineages representing intermediate stages of
615 multicellular evolution. Further investigation of TE organization and regulatory architecture
616 in *Pigoraptor* may therefore provide important insight into evolutionary transitions preceding
617 more stable multicellular systems.

bioRxiv preprint doi: https://doi.org/10.64898/2026.06.24.734286; this version posted June 25, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license.

Despite clear differences in overall TE proportion, TE diversity analyses revealed no universal relationship between multicellularity and increased TE diversity. Instead, Shannon and Simpson indices demonstrated strongly lineage-specific patterns. Within Heterolobosea, *Acrasis kona* exhibited substantially greater TE diversity than the compared *Naegleria* species, whereas *Tetramitus jugosus* displayed relatively high diversity despite very low total TE abundance. In Chlorophyceae, the unicellular alga *Chlamydomonas reinhardtii* exhibited the highest overall TE diversity among all examined assemblies despite containing lower total TE abundance than *Volvox carteri*. Similarly, within Filasterea, TE diversity patterns reflected distinct lineage-specific TE histories rather than a simple relationship with developmental complexity. Together, these results indicate that TE expansion within protist genomes frequently reflects amplification of particular TE lineages rather than uniform diversification across all TE categories.

This interpretation is further supported by the TE compositional analyses. The PCA demonstrated that species frequently diverged through expansion of specific TE groups rather than through complete restructuring of their TE repertoires. For example, the large TE burden observed in *Volvox carteri* was driven primarily by LTR retrotransposons, whereas rolling-circle elements contributed strongly to the distinctive TE landscape of *Capsaspora owczarzaki*. Conversely, *Chlamydomonas reinhardtii* retained a comparatively balanced TE composition despite lower total TE abundance. These findings suggest that TE evolutionary trajectories differ substantially among protist lineages and are shaped by lineage-specific amplification, retention, and turnover dynamics rather than by a single conserved genomic response associated with multicellularity.

Interestingly, not all aggregative systems exhibited substantial TE expansion. Although *Fonticula alba* forms multicellular fruiting bodies, its genome remains comparatively TE-poor and compositionally similar to several unicellular taxa, although some underestimation of TE content may reflect assembly incompleteness or fragmentation. This result indicates that transitions toward cooperative or multicellular behaviors can occur through distinct genomic routes and do not require major TE proliferation. Together, these observations support a model in which TE accumulation frequently accompanies multicellular or aggregative lifestyles but is not itself a prerequisite for the evolution of multicellular organization.

bioRxiv preprint doi: https://doi.org/10.64898/2026.06.24.734286; this version posted June 25, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license.

TE density and developmental regulation across protist systems

Comparative TE-density analyses revealed even stronger lineage-specific patterns. In *Acrasis kona*, upregulated and downregulated genes during aggregation versus germination transitions consistently exhibited relatively TE-depleted upstream environments compared to non-differentially expressed genes. Although the absolute differences in TE density were modest, these patterns were significant and remained consistent across multiple statistical approaches. Meanwhile, aggregation versus growth comparisons exhibited almost identical TE-density values among regulatory categories and produced no significant differences. This suggests that TE depletion is not uniformly associated with all developmental transitions in *A. kona*, but instead characterizes specific regulatory programs linked to aggregation and germination.

At first glance, the relatively low TE density surrounding regulated genes in *A. kona* may appear unexpected given extensive evidence showing that TEs can contribute to regulatory innovation in many eukaryotic systems (Chuong et al., 2017; Sundaram & Wysocka, 2020). However, even genes classified as TE-associated in *A. kona* remained only modestly enriched for nearby TE sequence in absolute terms, generally differing by only a few percent of upstream sequence composition. Thus, the observed patterns likely reflect subtle but evolutionarily meaningful differences in local genomic environments rather than extensive TE-driven restructuring of promoters or enhancers.

One possible interpretation is that genes requiring rapid and coordinated transcriptional responses during aggregation occupy comparatively TE-poor regulatory neighborhoods in order to preserve regulatory precision. Numerous studies across animals, fungi, and plants demonstrate that TE insertions near promoters or enhancers can alter transcription factor binding, chromatin accessibility, or local regulatory interactions (Feschotte, 2008; Bourque et al., 2018). Developmentally responsive genes may therefore experience stronger selective pressure inhibiting nearby TE accumulation because even modest regulatory perturbations could disrupt coordinated multicellular behavior. Within this framework, the TE-depleted regulatory landscapes observed in *A. kona* may reflect evolutionary constraints associated with maintaining robust aggregation-associated transcriptional programs. Future analyses examining gene family expansion and TE-

bioRxiv preprint doi: https://doi.org/10.64898/2026.06.24.734286; this version posted June 25, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license.

679 associated duplication events may further clarify whether TEs contribute indirectly to
680 multicellular evolution through the generation of novel gene copies and regulatory diversity.

681 Notably, *Capsaspora owczarzaki* exhibited comparatively weak and inconsistent TE-
682 density organization across developmental states involving aggregation. Some comparisons
683 showed slightly elevated TE density near upregulated genes, whereas others revealed only
684 minimal differences among regulatory categories. None of these contrasts reached strong
685 statistical significance. These results suggest that TE-associated regulatory architecture in
686 *Capsaspora* may be substantially less structured than in *Acrasis*. These differences may
687 reflect lineage-specific variation in regulatory organization and developmental gene
688 expression programs across aggregative protists.

689 The most contrasting pattern was observed in *Volvox carteri*. Unlike *A. kona*, somatic-
690 associated genes in *Volvox* exhibited comparatively elevated upstream TE density relative to
691 reproductive-associated genes. Importantly, this signal was strongest when all genes were
692 included and became substantially weaker when analyses were restricted to TE-positive loci.
693 This indicates that the primary pattern in *Volvox* is driven largely by whether genes occur
694 near TEs at all rather than by large differences in TE density once nearby TEs are already
695 present. Although the observed differences remained modest in absolute magnitude, the
696 consistent enrichment of TEs near somatic-associated genes suggests that differentiated
697 multicellularity in volvocine algae may involve a distinct relationship between
698 developmental regulation and genomic TE environments.

699 These results indicate that TE-associated regulatory organization is not evolutionarily
700 conserved across protist developmental systems. Instead, different lineages appear to have
701 evolved distinct genomic relationships with nearby TEs. In some systems, such as *Acrasis*
702 *kona*, developmental regulation is associated with relative exclusion of TEs from upstream
703 regulatory regions, whereas in others, such as *Volvox carteri*, developmental differentiation
704 may tolerate or even coincide with comparatively TE-enriched genomic environments.
705 Meanwhile, *Capsaspora owczarzaki* appears to occupy an intermediate condition
706 characterized by comparatively weak TE-density organization.

707 Importantly, the TE-density differences observed across all systems remained
708 relatively subtle in absolute magnitude, often differing by only approximately 2–4% of

bioRxiv preprint doi: https://doi.org/10.64898/2026.06.24.734286; this version posted June 25, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license.

upstream sequence composition. Consequently, these results should not be interpreted as evidence of extensive promoter restructuring driven by TEs. Rather, they support a model in which relatively small shifts in genomic TE organization may contribute to broader differences in regulatory architecture when acting across thousands of genes over evolutionary timescales.

Collectively, these findings emphasize that the contribution of transposable elements to developmental evolution is highly context-dependent. In some lineages, TEs may primarily function as genomic elements that must be selectively excluded from sensitive regulatory loci, whereas in others they may become integrated into broader regulatory landscapes associated with cellular differentiation and developmental complexity. These contrasting patterns illustrate that the evolutionary relationship between TEs and multicellularity cannot be explained through a single universal model but instead reflects diverse lineage-specific trajectories of genome evolution and regulatory organization.

Conclusion

This study combined comparative transposable element annotation with genome-wide analyses of TE density surrounding developmentally regulated genes across multiple protist lineages exhibiting distinct developmental and cellular organizations. Comparative analyses revealed that species exhibiting aggregative, cooperative, or differentiated developmental states generally contained higher overall TE proportions than representative unicellular taxa, although this trend was not universal across all lineages. TE diversity and TE composition additionally exhibited strong lineage-specific patterns, indicating that TE evolution in protists is shaped by distinct genomic trajectories rather than by a single conserved relationship with multicellularity.

Comparative TE-density analyses further demonstrated that relationships between TEs and developmental regulation differ substantially among protist systems. *Acrasis kona* exhibited reduced TE density near genes regulated during aggregation versus germination transitions, whereas *Volvox carteri* displayed comparatively elevated TE density near somatic-associated loci. At the same time, *Capsaspora owczarzaki* showed relatively weak and inconsistent TE-density organization across developmental states. Together, these

bioRxiv preprint doi: https://doi.org/10.64898/2026.06.24.734286; this version posted June 25, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license.

findings indicate that TE-associated regulatory landscapes are highly lineage-specific and do not follow a universal pattern associated with multicellular evolution.

Overall, this work emphasizes the importance of integrating comparative genomics and transcriptomics to better understand how transposable elements contribute to genome organization, regulatory architecture, and the evolution of developmental complexity across diverse eukaryotic lineages.

Acknowledgments

This work utilized computational resources provided by the Texas Tech University High Performance Computing Center (TTU HPCC). PromethION sequencing for *Acrasis kona* was supported by the National Science Foundation, Division of Environmental Biology, Biodiversity and Systematics Cluster (grant DEB-2100888). Additional institutional support was provided by the Department of Biological Sciences at Texas Tech University.

Data Availability

The genome assemblies analyzed in this study were obtained from publicly available repositories, including NCBI GenBank, the European Nucleotide Archive (ENA), and associated BioProjects under the accession numbers reported in Table 1. The newly generated Acrasis kona genome assembly has been deposited in NCBI under BioProject accession PRJNA1447523, BioSample accession SAMN57028165, and Whole Genome Shotgun accession JBWXZU000000000.

Transcriptomic datasets used for differential gene expression analyses were obtained from publicly available repositories. Accession numbers for all RNA-seq datasets are provided in Supplementary Table S1. The curated transposable element classification dataset generated during this study is available as Supplementary Table S2, and the complete curated transposable element consensus library used for RepeatMasker analyses is available as Supplementary File S1.

Conflict of Interest

The authors declare no competing interests.

bioRxiv preprint doi: https://doi.org/10.64898/2026.06.24.734286; this version posted June 25, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license.

766 **Bibliography:**

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

768 Bourque, G., Burns, K. H., Gehring, M., Gorbunova, V., Seluanov, A., Hammell, M.,
769 Imbeault, M., Izsvák, Z., Levin, H. L., Macfarlan, T. S., Mager, D. L., & Feschotte,
770 C. (2018). Ten things you should know about transposable elements. *Genome*
771 *Biology*, 19(1), 199. https://doi.org/10.1186/s13059-018-1577-z

772 Brown, M. W., Spiegel, F. W., & Silberman, J. D. (2009). Phylogeny of the "forgotten"
773 cellular slime mold, *Fonticula alba*, reveals a key evolutionary branch within
774 Opisthokonta. *Molecular Biology and Evolution, 26*(12), 2699–2709.
775 https://doi.org/10.1093/molbev/msp185

776 Brown, M. W., Silberman, J. D., & Spiegel, F. W. (2012). *A contemporary evaluation of the*
777 *acrasids (Acrasidae, Heterolobosea, Excavata). European Journal of Protistology*,
778 48(2), 103–123. https://doi.org/10.1016/j.ejop.2011.10.001

779 Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., & Madden,
780 T. L. (2009). BLAST+: Architecture and applications. *BMC Bioinformatics*, 10, 421.
781 https://doi.org/10.1186/1471-2105-10-421

782 Casacuberta, E., & González, J. (2013). The impact of transposable elements in
783 environmental adaptation. *Molecular Ecology*, 22(6), 1503–1517.
784 https://doi.org/10.1111/mec.12170

785 Chuong, E. B., Elde, N. C., & Feschotte, C. (2017). Regulatory activities of transposable
786 elements: from conflicts to benefits. *Nature Reviews Genetics*, 18(2), 71-86.
787 https://doi.org/10.1038/nrg.2016.139

788 Correa, M., Lerat, E., Birmelé, E., Samson, F., Bouillon, B., Normand, K., & Rizzon, C.
789 (2021). *The transposable element environment of human genes differs according to*
790 *their duplication status and essentiality*. Genome Biology and Evolution, 13(5),
791 evab062. https://doi.org/10.1093/gbe/evab062

792 Dereeper, A., Allouch, N., Guerlais, V., Garnier, M., Ma, L., De Jonckheere, J. F., Joseph, S.
793 J., Ali, I. K. M., Talarmin, A., & Marcelino, I. (2023). *Naegleria genus pangenome*

bioRxiv preprint doi: https://doi.org/10.64898/2026.06.24.734286; this version posted June 25, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license.

reveals new structural and functional insights into the versatility of these free-living amoebae. BMC Genomics, 24(1), 119. https://doi.org/10.1186/s12864-023-09179-9

Dubin, M. J., Mittelsten Scheid, O., & Becker, C. (2018). Transposons: a blessing curse. *Current Opinion in Plant Biology, 42*, 23–29. https://doi.org/10.1016/j.pbi.2018.01.003

Edgar, R. C. (2010). *Search and clustering orders of magnitude faster than BLAST.* Bioinformatics, 26(19), 2460–2461. https://doi.org/10.1093/bioinformatics/btq461

Elbarbary, R. A., Lucas, B. A., & Maquat, L. E. (2016). Retrotransposons as regulators of gene expression. *Science, 351*(6274), aac7247. https://doi.org/10.1126/science.aac7247

Etchegaray, E., & Naville, M. (2021). Transposable element–derived sequences in vertebrate regulatory landscapes. *Mobile DNA, 12*, 6. https://doi.org/10.1186/s13100-020-00229-5

Fairclough, S. R., Chen, Z., Kramer, E., Zeng, Q., Young, S., Robertson, H. M., ... King, N. (2013). Premetazoan genome evolution and the regulation of cell differentiation in *Salpingoeca rosetta*. *Genome Biology*, 14(2), R15. https://doi.org/10.1186/gb-2013-14-2-r15

Feng, J., Meyer, C. A., Wang, Q., Liu, J. S., Shirley Liu, X., & Zhang, Y. (2012). *GFOLD: A generalized fold change for ranking differentially expressed genes from RNA-seq data.* Bioinformatics, 28(21), 2782–2788.

Feschotte, C. (2008). Transposable elements and the evolution of regulatory networks. *Nature Reviews Genetics*, 9(5), 397–405. https://doi.org/10.1038/nrg2337

Fiston-Lavier, A.-S., Quesneville, H., & Vejnar, C. (2012). Transposable element sequence evolution is influenced by gene context. *arXiv preprint arXiv:1209.0176*. https://doi.org/10.48550/arXiv.1209.0176

Flynn, J. M., Hubley, R., Goubert, C., Rosen, J., Clark, A. G., Feschotte, C., & Smit, A. F. A. (2020). *RepeatModeler2 for automated genomic discovery of transposable element families.* PNAS, 117(17), 9451–9457. https://doi.org/10.1073/pnas.1921046117

bioRxiv preprint doi: https://doi.org/10.64898/2026.06.24.734286; this version posted June 25, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license.

822 Fortunato, A. S., Adamska, M., Adamski, M., ... Degnan, B. M. (2014). Calcisponges have a
823 ParaHox gene and dynamic expression of dispersed NK homeobox genes. *Nature*,
824 514(7524), 620–623. https://doi.org/10.1038/nature13881

825 Fritz-Laylin, L. K., Ginger, M. L., Walsh, C., Dawson, S. C., Fulton, C., & Cande, W. Z.
826 (2010). *The Naegleria genome: A free-living microbial eukaryote lends unique*
827 *insights into core eukaryotic cell biology. Cell*, 140(5), 631–642.
828 https://doi.org/10.1016/j.cell.2010.01.032

829 Gebrie, S. A. (2023). Transposable elements as essential elements in the control of gene
830 expression. *Mobile DNA, 14*, 4.
831 https://doi.org/10.1186/s13100-023-00297-3

832 Gabriel, L., Hoff, K. J., Brůna, T., Borodovsky, M., & Stanke, M. (2023). *BRAKER3: Fully*
833 *automated genome annotation using RNA-Seq and protein evidence.* Genome
834 Biology, 24(1), 34.

835 Gurevich, A., Saveliev, V., Vyahhi, N., & Tesler, G. (2013). QUAST: Quality assessment tool
836 for genome assemblies. *Bioinformatics*, 29(8), 1072–1075.
837 https://doi.org/10.1093/bioinformatics/btt086

838 Hanschen, E. R., Chen, Z., Pruitt, J. N., & Ferris, P. J. (2017). Whole transcriptome RNA-Seq analysis
839 reveals extensive cell type-specific compartmentalization in *Volvox carteri*. *BMC Biology, 15*,
840 111. https://doi.org/10.1186/s12915-017-0450-y

841 Hanschen, E. R., Herron, M. D., Wiens, J. J., Nozaki, H., & Michod, R. E. (2018).
842 Multicellularity drives the evolution of sexual traits. *The American Naturalist, 192*(3),
843 E93–E105. https://doi.org/10.1086/698301

844 Hehenberger, E., Tikhonenkov, D. V., Kolisko, M., del Campo, J., Esaulov, A. S., Mylnikov,
845 A. P., & Keeling, P. J. (2017). *Novel predators reshape holozoan phylogeny and reveal*
846 *the presence of two distinct modes of multicellularity. Current Biology*, 27(13), 2043–
847 2050. https://doi.org/10.1016/j.cub.2017.06.006

848 Hehenberger, E., Tikhonenkov, D. V., Kolisko, M., Del Campo, J., Esaulov, A. S., Mylnikov,
849 A. P., & Keeling, P. J. (2017). *Insights into the origin of metazoan multicellularity*

bioRxiv preprint doi: https://doi.org/10.64898/2026.06.24.734286; this version posted June 25, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license.

from predatory unicellular relatives of animals. BMC Biology, 15, 6. https://doi.org/10.1186/s12915-017-0357-x

Herron, M. D. (2016). *Origins of multicellular complexity: Volvox and the volvocine algae. Molecular Ecology*, 25(6), 1213-1223. https://doi.org/10.1111/mec.13551

Hu, K., Ni, P., Xu, M., Zou, Y., Chang, J., Gao, X., Li, Y., Ruan, J., Hu, B., & Wang, J. (2024). HiTE: A fast and accurate dynamic boundary adjustment approach for full-length transposable element detection and annotation. *Nature Communications, 15*, 5573. https://doi.org/10.1038/s41467-024-49912-8

Jachowicz, J. W., Bing, X., Pontabry, J., Bošković, A., Rando, O. J., & Torres-Padilla, M. E. (2017). LINE-1 activation after fertilization regulates global chromatin accessibility in the early mouse embryo. *Nature Genetics, 49*(10), 1502–1510. https://doi.org/10.1038/ng.3945

Jones, P., et al. (2014). *InterProScan 5: Genome-scale protein function classification.* Bioinformatics, 30(9), 1236–1240

Kim, K.-S., Kustu, S., & Inwood, W. (2006). *Natural history of transposition in the green alga Chlamydomonas reinhardtii: Use of the AMT4 locus as an experimental system. Genetics, 173*(4), 2005–2019. https://doi.org/10.1534/genetics.106.058263

King, N. (2004). The unicellular ancestry of animal development. *Developmental Cell*, 7(3), 313–325. https://doi.org/10.1016/j.devcel.2004.08.010

King, N., Westbrook, M. J., Young, S. L., Kuo, A., Abedin, M., Chapman, J., ... Rokhsar, D. S. (2008). The genome of the choanoflagellate *Monosiga brevicollis* and the origin of metazoans. *Nature*, 451(7180), 783–788. https://doi.org/10.1038/nature06617

Klein, B., Wibberg, D., y Hallmann, A. (2017). Whole transcriptome RNA-Seq analysis reveals extensive cell type-specific compartmentalization in *Volvox carteri*. *BMC Biology*, *15*(111). https://doi.org/10.1186/s12915-017-0450-y

Kolmogorov, M., Yuan, J., Lin, Y., & Pevzner, P. A. (2019). Assembly of long, error-prone reads using repeat graphs. *Nature Biotechnology*, 37(5), 540–546. https://doi.org/10.1038/s41587-019-0072-8

bioRxiv preprint doi: https://doi.org/10.64898/2026.06.24.734286; this version posted June 25, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license.

878 Langmüller, A. M., Nolte, V., Dolezal, M., & Schlötterer, C. (2023). *The genomic distribution*
879 *of transposable elements is driven by spatially variable purifying selection*. Nucleic
880 Acids Research, 51(17), 9203–9213. https://doi.org/10.1093/nar/gkad635

881 Lefebvre, P. A., & Silflow, C. D. (1999). *Chlamydomonas: The cell and its genomes.*
882 Genetics, 151(1), 9–14. https://doi.org/10.1093/genetics/151.1.9

883 Li, H. (2018). *Minimap2: Pairwise alignment for nucleotide sequences.* Bioinformatics,
884 34(18), 3094–3100.

885 Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., … & 1000 Genome
886 Project Data Processing Subgroup. (2009). The Sequence Alignment/Map format and
887 SAMtools. *Bioinformatics*, 25(16), 2078–2079.

888 https://doi.org/10.1093/bioinformatics/btp352

889 Li, X., Yang, Z., & Liu, Y. (2024). *HELIANO: A deep-learning-based framework for accurate*
890 *identification and annotation of Helitron transposons.* Bioinformatics, 40(2),
891 btae012.

892 Liao, Y., Smyth, G. K., & Shi, W. (2014). *featureCounts: An efficient general purpose*
893 *program for assigning sequence reads to genomic features.* Bioinformatics, 30(7),
894 923–930.

895 Liechti, N., Schürch, N., Bruggmann, R., & Wittwer, M. (2018). The genome of *Naegleria*
896 *lovaniensis*, the basis for a comparative approach to unravel pathogenicity factors of
897 the human pathogenic amoeba *N. fowleri*. *BMC Genomics, 19*, 654.
898 https://doi.org/10.1186/s12864-018-4994-1

899 López-Escardó, D., Torruella, G., Moreira, D., López-García, P., & Ruiz-Trillo, I. (2017).
900 *Parvularia atlantis gen. et sp. nov., a nucleariid filose amoeba (Holomycota,*
901 *Opisthokonta)*. *Journal of Eukaryotic Microbiology, 65*(2), 170–181.
902 https://doi.org/10.1111/jeu.12447

903 Love, M. I., Huber, W., & Anders, S. (2014). Moderated estimation of fold change and
904 dispersion for RNA-seq data with DESeq2. *Genome Biology, 15*(12), 550.
905 https://doi.org/10.1186/s13059-014-0550-8

bioRxiv preprint doi: https://doi.org/10.64898/2026.06.24.734286; this version posted June 25, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license.

906 Manni, M., Berkeley, M. R., Seppey, M., Simão, F. A., & Zdobnov, E. M. (2021). BUSCO
907 Update: Novel and streamlined workflows along with broader and deeper
908 phylogenetic coverage for scoring of eukaryotic, prokaryotic, and viral genomes.
909 *Molecular Biology and Evolution*, 38(10), 4647–4654.
910 https://doi.org/10.1093/molbev/msab199

911 McGowan, J., et al. (2025). Comparative single-cell genomics of two uncultivated *Naegleria*
912 spp. *mSphere, 10*(x), e00352-25. https://doi.org/10.1128/msphere.00352-25b

913 McKinney, W. (2010). *Data structures for statistical computing in Python.* In Proceedings of
914 the 9th Python in Science Conference (SciPy 2010), 51–56.

915 Merchant, S. S., Prochnik, S. E., Vallon, O., Harris, E. H., Karpowicz, S. J., Witman, G. B.,
916 Terry, A., Salamov, A., Fritz-Laylin, L. K., Maréchal-Drouard, L., et al. (2007). *The*
917 *Chlamydomonas genome reveals the evolution of key animal and plant functions.*
918 *Science*, 318(5848), 245–250. https://doi.org/10.1126/science.1143609

919 Mokhtar, M. M., & El Allali, A. (2023). *MegaLTR: a web server and standalone pipeline for*
920 *detecting and annotating LTR-retrotransposons in plant genomes.* Frontiers in Plant
921 Science, 14, 1237426.

922 Moroz, L. L., Kocot, K. M., Citarella, M. R., Dosung, S., Norekian, T. P., Povolotskaya, I.
923 S., ... Kohn, A. B. (2014). The ctenophore genome and the evolutionary origins of
924 neural systems. *Nature*, 510(7503), 109–114. https://doi.org/10.1038/nature13400

925 Muotri, A. R., Chu, V. T., Marchetto, M. C., Deng, W., Moran, J. V., & Gage, F. H. (2005).
926 Somatic retrotransposition alters the genetic landscape of the human brain. *Nature,*
927 *435*(7044), 903–910. https://doi.org/10.1038/nature03663

928 National Center for Biotechnology Information (NCBI). (n.d.). *Naegleria gruberi* genome
929 assembly (GCF_000004985.1). Recuperado de
930 https://www.ncbi.nlm.nih.gov/datasets/genome/GCF_000004985.1/

931 National Center for Biotechnology Information (NCBI). (n.d.). *Naegleria fowleri* genome
932 assembly (GCA_000499105.1). Recuperado de
933 https://www.ncbi.nlm.nih.gov/datasets/genome/GCA_000499105.1/

bioRxiv preprint doi: https://doi.org/10.64898/2026.06.24.734286; this version posted June 25, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license.

Nawrocki, E. P., & Eddy, S. R. (2013). *Infernal 1.1: 100-fold faster RNA homology searches.* Bioinformatics, 29(22), 2933–2935.

Nellåker, C., Keane, T. M., Yalcin, B., Wong, K., Agam, A., Belgard, T. G., Flint, J., Adams, D. J., & Jackson, I. J. (2012). *The genomic landscape shaped by selection on transposable elements across 18 mouse strains*. Genome Biology, 13(6), R45. https://doi.org/10.1186/gb-2012-13-6-r45

Ocaña-Pallarès, E., Williams, T. A., López-Escardó, D., Arroyo, A. S., Pathmanathan, J. S., Bapteste, E., Tikhonenkov, D. V., Keeling, P. J., Szöllősi, G. J., & Ruiz-Trillo, I. (2022). Divergent genomic trajectories predate the origin of animals and fungi. *Nature, 609*(7928), 747–753. https://doi.org/10.1038/s41586-022-05110-4

Orozco-Arias, S., Liu, J., Tabares-Soto, R., Ceballos, D., Domingues, D. S., Garavito, A., Ming, R., & Guyot, R. (2018). *Inpactor: Integrated and Parallel Analyzer and Classifier of LTR Retrotransposons and Its Application for Pineapple LTR Retrotransposons Diversity and Dynamics.* Biology (Basel), 7(2), 32.

Oxford Nanopore Technologies. (2021). *Medaka: sequence correction provided by ONT Research.* Retrieved from https://github.com/nanoporetech/medaka

Pánek, T., et al. (2017). *Phylogeny of Heterolobosea revealed by multigene analyses*. *Molecular Phylogenetics and Evolution*, 112, 70–80.

Parfrey, L. W., & Lahr, D. J. G. (2013). Multicellularity arose several times in the evolution of eukaryotes. *BioEssays, 35*(4), 339–347. https://doi.org/10.1002/bies.201200143

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., & Duchesnay, É. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research, 12*, 2825–2830.

Percharde, M., Lin, C.-J., Yin, Y., Guan, J., Peixoto, G. A., Bulut-Karslioglu, A., Biechele, S., Huang, B., Shen, X., & Ramalho-Santos, M. (2018). A LINE1–nucleolin partnership regulates early development and ESC identity. *Cell, 174*(2), 391–405.e19. https://doi.org/10.1016/j.cell.2018.05.043

bioRxiv preprint doi: https://doi.org/10.64898/2026.06.24.734286; this version posted June 25, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license.

962 Pereira, V., Enard, D., & Eyre-Walker, A. (2009). The effect of transposable element
963 insertions on gene expression evolution in rodents. *PLoS ONE, 4*(2), e4321.
964 https://doi.org/10.1371/journal.pone.0004321

965 Pourrajab, F., & Hekmatimoghaddam, S. (2021). Transposable elements, contributors in the
966 evolution of organisms (from an arms race to a source of raw materials). *Heliyon*, 7,
967 e06029. https://doi.org/10.1016/j.heliyon.2021.e06029

968 Putnam, N. H., Srivastava, M., Hellsten, U., Dirks, B., Chapman, J., Salamov, A., ... Rokhsar, D. S.
969 (2007). Sea anemone genome reveals ancestral eumetazoan gene repertoire and genomic
970 organization. *Science*, 317(5834), 86–94. https://doi.org/10.1126/science.1139158

971 Prochnik, S. E., Umen, J., Nedelcu, A. M., Hallmann, A., Miller, S. M., Nishii, I., Ferris, P.,
972 Kuo, A., Mitros, T., Fritz-Laylin, L. K., et al. (2010). *Genomic analysis of organismal*
973 *complexity in the multicellular green alga Volvox carteri. Science*, 329(5988), 223–
974 226. https://doi.org/10.1126/science.1188800

975 Qiu, Y.,, Nei, M., & Zhang, Y. (2021). Whole-genome variation of transposable element
976 insertions in maize. *G3: Genes, Genomes, Genetics, 11*(10), jkab238.
977 https://doi.org/10.1093/g3journal/jkab238

978 Quinlan, A. R., & Hall, I. M. (2010). BEDTools: A flexible suite of utilities for comparing
979 genomic features. *Bioinformatics*, 26(6), 841–842.
980 https://doi.org/10.1093/bioinformatics/btq033

981 Rebollo, R., Romanish, M. T., & Mager, D. L. (2012). Transposable elements: An abundant
982 and natural source of regulatory sequences for host genes. *Annual Review of Genetics,*
983 *46*, 21–42. https://doi.org/10.1146/annurev-genet-110711-155621

984 Rokas, A. (2008). The origins of multicellularity and the early history of the genetic toolkit
985 for animal development. *Annual Review of Genetics*, 42, 235–251.
986 https://doi.org/10.1146/annurev.genet.42.110807.091513

987 Ruiz-Trillo, I., Burger, G., Holland, P. W. H., King, N., Lang, B. F., Roger, A. J., & Gray, M.
988 W. (2007). The origins of multicellularity: A multi-taxon genome initiative. *Trends in*
989 *Genetics*, 23(3), 113–118.

bioRxiv preprint doi: https://doi.org/10.64898/2026.06.24.734286; this version posted June 25, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license.

990 Sebé-Pedrós, A., Irimia, M., del Campo, J., Parra-Acero, H., Russ, C., Nusbaum, C.,
991 Blencowe, B. J., & Ruiz-Trillo, I. (2013). Regulated aggregative multicellularity in a
992 close unicellular relative of metazoa. ***eLife***, *2*, e01287.
993 https://doi.org/10.7554/eLife.01287

994 Sebé-Pedrós, A., Degnan, B. M., & Ruiz-Trillo, I. (2017). The origin of Metazoa: A
995 unicellular perspective. *Nature Reviews Genetics, 18*(8), 498–512.
996 https://doi.org/10.1038/nrg.2017.21

997 Shannon, C. E. (1948). A mathematical theory of communication. Bell System Technical
998 Journal, 27(3), 379–423.

999 Sheikh, S., Brown, M. W., Kang, S., et al. (2024). *The Acrasis kona genome and*
1000 *developmental transcriptomes reveal deep origins of eukaryotic multicellular*
1001 *pathways. Nature Communications*, 15, 269. https://doi.org/10.1038/s41467-024-
1002 45190-7

1003 Shen, W., Le, S., Li, Y., & Hu, F. (2016). SeqKit: A cross-platform and ultrafast toolkit for
1004 FASTA/Q file manipulation. *PLOS ONE*, 11(10), e0163962.
1005 https://doi.org/10.1371/journal.pone.0163962

1006 Simpson, E. H. (1949). Measurement of diversity. Nature, 163(4148), 688.

1007 Smit, A. F. A., Hubley, R., & Green, P. (2013–2024). *RepeatMasker Open-4.0.*
1008 http://www.repeatmasker.org

1009 Sundaram, V., & Wysocka, J. (2020). Transposable elements as a potent source of diverse
1010 cis-regulatory sequences in mammalian genomes. *Philosophical Transactions of the*
1011 *Royal Society B: Biological Sciences*, 375(1795), 20190347.
1012 https://doi.org/10.1098/rstb.2019.0347

1013 Suga, H., Chen, Z., de Mendoza, A., Sebé-Pedrós, A., Brown, M. W., Kramer, E., Carr, M.,
1014 Kerner, P., Vervoort, M., Sánchez-Pons, N., Torruella, G., Derelle, R., Manning, G.,
1015 & Ruiz-Trillo, I. (2013). *The Capsaspora genome reveals a complex unicellular*
1016 *prehistory of animals. Nature Communications*, 4, 2325.
1017 https://doi.org/10.1038/ncomms3325

bioRxiv preprint doi: https://doi.org/10.64898/2026.06.24.734286; this version posted June 25, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license.

1018 Srivastava, M., Begovic, E., Chapman, J., Putnam, N. H., Hellsten, U., Kawashima, T., ...
1019 Rokhsar, D. S. (2008). The Trichoplax genome and the nature of placozoans. *Nature*,
1020 454, 955–960. https://doi.org/10.1038/nature07191

1021 Srivastava, M., Simakov, O., Chapman, J., Fahey, B., Gauthier, M. E. A., Mitros, T., ...
1022 Rokhsar, D. S. (2010). The *Nematostella vectensis* genome and the evolution of
1023 multicellularity. *Nature*, 468(7326), 693–698. https://doi.org/10.1038/nature09633

1024 Teresi, S. J., Teresi, M. B., & Edger, P. P. (2022). *TE Density: A tool to investigate the biology*
1025 *of transposable elements.* Mobile DNA, 13, 11. https://doi.org/10.1186/s13100-022-
1026 00268-1

1027 Tice, A. K., Brown, M. W., & colleagues. (2022). *The cellular slime mold Fonticula alba*
1028 *forms a dynamic, multicellular collective while feeding on bacteria*. *Current Biology,*
1029 *32*(9), 1961–1973.e4. https://doi.org/10.1016/j.cub.2022.03.018

1030 Ullastres, A., Méndez-González, J., & González, J. (2021). Regulatory regions in natural
1031 transposable element insertions drive inter-individual differences in response to
1032 infection in *Drosophila*. *Genome Biology, 22*, 280.
1033 https://doi.org/10.1186/s13059-021-02471-3

1034 Umen, J. G. (2014). Green algae and the origins of multicellularity in the plant kingdom.
1035 *Cold Spring Harbor Perspectives in Biology, 6*(11), a016170.
1036 https://doi.org/10.1101/cshperspect.a016170

1037 Virtanen, P., Gommers, R., Oliphant, T. E., et al. (2020). *SciPy 1.0: Fundamental algorithms*
1038 *for scientific computing in Python.* Nature Methods, 17(3), 261–272.
1039 https://doi.org/10.1038/s41592-019-0686-2

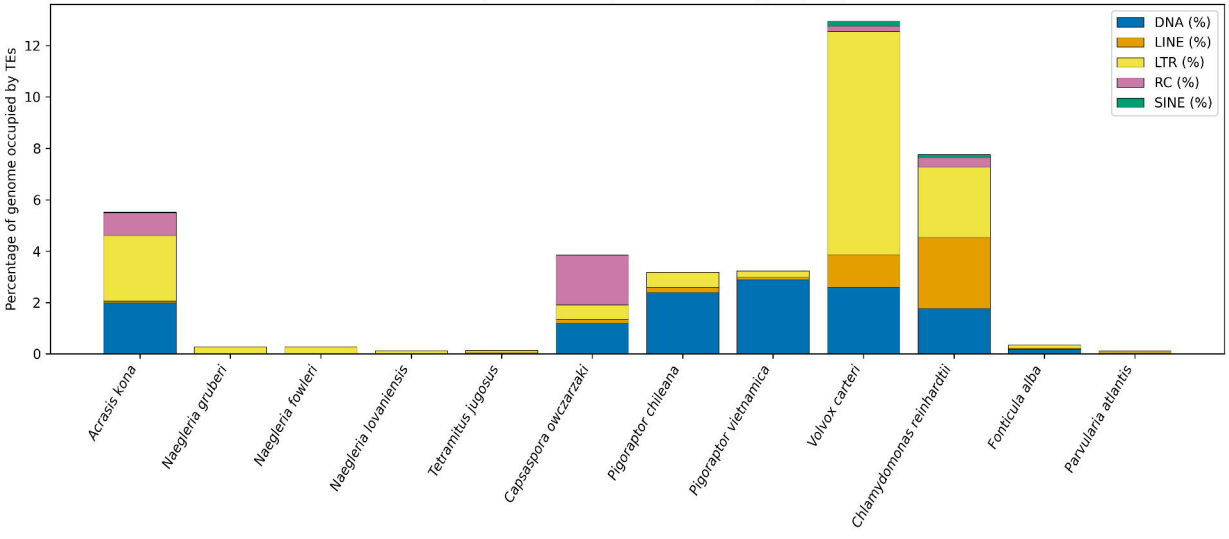
1040 Watson, J. D. (2014). *Molecular Biology of the Gene* (7th ed.). Pearson.

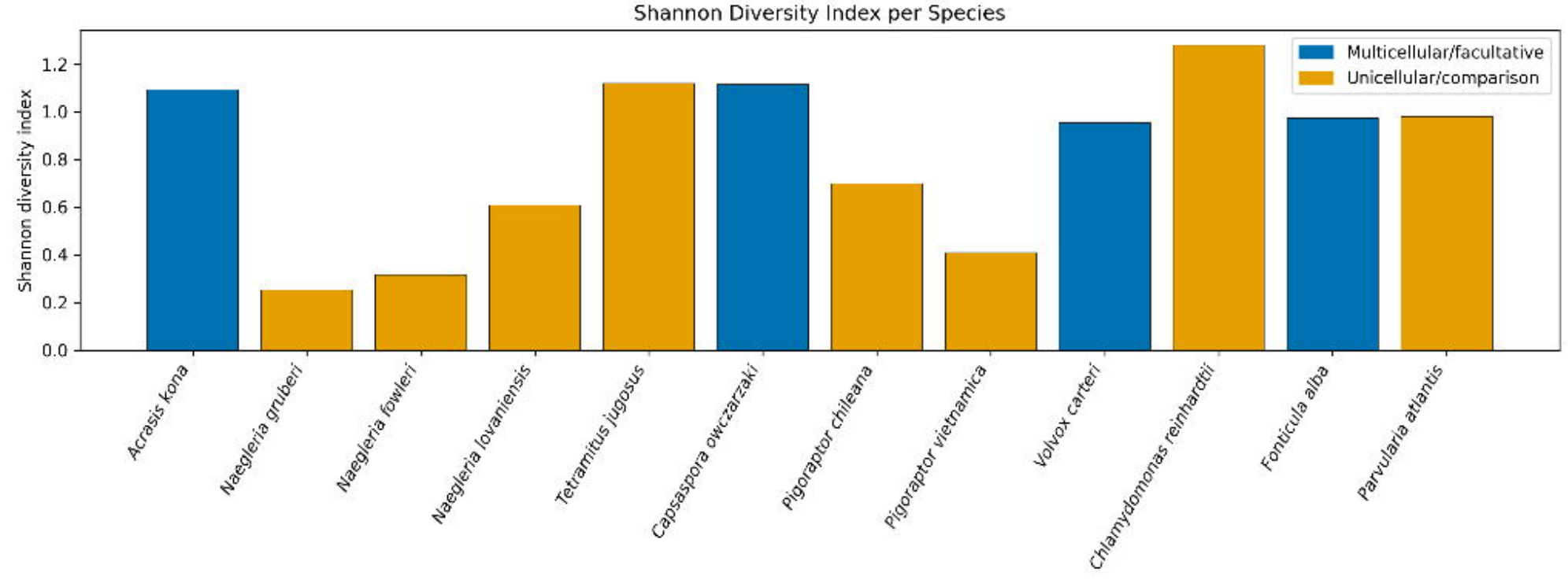
1041 Wessler, S. R. (2006). Transposable elements and the evolution of eukaryotic genomes.
1042 *Proceedings of the National Academy of Sciences*, 103(47), 17600–17601.
1043 https://doi.org/10.1073/pnas.0607612103

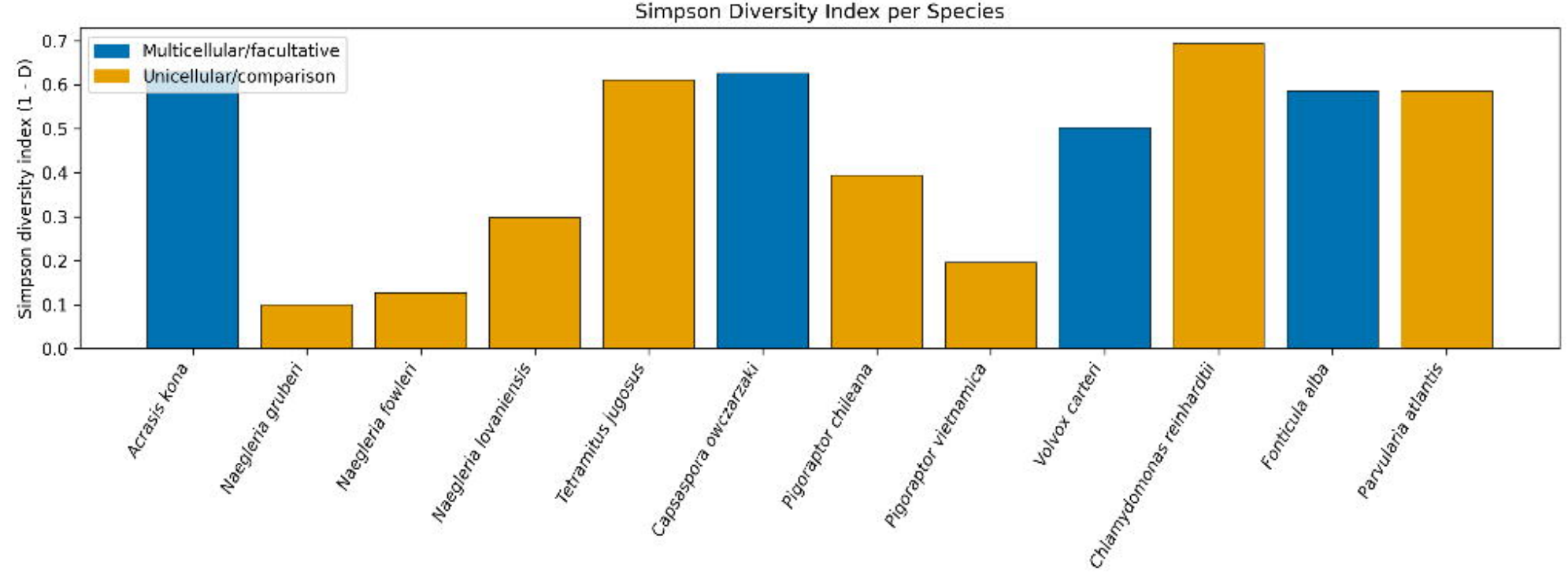
1044 Wilcoxon, F. (1945). *Individual comparisons by ranking methods.* Biometrics Bulletin, 1(6),
1045 80–83. https://doi.org/10.2307/3001968

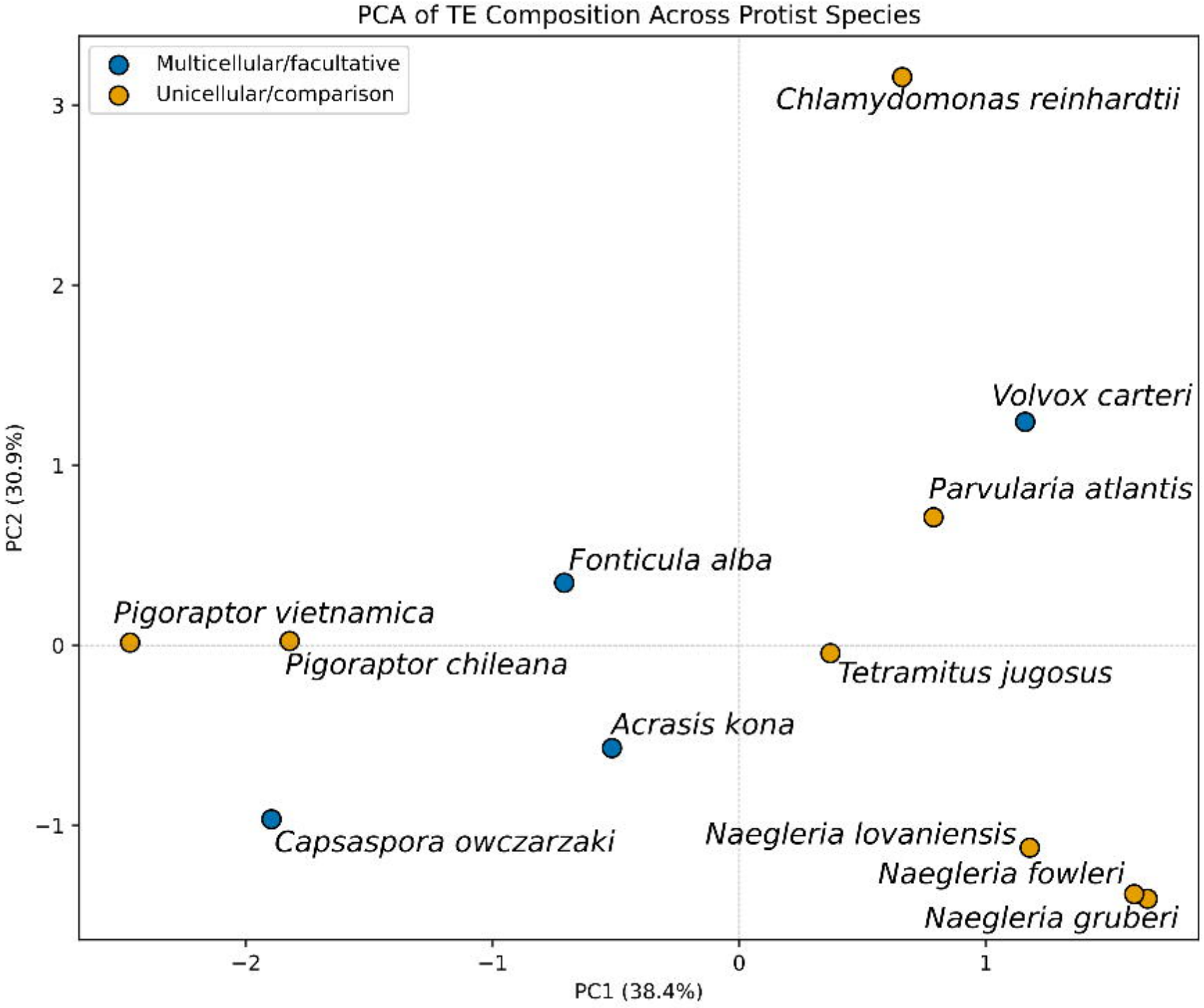
bioRxiv preprint doi: https://doi.org/10.64898/2026.06.24.734286; this version posted June 25, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license.

Zysset-Burri, D. C., Müller, N., Beuret, C., Heller, M., Schürch, N., Gottstein, B., & Wittwer, M. (2014). Genome-wide identification of pathogenicity factors of the free-living amoeba *Naegleria fowleri*. *BMC Genomics, 15*, 496. https://doi.org/10.1186/1471-2164-15-496


Genomic Proportion of Transposable Element Types per Species
Percentage of genome occupied by TEs
0
2
4
6
8
10
12
DNA (%)
LINE (%)
LTR (%)
RC (%)
SINE (%)
Acrasis kona
Naegleria gruberi
Naegleria fowleri
Naegleria lovaniensis
Tetramitus jugosus
Capsaspora owczarzaki
Pigoraptor chileana
Pigoraptor vietnamica
Volvox carteri
Chlamydomonas reinhardtii
Fonticula alba
Parvularia atlantis


Shannon Diversity Index per Species
Shannon diversity index
0.0
0.2
0.4
0.6
0.8
1.0
1.2
Multicellular/facultative
Unicellular/comparison
Acrasis kona
Naegleria gruberi
Naegleria fowleri
Naegleria lovaniensis
Tetramitus jugosus
Capsaspora owczarzaki
Pigoraptor chileana
Pigoraptor vietnamica
Volvox carteri
Chlamydomonas reinhardtii
Fonticula alba
Parvularia atlantis


Simpson Diversity Index per Species
Simpson diversity index (1 - D)
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
Multicellular/facultative
Unicellular/comparison
Acrasis kona
Naegleria gruberi
Naegleria fowleri
Naegleria lovaniensis
Tetramitus jugosus
Capsaspora owczarzaki
Pigoraptor chileana
Pigoraptor vietnamica
Volvox carteri
Chlamydomonas reinhardtii
Fonticula alba
Parvularia atlantis


PCA of TE Composition Across Protist Species
Multicellular/facultative
Unicellular/comparison
Chlamydomonas reinhardtii
Volvox carteri
Parvularia atlantis
Fonticula alba
Pigoraptor vietnamica
Pigoraptor chileana
Tetramitus jugosus
Acrasis kona
Capsaspora owczarzaki
Naegleria lovaniensis
Naegleria fowleri
Naegleria gruberi
PC1 (38.4%)
PC2 (30.9%)
3
2
1
0
−1
−2
−1
0
1

bioRxiv preprint doi: https://doi.org/10.64898/2026.06.24.734286; this version posted June 25, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license.

Table 1. Genome assemblies analyzed in this study: Genome assemblies representing unicellular, multicellular, and facultatively multicellular protists were used to compare TE diversity and proportion across lineages. The *Acrasis kona* genome was newly assembled in this study from ONT PromethION data, while all other assemblies were retrieved from the NCBI Genome database. Accession numbers for each genome are provided.

Taxonomic Association	Species	Lifestyle	Genome Accession
Chlorophyceae	*Chlamydomonas reinhardtii*	Unicellular	GCA_000002595.3
	Volvox carteri	Multicellular	GCA_000143455.1
Cristidiscoidea	*Fonticula alba*	Multicellular facultative	GCA_000388065.2
	Parvularia atlantis	Unicellular	GCA_943704415.1
Filasterea	*Capsaspora owczarzaki*	Multicellular facultative	GCA_000151315.2
	Pigoraptor chileana	Unicellular	PRJEB52884
	Pigoraptor vietnamica	Unicellular	PRJEB52884
Heterolobosea	*Acrasis kona*	Multicellular facultative	This study (ONT PromethION assembly)
	Tetramitus jugosus	Unicellular	PRJEB50460
	Naegleria gruberi	Unicellular	GCA_000004985.1
	Naegleria fowleri	Unicellular	GCA_014843625.1
	Naegleria lovaniensis	Unicellular	GCA_022530875.1

bioRxiv preprint doi: https://doi.org/10.64898/2026.06.24.734286; this version posted June 25, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license.

Table 4. Upstream TE-density patterns across developmental regulatory comparisons in protist systems. Mean upstream TE-density values calculated within 5 kb windows surrounding genes belonging to distinct developmental regulatory categories in *Acrasis kona*, *Capsaspora owczarzaki*, and *Volvox carteri*. All analyses were performed using the complete set of annotated genes ("All genes" dataset). Comparisons include aggregation-associated transitions in *A. kona* and *C. owczarzaki* as well as reproductive versus somatic differentiation in *V. carteri*. For each comparison, the table reports the number of genes assigned to each regulatory category, mean TE-density values, overall TE-density pattern, and statistical significance. The results reveal lineage-specific relationships between TE density and developmental gene regulation, including TE depletion near regulated genes in *A. kona*, relatively weak and variable patterns in *C. owczarzaki*, and increased TE density near somatic-associated genes in *V. carteri*.

Species	Developmental comparison	Dataset	Regulated gene category	Mean TE density	Main pattern	Significance
Acrasis kona	Aggregation vs Growth	All genes	Upregulated	0.046	weak differences	NS
			Downregulated	0.044		
			NDE	0.045		
Acrasis kona	Aggregation vs Germination	All genes	Upregulated	0.032	NDE > regulated	p < 1e-5
			Downregulated	0.026		
			NDE	0.053		
Capsaspora owczarzaki	Aggregative vs Adherent	All genes	Upregulated	0.04	Upregulated > NDE	NS
			Downregulated	0.029		
			NDE	0.032		
Capsaspora owczarzaki	Aggregative vs Filopodial	All genes	Upregulated	0.031	weak differences	NS
			Downregulated	0.033		
			NDE	0.032		
Volvox carteri	Reproductive vs Somatic	All genes	Reproductive-upregulated	0.067	Somatic > NDE > Reproductive	p < 0.05
			Somatic-upregulated	0.1		
			NDE	0.092		

bioRxiv preprint doi: https://doi.org/10.64898/2026.06.24.734286; this version posted June 25, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license.

Table 2. Assembly quality metrics for *Acrasis kona:* The original assembly from Sheikh et al. (2024) is compared with the new long-read assembly generated in this study. Metrics include contiguity statistics (N50, L50), total assembly length, GC content, ambiguous base frequency, and BUSCO completeness scores. The new assembly shows major improvements in contiguity and completeness, reflecting more accurate recovery of genomic structure.

Category	Metric	*A. kona* (Sheikh et al., 2024)	*A. kona* (this study)
QUAST	N50	42,575	262,724
	N90	7,953	70,789
	auN	61,431	390,397
	L50	262	47
	L90	1,274	190
	# contigs ≥ 0 bp	2,039	363
	# contigs ≥ 1 kb	2,039	351
	# contigs ≥ 5 kb	1,874	336
	# contigs ≥ 10 kb	1,021	314
	# contigs ≥ 25 kb	467	268
	# contigs ≥ 50 kb	220	224
	Total contigs	2,039	359
	Largest contig	259,475 bp	1,572,856 bp
	Total length (≥0 bp)	44,020,442 bp	47,827,436 bp
	GC (%)	38.89%	38.80%
	N's per 100 kbp	7.56	0.21
BUSCO	Complete BUSCOs (C)	85.5% (218)	85.1% (217)
	Single-copy BUSCOs (S)	74.9% (191)	72.9% (186)
	Duplicated BUSCOs (D)	10.6% (27)	12.2% (31)
	Fragmented BUSCOs (F)	7.5% (19)	5.1% (13)
	Missing BUSCOs (M)	7.0% (18)	9.8 % (25)
	Total BUSCOs (n)	255	255

bioRxiv preprint doi: https://doi.org/10.64898/2026.06.24.734286; this version posted June 25, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license.

Table 3. Proportion of transposable elements by class across unicellular and multicellular protist genomes: Proportional genomic content of major transposable element (TE) classes—including DNA transposons, LINEs, LTR retrotransposons, rolling-circle Helitrons, and SINEs—is shown for each unicellular and multicellular (or aggregatively multicellular) protist species analyzed. Values represent the percentage of the genome masked by each TE class based on RepeatMasker screening using the curated TE library generated in this study. The table highlights substantial lineage-specific variation in both total TE content and class composition across protist genomes.

Species genome assemblies	Taxonomic Group	Type	LTR (%)	DNA (%)	LINE (%)	RC (%)	SINE (%)	Total (%)
Acrasis kona	Heterolobosea	Multicellular (fac.)	2.55	2.00	0.06	0.89	0.02	5.52
Naegleria gruberi	Heterolobosea	Unicellular	0.27	0.01	0.00	0.00	0.00	0.28
Naegleria fowleri	Heterolobosea	Unicellular	0.26	0.01	0.01	0.00	0.00	0.28
Naegleria lovaniensis	Heterolobosea	Unicellular	0.10	0.01	0.00	0.00	0.00	0.13
Tetramitus jugosus	Heterolobosea	Unicellular	0.08	0.03	0.02	0.01	0.00	0.14
Capsaspora owczarzaki	Filasterea	Multicellular (fac.)	0.57	1.20	0.14	1.94	0.00	3.86
Pigoraptor chileana	Filasterea	Unicellular	0.59	2.40	0.19	0.00	0.00	3.18
Pigoraptor vietnamica	Filasterea	Unicellular	0.24	2.89	0.10	0.00	0.00	3.24
Volvox carteri	Chlorophyceae	Multicellular	8.67	2.58	1.29	0.22	0.19	12.96
Chlamydomonas reinhardtii	Chlorophyceae	Unicellular	2.75	1.77	2.76	0.37	0.12	7.76
Fonticula alba	Cristidiscoidea	Multicellular (fac.)	0.13	0.19	0.04	0.00	0.00	0.36
Parvularia atlantis	Cristidiscoidea	Unicellular	0.06	0.02	0.03	0.00	0.00	0.11