

RESEARCH

Open Access



Chromosome-level genome assembly of *Eimeria tenella* at the single-oocyst level

Kaihui Zhang^{1,2,3}, Yudong Cai⁴, Yuancai Chen^{1,2,3}, Yin Fu^{1,2,3}, Ziqi Zhu^{1,2,3}, Jianying Huang^{1,2,3}, Huikai Qin^{1,2,3}, Qimeng Yang⁴, Xinmei Li⁴, Yayun Wu^{1,2,3}, Xun Suo⁵, Yu Jiang^{4*} and Longxian Zhang^{1,2,3*}

Abstract

Background *Eimeria* are obligate protozoan parasites, and more than 1,500 species have been reported. However, *Eimeria* genomes lag behind many other eukaryotes since obtaining many oocysts is difficult due to a lack of sustainable in vitro culture, highly repetitive sequences, and mixed species infections. To address this challenge, we used whole-genome amplification of a single oocyst followed by long-read sequencing and obtained a chromosome-level genome of *Eimeria tenella*.

Results The assembled genome was 52.13 Mb long, encompassing 15 chromosomes and 46.94% repeat sequences. In total, 7,296 protein-coding genes were predicted, exhibiting high completeness, with 92.00% single-copy BUSCO genes. To the best of our knowledge, this is the first chromosome-level assembly of *E. tenella* using a combination of single-oocyst whole-genome amplification and long-read sequencing. Comparative genomic and transcriptome analyses confirmed evolutionary relationship and supported estimates of divergence time of apicomplexan parasites and identified AP2 and Myb gene families that may play indispensable roles in regulating the growth and development of *E. tenella*.

Conclusion This high-quality genome assembly and the established sequencing strategy provide valuable community resources for comparative genomic and evolutionary analyses of the *Eimeria* clade. Additionally, our study also provides a valuable resource for exploring the roles of AP2 and Myb transcription factor genes in regulating the development of *Eimeria* parasites.

Keywords *Eimeria tenella*, Single oocyst, Genome, Repeat sequences, Third-generation sequencing, Comparative genomic, AP2, Myb

*Correspondence:

Yu Jiang

yu.jiang@nwafu.edu.cn

Longxian Zhang

zhanglx8999@henau.edu.cn

¹College of Veterinary Medicine, Henan Agricultural University, No. 15

Longzihu University Area, Zhengzhou New District, Zhengzhou

450046, P.R. China

²International Joint Research Laboratory for Zoonotic Diseases of Henan, Zhengzhou 450002, Henan Province, China

³Key Laboratory of Quality and Safety Control of Poultry Products (Zhengzhou), Ministry of Agriculture and Rural Affairs, Zhengzhou, P.R. China

⁴Key Laboratory of Animal Genetics, Breeding and Reproduction of Shaanxi Province, College of Animal Science and Technology, Northwest A&F University, No. 22, Xinong Road, Agricultural High-tech Industrial Demonstration Zone, Yangling 712100, China

⁵National Key Laboratory of Veterinary Public Health Security, Key Laboratory of Animal Epidemiology and Zoonosis of Ministry of Agriculture, National Animal Protozoa Laboratory, College of Veterinary Medicine, China Agricultural University, Beijing 100193, China



© The Author(s) 2025. **Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

Background

Eimeria are obligate protozoan parasites that have evolved to exhibit immense diversity in their host range, including mammals, birds, reptiles, fish, and amphibians [1, 2, 3, 4]. Currently, more than 1,500 species of *Eimeria* have been described, virtually all of which are restricted to a single host species [5, 6, 7]. Oocysts that infect birds and livestock are of great relevance, and coccidiosis caused by *Eimeria* species has caused great economic loss to the aquaculture industry and infected wild vertebrates are significant [8, 9, 10, 11, 12]. In domestic chickens (*Gallus gallus domesticus*), seven globally recognized *Eimeria* species infect distinct regions of the intestine, each causing varying degrees of pathology [7]. Understanding the biology of *Eimeria* and their genetic diversity, population structure, and capacity to evolve underpins the development of new drugs and vaccines. It is now more than 100 years since the significance of coccidiosis was first recognized in poultry, however notable gaps persist in our knowledge [5, 8].

High-quality genome sequencing is a powerful tool for studying *Eimeria* in biology, genetics, genome, and comparative genomics; however, studies on the *Eimeria* genome are relatively scarce. Genome sequencing started for *Eimeria* in 2002; after nearly 22 years of extensive research by scientists around the world, only 12 *Eimeria* species (*E. tenella*, *E. praecox*, *E. acervulina*, *E. maxima*, *E. necatrix*, *E. mitis*, *E. brunetti*, *E. falciformis*, *E. nieschulzi*, *E. zaria*, *E. nagambie*, and *E. outlata*) have been sequenced [13]. The landmark publication of seven *Eimeria* genome sequences in 2014 greatly boosted studies on *Eimeria*, related apicomplexans, and many protozoa [14]. The Darwin Tree of Life (<https://www.darwintreeoflife.org/>) and the Earth BioGenome Project (<https://www.earthbiogenome.org/>) have further promoted the progress of *Eimeria* genome research [15]. The genomes of *E. tenella* and *E. praecox* have reached a chromosome level in 2021 and 2023 [15].

The quality and number of *Eimeria* genomes have lagged behind many other eukaryotes for the following reasons: (1) 45–50% highly repetitive sequences (frequent occurrence of trimer CAG (and other derivatives) and heptamer AAACCCT/AGGGTTT) adding to the difficulty of genome assembly [16]; (2) a lack of sustainable in vitro culture cannot obtain a large number of single-species oocysts; (3) often a mixed infection with strict host specificity, and it is difficult to establish animal models and obtain a large number of single-species oocysts; and (4) the oocyst wall often adheres to food residues and microorganisms like bacteria and fungi, which can lead to DNA contamination and seriously affects genome assembly. Based on the genomic and biological characteristics of *Eimeria*, in 2014, Nair proposed using single-cell and low-input amplification-based genomics,

which can be considered for sequencing with little or no need for in vivo culture [17]. Therefore, additional tools are required if we want to make progress, such as single-cell sequencing and long-read sequencing.

Recent advances in whole-genome amplification (WGA) methods have enabled sequencing of picogram amounts of DNA in a single cell, thus providing a powerful way to obtain microbial genome sequences without cultivation [18]. Detected genetic variants at the single-cell level through WGA and sequencing are well established in human cell lines [19, 20, 21, 22, 23, 24, 25, 26]. To date, single-cell WGA and sequencing have been used successfully in ciliates, *Plasmodium* spp., *Cryptosporidium* spp., and *Leishmania donovani*, leading to important progress in understanding genome biology and evolution [27, 28, 29, 30]. Concurrently, with the progression of long-read sequencing technologies, specifically from companies such as Oxford Nanopore Technologies and Pacific Biosciences, genomic research has transitioned into the era of telomere-to-telomere (T2T) sequencing [31, 32]. High-fidelity (HiFi) sequence reads generated via circular consensus sequencing (CCS) are accurate (greater than 99.00%), especially for repeat sequence regions [33, 34]. Together these technologies enable us to overcome the key difficulties.

To address this, we developed a platform for single-oocyst, whole-genome sequencing using third-generation technology, enabling genome analysis of unculturable species or those with limited oocyst yields. First, *E. tenella* with a reference genome and transcriptome data (complete growth and development) was selected as the research object. Second, we performed single-oocyst WGA using the multiple displacement amplification (MDA) method in *E. tenella* and then used the amplified DNA to construct high-throughput libraries for whole-genome sequencing (WGS) using long-read sequencing technology (ONT and PacBio). Finally, through assembly, we obtained a high-quality genome, and through comparative genomics and transcriptomic analysis, we elucidated the genomic features and revealed the gene expression characteristics at different developmental stages, identifying key genes involved in the developmental process, thereby laying a foundation for an in-depth understanding of the *Eimeria* lifecycle.

Material & methods

Sample preparation

The *E. tenella* strain used in this study was the Houghton (H) strain donated by the State Key Laboratory of Zoon Protozoa, China Agricultural University. Sporulated oocysts were suspended in 2.50% K₂Cr₂O₇ solution and then washed with a sterile PBS solution.

Single-oocyst separation

Five microliters of purified *E. tenella* oocyst suspension was absorbed using a 10 µL pipette and dripped onto a glass Petri dish. Under an inverted Olympus microscope at 60 × magnification (OLYMPUS-BX53, Japan), a single oocyst of *E. tenella* was isolated using a three-axis hydraulic micromanipulator (World Precision Instruments, Inc., USA). In this study, only one oocyst was

selected, the oocyst wall was punctured with a capillary needle of approximately 40 µm, and the four sporocysts were released. The four sporocysts were transferred into a PCR tube containing 4 µL of PBS buffer (Fig. 1).

Single-oocyst WGA, library construction, and sequencing

Single-cell WGA was performed using a REPLI-g@Single Cell Kit (QIAGEN, Germany). WGA products were

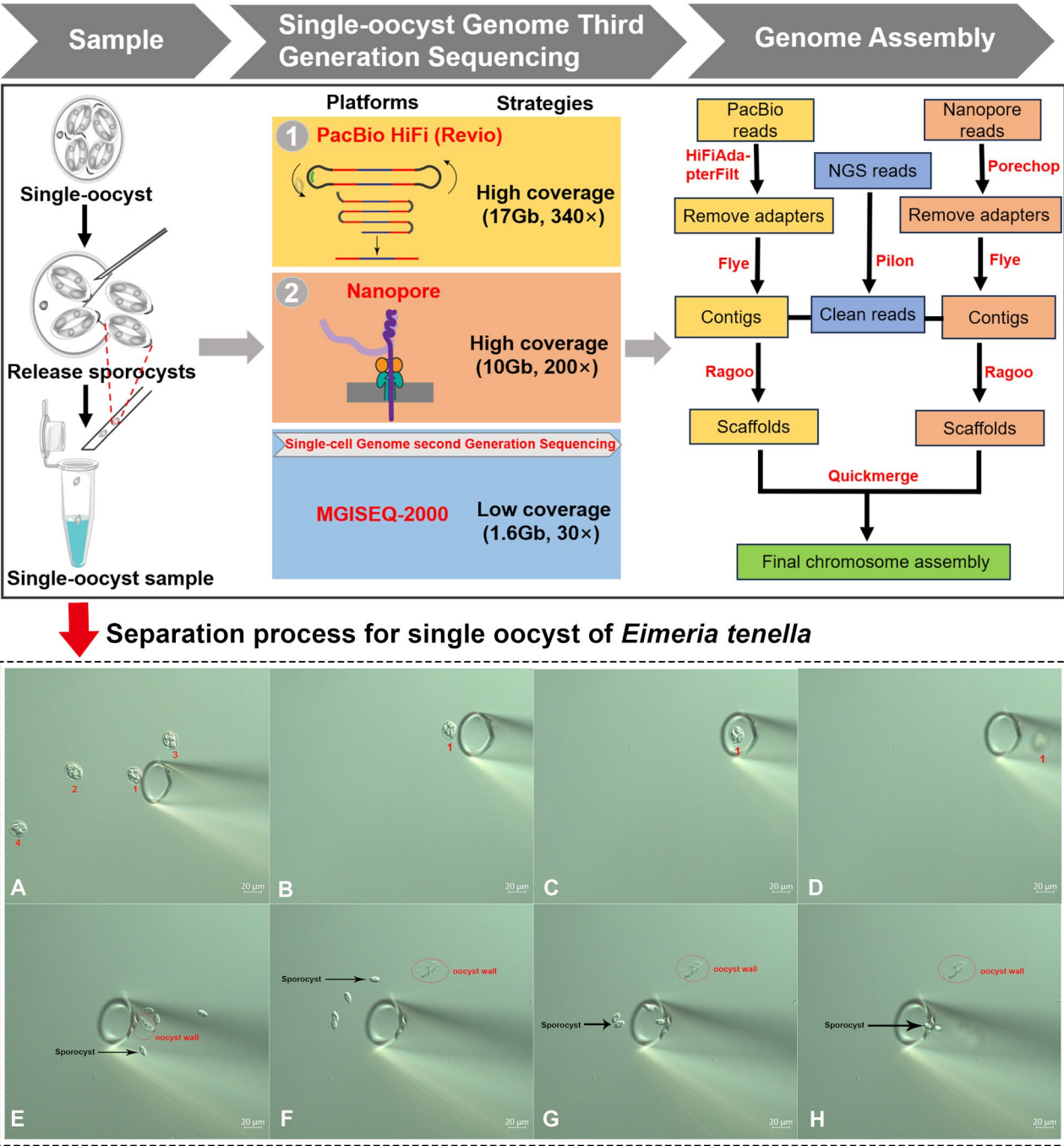


Fig. 1 Assembly workflow of single-oocyst genomes of *E. tenella*. The workflow illustrates the preparation of a single oocyst of the *E. tenella* sample, sequencing platforms, and assemblers

purified using Agencourt AMPure XP beads (BECKMAN, USA) to remove dNTP, primers, primer dimers, salt ions, and other impurities from the amplified products. To reduce exogenous DNA contamination, we cleaned the bench using anhydrous ethanol and treated the MDA reagents with UV irradiation for 30 min before amplifying the single-oocyst sample. Both MDA amplification and DNA purification tests were performed in a 1300 Series II Class A2 biosafety cabinet (Thermo Fisher Scientific). NanoDrop One (Thermo Fisher Scientific, USA) was used to detect sample concentrations C (ng/ μ L), A260/A280, and A260/A230. Qubit 3.0 (Invitrogen, USA) was used to quantify the WGA products, and the Nc/Qc (NanoDrop/Qubit) ratio was calculated according to the concentrations detected by the two instruments. At the same time, 1 μ L of the amplification product was diluted with double steamed water at 1:100, and then the content of *E. tenella* genomic DNA was detected by qPCR based on the *SSU* rRNA gene of *E. tenella* (Supplemental Material). The content of *E. tenella* genomic DNA in the diluent was characterized by the Ct values obtained using a fluorescent quantitative PCR instrument (qTOWER3 G, Analytikjena, Germany).

Single-oocyst WGA products library construction and sequencing

High-quality amplified DNA was used to construct a genomic library, which was size-selected using BluePippin (Sage Science, USA). The purified and size-selected library was then sequenced on the PacBio Revio equipment platform (HiFi) in continuous long-read mode (Pacific Biosciences, USA) and on the PromethION sequencer (ONT, UK) following the manufacturer's instructions (Supplemental Material). A total of 17 Gb (340 $\times$ coverage) PacBio HiFi and 10 Gb (200 $\times$ coverage) ONT long sequencing reads were obtained after removing adaptor reads. For short-read sequencing, library preparation was performed with 50 ng of fragmented DNA using the MGIEasy Universal DNA Library Prep Kit (MGI, Shenzhen, China) and sequenced on the MGISEQ-2000 platform (BGI, Shenzhen, China) (Supplemental Material). Approximately 1.6 Gb (32 $\times$ coverage) of 150-bp paired-end reads (clean data) were generated using the MGI sequencing platform (Fig. 1).

Genome assembly

The adapters in the HiFi and ONT raw data were removed using HiFiAdapterFilt v.2.0.1 [35] and Porechop v.0.3.2 [36] to trim adapters and low-quality bases in the short reads. First, based on ONT and PacBio reads, genome assembly was performed using Flye v.2.9.2 [37] with the default parameters. RaGOO v.1.1 [38] was used to map the contigs to the chromosomes of closely related species (PRJEB43184). Additionally, to enhance the

contiguity of the assembly, the genome assemblies generated by the two platforms were merged using Quickmerge v.0.3 (<https://github.com/mahulchak/quickmerge>). Finally, they were polished using Pilon v.1.24 (<https://github.com/broadinstitute/pilon>) based on short sequencing reads [39] (Fig. 1). Benchmarking Universal Single-Copy Orthologs (BUSCO) v.5.4.6 [40] was used to evaluate the completeness of the *E. tenella* genome assembly against the Coccidia_odb10 database.

Repetitive elements annotation

A combination of de novo and homolog search strategies was used to identify and annotate the repeat sequences in the *E. tenella* genome. RepeatScout v.1.0.5 [41] identified duplicate sequences to remove LTR-type data and then merged the results with LTR-FINDER v.1.05 [42] and LTR-harvest [43]. The identified repeats and the Repeat database were merged as the final repeat sequence library, followed by classification into different repeat categories using the PASTECClassifier.py script included in REPET v.2.5 [44]. The *E. tenella* genome assembly was investigated for TEs using RepeatMasker v.4.0.6, with a final repeat sequence library [45].

Gene prediction and annotation

Protein-coding genes were predicted by integrating ab initio methods, homology alignment data, and transcriptomic data. Transcriptomic data for *E. tenella* from unsporulated oocysts (SRX15012440), sporulated oocysts (SRX15012437), sporozoites (SAMEA3249300, SAMEA3249299 and SAMEA3249304), merozoites (SAMEA3249301 and SAMEA3249306), and gametocytes (SAMEA3249302, SAMEA3249305 and SAMEA3276969) (Project: PRJEB8442, <https://www.ebi.ac.uk/ena/browser/view/PRJEB8442>; <http://www.ncbi.nlm.nih.gov/sra>) were downloaded from the NCBI and ENA databases. For ab initio methods, PASA v.2.4.0 [46] was applied to produce candidate gene structures, which could be applied to obtain a set of gene structures for training SNAP (v.2013-11-29) [47], Augustus v.3.3.3 [48], GenomeThreader v.1.6.1 [49], and GlimmerHMM v.3.0.4 [50]. Subsequently, using trained gene models, Augustus v.3.3.3 [48] and GlimmerHMM v.3.0.4 [50] were used to forecast the gene structure. Lutoff v.1.6.3 was used for annotation based on the reference genome annotation result file [51]. Gene models derived from ab initio and homologous alignment data were integrated into a non-repetitive gene set using EvidenceModeler v.1.1.1 [52], and protein-coding genes were predicted.

The predicted protein sequences were functionally annotated through searching against 18 databases using InterProScan v.5 [53], including CDD [54], Coils [55], Gene Ontology [56], Gene3D [57], Hamap [58], MobiDBLite [59], PANTHER [60], Pfam [61], Phobius [62], PIR

[63], PRINTS [64], ProSite [65], SFLD [66], SignalP [67], SMART [68], SUPERFAMILY [69], TIGRFAM [70], and TMHMM [71].

Circumscribing species

The whole-genome average nucleotide identity (ANI) and alignment coverage between *Eimeria* spp., about *E. tenella*_1 (in the present study), *E. tenella* (PRJEB43184), *E. necatrix* Houghton (PRJEB4833), *E. maxima* Weybridge (PRJEB4864), *E. acervulina* Houghton (PRJEB4832), *E. brunetti* Houghton (PRJEB4834), *E. mitis* Houghton (PRJEB4835), *E. praecox* (PRJEB71489), *E. nagambie* (PRJEB40060), *E. lata* (PRJEB40060), *E. zaria* (PRJEB40060), *E. falciformis* Bayer Haberkorn 1970 (PRJNA232109), and *E. nieschulzi* Landers (PRJNA258495) were calculated using FastANI v.0.2.10 [72] with the default parameters.

Collinearity analysis

Three species with completed genomes, namely *E. tenella* (in this study), *E. tenella* (PRJEB43184), *E. tenella* APU2 (PRJNA929509) and *E. praecox* (PRJEB71489), were selected for collinearity analysis. MCScanX [73] was used to identify the homologous scaffolds and gene synteny. Pairwise blocks were defined as at least five homologous genes in a 25-gene size window.

Phylogenetic analysis and divergence time estimation

The orthologous gene copies that were present in 13 species of apicomplexan parasites, namely *P. falciparum* 3D7 (PRJNA13173), *B. bovis* T2Bo (PRJNA18731), *C. parvum* C1HN (PRJNA1045063), *T. gondii* ME49 (PRJNA28893), *Cystoisospora suis* Wien I (PRJNA341953), *Cyclospora cayetanensis* NF1_C8 (PRJNA357479), *E. tenella* (in present study), *E. necatrix*, *E. maxima*, *E. acervulina*, *E. brunetti*, *E. mitis*, and *E. praecox*, were identified and aligned using OrthoFinder v.2.5.4 [74]. The aligned sequences were concatenated using custom scripts to generate FASTA files for further phylogenetic analyses. Maximum likelihood phylogenetic trees were generated using RAxML v.4.4 [75] with the best-fit model LG + F + R5. ITOL was used to visualize and edit phylogenetic tree-labels (<http://itol2.embl.de/upload.cgi>). The divergence time for apicomplexan parasites was estimated using the MCMCTree [76] program with two correlated time points: 817 million years ago (Mya, divergence time between *P. falciparum* and *T. gondii*, ranging from 580 to 817 Mya), 545 Mya (divergence time between *B. bovis* and *C. parvum*, ranging from 420.3 to 580 Mya) and 499 Mya (divergence time between *T. gondii* and *C. parvum*, ranging from 420 to 580 Mya) [77–78].

Ortholog group identification and gene family expansion and contraction analysis

The apicomplexan protein sequences were downloaded from the NCBI database. The orthologous groups across 13 species, namely *P. falciparum*, *B. bovis*, *C. parvum*, *T. gondii*, *C. suis*, *C. cayetanensis*, *E. tenella*, *E. necatrix*, *E. maxima*, *E. acervulina*, *E. brunetti*, *E. mitis*, and *E. praecox*, were identified using OrthoFinder v.2.5.4 [74], which is a practical, fast, accurate, and comprehensive tool for comparative genomes. The identified ortholog groups were used for further analysis of gene family expansion and contraction with cafe5 v.5.0.0 [79].

Ka/Ks analysis

The nonsynonymous (*Ka*) and synonymous (*Ks*) substitution rates and positive selection strength (*Ka/Ks*) were calculated using the *KaKs_Calculator* v.2.0 [80]. First, reciprocal BLAST was used to run pairwise alignments of the *E. tenella* genome with *C. cayetanensis* and *E. maxima*; the e-value was set to 1e-5, and the number of hits for each pair of species was set to 5. Second, each pairwise protein sequence was aligned using MUSCLE [81], and pairwise nucleotide sequence alignments were generated by transforming protein alignments into codon alignments using ParaAT [82]. Third, *Ka/Ks* ratios were calculated based on pairwise codon alignments using *KaKs_Calculator*, and the *KaKs_Calculator* models were invoked from the PAML.M0 model (Branch site model) used in this study.

Identification Of AP2 and Myb gene family in *Eimeria* spp. Of chicken and gene families phylogenetic analysis

To investigate the type and number of AP2 and Myb-related gene families in each *Eimeria* spp., the target proteins of the AP2 and Myb gene families retrieved from *T. gondii* in ToxoDB (<https://toxodb.org/>) were first downloaded as reference data [83]. The AP2 and Myb gene families were identified by BLASTP searches using the above reference data (E-value < 1e-5). At the same time, the HMMER (<http://hmmer.org/>) was also used to search these two gene families by building a Hidden Markov Model (HMM) profile [85]. Finally, the filtered AP2 and Myb proteins were merged respectively from the results of the above two methods and used for further analysis. The AP2 and Myb gene family protein sequences of *Eimeria* spp. were used for phylogenetic analyses. Protein sequences were aligned for each dataset using the MUSCLE [81] algorithm on the GUIDANCE web server with default parameters. A phylogenetic tree was constructed using the maximum likelihood (ML) method implemented in IQ-TREE with 1,000 bootstrap replicates [87, 88]. The best-fit amino acid substitution model for ML analysis was MFP.

Chromosome distribution, gene structure, conserved domain, and conserved motif analysis

The chromosomal distribution of these genes was obtained from the genome annotation information. The exon-intron organization and splicing phase of these predicted AP2 and Myb genes were also investigated based on the annotation file of the *Eimeria* spp. genome and then graphically displayed using the Gene Structure Display Server [84]. The conserved domains were predicted using the Hmmer web server [85]. Finally, conserved protein motifs were analyzed using MEME [86], with an optimum motif width ranging from 6 to 200 and a maximum number of 10.

Gene expression of EtAP2 and EtMyb gene families

The high-throughput RNA sequencing data of *E. tenella* were retrieved and downloaded from the SRA database (Project: PRJEB8442, <https://www.ebi.ac.uk/ena/browser/view/PRJEB8442>; <http://www.ncbi.nlm.nih.gov/sra>) and used to detect the differential expression of the EtAP2 and EtMyb genes. A total of five RNA datasets from different developmental stages were used, namely unsporulated, sporulated, sporozoites, merozoites, and gametocytes. Hisat2 v.2.2.1 [89] was used to compare transcriptome data to the genome of *E. tenella*, featureCounts (<https://github.com/apeltzer/IGCG-featureCounts>) was used for gene quantitative analysis, and DESeq2 v.1.4 (<https://github.com/thelovelab/DESeq2>) was used for differential expression analysis.

Results

Genome assembly and annotation

The final assembly had a total length of 52.13 Mb in 15 chromosomal scaffolds and one mitochondrial scaffold, which is consistent with previous studies on *E. tenella* genome size. Fifteen chromosome-level scaffolds lengths are as follows: 0.88 Mb (Chromosome (Chr) 1), 1.11 Mb (Chr 2), 1.83 Mb (Chr 3), 1.80 Mb (Chr 4), 2.84 Mb (Chr 5), 3.26 Mb (Chr 6), 3.61 Mb (Chr 7), 3.76 Mb (Chr 8), 3.54 Mb (Chr 9), 3.93 Mb (Chr 10), 4.14 Mb (Chr 11), 3.99 Mb (Chr 12), 4.85 Mb (Chr 13), 5.82 Mb (Chr 14), and 6.76 Mb (Chr 15), and there is certain collinearity between chromosomes (Fig. 2). The total scaffold N50 and GC contents of the genome were 3.92 Mb and 51.61%, respectively (Table 1). We re-mapped the PacBio HiFi reads to the *E. tenella* assembly and found a uniformly high coverage across nearly all genomic regions. The assembly had a BUSCO completeness score of 92.00%, using the Coccidia_odb10 reference set. All these results confirmed the overall accuracy of the assembly (Fig. 3a and S1; Table 1).

A total of 24.47 Mb of sequences (46.94%) were annotated as TEs, 7,296 protein-coding genes were predicted and 5,484 (75.16%) genes were annotated in the assembly

of *E. tenella* (Table S1). The protein-coding genes of *E. tenella* were functionally annotated (783 KEGG pathways and 54 GO categories) (Table S2, Fig. S2), which provides a systematic understanding of the biological functions of *E. tenella* genes and helps to elucidate their roles in host infection and the parasite's lifecycle. Collectively, the results illustrate the high quality, reliability, and accuracy of the assembly.

Analysis of repeat sequences

A total of 24.47 Mb of sequences, representing 46.94% of the *E. tenella* assembly, were annotated as repeat, which included 128,513 elements (Table S1). This repeat sequences content was higher than that previously reported for *E. tenella* (38.94%) and *E. maxima* (37.78%). Approximately 33.92% of the *E. tenella* genome was annotated as containing transposon elements (TEs), including 1.45% retrotransposons and 19.87% DNA transposons. Unclassified DNA transposons were the dominant TEs, followed by Retroelements. The *Gypsy* retrotransposon element was the second most abundant TE, constituting 2.35% of annotated TE content. Cross-genome comparisons with *E. brunetti* and *E. praecox* showed that LTR-RTs, especially *Gypsy* elements, contributed the most to the genome expansion of *E. praecox* (Fig. 3b).

A distinct bimodal distribution was observed for *Gypsy* insertion times in *E. tenella*, whereas a trimodal distribution was observed for *E. praecox* (Fig. 3c). *E. tenella* had a comparatively high proportion of recent LTR-RT insertions, with the peak of amplification appearing around 0.1 My; the other peak occurred at approximately 0.45 My (Fig. 3c). At the superfamily level, we found very recent bursts of unknown elements in *E. tenella* and bursts of *Gypsy* element at 0.075–0.15 My, while amplifications of unknown and *Gypsy* retrotransposons dominantly shaped the bimodal distribution pattern burst dynamics. We also found very recent bursts of *Gypsy* elements in *E. praecox* (Fig. 3b, c).

A robust genome-based taxonomy

We found a clear whole-genome average nucleotide identity (whole-genome ANI) discontinuity among the 13 *Eimeria* species: >99.91% or <91.08%; among them, the ANI score of *E. tenella* and *E. necatrix* was 91.08%, that of *E. maxima* and *E. lata* was 90.63%, and those of other *Eimeria* species in chicken ranged from 77.35 to 82.77%. The ANI score of the two *Eimeria* species in the mouse was 77.69%. The range of ANI scores between *Eimeria* species of chicken and mouse *Eimeria* species was 77.05–78.46%, suggesting that the ANI was also suitable for *Eimeria* species (Fig. 3d, e).

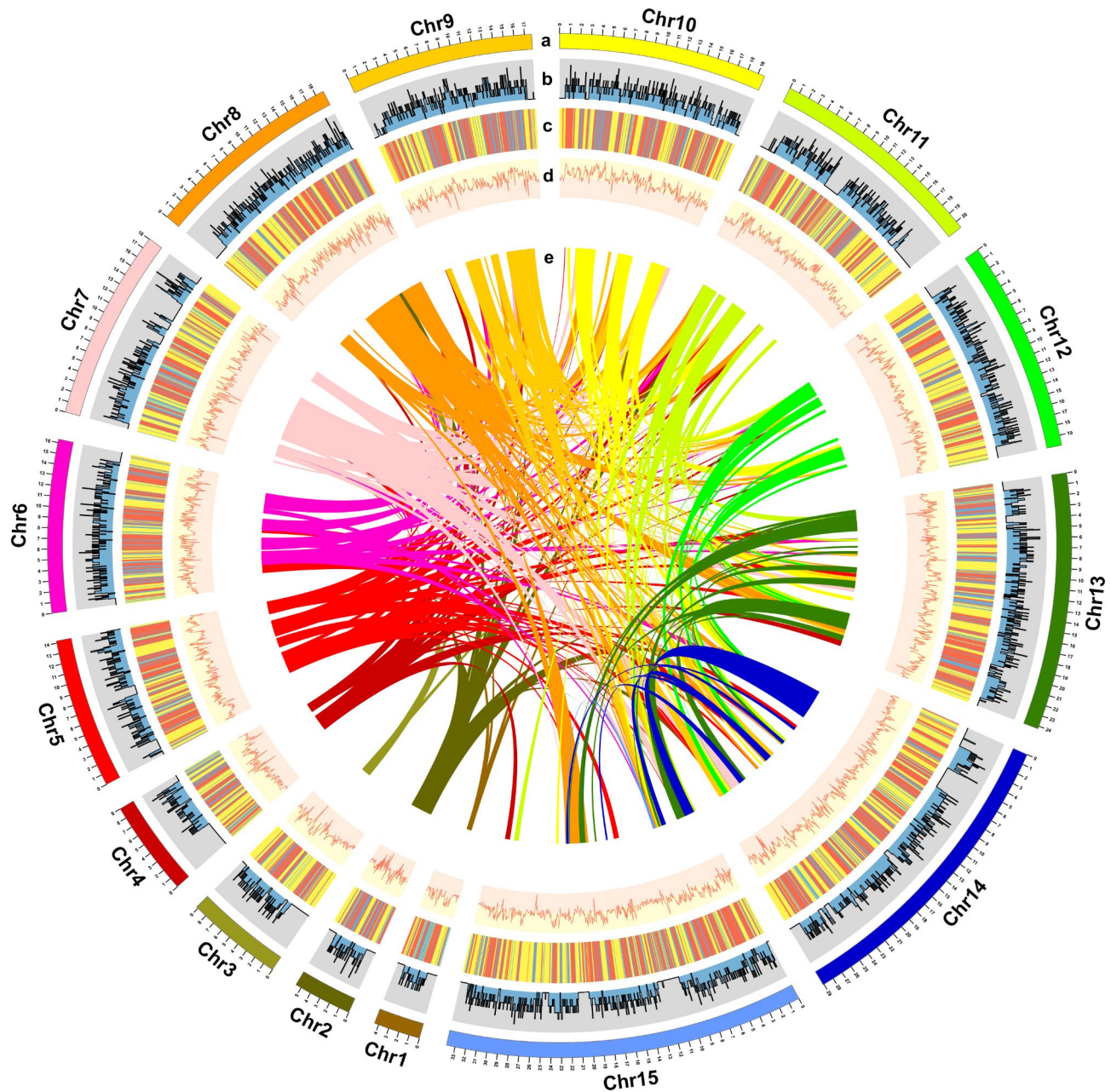


Fig. 2 Genomic features of *E. tenella*. Circos display of important features of the assembled *E. tenella* genome. The five layers depict the chromosome names and sizes, the scale unit is 0.2 Mb. (a), gene density along each chromosome, black: low gene distribution, blue: high gene distribution (b), repeat sequence density along each chromosome, red: high repeat density, blue: medium repeat density, yellow: low repeat density (c), distribution of GC content in each chromosome (d), links between syntenic genes (e)

Orthology, collinearity, and phylogenetic analyses

In *P. falciparum*, *C. suis*, *C. parvum*, *T. gondii* and *E. tenella* species, a total of genes (8,740) were clustered into (6,691) orthologous groups, of which 2,112 (31.56%) genes were shared across all five apicomplexan species (Fig. 4a). A total of 6,320 genes were clustered into 5,276 orthologous groups, of which 2,857 (54.15%) genes were shared across all five *Eimeria* species, representing a pan-*Eimeria* conserved orthogroup (Fig. 4a). Furthermore,

synteny analysis indicated that *E. tenella* and the reference genome (*E. tenella*) barely exhibited disparities in gene content within comparable regions (Fig. 4b, S5). The observed differences primarily pertain to minor structural variations such as the slightest inversions. However, the genes of *E. tenella* and *E. praecox* displayed no significant collinear relationships, with a rearrangement of large fragments (Fig. 4b).

Table 1 Comparison between the assembled and published *E. tenella* reference genomes

Statistic	<i>E. tenella</i> (This study)	<i>E. tenella</i> (Published in 2023)	<i>E. tenella</i> (Published in 2021)	<i>E. tenella</i> (Published in 2013)
Number of oocysts	One	Population	Population	Population
Sequencing platform	PB + ONT + MGI	ONT + Illumina	PB + Hic + 10X Genomics Illumina	Illumina
Genome assembly				
Number of scaffolds	15	15	33	4,664
Genome size (bp)	51,908,535	51,306,473	53,250,139	51,859,670
Largest contig (bp)	6,727,460	6,756,684	6,779,867	1,462,819
N50 (bp)	3,921,268	3,922,363	4,007,228	200,914
N90 (bp)	1,828,249	1,771,817	1,948,711	3,069
GC (%)	51.61	51.5	51.5	51.5
Size of total repeat sequences (%)	47.17	-	-	38.94
Genome annotation				
Number of predicted genes	6,095	-	7,268	8,603
Number of annotated genes	5,484	-	-	-
Genome completeness				
Complete BUSCOs (%)	92.0	98.8	98.8	93.0
Complete and single-copy BUSCOs (%)	91.4	98.8	98.4	92.8
Complete and duplicated BUSCOs (%)	0.6	0	0.4	0.2
Fragmented BUSCOs (%)	0.1	0.8	0.4	5.2
Missing BUSCOs (%)	0.7	0.4	0.8	1.8
Total Lineage BUSCOs	502	502	502	502

Note: Pacbio, PB; Nanopore ONT
Unknown information denoted by hyphen (“-”)

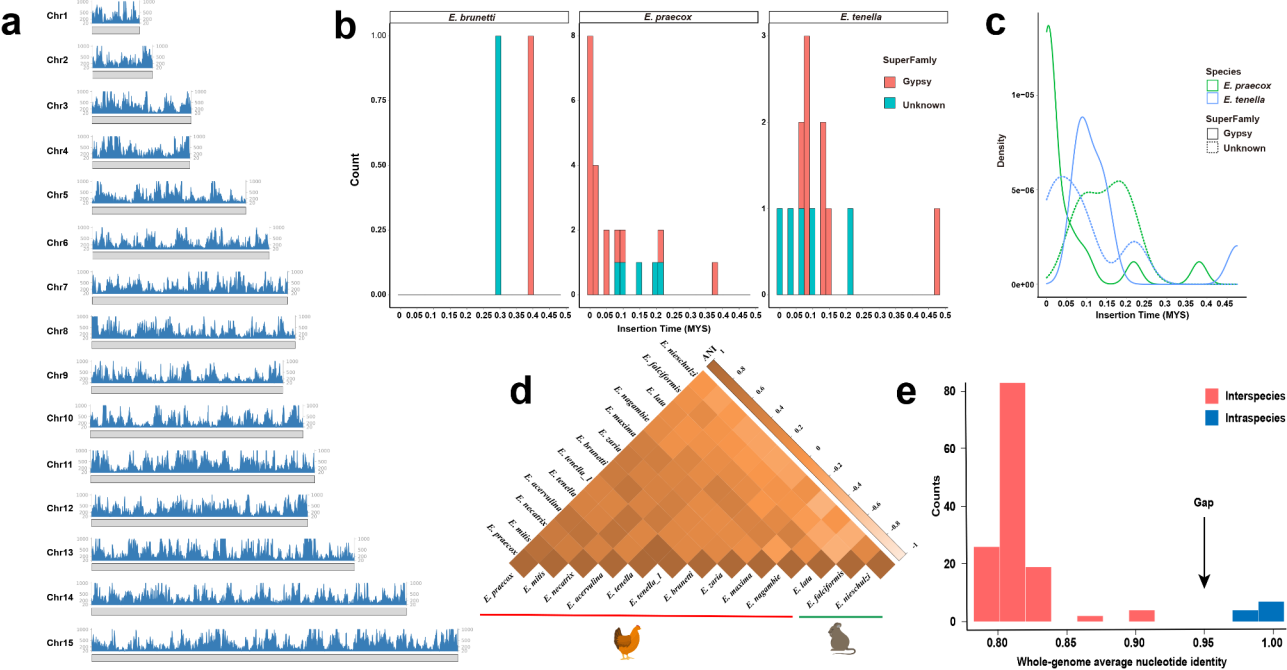


Fig. 3 *E. tenella* genome-based taxonomy, sequencing depth, and analysis of TEs. Genome sequencing depth by PacBio platform (a), insertion bursts of Gypsy and unknown elements in *E. tenella*, *E. brunetti*, and *E. praecox* (b), temporal patterns of LTR-RT insertion bursts in *E. tenella* as compared with those in *E. brunetti* and *E. praecox* (c), heatmap showing the average nucleotide identity (ANI) among the 13 *Eimeria* species (d), ANI distributions among the 13 *Eimeria* species (e)

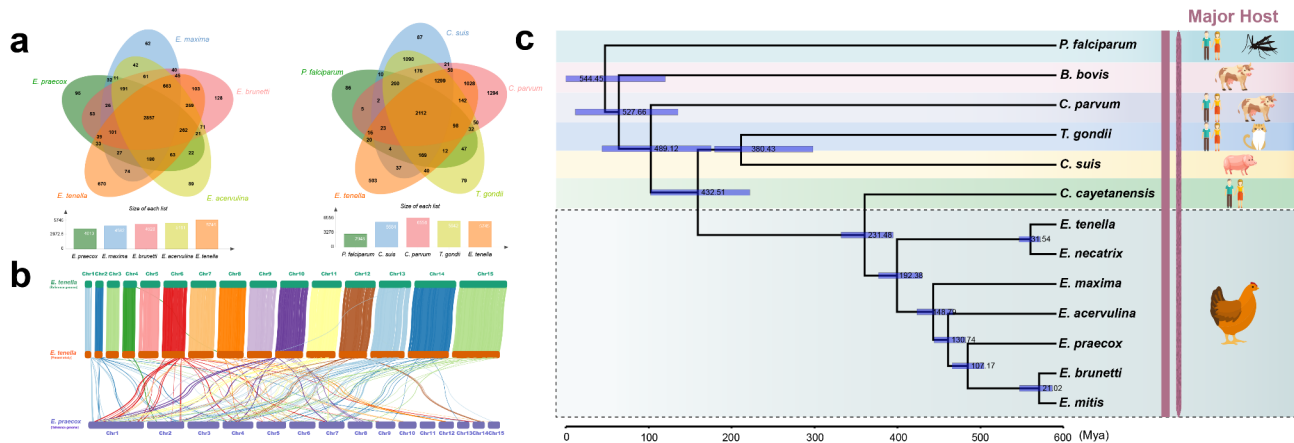


Fig. 4 Gene cluster, collinearity, and phylogenetic analysis of apicomplexa. Orthology between species and unique *E. tenella* proteins. Venn diagram showing orthologous groups in *E. tenella* and nine other apicomplexan species (a), *E. tenella* is collinear with the reference genome and *E. praecox* genes (b), estimated divergence time across apicomplexan parasites. Species dates were determined using 150 orthologous groups across 13 apicomplexan parasites. Two fossil times were included to calibrate the split between these species. 95% confidence intervals for each node are shown in the heatmaps (c)

The species tree was also inferred using phylogenomic analysis of 150 single-copy orthologous genes identified using OrthoFinder v2.5.4 [74] in 13 apicomplexa species. *P. falciparum* was used as an outgroup to reconstruct ML phylogenetic trees. Each species formed a monophyletic clade with sister-group relationships among clades (*T. gondii* and *C. suis*). *Eimeria* spp. are paraphyletic, with sister-group relationships among clade (*E. tenella* and *E. necatrix*) and (*E. brunetti* and *E. mitis*) (Fig. 4c). In the phylogenetic tree of apicomplexan parasites, the *Eimeria* genus is positioned on a relatively lower branch, indicating a distant evolutionary relationship with species like *P. falciparum*, *B. bovis*, *T. gondii*, and *C. parvum* (Fig. 4c). However, it is more closely related to the host-specific species *C. suis* and *C. cayetanensis* (Fig. 4c). This result suggests that these species may have undergone similar selective pressures during evolution, particularly in terms of host specificity and adaptation of parasitic mechanisms. *E. tenella* was derived from a common ancestor with other parasites approximately 231.48 million years ago (Mya) (Fig. 4c). This means *E. tenella* shares a distant common ancestor with other parasites, dating back to the early Triassic period (the early evolution of dinosaurs). Furthermore, *E. tenella* and other *Eimeria* species are thought to have split around 192.38 Mya (Fig. 4c). This divergence event likely marks the point at which *Eimeria* species began to diverge from a common ancestor, gradually adapting to host environments and forming distinct species. The relatively recent timing of this divergence suggests that *E. tenella* and other *Eimeria* species share a more recent common ancestor and have undergone similar evolutionary pressures in terms of host specificity and parasitic mechanisms.

Multigene families and adaptive evolution of *Eimeria tenella*

All apicomplexan parasites species demonstrate a high level of gene family stability, and the high proportion of stable gene families helps them maintain long-term survival within the host, ensuring the stability of their physiological functions (Fig. 5a). *C. parvum* exhibits a significant contraction of gene families, far surpassing other species (with 1339 contracted gene families), reflecting its genome streamlining in adaptation to its host and lifestyle (Fig. 5a). This genome simplification is closely related to its high adaptability to the host environment, allowing *C. parvum* to achieve maximum survival and reproductive capacity with the minimal genetic burden. In contrast, *Cystoisospora suis* shows a substantial expansion of gene families, far exceeding other species (with 1556 expanded gene families), indicating significant genomic expansion to adapt to the microenvironment within its pig host, thereby enhancing its survival and reproductive abilities within the host (Fig. 5a).

In *Eimeria* species, most exhibit a certain degree of gene family expansion, with *E. necatrix* and *E. maxima* standing out, suggesting a strong demand for adaptation to host immune responses and increased infection capability (Fig. 5a). In contrast, *E. tenella* shows relatively fewer expanded gene families (48 expanded), such as those involved in transcription (AP2 domain transcription factor and DEAD-like helicase superfamily), translation (lysyl-tRNA synthetase, class-I aminoacyl-tRNA synthetase family, RNA methyltransferase, and ribosomal protein L15), post-transcriptional modification (phosphatase 2 C), and protein degradation (S8 family peptidase and ubiquitin-transferase domain-containing protein). These expanded genes are primarily involved in various biological processes, including growth and

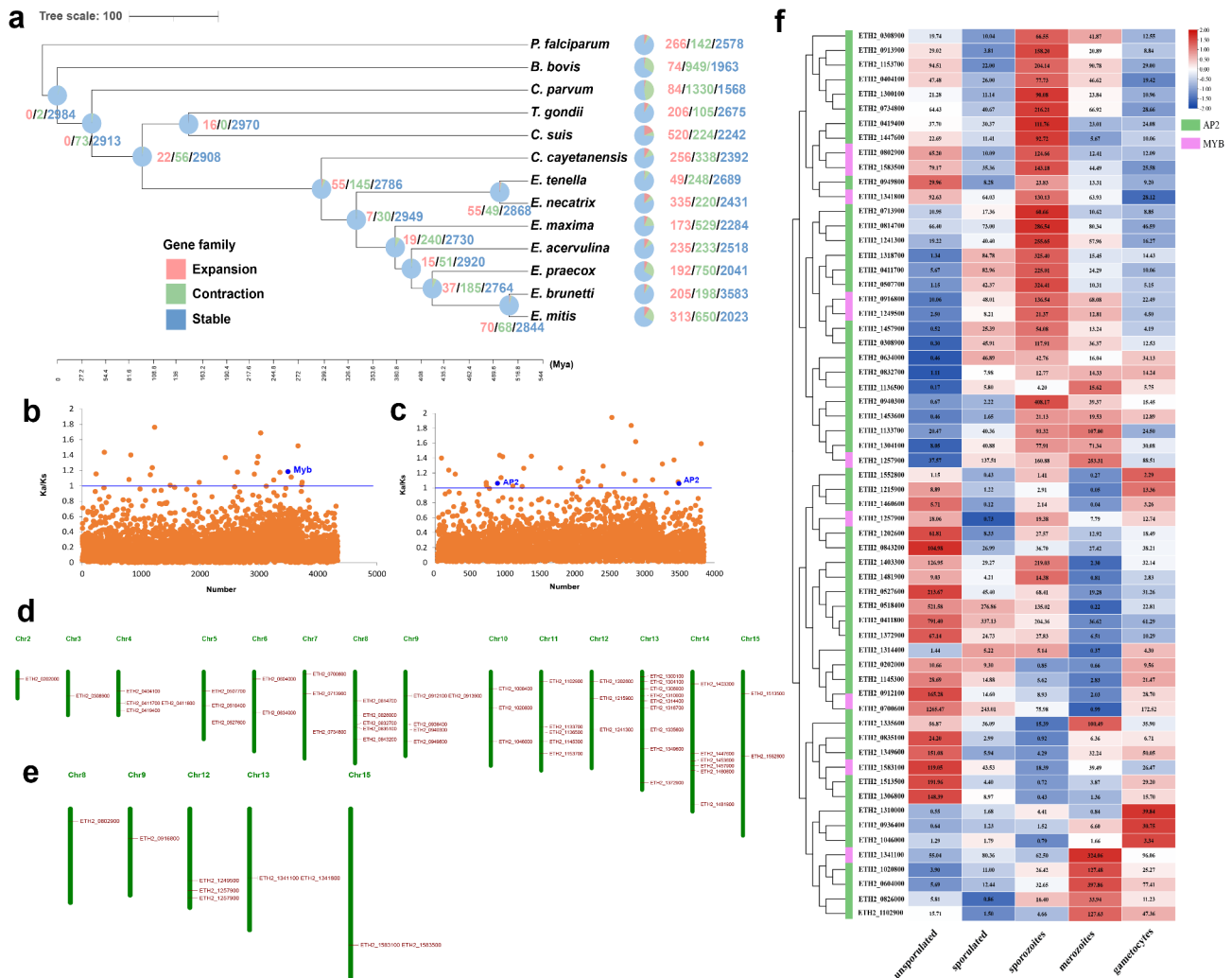


Fig. 5 Evolution of gene families, adaptive evolution in apicomplexan parasites, chromosomal distribution, and expression of EtAP2 and EtMyb gene families. Dynamics of gene family size in apicomplexa parasite genomes. The numbers below the branches indicate gene family expansions/contractions/stable, and the numbers above the branches show gene gains/losses/retention (**a**), Ka/Ks comparisons of the *E. tenella* genome with *C. cayetanensis* (**b**), Ka/Ks comparisons of the *E. tenella* genome with *E. maxima* (**c**), schematic representations of the chromosomal distribution of EtAP2 genes. The chromosome number is indicated at the upside of each chromosome (**d**), schematic representations of the chromosomal distribution of EtMyb genes, the chromosome number is indicated at the upside of each chromosome (**e**), the correlation between the gene expression patterns of 49 EtAP2 and 7 EtMyb genes (**f**), the scale in the figure represents the Log2(TPM values) after column normalization processing, where the depth of the color indicates the degree of correlation. The darker the color, the stronger the correlation, ranging from -2 (completely negatively correlated) to 2 (completely positively correlated)

development, host invasion, metabolism, catalysis, and ATP energy acquisition. These processes are crucial for *E. tenella*'s infection mechanisms, host adaptation, lifecycle development, and energy acquisition, enabling it to survive and rapidly reproduce within the host, thus completing its lifecycle. Most *Eimeria* species also show a high level of gene family contraction, especially *E. tenella* and *E. acervulina*, which may be associated with their specific adaptation to the host, the elimination of redundant genes, and the optimization of their genome structure (Fig. 5a). Despite significant gene expansion in *E. necatrix* and *E. maxima*, their gene family contraction is relatively limited, which may indicate that their genomes maintain

a higher complexity to adapt to the immune challenges of different hosts.

We discovered species adaptations using pairwise Ka/Ks comparisons of the *E. tenella* genome with those of *C. cayetanensis* and *E. maxima*. Within the *E. tenella* genome, we found that 22 and 27 of these families had higher Ka/Ks ratios than other genes (Fig. 5b, c). We found that some genes received selection pressure, including NUF2, protein serine/threonine phosphatase, AP2, and Myb (Fig. 5b, c).

Identification, phylogenetic relationship, motif, protein domain, and gene structure analyses of the AP2 and Myb gene families in seven *Eimeria* spp. from chicken

In total, 284 AP2 and 60 Myb protein sequences were identified in the genomes of seven *Eimeria* spp. from chickens, and the AP2 and Myb gene numbers of each species ranged from 22 to 53 and 4 to 12, respectively (Table S3 and S4). We named the 284 AP2 and 60 Myb genes based on their location on the chromosomes or scaffold of seven *Eimeria* spp. from chickens (Tables S3 and S4). The 284 predicted AP2 and 60 predicted Myb proteins varied in length, molecular weight (MV), and isoelectric points (PIs) (Tables S3 and S4). To further study the evolutionary relationships of the AP2 and Myb genes, we constructed phylogenetic trees using their respective protein sequences of these seven *Eimeria* spp. from chicken and analyzed the protein characteristics, including conserved motifs, protein domains, and corresponding exon organization (Figs. S3 and S4). The information obtained from the phylogenetic tree showed that the AP2 and Myb genes from the different species were also divided into many groups, whereas all AP2 and Myb genes were scattered (Figs. S3 and S4).

Chromosomal distribution and differential expression of the EtAP2 and EtMyb genes

In the present study, chromosome mapping of EtAP2 and EtMyb was performed using the *E. tenella* genome. A total of 52 AP2 and 9 EtMyb genes were evenly distributed on chromosomes 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, and 15, and nine AP2 and three Myb transcription factors were clustered on chromosomes 13 and 12, respectively (Fig. 5d, e). To investigate the expression patterns of EtAP2 and EtMyb further, we analyzed the transcriptome data of the five stages during the growth and development of *E. tenella*. The results showed high variance in the expression levels of EtAP2 and EtMyb. Regarding EtAP2, 10 AP2 genes showed no obvious high expression during the whole growth and development process of *E. tenella*, most of the AP2 genes were highly expressed in the asexual reproductive stage, and only four AP2 genes were highly expressed in the sexual reproductive stage (Fig. 5f). A total of 17, 9, 22, 13, and 5 AP2 genes were highly expressed in unsporulated oocysts, sporulated oocysts, sporozoites, merozoites, and gametocytes, respectively. Among them, two AP2 genes (ETH2_0518400 and ETH2_0411800) were significantly overexpressed during the entire growth stage, especially in the unsporulated oocysts stage (Fig. 5f). The ETH2_0940300 AP2 gene was highly expressed only at the sporozoite stage (Fig. 5f). The ETH2_0604000 AP2 gene was highly expressed only at the merozoite stage (Fig. 5f).

Four proteins showed relatively high expression at multiple stages during the growth and development of *E. tenella*, including ETH2_1257900, ETH2_0916800, ETH2_1341800, and ETH2_1583500 (Fig. 5f). However, two genes, namely ETH2_1257900 and ETH2_1249500, showed low expression in all stages during the growth and development of *E. tenella* (Fig. 5f). Furthermore, ETH2_0802900 was significantly expressed only in unsporulated and sporozoite stages. The ETH2_1257900, ETH2_0916800, ETH2_1341800, ETH2_1583500, and ETH2_0802900 genes were significantly expressed in the sporozoite stage, and their expression levels decreased after the sporozoite stage (Fig. 5f). Interestingly, with the growth and development of *E. tenella*, the expression level of ETH2_1257900 increased gradually and was the highest in the merozoite stage (Fig. 5f).

Discussion

High-quality genomes provide valuable data for comparative analyses to gain insights into the biological features of *Eimeria* species. In this study, we developed a platform for whole-genome, third-generation sequencing of a single oocyst. High genome completeness (average 92.00%) and high precision are encouraging, suggesting that it is possible to obtain high-quality genome data from a single oocyst of *Eimeria* regarding both coverage and precision. We obtained a genome at the chromosome level of *E. tenella* and successfully assembled 15 chromosome-level scaffolds. However, pulsed-field gel electrophoresis and electron microscopy results indicated that *E. tenella* had at least 14 chromosomes, with size ranging from 1 to 7 Mb [90, 91]. Interestingly, recent applications of third-generation ONT, PacBio, and Hi-C sequencing to *E. tenella* and *E. necatrix* have increased their karyotypes from 14 to 15 chromosomes [15]. Conversely, closely related coccidia species, *Neospora caninum* and *T. gondii*, have exhibited a reduction in their karyotype count, decreasing from 14 to 13 chromosomes [92, 93]. These findings challenge previous assertions regarding the haploid chromosome number of 14 for *Eimeria* species, suggesting that further investigation is necessary to accurately determine and understand the karyotype complexity within this group of parasites.

Since the first genome of *E. tenella* was published in 2014 [14, 15], three genome versions of *E. tenella* have been published and updated, which constitute valuable genomic resources for the study of *E. tenella*. When comparing our results to the genome assemblies of *E. tenella* published in 2021 and 2023, we observed similarities in genome size, GC content, number of scaffolds, N50 size, and genome completeness. However, notable improvements were observed when comparing our results to the 2014 assembly. Specifically, the number of scaffolds decreased dramatically from 4,664 to 15, indicating the

achievement of a chromosomal-level draft of the genome. The N50 increased from 200,914 bp to 3,921,268 bp, and the size of repeat sequences in this study was higher than that in previous studies (Table 1). However, the number of genes predicted in this study (7,296) was lower than that reported in previous studies (8,603) (Table 1). This discrepancy may be attributed to the high quality and continuity of the genome assembly in our study, which likely reduced errors and duplicate gene predictions. And, we think the genome of *E. tenella* reached chromosomal level at 2021, 2023 and in the present study, which primarily relies on long-read sequencing technology capable of spanning repetitive sequence regions.

In 2022, Tang assembled the human genome in 1, 10, 20 and 30 cells to investigate the limitations of human genome assembly using a minimal amount of single-cell genome sequencing data. The results revealed that BUSCO evaluation completeness was 13.00%, 81.20%, 90.80%, and 91.50% [94]. In apicomplexan parasites, The BUSCO evaluation completeness of single-amplified genomes from *Plasmodium* spp. and *Cryptosporidium* oocysts were 62.00% and 81%, respectively [28]. We identified 462 (92.00%) of 502 conserved BUSCO genes (Table 1). Thus, the *E. tenella* genome sequence is of high quality in both intergenic and genic regions because an *E. tenella* oocyst contains eight sporozoites, and the sporozoite is the clonal unit.

Additionally, in apicomplexan parasites, repeat sequences have been observed in *Eimeria*, *Plasmodium* (*P. gallinaceum* and *P. relictum*), and *Babesia* spp [30, 95, 96]. The retrotransposons identified in *Eimeria* spp. were mainly classified as belonging to the LTR/*Gypsy* family. In the present study, repeated sequences of several *Eimeria* species were analyzed, and the results showed that recent large-scale bursts of *Gypsy* retroelements may have contributed directly to the *E. praecox* genome expansion (Fig. 3b, c). All transposable elements (TEs) were inserted within the past 0.5 million years (Mya), indicating that after speciation events among the three *Eimeria* species, at least one significant amplification of TEs occurred in each genome. Intriguingly, in *E. tenella*, unclassified DNA transposons were the major TEs that recently expanded (Fig. 3b, c). However, we do not know whether this affects the gene structure and function of the genome. Therefore, additional studies are necessary to elucidate the potential role of transposable elements (TEs) and DNA transposons in the evolution of *E. tenella*. Furthermore, it would be valuable to investigate whether TE and DNA transposon insertion polymorphisms can serve as markers to trace the origin and phylogenetic relationships of *E. tenella*.

Robust taxonomy is essential for communicating scientific results and describing biodiversity [96, 97, 98]. Species identification and classification of *Eimeria* mainly

depend on morphological and molecular biological identification (SSU rRNA, ITS, and COI gene loci). However, the morphological structure and molecular sequences of *Eimeria* species exhibit a high degree of similarity, which poses challenges in accurately identifying and distinguishing between species. Consequently, this similarity has led to confusion and ambiguity, with multiple names being assigned to the same *Eimeria* species. Fortunately, the emergence of whole-genome ANI tools has provided favorable evidence for *Eimeria* species identification and classification. Whole-genome ANI has been widely used to quantitatively circumscribe species of prokaryotes [97, 99]. We found a clear ANI discontinuity among the 13 genomes of *Eimeria*: >99.91% or <91.08% (Fig. 3d, e), suggesting that ANI was also suitable for *Eimeria*. For *Eimeria*, ANI >97.00% and <92.00% were considered intraspecies and interspecies boundaries, respectively. These boundaries of *Eimeria* are much clearer and greater than those of prokaryotes (with ANI >95.00% and <95.00% as intraspecies and interspecies boundaries, respectively), suggesting that intraspecific genetic conservation and interspecific genetic divergence of *Eimeria* are greater than those of prokaryotes. In addition, the ANI results of the genome in this study further confirmed that the three *Eimeria* species (*E. lata*, *E. nagambie*, *E. zaria*) discovered in 2021 are new *Eimeria* species [100]. Intriguingly, the ANI scores of *E. tenella* and *E. necatrix* were 91.08%, and the two species had similar virulence intensities and were the most closely related. Therefore, ANI scores for *E. maxima* and *E. lata* were 90.63%, suggesting that *E. maxima* and *E. lata* may be closely related to each other and have similar biological functions.

Based on the interspecific collinearity analysis, it was evident that the collinearity between the *Eimeria* species was low, indicating significant differences in chromosome structure, which is similar to previous research [14] (Fig. 4b). In addition, ignoring the fact that the *E. tenella* genome contains 45–55% repetitive sequences, we believe that differences in parasitic sites, species evolution, and pathogenicity may be one of the reasons for this result. Our research revealed minimal disparities in gene content within comparable regions of *C. parvum* and *C. tyzzeri*, primarily involving minor structural variations such as small translocations and inversions. However, large fragments have been rearranged in the genes of *C. parvum* and *C. muris* (unpublished data). The parasitic site of *E. tenella* is the cecum, and that of *E. praecox* is the first third of the small intestine. Additionally, *E. tenella* and *E. praecox* had a relatively long evolutionary distance from each other, and *E. tenella* and *E. necatrix* diverged approximately 192.38 Mya (Fig. 4c). Moreover, previous research has shown that synteny between the genomes of *E. tenella* and *E. necatrix* is extensive, but

that between the genome of *E. tenella* and those of *E. maxima* and *E. acervulina* is notably less, which also supported this result [14]. Finally, we believe that more genomic studies on *Eimeria* are needed to better elucidate the synteny among their chromosomes and to confirm our hypothesis.

Phylogenetic analysis of protein sequences confirmed the close relatedness of *C. cayetanensis* to *Eimeria* spp. and between *Eimeria* chicken species, which was consistent with other studies [14, 77, 79, 101]. Our comparative analysis revealed that *E. tenella* was derived from a common ancestor of *C. cayetanensis* around 231.48 Mya (Fig. 4c). Thus, *C. cayetanensis* is similar to *E. tenella* in terms of genome organization, metabolic capabilities, and potential invasion mechanisms. We also investigated the differences between the seven *Eimeria* spp. that infect chickens. The phylogenetic relationship between *Eimeria* spp. appears to be related to their pathogenicity. *E. tenella* and *E. necatrix* with high pathogenicity clustered in a small branch (diverged approximately 31.54 Mya), *E. brunetti* and *E. mitis* with certain pathogenicity clustered in a small branch (diverged approximately 21.02 Mya), with low pathogenicity of *E. maxima* and *E. acervulina*, *E. praecox* with little or no pathogenic were in three separate branches (Fig. 4c). The collinearity results of *Eimeria* spp. from other studies and in the present study all supported this phylogenetic relationship [14] (Fig. 4b).

Transcription factors play major roles in the transcriptional regulation of developmental stages in eukaryotes. As apicomplexan parasites, *Eimeria* spp. have complex life cycles that are characterized by several differentiation stages reflected in transcriptome changes [102]. Despite extensive research on ApiAP2 transcription factors in apicomplexan parasites, such as *Plasmodium* spp., *T. gondii*, and *Cryptosporidium* spp [103, 104, 105, 106, 107], their functions in *Eimeria* spp. have not been extensively studied to date. Previous research showed that EnApiAP2 plays an important role in the growth, development, and sexual evolution of *E. necatrix* [108–109]. And AP2 (ETH2_0411800) is not essential for the growth and development of *E. tenella* [110]. Our results showed that the AP2 transcription factor gene family not only expanded but may also be under diversifying selection, which suggests that the AP2 gene family could be important for host-parasite interactions (Fig. 5c). We also identified 284 AP2 transcription factors of chicken *Eimeria* spp. using a bioinformatics approach, and a phylogenetic analysis of AP2 domain sequences showed that *Eimeria* spp. clustered in different clades, which is consistent with previous research (Fig. S3 and Table S4) [111]. The ETH2_1513500, ETH2_0940300, and ETH2_0604000 genes were found to be mainly expressed during the unsporulated, sporozoite, and merozoite stages, respectively, indicating that it is a sporogony and asexual

reproduction stage-specific gene in *E. tenella* (Fig. 5f). Most AP2 genes were significantly expressed during the asexual stage, suggesting that the AP2 gene family may play an indispensable role in regulating the growth and development of *E. tenella*.

Compared to a previous study [110], the number of identified AP2 protein family member was similar; however, there are some differences in the expression dynamics results (Table S3). For example, in this study, 17 genes were identified as highly expressed in the unsporulated stage, which is three more than in previous study [110]. Similarly, the number of highly expressed genes in the sporozoite stage (21) is significantly higher than the 6 reported previously, whereas the number of genes expressed in the gametocyte stage (5) is much lower than the 15 reported in earlier studies (Fig. 5f) [110]. We believe that the differences in expression dynamics results may be attributed to various factors, such as sample preparation, library construction, and the choice of sequencing platforms. Nevertheless, the findings from multiple studies on the expression of the EtAP2 gene family provide a solid foundation for further in-depth functional research. Therefore, future studies on EtAP2 gene expression and function will require greater focus and involvement from researchers.

Recently, significant advancements have been achieved in the research of the functional role of the Myb gene family within the *T. gondii*, *P. falciparum*, *C. parvum*, and *Giardia intestinalis* genomes [112, 113, 114, 115]. Our results showed that the Myb transcription factor gene family may also be under diverse selection. However, there are no data on the Myb gene family in the *Eimeria* species. In the present study, the number of genes containing Myb domains in *Eimeria* varied from 4 to 12 (Table S4). There was a negative correlation between the number of Myb genes and the genome size of *Eimeria* spp. of chickens. Through the Myb gene phylogenetic tree consisting of *Eimeria* spp. of chicken, we noticed that the Myb genes of *Eimeria* were scattered among many groups, suggesting that they may not have evolved through genome-wide replication (Fig. S4). The distribution of Myb genes across multiple branches indicates that these genes may have undergone a diversified evolutionary process in different species of *Eimeria*. At the same time, it may suggest that the Myb genes in different *Eimeria* species have functionally diverged, potentially adapting to different ecological environments or host-specific requirements. Further findings suggest that the EtMyb gene family is differentially expressed at dramatically different life stages in *E. tenella* and may act as transcriptional regulators for the progression of *E. tenella* throughout the life cycle. The significant expression of the five EtMyb gene families during the sporozoite stage and their subsequent decrease in expression after

this stage suggests their potential role in regulating genes critical for the early stages of *E. tenella* development and host infection. This stage-specific expression pattern indicates that these transcription factors may be essential for processes like sporozoite activation, host-cell invasion, or adaptation to the host environment. Understanding this regulatory mechanism could provide insights into the molecular basis of the parasite's lifecycle and help identify potential targets for interventions to disrupt the infection process.

Conclusion

In summary, we developed a platform for whole-genome, third-generation sequencing of a single oocyst and generated the genome at the chromosome level. This provides an effective method for WGS of more than 1500 *Eimeria* species for which population oocysts cannot be obtained. Furthermore, we effectively prove that the genome obtained from a single oocyst of *Eimeria* is of high quality and accuracy. Through comparative genomics and transcriptomic analysis, we understand the genome characteristics and have identified AP2 and Myb gene families that may perform roles in various biological processes. It lays a foundation for the later function mining research.

Technical validation

The assembly was evaluated using two criteria: mapping of long sequencing reads and BUSCO assessment. PacBio HiFi and ONT long reads were aligned using minimap2 v.2.24. The mapping rates for the HiFi and ONT reads were 99.61% and 99.24%, respectively. Moreover, presence of highly conserved eukaryotic orthologs via BUSCO (Table 1). Overall, these assessments independently confirmed the accuracy and completeness of genome assembly.

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12864-025-11423-1>.

Supplementary Material 1
Supplementary Material 2
Supplementary Material 3
Supplementary Material 4
Supplementary Material 5
Supplementary Material 6

Acknowledgements

We thank LetPub (www.letpub.com) for its linguistic assistance during the preparation of this manuscript.

Author contributions

LZ, YJ, and XS designed the study. KZ, ZZ, YF, and JH performed the experiments. KZ, YC, QY, and XL conducted the data analyses. KZ, YC, HQ, YW,

and LZ created the figures and wrote the first draft. All authors revised the manuscript. LZ funded the work.

Funding

This research was funded by the National Key Research and Development Plan Project (2023YFD1801200), NSFC-Henan Joint Fund Key Project (U1904203), and Leading Talents of the Central Plains Thousand Talents Program (19CZ0122).

Data availability

The associated sequencing data can be access through National Genomics Data Center, China National center for Bioinformation (PRJCA028428, <http://ngdc.cncb.ac.cn/>). MGI short reads raw sequencing data access link for reviewers: <https://ngdc.cncb.ac.cn/gsa/s/RETf4QaG>. PacBio HiFi long reads raw sequencing data access link for reviewers: <https://ngdc.cncb.ac.cn/gsa/s/6i6V9pEe>. ONT long reads raw sequencing data access link for reviewers: <https://ngdc.cncb.ac.cn/gsa/s/wi355OU8>. And the whole-genome assembly (accession GWHEUVT00000000.1) of the *E. tenella* Houghton strain can be access through National Genomics Data Center, China National Centre for Bioinformation/Beijing Institute of Genomics, Chinese Academy of Sciences (PRJCA028428, <https://ngdc.cncb.ac.cn/gwh/Assembly/reviewersPage/rubhUHBkfJWacQgqdaowMAVgiCgCkOcJIMApAQfclNVODBDKdyvES0eXuEocKY>).

Code availability

No custom code was used in this study. Data analyses were performed using the standard bioinformatics tools specified in the methods.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Received: 26 September 2024 / Accepted: 28 February 2025

Published online: 17 March 2025

References

1. Jirku M, Obornik M, Lukes J, Modry DA. Model for taxonomic work on homoxenous coccidia: redescription, host specificity, and molecular phylogeny of *Eimeria Ranae* Dobell, 1909, with a review of anuran-host *Eimeria* Apicomplexa: Eimeriorina. J Eukaryot Microbiol. 2009;56:39–51. <https://doi.org/10.1111/j.1550-7408.2008.00362.x>
2. Gibson-Kueh S, Yang R, Thuy NT, Jones JB, Nicholls P, Ryan U. The molecular characterization of an *Eimeria* and *Cryptosporidium* detected in Asian Seabass (*Lates calcarifer*) cultured in Vietnam. Vet Parasitol. 2001;181:91–6. <https://doi.org/10.1016/j.vetpar.2011.05.004>
3. Yang R, Fenwick S, Potter A, Elliot A, Power M, Beveridge I, et al. Molecular characterization of *Eimeria* species in macropods. Exp Parasitol. 2012;132. <https://doi.org/10.1016/j.exppara.2012.07.003>
4. Chapman HD, Barta JR, Blake D, Gruber A, Jenkins M, Smith NC, et al. A selective review of advances in coccidiosis research. Adv Parasitol. 2013;83:93–171. <https://doi.org/10.1016/B978-0-12-407705-8.00002-1>
5. Jirku M, Kvicerova J, Modry D, Hypsa V. Evolutionary plasticity in coccidia–striking morphological similarity of unrelated coccidia (Apicomplexa) From related hosts: *Eimeria* spp. From African and Asian pangolins (*Mammalia: Pholidota*). Protist. 2013;164:470–81. <https://doi.org/10.1016/j.protis.2013.04.01>
6. Kvicerova J, Hypsa V. Host-parasite incongruences in rodent *Eimeria* suggest significant role of adaptation rather than cophylogeny in maintenance of host specificity. PLoS ONE. 2013;8:e63601. <https://doi.org/10.1371/journal.pone.0063601>
7. Beck HP, Blake D, Dardé ML, Felger I, Pedraza-Díaz S, Javier RC, et al. Molecular approaches to diversity of populations of apicomplexan parasites. Int J Parasitol. 2009;39(2). <https://doi.org/10.1016/j.ijpara.2008.10.001>

8. Jeanes C, Vaughan-Higgins R, Green RE, Sainsbury AW, Marshall RN, Blake DP. Two new *Eimeria* species parasitic in corncrakes (*Crex crex*) (Gruiformes: Rallidae) in the united Kingdom. *J Parasitol*. 2013;99:634–8. <https://doi.org/10.1645/12-52.1>
9. Sarah CLK, Andy F, Owen LP, Trevor RJ, Rebecca B, Amy B. Stability of within-host-parasite communities in a wild mammal system. *Proc Biol Sci*. 2013;280:0. <https://doi.org/10.1098/rspb.2013.0598>
10. Lassen B, Ostergaard S. Estimation of the economical effects of *Eimeria* infections in Estonian dairy herds using a stochastic model. *Prev Vet Med*. 2013;106:258–65. <https://doi.org/10.1016/j.prevetmed.2012.04.005>
11. Blake DP, Tomley FM. Securing poultry production from the ever-present *Eimeria* challenge. *Trends Parasitol*. 2013;30:12–9. <https://doi.org/10.1016/j.pt.2013.10.003>
12. Dalloul RA, Lillehoj HS. Poultry coccidiosis: recent advancements in control measures and vaccine development. *Expert Rev Vaccines*. 2006;5:143–63. <https://doi.org/10.1586/14760584.5.1.143>
13. Chapman HD, Shirley MW. The Houghton strain of *Eimeria tenella*: a review of the type strain selected for genome sequencing. *Avian Pathol*. 2003;32:115–27. <https://doi.org/10.1080/0307945021000071588>
14. Reid AJ, Blake DP, Ansari HR, Billington K, Browne HP, Bryant J, et al. Genomic analysis of the causative agents of coccidiosis in domestic chickens. *Genome Res*. 2014;11(10):1676–85. <https://doi.org/10.1101/gr.168955.113>
15. Erik A, Ulrike B, Damer B, Alexander D, Michelle S, Craig C, et al. The complete genome sequence of *Eimeria tenella* (Tyzzer 1929), a common gut parasite of chickens. *Wellcome Open Res*. 2021;9:6:225. <https://doi.org/10.12688/wellcomeopenres.17100.1>
16. Blake DP. *Eimeria* genomics: where are we now and where are we going? *Vet Parasitol*. 2015;212:68–74. <https://doi.org/10.1016/j.vetpar.2015.05.007>
17. Nair S, Nkhoma SC, Serre D, Zimmerman PA, Gorena K, et al. Single-cell genomics for dissection of complex malaria infections. *Genome Res*. 2014;24:1028–38. <https://doi.org/10.1101/gr.168286.113>
18. Chen CY, Xing D, Tan LZ, Li H, Zhou GY, Huang L, et al. Single-cell whole-genome analyses by linear amplification via transposon insertion (LIANTI). *Science*. 2017;6334189. <https://doi.org/10.1126/science.aak9787>
19. Tang F, Barbacioru C, Wang Y. mRNA-Seq whole transcriptome analysis of a single cell. *Nat Methods*. 2009;6:377–82. <https://doi.org/10.1038/nmeth.1315>
20. Ilya MF, Johanna G, Maxim I, Hugo BB, Sergey VU, Nezar A, et al. Single-nucleus Hi-C reveals unique chromatin reorganization at oocyte-to-zygote transition. *Nature*. 2017;544:110–4. <https://doi.org/10.1038/nmeth.1315>
21. Tan L, Xing D, Chang CH, Heng L, Xie XS. Three-dimensional genome structures of single diploid human cells. *Science*. 2018;361:924–8.
22. Granja JM, Corces MR, Pierce SE, Bagdatli ST, Choudhry H, Chang HY, et al. ArchR is a scalable software package for integrative single-cell chromatin accessibility analysis. *Nat Genet*. 2021;53:403–11. <https://doi.org/10.1038/s41588-021-00790-6>
23. Buenrostro JD, Wu B, Litzenburger UM, Ruff D, Gonzales ML, Snyder MP, et al. Single-cell chromatin accessibility reveals principles of regulatory variation. *Nature*. 2015;523:486–90. <https://doi.org/10.1038/nature14590>
24. Guo HS, Zhu P, Wu XL, Li XL, Wen L, Tang FC. Single-cell methylome landscapes of mouse embryonic stem cells and early embryos analyzed using reduced representation bisulfite sequencing. *Genome Res*. 2013;23:2126–35. <https://doi.org/10.1101/gr.161679.113>
25. Chen S, Lake BB, Zhang K. High-throughput sequencing of the transcriptome and chromatin accessibility in the same cell. *Nat Biotechnol*. 2019;37:1452–7.
26. Dean FB, Nelson JR, Giesler TL, Lasken RS. Rapid amplification of plasmid and phage DNA using phi 29 DNA polymerase and multiply-primed rolling circle amplification. *Genome Res*. 2001;11:1095–9. <https://doi.org/10.1101/gr.180501>
27. Chen K, Wang GY, Xiong J, Jiang CQ, Miao W. Exploration of genetic variations through single-cell whole-genome sequencing in the model ciliate tetrahymena thermophila. *J Eukaryot Microbiol*. 2019;66(6):954–65. <https://doi.org/10.1111/jeu.12746>
28. Karin T, Björn H, Anna-Maria D, Cecilia A, Romanico A, Mikael H, et al. *Cryptosporidium* as a testbed for single cell genome characterization of unicellular eukaryotes. *BMC Genomics*. 2016;23:17471. <https://doi.org/10.1186/s12864-016-2815-y>
29. Dia A, Cheeseman IH. Single-cell genome sequencing of protozoan parasites. *Trends Parasitol*. 2021;37:803–14.
30. Gabriel HN, Pieter M, Hideo I, Ilse M, Nada K, Akila Y, et al. High throughput single-cell genome sequencing gives insights into the generation and evolution of mosaic aneuploidy in *Leishmania donovani*. *Nucleic Acids Res*. 2021;50(1):293–305. <https://doi.org/10.1093/nar/gkab1203>
31. Goodwin S, McPherson JD, McCombie WR. Coming of age: ten years of next-generation sequencing technologies. *Nat Rev Genet*. 2016;17:333–51. <https://doi.org/10.1038/nrg.2016.49>
32. Wenger AM, Peluso P, Rowell WJ, Chang PC, Hall RJ, Concepcion GT. Accurate circular consensus long-read sequencing improves variant detection and assembly of a human genome. *Nat Biotechnol*. 2019;37:1155–62. <https://doi.org/10.1038/s41587-019-0217-9>
33. Burton JN, Adey A, Patwardhan RP, Qiu R, Kitzman JO, Shendure J. Chromosome-scale scaffolding of de novo genome assemblies based on chromatin interactions. *Nat Biotechnol*. 2013;31:1119–25. <https://doi.org/10.1038/nbt.2727>
34. Kaplan N, Dekker J. High-throughput genome scaffolding from in vivo DNA interaction frequency. *Nat Biotechnol*. 2013;31:1143–7. <https://doi.org/10.1038/nbt.2768>
35. Sim SB, Corpuz RL, Simmonds TJ, Scott MG. HiFiAdapterFilter, a memory efficient read processing pipeline, prevents occurrence of adapter sequence in PacBio HiFi reads and their negative impacts on genome assembly. *BMC Genomics*. 2022;23(1):157. <https://doi.org/10.1186/s12864-022-08375-1>
36. Bonenfant Q, Noé L, Touzet H. Porechop_ABL: discovering unknown adapters in Oxford nanopore technology sequencing reads for downstream trimming. *Bioinform Adv*. 2023;3(1):vbac085. <https://doi.org/10.1093/bioadv/vbac085>
37. Mikhail K, Derek MB, Bahar B, Alexey G, Mikhail R, Sung BS, et al. MetaFlye: scalable long-read metagenome assembly using repeat graphs. *Nat Methods*. 2020;17(11):1103–10. <https://doi.org/10.1038/s41592-020-00971-x>
38. Michael A, Sebastian S, Srividya R, Xingang W, Sara G, Fritz JS, et al. RaGOO: fast and accurate reference-guided scaffolding of draft genomes. *Genome Biol*. 2019;20(1):224. <https://doi.org/10.1186/s13059-019-1829-6>
39. Walker BJ, Abeel T, Shea T, Priest M, Abouelliel A, Sakthikumar S, et al. Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PLoS ONE*. 2014;9(11):e112963. <https://doi.org/10.1371/journal.pone.0112963>
40. Simão FA, Waterhouse RM, Panagiotis I, Kriventseva EV, Zdobnov EM. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics*. 2015;31:3210–2.
41. Price AL, Jones NC, Pevzner PA. De Novo identification of repeat families in large genomes. *Bioinformatics*. 2005;21(1):i351–8.
42. Xu Z, Wang H. LTR_FINDER: an efficient tool for the prediction of full-length LTR retrotransposons. *Nucleic Acids Res*. 2007;35(2):265–8. <https://doi.org/10.1093/nar/gkm286>
43. Ellinghaus D, Kurtz S, Willhoeft U. LTRharvest, an efficient and flexible software for de novo detection of LTR retrotransposons. *BMC Bioinformatics*. 2008;14:9:18. <https://doi.org/10.1186/1471-2105-9-18>
44. Edgar RC, Myers EW. PILER: identification and classification of genomic repeats. *Bioinformatics*. 2005;21(1):152–8. <https://doi.org/10.1093/bioinformatics/bti1003>
45. Tarailo-Graovac M, Chen N. Using repeatmasker to identify repetitive elements in genomic sequences. *Curr Protoc Bioinf*. 2009;25(1):410.
46. Haas BJ, Delcher AL, Mount SM, Wortman JR, Smith JRK, Hannick LI. Improving the Arabidopsis genome annotation using maximal transcript alignment assemblies. *Nucleic Acids Res*. 2003;31:5654–66. <https://doi.org/10.1093/nar/gkg770>
47. Korf I. Gene finding in novel genomes. *BMC Bioinformatics*. 2004;5:59.
48. Mario S, Burkhard M. AUGUSTUS: a web server for gene prediction in eukaryotes that allows user-defined constraints. *Nucleic Acids Res*. 2005;33:465–7. <https://doi.org/10.1093/nar/gki458>
49. Gremme G, Brendel V, Sparks ME, Kurtz S. Engineering a software tool for gene structure prediction in higher organisms. *Inf Softw Tech*. 2005;47:965–78.
50. Majoros WH, Pertea M, Salzberg SL. TigrScan and GlimmerHMM: two open source Ab initio eukaryotic gene-finders. *Bioinformatics*. 2004;20:2878–9. <https://doi.org/10.1093/bioinformatics/bth315>
51. Alaina S, Salzberg SL. LiftOff: accurate mapping of gene annotations. *Bioinformatics*. 2021;19(12):37. <https://doi.org/10.1093/bioinformatics/btaa1016>
52. Haas BJ, Salzberg STL, Zhu W, Pertea M, Allen JE, Orvis J, et al. Automated eukaryotic gene structure annotation using evidencemodeler and the program to assemble spliced alignments. *Genome Bio*. 2008;9:R7. <https://doi.org/10.1186/gb-2008-9-1-r7>
53. Jones P, Binns D, Chang HY, Fraser M, Li W, McAnulla C, et al. InterProScan 5: genome-scale protein function classification. *Bioinformatics*. 2014;30:1236–40. <https://doi.org/10.1093/bioinformatics/btu031>

54. Marchler-Bauer A, Zheng C, Chitsaz F, Derbyshire MK, Geer LY, Geer RC, et al. CDD: conserved domains and protein three-dimensional structure. *Nucleic Acids Res.* 2014;41:D348–52. <https://doi.org/10.1093/nar/gks1243>
55. Fitzkee NC, Fleming PJ, Rose GD. The protein coil library: a structural database of nonhelix, nonstrand fragments derived from the PDB. *Proteins.* 2005;58:852–4. <https://doi.org/10.1002/prot.20394>
56. Michael A, Catherine AB, Judith AB, David B, Heather B, Allan PD, et al. Gene ontology: tool for the unification of biology. *Nat Genet.* 2000;25:25–9. <https://doi.org/10.1038/75556>
57. Corin Y, Jonathan L, Adam R, Paul K, Nigel M, Liu L, et al. Gene3D: comprehensive structural and functional annotation of genomes. *Nucleic Acids Res.* 2008;36:D414–8.
58. Tania L, Auchincloss AH, Elisabeth C, Guillaume K, Karine M, Catherine R, et al. HAMAP: a database of completely sequenced microbial proteome sets and manually curated microbial protein families in UniProtKB/Swiss-Prot. *Nucleic Acids Res.* 2009;37:D471–8. <https://doi.org/10.1093/nar/gkn661>
59. Necci M, Piovesan D, Clementel D, Dosztányi Z, Tosatto SCE. MobiDB-lite 3.0: fast consensus annotation of intrinsic disorder flavors in proteins. *Bioinformatics.* 2021;36:5533–4. <https://doi.org/10.1093/bioinformatics/btaa1045>
60. Mi H, Poudel S, Muruganujan A, Casagrande JT, Thomas PD. PANTHER version 10: expanded protein families and functions, and analysis tools. *Nucleic Acids Res.* 2016;44:D336–42. <https://doi.org/10.1093/nar/gkv1194>
61. Finn RD. Pfam: the protein families database. *Nucleic Acids Res.* 2013;42:D222–30.
62. Käll L, Krogh A, Sonnhammer EL. Advantages of combined transmembrane topology and signal peptide prediction—the phobius web server. *Nucleic Acids Res.* 2007;35:W429–32. <https://doi.org/10.1093/nar/gkm256>
63. Barker WC, Garavelli JS, Mcgarvey PB, Marzec CR, Orcutt BC, Srinivasarao GY. The PIR-international protein sequence database. *Nucleic Acids Res.* 1999;27:39–43.
64. Attwood TK. The PRINTS database: a fine-grained protein sequence annotation and analysis resource—its status in 2012. *Database (Oxford).* 2012;bas019. <https://doi.org/10.1093/database/bas019>
65. Sigrist CJA, Lorenzo C, Edouard DC, Langendijk-Genevaux PS, Virginie B, Amos B. PROSITE, a protein domain database for functional characterization and annotation. *Nucleic Acids Res.* 2010;38:D161–6. <https://doi.org/10.1093/nar/gkp885>
66. Eyal A, Shoshana B, Almonacid DE, Barber AE, Custer AF, Hicks MA. The structure-function linkage database. *Nucleic Acids Res.* 2014;42:D521–30.
67. Teufel F, Armenteros JJA, Johansen AR, Gislason MH, Pihl SI, Tsirigos KD, et al. SignalP 6.0 predicts all five types of signal peptides using protein language models. *Nat Biotechnol.* 2022;40:1023–5. <https://doi.org/10.1038/s41587-021-01156-3>
68. Letunic I, Bork P. 20 years of the SMART protein domain annotation resource. *Nucleic Acids Res.* 2018;46:D493–6. <https://doi.org/10.1093/nar/gkx922>
69. Derek W, Martin M, Christine V, Cyrus C, Julian G. The SUPERFAMILY database in 2007: families and functions. *Nucleic Acids Res.* 2007;35:D308–13.
70. Haft DH, Selengut JD, White O. The TIGRFAMs database of protein families. *Nucleic Acids Res.* 2003;31:371–3.
71. Krogh A, Larsson B, Heijne G, Sonnhammer ELL. Bioinformatics S. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *J Mol Biol.* 2001;305:567–80. <https://doi.org/10.1006/jmbi.2000.4315>
72. Chirag J, Luis M, Rodriguez -RAM, Phillippy. High throughput ANI analysis of 90K prokaryotic genomes reveals clear species boundaries. *Nat Commun.* 2018;9:5114. <https://doi.org/10.1038/s41467-018-07641-9>
73. Wang YP, Tang HB, Debarry JD, Tan X, Li JP, Wang XY, et al. MCSanX: a toolkit for detection and evolutionary analysis of gene synteny and collinearity. *Nucleic Acids Res.* 2012;8(7):e49. <https://doi.org/10.1093/nar/gkr1293>
74. Emms DM, Kelly S. OrthoFinder: solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. *Genome Biol.* 2015;16:157. <https://doi.org/10.1186/s13059-015-0721-2>
75. Yang Z. PAML 4: phylogenetic analysis by maximum likelihood. *Mol Biol Evol.* 2007;24:1586–91.
76. Huelsenbeck JP, Ronquist F. MrBayes: Bayesian inference of phylogenetic trees. *Bioinformatics.* 2001;17:754–5.
77. Wang J, Chen K, Yang J, Zhang S, Li Y, Liu G, et al. Comparative genomic analysis of *Babesia Duncan* responsible for human babesiosis. *BMC Biol.* 2022;20(1):1–14. <https://doi.org/10.1186/s12915-022-01361-9>
78. Liu Q, Guan XA, Li DF, Zheng YX, Wang S, Xuan XN, et al. *Babesia Gibsoni* Whole-Genome sequencing, assembling, annotation, and comparative analysis. *Microbiol Spectr.* 2023;11(4):0. <https://doi.org/10.1128/spectrum.00721-23>
79. Fábio KM, Vanderpool D, Fulton B, Matthew WH, Author N. CAFE 5 models variation in evolutionary rates among gene families. *Bioinformatics.* 2020;36(22–23):5516–8. <https://doi.org/10.1093/bioinformatics/btaa1022>
80. Wang D, Zhang Y, Zhang Z, Zhu J, Yu J. KaKs_Calculator 2.0. A toolkit incorporating gamma-series methods and sliding window strategies. *Genomics Proteom Bioinform.* 2010;8(1):77–80. [https://doi.org/10.1016/S1672-0229\(10\)60008-3](https://doi.org/10.1016/S1672-0229(10)60008-3)
81. Robert CE. MUSCLE: a multiple sequence alignment method with reduced time and space complexity. *BMC Bioinformatics.* 2004;5:0. <https://doi.org/10.1186/1471-2105-5-113>
82. Zhang Z, Xiao J, Wu J, Zhang H, Liu G, Wang X. ParaAT. A parallel tool for constructing multiple protein-coding DNA alignments. *Biochem Biophys Res Commun.* 2012;419(4):779–81. <https://doi.org/10.1016/j.bbrc.2012.02.101>
83. Harb OS, Kissinger JC, Roos DS. ToxoDB: the functional genomic resource for *Toxoplasma* and related organisms *. *Toxoplasma gondii*. (Third Edition). 2020;1021–41. <https://doi.org/10.1016/B978-0-12-815041-2.00023-2>
84. Bo H, Jin J, Guo A-Y, Zhang H, Jingchu Luo and, Gao G. GSDS 2.0: an upgraded gene feature visualization server. *Bioinformatics.* 2015;31(8):1296–7. <https://doi.org/10.1093/bioinformatics/btu817>
85. Simon CP, Aurélien L, Sean RE, Youngmi P, Rodrigo L, Robert DF. HMMER web server: 2018 update. *Nucleic Acids Res.* 2018;46(0):0. <https://doi.org/10.1093/nar/gky448>
86. Timothy LB, Mikael B, Fabian AB, Martin F, Charles EG, Luca C, et al. MEME SUITE: tools for motif discovery and searching. *Nucleic Acids Res.* 2009;37(0):0. <https://doi.org/10.1093/nar/gkp335>
87. Di MA, Citores L, Russo R, Iglesias R, Ferreras JM. Sequence comparison and phylogenetic analysis by the maximum likelihood method of ribosome-inactivating proteins from angiosperms. *Plant Mol Biol.* 2014;85(6):575–8. <http://doi.org/10.1007/s11103-014-0204-y>
88. Nguyen LT, Schmidt HA, Haeseler A, Minh BQ. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol Biol Evol.* 2015;32:268–74. <https://doi.org/10.1093/molbev/msu300>
89. Kim D, Paggi JM, Park C, Bennett C, Salzberg SL. Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nat Biotechnol.* 2019;37:907–15. <https://doi.org/10.1038/s41587-019-0201-4>
90. Del CE, Pages M, Gallego M, Monteagudo L, Sánchez-Acedo C. Synaptonemal complex karyotype of *Eimeria tenella*. *Int J Parasitol.* 2005;35:1445–51. <https://doi.org/10.1016/j.ijpara.2005.06.009>
91. Shirley M. The genome of *Eimeria tenella*: further studies on its molecular organisation. *Parasitol Res.* 1994;80:366–73.
92. Khan A, Fujita AW, Randle N, Regidor-Cerrillo J, Shaik JS, Shen K, et al. Global selective sweep of a highly inbred genome of the cattle parasite *Neospora Caninum*. *Proc Natl Acad Sci USA.* 2019;116(15):22764–73. <https://doi.org/10.1073/pnas.1920070116>
93. Brooks CF, Francia ME, Gissot M, Croken MM, Kim K, Striepen B. *Toxoplasma gondii* sequesters centromeres to a specific nuclear region throughout the cell cycle. *Proc Natl Acad Sci USA.* 2011;108(9):3767–72. <https://doi.org/10.1073/pnas.1006741108>
94. Xie HL, Li W, Hu YQ, Yang C, Lu JS, Guo YQ, et al. *De Novo* assembly of human genome at single-cell levels. *Nucleic Acids Res.* 2022;22(13):7479–92. <https://doi.org/10.1093/nar/gkac586>
95. Bohme U, Otto TD, Cotton JA, Steinbiss S, Sanders M, Oyola SO, et al. Complete avian malaria parasite genomes reveal features associated with lineage-specific evolution in birds and mammals. *Genome Res.* 2018;28:547–60. <https://doi.org/10.1101/gr.218123.116>
96. Wang JM, Chen K, Ren QY, Zhang SD, Yang JF, Wang YB, et al. Comparative genomics reveals unique features of two *Babesia motasi* subspecies: *Babesia motasi lintanensis* and *Babesia motasi hebeiensis*. *Int J Parasitol.* 2023;53:265–83. <https://doi.org/10.1016/j.ijpara.2023.02.005>
97. Jain C, Rodriguez-R LM, Phillippy AM, Konstantinidis KT, Aluru S. High throughput ANI analysis of 90K prokaryotic genomes reveals clear species boundaries. *Nat Commun.* 2018;9:5114. <https://doi.org/10.1038/s41467-018-07641-9>
98. Parks DH, Chuvochina M, Chaumeil PA, Rinke C, Mussig AJ, Hugenholtz P. A complete domain-to-species taxonomy for bacteria and archaea. *Nat Biotechnol.* 2020;38:1079–86. <https://doi.org/10.1038/s41587-020-0539-7>
99. Parks DH, Chuvochina M, Waite DW, Rinke C, Skarshewski A, Chaumeil PA. A standardized bacterial taxonomy based on genome phylogeny substantially revises the tree of life. *Nat Biotechnol.* 2018;36:996–1004. <https://doi.org/10.1038/nbt.4229>

100. Blake DP, Vrba V, Xia D, Jatau ID, Spiro S, Nolan MJ, et al. Genetic and biological characterisation of three cryptic *Eimeria* operational taxonomic units that infect chickens (*Gallus gallus domesticus*). *Int J Parasitol*. 2021;51(8):621–34. <https://doi.org/10.1016/j.ijpara.2020.12.004>
101. Liu S, Wang L, Zheng H, Xu Z, Roellig DM, Li N. Comparative genomics reveals *Cyclospora cayentanensis* possesses coccidia-like metabolism and invasion components but unique surface antigens. *BMC Genomics*. 2016;17(1):316. <https://doi.org/10.1186/s12864-016-2632-3>
102. Balaji S, Babu MM, Iyer LM, Aravind L. Discovery of the principal specific transcription factors of apicomplexa and their implication for the evolution of the AP2-integrase DNA binding domains. *Nucleic Acids Res*. 2005;33:3994–4006.
103. Tandel J, Walzer KA, Byerly JH, Pinkston B, Beiting DP, Striepen B. Genetic ablation of a female-specific Apetala 2 transcription factor blocks oocyst shedding in *Cryptosporidium parvum*. *mBio*. 2023;4:e03261–22. <https://doi.org/10.1128/mbio.03261-22>
104. Josling GA, Russell TJ, Venezia J, Orchard L, van Biljon R, Painter HJ, et al. Dissecting the role of PfAP2-G in malaria gametocytogenesis. *Nat Commun*. 2020;11:1503. <https://doi.org/10.1038/s41467-020-15026-0>
105. Shang X, Wang C, Fan Y, Guo G, Wang F, Zhao Y, et al. Genome-wide landscape of ApiAP2 transcription factors reveals a heterochromatin-associated regulatory network during *Plasmodium falciparum* blood stage development. *Nucleic Acids Res*. 2022;50:3413–31. <https://doi.org/10.1093/nar/gkac176>
106. Farhat DC, Hakimi MA. The developmental trajectories of *Toxoplasma* stem from an elaborate epigenetic rewiring. *Trends Parasitol*. 2022;38:37–53.
107. Wang C, Hu D, Tang X, Song X, Wang S, Zhang S, et al. Internal daughter formation of *Toxoplasma gondii* tachyzoites is coordinated by transcription factor TgAP2IX-5. *Cell Microbiol*. 2021;23:e13291. <https://doi.org/10.1111/cmi.13291>
108. Wang LY. Recombinant expression of *Eimeria necatrix* EnApiAP2 protein and its combined immune protective effect with rEtGAM. *Yangzhou Univ*. 2022;1:41. <https://doi.org/10.27441/d.cnki.gyzdu.2022.002489>
109. Cai WM, Feng QQ, Wang LY, Su SJ, Hou ZF, Liu DD, et al. Localization in vivo and in vitro confirms EnApiAP2 protein encoded by ENH_00027130 as a nuclear protein in *Eimeria necatrix*. *Front Cell Infect Microbiol*. 2023;5:131305727. <https://doi.org/10.3389/fcimb.2023.1305727>
110. Chen LL, Tang XM, Sun P, Hu DD, Zhang YY, Wang CY, et al. Comparative transcriptome profiling of *Eimeria tenella* in various developmental stages and functional analysis of an ApiAP2 transcription factor exclusively expressed during sporogony. *Parasites Vectors*. 2023;7(1):19. <https://doi.org/10.1186/s13071-023-05828-8>
111. Morrison DA, Bornstein S, Thebo P, Wernery U, Kinne J, Mattsson JG. The current status of the small subunit rRNA phylogeny of the coccidia (Sporozoa). *Int J Parasitol*. 2004;34:501–14. <https://doi.org/10.1016/j.ijpara.2003.11.006>
112. Waldman BS, Schwarz D, Li MHW, Saeij J, Lourido S. Identification of a master regulator of differentiation in *Toxoplasma*. *Cell*. 2020;180:0. <https://doi.org/10.1016/j.cell.2019.12.013>
113. Walzer KA, Tandel J, Byerly JH, Daniels AM, Gullicksrud JA, Whelan EC, et al. Transcriptional control of the *Cryptosporidium* life cycle. *Nature*. 2024;630:0. <https://doi.org/10.1038/s41586-024-07466-1>
114. Boschet C, Gissot M, Briquet S, Hamid Z, Claudel-Renard C, Vaquero C. Characterization of PfMyb1 transcription factor during erythrocytic development of 3D7 and F12 *Plasmodium falciparum* clones. *Mol Biochem Parasitol*. 2004;138(1):159–63. <https://doi.org/10.1016/j.molbiopara.2004.07.011>
115. Cho CC, Su LH, Huang YC, Pan YJ, Sun CH. Regulation of a Myb transcription factor by cyclin-dependent kinase 2 in *Giardia lamblia*. *J Biol Chem*. 2012;287(6):3733–50. <https://doi.org/10.1074/jbc.M111.298893>

Publisher's note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.