BMC Genomics

RESEARCH Open Access

Chromosome-scale nuclear genome and proteome of *Anoplocephala perfoliata* elucidate lineage-specific features of a 'neglected' equine tapeworm

Neil D. Young[1*], Tao Wang[1], Ching-Seng Ang[2], Dilara Lale[3,4], Hans-Peter Fuehrer[5], Sunita B. Sumanam[1], Pasi K. Korhonen[1], Bill C. H. Chang[1] and Robin B. Gasser[1*]

Abstract

Background *Anoplocephala perfoliata* is the most prevalent and pathogenic tapeworm (cestode) of horses worldwide, yet it remains molecularly understudied. Here, we present the mitochondrial and chromosome-scale nuclear genomes and matched somatic proteome for this parasite, establishing the first high-resolution molecular resource for the family Anoplocephalidae.

Results This parasite was first characterised morphologically and then by its mitochondrial genome (size: 13,776 bp). Its complete nuclear genome (size: 372.3 Mb) was assembled and characterised; it encodes 9,711 protein-coding genes, 78.2% of which were functionally annotated and ~80% supported by transcriptomic evidence. Proteomic analysis confirmed 758 proteins in previously-analysed excretory/secretory (ES) products from adult worms, including highly expressed components of the ubiquitin–proteasome system, stress response families – e.g., translationally controlled tumour proteins (TCTPs) and universal stress proteins (USPs) – and cytoskeletal scaffolds. Approximately 6.5% of the genome contains retroelements, predominantly LINEs. Comparative genomic analyses revealed a relatively conserved synteny with members of the family Taeniidae (*Echinococcus granulosus*, *E. multilocularis* and *Taenia multiceps*) and a pronounced structural divergence from *Hymenolepis microstoma* (Hymenolepididae), reflecting mosaic genome evolution within the order Cyclophyllidea. Classification of proteins inferred from the genome identified GTPases, kinases, peptidases and secretome-associated proteins among the most abundant groups. A subset of proteins exhibited signal peptides or extracellular localisation, suggesting their role as parasite-derived proteins (PDPs) involved in host–parasite communication and immune evasion.

Conclusion This integrated genomic and proteomic framework reveals lineage-specific molecular adaptations in *A. perfoliata* and provides a foundation for future functional and translational investigations of this and closely related cestodes.

*Correspondence:
Neil D. Young
nyoung@unimelb.edu.au
Robin B. Gasser
robinbg@unimelb.edu.au

Full list of author information is available at the end of the article

Keywords *Anoplocephala perfoliata*, Anoplocephalidae, Equine tapeworm, Cestode, Molecular biology, Chromosome-scale genome, Proteome, Comparative genomics, Host–parasite interactions

Background

Tapeworms (class Cestoda; cestodes) are obligate parasitic flatworms that infect all major vertebrate groups, including humans and livestock [1, 2]. Within this class, members of the order Cyclophyllidea account for most described species, and some are responsible for medically and economically important cestodiases worldwide. These parasites exhibit complex life cycles involving definitive and intermediate hosts and display remarkable morphological and developmental adaptations to life inside vertebrate or invertebrate hosts. Their impact on animal health and productivity is substantial, particularly in domestic species such as sheep, cattle and horses, where infection with larval or adult stages can cause chronic disease, productions losses and present an economic burden.

Cestodes are distinguished by a range of unique biological features that reflect their parasitic lifestyle. These include the loss of a digestive tract and reliance on nutrient uptake via a syncytial tegument, the ability to produce segments via a proliferative neck region and highly specialised larval forms adapted for tissue invasion and immune evasion. Molecular interactions at the host–parasite interface – particularly those involving parasite-derived proteins (PDPs) [3] – are central to their survival and persistence. Understanding the genetic and developmental foundations of these traits is critical for elucidating the evolution of parasitism and identifying molecular targets for intervention.

In recent years, the availability of high-quality genome assemblies for selected cestodes – such as *Echinococcus granulosus*, *E. multilocularis*, *Taenia solium*, *T. multiceps* and *Hymenolepis microstoma* – has advanced our understanding of flatworm evolution, gene content and metabolic reduction [4–8]. These resources have revealed lineage-specific patterns of gene expansion and contraction, the apparent loss of canonical metabolic pathways and the diversification of gene families associated with host interaction. However, while genomic data have improved markedly, functional datasets remain sparse across most groups of cestodes, limiting our ability to infer gene function and dissect parasite biology at a mechanistic level.

In particular, the family Anoplocephalidae remains conspicuously absent from current genomic databases, including WormBase Parasite (https://parasite.wormbase.org/) and National Center for Biotechnology Information (NCBI; https://www.ncbi.nlm.nih.gov/). This large and diverse group of tapeworms parasitises herbivorous mammals, including ruminants, camelids and equids. *Anoplocephala perfoliata*, the commonest equine tapeworm worldwide, inhabits the ileocecal junction and caecum of equids including horses, and is frequently associated with mucosal inflammation, ulceration, colic and/or ileo-caecal intussusception, particularly in young or heavily infected animals (e.g. [9, 10]). Its life cycle, involving oribatid mites as intermediate hosts, is well characterised, yet no genomic data are available for this species or its relatives. This absence of information has hindered efforts to place members of the Anoplocephalidae in a broader comparative framework and to understand their specific biological adaptations relative to other tapeworms.

Cestodes also present important models for exploring how developmental pathways are modified and simplified in parasitic lineages. Although they belong to the superphylum Lophotrochozoa, tapeworms lack a through-gut and exhibit profound reductions in organ systems and body plan complexity. Nevertheless, they retain many conserved elements of the metazoan developmental tool kit, including the Wnt, TGF-β, FGF, Notch, Hedgehog, homeobox, Sox and Fox gene families. These genes coordinate patterning, segmentation and tissue differentiation across metazoans and are implicated in unique pathways and processes in flatworms such as strobilation and neoblast-driven regeneration [11–14]. Investigating these gene families in neglected groups such as the Anoplocephalidae would provide an opportunity to assess how developmental systems are remodelled in parasitic lineages.

Another significant area of interest lies in the characterisation of parasite-derived proteins (PDPs) – a broad set of molecules secreted or displayed by the parasite that mediate immune evasion, nutrient uptake, host tissue modulation or pathogenic effects. This group includes classical excretory/secretory (ES) proteins, surface-associated molecules and vesicle-contained cargo [15]. Comparative analyses across cestodes have shown that PDP repertoires are often expanded and exhibit lineage-specific profiles related to host range, tissue tropism and life cycle stage [4, 8]. However, the diversity and biological significance of PDPs in many taxa, including *A. perfoliata*, remain unexplored.

In this context, expanding genomic resources to include *A. perfoliata* will assist in resolving some evolutionary relationships within the Cyclophyllidea and for investigating the molecular basis of parasitism in equines. Moreover, comparative analyses, incorporating representatives from other cyclophyllideans – such as *E. granulosus*, *E. multilocularis*, *Taenia solium*, *T. multiceps* and

Hymenolepis microstoma – should enable deeper insight into the evolutionary origins, developmental trajectories and host-parasite interactions that shape cestode biology.

Results

Genetic characterisation

The complete circular mitochondrial (mt) genome and the linear nuclear ribosomal RNA gene cluster derived from the morphologically identified specimen of *A. perfoliata* (Fig. 1) were 13,776 bp and 8,553 bp in size/length, respectively (Fig. 2A, B). The mt genome (designated mt-Ap-Vienna-nano.v1; GenBank accession PV861922) is notably smaller than that of a previously reported mt genome associated with *A. perfoliata* from a donkey (14,459 bp; GenBank accession no. NC_028425.1; ref [16]), but very similar in length and the same in gene order to a published mt genome assigned to *A. magna* also from a donkey (13,759 bp; accession NC_031801.1; ref [17]). Comparative analysis at the nucleotide level revealed that this sequence was more similar (86.2% identity) to NC_031801.1 (*A. magna*) than to NC_028425.1 (*A. perfoliata*) (84.2% identity) – with variability concentrated in the genes *nad*4, *nad*5 and *nad*6. Phylogenetic analysis using publicly available mt cytochrome *c* oxidase subunit 1 gene (*cox*1) sequences representing *A. perfoliata* (Fig. 2C) supported the grouping of the present worm with European and Japanese *A. perfoliata* specimens to the exclusion of Chinese and Australian representatives. Given that Guo [16, 17] did not describe the diagnostic morphological features of the sequenced specimens, it is plausible that the *A. magna* designation might reflect a misidentification.

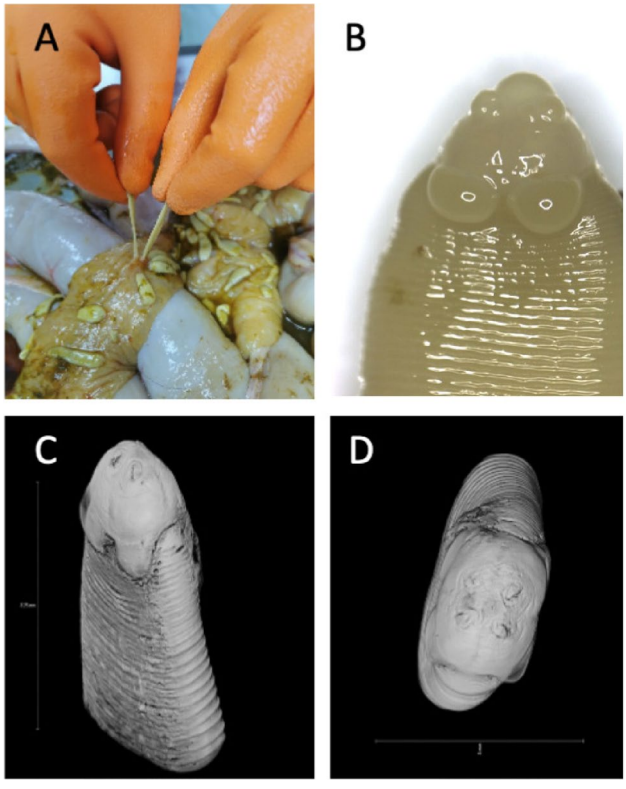

Fig. 1 *Anoplocephala* from a Shetland pony mare. Individual worms were detected on the intestinal wall at *post-mortem* (**a**). Using a dissecting microscope, adult specimens were identified morphologically as *A. perfoliata*, based on the presence of lappets under the suckers (**b**). Computed tomographic imaging was used to enhance the resolution of morphological structures (**c** and **d**); the suckers are contracted because the worms were not adequately relaxed in tap water prior to fixation in ethanol (70% v/v), but the lappets are readily seen

A high-quality nuclear genome

The nuclear genomic data of *A. perfoliata* (Table S1) was assembled into 163 contigs representing a genome size of ~ 372.3 Mb. The assembly was highly contiguous (N50 = 43.3 Mb; maximum contig = 57 Mb) and almost complete (only 0.06% unresolved bases) (cf. Table 1). Benchmarking Universal Single-Copy Orthologs (BUSCO; ref [18]) assessment using 954 conserved metazoan genes identified 68.7% of genes as complete, with 65.6% as single-copy orthologues and 3.1% duplicated. These statistics are typical for parasitic flatworms and provides a strong platform for downstream applications.

Repeat element analysis revealed 76,763 retroelements (6.55% of the genome), with long terminal repeats (LTRs; 66,757 elements; 5.01% of genome) and long-interspersed nuclear elements (LINEs; 9,406 elements; 1.52% of genome) being the most prevalent (Table S2). Short-interspersed nuclear elements (SINEs) and Penelope-like elements were detected at much lower levels – 0.02% and 0.05%, respectively. Other repeat families (e.g., CRE/SLACS) were absent. In addition, 149,968 DNA transposons (27.35% of genome) were predicted, with the mariner-like Tc1-IS630-Pogo elements being most abundant ($n = 33,949$, equating to 7.93% of the genome). These findings suggest a moderate repetitive landscape, with a strong bias towards DNA transposons and LINEs and low expansion of other repeat classes. Additional repeat classification statistics are presented in Table S2.

Gene prediction and annotation support

A total of 9,711 protein-coding genes were predicted from the nuclear genome (mean gene length: 7,428 bp), linked to 10,882 transcripts (mean transcript length: 1,874 bp; average of 6.8 exons per transcript). Untranslated regions were present in 7,606 transcripts. Overall, coding regions comprised 4.8% of the genome. Annotation was supported by transcriptional evidence for 7,756 genes and proteomic evidence for 1,648 genes, yielding 79.9% support overall. Functional annotation indicated that 7,595 genes (78.2%) had orthologues in eggNOG, and 6,089 (62.7%) were linked to KEGG orthologous groups, 3,985 (41.0%) mapped to KEGG pathways, and

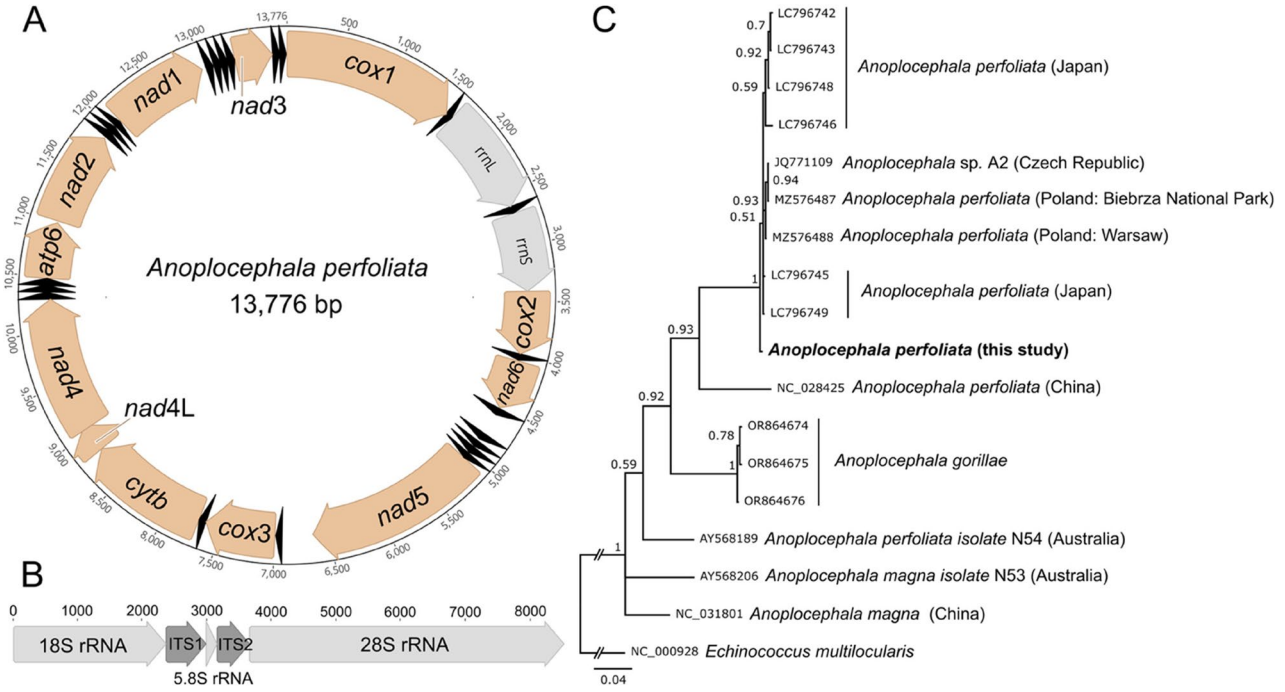

Fig. 2 The mitochondrial genome (**A**) derived from the adult specimen of *Anoplocephala perfoliata* obtained from the present case (Shetland pony mare) – identified morphologically as *A. perfoliata*. Long (*rrn*L) and short (*rrn*S) ribosomal RNA gene subunits are indicated in light grey, the protein-coding genes in orange and tRNA in black. The direction of gene transcription is indicated. Nuclear ribosomal RNA gene cluster (**B**) derived from the same adult specimen of *A. perfoliata*. The ribosomal subunit genes (18 S and 28 S) are indicated in light grey, and the internal transcribed spacers (ITS-1 and ITS-2) in dark grey. A phylogenetic tree (**C**) constructed by Bayesian inference from partial cytochrome *c* oxidase subunit 1 gene (*cox*1) nucleotide sequence data to show the relationship of the present *A. perfoliata* specimen with *Anoplocephala* specimens from other countries characterised previously by *cox*1 sequencing (accession codes indicated at the branches

Table 1 Genomic assembly statistics for the nuclear genome of *Anoplocephala perfoliata*

Description	
Total number of contigs	163
Total sequence length (bp)	372,309,905
N50 (bp)	43,356,837
Min/Max (bp)	1000–57,085,201
Nucleotide gaps	0.06%
Genome BUSCO (*n* = 954; Metazoa) [a]	
Complete	656 (68.7%)
Complete and single-copy BUSCOs	626 (65.6%)
Complete and duplicated	30 (3.1%)
"Fragmented"	49 (5.1%)
"Missing"	249 (26.2%)

[a] Seppey et al. [18]

5,890 (60.7%) were assigned Gene Ontology (GO) terms. BUSCO analysis of the predicted proteome yielded results consistent with the genome-level assessment: 66.3% complete, including 596 single-copy and 36 duplicated orthologues, and 2.8% fragmented (Table 2).

Orthology, phylogeny and synteny

Orthology profiling using 2,843 platyhelminth-specific HOGs identified 2,074 (72.95%) single-copy and 317 (11.15%) duplicated genes, with 452 (15.9%) missing. Only 9.07% of duplications were unexpected. Lineage placement analysis showed that 7,674 proteins (79.02%) matched the expected phylogenetic profile for *A. perfoliata*, with 1,854 partial and 466 fragmented (but consistent) hits. Only 5.6% of proteins showed inconsistent placement; no contaminants were detected. Most proteins (84.63%) aligned with orthologues of other cyclophyllidean representatives (Table 3; refs [5, 7, 19–22]). Comparative analysis with *H. microstoma*, *T. multiceps*, *E. multilocularis* and *E. granulosus* confirmed expected

Table 2 Statistics relating to the nuclear genome of *Anoplocephala perfoliata*

Description	
Number of genes (average length)	9711 (7428)
Number of mRNA (average length)	10,882 (1874)
Number of coding domains (average length)	10,882 (1874)
Number of mRNAs (≥ 1 untranslated region, UTR)	7606
Average exons per mRNA	6.8
Proportion of genome that is protein coding	4.8%
Number of genes with transcriptomic/proteomic support	7756/758
Number of genes with:	
eggNOG annotation	7595 (78.2%)
KEGG orthologous term	6089 (62.7%)
KEGG pathway term	3985 (41.0%)
Gene Ontology term	5890 (60.7%)
Protein BUSCO (Metazoa; 954 groups)	
Complete	632 (66.3%)
Complete and single-copy	596 (62.5%)
Complete and duplicated	36 (3.8%)
Fragmented	27 (2.8%)
Missing	295 (30.9%)
OMark HOGS (Platyhelminthes: 2843)	
Single:	2074 (72.95%)
Duplicated:	317 (11.15%)
Duplicated, unexpected:	258 (9.07%)
Duplicated, expected:	59 (2.08%)
Missing:	452 (15.90%)
Consistent lineage placements	
Total consistent	7674 (79.02%)
Consistent, partial hits	1854 (19.09%)
Consistent, fragmented: 466 (4.49%)	437 (4.50%)
Total inconsistent lineage placements	544 (5.60%)
Total contaminants:	0 (0.00%)
Detected Clade: Taeniidae	
Number of associated query proteins:	8218 (84.63%)

Table 3 Contiguous genomes representing members of the order cyclophyllidea used for synteny and comparative genomic analyses in the present study

Species	Family	Reference(s)	Bioproject; assembly	Genome size (Mb)	BUSCO [a] completeness (%)
Anoplocephala perfoliata	Anoplocephalidae	This study	PRJNA1146763; Chromosome-level	372.31	66.3/68.7
Echinococcus granulosus	Taeniidae	[19]	PRJNA754835; Chromosome-level	172.98	76.9/67.8
Echinococcus multilocularis	Taeniidae	[5]	PRJEB122; Chromosome-level	114.96	79.8/67.4
Taenia multiceps	Taeniidae	[7]	PRJNA307624; Hi-C enhanced	240.61	63.5/66.4
Hymenolepis microstoma	Hymenolepididae	[20, 21, 22]	PRJEB124; Chromosome-level	168.94	78.4/74.7

[a] Annotated genes/genome assembly

phylogenetic relationships (Fig. 3A, left) and revealed distinct patterns of synteny. *A. perfoliata* showed strong collinearity with *T. multiceps*, *E. multilocularis* and *E. granulosus*, but more fragmented alignment with *H. microstoma* (Fig. 3A). For example, the content of chromosome 7 of *A. perfoliata* mapped exclusively to chromosome 6 of *H. microstoma* but was associated with two separate chromosomes (cm038617.1 and cm038619.1) in *E. granulosus* (Fig. 3B & C). In addition, the content of chromosome 8 of *A. perfoliata* mapped to two separate chromosomes in both *H. microstoma* (chromosomes 1 and 3) and *E. granulosus* (cm038613.1 and cm038620.1) (Fig. 3B & C). Genes in syntenic regions numbered 8,198 with *E. multilocularis* (84.4%) and 6,398 with *T. multiceps* (65.9%). Syntenic alignment spanned 84.4% of the *A. perfoliata* genome with *E. multilocularis*, 79.5% with *E. granulosus*, 80.1% with *H. microstoma*, and 65.8% with *T. multiceps* (Fig. 3A). Shared syntenic blocks ranged from 31 with *E. multilocularis* to 150 with *E. granulosus*.

Proteomic profiling of adult *A. perfoliata*

High-resolution proteomic analyses of adult *A. perfoliata* identified and quantified 1,648 proteins from somatic extracts, and 758 unique excretory/secretory (ES) proteins (from a published dataset; ref [23]) (see Fig. 4; Table 4). Of these, 667 proteins (40.5% of the somatic set and 88.0% of the ES set) were common to both compartments, reflecting substantial overlap in functional annotation.

Characterisation of the somatic proteome revealed an enrichment for proteins involved in binding (GO:0005488) and catalytic activity (GO:0003824), including subclasses associated with compound binding, protein binding, hydrolase activity and transferase activity. KEGG annotation mapped the somatic proteome to four core biological systems – genetic information processing, metabolism, cellular processes and environmental information processing – with the largest group comprising genes linked to transcription, translation and protein turnover (e.g. ribosome, spliceosome, RNA degradation pathways) (Tables S3 and S4).

The secretome was defined using high-confidence detection across biological replicates (PXD029144), with 611 ES proteins consistently identified. Protein abundances, estimated via iBAQ, spanned over five orders of magnitude ($\sim 4.6 \times 10^3$ to $\sim 1.6 \times 10^9$), with a median of $\sim 3.2 \times 10^6$ and a mean of $\sim 2.3 \times 10^7$. Annotation using Pfam and eggNOG revealed that 89.1% of ES proteins

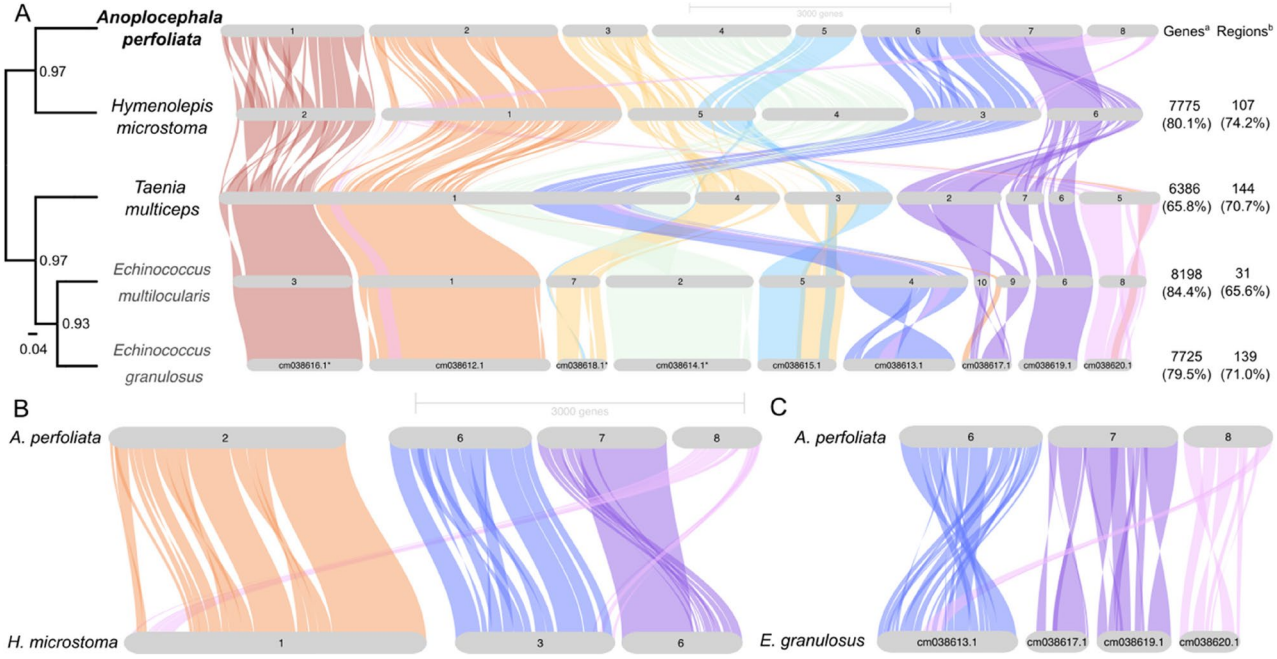

Fig. 3 Comparative synteny and chromosomal contiguity among select cyclophyllidean cestodes for which high quality genomes are publicly available. **A** Left: Phylogenetic tree of five cyclophyllidean cestodes inferred from genome-wide alignments of orthologous single copy genes. Branch support is shown as probability, and scale bar is in substitutions per site. Right: GENESPACE plot depicting syntenic relationship of the eight inferred chromosomes of *Anoplocephala perfoliata* to corresponding syntenic genomic regions in *Hymenolepis microstoma*, *Taenia multiceps*, *Echinococcus multilocularis* and *E. granulosus*. Ribbons represent syntenic regions, with colour-coded flows illustrating conserved synteny with *A. perfoliata* chromosomes (as the reference genome). [a] Total number of *A. perfoliata* genes within species-specific syntenic regions. Percentage of the total gene set is shown in brackets. [b] Total number of *A. perfoliata* syntenic regions shared with each species. Percentage of the total *(A) perfoliata* genome size within syntenic regions is shown in brackets. **B** and **C** GENESPACE plots depicting relationships between select chromosomes of *A. perfoliata* and those of *H. microstoma* and *E. granulosus*, respectively, illustrating marked differences

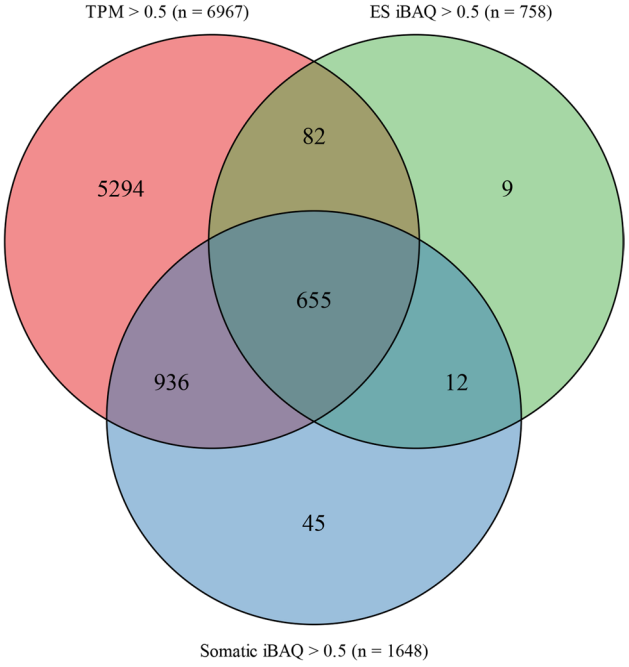

Fig. 4 A Venn diagram showing the numbers of somatic and excretory/secretory (ES) proteins represented in the transcriptome of *Anoplocephala perfoliatia* in the present study. Thresholds of > 0.5 transcripts per million (TPM) and intensity-based absolute quantification (iBAQ) were applied to respective transcriptomic and proteomic data sets

carried identifiable domains, and 90.2% matched known orthologous groups, indicating relative evolutionary conservation of key secretory functions. Integration with the broader genome-wide annotation (Table S3) showed that ES proteins predominantly mapped to core functional categories, including post-translational modification and chaperone systems (COG O; 134 proteins), signalling pathways (COG T; 65), intracellular trafficking and secretion (COG U; 52) and cytoskeletal organisation (COG Z; 60), consistent with roles in host interactions, environmental sensing and tissue invasion. Only a minority of ES proteins (9.8%; 74) lacked COG assignments, suggesting that most components of the secretome are evolutionarily conserved rather than lineage-specific innovations.

Proteasome subunits and threonine-type endopeptidases were prominent among the secreted proteins, reflecting an active ubiquitin–proteasome system. Universal stress proteins (USPs) and translationally controlled tumour proteins (TCTPs) were also detected, consistent with roles in resilience under host-induced stress, although neither group featured among the most abundant proteins. Scaffold and adaptor proteins containing BAR, SH3 and WD40 domains were well represented and might support cytoskeletal organisation, membrane remodelling and signal integration at the host–parasite interface. Another defined subset of ES and surface-associated proteins comprised potential parasite-derived molecules (PDMs), including 59 proteins with predicted signal peptides. These proteins, confirmed by mass spectrometry, represent classical PDMs implicated in immune modulation, host tissue invasion and extracellular communication [15, 24].

Finally, integration of transcriptomic and proteomic datasets revealed that 96.5–97.2% of somatic and ES proteins had corresponding transcripts (Fig. 4). The transcriptome was inferred to encode > 6,900 proteins, 5,294 (~ 76%) of which were not detected proteomically. When corresponding data were available, transcription levels correlated with protein abundances for the somatic (Fig. S1A) and ES (Fig. S1B) proteomes. In addition, transcription levels were higher for genes encoding somatic proteins with supporting proteomic evidence than for those without (Fig. S1C), consistent with transcriptional breadth exceeding proteomic sensitivity. The intersection of all three datasets included 655 proteins, representing a high-confidence core proteome. Only 45 and nine proteins were unique to the somatic and ES datasets, respectively, indicating minimal compartment-specific differences. Taken together, these findings provide a robust, multi-layered characterisation of the *A. perfoliata* proteome and offer insight into parasite biology, stress responses and host–parasite interaction.

Functional classification of protein groups

To define the functional systems represented in the *A. perfoliata* genome, all 9,711 predicted proteins were assigned to KEGG BRITE categories and cross-referenced with transcriptomic, somatic proteomic and ES proteomic data (Table S4). Proteins were grouped into enzymes, genetic information processing and signalling/cellular categories, and quantified by transcriptional support, proteome detection and SCO presence in the ES fraction.

Peptidases and inhibitors ($n = 209$) were strongly supported; 94.7% were transcribed, 41.6% were somatic, 26.8% ES and 33 single copy orthologs (SCOs). Protein kinases ($n = 250$) and phosphatases ($n = 185$) were also well transcribed (> 90%) but had limited ES representation (4.4%, 15.1%). Amino acid–related enzymes ($n = 30$), lipid biosynthesis proteins ($n = 23$) and prenyltransferases ($n = 15$) showed full transcriptional support with 8–30% detection in the ES fraction. Glycosyltransferases ($n = 62$) were not detected in the ES proteome.

Genetic information processing genes were highly transcribed and variably represented in the proteomes. Membrane trafficking ($n = 566$) and chromosome-associated proteins ($n = 453$) were detected in 39.0% and 21.4% of the somatic proteome, and 23.7% and 9.5% were represented in the ES proteome, respectively. Proteasome components ($n = 48$), chaperones ($n = 109$) and translation factors ($n = 58$) were enriched in the ES proteome (66.7%, 34.9% and 20.7%) and included 17, 26 and

Table 4 Summary of the key KEGG BRITE protein groups initially inferred from the gene set of *Anoplocephala perfoliata*, supported by transcription, somatic protein and/or excretory/secretory (ES) evidence. Proteins were classified into broad functional groups based on KEGG BRITE terms inferred from annotation fields in Table S3. This categorisation provides an overview of conserved molecular systems, including enzymes, genetic information processing and signalling components. Some proteins were assigned to multiple groups based on overlapping annotations. For reference, individual BRITE terms and gene-level assignments are provided in Table S3 and can be replicated using the KEGG mapper tool (https://www.genome.jp/kegg/mapper/reconstruct.html)

Classification[a]	Proteins inferred from the genome (%)	Evidence of gene trans-cription (%)	Somatic proteins identified (%)	ES proteins identified [b] (%)	Single copy orthologues (SCOs) encoding ES proteins (%)
Metabolism					
ko01001 Protein kinases	250	226 (90.4)	39 (15.6)	11 (4.4)	5 (2)
ko01002 Peptidases and inhibitors	209	198 (94.7)	87 (41.6)	56 (26.8)	33 (15.8)
ko01009 Protein phosphatases and associated proteins	185	169 (91.4)	48 (25.9)	28 (15.1)	12 (6.5)
ko01003 Glycosyltransferases	62	56 (90.3)	7 (11.3)	(0)	(0)
ko01007 Amino acid related enzymes	30	30 (100)	25 (83.3)	9 (30)	3 (10)
ko01004 Lipid biosynthesis proteins	23	23 (100)	5 (21.7)	2 (8.7)	2 (8.7)
ko01006 Prenyltransferases	15	15 (100)	2 (13.3)	4 (26.7)	4 (26.7)
Genetic information processing					
ko04131 Membrane trafficking	566	522 (92.2)	221 (39)	134 (23.7)	80 (14.1)
ko03036 Chromosome and associated proteins	453	405 (89.4)	97 (21.4)	43 (9.5)	22 (4.9)
ko03000 Transcription factors	278	211 (75.9)	16 (5.8)	6 (2.2)	4 (1.4)
ko04121 Ubiquitin system	266	240 (90.2)	54 (20.3)	29 (10.9)	14 (5.3)
ko03041 Spliceosome	222	217 (97.7)	60 (27)	12 (5.4)	7 (3.2)
ko03019 Messenger RNA biogenesis	220	203 (92.3)	71 (32.3)	31 (14.1)	19 (8.6)
ko03009 Ribosome biogenesis	162	156 (96.3)	29 (17.9)	9 (5.6)	4 (2.5)
ko03400 DNA repair and recombination proteins	155	141 (91)	30 (19.4)	13 (8.4)	9 (5.8)
ko03029 Mitochondrial biogenesis	151	145 (96)	36 (23.8)	11 (7.3)	5 (3.3)
ko03021 Transcription machinery	136	124 (91.2)	14 (10.3)	4 (2.9)	3 (2.2)
ko03110 Chaperones and folding catalysts	109	105 (96.3)	62 (56.9)	38 (34.9)	26 (23.9)
ko03016 Transfer RNA biogenesis	92	88 (95.7)	34 (37)	13 (14.1)	7 (7.6)
ko03011 Ribosome	91	87 (95.6)	41 (45.1)	6 (6.6)	2 (2.2)
ko03032 DNA replication proteins	76	65 (85.5)	7 (9.2)	0 (0)	0 (0)
ko03012 Translation factors	58	56 (96.6)	28 (48.3)	12 (20.7)	8 (13.8)
ko03051 Proteasome	48	46 (95.8)	39 (81.3)	32 (66.7)	17 (35.4)
Signalling and cellular processes					
ko04147 Exosome	311	301 (96.8)	219 (70.4)	146 (46.9)	82 (26.4)
ko02000 Transporters	196	185 (94.4)	29 (14.8)	12 (6.1)	4 (2)
ko04812 Cytoskeleton proteins	183	157 (85.8)	94 (51.4)	64 (35)	33 (18)
ko04040 Ion channels	125	92 (73.6)	4 (3.2)	5 (4)	4 (3.2)
ko03037 Cilium and associated proteins	112	84 (75)	24 (21.4)	11 (9.8)	9 (8)
ko04031 GTP-binding proteins	80	77 (96.3)	49 (61.3)	28 (35)	20 (25)
ko04090 CD molecules	54	49 (90.7)	14 (25.9)	3 (5.6)	2 (3.7)
ko04515 Cell adhesion molecules	48	34 (70.8)	10 (20.8)	1 (2.1)	1 (2.1)
ko04030 G protein-coupled receptors	46	39 (84.8)	0 (0)	0 (0)	0 (0)
ko04990 Domain-containing proteins not elsewhere classified	36	36 (100)	14 (38.9)	9 (25)	6 (16.7)
ko00536 Glycosaminoglycan binding proteins	32	29 (90.6)	14 (43.8)	10 (31.3)	5 (15.6)
ko03310 Nuclear receptors	18	12 (66.7)	1 (5.6)	0 (0)	0 (0)
ko00537 Glycosylphosphatidylinositol - GPI -anchor	15	14 (93.3)	3 (20)	1 (6.7)	1 (6.7)
ko04091 Lectins	14	11 (78.6)	3 (21.4)	2 (14.3)	1 (7.1)
ko00535 Proteoglycans	13	11 (84.6)	3 (23.1)	3 (23.1)	3 (23.1)
ko02044 Secretion system	11	10 (90.9)	3 (27.3)	1 (9.1)	1 (9.1)
ko04052 Cytokines and neuropeptides	10	9 (90)	1 (10)	1 (10)	1 (10)

[a] Some proteins belonged to multiple categories. Table S4 can be used to replicate this protein family summary table and assign

[b] Data derived from a published article [23]

BRITE family annotation to each protein using the KEGG mapper tool (https://www.genome.jp/kegg/mapper/reconstruct.html)

8 SCOs, respectively. Transcription factors ($n = 278$) and DNA replication proteins ($n = 76$) were minimally represented in the ES fraction (< 3%).

Signalling and structural systems were prominent. Exosome-related proteins ($n = 311$) were broadly supported across all modalities (46.9% ES; 82 SCOs). Cytoskeletal proteins ($n = 183$) and GTP-binding proteins ($n = 80$) were well supported (35–25% ES, with 33 and 20 SCOs, respectively). Ion channels ($n = 125$) and GPCRs ($n = 46$) were largely absent from proteomic datasets. Smaller groups such as cell adhesion molecules, cytokines, glycosaminoglycan-binding proteins and lectins were variably detected in the ES proteome. CD molecule–like proteins ($n = 54$) and GPI-anchored proteins ($n = 15$) were identified; these likely reflect structural or immunomodulatory proteins inferred from domain-based annotations (e.g., ko04090 and ko00537).

Discussion

The definition of a chromosome-scale nuclear genome and matched proteome for *A. perfoliata* represents a critical advance in understanding the molecular biology of anoplocephaline tapeworms. As the first such resource for the family Anoplocephalidae, this dataset fills a critical gap in the genomics of the Cyclophyllidea and provides a foundation for interpreting the molecular basis of host–parasite interactions, parasite adaptation and pathogenesis.

Comparative mitogenomic analyses positioned *A. perfoliata* within a lineage comprising European and Japanese representatives of the species, clearly genetically distinct from Chinese and Australian isolates (Fig. 2), and the nuclear genomic comparison revealed conserved chromosomal architecture interspersed with substantial rearrangements (Fig. 3A-C). The latter mosaic synteny is consistent with the dynamic nature of flatworm genomes and may reflect selective pressures imposed by distinct life cycles and host niches. Repeat content, primarily composed of LINE elements, was moderate but well defined, further supporting the structural integrity of the genome. BUSCO metrics (68.7% completeness) and functional annotation (>78% of genes) confirm the quality and representativeness of the nuclear genome.

Proteomic profiling of somatic tissues of adult *A. perfoliata* revealed strong enrichment in molecular systems essential for intracellular stability and stress mitigation, including protein folding, degradation and cytoskeletal organisation. The prominence of proteasome subunits, universal stress proteins (USPs) and translationally controlled tumour proteins (TCTPs) might reflect cellular adaptations to the dynamic, immunologically active environment in the host gut. Similarly, WD40, BAR and SH3 domain–containing proteins indicate the importance of membrane scaffolding and signalling cross-talk in parasite physiology. These proteins may underpin structural plasticity and signalling responsiveness at the host–parasite interface.

Within the proteome, a subset of proteins carried predicted signal peptides or were annotated as extracellular, consistent with parasite-derived molecules (PDPs) being involved in host–parasite communication. Notably, several proteins were annotated with KEGG BRITE terms associated with exosome biogenesis or vesicle trafficking (e.g., ko04147), indicating a capacity for vesicle-mediated delivery of effectors into the host environment. The enrichment of proteins linked to cytoskeletal dynamics, vesicle transport and proteasome function – many of which are also detected in the excretory/secretory (ES) proteome – supports the existence of an active extracellular vesicle (EV) system. This finding is consistent with secretory strategies observed in other parasitic flatworms and suggests that *A. perfoliata* might deploy vesicle-contained molecular cargo to modulate host tissue and immune responses.

Proteins including proteasome subunits, TCTPs and USPs, all detected in high abundance and with transcriptomic support, have been proposed to localise within helminth EVs, and may serve roles in immune modulation or stress adaptation [25, 26]. While their presence in cestode EVs remains to be confirmed experimentally, these findings are consistent with proteomic studies of trematodes and nematodes, and suggest that *A. perfoliata* may exploit a similar strategy, including tetraspanins, 14-3-3 proteins, Rab GTPases and HSP70, all of which have been reported in EVs of flatworms including *E. granulosus* and *E. multilocularis* [27–31].

Additionally, the abundant presence of scaffold proteins with WD40, SH3 and BAR domains supports a structural and mechanistic basis for vesicle formation and trafficking. These domains are integral to membrane remodelling, curvature sensing and protein–protein interactions, and their presence, together with annexins, actins, enolases and GAPDH-like proteins, further supports the conservation of EV cargo components for cestodes and other flatworms [27–29, 32].

The integration of multi-omic layers identified a core proteome (>650 proteins) consistently detected across the transcriptomic, somatic and ES datasets. These molecules, enriched in chaperones, cytoskeletal elements, metabolic enzymes and stress response proteins, represent ideal candidates for further experimental characterisation. Many were also represented in KEGG BRITE and COG functional categories, suggesting their central role in parasite viability.

While this study establishes a foundational molecular resource for *A. perfoliata*, several avenues remain open for exploration. Experimental validation of predicted parasite-derived proteins, particularly through the

isolation and proteomic characterisation of EVs from *A. perfoliata*, would provide direct evidence for their roles in host–parasite signalling. Similarly, comparative transcriptomic or proteomic profiling of larval stages, such as the cysticercoid, could uncover stage-specific molecules involved in developmental transitions, host tissue tropism and/or immune evasion. Unannotated or hypothetical proteins with strong protein expression signals in the secretome also warrant prioritisation for functional characterisation, particularly those without COG or KEGG annotations.

Given the role of *A. perfoliata* in equine colic and ileocaecal intussusception, and its broad global distribution, the present datasets also hold translational potential. Secreted and surface-associated proteins with strong evidence of expression and immunogenic domain features could be pursued as serodiagnostic targets or subunit vaccine candidates. Moreover, the shared molecular architecture observed between *A. perfoliata* and taeniid cestodes (e.g., *Echinococcus* and *Taenia*) provides opportunities for cross-applicable interventions, particularly in veterinary and medical contexts. The availability of a high-quality genome and proteome now enables targeted studies of cestode biology and host–parasite interactions in a neglected, yet clinically highly relevant parasite.

In conclusion, this study delivers the first chromosome-scale nuclear genome and matched proteome for *A. perfoliata*, a tapeworm of major clinical significance in equine medicine. The genome provides a structurally robust and functionally rich platform for molecular investigations, while the proteomic data offer new insight into core biological systems and parasite-derived molecules potentially involved in host–parasite communication. Integration of transcriptomic, somatic and excretory/secretory proteome datasets defines a core proteome, including subsets of molecules of potential diagnostic, vaccine or therapeutic value. Together, these resources lay a solid foundation for future experimental and comparative research on cestode biology and might contribute to improved control of a significant parasitic infection/disease of horses.

Methods

Parasite

Live, adult *A. perfoliata* worms were collected from the intestine during *post-mortem* examination of a 24-year-old Shetland pony mare in Austria. Morphological assessment, including computed tomography (CT) imaging (following staining with 1% phosphotungstic acid; ref [33]), revealed the characteristic presence of lappets beneath the suckers (Fig. 1; cf [1, 34, 35], allowing each worm for molecular investigation to be identified to a species level. Individual worms were preserved in 70% ethanol or RNAlater (Thermo Fisher Scientific, Waltham, MA, USA) and shipped to the University of Melbourne for molecular analyses.

Nucleic acid isolation, library construction and sequencing

A neck region (8 mm; ~75 mg) of an RNAlater-preserved adult worm was used to isolate high-quality total genomic DNA (150 ng) using the Nanobind Tissue Big DNA kit (PacBio, Menlo Park, CA, USA). The integrity of DNA was verified using 4200 TapeStation System Genomic DNA ScreenTape Assay (Agilent Technologies, Waldbronn, Germany). A SMRTbell DNA library was constructed from 15 µg of Repli-G (Qiagen) whole genome amplified (WGA) DNA and sequenced using the PacBio-HiFi Revio sequencing platform (PacBio, Menlo Park, CA, USA). In addition, long WGA DNA reads were sequenced using Oxford Nanopore platform. DNA (1 µg) was used to create a DNA library using a nanopore DNA sequencing kit (SQK-LSK114; Oxford Nanopore Technologies, Oxford, UK). This library was sequenced in a MinION sequencer (Oxford Nanopore Technologies) for 72 h using an R9.4.1 flow cell (FLO-MIN106). Following sequencing, bases were 'called' from raw POD5 data using the program Dorado v0.5.0 (Oxford Nanopore Technologies) using the 'super accurate' model and stored in the FASTQ format. High-throughput chromosome conformation capture (Hi-C) sequencing was performed using the Arima-HiC + for high coverage kit according to manufacturer's instructions (Arima Genomics, Carlsbad, CA, USA). In brief, 75 mg of ethanol-preserved proglottid tissue was cross-linked and restriction digested using the Arima enzyme mix (L/N2309050008; part no. A160162 v01; Arima Genomics, CA, USA). The proximally-ligated DNA was fragmented to 550–660 bp using an ultra-sonicator (M220; Covaris, USA) and used to construct an indexed Illumina library. The quality of the library was assessed using the TapeStation (as described above) and paired-end sequenced in one lane of a Nova Seq X sequencing platform (10B, 300 cycles; Illumina, USA).

Assembly of the mt genome and associated analyses

The complete mt genome, designated as mt–Ap_Viennanano.v1, was assembled using long-read sequencing data with FLYE v.2.6 [36], applying the --nano-raw option. Initial error correction of the long-read sequence data was conducted using the medaka_consensus tool from the Medaka package v.0.10.0 (https://github.com/nanoporetech/medaka). Subsequently, the long reads were aligned to the assembled mt genome with minimap2 v.2.0 [37], and genome coverage was assessed using the mpileup tool in the SAMtools package v.1.9 [38]. The annotation of protein-coding gene regions was initially performed using the MITOS webserver [39] with the mt genetic code for invertebrates (https://www.ncbi.nlm.nih

.gov/Taxonomy/Utils/wprintgc.cgi; translation_table_5). Each protein gene's open reading frame (ORF) was then manually curated using Geneious v.11.1.5 [40] and gene regions identified (Table 1) based on comparison with published mt genomes purported to represent *A. perfoliata* and *A. magna* (accession codes NC_028425.1 and NC_031801.1). The complete mt genome sequence was submitted to the GenBank database under accession PV861922.

The genetic relationship of the *A. perfoliata* specimen studied here with others from different countries was inferred using publicly accessible mt *cox*1 sequences, with *E. multilocularis* as an outgroup (Fig. 2C). Nucleotide sequences were aligned over a consensus length 598 bp using MAFFT, and data were subjected to analysis by Bayesian inference (BI) in MrBayes v3.2.6 [41] using the GTR + I + G substitution model for 1,000,000 generations. A consensus tree was constructed from the final 75% of trees.

Assembly of the nuclear genome

To obtain a draft nuclear genome for *A. perfoliata*, PacBio and Oxford Nanopore long- and Hi-C short-reads were assembled using hifiasm v.0.19.8 [42]; https://github.com/chhylp123/hifiasm). Then, haplotypic sequences were removed from the assembly using purge_haplotigs v.1.1.0 [43]. The Hi-C reads were mapped to assembled contigs using Chromap v.0.2.5 [44] with the "--preset hic" option. Mapped Hi-C reads were then used to scaffold contigs using YaHS v.1.1 [45]. Curation of inferred chromosomes was undertaken using Juicebox v.2.20.00 [46], and final contact maps were visualised using HiContacts v.1.0 [47]. Gaps in scaffolds were closed using PacBio long reads employing the program DENTIST v.4.0.0 [48]. At each step, assembly results were assessed using BUSCO v.5.1.2 [18].

Isolation of RNA and sequencing

Total RNA was isolated from 50 mg of RNAlater-preserved proglottid tissue using the TriPure Isolation Reagent (Sigma Aldrich) and DNase-treated using a TURBO DNA-free™ kit (Thermo Fisher Scientific). The size, integrity and amount of RNA were determined using a 4200 TapeStation System RNA ScreenTape Assay (Agilent Technologies, Waldbronn, Germany) and a Qubit® 3.0 fluorometer RNA High Sensitivity Assay (Life Technologies, Carlsbad, California, USA). Messenger RNA (mRNA) was purified from total RNA using the Dynabeads® mRNA Purification Kit (Thermo Fisher Scientific). The mRNA was reverse transcribed, and a long-read library was prepared using the Oxford Nanopore cDNA-sequencing kit (SQK-PCS114; Oxford Nanopore Technologies), according to the manufacturer's instructions. The prepared library was sequenced on a Promethion sequencer (Oxford Nanopore Technologies) for 24 h using an R10.4.1 flow cell (FLO-PRO114M). Following sequencing, bases were 'called' and stored as described for the DNA library.

Gene prediction and annotation

First, custom repeat models, inferred from the *A. perfoliata* nuclear genome assembly using RepeatModeler v.2.0.4 [49], were annotated using Terrier v.0.2.0 [50]. The final library was used to soft-mask repeats in the assembled genome using RepeatMasker v.4.1.5 [51]. Gene models were predicted from the soft-masked genome using Braker v.3.0.3 [52, 53] and GUSHRC v.1.6.2 [54, 55]. Long-read cDNA sequences were mapped to the genome assembly using minimap2 v.2.26 [38] and SAMtools v.1.17 [38] and using the "-ax splice -N 1" options. A cDNA alignment BAM file and the UniProtKB/Swiss-Prot [56] protein sequences (downloaded in May 2024) were used in Braker in ETP mode. The completeness of the gene set was assessed using BUSCO v.5.1.2 [18] and OMark v.0.3.0 [57]. Genes and inferred proteins were annotated using eggNOG-mapper v.2.1.9 [58].

Synteny and comparative genomic analyses

Syntenic relationships were assessed between the nuclear genome of *A. perfoliata* and the most contiguous nuclear genomes of cestodes currently available: *E. granulosus*, *E. multilocularis*, *H. microstoma* and *T. multiceps* (Cyclophyllidea) (Table 3). Single-copy genes and one-to-one orthologs were identified in a pairwise manner among the five species of cestodes using OrthoFinder v.2.5.4 [59]. Subsequently, GENESPACE v.1.2.3 [60] was employed to assess and display synteny and orthology across these genomes. In addition, single-copy genes and one-to-one orthologs were identified among the five cyclophyllidean gene sets using OrthoFinder v.2.5.4.

Mass spectrometry analysis of the somatic proteome

Proteins were extracted from four distinct sections (each comprising 3–4 proglottids; 1 mm long x 5 mm wide) from an adult *A. perfoliata*. These four samples (replicates) were each homogenised in 8 M urea with 100 mM triethylammonium bicarbonate (pH 8.5) and sonicated on ice (8 × 20 s pulses at 20 kHz). Lysates were treated with protease inhibitors, incubated at 23 °C and centrifuged at 12,000 ×g for 30 min at 4 °C. Protein concentrations were measured using a BCA assay.

For in-solution digestion, 40 µg of protein per sample were reduced with 10 mM TCEP at 55 °C for 45 min, alkylated with 55 mM iodoacetamide at 22 °C in the dark for 30 min, and digested overnight at 37 °C using a Lys-C/trypsin mix. Peptides were acidified, purified using Oasis HLB cartridges. Samples were freeze-dried and reconstituted in 2% acetonitrile with 0.05% trifluoroacetic acid

prior to liquid chromatography-tandem mass spectrometry (LC-MS/MS) analysis. LC-MS/MS was performed on an Orbitrap Ascend mass spectrometer coupled to a Dionex Ultimate 3000 RSLC nano-HPLC. Peptides were separated on an Acclaim PepMap RSLC C18 column using a 3–40% gradient of solvent B (acetonitrile with 0.1% formic acid and 5% DMSO) over 82 min. MS1 scans were acquired at 120,000 resolution. The top-speed data-dependent acquisition method (3 s cycle) was used for precursor selection (charge states 2–7), followed by HCD fragmentation and MS2 scans at 15,000 resolution. Dynamic exclusion was set to 30 s. Full MS scans were acquired from 350 to 1400 m/z at 120,000 resolution (AGC target 250%, IT 50 ms), followed by 50 staggered isolation windows (13.7 Da with 1 Da overlap). MS2 scans were collected at 30,000 resolution (AGC 2000%, IT 55 ms) with a normalised collision energy of 30%.

Protein identification and quantification were performed using Spectronaut (v19.4) employing a data-independent acquisition (DIA) workflow against the *A. perfoliata* proteome predicted from the nuclear genome. Trypsin specificity (≤ 2 missed cleavages), fixed carbamidomethylation (C) and variable oxidation (M) and N-terminal acetylation were used. FDR thresholds were set at 1% for protein and PSM levels. MS2-level quantification was performed with automatic cross-run normalisation. Statistical analyses were conducted in Perseus (v1.6.1.1). Only proteins identified in four biological replicates were retained. Mass spectrometry proteomics data have been deposited in the ProteomeXchange Consortium via PRIDE (identifier PXD067162).

Identification of ES proteins

Raw MS data from a previous data-dependent acquisition (DDA) based ES protein analysis of *A. perfoliata* [23] was obtained from PRIDE (identifier PXD029144). and processed using MaxQuant (v.2.4.0.0) [61] for the identification of proteins in the *A. perfoliata* proteome predicted from genes of the nuclear genome characterised here. The search parameters were based on Hautala et al. [23] – using a precursor tolerance of 20 ppm, MS/MS tolerance of 4.5 ppm, fixed modifications of carbamidomethylation of cysteine (+ 57 Da) and methionine oxidation (+ 16 Da). Only proteins identified in all four biological replicates representing at least one sample set (cf [23]) were retained. FDR thresholds were set at 1% at both the peptide and protein levels.

Bioinformatic analyses of protein data sets

The somatic proteomic data produced here (identifier PXD067162) as well as a previously published ES protein data (identifier PXD029144; ref [23]) were each analysed using a suite of tools – functional annotation of differentially expressed proteins was refined using EggNOG-mapper v2.1.9 [58]; pathway assignment was based on KEGG orthology (KO) terms retrieved from the Kyoto Encyclopedia of Genes and Genomes (KEGG) [62]; and protein sets were compared using a Venn diagram.

Abbreviations

Ap	*Anoplocephala perfoliata*
bp	Base pairs
BRITE	Biological Relationship and Information Transmission Entity
BUSCO	Benchmarking Universal Single-Copy Orthologs
CD	Cluster of differentiation
CT	Computed tomography
DDA	Data-dependent acquisition
DIA	Data-independent acquisition
DNA	Deoxyribonucleic acid
ES	Excretory/secretory
EV	Extracellular vesicle
FDR	False discovery rate
FGF	Fibroblast growth factor
GO	Gene Ontology
GPCR	G protein-coupled receptor
HOG	Hierarchical orthologous group
HSP	Heat shock protein
iBAQ	Intensity-based absolute quantification
ITS	Internal transcribed spacer
KEGG	Kyoto Encyclopedia of Genes and Genomes
KO	KEGG Orthology
LC–MS/MS	Liquid chromatography tandem mass spectrometry
LINE	Long interspersed nuclear element
LTR	Long terminal repeat
mRNA	Messenger RNA
Mt	Mitochondrial
ORF	Open reading frame
PacBio	Pacific Biosciences
PDP	Parasite-derived protein
PDM	Parasite-derived molecule
PRIDE	Proteomics Identifications Database
RNA	Ribonucleic acid
RNA-seq	RNA sequencing
RIN	RNA integrity number
SCO	Single-copy orthologue
SINE	Short interspersed nuclear element
SMRT	Single-molecule real-time (sequencing)
TCTP	Translationally controlled tumour protein
TPM	Transcripts per million
UTR	Untranslated region
USP	Universal stress protein
WGA	Whole genome amplification

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12864-026-12554-9.

Supplementary Material 1.

Supplementary Material 2.

Acknowledgements

We thank Maria Unterköfler (Parasitology Unit, Veterinary Medical University, Vienna, Austria) for identifying the adult specimen of *A. perfoliata*.

Authors' contributions

D.L. and H.-P.F. collected the samples and provided the case history. N.D.Y. and R.B.G. designed the molecular study. S.B.S. and N.D.Y. conducted the genomic and transcriptomic analyses. T.W. and C.-S.A. carried out the proteomic analysis, with bioinformatic support from P.K.K. N.D.Y. conducted the bioinformatic analyses of genomic and transcriptomic datasets. B.C.H.C. provided infrastructure support. R.B.G., N.D.Y. and T.W. drafted the initial

manuscript with inputs from other authors. N.D.Y., B.C.H.C. and R.B.G. supervised the project. All authors read and approved the final manuscript.

Funding
This project was supported through a Future Fellowship (FT230100559 to NDY) and grants (LP180101334 to NDY; LP180101085 and LP220200614 to RBG) from the Australian Research Council (ARC).

Data availability
The genome bioproject is accessible via National Center for Biotechnology Information (NCBI) –PRJNA1146763. Raw sequence data are available via the NCBI Sequence Read Archive – BioProject PRJNA1146763. Proteomic data are available from the ProteomeXchange Consortium via PRIDE –PXD067162 and PXD029144.

Declarations

Ethics approval and consent to participate
Not applicable.

Consent for publication
Not applicable.

Competing interests
The authors declare no competing interests.

Author details
[1]Department of Veterinary Biosciences, Melbourne Veterinary School, The University of Melbourne, Parkville, VIC 3010, Australia
[2]Bio21 Mass Spectrometry and Proteomics Facility, The University of Melbourne, Parkville, VIC 3010, Australia
[3]University Equine Hospital, Clinical Unit of Equine Internal Medicine, University of Veterinary Medicine Vienna, Vienna, Austria
[4]Institute of Virology, Vetsuisse Faculty, University of Zurich, Zurich, Switzerland
[5]Institute of Parasitology, University of Veterinary Medicine Vienna, Vienna 1210, Austria

Received: 30 September 2025 / Accepted: 13 January 2026
Published online: 21 January 2026

References

1. Beveridge I, Family Anoplocephalidae Cholodkovsky, 1902. In: Khalil LF, Jones A, Bray RA, editors. Keys to the cestode parasites of vertebrates. Wallingford: CAB International; 1994. pp. 315–366. ISBN 0851988792.
2. Caira JN, Jensen K. Planetary biodiversity inventory (2008–2017): Tapeworms from vertebrate bowels of the Earth. The University of Kansas, Natural History Museum, Special Publication No. 25. 2017. pp 1-463. ISBN 9780998857803.
3. Geary TG. New paradigms in research on *Dirofilaria immitis*. Parasites Vectors. 2023;16:247.
4. Olson PD, Zarowiecki M, Kiss F, Brehm K. Cestode genomics – progress and prospects for advancing basic and applied aspects of flatworm biology. Parasite Immunol. 2012;34:130–50.
5. Tsai IJ, Zarowiecki M, Holroyd N, Garciarrubio A, Sánchez-Flores A, Brooks KL, Tracey A, Bobes RJ, Fragoso G, Sciutto E, Aslett M, Beasley H, Bennett HM, Cai X, Camicia F, Clark R, Cucher M, De Silva N, Day TA, Deplazes P, Estrada K, Fernández C, Holland PWH, Hou J, Hu S, Huckvale T, Hung SS, Kamenetzky L, Keane JA, Kiss F, Koziol U, Lambert O, Liu K, Luo X, Luo Y, Macchiaroli N, Nichol S, Paps J, Parkinson J, Pouchkina-Stantcheva N, Riddiford N, Rosenzvit M, Salinas G, Wasmuth JD, Zamanian M, Zheng Y, Taenia solium Genome Consortium, Cai J, Soberón X, Olson PD, Laclette JP, Brehm K, Berriman M. The genomes of four tapeworm species reveal adaptations to parasitism. Nature. 2013;496:57–63.
6. Zheng H, Zhang W, Zhang L, Zhang Z, Li J, Lu G, Zhu Y, Wang Y, Huang Y, Liu J, Kang H, Chen J, Wang L, Chen A, Yu S, Gao Z, Jin L, Gu W, Wang Z, Zhao L, Shi B, Wen H, Lin R, Jones MK, Brejova B, Vinar T, Zhao G, McManus DP, Chen Z, Zhou Y, Wang S. The genome of the hydatid tapeworm *Echinococcus granulosus*. Nat Genet. 2013;45:1168–75.
7. Li W, Liu B, Yang Y, Ren Y, Wang S, Liu C, Zhang N, Qu Z, Yang W, Zhang Y, Yan H, Jiang F, Li L, Li S, Jia W, Yin H, Cai X, Liu T, McManus DP, Fan W, Fu B. The genome of tapeworm *Taenia multiceps* sheds light on Understanding parasitic mechanism and control of coenurosis disease. DNA Res. 2018;25:499–510.
8. International Helminth Genomes Consortium. Comparative genomics of the major parasitic worms. Nat Genet. 2019;51:163–74.
9. Williamson RMC, Gasser RB, Middleton D, Beveridge I. The distribution of *Anoplocephala perfoliata* in the intestine of the horse and associated pathological changes. Vet Parasitol. 1997;73:225–41.
10. Gasser RB, Williamson RMC, Beveridge I. *Anoplocephala perfoliata* of horses—significant scope for further research, improved diagnosis and control. Parasitology. 2005;131:1–13.
11. Riddiford N, Olson PD. Wnt gene loss in flatworms: taxonomic and phylogenetic implications. Evol Dev. 2011;13:176–87.
12. Thompson RCA, Lymbery AJ. Let's not forget the thinkers: the legacy of Desmond Smyth and fundamental questions in parasite biology. Trends Parasitol. 2013;29:581–4.
13. Koziol U, Brehm K. Recent advances in *Echinococcus* genomics and stem cell research. Vet Parasitol. 2015;213:92–102.
14. Koziol U. Evolutionary developmental biology of tapeworms: recent progress and future perspectives. EvoDevo. 2017;8:12.
15. Hewitson JP, Grainger JR, Maizels RM. Helminth immunoregulation: the role of parasite secreted proteins in modulating host immunity. Mol Biochem Parasitol. 2009;167:1–11.
16. Guo A. The complete mitochondrial genome of *Anoplocephala perfoliata*, the first representative for the family anoplocephalidae. Parasites Vectors. 2015;8:549.
17. Guo A. Complete mitochondrial genome of *Anoplocephala magna* solidifying the species. Korean J Parasitol. 2016;54:369–73.
18. Seppey M, Manni M, Zdobnov EM. BUSCO: assessing genome assembly and annotation completeness. Methods Mol Biol. 2019;1962:227–45.
19. Korhonen PK, Kinkar L, Young ND, Cai H, Lightowlers MW, Gauci C, Jabbar A, Chang BCH, Wang T, Hofmann A, Koehler AV, Li J, Li J, Wang D, Yin J, Yang H, Jenkins DJ, Saarma U, Laurimäe T, Rostami-Nejad M, Irshadullah M, Mirhendi H, Sharbatkhori M, Ponce-Gordo F, Simsek S, Casulli A, Zait H, Atoyan H, de la Rue ML, Romig T, Wassermann M, Aghayan SA, Gevorgyan H, Yang B, Gasser RB. Chromosome-scale *Echinococcus granulosus* (genotype G1) genome reveals the Eg95 gene family and conservation of the EG95-vaccine molecule. Commun Biol. 2022;5:199.
20. Olson PD, Zarowiecki M, James K, Baillie A, Bartl G, Burchell P, Chellappoo A, Jarero F, Tan LY, Holroyd N, Berriman M. Genome-wide transcriptome profiling and Spatial expression analyses identify signals and switches of development in tapeworms. EvoDevo. 2018;9:21.
21. Olson PD, Tracey A, Baillie A, James K, Doyle SR, Buddenborg SK, Rodgers FH, Holroyd N, Berriman M. Complete representation of a tapeworm genome reveals chromosomes capped by centromeres, necessitating a dual role in segregation and protection. BMC Biol. 2020;18:165.
22. Cunningham LJ, Olson PD. Description of *Hymenolepis microstoma* (Nottingham strain): a classical tapeworm model for research in the genomic era. Parasites Vectors. 2010;3:123.
23. Hautala K, Pursiainen J, Näreaho A, Nyman T, Varmanen P, Sukura A, Nielsen MK, Savijoki K. Label-free quantitative proteomics and Immunoblotting identifies immunoreactive and other excretory-secretory (E/S) proteins of *Anoplocephala perfoliata*. Front Immunol. 2022;13:1045468.
24. Rogers M, Kmath S, McManus D, Jones M, Gordon C, Navarro S. *Schistosoma* excretory/secretory products: an untapped library of tolerogenic immunotherapeutics against food allergy. Clin Transl Immunol. 2024;13:e70007.
25. Drurey C, Maizels RM. Helminth extracellular vesicles: interactions with the host immune system. Mol Immunol. 2021;137:124–33.
26. Marcilla A, Trelis M, Cortés A, Sotillo J, Cantalapiedra F, Minguez MT, Valero ML, del Sánchez MM, Muñoz-Antolí C, Toledo R, Bernal D. Extracellular vesicles from parasitic helminths contain specific excretory/secretory proteins and are internalised in intestinal host cells. PLoS ONE. 2012;7:e45974.
27. Cwiklinski K, de la Torre-Escudero E, Trelis M, Bernal D, Dufresne PJ, Brennan GP, O'Neill SM, Tort JF, Paterson S, Marcilla A, Dalton JP. The extracellular vesicles of the helminth pathogen *Fasciola hepatica*: biogenesis, protein cargo and functional biology. Mol Cell Proteom. 2015;14:3258–73.
28. Nowacki FC, Swain MT, Klychnikov OI, Niazi U, Ivens A, Quinnell RJ, Maizels RM, Hall R. Protein and small noncoding RNA-enriched extracellular vesicles are released by the pathogenic blood fluke *Schistosoma mansoni*. J Extracell Vesicles. 2015;4:28665.

29. Zheng H, Zhang W, Zhang L, Zhang Z, Li J, Lu G, Zhu Y, Wang Y, Liu J, Zhang B, Zhou Y, Zheng Y, Wang S, Yan X, Wang X, Yu S, Lin R, Zhao W, McManus DP, Wen H. Regulatory effects of *Echinococcus multilocularis* extracellular vesicles on human monocytes and macrophages. Int J Parasitol. 2017;47:675–85.
30. Siles-Lucas M, Sánchez-Ovejero C, González-Sánchez M, González E, García MP, Zafra R, Pérez-Arellano JL, Martínez-Sernández V, Muñoz C, Gárate T. Extracellular vesicles from *Echinococcus granulosus* metacestodes: analysis of protein content and potential functions. PLoS Negl Trop Dis. 2018;12:e0007032.
31. Liu F, Li X, Wang L, Zhao Y, Zhang Z, Hou J, Wen H, Craig PS, Giraudon P, Vuitton DA, Wang J. Proteomic profiling of serum extracellular vesicles identifies diagnostic markers for echinococcosis. PLoS Negl Trop Dis. 2022;16:e0010814.
32. Siles-Lucas M, Morchón R, Simon F, Manzano-Román R. Cestode parasites release extracellular vesicles with microRNAs and immunodiagnostic protein cargo. Int J Parasitol. 2017;47:675–86.
33. Penna-Gonçalves V, Willmott NJ, Kelly MBJ, Black JR, Lowe EC, Herberstein ME. Comparing MicroCT staining and scanning methodology for brain studies in various sizes of spiders. J Comp Neurol. 2025;533:e70017. https://doi.org/10.1002/cne.70017.
34. Kahane Z. Anatomie von *Taenia perfoliata* Göze, als Beitrag Zur Kenntnis der Zestoden. Z Wiss Zool. 1880;34:175–254.
35. Spasskii AA. Anoplocephalate tapeworms of domestic and wild animals. In: Skrjabin KI, editor. *Essentials of Cestology*. Vol. 1. Israel Program for Scientific Translations, Jerusalem (English translation, 1966); 1951. pp 1-783.ISBN 9780706505785.
36. Kolmogorov M, Yuan J, Lin Y, Pevzner PA. Assembly of long, error-prone reads using repeat graphs. Nat Methods. 2019;16:753–60.
37. Li H. Minimap2: pairwise alignment for nucleotide sequences. Bioinformatics. 2018;34:3094–100.
38. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R. The sequence alignment/map format and samtools. Bioinformatics. 2009;25:2078–9.
39. Bernt M, Donath A, Jühling F, Externbrink F, Florentz C, Fritzsch G, Pütz J, Middendorf M, Stadler PF. MITOS: improved de novo metazoan mitochondrial genome annotation. Mol Phylogenet Evol. 2013;69:313–9.
40. Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, Buxton S, Cooper A, Markowitz S, Duran C, Thierer T, Ashton B, Meintjes P, Drummond A. Geneious basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics. 2012;28:1647–9.
41. Ronquist F, Teslenko M, van der Mark P, Ayres DL, Darling A, Höhna S, Larget B, Liu L, Suchard MA, Huelsenbeck JP. MrBayes 3.2: efficient bayesian phylogenetic inference and model choice across a large model space. Syst Biol. 2012;61:539–42.
42. Cheng H, Concepcion GT, Feng X, Zhang H, Li H. Haplotype-resolved de novo assembly using phased assembly graphs with hifiasm. Nat Methods. 2021;18:170–5.
43. Roach MJ, Schmidt SA, Borneman AR. Purge haplotigs: allelic contig reassignment for third-gen diploid genome assemblies. BMC Bioinformatics. 2018;19:460.
44. Zhang H, Song L, Wang X, Cheng H, Wang C, Meyer CA, Liu T, Tang M, Aluru S, Yue F, Liu XS, Li H. Fast alignment and preprocessing of chromatin profiles with chromap. Nat Commun. 2021;12:6566.
45. Zhou C, McCarthy SA, Durbin R. YaHS: yet another Hi-C scaffolding tool. Bioinformatics. 2023;39:btac808.
46. Durand NC, Shamim MS, Machol I, Rao SSP, Huntley MH, Lander ES, Aiden EL. Juicer provides a one-click system for analyzing loop-resolution Hi-C experiments. Cell Syst. 2016;3:95–8.
47. Serizay J, Matthey-Doret C, Bignaud A, Baudry L, Koszul R. Orchestrating chromosome conformation capture analysis with bioconductor. Nat Commun. 2024;15:1072.
48. Ludwig A, Pippel M, Myers G, Hiller M. DENTIST – using long reads for closing assembly gaps at high accuracy. Gigascience. 2022;11:giab100.
49. Flynn JM, Hubley R, Goubert C, Rosen J, Clark AG, Feschotte C, Smit AFA. RepeatModeler2 for automated genomic discovery of transposable element families. Proc Natl Acad Sci U S A. 2020;117:9451–7.
50. Turnbull R, Young ND, Tescari E, Skerratt LF, Kosch TA. Terrier: a deep learning repeat classifier. Brief Bioinform. 2025;26:bbaf442.
51. Tarailo-Graovac M, Chen N. Using RepeatMasker to identify repetitive elements in genomic sequences. Curr Protoc Bioinform. 2009;4:4.10.1–4.10.14.
52. Hoff KJ, Lomsadze A, Borodovsky M, Stanke M. Whole-genome annotation with BRAKER. Methods Mol Biol. 2019;1962:65–95.
53. Brůna T, Hoff KJ, Lomsadze A, Stanke M, Borodovsky M. BRAKER2: automatic eukaryotic genome annotation with GeneMark-EP + and AUGUSTUS supported by a protein database. NAR Genom Bioinform. 2021;3:lqaa108.
54. Keilwagen J, Hartung F, Paulini M, Twardziok SO, Grau J. Combining RNA-seq data and homology-based gene prediction for plants, animals and fungi. BMC Bioinformatics. 2018;19:189.
55. Keilwagen J, Hartung F, Grau J. GeMoMa: homology-based gene prediction utilizing intron position conservation and RNA-seq data. Methods Mol Biol. 2019;1962:161–77.
56. UniProt Consortium. UniProt: the universal protein knowledgebase in 2023. Nucleic Acids Res. 2023;51:D523–31.
57. Nevers Y, Warwick Vesztrocy A, Rossier V, Train C-M, Altenhoff A, Dessimoz C, Glover NM. OMark: quality assessment of gene repertoire annotations. Nat Biotechnol. 2025;43:124–33.
58. Cantalapiedra CP, Hernández-Plaza A, Letunic I, Bork P, Huerta-Cepas J. eggNOG-mapper v2: functional annotation, orthology assignments and domain prediction at the metagenomic scale. Mol Biol Evol. 2021;38:5825–9.
59. Emms DM, Kelly S. OrthoFinder: solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. Genome Biol. 2015;16:157.
60. Lovell JT, Sreedasyam A, Schranz ME, Wilson M, Carlson JW, Harkess A, Emms D, Goodstein DM, Schmutz J. GENESPACE tracks regions of interest and gene copy number variation across multiple genomes. eLife. 2022;11:e78526.
61. Tyanova S, Temu T, Cox J. The MaxQuant computational platform for mass spectrometry-based shotgun proteomics. Nat Protoc. 2016;11:2301–19.
62. Kanehisa M, Sato Y, Kawashima M, Furumichi M, Tanabe M. KEGG as a reference resource for gene and protein annotation. Nucleic Acids Res. 2016;44:D457–62.

Publisher's note