Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Conserved enrichment of metabolically unstable proteins in eggs of Echinococcus granulosus and Echinococcus multilocularis: a comparative in silico screen and functional candidate resource

Shuai Han
Fudan University
Mingqi Cai
Fudan University
Cun Yi
Fudan University
Chuizhao Xue
xuecz@nipd.chinacdc.cn

National Institute of Parasitic Diseases, Chinese Center for Disease Control and Prevention; Chinese Center for Tropical Diseases Research
Jipeng Wang
Fudan University
Wei Hu
Fudan University

Research Article

Keywords: Echinococcus granulosus, Echinococcus multilocularis, Echinococcosis, tapeworm eggs, metabolically unstable proteins, functional enrichment

Posted Date: November 25th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11476386/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Abstract

Background

Eggs are the transmission stage of *Echinococcus*, surviving for extended periods in the environment while retaining the capacity for rapid hatching after ingestion by the intermediate host. Predicted metabolically unstable proteins (MUP) may contribute to this transmission-relevant biology, but their functional characteristics across the two most important species, *Echinococcus granulosus* (*E.granulosus)* and *Echinococcus multilocularis* (*E.multilocularis)*, and the robustness of any cross-species differences, remain unresolved.

Results

Using literature-derived keywords, we retrieved and curated a non-redundant set of 1,001 egg-associated protein sequences across seven *Echinococcus* species (including 644 from *E.granulosus* and 319 from *E.multilocularis*) from NCBI. Instability indices (Biopython) classified proteins with values > 40 as MUP; eggNOG-mapper, KOFAM, and InterProScan provided functional annotations, and enrichment was assessed with clusterProfiler against each species' annotated egg proteome. The proportion of MUP did not differ between species (59.78% vs. 57.37%, respectively; $\chi^2 = 0.42$, $P = 0.517$). In *E. granulosus* eggs, MUP were significantly enriched in transcription regulation, signal transduction regulation, and the tropomyosin and LDL receptor-like superfamily domains. *E. multilocularis* MUP showed 29 terms (15 after redundancy removal; microtubule binding, protein kinase binding, antigen processing/presentation), none surviving multiple-testing correction, with closely matched shared fold enrichment (protein kinase binding: 1.365 vs. 1.391). Bootstrap down-sampling of *E. granulosus* to *E. multilocularis*' sample size and annotation coverage abolished significance for nearly all enriched terms (per-term ≤ 4.4% across 1,000 runs), and 95% confidence intervals for shared-term fold enrichment overlapped completely between species. Cross-validation against published developmental datasets supported egg-stage relevance for the majority of flagged candidates: 151 of 182 mapped *E. multilocularis* MUP (82.5%) are transcribed in the oncosphere, and an *E. granulosus* egg-stage proteome confirms protein-level expression.

Conclusions

The functional landscape of egg-associated MUP is largely conserved between *E. granulosus* and *E. multilocularis*, apparent interspecific divergence reflects differential annotation coverage and statistical power, not genuine species-specific specialization. We provide prioritized, computationally flagged labile candidates—transcription/signaling regulators, tropomyosin, LDL receptor-like domains—whose contribution to egg persistence and transmission remains to be established experimentally. Larger *E. multilocularis* sampling is needed to resolve enrichment; the balanced framework is reusable for cross-species egg proteome studies.

1. Background

Echinococcosis is a zoonotic parasitic disease caused by the larvae of *Echinococcus* tapeworms. It poses a serious threat to human health and the livestock industry. Among *Echinococcus* species, *Echinococcus granulosus (E. granulosus)* and *Echinococcus multilocularis (E. multilocularis)* are the two primary pathogens, causing cystic echinococcosis and alveolar echinococcosis, respectively [1]. The two species differ significantly in morphology, life cycle, host range, and pathogenic mechanisms [1].

Following expulsion from the definitive host, *Echinococcus* eggs must survive in the external environment for extended periods—ranging from days to years—until ingested by an intermediate host [1, 2]. Studies on taeniid eggs have demonstrated that survival is highly temperature-dependent, with some embryos surviving more than 200 days at 7°C but less than 10 days at 38°C [3]. Similarly, Bucur et al. demonstrated that *Taenia saginata* eggs can remain viable and activatable for up to six months at 5°C in water and silt, whereas storage at 20°C significantly reduced their survival [4]. Moreover, Sánchez Thevenet et al. showed that *Taenia hydatigena* eggs could survive up to 365 days under various relative humidity conditions, with egg clumping and triacylglycerol mobilization serving as key desiccation-resistance mechanisms [5]. Classic studies by Laws further established desiccation as the dominant natural factor limiting *taeniid* egg survival, with notable species-specific differences in susceptibility [6]. During this diapause-like stage, eggs must maintain structural integrity and proteostasis while preserving limited energy reserves, yet they must remain poised for rapid hatching and oncosphere development within minutes upon host ingestion. This remarkable combination of long-term environmental persistence and rapid developmental reactivation is a classic example of an adaptive trade-off: the egg must minimize metabolic expenditure during external sojourn to avoid depletion of finite resources, while retaining the capacity for instantaneous activation when host cues are perceived. Such dormancy strategies, including diapause and quiescence, are critical physiological mechanisms that enable parasites to cope with environmental extremes while maintaining the capacity for rapid response to host signals [7]. Such a strategy is central to the transmission success of *Echinococcus* and other environmentally transmitted parasites.

Protein stability in the cellular context is determined not only by thermodynamic resistance to denaturation but also by intracellular half-life—the duration for which a protein persists before being degraded. Proteins with short half-lives, i.e., those predicted to be metabolically unstable, are often enriched in regulatory networks, as their rapid turnover enables precise temporal control of signaling cascades, swift clearance of inhibitory factors upon activation signals, and removal of damaged polypeptides under stress [8, 9]. In parasitic organisms, such dynamic regulation is particularly critical for balancing prolonged environmental survival with rapid developmental transitions upon host entry [10, 11]. In *Echinococcus*, Hemer et al. demonstrated that host insulin at physiological concentrations significantly stimulates larval stem cell proliferation and metacestode formation through evolutionarily conserved insulin signalling pathways, highlighting the importance of hormonal host-parasite cross-communication in regulating parasite development and physiology [12]. Recent integrative studies in *Leishmania* have demonstrated that stage differentiation relies heavily on coordinated post-

transcriptional regulation, including differential mRNA turnover and protein degradation, rather than genomic variation [13]. Although direct experimental evidence in *Echinococcus* eggs remains scarce, this principle provides a compelling framework for hypothesis generation. From an adaptive perspective, the enrichment of metabolically unstable proteins (MUP) in eggs may represent an evolutionary solution to the conflicting demands of long-term quiescence and instantaneous reactivation: these proteins can be quickly degraded when not needed (saving energy and preventing aberrant signaling) and rapidly resynthesized or activated when the host environment triggers hatching, thus providing both economy and responsiveness.

Despite the central role of eggs in *Echinococcus* transmission, our understanding of protein turnover dynamics in this life stage remains extremely limited, and functional differences in egg-associated MUP between species have rarely been reported. Advances in bioinformatics methods have facilitated large-scale analyses of protein stability [14, 15]. The instability index, originally defined by Guruprasad et al. based on dipeptide composition to predict *in vivo* intracellular half-life, has been widely used to classify proteins as stable or unstable [16, 17]. Functional annotations [18–20], combined with enrichment analysis using clusterProfiler [21], can systematically elucidate the functional preferences of specific protein sets [22].

The eggs of *E. granulosus* and *E. multilocularis* exhibit similar survival capacities under laboratory conditions [2], yet the two species differ significantly in their geographic distribution and life-history traits. *E. granulosus* is distributed globally across diverse climatic zones and utilizes a broad spectrum of domestic and wild ungulates as intermediate hosts, thereby exposing its eggs to highly variable external environments. In contrast, *E. multilocularis* is restricted mainly to temperate and boreal regions of the Northern Hemisphere and depends largely on small rodent hosts [23]. The availability of genomic resources for both species has enabled comparative analyses of their molecular repertoires [24], revealing lineage-specific differences in gene families potentially involved in host-parasite interactions [25]. These distinct ecological constraints could, in principle, shape species-specific egg traits; however, whether they translate into divergent functional enrichment of egg-associated MUP remains unknown. Direct cross-species enrichment comparisons are further complicated by differences in genome annotation coverage and statistical power between species, which can create apparent but artefactual divergence. Despite the conserved biological characteristics of tapeworm eggs that may maintain similar overall proportions of egg-associated MUP between the two species, it remains unclear whether their functional enrichment patterns have diverged correspondingly to accommodate species-specific adaptive strategies for environmental persistence and host-triggered developmental activation.

On this basis, we asked whether egg-associated MUP repertoires differ between the two species and, crucially, whether any observed differences are biologically real or artefactual. To address this, we performed a comparative in silico screen of egg-associated proteins from both species, integrating stability prediction with functional enrichment against a species-matched egg- proteins backgrounds. We applied bootstrap down-sampling to equalize sample size and annotation coverage between species. Our aims were to determine whether the functional landscape of egg-associated MUP is conserved or

species-specific, to provide a prioritized resource of computationally flagged labile egg candidates for experimental follow-up, and to deliver a balanced framework reusable for cross-species egg proteome comparisons.

2. Methods

2.1 Literature search and keyword extraction

The following search strategy was used to search PubMed:

(((egg[Title]) AND (protein[Title])) OR ((egg[Title/Abstract]) AND (protein[Title/Abstract]))) AND (parasite). Because direct studies on *Echinococcus* egg proteins are scarce, we searched PubMed for studies on egg proteins from parasites in general, and extracted protein names or functions as keywords for subsequent screening of *Echinococcus* protein sequences.

Keywords referring to the names or functions of parasite egg proteins were extracted from the titles and abstracts of the retrieved articles. The extracted keywords were “Major Egg Antigen,” “MEA,” “Em18,” “AntigenB,” “AgB,” “Em95,” “Eg95,” “P29,” “P30,” “immunogen,” “allergen,” “diagnostic,” “serological,” “antibody,” “epitope,” “reactive,” “vaccine candidate,” “vaccine,” “lipoprotein,” “heat shock,” “HSP,” “paramyosin,” “tropomyosin,” “actin,” “enolase,” “gapdh,” “biomarker,” “Egg shell protein,” “Ovo,” “Chitinase,” “Chitin synthase,” “KDPG,” “Tyrosine rich,” and “Tyrosine-rich”.

2.2 Protein sequence retrieval and preliminary screening

All protein sequences from the genus *Echinococcus* were downloaded from the NCBI Protein Database. Sequence titles and annotations were matched against the keywords listed above, and sequences containing at least one keyword were retained. Keywords were derived from PubMed studies on parasite egg proteins (Section 2.1) and then used to screen *Echinococcus* protein sequences downloaded from NCBI.

2.3 Quality filtering

Sequences obtained from preliminary screening were first grouped by locus or gene name. For sequences lacking both identifiers, we grouped them by their protein description (the definition line in the FASTA header) when available; sequences with no shared description were treated as independent entries and retained without redundancy removal. Within each defined group, the longest sequence was kept as the representative to remove redundancy. Subsequently, sequences containing non-standard amino acid residues (e.g., “B” or “X”) and sequences labeled as “syntheticconstruct” were excluded, yielding a high-quality, non-redundant protein sequence set for subsequent analyses.

2.4 Calculation of instability index

The instability index, which predicts intracellular half-life based on dipeptide composition, was calculated for each protein using the Bio.SeqUtils.ProtParam module in Biopython. This index is based on the

composition of dipeptides in the amino acid sequence. A threshold of 40 was used to distinguish between stable and unstable proteins [17]. In this study, “metabolically unstable proteins” refers to proteins with an instability index > 40, predicted from dipeptide composition. The same computational approach has been applied in recent studies of *Echinococcus* recombinant antigens and protein families, confirming its utility for parasite protein stability prediction [25, 26]. Proteins with index values > 40 were classified as metabolically unstable, whereas those with values ≤ 40 were classified as stable. The proportions of egg-associated MUP were compared between the two species using a chi-square test, with $\alpha = 0.05$.

2.5 Functional annotation

GO annotation of the protein sequences was performed using eggNOG-mapper v2.1.13 with default parameters, KEGG annotation using KofamScan v1.3.0–2 with the default E-value threshold, and IPR domain annotation using InterProScan v5.77-108.0.

2.6 Analysis of GO functional preferences in egg protein subgroups

Egg-associated proteins identified through literature-based keyword screening were analyzed to elucidate the functional preferences of MUP within the egg protein dataset, rather than to examine enrichment relative to all proteins encoded by each species’ genome. Using all proteins encoded by each species' genome as the background would introduce many genes specifically expressed during adult and larval stages that are unrelated to egg biology, thereby diluting egg-specific functional signals and confounding cross-species comparisons. Therefore, the following approach, hereafter termed “subgroup proportion analysis,” was adopted to analyze distributional preferences among egg protein subgroups.

2.6.1 Definition of background and target sets

The background set comprised all annotated egg-associated proteins from the same species, including MUP (set A) and stable proteins (set B); thus, the background set = $A \cup B$. Hereafter, “all annotated proteins of the same species” refers to this egg-associated background set. Egg-associated proteins without functional annotations were excluded from the statistical analyses, summarized separately, and discussed in the study limitations as a potential source of bias in the enrichment results.

In this study, “enrichment” refers specifically to a functional category containing a significantly higher proportion of MUP than the background set of annotated egg-associated proteins, as determined using the hypergeometric test with Benjamini-Hochberg (BH) correction for multiple testing. This analysis was limited to comparisons within the egg-associated protein subset and does not represent distribution relative to all proteins encoded by the species’ genome or directly imply that these proteins perform the corresponding functions in eggs.

2.6.2 Statistical enrichment analysis and significance thresholds

GO enrichment was analyzed using the R package clusterProfiler (v4.4.2) and the hypergeometric test, with BH correction for multiple testing. The analysis was limited to species with more than 15 annotated proteins in the target set to avoid statistical instability caused by small sample sizes. Terms with a raw *P* value < 0.05 were considered to show a trend toward enrichment, whereas those with a BH-adjusted *P* value < 0.05 were considered significantly enriched.

For each eligible species, a GO term-protein ID mapping table (TERM2GENE) was constructed as input for the enricher() function. GO term descriptions were obtained from GO.db v3.20.0, and obsolete terms were removed.

2.6.3 Annotation filtering and hierarchy pruning

Because GO annotations are transferred across species based on sequence homology, some terms may refer to anatomical structures or physiological processes unique to higher animals, such as visual organ development or neurobehavioral processes, for which *Echinococcus* species may lack direct functional counterparts. Therefore, significantly enriched GO terms were manually reviewed. The review process was as follows: two researchers independently reviewed the descriptions of all significantly enriched GO terms and evaluated them according to predefined criteria. Specifically, a term was considered to have doubtful biological relevance if it referred explicitly to an anatomical structure or physiological process, such as the eye, heart, or behavior, for which tapeworms lack a corresponding tissue or function. The classification was accepted when the two researchers agreed. Disagreements were resolved through discussion. Terms classified as having doubtful biological relevance were retained in full in the supplementary materials but were excluded from the main table and core conclusions.

To address hierarchical redundancy among parent and child terms in the GO directed acyclic graph, GO.db was used to identify the ancestors of each GO term. When multiple significantly enriched terms occurred within the same hierarchical branch, we retained the most specific term at the bottom of the hierarchy and excluded its ancestor terms.

2.6.4 Parallel control analysis

A parallel enrichment analysis was conducted using stable proteins for each species as the target set and all annotated egg-associated proteins from the same species as the background set, following the methods described in Sections 2.6.1–2.6.3. This analysis served as an internal control to distinguish the background functional patterns of egg-associated proteins from enrichment signals specific to MUP.

2.7 KEGG pathway subgroup proportion analysis

With egg-associated MUP as the target set and all annotated proteins of the same species as the background set, the enricher() function in clusterProfiler was used to perform KEGG pathway enrichment analysis for each species with more than 15 annotated proteins in the target set. The criteria for defining enrichment trends and significant enrichment were the same as those described in Section 2.6.2. Unlike GO terms, KEGG pathways do not have a directed acyclic graph hierarchy. Therefore, no hierarchy-based pruning was performed. However, pathway terms in the annotation results that were clearly associated

with anatomical structures specific to higher animals, such as cardiac muscle or the nervous system, and for which corresponding functions in tapeworms were not supported by the literature were manually reviewed and filtered according to the principles described in Section 2.6.3.

2.8 IPR domain subgroup proportion analysis

With egg-associated MUP as the target set and all annotated proteins of the same species as the background set, the enricher() function in clusterProfiler was used to perform IPR domain enrichment analysis for each species with more than 15 annotated proteins in the target set. The criteria for defining enrichment trends and significant enrichment were the same as those described in Section 2.6.2. IPR domains have hierarchical relationships, such as superfamily–family relationships. Because this analysis aimed to identify all enriched domain signals in MUP regardless of their hierarchical position, no hierarchy-based pruning was performed.

2.9 Cross-species comparison strategy

Given the differences between the two species in annotation completeness, the number of egg-associated MUP, and the number of terms tested, direct comparison of raw or BH-adjusted *P* values would not be appropriate. Therefore, fold enrichment was used as the primary metric for cross-species comparisons to assess the enrichment intensity of GO terms, KEGG pathways, and IPR domains showing enrichment trends or significant enrichment. Fold enrichment was calculated as follows: (number of proteins annotated to a given term in the target set / total number of proteins in the target set) / (number of proteins annotated to that term in the background set / total number of proteins in the background set). Higher values indicated greater enrichment intensity.

2.10 Sensitivity analysis

To determine whether observed interspecific differences in GO enrichment were robust or instead reflected disparities in sample size and annotation coverage, we performed bootstrap-based sensitivity analyses in R (1,000 iterations).

Single-species bootstrap. For each GO term enriched in *E. multilocularis*, its MUP set was resampled with replacement and enrichment recomputed against the species-matched background, recording the proportion of runs in which the term remained significant.

Cross-species equalization. *E. granulosus* MUP were down-sampled to the sample size and annotation coverage of *E. multilocularis*, and enrichment was recomputed for the GO terms shared between the two species, recording per-term bootstrap significance rates and the distribution of fold enrichment.

Sample-size gradient. *E. granulosus* MUP were down-sampled to a series of progressively smaller sample sizes, and the number of significantly enriched terms was recorded at each size.

Shared-term confidence intervals. For GO terms enriched in both species, bootstrap 95% confidence intervals of fold enrichment were computed and assessed for overlap.

Threshold sensitivity. Proteins were reclassified under alternative instability-index cut-offs, and enrichment was recomputed for each species.

2.11 Cross-validation of egg-stage expression

To evaluate whether computationally flagged egg-associated MUP candidates are expressed at the egg/oncosphere stage, candidates were cross-validated against published developmental datasets.

E. multilocularis (quantitative). The *E. multilocularis* MUP sequences were searched against the *E. multilocularis* proteome (PRJEB122, WormBase ParaSite release WBPS19) with diamond v2.1.9 (blastp mode), requiring E-value < 1e-5 and retaining the single best (highest-identity) hit per query as the orthologous mapping. Mapped EmuJ identifiers were joined to the oncosphere transcriptome of a published developmental study [27]; an ortholog was scored as expressed when RPKM > 0 in any oncosphere sample. Queries mapping to an EmuJ identifier absent from the transcriptome lookup table were classified separately and not counted as negatives.

E. granulosus (qualitative). The published *E. granulosus* egg-stage iTRAQ proteome [28] does not publicly release a complete differential-protein table; *E. granulosus* MUP were therefore cross-checked against this proteome qualitatively, without a per-candidate coverage analysis.

3. Results

A total of 1,431 articles were retrieved from PubMed. All 80,026 protein sequences from the genus *Echinococcus* were downloaded from the NCBI Protein Database. After literature-based keyword screening, redundancy removal, and quality filtering, 1,001 non-redundant egg-associated protein sequences were obtained. The sequences were then classified according to thermal stability and subjected to functional annotation.

3.1 Stability classification and species distribution

Among the 1,001 sequences analyzed, 573 (57.24%) were classified as metabolically unstable and 428 (42.76%) as stable. Table 1 shows the distribution by species: *E. granulosus* had 385 egg-associated MUP (59.78%; 385/644), whereas *E. multilocularis* had 183 (57.37%; 183/319). The chi-square test showed no significant difference in the proportion of MUP between the two species ($\chi^2 = 0.420$, $P = 0.517$). Each remaining species—*E. canadensis*, *E. oligarthrus*, *E. ortleppi*, *E. vogeli*, and *E. equinus*— had fewer than 15 target proteins and therefore did not meet the analysis threshold specified in Section 2.6.2. These species were therefore excluded from subsequent functional annotation and subgroup proportion analyses.

Table 1
Numbers and proportions of stable and metabolically unstable proteins by species

Species	Stable		Metabolically unstable		Total
	No. of sequences	Proportion (%)	No. of sequences	Proportion (%)	
E. granulosus	259	40.22	385	59.78	644
E. multilocularis	136	42.63	183	57.37	319
*E. canadensis**	11	78.57	3	21.43	14
*E. oligarthrus**	8	100.00	0	0.00	8
*E. ortleppi**	6	85.71	1	14.29	7
*E. vogeli**	6	100.00	0	0.00	6
*E. equinus**	2	66.67	1	33.33	3
Total	428	42.76	573	57.24	1,001

* Not included in subsequent analyses because the number of sequences was below the analysis threshold (15 proteins).

3.2 Functional annotation and enrichment characteristics

Among the 1,001 egg-associated protein sequences, functional annotation coverage differed significantly between *E. granulosus* and *E. multilocularis*. GO annotation coverage was 79.81% (514/644) for *E. granulosus* and 70.22% (224/319) for *E. multilocularis* (χ^2 = 10.438, df = 1, P = 0.001); KEGG annotation coverage was 83.39% (537/644) and 70.85% (226/319), respectively (χ^2 = 19.627, df = 1, P < 0.001); and IPR domain annotation coverage was 94.88% (611/644) and 90.60% (289/319), respectively (χ^2 = 5.711, df = 1, P = 0.02).

3.2.1 GO enrichment characteristics of *E. granulosus*

All annotated proteins were associated with 1,627 GO terms. Enrichment analysis showed that 161 terms had enrichment trends among egg-associated MUP (P < 0.05). After BH correction, 20 terms remained significantly enriched (BH-adjusted P < 0.05). To address hierarchical redundancy in the GO directed acyclic graph, the most specific terms were retained and their ancestor terms were removed as described in Section 2.6.3. For example, GO:0019900 (kinase binding) was removed because it is the parent of GO:0019901 (protein kinase binding). This reduced the 20 significant terms to 10 non-redundant terms. After manual filtering for biological relevance, GO:0007610 (behavior) and GO:0043010 (camera-type eye development) were removed because corresponding anatomical structures or

physiological functions have not been demonstrated in *E. granulosus*. Eight terms were retained in Table 2. All 20 significant terms, including those excluded by either pruning or manual filtering, are provided in S1 Table.

Table 2

GO terms significantly enriched among egg-associated metabolically unstable proteins in *E. granulosus*

GO ID	Description	Target protein ratio	Background protein ratio	Fold enrichment	BH-adjusted *P* value
GO:0019901	Protein kinase binding	41/359	43/514	1.365	0.004
GO:0043393	Regulation of protein binding	31/359	32/514	1.387	0.014
GO:0045944	Positive regulation of transcription by RNA polymerase II	33/359	35/514	1.350	0.034
GO:0003713	Transcription coactivator activity	22/359	22/514	1.432	0.034
GO:0042326	Negative regulation of phosphorylation	27/359	28/514	1.381	0.038
GO:0010256	Endomembrane system organization	26/359	27/514	1.379	0.044
GO:0034654	Nucleobase-containing compound biosynthetic process	31/359	33/514	1.345	0.044
GO:0051052	Regulation of DNA metabolic process	20/359	20/514	1.432	0.044

Note: The target protein ratio represents the number of metabolically unstable proteins annotated to a given GO term divided by the total number of GO-annotated metabolically unstable proteins (359). The background protein ratio represents the number of egg-associated proteins annotated to that term divided by the total number of GO-annotated egg-associated proteins (514). Hierarchical redundancy was removed by excluding parent GO terms; for example, GO:0019900 (kinase binding) was removed because it is the parent of GO:0019901 (protein kinase binding). Following manual filtering for biological relevance, GO:0007610 (behavior) and GO:0043010 (camera-type eye development) were excluded from the table because corresponding anatomical structures or physiological functions have not been demonstrated in *E. granulosus*. The complete list of terms, including those excluded, is provided in S1 Table. BH, Benjamini–Hochberg; EG, *Echinococcus granulosus*.

Significant enrichment was observed mainly in signal transduction and binding functions—GO:0019901 (protein kinase binding; 41 proteins) and GO:0043393 (regulation of protein binding; 31 proteins)—transcriptional regulation—GO:0045944 (positive regulation of transcription by RNA polymerase II; 33 proteins) and GO:0003713 (transcription coactivator activity; 22 proteins)—and negative regulation of phosphorylation (GO:0042326; 27 proteins). Endomembrane system organization (GO:0010256; 26

proteins; BH-adjusted *P* = 0.044), nucleobase-containing compound biosynthetic process (GO:0034654; 31 proteins; BH-adjusted *P* = 0.044), and regulation of DNA metabolic process (GO:0051052; 20 proteins; BH-adjusted *P* = 0.044) were also significantly enriched in the target protein set.

The significantly enriched terms represented three core functional modules: (1) transcriptional regulation—GO:0045944 (positive regulation of transcription by RNA polymerase II) and GO:0003713 (transcription coactivator activity); (2) signal transduction regulation—GO:0019901 (protein kinase binding), GO:0043393 (regulation of protein binding), and GO:0042326 (negative regulation of phosphorylation); and (3) basal metabolism and maintenance of cellular structure—GO:0010256 (endomembrane system organization), GO:0034654 (nucleobase-containing compound biosynthetic process), and GO:0051052 (regulation of DNA metabolic process). These patterns suggest that MUP in *E. granulosus* may be involved primarily in the rapid adjustment of gene expression programs and the precise temporal regulation of signaling pathways.

3.2.2 GO enrichment characteristics of *E. multilocularis*

Annotated *E. multilocularis* proteins were associated with 1,121 GO terms. Enrichment analysis showed that 29 GO terms had enrichment trends among egg-associated MUP (raw $P < 0.05$; S2 Table), but none remained significantly enriched after BH correction (BH-adjusted $P > 0.05$). To address hierarchical redundancy, parent terms were removed as described in Section 2.6.3, resulting in 15 non-redundant terms with enrichment trends. These terms were primarily related to transcriptional regulation, including GO:0045893 (positive regulation of DNA-templated transcription); cytoskeletal and vesicular transport, including GO:0008017 (microtubule binding) and GO:0006888 (endoplasmic reticulum to Golgi vesicle-mediated transport); signal transduction, including GO:0019901 (protein kinase binding) and GO:0031267 (small GTPase binding); immune-related processes, including GO:0019886 (antigen processing and presentation of exogenous peptide antigen via MHC class II); and organelle localization and development-related processes, including GO:0051640 (organelle localization) and GO:0009887 (animal organ morphogenesis) (Table 3).

Enrichment trends among *E. multilocularis* MUP were dispersed, and no functional category reached statistical significance after multiple-testing correction.

Table 3

GO terms showing enrichment trends among egg-associated metabolically unstable proteins in *E. multilocularis*

GO ID	Description	Target protein ratio	Background protein ratio	Fold enrichment	Raw *P* value
GO:0008017	microtubule binding	11/151	11/224	1.483	0.012
GO:0098687	chromosomal region	11/151	11/224	1.483	0.012
GO:0019901	protein kinase binding	15/151	16/224	1.391	0.013
GO:0045893	positive regulation of DNA-templated transcription	18/151	20/224	1.335	0.017
GO:0005869	dynactin complex	10/151	10/224	1.483	0.018
GO:0019886	antigen processing and presentation of exogenous peptide antigen via MHC class II	10/151	10/224	1.483	0.018
GO:0031267	small GTPase binding	10/151	10/224	1.483	0.018
GO:0006888	endoplasmic reticulum to Golgi vesicle-mediated transport	14/151	15/224	1.385	0.019
GO:0034654	nucleobase-containing compound biosynthetic process	13/151	14/224	1.377	0.027
GO:0007010	cytoskeleton organization	36/151	45/224	1.187	0.030
GO:0008150	biological process	120/151	169/224	1.053	0.034
GO:0003674	molecular function	116/151	163/224	1.056	0.037
GO:0043229	intracellular organelle	114/151	160/224	1.057	0.039
GO:0051640	organelle localization	12/151	13/224	1.369	0.039

Note: The target protein ratio represents the number of metabolically unstable proteins annotated to a given GO term divided by the total number of GO-annotated metabolically unstable proteins (151). The background protein ratio represents the number of egg-associated proteins annotated to that term divided by the total number of GO-annotated egg-associated proteins (224). The listed terms showed enrichment trends based on raw *P* values < 0.05, but none remained significantly enriched after BH correction. Because none of the terms remained significant after BH correction, the biological-relevance filtering described in Section 2.6.3 was not applied to these trend-level terms. Hierarchical redundancy was removed by excluding parent GO terms as described in Section 2.6.3. The complete results are provided in S2 Table. BH, Benjamini–Hochberg; EM, *Echinococcus multilocularis*.

GO ID	Description	Target protein ratio	Background protein ratio	Fold enrichment	Raw *P* value
GO:0009887	animal organ morphogenesis	21/151	25/224	1.246	0.044

Note: The target protein ratio represents the number of metabolically unstable proteins annotated to a given GO term divided by the total number of GO-annotated metabolically unstable proteins (151). The background protein ratio represents the number of egg-associated proteins annotated to that term divided by the total number of GO-annotated egg-associated proteins (224). The listed terms showed enrichment trends based on raw *P* values < 0.05, but none remained significantly enriched after BH correction. Because none of the terms remained significant after BH correction, the biological-relevance filtering described in Section 2.6.3 was not applied to these trend-level terms. Hierarchical redundancy was removed by excluding parent GO terms as described in Section 2.6.3. The complete results are provided in S2 Table. BH, Benjamini–Hochberg; EM, *Echinococcus multilocularis*.

3.2.3 KEGG pathway subgroup proportion analysis

Annotated *E. granulosus* proteins were assigned to 52 KEGG pathways. Enrichment analysis showed that three pathways had enrichment trends among egg-associated MUP ($P < 0.05$; S3 Table). After manual filtering for biological relevance according to the principles described in Sections 2.6.3 and 2.7, map04261 (adrenergic signaling in cardiomyocytes) was excluded due to its lack of anatomical and physiological correspondence in tapeworms, leaving one pathway with an enrichment trend: map05206 (microRNAs in cancer). After BH correction, none of the pathways remained significantly enriched (all BH-adjusted *P* values > 0.05).

Annotated *E. multilocularis* proteins were assigned to 39 KEGG pathways. No pathways showed enrichment trends (raw $P < 0.05$) or significant enrichment (BH-adjusted $P < 0.05$). All annotated pathways are provided in S4 Table.

3.2.4 IPR domain subgroup proportion analysis

Annotated *E. granulosus* proteins were assigned to 25 IPR domains. Enrichment analysis showed that three domains had enrichment trends among MUP, two of which were significantly enriched (BH-adjusted $P < 0.05$; Table 4): tropomyosin (IPR000533; 12 proteins; BH-adjusted $P = 0.041$) and the LDL receptor-like superfamily (IPR036055; 11 proteins; BH-adjusted $P = 0.041$). As described in Section 2.8, IPR domains were not subjected to hierarchy-based pruning. Therefore, all three domains were retained. All domains showing enrichment trends are provided in S5 Table.

Annotated *E. multilocularis* proteins were assigned to 10 IPR domains (S5 Table). However, no domains showed enrichment trends (raw $P < 0.05$) or significant enrichment (BH-adjusted $P < 0.05$).

Table 4
IPR domains significantly enriched among egg-associated metabolically unstable proteins in *E. granulosus*

IPR ID	Description	Target protein ratio	Background protein ratio	Fold enrichment	BH-adjusted *P* value
IPR000533	Tropomyosin	12/365	12/611	1.674	0.041
IPR036055	LDL receptor-like superfamily	11/365	11/611	1.674	0.041

Note: The target protein ratio represents the number of metabolically unstable proteins assigned to a given IPR domain divided by the total number of IPR-annotated metabolically unstable proteins (365). The background protein ratio represents the number of egg-associated proteins assigned to that domain divided by the total number of IPR-annotated egg-associated proteins (611). IPR domains were not subjected to hierarchy-based pruning, as described in Section 2.8. All domains showing enrichment trends are provided in S5 Table. BH, Benjamini–Hochberg; IPR, InterPro.

3.2.5 Parallel control analysis of stable proteins

To determine whether the functional enrichment observed among egg-associated MUP was a general background characteristic of egg-associated proteins, stable proteins were analyzed using the same subgroup proportion analysis, as described in Section 2.6.4. After the applicable post-enrichment processing, stable proteins in *E. granulosus* showed significant enrichment in 205 GO terms, 40 KEGG pathways, and 8 IPR domains, whereas those in *E. multilocularis* showed significant enrichment in 198 GO terms, 38 KEGG pathways, and 6 IPR domains (BH-adjusted $P < 0.05$). The complete results are provided in S6-S8 Tables.

However, the enriched functional categories differed markedly between stable and MUP. Of note, unlike the metabolically unstable protein results, the stable protein enrichments presented below did not undergo biological-relevance filtering. Accordingly, some terms included in the stable-protein results lack clear anatomical or physiological correspondence in tapeworms (e.g., those related to vertebrate-specific structures or immune functions) and should not be interpreted as evidence of genuine biological function; they are presented here for complete transparency of the unfiltered enrichment output. The enriched GO terms for stable proteins in *E. granulosus* primarily included ATP metabolic process, ubiquitin protein ligase binding, myelin sheath, neutrophil degranulation, and protein folding chaperone, whereas those for *E. multilocularis* included anion binding, positive regulation of immune system process, ion binding, and de novo protein folding. For KEGG pathways, stable proteins from both species were consistently enriched in infection- and immunity-related pathways, including map04612 (antigen processing and presentation). For IPR domains, stable proteins in *E. granulosus* were significantly enriched in Taeniidae antigen and several heat shock protein 70 family-related domains, whereas those in *E. multilocularis* were enriched in ATPase, nucleotide binding domain, and Taeniidae antigen.

The functional categories enriched among stable proteins showed almost no overlap with those enriched among MUP, including transcriptional regulation, kinase binding, negative regulation of phosphorylation, tropomyosin, and the LDL receptor-like superfamily (Sections 3.2.1 and 3.2.4). Therefore, the parallel control analysis demonstrated that the functional enrichment observed among MUP was not a general background characteristic of egg-associated proteins but was specific to the metabolically unstable protein subgroup.

3.3 Cross-species comparison and robustness of GO enrichment

Among the 175 GO terms (S9 Table) with raw $P < 0.05$ in *E. granulosus* or *E. multilocularis*, 15 were shared between two species (Table 5). Of these, eight were significant in *E. granulosus* and trend-level in *E. multilocularis*, whereas the remaining seven showed trend-level in both species. The shared terms were concentrated in transcriptional regulation (GO:0045893, GO:0009891, GO:0010557, GO:0031328, GO:1902680), signal transduction/binding (GO:0019901, GO:0019900), nucleobase-containing compound biosynthesis (GO:0034654), cytoskeleton organization (GO:0007010), microtubule binding (GO:0008017), and chromosomal region (GO:0098687). Four broad or root-level GO terms (GO:0003674, GO:0008150, GO:0043226, GO:0043229) were also shared but were not interpreted as specific biological signals. Mean fold enrichment for shared terms was comparable between species (1.27 vs. 1.29), indicating conserved enrichment intensity for these functions.

Table 5
Fold enrichment of GO terms shared between *E. granulosus* and *E. multilocularis*.

ID	Description	FoldEnrichment		Status	
		E. granulosus	***E. multilocularis***	***E. granulosus***	***E. multilocularis***
GO:0009891	positive regulation of biosynthetic process	1.392	1.381	Sig	Trend
GO:0010557	positive regulation of macromolecule biosynthetic process	1.392	1.377	Sig	Trend
GO:0031328	positive regulation of cellular biosynthetic process	1.392	1.381	Sig	Trend
GO:1902680	positive regulation of RNA biosynthetic process	1.386	1.354	Sig	Trend
GO:0045893	positive regulation of DNA-templated transcription	1.381	1.335	Sig	Trend
GO:0019900	kinase binding	1.365	1.391	Sig	Trend
GO:0019901	protein kinase binding	1.365	1.391	Sig	Trend
GO:0034654	nucleobase-containing compound biosynthetic process	1.345	1.377	Sig	Trend
GO:0003674	molecular_function	1.047	1.056	Trend	Trend
GO:0007010	cytoskeleton organization	1.143	1.187	Trend	Trend
GO:0008150	biological_process	1.046	1.053	Trend	Trend
GO:0008017	microtubule binding	1.432	1.483	Trend	Trend
GO:0098687	chromosomal region	1.312	1.483	Trend	Trend
GO:0043226	organelle	1.039	1.057	Trend	Trend

Note: Shared GO terms are those with raw $P < 0.05$ in both species. Status indicates enrichment before GO hierarchy pruning and biological-relevance filtering. Significant: BH-adjusted $P < 0.05$. Trend: raw $P < 0.05$ but BH-adjusted $P \geq 0.05$. No term was significant in *E. multilocularis*. After hierarchy pruning and biological-relevance filtering, eight terms remained significantly enriched in *E. granulosus* (Table 2); among the shared terms in this table, only GO:0019901 and GO:0034654 remained significant after pruning.

ID	Description	FoldEnrichment		Status	
		E. granulosus	***E. multilocularis***	***E. granulosus***	***E. multilocularis***
GO:0043229	intracellular organelle	1.039	1.057	Trend	Trend

Note: Shared GO terms are those with raw $P < 0.05$ in both species. Status indicates enrichment before GO hierarchy pruning and biological-relevance filtering. Significant: BH-adjusted $P < 0.05$. Trend: raw $P < 0.05$ but BH-adjusted $P \geq 0.05$. No term was significant in *E. multilocularis*. After hierarchy pruning and biological-relevance filtering, eight terms remained significantly enriched in *E. granulosus* (Table 2); among the shared terms in this table, only GO:0019901 and GO:0034654 remained significant after pruning.

Among the GO terms significantly enriched in *E. granulosus* after BH correction (Table 2), two, GO:0019901 and GO:0034654, also showed enrichment trends in *E. multilocularis*; the remaining six did not reach raw $P < 0.05$ in *E. multilocularis*.

Because more terms reached significance in *E. granulosus* than in *E. multilocularis* (146 vs. 14 at raw $P < 0.05$), we tested whether this apparent asymmetry reflected genuine species-specific specialization or differential annotation coverage and statistical power (Table 6 for Summary, S10-S14 Tables for details). When *E. granulosus* MUP were down-sampled to the *E. multilocularis* sample size and annotation coverage, per-term bootstrap significance rates across the eight shared key GO terms were ≤ 4.4% (fold-enrichment medians 1.34–1.43). Along a gradient of decreasing sample size (644, 500, 400, 319, 250, 183), the number of significantly enriched terms fell from 8 to 7, 6, 4, 2, and 0, respectively. Bootstrap 95% confidence intervals of fold enrichment for all eight shared terms overlapped completely between the two species, and bootstrap resampling of the six *E. multilocularis* enrichment-trend terms yielded a 0% significance rate. Reclassification under alternative instability-index thresholds (35, 40, 45) produced 1, 8, and 6 significant terms, respectively, in *E. granulosus* and none in *E. multilocularis* at any threshold. Together, these analyses indicate that the apparent interspecific differences in GO enrichment reflect differential annotation coverage and statistical power rather than genuine species-specific specialization, and that the functional landscape of egg-associated MUP is largely conserved between the two species.

Table 6
Summary of sensitivity analyses for the cross-species enrichment comparison

Analysis	Setting	Result
Cross-species equalization	E. granulosus MUP down-sampled to E. multilocularis sample size & annotation coverage; 8 shared key GO terms	Per-term bootstrap significance ≤ 4.4% (fold-enrichment median 1.34–1.43)
Sample-size gradient	E. granulosus down-sampled to 644 / 500 / 400 / 319 / 250 / 183	No. significant terms: 8 / 7 / 6 / 4 / 2 / 0
Single-species bootstrap	6 E. multilocularis enrichment-trend terms, 1,000 runs	0% significance rate
Shared-term fold CI overlap	8 shared GO terms, bootstrap 95% CI	Overlap = true for all
Threshold sensitivity	Instability-index cut-offs 35 / 40 / 45	E. granulosus: 1 / 8 / 6; E. multilocularis: 0 / 0 / 0

3.4 Cross-validation of egg-stage expression of MUP candidates

To test whether the computationally flagged candidates are actually expressed at the egg/oncosphere stage, we cross-validated them against published developmental datasets (Fig. 1). Of the 183 *E. multilocularis* MUP candidates, 182 mapped to unambiguous EmuJ orthologs in the *E. multilocularis* proteome (median 100% sequence identity; all E-value < 1e-5); 151 (82.5%) of these orthologs were expressed in the oncosphere transcriptome (RPKM > 0 in any of three oncosphere samples), 17 were undetectable, and 14 mapped to EmuJ identifiers absent from the transcriptome lookup table and were therefore not scored. These results support egg/oncosphere-stage relevance for the majority of the *E. multilocularis* MUP candidates. For *E. granulosus*, the published egg-stage proteome [28] corroborates protein-level expression at the oncosphere stage, although the absence of a complete public protein table precluded a quantitative per-candidate analysis.

4. Discussion

To the best of our knowledge, this is the first study to systematically compare the functional enrichment patterns of MUP in the eggs of *E. granulosus* and *E. multilocularis*. The overall proportion of MUP did not differ significantly between the eggs of the two species (*E. granulosus*, 59.8%; *E. multilocularis*, 57.4%; $\chi^2 = 0.42$, $P = 0.517$), suggesting that the abundance of proteins with high instability index among egg-associated proteins may be evolutionarily constrained, possibly by fundamental requirements for protein homeostasis during prolonged environmental persistence. However, the functional profiles of egg-associated MUP appeared to differ between the species: MUP in *E. granulosus* eggs were significantly enriched in transcription regulation, signal transduction regulation, and the tropomyosin and LDL receptor-like superfamily domains, whereas no term reached significance in *E. multilocularis* after BH correction, although enrichment trends were observed. Superficially, this apparent asymmetry suggests

species-specific functional specialization. However, as detailed below, a balanced down-sampling analysis showed that this apparent divergence reflects differential annotation coverage and statistical power rather than genuine species-specific specialization; the functional landscape of egg-associated MUP is therefore largely conserved between the two species.

Significant enrichment in transcriptional regulation and signal transduction was a prominent feature of MUP in *E. granulosus* eggs. The enrichment of GO:0045944 (positive regulation of transcription by RNA polymerase II) and GO:0003713 (transcription coactivator activity) suggests transcription-related proteins are overrepresented among MUP in *E. granulosus* eggs. Although mature tapeworm eggs are quiescent and have low global transcriptional activity, metabolically unstable transcriptional regulators may serve as maternal reserve proteins whose short half-life ensures their rapid clearance when not needed, thereby preventing sustained activation of transcriptional programs; upon hatching signals, however, these factors can be rapidly resynthesized to support the swift initiation of gene expression programs required for oncosphere development. After being expelled from the definitive host, the eggs must survive for extended periods, ranging from months to years, in the external environment, where they encounter stresses such as fluctuations in temperature and humidity. Furthermore, during prolonged environmental survival, the rapid turnover of DNA repair-related regulatory factors may facilitate dynamic adjustment of repair complex composition in response to accumulating stress, contributing to genome integrity maintenance. The enrichment of kinase-binding and phosphorylation-regulatory proteins among egg-associated MUP is consistent with the broader paradigm that rapid turnover of signalling regulators is critical for developmental transitions in parasitic helminths. For instance, insulin signalling has been shown to drive stem cell proliferation and larval growth in *Echinococcus* through PI3K/Akt pathway activation [12], underscoring the importance of dynamic phosphorylation-dependent signalling networks in parasite development.

The enrichment of protein kinase binding (GO:0019901) and regulation of protein binding (GO:0043393) suggests that signal-transduction regulatory proteins are overrepresented among MUP in *E. granulosus* eggs. These short-half-life kinase-interacting proteins may participate in transient signaling complexes that sense environmental cues—such as temperature shifts, osmotic changes, or host-derived stimuli—and transduce them into adaptive responses. The rapid degradation of kinase-binding proteins ensures that signaling complexes are promptly cleared upon completion of their mission, preventing sustained pathway activation and unnecessary energy expenditure. The concurrent enrichment of negative regulation of phosphorylation (GO:0042326) and kinase-binding terms suggests a balance between pathway activation and inhibition. This pattern, together with the enrichment of kinase-binding terms, suggests that protein degradation and dephosphorylation may act synergistically to achieve rapid signal resetting. Such bidirectional regulation may allow precise control of signaling pathways, enabling the eggs to respond appropriately rather than excessively to environmental changes.

The enrichment of endomembrane system organization (GO:0010256) suggests the presence of an active membrane-vesicle trafficking system within *E. granulosus* eggs [29]. This system may be involved in the packaging and release of secretory proteins, such as hatching enzymes or protective shell

proteins, thereby providing a material basis for invasive activity after prolonged environmental survival. Transcriptomic evidence from *E. granulosus* protoscoleces during encystation has implicated dynein light chain, Ras-related proteins, and aquaporin 4 in vesicle-mediated transport and water regulation [29], suggesting that similar trafficking mechanisms may operate in eggs to maintain cellular homeostasis under environmental stress. The concurrent significant enrichment of regulation of DNA metabolic process (GO:0051052) and nucleobase-containing compound biosynthetic process (GO:0034654) suggests that egg-associated MUP include regulatory factors associated with DNA replication and repair, as well as nucleotide synthesis. During their prolonged environmental survival, tapeworm eggs may be exposed to DNA-damaging factors, such as ultraviolet (UV) radiation and oxidative stress. The rapid turnover of these regulatory proteins may facilitate dynamic adjustment of DNA-repair complex composition in response to accumulating stress, allowing the eggs to modulate repair capacity without sustained commitment of specific regulatory factors.

GO:0007610 (behavior) and GO:0043010 (camera-type eye development) were also significantly enriched. However, these terms lack clear anatomical and physiological counterparts in tapeworms, and their biological significance remains uncertain. Therefore, they were excluded from Table 2 and were not considered part of the core findings; the complete results are provided in S1 Table. Nevertheless, their enrichment may reflect unknown functions related to stress responses or chemotaxis. Further experimental studies are required to evaluate this possibility.

Tropomyosin (IPR000533) is a cytoskeletal protein involved primarily in actin-filament stabilization and cell contraction. In tapeworm eggs, tropomyosin may contribute to the development of muscle precursors in oncospheres or to the mechanical resilience of the embryonic membrane. Its predicted instability may allow rapid turnover of cytoskeletal regulatory proteins during prolonged dormancy, maintaining structural integrity while preparing for cytoskeletal remodeling upon hatching. Tropomyosin is an immunogenic antigen in several parasitic species, including *Taenia solium* and *Trichinella* species. Its metabolic instability may have implications for the stability and reproducibility of diagnostic reagents based on this protein and should therefore be considered in their future development. This observation is corroborated by proteomic data showing that paramyosin, a tropomyosin-family protein, is significantly upregulated in the *E. granulosus* oncosphere compared with other developmental stages [28]. The LDL receptor-like superfamily (IPR036055) is generally involved in lipoprotein uptake and signal transduction. Eggs cannot obtain nutrients outside the host; however, upon entry into the host intestine, LDL receptor-like proteins may assist oncospheres in acquiring host-derived lipids during early development. The short half-life of such receptor proteins may help limit their unnecessary accumulation during environmental survival, thereby conserving energy, while allowing rapid synthesis of new receptors upon host entry to respond to nutritional signals.

For KEGG pathways, microRNAs in cancer (map05206) showed an enrichment trend in *E. granulosus* (raw $P < 0.05$) but was not significant after BH correction. Although this pathway is cancer-related, its core components, including those of the MAPK and PI3K-Akt pathways, are highly conserved in

tapeworms. This enrichment trend may reflect regulatory demands of microRNA-related processes in eggs rather than direct involvement in cancer-associated processes.

Although the overall proportion of MUP did not differ significantly between the two species, their functional enrichment patterns appeared to differ markedly. In *E. granulosus*, MUP were significantly enriched in specific functions, including transcriptional regulation, signal transduction, and cytoskeletal remodeling (BH-adjusted $P < 0.05$), whereas enrichment trends in *E. multilocularis* were more dispersed, with no functional category reaching statistical significance.

Two non-mutually-exclusive explanations could account for this apparent asymmetry: genuine species-specific specialization, or differential annotation coverage and statistical power. GO and KEGG annotation coverage for egg-associated proteins was significantly lower in *E. multilocularis* (70.22% and 70.85%, respectively) than in *E. granulosus*(79.81% and 83.39%, respectively; GO: $\chi^2 = 10.438$, $P = 0.001$; KEGG: $\chi^2 = 19.627$, $P < 0.001$). *E. multilocularis* also had fewer egg-associated MUP than *E. granulosus* (183 vs. 385). Together, these factors may have reduced statistical power. Despite 29 raw $P < 0.05$ terms, none survived BH correction in *E. multilocularis* (Table 3). The cross-species comparison showed that the fold enrichment values of shared terms were generally similar between the species (Table 5). For example, the values for GO:0019901 (protein kinase binding) were 1.365 in *E. granulosus* and 1.391 in *E. multilocularis*. This pattern suggests that the lack of significant enrichment in *E. multilocularis* may reflect limited statistical power rather than necessarily weaker functional enrichment among its egg-associated MUP. To distinguish between these explanations, we applied a balanced down-sampling framework (Methods 2.10). When *E. granulosus* MUP were down-sampled to the *E. multilocularis* sample size and annotation coverage, per-term bootstrap significance rates across the eight key shared GO terms were ≤ 4.4% (S10 Table), and the number of significantly enriched terms fell monotonically with decreasing sample size (8, 7, 6, 4, 2 and 0 at 644, 500, 400, 319, 250 and 183 proteins, respectively; S11 Table). Bootstrap 95% confidence intervals of fold enrichment for all eight shared terms overlapped completely between the species (S13 Table), and bootstrap resampling of the *E. multilocularis* enrichment-trend terms yielded a 0% significance rate (S12 Table). Reclassification under alternative instability-index thresholds (35, 40 and 45) affected only *E. granulosus* (1, 8 and 6 significant terms, respectively), not *E. multilocularis* (S14 Table). Together, these analyses indicate that the apparent interspecific differences in enrichment reflect differential annotation coverage and statistical power rather than genuine species-specific specialization, and that the functional landscape of egg-associated MUP is largely conserved between the two species.

To determine whether the prioritized candidates are transcribed at the transmission-relevant stage, we cross-validated the *E. multilocularis* candidates against the oncosphere transcriptome of Huang et al. [27]. Of 183 predicted MUP in *E. multilocularis* eggs, 182 had a significant best blastp hit (E < 1e-5) in the *E. multilocularis* proteome, and 151 (82.5%) of these were transcribed in the activated oncosphere. On the *E. granulosus* side, an egg-stage proteome further supported protein-level expression of regulatory proteins in the oncosphere stage[28], providing qualitative corroboration. These results indicate that the majority of computationally flagged candidates are transcribed at the egg/oncosphere stage, although

the labile candidates not detected in the transcriptome (17/182) may reflect stage- or condition-specific expression rather than true absence of expression.

The present study provides, to our knowledge, the first systematic catalog of predicted metabolically unstable proteins associated with *Echinococcus* eggs, together with a prioritized candidate list (S15 Table) spanning transcription/signal-transduction regulators, tropomyosin, and LDL receptor-like domains. It also introduces a reusable, annotation- and power-balanced framework for cross-species enrichment comparison, which should help future studies avoid mistaking differential annotation coverage for genuine biological divergence. We stop short of claiming a causal role for protein lability in egg persistence and transmission; the contribution of these candidates to egg survival and host entry remains a testable hypothesis rather than an established mechanism.

Of note, no molecular chaperone (e.g., HSP)-related functions appeared among the enriched terms in egg-associated MUP. This is consistent with their longer half-lives (classified as stable proteins) and further corroborates that MUP are primarily enriched in regulatory rather than housekeeping functions. Indeed, heat shock proteins are classified as stable proteins (instability index ≤ 40) despite their high expression in oncospheres [28], consistent with their absence from the metabolically unstable protein enrichment results.

Nevertheless, this study has certain limitations. Because direct studies on *Echinococcus* egg proteins are limited, the keyword list was derived from parasite egg protein studies in general. This approach may miss *Echinococcus*-specific egg proteins and may include proteins that are not expressed in *Echinococcus* eggs. The curated dataset should therefore be considered a hypothesis-generating resource rather than a comprehensive, experimentally verified egg proteome. The background set for subgroup proportion analysis was limited to proteins with GO, KEGG, or IPR domain annotations, rather than all proteins encoded by the genome. A similar strategy (enriching against a stage-matched background rather than the whole-genome background) has been applied in proteomic studies of other parasitic helminth eggs, such as *Schistosoma japonicum*, to minimize dilution by stage-irrelevant proteins [30]. Some unannotated proteins may have important functions but were excluded from the analysis, potentially limiting the comprehensiveness of the enrichment results. The annotation gap is not unique to our dataset; transcriptomic analyses of *E. multilocularis* oncospheres have similarly identified numerous uncharacterised transcripts [27]. Second, the instability index was predicted from amino acid dipeptide composition and reflects trends in *in vivo* half-life rather than experimentally determined degradation rates. Tissue- or stage-specific differences in proteasome activity may exist within the eggs, potentially causing deviations between predicted and actual half-lives under different physiological states. Furthermore, the threshold of 40 is somewhat arbitrary, and converting continuously varying stability values into binary categories may have resulted in information loss. Although BH correction reduces the false discovery rate, it may also fail to detect genuine enrichments with smaller effect sizes. The enrichment trends observed in *E. multilocularis* may fall into this category and should therefore be re-evaluated using a larger sample size. Third, the keyword-based screening strategy inherently favours proteins with known antigenic, serodiagnostic, or immunogenic properties, reflecting the current state of

the literature in which egg-associated proteins of *Echinococcus* have been predominantly investigated as diagnostic markers or vaccine candidates. Our dataset therefore represents a curated collection of egg-associated immunogenic proteins rather than a comprehensive egg proteome. Functional categories less commonly targeted in antigen-discovery studies, such as core metabolic enzymes or structural components, may be underrepresented, and this selection bias should be considered when interpreting the apparent prominence of regulatory and signalling-related enrichment terms. The literature-based keywords were derived primarily from known egg antigens and structural proteins. Consequently, some unknown egg-specific proteins may have been omitted. In addition, keyword matching could not confirm that all screened sequences were highly expressed during the egg stage, because some may also be expressed during the adult or larval stages. Finally, with respect to the cross-validation, absence of a transcriptomic signal does not prove the absence of protein expression, and 14 *E. multilocularis* candidates with no look-up entry in the oncosphere transcriptome were treated as missing rather than as negative evidence; on the *E. granulosus* side, the complete differential protein table underlying [28] is not publicly available, so our protein-level corroboration there remains qualitative rather than a fully comparable cross-check.

In conclusion, the overall proportion of egg-associated MUP is conserved between the eggs of *E. granulosus* and *E. multilocularis*, and—after controlling for annotation coverage and statistical power—their functional landscape is largely conserved as well. The apparent interspecific differences in enrichment reported at face value reflect differential annotation completeness and statistical power (S10–S14 Tables) rather than genuine species-specific specialization. A majority of the prioritized candidates are transcribed at the oncosphere stage, and the candidate list (S15 Table) offers a focused resource for experimental follow-up. Larger *E. multilocularis* sampling will be required to resolve enrichment at full power, and our balanced comparison framework is reusable for cross-species egg proteome studies. Future work integrating protein-stability assays with measures of egg persistence and transmission will be needed to test the proposed functional hypotheses.

Declarations

Ethics approval and consent to participate

Not applicable. This study is an in silico bioinformatics analysis of publicly available protein sequence data and did not involve human participants, human data, human tissue, or animal experiments.

Consent for publication

Not applicable. This manuscript does not contain data from any individual person in any form.

Ethics approval and consent to participate

Not applicable. This study is an in silico bioinformatics analysis of publicly available protein sequence data and did not involve human participants, human data, human tissue, or animal experiments.

Availability of data and materials

All protein sequence data analysed in this study were retrieved from the publicly available NCBI Protein Database (https://www.ncbi.nlm.nih.gov/protein/)); the complete search strategy and the keyword list used to screen egg-associated proteins are described in the Materials and Methods section and are based on publicly available PubMed records. The *E. multilocularis* oncosphere transcriptome used for cross-validation is from the published study of Huang et al. (*PLoS Negl Trop Dis*. 2016;10(4): e0004634, https://doi.org/10.1371/journal.pntd.0004634); the *E. granulosus* egg-stage iTRAQ proteome is from the published study of Li et al. (*Parasite*. 2021;28:15, DOI as given in the reference list). All data generated during this study, including the non-redundant egg-associated protein dataset, functional annotation results, and enrichment outputs, are included in this article and its Supporting Information files (S1–S15 Tables). No restricted or proprietary data were used.

Competing interests

The authors declare that they have no competing interests.

Funding

This work was supported by National Key Research and Development Program of China

[grant 2021YFC2300800]. The funders had no role in the study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Author Contributions

Conceptualization: SH CX. Data curation: SH. Formal analysis: MC CX. Funding acquisition: JW. Methodology: SH CX JW WH. Project administration: WH. Resources: SH CX JW. Software: CY. Supervision: WH. Visualization: SH MC JW WH. Writing – original draft: SH. Writing – review & editing: SH CX JW WH.

Acknowledgements

Not applicable.

References

1. Eckert J et al. WHO/OIE Manual on echinococcosis in humans and animals: A public health problem of global concern. 2001. Available from: https://www.who.int/publications/i/item/929044522X
2. Barosi R, Umhang G. Presence of Echinococcus eggs in the environment and food: a review of current data and future prospects. Parasitology. 2024;151(13):1416–31.
3. Gemmell MA, Taeniidae. Modification to the life span of the egg and the regulation of tapeworm populations. Exp Parasitol. 1977;41(2):314–28.

4. Bucur I, Gabriël S, Van Damme I, Dorny P, Vang Johansen M. Survival of Taenia saginata eggs under different environmental conditions. Vet Parasitol. 2019;266:88–95.
5. Sánchez Thevenet P, Alvarez HM, Basualdo JA. Survival, physical and physiological changes of Taenia hydatigena eggs under different conditions of water stress. Exp Parasitol. 2017;177:47–56.
6. Laws GF. Physical factors influencing survival of taeniid eggs. Exp Parasitol. 1968;22(2):227–39.
7. Aleuy OA, Kutz S. Adaptations, life-history traits and ecological mechanisms of parasites to survive extremes and environmental unpredictability in the face of climate change. Int J Parasitol Parasites Wildl. 2020;12:308–17.
8. Becher I, et al. Pervasive Protein Thermal Stability Variation during the Cell Cycle. Cell. 2018;173(6):1495–e150718.
9. Geng P, et al. A thermosensor FUST1 primes heat-induced stress granule formation via biomolecular condensation in Arabidopsis. Cell Res. 2025;35(7):483–96.
10. Roquis D, et al. The Epigenome of Schistosoma mansoni Provides Insight about How Cercariae Poise Transcription until Infection. PLoS Negl Trop Dis. 2015;9:e0003853. 10.1371/journal.pntd.0003853.
11. Briquet S, et al. Preparing for Transmission: Gene Regulation in Plasmodium Sporozoites. Front Cell Infect Microbiol. 2021;10:618430.
12. Hemer S, Konrad C, Spiliotis M, Koziol U, Schaack D, Förster S, et al. Host insulin stimulates Echinococcus multilocularis insulin signalling pathways and larval development. BMC Biol. 2014;12:5.
13. Pescher P, Ballez O, Crepieux G, Späth GF, Hotez PJ, Goyal G, et al. Five-layer systems analysis of Leishmania stage differentiation reveals an essential role for protein degradation in parasite development. eLife. 2026;14:e111115.
14. Li Q, et al. Bioinformatic analysis of the molecular characteristics of major egg antigens from Echinococcus. China Trop Med. 2022;22(7):611–6.
15. Marabotti A, Scafuri B, Facchiano A. Predicting the stability of mutant proteins by computational approaches: an overview. Brief Bioinform. 2021;22(3):bbaa074.
16. Guruprasad K, Reddy BVB, Pandit MW. Correlation between stability of a protein and its dipeptide composition: a novel approach for predicting in vivo stability of a protein from its primary sequence. Protein Eng. 1990;4(2):155–61.
17. Gasteiger E, et al. Identification and Analysis Tools on the ExPASy Server. In: Walker JM, editor. The proteomics protocols handbook. Totowa (NJ): Humana; 2005. pp. 571–607.
18. Ashburner M, et al. Gene Ontology: tool for the unification of biology. Nat Genet. 2000;25(1):25–9.
19. Paysan-Lafosse T, et al. InterPro in 2022. Nucleic Acids Res. 2023;51(D1):D418–27.
20. Kanehisa M, Goto S. KEGG: kyoto encyclopedia of genes and genomes. Nucleic Acids Res. 2000;28(1):27–30.

21. Yu G, et al. clusterProfiler: an R package for comparing biological themes among gene clusters. Omics. 2012;16(5):284–7.
22. Laukens K, Naulaerts S, Berghe W. Bioinformatics approaches for the functional interpretation of protein lists: From ontology term enrichment to network analysis. Proteomics. 2015;15(5–6):981–96.
23. Deplazes P, et al. Chapter Six - Global Distribution of Alveolar and Cystic Echinococcosis. In: Deplazes P, et al. editors. Advances in Parasitology. Academic; 2017. pp. 315–493.
24. Zheng H, Zhang W, Zhang L, Zhang Z, Li J, Lu G, et al. The genome of the hydatid tapeworm Echinococcus granulosus. Nat Genet. 2013;45(10):1168–75.
25. Zhang H, Tian M, Qi W, Wu J, Zheng H, Guo G, et al. Bioinformatic comparison of Kunitz protease inhibitors in Echinococcus granulosus sensu stricto and E. multilocularis and the genes expressed in different developmental stages of E. granulosus s.s. BMC Genomics. 2021;22(1):887.
26. Zhou Y, Zhao S, Li Y, Yu M, Zheng J, Gong Q, et al. Design and functional preliminary investigation of recombinant antigen EgG1Y162–EgG1Y162 against Echinococcus granulosus. Open Life Sci. 2023;18(1):2020558.
27. Huang F, Dang Z, Suzuki Y, Horiuchi T, Yagi K, Kouguchi H, et al. Analysis on gene expression profile in oncospheres and early stage metacestodes from Echinococcus multilocularis. PLoS Negl Trop Dis. 2016;10(4):e0004634.
28. Li X, Jiang S, Wang X, Hui W, Jia B. iTRAQ-based comparative proteomic analysis in different developmental stages of Echinococcus granulosus. Parasite. 2021;28:15.
29. Fan J, Wu H, Li K, Liu X, Tan Q, Cao W, et al. Transcriptomic features of Echinococcus granulosus protoscolex during the encystation process. Korean J Parasitol. 2020;58(3):287–99.
30. De Marco Verissimo C, Potriquet J, You H, McManus DP, Mulvenna J, Jones MK. Qualitative and quantitative proteomic analyses of Schistosoma japonicum eggs and egg-derived secretory-excretory proteins. Parasit Vectors. 2019;12(1):173.

Figures

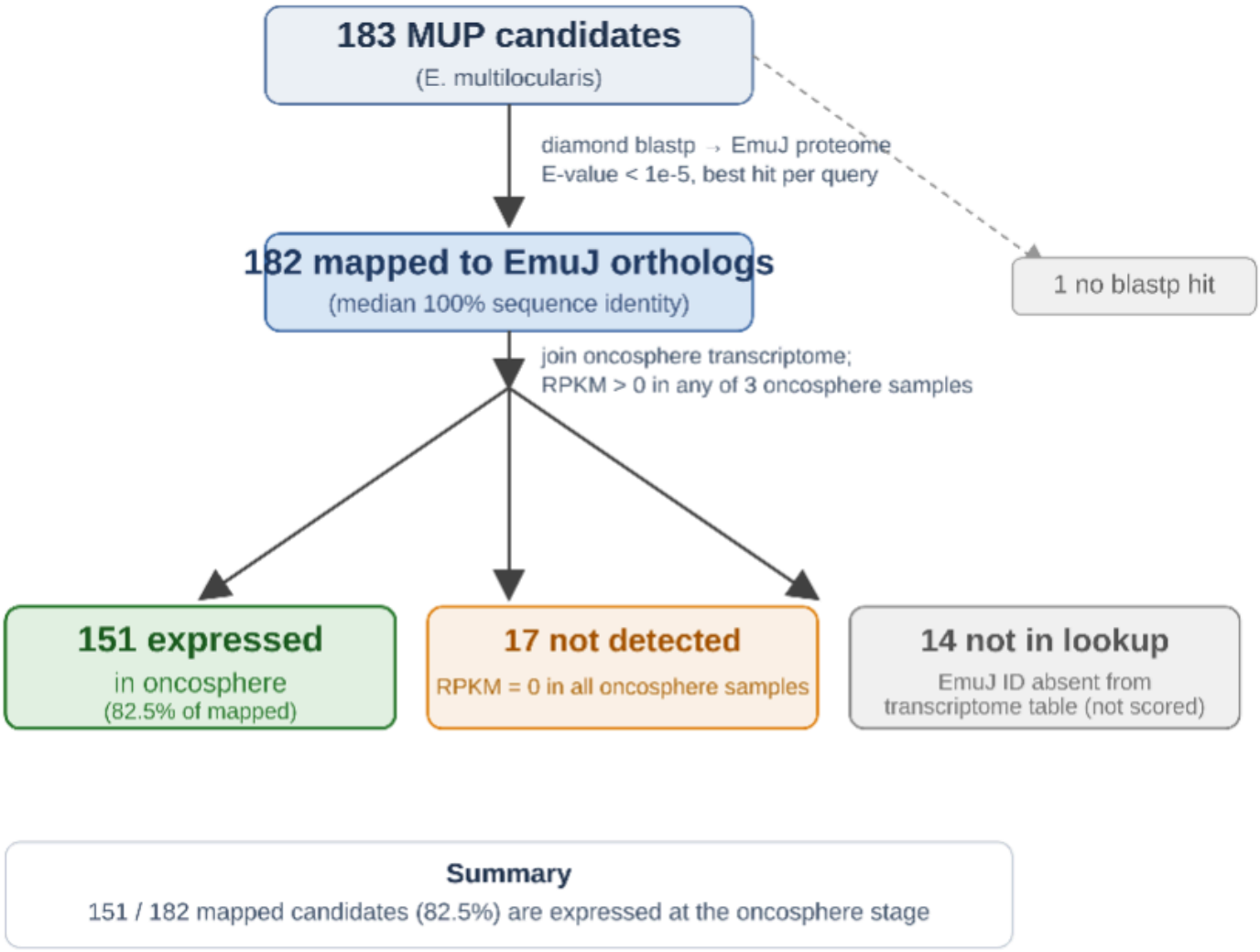


Figure 1

Cross-validation of *E. multilocularis* MUP candidates against the oncosphere transcriptome. Metabolically unstable proteins (MUP) of *E. multilocularis* (n = 183) were mapped to the EmuJ proteome by BLASTP (E-value < 1e-5) and their encoding genes were cross-checked against the published oncosphere transcriptome (Huang et al., 2016). Of 182 successfully mapped candidates, 151 (82.5%) were transcribed in the oncosphere; 17 were not detected, and 14 were absent from the transcriptome query table.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- S1Table.docx
- S2Table.docx
- S3Table.docx

- S4Table.docx
- S5Table.docx
- S6Table.docx
- S7Table.docx
- S8Table.docx
- S9Table.docx
- S10Table.docx
- S11Table.docx
- S12Table.docx
- S13Table.docx
- S14Table.docx
- S15Table.docx
- Supportinginformationcaptions.docx