scientific data

Check for updates

OPEN

DATA DESCRIPTOR

De novo transcriptome assembly of a pigmentation-variable population of cave snail *Physella* sp

Magdalena Grgić[1], Robert G. Weck[2], Gaj Keresteš[1,3], Lara Kalogjera[1,4] & Helena Bilandžija[1]✉

Subterranean ecosystems are among the most widespread yet least explored environments on Earth. Species that are not obligate cave dwellers are especially understudied, due to the predominant research focus on obligate subterranean organisms. Although aquatic snails represent one of the most species-rich groups in groundwater, no genomic resources have been available for these taxa until now. To fill this gap, we generated a reference transcriptome for a cave-dwelling snail of the genus *Physella*. The *Physella* population from Fogelpole Cave, IL, USA, exhibits heteromorphic phenotypes, with individuals ranging from fully pigmented to completely lacking pigment. The reference transcriptome generated in this study shows high BUSCO scores, high mapping rates and robust overall quality metrics. These datasets provide a valuable resource for investigating molecular changes associated with the early stages of cave adaptation and the evolution of specialized cave traits, especially the loss of pigmentation.

Background & Summary

Mollusks, the second most biodiverse group of animals on Earth, are an ecologically diverse phylum. Mollusks have successfully colonized nearly every habitat, from the deep sea to terrestrial environments, and even caves. Caves are among the most extreme and least explored habitats on Earth. They are characterized by strong selective pressures, including permanent darkness, limited food resources, and disrupted daily and annual rhythms, which lead to the evolution of specialized traits in cave-dwelling species[1,2]. The most famous is the loss of eyes and pigmentation, but notable changes in metabolism, circadian rhythms, reproduction, nervous system and behaviors also accompany life in caves[1,2]. The presence of similar selection pressures in different caves makes cave animals a prime example of convergent evolution in nature and excellent models for evolutionary studies.

Gastropods are the most abundant subterranean mollusks, as only a few bivalve species are found in caves[3,4]. Obligate cave snails, with 450 described species, represent the second most abundant group in groundwater ecosystems, after crustaceans[5]. However, snails remain one of the most understudied groups in caves[5–7]. They are also highly threatened, as freshwater snails have the highest documented extinction rate of any taxonomic group[7]. Given the general lack of knowledge on subterranean fauna[8], the importance of snails in groundwater communities, and their elevated risks of extinction, they are one of the highest-priority groups for research.

While obligate cave fauna, species that complete their life cycle in caves, are understudied and largely neglected in conservation and management policies[8], facultative cave species (troglophiles) attract even less attention. Although troglophiles are frequently encountered in caves, they lack prominent cave adaptations that attract researchers and the general public, so the knowledge gap regarding their biology and ecology is even greater. One such example are the cave-dwelling populations of snails in the genus *Physella*, which inhabit several caves in southwestern Illinois, USA (Fig. 1a)[9].

The taxonomy of Physella (family Physidae) is well known to be problematic, largely due to the reliance on shell morphology, which is highly plastic and environmentally influenced[10]. Molecular studies have shown limited genetic divergence among morphologically distinct taxa, alongside cases of cryptic diversity, complicating species-level identification[11]. These challenges are particularly pronounced in cave populations, and

[1]Ruđer Bošković Institute, Bijenička cesta 54, Zagreb, 10000, Croatia. [2]Southwestern Illinois College, Belleville, Illinois, 62221, USA. [3]Present address: University of Zagreb, Faculty of Science, Horvatovac 102a, Zagreb, 10000, Croatia. [4]Present address: Institute of Microbiology and Microbial Biotechnology, Department of Biotechnology and Food Science, BOKU University, 1190, Vienna, Austria. ✉e-mail: hbilandz@irb.hr

genomic data are essential for resolving taxonomic placement and understanding the evolutionary status of these populations.

Fogelpole Cave in Monroe County, Illinois is home to a highly polymorphic population of *Physella* snails, with a spectrum of phenotypes ranging from individuals lacking any obvious pigment to fully dark and pigmented snails[12]. The polymorphic nature of the Fogelpole Cave *Physella*, coupled with the relative ease with which laboratory colonies of these snails can be established[13], creates a unique opportunity to study the genetic and evolutionary mechanisms underlying the early stages of cave adaptation, making it an ideal system for investigating colonization of subterranean habitats.

To gain a mechanistic understanding of the evolutionary processes driving cave adaptations, the availability of genomic data is essential. While genomic resources exist for cave-adapted fish, salamanders, and crustaceans[14–23], mollusks remain a major gap in subterranean genomics. To address this gap, we have generated the first genomic data for any cave snail. The transcriptomic datasets of *Physella* sp. from Fogelpole Cave will provide a crucial foundation for understanding the evolution of cave-dwelling species and the genetic basis of specialized cave traits in a new phylum. These datasets are particularly valuable for understanding the colonization of caves and the early stages of adaptation to life underground. In particular, they can be used to study the genetic basis of pigmentation loss, one of the most common cave adaptations, whose molecular mechanisms remain largely elusive[23–25], with the exception of cavefish *Astyanax mexicanus*[26]. In addition to this vertebrate model, the crustacean *Asellus aquaticus* has emerged as a powerful invertebrate system for studying cave evolution, with recent work providing important insights into the genetic architecture and evolutionary pathways underlying eye and pigmentation loss[27–30]. Despite these advances, comparable genomic resources are lacking for mollusks.

Moreover, melanogenesis is implicated in a range of biological functions beyond body pigmentation across diverse taxa[31,32] and numerous pleiotropic effects associated with melanin loss have been discovered in *Astyanax*[33–35]. *Physella* therefore represents an alternative model system for investigating the evolution of pigmentation loss and the potential role of natural selection favoring albinism through pleiotropy. Finally, the availability of genomic datasets for different morphs of this species also enables investigation of the genetic basis of other cave-related changes in *Physella*, including reduced growth, delayed sexual maturity and decreased fecundity[12].

Methods

Snail collection and hybrid crosses. Adult snails of different morphotypes were collected from Fogelpole Cave, Illinois, USA, and brought to the laboratory to establish breeding cultures. In the laboratory, snails were maintained at room temperature (20–22 °C) in containers with regular water changes and were fed ad libitum to ensure stable growth and reproduction. Detailed protocols for cultivating snails in the laboratory are described elsewhere[13].

Parental crosses between pigmented and depigmented individuals were initiated by pairing immature snails (ca. 2 mm shell length) in small containers. After reaching reproductive maturity, parent snails laid eggs, and F1 embryos were isolated. Physids are hermaphrodites and capable of self-fertilization, though they typically outcross when mates are available[36]. To ensure that F1 offspring were the result of outcrossing, the depigmented parents from each cross were periodically isolated and allowed to deposit eggs. Developing embryos were examined prior to hatching, and eye pigmentation was used as an indicator of successful outcrossing. Given that parental individuals exhibited stable pigmentation phenotypes and that previous studies have demonstrated a simple Mendelian basis of this trait[12], the appearance of pigmentation in offspring derived from depigmented individuals was interpreted as evidence of hybrid origin. However, low levels of residual heterozygosity or rare deviations from expected genotype–phenotype relationships cannot be entirely excluded. To produce the F2 generation, juvenile F1 snails were paired in separate containers. Following maturation, egg masses produced by these crosses were collected, and resulting F2 offspring were reared until they reached a size suitable for phenotypic assessment and downstream tissue sampling.

Sampling and sample groups. Two sample groups were used for transcriptome sequencing. The first group consisted of 10 adult F2 individuals (5 lightly pigmented snails and 5 snails with dark pigmentation) derived from a cross between an albino and a pigmented snail, as described above. Snails were starved for a week before sampling to empty their guts and reduce contamination with non-target taxa. After a week they were flash-frozen in liquid nitrogen and kept at −80 °C until RNA isolation.

The second sampling group consisted of the foot tissue of albino and pigmented individuals from laboratory culture. Animals were anesthetized with MS222 (3 g/L) which caused foot extension. The posterior part of the foot was quickly severed and immediately stored in DNA/RNA Shield (cat. # R1100-250, Zymo Research, USA) until RNA isolation. Since the mass of tissue from each individual was minute, the foot from 3 individuals were pooled in one sample. In total, 4 such samples were prepared from albino individuals and 4 from pigmented ones.

Species identification. To identify the species in Fogelpole Cave, we sequenced 16S and CO1 phylogenetic markers[37–42] and reconstructed the phylogeny of Physidae (Fig. 1b). We used the following primers: ColFol-rev 5′-TAAACTTCNGGRTGNCCAAAAAATCA-3′ and ColFol-for 5′-TTTCAACAAATCATAARGAYATYGG-3′ for CO1 and 16Sar-L 5′-CGCCTGTTTATCAAAAACAT-3′ and 16SbR_Isopoda 5′-CCGGTCTGAACTCAGATCACG-3′ for 16S. Physid DNA from several localities including Fogelpole Cave and Losing Pond, Monroe County, IL and Jarun Lake, Zagreb, Croatia was isolated using DNeasy Blood & Tissue Kit (cat. # 69504, Qiagen, Germany), following the manufacturer's protocol. Thermal cycling conditions were: initial denaturation at 94 °C for 3 min, followed by 35 cycles of denaturation at 94 °C for 45 seconds, annealing at 45 °C for COI and 55.3 °C for 16S for 45 seconds, and extension at 72 °C for 1 minute, concluding with a 5-minute final extension at 72 °C. The PCR mixture consisted of 20 uL 2x PCR mix (ReadyMix Taq PCR Reaction Mix, cat. # P4600, Sigma Aldrich, USA), 1.2 uL of forward and reverse primers, 12.4 uL

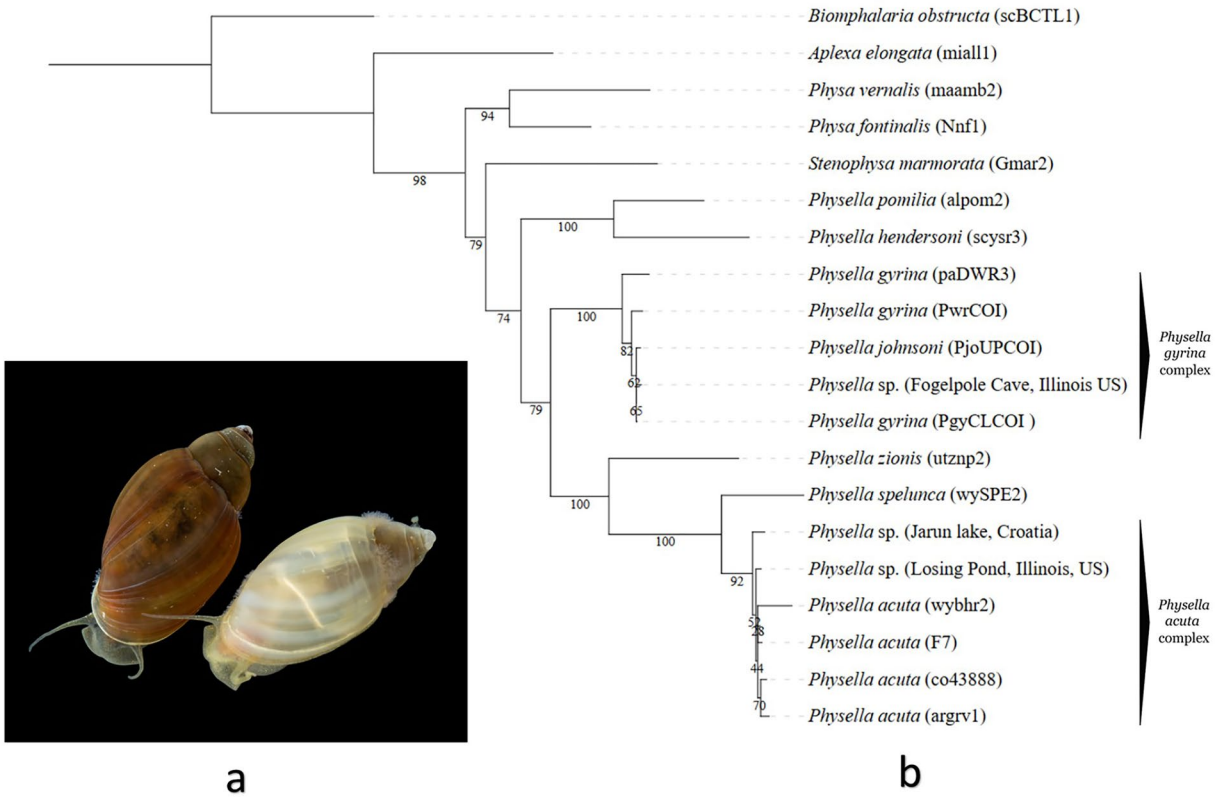


Fig. 1 Model species and its phylogenetic position. (**a**) Two differently pigmented morphotypes of *Physella* sp. from Fogelpole Cave. Photo credit: Tin Rožman. (**b**) Maximum-likelihood (ML) tree of Physidae based on 16S and CO1 phylogenetic markers. Branch support values are based on 1000 ultrafast bootstrap (UFBoot) replicates. *Physella* sp. from Fogelpole Cave is placed within the *Physella gyrina* clade.

H2O, 1.2 uL Mg^{2+} and 4 uL of DNA (in total 80 ng). For purification, 10 uL of PCR product was mixed with 0.2 uL Exonuclease I and 1 uL FastAP Thermosensitive Alkaline Phosphatase (cat. # EN0582 and EF0651, Thermo Fisher Scientific, USA) and incubated at 37 °C for 45 min followed by 80 °C for 15 min. PCR products were sequenced using the Sanger method.

16S and CO1 sequences from different physid species and outgroup were retrieved from the NCBI[43–76] and processed using Geneious Prime software 2024.0.7. Sequences were aligned with MUSCLE 5.1[77] using the Super5 algorithm. Alignment ends were trimmed, and sites with 50% gaps were masked. Resulting alignments were used to construct a maximum-likelihood (ML) tree with IQtree 2.2.6[78–80] using ModelFinder and a partition model for each marker. Branch support values were estimated using 1000 ultrafast bootstrap (UFBoot) replicates, and the resulting tree was visualized in iTOL (online version v7; accessed 10/3/2025) and rooted with *Biomphalaria obstructa* as an outgroup.

Since snails within Physidae have inconsistent use of species names in the literature and several taxonomic revisions were conducted thus far[10,11,81,82], here we used current species names from MolluscaBase (https://www.molluscabase.org/, accessed on 10/3/2025) for species designations in the phylogenetic tree. A detailed list of sequence accession numbers and original identifiers of specific isolates with their NCBI names and their names in MolluscaBase can be found in Table 1.

RNA extraction, library preparation and sequencing. Whole RNA from 18 samples was extracted using SPLIT RNA Extraction Kit (cat. # SKU: 008.48, Lexogen, Austria) following the suggested protocol. RNA concentration was measured using spectrophotometry, and the quality and integrity of the isolated RNA was assessed using gel electrophoresis. During protocol optimization for RNA isolation, gel electrophoresis as a qualitative assessment method for RNA integrity was confirmed with Agilent 2100 Bioanalyzer traces by evaluating RNA profiles. The success of DNase treatment (TURBO DNase, cat. # AM2239, Thermo Fisher Scientific, USA) in removing potential DNA contamination was confirmed by the absence of CO1 amplification using primers HCO2198coi (5′-TANACYTCNGGRTGNCCRAARAATCA-3′) and LepF1 (5′-ATTCAACCATCATAAAGATATTGG-3′). The thermal cycling protocol was as follows: initial denaturation at 95 °C for 15 min, followed by 40 cycles of denaturation at 94 °C for 30 seconds, annealing at 45 °C for 1 minute, and extension at 72 °C for 90 seconds, concluding with a 10-minute final extension at 72 °C. Final RNA concentrations were measured using a fluorometric assay (QFX Fluorometer and DeNovix RNA Quantification Kit, cat. # KIT-RNA-2-NS, DeNovix, USA).

Name related to NCBI accession number	MolluscaBase name (accessed on 10/3/2025)	NCBI Accession number - 16S	NCBI Accession number - CO1	Sample identity
Physella johnsoni	*Physella johnsoni* Clench, 1926	AF346750.1	AF346738.1	PjoUPCOI
Physella gyrina	*Physella gyrina* Say, 1821	AF346755.1	AF346743.1	PgyCLCOI
Physella wrighti	*Physella gyrina* Say, 1821	AF346757.1	AF346745.1	PwrCOI
Physella virgata	*Physella acuta* Draparnaud, 1805	AY651209.1	AY651170.1	argrv1
Physella anatina	*Physella acuta* Draparnaud, 1805	AY651214.1	AY651175.1	co43888
Physa acuta	*Physella acuta* Draparnaud, 1805	AY651224.1	AY651186.1	F7
Physa fontinalis	*Physa fontinalis* Linnaeus, 1758	AY651227.1	AY651189.1	Nnf1
Physella hendersoni	*Physella hendersoni* Clench, 1925	AY651234.1	AY651196.1	scysr3
Petrophysa zionis	*Physella zionis* Pilsbry, 1926	AY651236.1	AY651198.1	utznp2
Physa acuta	*Physella acuta* Draparnaud, 1805	AY651241.1	AY651203.1	wybhr2
Physa vernalis	*Physa vernalis* D. W. Taylor & Jokinen, 1984	EU038329.1	EU038376.1	maamb2
Physella spelunca	*Physella spelunca* R. D. Turner & Clench, 1974	AY651243.1	AY651205.1	wySPE2
Physella ancillaria	*Physella gyrina* Say, 1821	EU038344.1	EU038391.1	paDWR3
Physella pomilia	*Physella pomilia* Conrad, 1834	EU038307.1	EU038354.1	alpom2
Stenophysa marmorata	*Stenophysa marmorata* Guilding, 1828	EU038323.1	EU038370.1	Gmar2
Aplexa elongata	*Aplexa prisca elongata* Noulet, 1854	EU038330.1	EU038377.1	miall1
Biomphalaria obstructa	*Biomphalaria obstructa* Morelet, 1849	AY651245.1	AY651207.1	scBCTL1

Table 1. List of accession numbers of sequences and their codes used for construction of Physidae phylogenetic tree in this study.

Libraries were prepared in-house using CORALL mRNA-Seq V2 (cat. # 177.96, Lexogen, Austria) using the Long Insert Size (RTL) protocol to produce strand-specific paired-end libraries. All libraries were inspected on an Agilent 2100 Bioanalyzer and quantified using a fluorometric approach (QFX Fluorometer and DeNovix dsDNA Ultra High Sensitivity Assay, cat. # DSDNA-ULTRA-2, DeNovix). Prepared libraries were multiplexed and sequenced in PE-150 mode on Illumina NovaSeq X plus by Novogene Europe, Cambridge, UK.

Bioinformatics pipeline. If not stated otherwise all analyses were conducted on the computing clusters Supek and Padobran of University Computing Center, University of Zagreb, Croatia. A summary of the workflow is shown in Fig. 2.

Read processing and *de novo* assembly. Raw reads were quality assessed using FastQC 0.12.1[83] and MultiQC 1.19[84], and trimmed for adapters, UMIs, polyX, and low-quality bases using fastp 0.23.4[85,86]. Clean reads were rechecked for quality before continuing the analyses. A reference transcriptome was generated from all samples with Trinity software 2.15.1[87]. Quality of the assembled transcriptome was assessed by Transrate 1.0.3[88] and completeness of the assembly was determined with BUSCO 5.8.0[89] using the metazoan gene set (metazoa_odb10) on the European Galaxy server[90].

Removing redundancy and mapping. Two approaches were used to reduce the redundancy of the transcriptome. The first was transcript clustering based on homology, which was done using the cd-hit-est option of the CD-HIT 4.8.1[91,92], with the homology threshold set to 90%. The second was a clustering step performed with Corset 1.07[93] with default parameters, which included filtering out transcripts with low read support (less than 10 reads per transcript). To assess whether the filtered transcriptome was over-filtered, reads were mapped back to the transcriptome using Salmon 1.10.2[94]. Read-count data produced with Corset[93] were used for principal component analysis (PCA) with the plotPCA function of DESeq 2 1.44.0[95] in R 4.4.1[96]. The quality and completeness of the final reference transcriptome was re-assessed with Transrate[88] and BUSCO[89] scores in the same way as earlier.

Protein prediction and functional annotation. Open reading frames (ORF) and candidate coding regions were predicted with TransDecoder 5.7.1 (Haas, BJ. https://github.com/TransDecoder/TransDecoder) and only the longest ORF per each transcript was retained. Predicted peptide sequences for the candidate ORFs were used for functional annotation using multiple databases. Ultra-sensitive mode of DIAMOND v2.1.8.162[97] was used to obtain annotations from UniProt's SwissProt and NCBI's NR (both databases downloaded in February 2024) using the BLASTP method. hmmscan from HMMER 3.4 (http://hmmer.org/) was used to detect protein domains from the PFAM database (downloaded in September 2023). EggNOG Mapper 5.0.2[98,99] was used for functional annotation of proteins in the Gene Ontology database and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway database. TMHMM 2.0[100–102] was used to detect proteins with transmembrane helices. Both eggNOG Mapper[98,99] and TMHMM[100–102] were run on the European Galaxy server[90].

Ethics declarations. Snails from Fogelpole Cave were collected under Illinois Nature Preserves Commission Permit #2020 INPC ROBERT WECK FOGELPOLE CAVE.

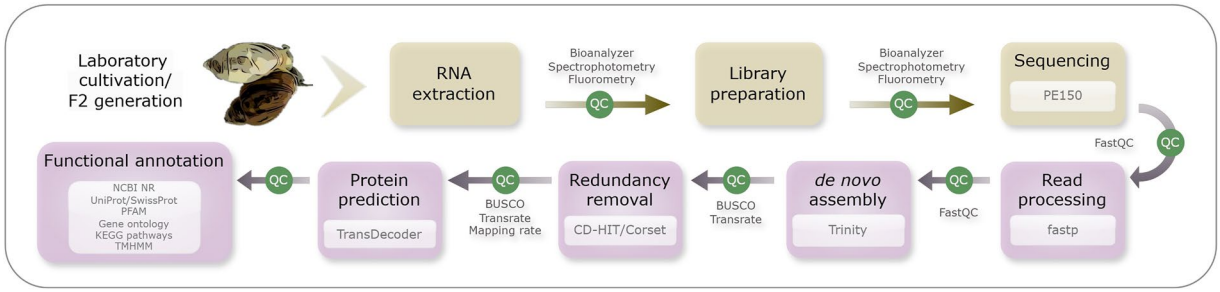


Fig. 2 Schematic representation of experimental and computational workflow.

Sample name	NCBI Accession Number	Pigmentation	Tissue	N of individuals	Raw reads (M)	Total raw reads (M)	Clean reads (M)	Total clean reads (M)	GC (%)	Retained reads (%)	Q30 (clean reads) (%)
PF1	SAMN48289483	Albino	Foot	3	14,038,298	107,606,905	13,307,255	102,404,012	40	94.79	93.70
PF2	SAMN48289484	Albino	Foot	3	13,578,197		12,719,656		39	93.68	93.41
PF3	SAMN48289485	Albino	Foot	3	15,764,232		15,026,613		40	95.32	93.86
PF4	SAMN48289486	Albino	Foot	3	9,742,326		9,170,288		39	94.13	93.12
PF5	SAMN48289487	Pigmented	Foot	3	15,823,321		15,221,557		41	96.20	93.99
PF6	SAMN48289488	Pigmented	Foot	3	12,760,731		12,240,670		40	95.92	93.81
PF7	SAMN48289489	Pigmented	Foot	3	15,887,703		15,181,685		40	95.56	93.85
PF8	SAMN48289490	Pigmented	Foot	3	10,012,097		9,536,288		40	95.25	94.07
PW1	SAMN48289491	Pigmented	Whole-body	1	19,645,136	237,747,189	18,357,549	222,828,440	41	93.45	93.79
PW2	SAMN48289492	Pigmented	Whole-body	1	25,502,398		23,846,524		41	93.51	93.43
PW3	SAMN48289493	Albino	Whole-body	1	27,569,978		26,073,006		41	94.57	93.88
PW4	SAMN48289494	Pigmented	Whole-body	1	25,215,265		23,611,845		42	93.64	93.72
PW5	SAMN48289495	Albino	Whole-body	1	21,869,923		20,543,020		42	93.93	93.75
PW6	SAMN48289496	Pigmented	Whole-body	1	23,178,484		21,787,690		40.5	94.00	93.75
PW7	SAMN48289497	Albino	Whole-body	1	22,015,528		20,628,596		41	93.70	93.28
PW8	SAMN48289498	Albino	Whole-body	1	23,454,935		21,919,446		41	93.45	93.14
PW9	SAMN48289499	Albino	Whole-body	1	23,036,248		21,616,017		41	93.83	93.36
PW10	SAMN48289500	Pigmented	Whole-body	1	26,259,294		24,444,747		41	93.09	93.21

Table 2. Summary of RNA-seq data generated in this study.

Data Records

Raw sequencing reads have been deposited in the NCBI Sequence Read Archive under study accession number SRP582749[103]. The respective BioSample accession numbers are listed in Table 2. Transcriptome with predicted proteins is deposited in the NCBI Transcriptome Shotgun Assembly (TSA) database under TSA accession number GLNG00000000[104]. Raw Trinity assembly, non-redundant transcriptome and annotations are deposited in the FigShare collection[105]. 16S and CO1 nucleotide sequences are deposited in NCBI GenBank under the accession numbers PV593642[37], PV593640[38], PV594015[39], PV593641[40], PV594016[41], and PV594017[42].

Technical Validation

Species identification. In total, previously published 16S and CO1 sequences from 16 physid species and one outgroup species[43–76], and snails from three localities in the US and Croatia[37–42], including Fogelpole Cave were used for phylogeny reconstruction. The resulting ML phylogenetic tree successfully placed snails from Fogelpole Cave into Physidae (Fig. 1b). More specifically, Fogelpole Cave snails nested within *Physella gyrina* complex, which was separated from *Physella acuta* complex. Basic tree topography is consistent with the previous phylogenetic analyses on Physidae[11,106]. Because more data is needed to resolve *Physella gyrina* complex and this is not within the scope of our study, we decided to use the name *Physella* sp. for the Fogelpole Cave population. Species *Physella gyrina* and *Physella acuta* appear several times in the tree because many species were synonymized with either of the two. Original identifiers are added to the names in the tree to distinguish between different samples.

RNA extraction, library preparation and sequencing. RNA samples with high integrity and no detected DNA with PCR were further used for library preparation and sequencing. For in-house prepared libraries, instructions from the library preparation kit were followed. Quality controls of the prepared libraries included meeting library concentration criteria and the expected library profile on the Agilent 2100 Bioanalyzer trace results.

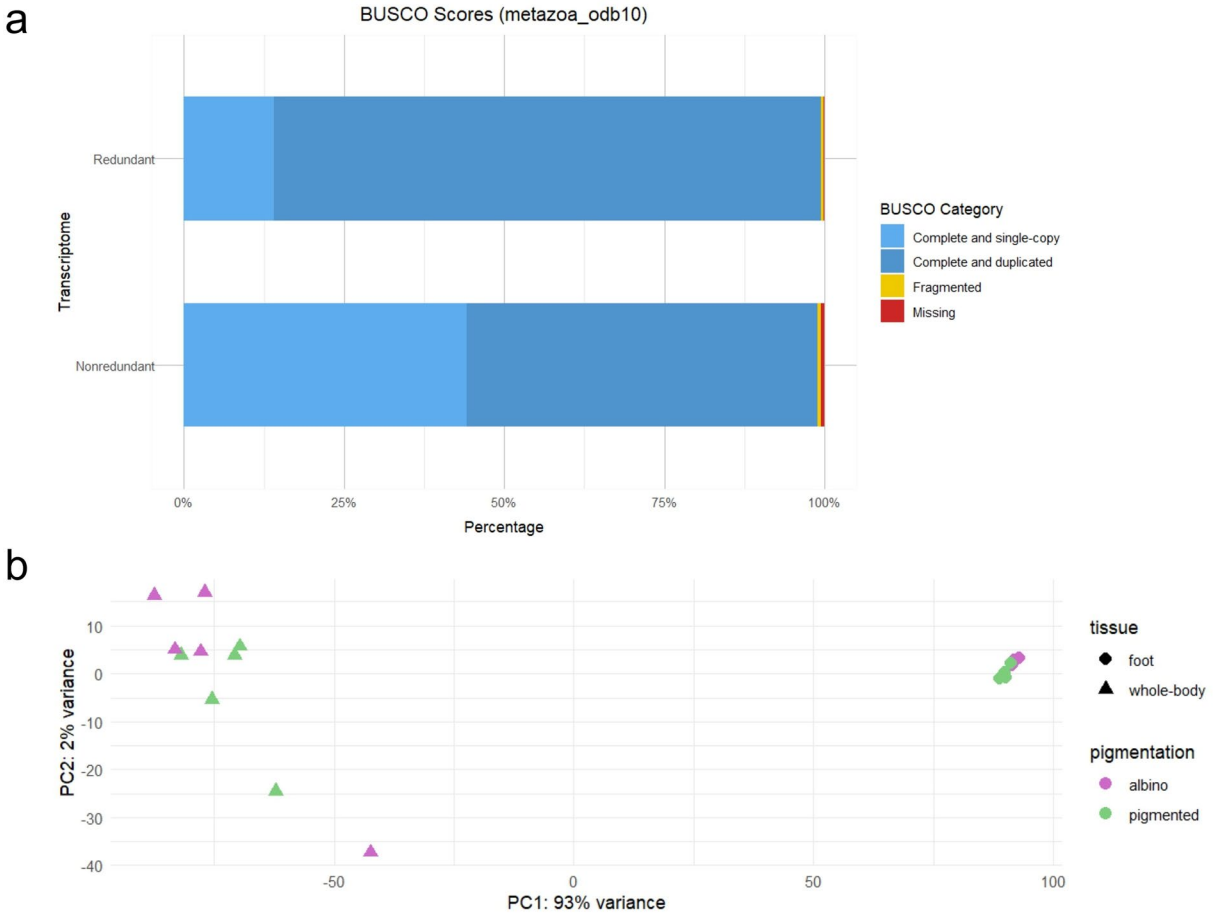


Fig. 3 Transcriptome quality assessments. (**a**) Benchmarking Universal Single-Copy Orthologs (BUSCO) scores of redundant and non-redundant transcriptomes using metazoan (n:954) gene set. (**b**) Principal component analysis (PCA) plot based on the top 500 most variable genes. The proportion of variance explained by each principal component is provided on each axis.

Read processing and *de novo* assembly. Detailed information on sequencing output and quality parameters for the samples used in the *de novo* transcriptome assembly is given in Table 2. In short, approximately 345 million paired-end reads were sequenced from 18 samples, ranging from 9 to 15 million raw reads per sample for foot tissue, and 18–26 million raw reads for whole-body samples. More than 325 million reads (94.2%) passed quality control and were used for *de novo* assembly with Trinity. The initial Trinity assembly consisted of 1,031,759 transcripts with an average mapping rate of 91.87% across all samples. The N50 value of the Trinity transcriptome was 1239, indicating good contiguity of the assembly. BUSCO scores (Fig. 3a) (metazoa; n:954) showed a high level of completeness (99.5% of genes being complete, 0.3% being fragmented and 0.2% missing).

Removing redundancy and mapping. CD-HIT reduced the number of initial transcripts to 662,467 and Corset further clustered them into 360,796 final clusters containing 411,434 transcripts in the final, non-redundant, assembly (Table 3). This is similar to the reference transcriptome of closely related *Physella acuta* which contains 325,563 transcripts and has a BUSCO score of 98%[107]. Both of these are within the range of published transcriptome assemblies for other gastropod species[108–112] and bivalves[113]. BUSCO scores for the non-redundant transcriptome show a high level of completeness (98.8% complete, 0.6% fragmented and 0.5% missing genes) (Table 3), indicating that redundancy removal did not significantly affect assembly quality in terms of completeness, while it did reduce the number of duplicated BUSCO genes from 85.4% to 54.7% (Fig. 3a). Although mapping rates to the non-redundant assembly were on average 86.77% for whole-body samples, and 87% for foot samples, which is slightly lower than the initial mapping rates, these percentages are still well above the generally accepted and suggested percentage of mapping rates for *de novo* transcriptomes of 80%[114]. PCA based on the top 500 most variable genes distinguished two sample groups based on the tissue type, as expected, with the first component accounting for 93% of the variance (Fig. 3b). Additionally, whole-body samples show higher variability compared to foot samples.

Protein prediction and functional annotation. TransDecoder predicted proteins from 60,505 transcripts in the final, non-redundant, transcriptome (Table 4). Of these, 48% were complete containing both start

	Trinity transcriptome	Non-redundant transcriptome
Assembly statistics		
N of transcripts	1,031,759	411,434
N bases	803,523,310	388,113,683
Min. length (bp)	167	179
Max. length (bp)	29,847	29,847
Mean length (bp)	779	943
N of transcripts of 1 K (bp)	204,829	109,818
N of transcripts over 10 K (bp)	1,112	555
N90	310	419
N70	633	776
N50	1,239	1,324
N30	2,318	2,333
N10	4,991	4,994
GC (%)	38	38
Transcriptome completeness		
Complete (Single + Duplicated)	99.50%	98.80%
Single	14.00%	44.10%
Duplicated	85.40%	54.70%
Fragmented	0.30%	0.60%
Missing	0.20%	0.50%
Mapping rates		
Overall mapping rate: whole-body samples	91.87%	86.77%
Overall mapping rate: foot samples		87%

Table 3. Basic statistics, BUSCO scores and mapping rates of Trinity transcriptome and non-redundant transcriptome.

	Non-redundant transcriptome
Protein prediction	
Number of predicted proteins	60,505
Number of complete predicted proteins	29,051
Number of internal predicted proteins	10,524
Number of 5′ partial predicted proteins	12,218
Number of 3′ partial predicted proteins	8,712
Min. length (aa)	86
Max. length (aa)	9,438
Mean length (aa)	317.7
Median length (aa)	199
Functional annotation	
NCBI NR	41,846
UniProt/SwissProt	26,858
PFAM	47,160
Gene Ontology	20,610
KEGG	21,181
TMHMM	11,955

Table 4. Protein prediction statistics and number of proteins functionally annotated in each database used in this study.

and stop codons, 17.39% were internal, lacking both 5′ and 3′ ends, 20.19% were 5′ partial and 14.4% were 3′ partial. Predicted protein length ranged from 86 to 9,438 aa (amino acids), with a mean length of 317.7 aa. Among predicted proteins, 26,858 were annotated in the UniProt/SwissProt database, and 41,846 were annotated in NCBI's NR database. EggNOG Mapper detected 31,792 proteins, of which 20,610 were annotated in Gene Ontology and 21,181 were assigned KEGG orthologues. A HMMER search with a cutoff E-value of 1e-5 detected 47,160 protein domains in the PFAM database. TMHMM identified 11,955 possible transmembrane proteins, defined as those containing at least one transmembrane helix of at least 18 amino acids.

Data availability

All data reported in this study are publicly available in online repositories. Raw sequencing reads have been deposited in the NCBI Sequence Read Archive under study accession number SRP582749103. The final curated transcriptome assembly and predicted proteins have been deposited in the NCBI Transcriptome Shotgun Assembly database under TSA accession number GLNG00000000104. The raw Trinity assembly, non-redundant transcriptome assembly, and functional annotations are available in the FigShare collection[105]. Mitochondrial 16S and CO1 nucleotide sequences have been deposited in NCBI GenBank under accession numbers PV59364237, PV59364038, PV59401539, PV59364140, PV59401641, and PV59401742.

Code availability

Software used, versions and parameters differing from default settings are listed below.

1. Geneious Prime software 2024.0.7.
2. MUSCLE 5.1[77].
3. IQtree 2.2.6[78–80].
4. iTOL (online version v7): https://itol.embl.de/, March 2025.
5. FastQC 0.12.1[83].
6. MultiQC 1.19[84].
7. fastp 0.23.4[85,86]: -x -g -5 -3 -n 5 -y -Y 20–detect_adapter_for_pe -p -U–umi_loc read1–umi_len 12.
8. Trinity 2.15.1[87]:–seqType fq–max_memory 240 G–CPU 64–SS_lib_type FR.
9. BUSCO v5.8.0[89]; Galaxy Version 5.8.0 + galaxy1:–lineage_dataset 'metazoa_odb10'.
10. Transrate 1.0.3[88].
11. CD-HIT 4.8.1[91,92]: -c 0.9.
12. Corset 1.07[93].
13. Salmon 1.10.2[94].
14. DESeq. 2::plotPCA 1.44.0[95] & R 4.4.1[96].
15. TransDecoder 5.7.1 (Haas, BJ. https://github.com/TransDecoder/TransDecoder):–single_best_only.
16. DIAMOND v2.1.8.162[97]: blastp–ultra-sensitive.
17. HMMER 3.4 (http://hmmer.org/): hmmscan -E 1e-5–tblout.
18. eggNOG Mapper 2.1.8[98,99]/eggNOG Database 5.0.2; Galaxy Version 2.1.8 + galaxy4[90]:–sensmode very-sensitive.
19. TMHMM 2.0; Galaxy Version 0.0.17[100,101,115].

Received: 12 May 2025; Accepted: 13 June 2026;
Published online: 27 June 2026

References

1. Culver, D. C. & Pipan, T. *The Biology of Caves and Other Subterranean Habitats*, https://doi.org/10.1093/oso/9780198820765.001.0001 (Oxford University Press, 2019).
2. White, W. B., Culver, D. C. & Pipan, T. *Encyclopedia of Caves*. https://doi.org/10.1016/C2017-0-01162-X (Elsevier, 2019).
3. Bilandžija, H., Morton, B., Podnar, M. & Ćetković, H. Evolutionary history of relict Congeria (Bivalvia: Dreissenidae): unearthing the subterranean biodiversity of the Dinaric Karst. *Frontiers in Zoology* **10**, 5 (2013).
4. Simone, L. R. L. & Ferreira, R. L. Eupera troglobia sp. nov.: the first troglobitic bivalve from the Americas (Mollusca, Bivalvia, Sphaeriidae). *Subterranean Biology* **42**, 165–184 (2022).
5. Prié, V. Molluscs. in *Encyclopedia of Caves* (eds White, W. B., Culver, D. C. & Pipan, T.) 725–731 (Elsevier, 2019).
6. Marmonier, P. *et al*. Groundwater biodiversity and constraints to biological distribution. in *Groundwater Ecology and Evolution* 113–140, https://doi.org/10.1016/B978-0-12-819119-4.00003-2 (Elsevier, 2023).
7. Gladstone, N. S. *et al*. Subterranean freshwater gastropod biodiversity and conservation in the United States and Mexico. *Conservation Biology* **36** (2022).
8. Mammola, S. *et al*. Scientists' Warning on the Conservation of Subterranean Ecosystems. *BioScience* **69**, 641–650 (2019).
9. Lewis, J. J., Moss, P., Tecic, D. & Nelson, M. E. A conservation focused inventory of subterranean invertebrates of the southwest Illinois Karst. *Journal of Cave and Karst Studies of the National Speleological Society* **65**, 9–21 (2003).
10. Taylor, D. W. Introduction to Physidae (Gastropoda: Hygrophila); biogeography, classification, morphology. *Rev. Biol. Trop.* **51**, 1–287 (2003).
11. Wethington, A. R. & Lydeard, C. A molecular phylogeny of Physidae (Gastropoda: Basommatophora) based on mitochondrial DNA sequences. *Journal of Molluscan Studies* **73**, 241–257 (2007).
12. Weck, R. G. & Taylor, S. J. Life history studies of a cave-dwelling population of Physa snails (Gastropoda: Basommatophora: Physidae) from southwestern Illinois. *Speleobiology Notes* **8**, 1–9 (2016).
13. Lukic, M. *et al*. A practical guide for the husbandry of cave and surface invertebrates as the first step in establishing new model organisms. *PLoS ONE* **19** (2024).
14. Warren, W. C. *et al*. A chromosome-level genome of Astyanax mexicanus surface fish for comparing population-specific genetic differences contributing to trait evolution. *Nature Communications* **12**, 1447 (2021).
15. Grant, E. H. C. *et al*. Speciation with gene flow in a narrow endemic West Virginia cave salamander (Gyrinophilus subterraneus). *Conservation Genetics* **23**, 727–744 (2022).
16. Lomheim, H. J. *et al*. Transcriptomic analysis of cave, surface, and hybrid samples of the isopod Asellus aquaticus and identification of chromosomal location of candidate genes for cave phenotype evolution. *EvoDevo* **14**, 9 (2023).
17. Jović, L., Bedek, J., Malard, F. & Bilandžija, H. De novo transcriptomes of cave and surface isopod crustaceans: insights from 11 species across three suborders. *Scientific Data* **11**, 595 (2024).
18. Stern, D. B. & Crandall, K. A. Phototransduction Gene Expression and Evolution in Cave and Surface Crayfishes. *Integrative and Comparative Biology* **58**, 398–410 (2018).
19. Balart-García, P. *et al*. Parallel and convergent genomic changes underlie independent subterranean colonization across beetles. *Nature Communications* **14**, 3842 (2023).
20. Carlini, D. B. & Fong, D. W. The transcriptomes of cave and surface populations of Gammarus minus (Crustacea: Amphipoda) provide evidence for positive selection on cave downregulated transcripts. *PLOS ONE* **12**, e0186173 (2017).

21. Wang, K. *et al*. Eyeless cave-dwelling *Leptonetela* spiders still rely on light. *Science Advances* **9** (2023).
22. Kostanjšek, R. *et al*. Toward the massive genome of *Proteus anguinus* —illuminating longevity, regeneration, convergent evolution, and metabolic disorders. *Annals of the New York Academy of Sciences* **1507**, 5–11 (2022).
23. Yang, J. *et al*. The Sinocyclocheilus cavefish genome provides insights into cave adaptation. *BMC Biology* **14**, 1 (2016).
24. Bilandžija, H., Ćetković, H. & Jeffery, W. R. Evolution of albinism in cave planthoppers by a convergent defect in the first step of melanin biosynthesis. *Evolution & Development* **14**, 196–203 (2012).
25. Protas, M. E., Trontelj, P. & Patel, N. H. Genetic basis of eye and pigment loss in the cave crustacean, *Asellus aquaticus*. *Proceedings of the National Academy of Sciences* **108**, 5702–5707 (2011).
26. Protas, M. E. *et al*. Genetic analysis of cavefish reveals molecular convergence in the evolution of albinism. *Nature Genetics* **38**, 107–111 (2006).
27. Mojaddidi, H., Fernandez, F. E., Erickson, P. A. & Protas, M. E. Embryonic origin and genetic basis of cave associated phenotypes in the isopod crustacean Asellus aquaticus. *Sci Rep* **8**, 16589 (2018).
28. Bakovic, V. *et al*. The genomics of phenotypically differentiated Asellus aquaticus cave, surface stream and lake ecotypes. *Molecular Ecology* **30**, 3530–3547 (2021).
29. Lomheim, H. J. *et al*. Comparative Embryology and Transcriptomics of Asellus infernus, an Isopod Crustacean From Sulfidic Groundwater. *Evol Dev* **27**, e70014 (2025).
30. Rodas, L. R. *et al*. Standing genetic variation as a potential mechanism of novel cave phenotype evolution in the freshwater isopod, Asellus aquaticus. *Evol Dev* **25**, 137–152 (2023).
31. Ducrest, A. L., Keller, L. & Roulin, A. Pleiotropy in the melanocortin system, coloration and behavioural syndromes. *Trends in Ecology and Evolution* **23**, 502–510 (2008).
32. Bilandžija, H., Laslo, M., Porter, M. L. & Fong, D. W. Melanization in response to wounding is ancestral in arthropods and conserved in albino cave species. *Scientific Reports* **7**, 17148 (2017).
33. Bilandžija, H., Ma, L., Parkhurst, A. & Jeffery, W. R. A Potential Benefit of Albinism in Astyanax Cavefish: Downregulation of the oca2 Gene Increases Tyrosine and Catecholamine Levels as an Alternative to Melanin Synthesis. *PLoS ONE* **8**, e80823 (2013).
34. Bilandžija, H., Abraham, L., Ma, L., Renner, K. J. & Jeffery, W. R. Behavioural changes controlled by catecholaminergic systems explain recurrent loss of pigmentation in cavefish. *Proceedings of the Royal Society B: Biological Sciences* **285** (2018).
35. O'Gorman, M. *et al*. Pleiotropic function of the oca2 gene underlies the evolution of sleep loss and albinism in cavefish. *Current Biology* **31**, 3694–3701.e4 (2021).
36. Dillon, R. T. & Wethington, A. R. The inheritance of albinism in a freshwater snail, Physa heterostropha. *The Journal of heredity* **83**, 208–10 (1992).
37. Grgic, M., Bilandzija, H. & Weck, R. G. Physella sp. isolate 1_Lo cytochrome c oxidase subunit I (COX1) gene, partial cds; mitochondrial. *GenBank* https://identifiers.org/ncbi/insdc:PV593642 (2025).
38. Grgic, M., Bilandzija, H. & Weck, R. G. Physella sp. isolate 1_Fo cytochrome c oxidase subunit I (COX1) gene, partial cds; mitochondrial. *GenBank* https://identifiers.org/ncbi/insdc:PV593640 (2025).
39. Grgic, M., Bilandzija, H. & Weck, R. G. Physella sp. isolate 1_Fo large subunit ribosomal RNA gene, partial sequence; mitochondrial. *GenBank* https://identifiers.org/ncbi/insdc:PV594015 (2025).
40. Grgic, M., Bilandzija, H. & Weck, R. G. Physella sp. isolate 1_Jarun cytochrome c oxidase subunit I (COX1) gene, partial cds; mitochondrial. *GenBank* https://identifiers.org/ncbi/insdc:PV593641 (2025).
41. Grgic, M., Bilandzija, H. & Weck, R. G. Physella sp. isolate 1_Jarun large subunit ribosomal RNA gene, partial sequence; mitochondrial. *GenBank* https://identifiers.org/ncbi/insdc:PV594016 (2025).
42. Grgic, M., Bilandzija, H. & Weck, R. G. Physella sp. isolate 1_Lo large subunit ribosomal RNA gene, partial sequence; mitochondrial. *GenBank* https://identifiers.org/ncbi/insdc:PV594017 (2025).
43. Remigio, E. A., Lepitzki, D. A. W., Lee, J. S. & Hebert, P. D. N. Physella gyrina Clear Spring population 16S ribosomal RNA gene, partial sequence; mitochondrial gene for mitochondrial product. *GenBank* https://identifiers.org/ncbi/insdc:AF346755 (2001).
44. Remigio, E. A., Lepitzki, D. A. W., Lee, J. S. & Hebert, P. D. N. Physella johnsoni Upper Spring population 16S ribosomal RNA gene, partial sequence; mitochondrial gene for mitochondrial product. *GenBank* https://identifiers.org/ncbi/insdc:AF346750 (2001).
45. Remigio, E. A., Lepitzki, D. A. W., Lee, J. S. & Hebert, P. D. N. Physella wrighti 16S ribosomal RNA gene, partial sequence; mitochondrial gene for mitochondrial product. *GenBank* https://identifiers.org/ncbi/insdc:AF346757 (2001).
46. Remigio, E. A., Lepitzki, D. A. W., Lee, J. S. & Hebert, P. D. N. Physella gyrina isolate PgyCLCOI cytochrome c oxidase subunit I (COI) gene, partial cds; mitochondrial gene for mitochondrial product. *GenBank* https://identifiers.org/ncbi/insdc:AF346743 (2016).
47. Remigio, E. A., Lepitzki, D. A. W., Lee, J. S. & Hebert, P. D. N. Physella johnsoni isolate PjoUPCOI cytochrome c oxidase subunit I (COI) gene, partial cds; mitochondrial gene for mitochondrial product. *GenBank* https://identifiers.org/ncbi/insdc:AF346738 (2016).
48. Remigio, E. A., Lepitzki, D. A. W., Lee, J. S. & Hebert, P. D. N. Physella wrighti cytochrome c oxidase subunit I (COI) gene, partial cds; mitochondrial gene for mitochondrial product. *GenBank* https://identifiers.org/ncbi/insdc:AF346745 (2016).
49. Wethington, A. R. & Guralnick, R. Biomphalaria obstructa isolate scBCTL1 16S ribosomal RNA gene, partial sequence; mitochondrial. *GenBank* https://identifiers.org/ncbi/insdc:AY651245 (2004).
50. Wethington, A. R. & Guralnick, R. Petrophysa zionis isolate utznp2 16S ribosomal RNA gene, partial sequence; mitochondrial. *GenBank* https://identifiers.org/ncbi/insdc:AY651236 (2004).
51. Wethington, A. R. & Guralnick, R. Physa acuta isolate F7 16S ribosomal RNA gene, partial sequence; mitochondrial. *GenBank* https://identifiers.org/ncbi/insdc:AY651224 (2004).
52. Wethington, A. R. & Guralnick, R. Physa acuta isolate wybhr2 16S ribosomal RNA gene, partial sequence; mitochondrial. *GenBank* https://identifiers.org/ncbi/insdc:AY651241 (2004).
53. Wethington, A. R. & Guralnick, R. Physa fontinalis isolate Nnf1 16S ribosomal RNA gene, partial sequence; mitochondrial. *GenBank* https://identifiers.org/ncbi/insdc:AY651227 (2004).
54. Wethington, A. R. & Guralnick, R. Physella anatina isolate co43888 16S ribosomal RNA gene, partial sequence; mitochondrial. *GenBank* https://identifiers.org/ncbi/insdc:AY651214 (2004).
55. Wethington, A. R. & Guralnick, R. Physella hendersoni isolate scysr3 16S ribosomal RNA gene, partial sequence; mitochondrial. *GenBank* https://identifiers.org/ncbi/insdc:AY651234 (2004).
56. Wethington, A. R. & Guralnick, R. Physella spelunca isolate wySPE2 16S ribosomal RNA gene, partial sequence; mitochondrial. *GenBank* https://identifiers.org/ncbi/insdc:AY651243 (2004).
57. Wethington, A. R. & Guralnick, R. Physella virgata isolate argrv1 16S ribosomal RNA gene, partial sequence; mitochondrial. *GenBank* https://identifiers.org/ncbi/insdc:AY651209 (2004).
58. Wethington, A. R. & Guralnick, R. Biomphalaria obstructa isolate scBCTL1 cytochrome oxidase subunit I (CO1) gene, partial cds; mitochondrial. *GenBank* https://identifiers.org/ncbi/insdc:AY651207 (2016).
59. Wethington, A. R. & Guralnick, R. Petrophysa zionis isolate utznp2 cytochrome oxidase subunit I (CO1) gene, partial cds; mitochondrial. *GenBank* https://identifiers.org/ncbi/insdc:AY651198 (2016).
60. Wethington, A. R. & Guralnick, R. Physa acuta isolate F7 cytochrome oxidase subunit I (CO1) gene, partial cds; mitochondrial. *GenBank* https://identifiers.org/ncbi/insdc:AY651186 (2016).

61. Wethington, A. R. & Guralnick, R. Physa acuta isolate wybhr2 cytochrome oxidase subunit I (CO1) gene, partial cds; mitochondrial. *GenBank* https://identifiers.org/ncbi/insdc:AY651203 (2016).
62. Wethington, A. R. & Guralnick, R. Physa fontinalis isolate Nnf1 cytochrome oxidase subunit I (CO1) gene, partial cds; mitochondrial. *GenBank* https://identifiers.org/ncbi/insdc:AY651189 (2016).
63. Wethington, A. R. & Guralnick, R. Physella anatina isolate co43888 cytochrome oxidase subunit I (CO1) gene, partial cds; mitochondrial. *GenBank* https://identifiers.org/ncbi/insdc:AY651175 (2016).
64. Wethington, A. R. & Guralnick, R. Physella hendersoni isolate scysr3 cytochrome oxidase subunit I (CO1) gene, partial cds; mitochondrial. *GenBank* https://identifiers.org/ncbi/insdc:AY651196 (2016).
65. Wethington, A. R. & Guralnick, R. Physella spelunca isolate wySPE2 cytochrome oxidase subunit I (CO1) gene, partial cds; mitochondrial. *GenBank* https://identifiers.org/ncbi/insdc:AY651205 (2016).
66. Wethington, A. R. & Guralnick, R. Physella virgata isolate argrv1 cytochrome oxidase subunit I (CO1) gene, partial cds; mitochondrial. *GenBank* https://identifiers.org/ncbi/insdc:AY651170 (2016).
67. Wethington, A. R. & Lydeard, C. Aplexa elongata isolate miall1 16S ribosomal RNA gene, partial sequence; mitochondrial. *GenBank* https://identifiers.org/ncbi/insdc:EU038330 (2017).
68. Wethington, A. R. & Lydeard, C. Aplexa elongata isolate miall1 cytochrome oxidase subunit I (CO1) gene, partial cds; mitochondrial. *GenBank* https://identifiers.org/ncbi/insdc:EU038377 (2017).
69. Wethington, A. R. & Lydeard, C. Physa vernalis isolate maamb2 16S ribosomal RNA gene, partial sequence; mitochondrial. *GenBank* https://identifiers.org/ncbi/insdc:EU038329 (2017).
70. Wethington, A. R. & Lydeard, C. Physa vernalis isolate maamb2 cytochrome oxidase subunit I (CO1) gene, partial cds; mitochondrial. *GenBank* https://identifiers.org/ncbi/insdc:EU038376 (2017).
71. Wethington, A. R. & Lydeard, C. Physella ancillaria isolate paDWR3 16S ribosomal RNA gene, partial sequence; mitochondrial. *GenBank* https://identifiers.org/ncbi/insdc:EU038344 (2017).
72. Wethington, A. R. & Lydeard, C. Physella ancillaria isolate paDWR3 cytochrome oxidase subunit I (CO1) gene, partial cds; mitochondrial. *GenBank* https://identifiers.org/ncbi/insdc:EU038391 (2017).
73. Wethington, A. R. & Lydeard, C. Physella pomilia isolate alpom2 16S ribosomal RNA gene, partial sequence; mitochondrial. *GenBank* https://identifiers.org/ncbi/insdc:EU038307 (2017).
74. Wethington, A. R. & Lydeard, C. Physella pomilia isolate alpom2 cytochrome oxidase subunit I (CO1) gene, partial cds; mitochondrial. *GenBank* https://identifiers.org/ncbi/insdc:EU038354 (2017).
75. Wethington, A. R. & Lydeard, C. Stenophysa marmorata isolate Gmar2 16S ribosomal RNA gene, partial sequence; mitochondrial. *GenBank* https://identifiers.org/ncbi/insdc:EU038323 (2017).
76. Wethington, A. R. & Lydeard, C. Stenophysa marmorata isolate Gmar2 cytochrome oxidase subunit I (CO1) gene, partial cds; mitochondrial. *GenBank* https://identifiers.org/ncbi/insdc:EU038370 (2017).
77. Edgar, R. C. High-accuracy alignment ensembles enable unbiased assessments of sequence homology and phylogeny, https://doi.org/10.1101/2021.06.20.449169 (2021).
78. Minh, B. Q. *et al.* IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the Genomic Era. *Molecular Biology and Evolution* **37**, 1530–1534 (2020).
79. Kalyaanamoorthy, S., Minh, B. Q., Wong, T. K. F., von Haeseler, A. & Jermiin, L. S. ModelFinder: fast model selection for accurate phylogenetic estimates. *Nature Methods* **14**, 587–589 (2017).
80. Hoang, D. T., Chernomor, O., von Haeseler, A., Minh, B. Q. & Vinh, L. S. UFBoot2: Improving the Ultrafast Bootstrap Approximation. *Molecular Biology and Evolution* **35**, 518–522 (2018).
81. George Ang Te. A systematic study of the family Physidae: Basommatophora, Pulmonata. (University of Michigan, 1978).
82. Burch, J. B. *North American Freshwater Snails: Introduction, Systematics, Nomenclature, Identification, Morphology, Habitats, Distribution*. **vol. 2** (Walkerana, Transactions of the POETS Society, 1988).
83. Andrews, S. FastQC: a quality control tool for high throughput sequence data. https://www.bioinformatics.babraham.ac.uk/projects/fastqc/ (2010).
84. Ewels, P., Magnusson, M., Lundin, S. & Käller, M. MultiQC: summarize analysis results for multiple tools and samples in a single report. *Bioinformatics* **32**, 3047–3048 (2016).
85. Chen, S., Zhou, Y., Chen, Y. & Gu, J. fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* **34**, i884–i890 (2018).
86. Chen, S. Ultrafast one-pass FASTQ data preprocessing, quality control, and deduplication using fastp. *iMeta* **2** (2023).
87. Grabherr, M. G. *et al.* Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nature Biotechnology* **29**, 644–652 (2011).
88. Smith-Unna, R., Boursnell, C., Patro, R., Hibberd, J. M. & Kelly, S. TransRate: reference-free quality assessment of de novo transcriptome assemblies. *Genome research* **26**, 1134–44 (2016).
89. Simão, F. A., Waterhouse, R. M., Ioannidis, P., Kriventseva, E. V. & Zdobnov, E. M. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* **31**, 3210–3212 (2015).
90. Abueg, L. A. L. *et al.* The Galaxy platform for accessible, reproducible, and collaborative data analyses: 2024 update. *Nucleic Acids Research* **52**, W83–W94 (2024).
91. Fu, L., Niu, B., Zhu, Z., Wu, S. & Li, W. CD-HIT: accelerated for clustering the next-generation sequencing data. *Bioinformatics* **28**, 3150–3152 (2012).
92. Li, W. & Godzik, A. Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics* **22**, 1658–1659 (2006).
93. Davidson, N. M. & Oshlack, A. Corset: Enabling differential gene expression analysis for de novo assembled transcriptomes. *Genome Biology* **15** (2014).
94. Patro, R., Duggal, G., Love, M. I., Irizarry, R. A. & Kingsford, C. Salmon provides fast and bias-aware quantification of transcript expression. *Nature Methods* **14**, 417–419 (2017).
95. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq. 2. *Genome Biology* **15**, 550 (2014).
96. R Core Team. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/. (2024).
97. Buchfink, B., Reuter, K. & Drost, H. G. Sensitive protein alignments at tree-of-life scale using DIAMOND. *Nature Methods* **18**, 366–368 (2021).
98. Cantalapiedra, C. P., Hernández-Plaza, A., Letunic, I., Bork, P. & Huerta-Cepas, J. eggNOG-mapper v2: Functional Annotation, Orthology Assignments, and Domain Prediction at the Metagenomic Scale. *Molecular Biology and Evolution* **38**, 5825–5829 (2021).
99. Huerta-Cepas, J. *et al.* eggNOG 5.0: a hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses. *Nucleic Acids Research* **47**, D309–D314 (2019).
100. Krogh, A., Larsson, B., von Heijne, G. & Sonnhammer, E. L. L. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *Journal of Molecular Biology* **305**, 567–580 (2001).
101. Sonnhammer, E. L., von Heijne, G. & Krogh, A. A hidden Markov model for predicting transmembrane helices in protein sequences. *Proceedings. International Conference on Intelligent Systems for Molecular Biology* **6**, 175–82 (1998).
102. Cock, P. J. A., Grüning, B. A., Paszkiewicz, K. & Pritchard, L. Galaxy tools and workflows for sequence analysis with applications in molecular plant pathology. *PeerJ* **1**, e167 (2013).

103. *NCBI Sequence Read Archive* https://identifiers.org/ncbi/insdc.sra:SRP582749 (2025).
104. Grgic, M., Weck, R. G., Kerestes, G., Kalogjera, L. & Bilandzija, H. TSA: Physella sp., transcriptome shotgun assembly. *GenBank*. https://identifiers.org/ncbi/insdc:GLNG00000000 (2026).
105. Grgić, M., Weck, R. G., Keresteš, G., Kalogjera, L. & Bilandžija, H. De novo transcriptome assembly of a pigmentation-variable population of cave snail Physella sp. https://doi.org/10.6084/m9.figshare.28956188 (2025).
106. Young, M. K., Smith, R., Pilgrim, K. L. & Schwartz, M. K. Molecular species delimitation refines the taxonomy of native and nonnative physinine snails in North America. *Scientific Reports* **11**, 21739 (2021).
107. Schultz, J. H., Bu, L., Kamel, B. & Adema, C. M. RNA-Seq: the Early Response of the Snail Physella acuta to the Digenetic Trematode Echinostoma paraensei. *Journal of Parasitology* **106**, 490 (2020).
108. Park, J. E. *et al*. Transcriptome sequencing of the endangered land snail Karaftohelix adamsi from the Island Ulleung: De novo assembly, annotation, valuation of fitness genes and SSR markers. *Genes & Genomics* **46**, 851–870 (2024).
109. Chetruengchai, W. *et al*. De novo genome assembly and transcriptome sequencing in foot and mantle tissues of Megaustenia siamensis reveals components of adhesive substances. *Scientific Reports* **14**, 13756 (2024).
110. Santos, C. A., Sonoda, G. G., Cortez, T., Coutinho, L. L. & Andrade, S. C. S. Transcriptome Expression of Biomineralization Genes in *Littoraria flava* Gastropod in Brazilian Rocky Shore Reveals Evidence of Local Adaptation. *Genome Biology and Evolution* **13**, (2021).
111. Yokomizo, T. & Takahashi, Y. Plasticity of circadian and circatidal rhythms in activity and transcriptomic dynamics in a freshwater snail. *Heredity* **132**, 267–274 (2024).
112. Klein, A. *et al*. Development and Interrogation of a Transcriptomic Resource for the Giant Triton Snail (Charonia tritonis). *Marine Biotechnology* **23**, 501–515 (2021).
113. Gomes-dos-Santos, A. *et al*. The gill transcriptome of threatened European freshwater mussels. *Scientific Data* **9**, 494 (2022).
114. Raghavan, V., Kraft, L., Mesny, F. & Rigerte, L. A simple guide to *de novo* transcriptome assembly and annotation. *Briefings in Bioinformatics* **23** (2022).
115. Kahsay, R. Y., Gao, G. & Liao, L. An improved hidden Markov model for transmembrane protein detection and topology prediction and its applications to complete genomes. *Bioinformatics* **21**, 1853–1858 (2005).

Acknowledgements

We wish to thank Tin Rožman, Marko Lukić, Nikolina Kuharić, and Iva Čupić for animal care and Tin Rožman for photos of animals. This research was financed by the Tenure Track Pilot Programme of the Croatian Science Foundation and the Ecole Polytechnique Fédérale de Lausanne and the Project TTP-2018-07-9675 Evolution in the dark, with funds of the Croatian-Swiss Research Programme, and the "Young Researchers' Career Development Project", DOK-2020-01, of the Croatian Science Foundation. The analyses were performed using the Advanced computing service provided by the University of Zagreb University Computing Centre - SRCE. The authors acknowledge the support of the Freiburg Galaxy Team: Person X and Björn Grüning, Bioinformatics, University of Freiburg (Germany) funded by the German Federal Ministry of Education and Research BMBF grant 031 A538A de.NBI-RBC and the Ministry of Science, Research and the Arts Baden-Württemberg (MWK) within the framework of LIBIS/de.NBI Freiburg.

Author contributions

M.G. design, wet lab, bioinformatic analyses, visualizations, original draft, reviewed and approved the final manuscript. R.G.W. field sampling, wet lab (hybrid crosses), reviewed and approved the final manuscript. G.K. wet lab (phylogeny), reviewed and approved the final manuscript. L.K. wet lab (phylogeny), reviewed and approved the final manuscript. H.B. conceptualization and design, supervision, funding, original draft, reviewed and approved the final manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to H.B.

Reprints and permissions information is available at www.nature.com/reprints.

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

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

© The Author(s) 2026