frontiers in Ecology and Evolution

TYPE Original Research
PUBLISHED 07 September 2026
DOI 10.3389/fevo.2026.1933916

OPEN ACCESS

EDITED BY
Peizhen Ma,
Chinese Academy of Fishery Sciences (CAFS), China

REVIEWED BY
Geraldine Dayrit,
University of the Philippines Manila College of Public Health, Philippines
João M. Moreno,
Universidade de Lisboa, Portugal

*CORRESPONDENCE
Mario Juan Gordillo-Pérez
mariojg755@gmail.com

RECEIVED 10 July 2026
REVISED 19 August 2026
ACCEPTED 21 August 2026
PUBLISHED 07 September 2026

CITATION
Gordillo-Pérez MJ, Reyes-Tur B, Chen Z, Sigwart J, Greve C, Heleven M, Smeets K and Beenaerts N (2026) Genome-wide identification and evolutionary characterization of predicted HSP70 proteins in the Cuban painted snail *Polymita picta*. *Front. Ecol. Evol.* 14:1933916. doi: 10.3389/fevo.2026.1933916

Genome-wide identification and evolutionary characterization of predicted HSP70 proteins in the Cuban painted snail *Polymita picta*

Mario Juan Gordillo-Pérez[1]*, Bernardo Reyes-Tur[2], Zeyuan Chen[3], Julia Sigwart[3], Carola Greve[3], Martijn Heleven[1], Karen Smeets[1] and Natalie Beenaerts[1]

[1]Centre for Environmental Sciences, Hasselt University, Hasselt, Belgium, [2]Department of Biology and Geography, University of Oriente, Santiago de Cuba, Cuba, [3]Senckenberg Research Institute and Museum of Natural History Frankfurt, Frankfurt, Germany

Climate change is expected to affect terrestrial ectotherms worldwide, including land snails, which are often characterized by limited dispersal capacity and high sensitivity to environmental fluctuations. Heat shock proteins of the 70 kDa family (HSP70) are highly conserved molecular chaperones involved in protein folding, cellular homeostasis and responses to environmental stress. Despite their biological importance, the diversity and evolution of HSP70 genes remain poorly explored in terrestrial gastropods. Here, we performed the first genome-based characterization of HSP70 genes in the Cuban painted snail *Polymita picta*, an endemic species of high conservation concern. Using sequence homology searches, conserved domain screening and comparative analyses, we identified ten non-redundant candidate HSP70 loci distributed across five genomic scaffolds, yielding ten predicted HSP70 protein sequences. The predicted proteins exhibited the characteristic physicochemical properties, conserved domains and motif architecture expected for cytosolic HSP70 proteins. Phylogenetic analyses grouped all *P. picta* sequences with cytosolic HSP70 homologues from other mollusks, whereas HSC70, HSP75 and HSP78 proteins formed distinct lineages. Pairwise sequence comparisons revealed high levels of amino acid conservation among paralogs, while Ka/Ks analyses indicated strong purifying selection across all sequence pairs. These results provide the first overview of the HSP70 repertoire currently identifiable in the *P. picta* genome and establish a molecular framework for future studies on stress physiology, gene expression and the evolutionary diversification of heat shock proteins in terrestrial mollusks.

KEYWORDS
conservation genomics, gastropods, gene family evolution, heat shock proteins, molecular chaperone

1 Introduction

Ectothermic organisms such as terrestrial mollusks are highly susceptible to fluctuations in temperature and humidity due to their limited dispersal capacity and physiological dependence on environmental conditions (Staikou et al., 2024; Noble et al., 2025). Habitat degradation, drought, and increasing temperatures constitute major threats affecting survival, reproduction, and longevity in terrestrial gastropods (Nicolai and Ansart, 2017). To cope with environmental stress, these organisms have evolved a wide range of physiological, cellular, and molecular responses, including aestivation, changes in tissue structure, mucus production, pigmentation, and the induction of molecular chaperones (Scheil et al., 2012; Schweizer et al., 2019; Tull et al., 2020).

Despite numerous studies addressing thermal effects on morphology and physiology in land snails, investigations focused on molecular mechanisms remain comparatively scarce (Schweizer et al., 2019). Among the molecular systems involved in cellular stress responses, the heat shock protein 70 (HSP70) family represents one of the most extensively studied groups of molecular chaperones. HSP70 proteins are ubiquitous and highly conserved across all domains of life (Singh et al., 2025). Based on molecular weight and functional characteristics, heat shock proteins are classified into several families, including small HSPs, HSP40, HSP70, HSP90, and HSP110 (Singh et al., 2025).

Within the HSP70 superfamily, several functionally distinct proteins have been described, including inducible HSP70s (also named HSPA1A), constitutively expressed HSC70s (also referred to as HSPA8/cognate heat shock protein), the predominantly constitutive mitochondrial HSP75 (HSPA9/mortalin/GRP75), and the endoplasmic reticulum HSP78 (HSPA5/BiP/GRP78), which performs essential housekeeping chaperone functions but can also be induced under endoplasmic reticulum stress; all of these proteins are characterized by a highly conserved N-terminal ATPase domain (Wang et al., 2021; Shao et al., 2022). More divergent lineages such as HSPA12A [heat shock protein family A (Hsp70) member 12A] and HSPA12B possess an unusual ATP-binding domain and show substantial sequence divergence from canonical HSP70 proteins (Han et al., 2003). Although HSPA12 proteins have recently been investigated in marine mollusks (i.e. Hu et al., 2022; Lu et al., 2024), their diversity and evolutionary relationships remain poorly understood across most gastropod lineages.

The HSP70 family plays essential roles in protein folding, proteostasis, protein translocation, and the prevention of protein aggregation (Wentink et al., 2026). In addition, HSP70 proteins participate in cellular responses to a wide range of environmental and physiological stressors (Singh et al., 2025). Structurally, canonical HSP70 proteins contain three major domains: a highly conserved N-terminal nucleotide-binding domain (NBD), a substrate-binding domain (SBD), and a variable C-terminal domain involved in co-chaperone interactions and subcellular targeting (Ciesielski et al., 2023; Melikov and Novak, 2024). In eukaryotic cells, HSP70 proteins have been reported in the cytoplasm, nucleus, endoplasmic reticulum, and mitochondria, with their subcellular localization largely determined by specific sequence motifs (Rosenzweig et al., 2019).

Until now, HSP70 proteins in terrestrial mollusks have been studied predominantly through protein expression analyses, particularly in Mediterranean species (e.g., Di Lellis et al., 2014; Staikou et al., 2024). The increasing availability of molluscan whole-genome assemblies has recently enabled genome-wide investigations of HSP70 gene families, revealing substantial variation in gene copy number, genomic organization, and evolutionary diversification among species (e.g. Hu et al., 2022; Zheng et al., 2023; Lu et al., 2024; Kim et al., 2024; Wu et al., 2025; Wang et al., 2025). However, comparable genome-wide studies remain scarce in terrestrial mollusks, and to our knowledge, have not been reported for Neotropical terrestrial gastropods.

In the Cuban archipelago, the genus *Polymita* Beck, 1837 comprises six endemic tree snail species: *P. muscarum*, *P. versicolor*, *P. venusta*, *P. sulphurosa*, *P. brocheri*, and *P. picta* (González-Guillén, 2021). These species are internationally recognized for their remarkable shell color polymorphism and are of considerable interest to evolutionary biologists and conservation practitioners (Alfonso and Berovides, 1993). All species of *Polymita* are considered threatened and are listed in Appendix I of the Convention on International Trade in Endangered Species (CITES, 2025; Maceira-Filgueira, 2016). Recently, genomic resources for *P. picta* have become available through the sequencing of its mitochondrial and nuclear genomes, providing new opportunities for evolutionary and comparative genomic research (Lewis et al., 2025; Reyes-Tur et al., 2025).

The aim of this study was to identify and characterize members of the HSP70 gene family in the recently assembled genome of *Polymita picta*. Specifically, we investigated their genomic distribution, physicochemical properties, conserved motifs, evolutionary relationships, and patterns of sequence divergence. To our knowledge, this represents the first genome-based characterization of HSP70 genes in a Neotropical terrestrial gastropod and provides a molecular framework for future studies linking HSP70 evolution and expression with responses to environmentally relevant stressors in land snails.

2 Materials and methods

2.1 Genome-wide identification of HSP70 family members in the Polymita picta genome

The identification of HSP70 family members in *P. picta* was based on sequence-homology and conserved-domain analyses, following approaches previously applied to molluscan genomes (Hu et al., 2022; Lu et al., 2024). Initial homology screening was performed using an earlier working version of the *P. picta* genomic dataset, whereas genomic coordinates and all final sequence analyses reported in this study refer to the subsequently established genome assembly v4665_2 (BioProject PRJNA1250545; GenBank WGS accession JBOZOB000000000.1).

A genome-wide TBLASTN v2.12.0 search was performed against the initial genomic dataset using 1,064 molluscan HSP protein sequences downloaded from NCBI as queries (downloaded

3 December 2024), using default search parameters. The search produced 78,994 raw alignments. Matches were filtered using an E-value threshold of $\leq 1 \times 10^{-5}$ and amino-acid identity $\geq 90\%$, resulting in 1,286 genomic alignments involving 472 of the original molluscan query proteins.

The 472 molluscan query proteins associated with the retained genomic alignments were subsequently screened for the characteristic HSP70 domain using *hmmscan* from HMMER v3.4 (Finn et al., 2011) against the Pfam-A database (El-Gebali et al., 2019). Matches to the HSP70 family profile hidden Markov model PF00012 were retained at an E-value threshold of $\leq 1 \times 10^{-5}$. Significant domain matches were parsed from the *domtblout* output and deduplicated, identifying 304 PF00012-positive molluscan query proteins. Thus, these 304 records represented HSP70-domain-containing query proteins associated with genomic TBLASTN hits, rather than 304 independently predicted *P. picta* HSP70 proteins.

Additional conserved-domain assessment was performed using the NCBI Batch Conserved Domain Search (CD-Search) against the Conserved Domain Database (CDD; Lu et al., 2020), using an E-value threshold of $\leq 1 \times 10^{-5}$. This analysis was used as an additional assessment of HSP70-family assignment among the query-associated records. Because multiple homologous molluscan query proteins frequently matched the same or overlapping genomic regions, the number of query-associated records was not interpreted as the number of HSP70 loci in the *P. picta* genome.

Candidate genomic regions identified in the initial dataset were subsequently mapped and curated against the established *P. picta* genome assembly v4665_2. Recurrent and overlapping matches supported by different molluscan HSP70 queries and mapping to the same genomic interval were consolidated as evidence for the same candidate region, whereas spatially distinct genomic positions were retained as separate candidate loci. Following preliminary consolidation and sequence assessment, candidate HSP70 regions were remapped to assembly v4665_2 to establish their final genomic locations. This mapping and subsequent curation resolved ten non-redundant candidate HSP70 loci: six located on scaffold ptg000811l_1 and one each on scaffolds ptg004791l_1, ptg010661l_1, ptg018870l_1, and ptg018978l_1. Genomic regions were additionally examined manually using Integrative Genomics Viewer (IGV) v2.19.3 (Thorvaldsdóttir et al., 2013).

Distinct candidates were retained when they occupied separate genomic positions and supported plausible HSP70 coding sequences. Because population-level or haplotype-resolved analyses were not performed, potentially allelic relationships among highly similar sequences could not be formally assessed. The retained candidates are therefore conservatively referred to as ten non-redundant genomic HSP70 loci based on their distinct positions in assembly v4665_2. These loci were subsequently subjected to sequence extraction, ORF assessment, and downstream characterization.

2.1.1 Identification of HSPA12-like sequences

Because HSPA12A and HSPA12B are highly divergent members of the HSP70 superfamily and have been reported in molluscan genomes (e.g., Hu et al., 2022; Lu et al., 2024), an additional targeted search was conducted to assess the presence of HSPA12-like sequences in *P. picta*. This search was performed independently of the primary HSP70 identification pipeline described above and directly against genome assembly v4665_2.

Representative molluscan HSPA12A and HSPA12B protein sequences retrieved from NCBI were used as queries in TBLASTN searches using BLAST+ v2.16.0 (Camacho et al., 2009). Matches with an E-value $\leq 1 \times 10^{-5}$ were retained and inspected for candidate HSPA12-like loci (Supplementary Data 1: http://doi.org/10.5281/zenodo.20664956). This complementary search was specifically intended to detect divergent HSP70-superfamily sequences that might not have been recovered by the stringent identity threshold applied during the primary screening.

2.1.2 Full-length sequence extraction and translation from genomic data

The ten non-redundant HSP70 loci retained after genomic mapping and curation were extracted from the *P. picta* genome assembly v4665_2. Their genomic coordinates were recorded in BED format, including scaffold, start and end positions, and strand orientation (Supplementary Data 2: http://doi.org/10.5281/zenodo.18494202). The final set comprised six loci on ptg000811l_1 and four dispersed loci located on ptg004791l_1, ptg010661l_1, ptg018870l_1, and ptg018978l_1.

Genomic sequence extraction and subsequent strand-aware processing were performed using BEDTools v2.30.0 (Quinlan and Hall, 2010) and custom R scripts (R Core Team, 2022). The genomic intervals corresponding to the ten retained loci were used as the reference coordinates for downstream extraction and translation.

To facilitate recovery of complete coding sequences, each candidate region was extended in a strand-aware manner by 50 bp at the 5′ end and 500 bp at the 3′ end. Sequences located on the negative strand were reverse complemented before translation. For each strand-oriented sequence, the three forward reading frames were evaluated. ORF selection prioritized a methionine-initiated ORF and, among plausible alternatives, the longest coding sequence with the fewest internal stop codons. The resulting ORFs were manually assessed together with the surrounding genomic sequence to identify plausible HSP70 coding sequences. Candidates lacking sufficient genomic and sequence support for assignment to the HSP70 family were excluded, whereas locus-specific sequence features, including apparent truncations or premature stop codons, were retained and reported when supported by the genomic sequence. The final predicted sequences, genomic coordinates, and strand orientations were retained for downstream analyses (Supplementary Data 3: http://doi.org/10.5281/zenodo.16744586).

Custom R scripts used for genomic sequence extraction, strand-aware extension, reading-frame evaluation, ORF selection, and protein translation were developed and refined with the assistance of ChatGPT (GPT-5.5, OpenAI, San Francisco, CA, USA). All generated code, outputs, and methodological decisions were subsequently reviewed, verified, and modified by the authors before implementation. This workflow was used to recover predicted HSP70 coding sequences associated with the ten curated genomic loci and was not intended to reconstruct complete exon–intron gene models.

2.2 Physicochemical characterization of *Polymita picta* HSP70 proteins

The physicochemical properties of the predicted *P. picta* HSP70 proteins (PpHSP70), including amino acid length, molecular weight (MW), theoretical isoelectric point (pI), instability index (II), aliphatic index (AI) and grand average of hydropathy (GRAVY), were calculated using the ExPASy ProtParam (Gasteiger et al., 2005). Subcellular localization and protein-sorting signals were predicted using DeepLoc 2.0 (Thumuluri et al., 2022). Signal peptide prediction was performed using SignalP v5.0 (Almagro et al., 2019). Conserved motifs were identified using MEME (Multiple Expectation Maximization for Motif Elicitation) within the MEME Suite (Bailey et al., 2015), with a maximum number of motifs set to 21 following the approach of Lu et al. (2024).

2.3 Motif analysis in *Polymita picta* HSP70 proteins and comparative sequences

A total of 34 protein sequences were included in motif analyses, comprising the 10 predicted HSP70 proteins from *P. picta* and representative HSP70 sequences from *Biomphalaria glabrata*, *Ostrea edulis*, *Magallana gigas*, *Mya arenaria*, *Aplysia californica*, *Dreissena polymorpha*, and *Crassostrea virginica*, together with one HSP70 sequence from *Drosophila melanogaster*. To facilitate comparisons among HSP70 subfamilies, sequences annotated as cytosolic HSP70, HSC70, HSP75, and HSP78 were included (Supplementary Data 4: https://doi.org/10.5281/zenodo.16744775). Species and sequences were selected based on the availability of annotated reference sequences representing the major HSP70 subfamilies included in the comparative analysis. The positions of the identified motifs were subsequently mapped onto the conserved structural architecture of HSP70 proteins to determine their association with the nucleotide-binding domain (NBD), its constituent subdomains (IA, IB, IIA, and IIB), the substrate-binding domain (SBD), and the C-terminal region (Supplementary Data 5: https://doi.org/10.5281/zenodo.21901174). Structural assignments were based on previously described HSP70 domain organization (Zhang et al., 2014).

2.4 Homology assessment and gene duplication analyses

Potential gene duplications were identified using two criteria: (i) amino acid sequence similarity greater than 70% and (ii) coverage of at least 70% of the longer sequence by the shorter aligned sequence (Yang et al., 2008; Lu et al., 2024). Protein sequences were aligned using MAFFT v7 (Katoh and Standley, 2013) and imported into R using the Biostrings package (Pagès et al., 2026). Sequence metadata, including scaffold identity and strand orientation, were parsed from FASTA headers and stored in a structured data frame.

Pairwise sequence comparisons were performed among all *Pp*HSP70 proteins. Amino acid identity was calculated as the proportion of identical residues relative to the total alignment length using custom R scripts. Patterns of sequence evolution among the predicted *Pp*HSP70 coding sequences were assessed through pairwise Ka/Ks analyses using KaKs_Calculator v3.0 (Zhang, 2022). The corresponding protein sequences were aligned using MAFFT v7, and the protein alignment together with the corresponding coding sequences was submitted to PAL2NAL (Suyama et al., 2006) to generate a codon-aware nucleotide alignment. The resulting alignment was converted to AXT format, and all 45 pairwise comparisons among the ten *Pp*HSP70 sequences were analyzed using the GY-HKY method, with HKY as the nucleotide substitution model. Ka/Ks ratios < 1, $= 1$, and > 1 were interpreted as being consistent with purifying selection, neutral evolution, and positive selection, respectively.

2.5 Phylogenetic analysis

For phylogenetic analyses, the identified *P. picta* HSP70 proteins were aligned together with the same set of comparative HSP70 sequences used for motif analyses to ensure consistency among comparative datasets. The major heat shock 70-kDa protein from *Drosophila melanogaster* was included as an external reference sequence and used to root the tree.

The best-fit amino acid substitution model was selected using ModelFinder implemented in IQ-TREE v2.3.6 under the "-m MFP" option (Kalyaanamoorthy et al., 2017). Maximum-likelihood phylogenetic reconstruction was subsequently performed using IQ-TREE (Nguyen et al., 2015; Minh et al., 2020).

Branch support was assessed using ultrafast bootstrap approximation (UFBoot) and the Shimodaira–Hasegawa approximate likelihood ratio test (SH-aLRT) with 10,000 replicates each (Guindon et al., 2010; Hoang et al., 2018). Nodes with UFBoot values ≥ 95 and SH-aLRT support values ≥ 80 were considered strongly supported (Cruz-Laufer et al., 2025). Phylogenetic trees were visualized and edited using FigTree v1.4.4.

3 Results

3.1 Genome-wide identification and physicochemical profiling of heat shock proteins 70 kDa in *Polymita picta*

The initial TBLASTN search using 1,064 molluscan HSP protein sequences recovered 78,994 alignments against the initial *P. picta* genomic dataset. Filtering at an E-value $\leq 1 \times 10^{-5}$ and amino-acid identity ≥ 90% retained 1,286 genomic alignments involving 472 of the original molluscan query proteins. HMMER screening subsequently identified 304 of these query proteins containing the HSP70 domain PF00012. Because multiple homologous molluscan query proteins frequently matched the same or overlapping genomic regions, these query-associated records were not interpreted as independent *P. picta* HSP70 loci. Candidate genomic regions were therefore consolidated according to genomic position and subsequently mapped and curated against genome assembly v4665_2. This procedure resolved ten non-redundant candidate HSP70 loci, which yielded ten predicted HSP70 protein sequences following sequence extraction and ORF assessment.

The ten predicted HSP70 proteins were located at distinct genomic positions and were named consecutively *Pp*HSP70_1 to *Pp*HSP70_10. Information associated with each identifier included scaffold, start and end coordinates, and strand orientation, with genomic positions reported in the text using 1-based coordinates e.g., *Pp*HSP70_1_ptg000811l_1:167165–168904 (–) (Supplementary Data 6: http://doi.org/10.5281/zenodo.16744894).

The ten predicted HSP70 proteins were distributed across five genomic scaffolds (Figure 1). *Pp*HSP70_1 to *Pp*HSP70_6 were clustered within scaffold ptg000811l_1, where four loci occurred on the negative strand and two on the positive strand. The remaining four loci were located individually on scaffolds ptg004791l_1, ptg010661l_1, ptg018870l_1, and ptg018978l_1.

A targeted search using representative molluscan HSPA12A and HSPA12B proteins recovered 56 significant TBLASTN matches against the *P. picta* genome assembly v4665_2. The strongest matches were concentrated within scaffold ptg008348l_1 and showed high similarity to gastropod HSPA12A proteins. The highest-scoring hit corresponded to HSPA12A from *Physella acuta*, exhibiting 92.0% amino-acid identity across 350 aligned residues (E-value = 0.0; bit score = 680). Similarly, HSPA12A from *Biomphalaria glabrata* showed 91.7% identity across 350 amino acids (E-value = 0.0; bit score = 677). Additional significant matches were recovered using HSPA12A sequences from *Berghia stephanieae* (86.0% identity) and *Haliotis discus* (77.8% identity), as well as HSPA12B sequences from *Babylonia areolata* (77.8% identity). Most high-confidence matches mapped to a restricted genomic region within scaffold ptg008348l_1, supporting the presence of at least one HSPA12-like locus in the current assembly. HSPA12-like sequences were not included among the ten predicted HSP70 proteins retained for downstream analyses because they were identified through the independent targeted search described in Section 2.1.1.

The ten predicted HSP70 proteins ranged from 635 to 672 amino acids in length, with an average length of approximately 641 amino acids (Table 1). Molecular weights ranged from 69,305.92 to 73,549.12 Da, with an average of 69,725.06 Da. The theoretical isoelectric points (pI) ranged from 5.58 to 5.80 (mean = 5.67), indicating a generally acidic character. All proteins exhibited negative GRAVY values (−0.47 to −0.38; mean = −0.44), indicating an overall hydrophilic character. The instability index ranged from 28.23 to 30.69 (mean = 29.24), classifying all proteins as stable according to ProtParam criteria. The aliphatic index ranged from 79.43 to 84.32 (mean = 81.15), values comparable to those reported for other molluscan HSP70 proteins (Table 1).

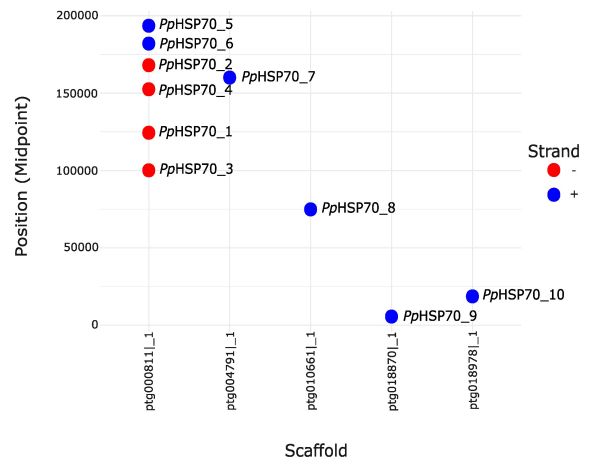

FIGURE 1
Scaffold distribution of predicted *Polymita picta* heat shock protein 70 kDa (HSP70) loci.

Most *P. picta* HSP70 proteins were predicted to be primarily localized in the cytoplasm, with prediction scores ranging from 0.47 to 0.78 (mean = 0.73) (Supplementary Data 7: https://doi.org/10.5281/zenodo.16745227). However, two proteins, *Pp*HSP70_3 and *Pp*HSP70_8, contained nuclear localization signals (NLS), suggesting potential dual localization in the cytoplasm and nucleus. Additionally, *Pp*HSP70_7 exhibited a predicted peroxisomal targeting signal and showed lower cytoplasmic probability (0.47) together with elevated probabilities for the endoplasmic reticulum (0.52), Golgi apparatus (0.47), and lysosome/vacuole (0.46). Predicted mitochondrial and plasma membrane localization probabilities were generally low across all proteins (mean < 0.34).

Signal peptide prediction using SignalP 5.0 indicated that none of the *Pp*HSP70 proteins possessed a classical N-terminal signal peptide for secretion through the Sec/SPI pathway. All sequences were classified as "OTHER" with high confidence (mean OTHER probability = 0.9975 ± 0.0003) (Supplementary Data 8: https://doi.org/10.5281/zenodo.16745405). Correspondingly, SP (Sec/SPI) probabilities were consistently low across all proteins (mean SP probability = 0.0025 ± 0.0003), indicating that none of the identified proteins are likely to be secreted through the classical secretory pathway.

3.2 Motif analysis of HSP70 family in *Polymita picta* and comparative sequences

A total of 21 conserved motifs were identified across the dataset, revealing differences in motif architecture among the major HSP70 subfamilies (Figure 2; Supplementary Data 5). Mitochondrial HSP70 proteins (HSP75) shared a unique N-terminal motif (motif 21) that was absent from all other sequences. HSP75 proteins also contained motif 18 within the ATPase region of the nucleotide-binding domain (NBD), whereas HSP78 proteins exhibited a distinct motif architecture in the corresponding region. Neither motif 18 nor motif 21 was detected in the cytosolic HSP70 proteins, including all *P. picta* sequences.

The three HSC70 proteins lacked motif 15 within the substrate-binding domain (SBD), whereas this motif was consistently present in the ten predicted *P. picta* HSP70 proteins and the inducible HSP70 sequence from *D. melanogaster*. Minor differences were observed in two short motifs (~8 amino acids) located near the IA actin-like ATPase subdomain of the NBD and in the central region of the SBD. Sequence alignment (Supplementary Data 9: https://doi.org/10.5281/zenodo.16746341) indicated that these differences corresponded primarily to scattered amino acid substitutions, most of which involved conservative substitutions.

Sequence comparisons were standardized using the conserved EEVD motif as a reference endpoint. Only *Pp*HSP70_7 contained a C-terminal extension of 30 residues downstream of the EEVD motif before the stop codon. In contrast, the predicted *Pp*HSP70_8 sequence terminated upstream of the expected C-terminal EEVD

TABLE 1 Physicochemical parameters of *Polymita picta* canonical heat shock protein 70 kDa.

Sequence	Number of amino acids	Molecular weight	Isoelectric point	GRAVY	Instability index	Aliphatic index
*Pp*HSP70_1	637	69536.36	5.66	-0.44	28.81	81.46
*Pp*HSP70_2	637	69550.26	5.58	-0.46	30.69	81
*Pp*HSP70_3	645	70355.17	5.59	-0.45	29.52	80.76
*Pp*HSP70_4	637	69607.35	5.66	-0.47	29.84	80.55
*Pp*HSP70_5	637	69580.28	5.58	-0.46	30.2	80.55
*Pp*HSP70_6	637	69607.35	5.66	-0.47	29.84	80.55
*Pp*HSP70_7	672	73549.12	5.8	-0.38	29.44	84.32
*Pp*HSP70_8	635	69305.92	5.76	-0.47	28.23	79.43
*Pp*HSP70_9	641	69925.64	5.69	-0.44	28.44	80.34
*Pp*HSP70_10	635	69317.22	5.77	-0.42	30.34	83.56

motif. Re-examination of the extended genomic region confirmed the presence of a premature stop codon at this position; however, the downstream genomic sequence retained the EEVD-encoding region in the same relative position observed in *Pp*HSP70_7 and *Pp*HSP70_9. In the region immediately upstream of EEVD (motif 16; Figure 2), short sequence variations were observed in *Pp*HSP70_7 (SGSSR), *Pp*HSP70_8 (AHSQ), and *Pp*HSP70_9 (AGSQ). In addition, a three-amino-acid insertion (QAT) was observed in the corresponding C-terminal region of *Pp*HSP70_10. The N-terminal region exhibited minor variation immediately following the initiating methionine. *Pp*HSP70_7 started with MT, *Pp*HSP70_8 with MS, and *Pp*HSP70_10 with MP, whereas all remaining proteins started with MA. Furthermore, *Pp*HSP70_10 lacked three amino acids in this region relative to the other paralogs.

3.3 Sequence identity and evolutionary constraints

Pairwise amino acid identity percentages were calculated to assess sequence similarity among the ten *P. picta* HSP70 proteins. As shown in Figure 3, proteins located within the same scaffold generally exhibited very high sequence identity (>97%). In contrast, lower identity values were observed among proteins located on different scaffolds. The lowest amino acid identity (83.26%) was observed between *Pp*HSP70_7 and *Pp*HSP70_10.

Pairwise Ka/Ks ratios were calculated to evaluate patterns of molecular evolution among the identified HSP70 proteins (Supplementary Data 10: https://doi.org/10.5281/zenodo.16751419). A total of 45 pairwise comparisons were analyzed. All estimated Ka/Ks values were below 1, indicating a predominance of synonymous substitutions relative to non-synonymous substitutions. The lowest ratio was observed between *Pp*HSP70_1 and *Pp*HSP70_10 (Ka/Ks = 0.0127), whereas the highest value was obtained for the comparisons *Pp*HSP70_3 versus *Pp*HSP70_4 and *Pp*HSP70_3 versus *Pp*HSP70_6 (Ka/Ks = 0.1642). The comparison between *Pp*HSP70_4 and *Pp*HSP70_6 returned a NaN value because no informative substitutions were available for Ka/Ks estimation.

3.4 Phylogenetic analysis of *Polymita picta* HSP70 and comparative sequences

Maximum-likelihood phylogenetic reconstruction recovered two major clades (Figure 4). The first clade grouped all *P. picta* sequences together with four cytosolic HSP70 proteins from the pulmonate gastropod *Biomphalaria glabrata*, receiving strong statistical support (SH-aLRT/UFBoot = 100/100). Within this clade, two principal subgroups were observed. One subgroup contained exclusively the four *B. glabrata* cytosolic HSP70 proteins, with internal nodes receiving very strong support (SH-aLRT/UFBoot = 100/100 and 98.8/100). The second subgroup contained all ten *P. picta* HSP70 proteins.

Relationships among the *P. picta* paralogs showed heterogeneous support. The position of *Pp*HSP70_10 relative to the remaining paralogs was weakly resolved, and the lowest support values within the *P. picta* clade were associated with nodes involving *Pp*HSP70_1, *Pp*HSP70_3, *Pp*HSP70_4, and *Pp*HSP70_6. In contrast, most remaining nodes exhibited moderate to strong support. The second major clade grouped HSC70, HSP75, and HSP78 proteins from the remaining molluscan species included in the analysis, also with strong statistical support (SH-aLRT/UFBoot = 99.4/100). Within this clade, an additional subgroup containing cytosolic HSP70 proteins from other molluscan taxa was recovered, although relationships among these sequences remained poorly resolved.

4 Discussion

This study provides the first genome-based characterization of HSP70 proteins in the Cuban painted snail *Polymita picta*. Using a combination of homology searches, domain prediction, motif analyses, and phylogenetic reconstruction, we identified ten high-confidence predicted HSP70 proteins and evaluated their molecular diversity and evolutionary relationships within a comparative molluscan framework.

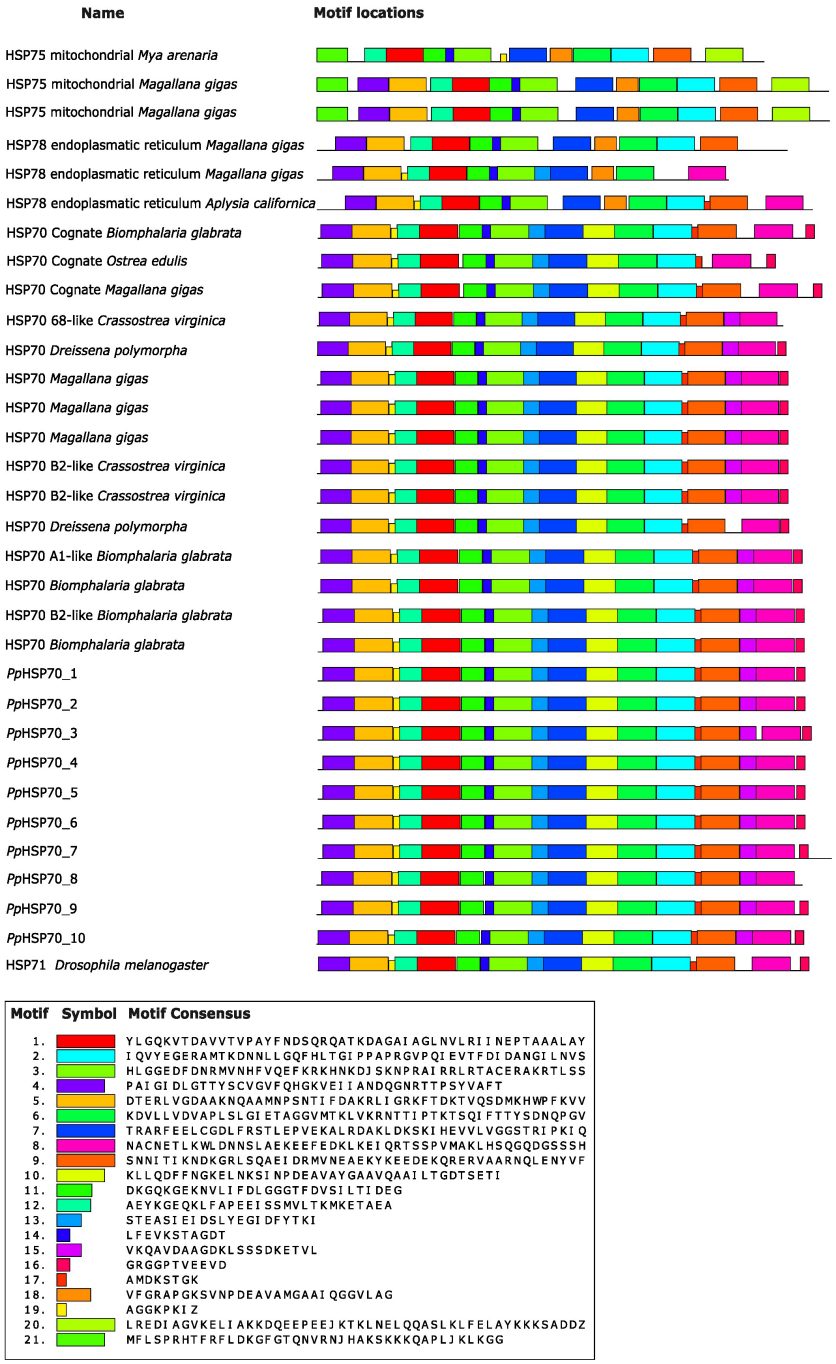

FIGURE 2
Distribution of conserved motifs of different subfamilies of heat shock proteins 70 kDa (HSP70) in *Polymita picta* and comparative sequences.

4.1 The genome-wide identification of HSP70s in *Polymita picta*

The heat shock proteins of 70 kDa (HSP70) constitute a well-characterized family of molecular chaperones involved in cellular homeostasis and stress responses from bacteria to mammals (Yu et al., 2021; Singh et al., 2025). However, studies addressing HSP70 diversity in mollusks have predominantly focused on species of economic importance, either as food resources or invasive taxa (Hu et al., 2022; Lu et al., 2024). Comparable genome-based studies in terrestrial gastropods remain extremely limited, hindering our understanding of the evolution and diversification of these proteins in land snails.

The recent availability of the *P. picta* genome assembly (Reyes-Tur et al., 2025) enabled this first systematic survey of HSP70 proteins in the species. The assembly comprises 1.85 Gbp distributed across 22,628 contigs, with a contig N50 of 124.2 Kbp and an estimated completeness of 88.4%. The relatively high heterozygosity reported for the genome (2.41%) is consistent with values observed in other molluscan species (Bean et al., 2022; Gundappa et al., 2022; Reyes-Tur et al., 2025). Given these assembly characteristics, the identified sequences were conservatively interpreted as non-

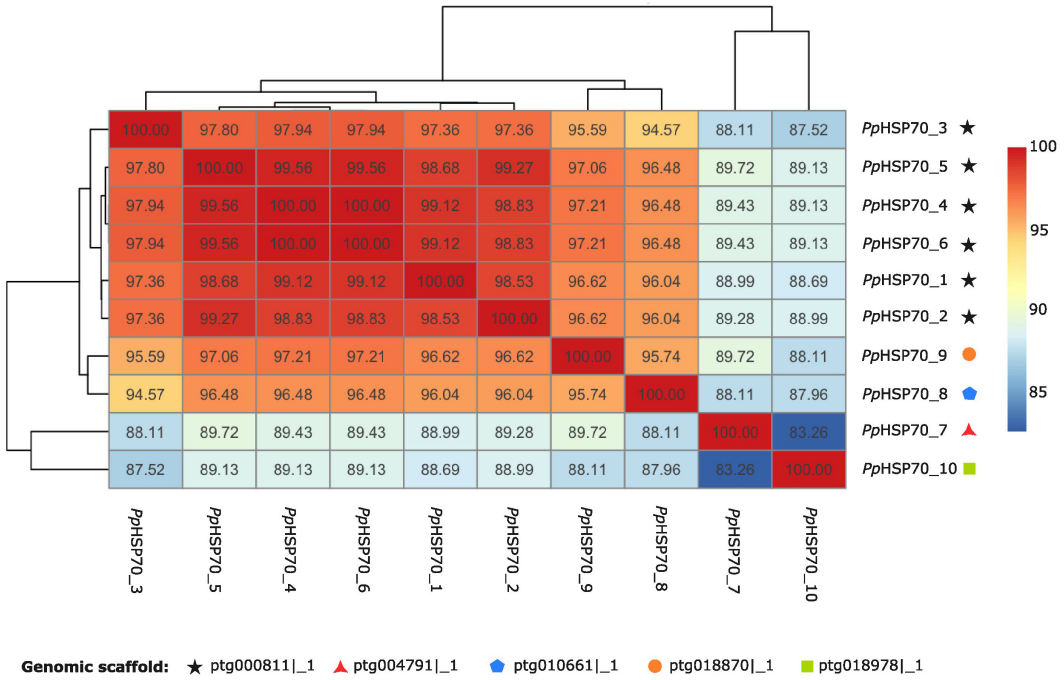

FIGURE 3
Matrix of identical positions (pID) percentages between *Polymita picta* heat shock protein 70 kDa (HSP70) sequences. Dendrograms represent hierarchical clustering based on pairwise sequence identity. Symbols indicate the genomic scaffold containing each HSP70 sequence, with identical symbols denoting sequences located on the same scaffold.

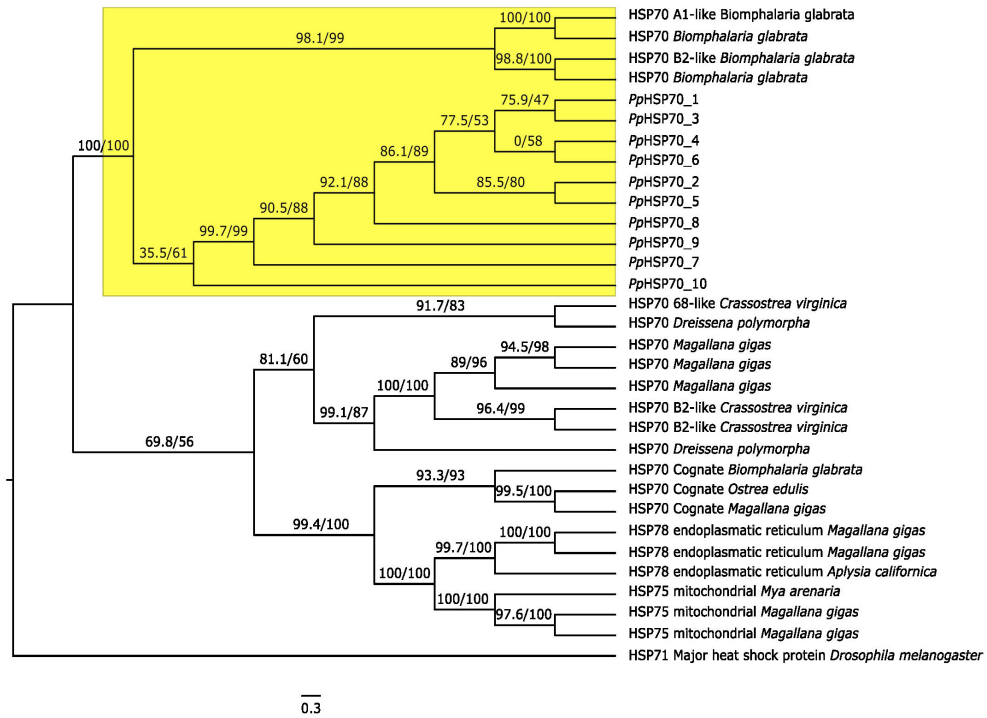

FIGURE 4
Maximum-likelihood phylogenetic reconstruction of *Polymita picta* heat shock protein 70 kDa (HSP70) amino acid sequences and representative molluscan orthologs. The clade containing all *P. picta* sequences is highlighted in yellow.

redundant genomic loci based on their distinct positions in the assembly; however, the possibility that some highly similar sequences represent allelic or haplotypic variation cannot be definitively excluded without population-level or haplotype-resolved genomic data.

A total of ten high-confidence predicted HSP70 proteins were identified in the genome. This number is comparable to that reported for several fish species and for the sea cucumber *Apostichopus japonicus* (Gao et al., 2018; Xu et al., 2018; Ma and Luo, 2020; Gao et al., 2022; Zheng et al., 2023). In contrast, substantially larger HSP70 repertoires have been reported in some marine bivalves, including 65 genes in the blood clam *Tegillarca granosa* (Kim et al., 2024) and more than one hundred HSP70-related sequences in the Pacific oyster *Magallana gigas* (Lu et al., 2024). These differences likely reflect lineage-specific patterns of gene family expansion and highlight the considerable variation in HSP70 copy number observed across molluscan taxa.

The ten identified HSP70 proteins were distributed across five genomic scaffolds. Four sequences (*Pp*HSP70_7 to *Pp*HSP70_10) were located on separate scaffolds, whereas *Pp*HSP70_1 to *Pp*HSP70_6 occurred in close proximity within scaffold ptg000811l_1. This distribution is consistent with the occurrence of both clustered and dispersed copies within the genome. Similar patterns of HSP70 gene organization have been reported in several molluscan genomes, including species exhibiting tandemly arranged HSP70 genes (Hu et al., 2022; Lu et al., 2024). However, given the scaffold-level nature of the current assembly, the precise genomic organization and evolutionary origin of these copies will require confirmation through chromosome-level assemblies and improved gene annotation.

A targeted search additionally recovered significant HSPA12-like matches within the *P. picta* genome. Most high-confidence hits were concentrated in scaffold ptg008348l_1 and exhibited strong similarity to HSPA12A proteins from other gastropods. Although HSPA12 proteins belong to the broader HSP70 superfamily, they constitute a highly divergent lineage and possess an atypical ATP-binding domain compared with canonical HSP70 proteins (Han et al., 2003; Lu et al., 2024). This sequence and domain divergence limits their direct comparability with the canonical HSP70 proteins analyzed here, particularly for analyses based on conserved motif architecture and phylogenetic relationships. Their inclusion within the same analytical framework could therefore confound the interpretation of evolutionary patterns inferred for the ten high-confidence canonical HSP70 proteins. For this reason, HSPA12-like sequences were identified and reported separately but were not incorporated into the comparative analyses of the canonical HSP70 dataset. Their evolutionary relationships and genomic organization warrant dedicated investigation in future studies.

Overall, *P. picta* HSP70 proteins exhibited a high degree of conservation in sequence length, molecular weight, isoelectric point, and hydropathic properties. All proteins displayed acidic pI values and negative GRAVY scores, consistent with the physicochemical characteristics commonly reported for HSP70 proteins in other taxa (Dayrit et al., 2024; Singh et al., 2025). Likewise, all sequences were classified as stable proteins according to their instability indices. Their relatively high aliphatic indices are also consistent with predicted thermostability, as a higher aliphatic index has been associated with increased thermostability of globular proteins (Ikai, 1980; Gasteiger et al., 2005). Although variation in aliphatic index among the *P. picta* paralogs was relatively small, the observed values are consistent with the predicted structural stability of these proteins. Hypothetically, these physicochemical properties could contribute to maintaining HSP70 structural stability and chaperone performance under elevated or fluctuating temperatures. However, this potential relationship cannot be inferred from sequence-based predictions alone and would require experimental validation under controlled thermal conditions. The predicted subcellular localization patterns suggest that the identified proteins are predominantly cytoplasmic chaperones. Nevertheless, *Pp*HSP70_3 and *Pp*HSP70_8 contained nuclear localization signals, indicating the possibility of dual cytoplasmic and nuclear localization. In addition, *Pp*HSP70_7 displayed targeting signals associated with multiple intracellular compartments, including the peroxisome, endoplasmic reticulum, Golgi apparatus, and lysosome/vacuole. Although these predictions require experimental validation, they suggest that the strong conservation of the core HSP70 molecular chaperone properties may coexist with some degree of intracellular specialization among paralogs. Thus, rather than indicating extensive functional divergence, differences in predicted subcellular targeting may reflect incipient subfunctionalization while retaining the conserved molecular chaperone function characteristic of the HSP70 family (Yu et al., 2021; Singh et al., 2025). Taken together, the localization predictions suggest that most *P. picta* HSP70 proteins are predominantly cytoplasmic chaperones, although several paralogs may also localize to additional intracellular compartments, including the nucleus and peroxisomes. Such patterns are consistent with the diverse intracellular distribution reported for HSP70 proteins in other eukaryotes (Hasnain and Kaneko, 2022; Kaneko, 2022). Importantly, none of the ten proteins contained a classical N-terminal signal peptide for secretion through the Sec/SPI pathway, consistent with an intracellular rather than secreted localization.

4.2 Motif analysis of HSP70 family in *Polymita picta* and comparative sequences

Conserved motifs are commonly associated with structural and functional features of proteins and can therefore provide information on the organization and differentiation of protein families (Lu et al., 2024). The motif architecture observed in *P. picta* and the comparative dataset is consistent with the canonical structural organization of HSP70 proteins. The presence of subfamily-specific motifs in HSP75 and HSP78, together with the absence of motif 15 in HSC70 proteins, highlights the structural differentiation among the major HSP70 lineages included in this study. In particular, the consistent presence of motif 15 in all inducible HSP70 proteins, including the ten predicted *P. picta* HSP70 proteins that clustered with the inducible cytosolic HSP70 group and the inducible HSP70 of *D. melanogaster*, is consistent with their sequence-based placement within the inducible cytosolic HSP70 group.

Most differences observed among *P. picta* paralogs consisted of scattered single amino acid substitutions that did not alter the overall motif architecture. Although *Pp*HSP70_8 exhibited minor variation in the region associated with motif 16, all identified

proteins retained the characteristic motif composition of HSP70 chaperones. Likewise, *Pp*HSP70_10 lacked several amino acids at the N-terminal region but retained the complete motif architecture associated with the nucleotide-binding domain. No major insertions or deletions affecting conserved domains were detected. All proteins contained the expected nucleotide-binding domain (NBD) and substrate-binding domain (SBD). The characteristic C-terminal EEVD motif associated with cytosolic HSP70 proteins was present in all predicted proteins except *Pp*HSP70_8. In this sequence, the predicted ORF terminated upstream of the expected EEVD motif because of a premature stop codon. However, re-examination of the downstream genomic region revealed the conserved EEVD-encoding sequence in the same relative position observed in closely related paralogs. Whether this interruption represents genuine sequence variation or a local limitation of the current genome assembly cannot be determined from the available genomic data. The overall conservation of motif composition suggests that the identified paralogs have maintained the core structural features required for HSP70 function despite sequence divergence (Yu et al., 2021; Singh et al., 2025).

The high degree of motif conservation observed among the ten proteins is consistent with the strong sequence similarity detected in pairwise comparisons and may reflect relatively recent duplication events, strong functional constraints, or a combination of both processes. However, the functional significance of the observed amino acid substitutions remains unknown and will require experimental validation through transcriptomic, proteomic, and expression-based studies. Comparisons with orthologous HSP70 proteins from other mollusks provide a useful framework for interpreting the evolutionary position of the *P. picta* sequences. Nevertheless, these comparisons remain constrained by the availability of annotated reference genomes and experimentally validated HSP70 proteins from terrestrial gastropods. As genomic resources continue to expand for non-model mollusks, future studies will be able to evaluate the conservation and diversification of HSP70 proteins in a broader phylogenetic context.

4.3 Sequence identity and evolutionary constraints

The amino acid sequences of *Pp*HSP70_1 to *Pp*HSP70_6, which are located within the same scaffold, exhibited very high levels of sequence similarity (>97%). This pattern, together with their close genomic proximity, is consistent with relatively recent local duplication events, although strong functional constraints may also contribute to maintaining their high sequence conservation. Although the current genome assembly does not allow definitive discrimination between duplicated loci and highly similar allelic variants, tandemly arranged HSP70 genes have been reported in several molluscan genomes. For example, HSP70 genes are unevenly distributed across the *Mercenaria mercenaria* genome, with a particularly high concentration on chromosome 7 and lineage-specific expansion of the HSPA12 subfamily (Hu et al., 2022). Similar patterns of HSP70 expansion and clustered or tandemly duplicated gene organization have been reported across several bivalve lineages (Jahan et al., 2023; Kim et al., 2024; Lu et al., 2024). Thus, the clustered distribution and high sequence identity of *Pp*HSP70_1–*Pp*HSP70_6 may represent a comparable pattern of local HSP70 expansion in *P. picta*. This interpretation is further supported by the phylogenetic reconstruction (Figure 4), in which *Pp*HSP70_1–*Pp*HSP70_6 form a well-supported group within the *P. picta* HSP70 lineage. The concordance between their genomic proximity, high sequence identity, and phylogenetic clustering is consistent with a lineage-specific local expansion of HSP70 copies. Although some relationships within this group showed lower node support, this likely reflects the very high sequence similarity among these paralogs and the consequently limited phylogenetic signal available to resolve their internal branching order. Therefore, while the available evidence supports a common evolutionary origin of this clustered group, the precise sequence and timing of individual duplication events cannot be resolved with the current scaffold-level assembly and phylogenetic data. Lower sequence identity values were observed among *Pp*HSP70_7–*Pp*HSP70_10; however, all pairwise comparisons still exceeded 83% amino acid identity. This level of conservation is consistent with the strong conservation observed among duplicated HSP70 family members in other mollusks (Lu et al., 2024) and indicates that all identified proteins belong to a closely related group of cytosolic HSP70 homologues.

All pairwise Ka/Ks ratios were below 0.16, indicating a predominance of synonymous over non-synonymous substitutions. Such values are consistent with predominantly strong purifying selection acting across the identified HSP70 coding sequences. Similar patterns of purifying selection have been reported across metazoan HSP70 lineages, consistent with strong functional constraints acting on these molecular chaperones (Yu et al., 2021; He et al., 2024). Gene duplication is widely recognized as an important source of genetic novelty and may contribute to the diversification of gene families through time (Kuzmin et al., 2022; Yanai, 2022). However, the uniformly low Ka/Ks values observed in *P. picta* indicate that the identified HSP70 proteins have retained a high degree of sequence conservation. In particular, the combination of high amino acid identity, conserved motif architecture, and low Ka/Ks ratios among the clustered copies is compatible with duplication followed by strong purifying selection, rather than extensive sequence divergence after duplication. No consistent association was apparent between Ka/Ks magnitude and basal or derived phylogenetic position (Figure 4), suggesting that purifying selection has acted broadly across the *P. picta* HSP70 lineage rather than being restricted to particular phylogenetic positions. Notably, *Pp*HSP70_4 and *Pp*HSP70_6 are identical at the amino acid level and occur as sister sequences in the phylogeny, a pattern consistent with a very recent duplication or exceptionally strong sequence conservation. However, because sequence identity and Ka/Ks estimates do not provide a direct measure of duplication time, the relative timing of these events cannot be established from the present data. Together with the observed motif conservation, these results suggest that the identified copies have maintained core chaperone functions and remain subject to strong evolutionary constraints.

Further studies incorporating chromosome-level assemblies, population-scale genomic sampling, transcriptomics, and expression profiling will be necessary to determine the evolutionary origin, genomic organization, and potential functional differentiation of individual HSP70 paralogs in *P. picta*.

4.4 Ecological and conservation perspectives

The identification of ten HSP70 proteins in *P. picta* provides a new molecular resource for investigating stress-response mechanisms in a threatened Cuban endemic painted snail. As a terrestrial ectotherm with limited mobility, *P. picta* cannot rapidly escape unfavorable environmental conditions, making cellular mechanisms involved in maintaining proteostasis potentially relevant to its responses to environmental stress. In this context, characterization of the HSP70 repertoire provides a genomic basis for formulating and testing hypotheses concerning how *P. picta* responds to environmentally relevant thermal variation. However, although HSP70 proteins are widely recognized as important components of cellular stress responses, the functional significance of the gene repertoire identified here remains unknown and cannot be inferred solely from sequence data. From an evolutionary perspective, the local clustering, high sequence identity, conserved motif architecture, and predominantly low Ka/Ks ratios observed among the *P. picta* paralogs suggest that expansion of this gene family has occurred while core chaperone functions remained under strong evolutionary constraint. Whether this genomic organization has ecological consequences –for example, through differences in expression among paralogs, tissue-specific responses, or differential activation under thermal stress– remains an experimentally testable question. The occurrence of multiple HSP70 copies in *P. picta* should not be interpreted as direct evidence of enhanced thermal tolerance, physiological plasticity, or resilience to climate change. Gene duplications are common evolutionary phenomena and may be associated with a variety of neutral, conserved, or lineage-specific processes (Kuzmin et al., 2022; Birchler and Yang, 2022). Experimental studies in other organisms have demonstrated that increased HSP70 copy number can produce both beneficial and detrimental effects depending on the environmental context (Roberts and Feder, 2000). Consequently, the relationship between HSP70 copy number and organismal performance cannot be assumed without functional validation.

The present study is based on a single genome assembly generated from one individual (Reyes-Tur et al., 2025). Therefore, the extent of intraspecific variation in HSP70 gene content, sequence diversity, and genomic organization remains unknown. This limitation is particularly relevant from a conservation-genomic perspective because determining whether the observed HSP70 repertoire is conserved across populations, or varies among geographically or environmentally differentiated populations, will require population-level sampling. Likewise, the targeted identification of HSPA12-like candidates suggests that additional members of the HSP70 superfamily may occur within the *P. picta* genome and deserve further investigation. Integrating population genomics with transcriptomics and gene-expression analyses under ecologically relevant thermal conditions would allow future studies to determine whether particular HSP70 paralogs show environment-dependent regulation and whether variation in the HSP70 repertoire is associated with differences among populations. Such approaches will provide a more complete understanding of HSP70 evolution in *Polymita* and establish a stronger framework for future conservation-oriented research in terrestrial gastropods.

5 Conclusion

This study presents the first genome-wide characterization of the canonical HSP70 repertoire in the Cuban endemic tree snail *P. picta*, identifying ten predicted cytosolic HSP70 homologues and characterizing their genomic organization and evolutionary relationships. The combined evidence from genomic distribution, sequence identity, conserved motif architecture, phylogenetic relationships, and Ka/Ks analyses is consistent with local expansion of HSP70 copies followed by predominantly strong purifying selection, indicating conservation of the core molecular chaperone architecture despite gene duplication. However, the current scaffold-level genome and single-individual sampling do not allow the timing of duplication events, population-level variation, or the functional differentiation of individual paralogs to be determined. Therefore, the present results should be regarded as a genomic framework for testing, rather than evidence of, ecological or physiological roles of individual HSP70 paralogs. Integrating these genomic data with population sampling, gene expression analyses, and experiments under ecologically relevant environmental conditions will be necessary to establish how HSP70 diversification relates to stress responses in *P. picta* and to evaluate its broader significance for the conservation biology of *Polymita*.

Data availability statement

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/Supplementary Material.

Ethics statement

The manuscript presents research that does not require ethical approval.

Author contributions

MG-P: Conceptualization, Formal analysis, Methodology, Visualization, Validation, Writing – original draft, Writing – review & editing, Resources, Data curation, Software, Investigation. BR-T: Writing – original draft, Supervision, Writing – review & editing. ZC: Software, Writing – original draft, Resources, Data curation, Methodology, Investigation, Writing – review & editing. JS: Conceptualization, Supervision, Writing – review & editing, Writing – original draft, Investigation, Resources, Validation, Formal analysis, Methodology. CG: Resources, Validation, Conceptualization, Writing – review & editing, Methodology, Formal analysis, Writing – original draft, Investigation, Supervision. MH: Formal analysis, Investigation, Writing – original draft, Writing – review & editing. KS: Supervision, Investigation, Resources, Writing – review & editing, Formal analysis, Writing – original draft. NB: Resources, Conceptualization,

Investigation, Writing – review & editing, Validation, Visualization, Formal analysis, Supervision, Methodology, Writing – original draft.

Funding

The author(s) declared that financial support was not received for this work and/or its publication.

Acknowledgments

The authors would like to thank Houman Kahroba for his valuable methodological guidance and insightful recommendations throughout the development of this study. We also thank Ivan Rodríguez López for his assistance in reviewing the manuscript's language, writing style, and overall readability. We are particularly grateful to Thierry Backeljau, Ann Cuypers, Joris M. Koene, Carolyn Hogg, Luke Silver and Andrea Schraven for their critical review and constructive comments on an earlier version of this work, which formed part of the first author's doctoral dissertation. Their feedback substantially contributed to improving the scientific quality and interpretation of the manuscript.

Conflict of interest

The author(s) declared that this work was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Generative AI statement

The author(s) declared that generative AI was used in the creation of this manuscript. The authors acknowledge the use of ChatGPT (GPT-5.5, OpenAI, San Francisco, CA, USA) during manuscript preparation for language editing, text refinement, assistance with R code development and troubleshooting, and methodological discussions. All outputs were critically reviewed, validated, and revised by the authors, who assume full responsibility for the content of the final manuscript.

Any alternative text (alt text) provided alongside figures in this article has been generated by Frontiers with the support of artificial intelligence and reasonable efforts have been made to ensure accuracy, including review by the authors wherever possible. If you identify any issues, please contact us.

Publisher's note

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fevo.2026.1933916/full#supplementary-material

References

Alfonso, M., and Berovides, V. (1993). Conservation problems of land snails in Cuba. *Tentacle* 3, 20–23.

Almagro, J. J., Tsirigos, K. D., and Sønderby, C. K. (2019). SignalP 5.0 improves signal peptide predictions using deep neural networks. *Nat. Biotechnol.* 37, 420–423. doi: 10.1038/s41587-019-0036-z

Bailey, T. L., Johnson, J., Grant, C. E., and Noble, W. S. (2015). The MEME suite. *Nucleic Acids Res.* 43, 39–49. doi: 10.1093/nar/gkv416

Bean, T. P., Tanguy, A., Peñaloza, C., Gundappa, M. K., Boutet, I., Houston, R. D., et al. (2022). Two parallel chromosome-level reference genomes to support restoration and aquaculture of European flat oyster *Ostrea edulis*. *Evol. Appl.* 15, 1709–1712. doi: 10.1111/eva.13465

Beck, C. H. H. (1837). *Index molluscorum praesentis aevi musei principis augustissimi Christiani Frederici. Fasciculus primus*. (Copenhagen: Hafniae), 1–100. Available online at: https://www.biodiversitylibrary.org/bibliography/77331.

Birchler, J. A., and Yang, H. (2022). The multiple fates of gene duplications: Deletion, hypofunctionalization, subfunctionalization, neofunctionalization, dosage balance constraints, and neutral variation. *Plant Cell* 34, 2466–2474. doi: 10.1093/plcell/koac076

Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., et al. (2009). BLAST+: architecture and applications. *BMC Bioinf.* 10, 421. doi: 10.1186/1471-2105-10-421

Ciesielski, S. J., Young, C., Ciesielska, E. J., and Ciesielski, G. L. (2023). The Hsp70 and JDP proteins: Structure-function perspective on molecular chaperone activity. *Enzymes* 54, 221–245. doi: 10.1016/bs.enz.2023.07.008

CITES (2025). "Appendices I, II and III," in *Convention on International Trade in Endangered Species of Wild Fauna and Flora* (Geneva, Switzerland: CITES Secretariat).

Cruz-Laufer, A. J., Vanhove, M. P., Bachmann, L., Barson, M., Bassirou, H., Bitja Nyom, A. R., et al. (2025). Adaptive evolution of stress response genes in parasites aligns with host niche diversity. *BMC Biol.* 23, 10. doi: 10.1186/s12915-024-02091-w

Dayrit, G. B., Burigsay, N. P. F., Vera Cruz, E. M., and Santos, M. D. (2024). *In silico* characterization and homology modeling of Nile tilapia (*Oreochromis niloticus*) Hsp70cBi and Hsp70cBc proteins. *Heliyon* 10, e32748. doi: 10.1016/j.heliyon.2024.e32748

Di Lellis, M. A., Sereda, S., Geißler, A., Picot, A., Arnold, P., Lang, S., et al. (2014). Phenotypic diversity, population structure and stress protein-based capacitoring in populations of *Xeropicta derbentina*, a heat-tolerant land snail species. *Cell. Stress Chaperones* 19, 791–800. doi: 10.1007/s12192-014-0503-x

El-Gebali, S., Mistry, J., Bateman, A., Eddy, S. R., Luciani, A., Potter, S. C., et al. (2019). The Pfam protein families database in 2019. *Nucleic Acids Res.* 47, 427–432. doi: 10.1093/nar/gky995

Finn, R. D., Clements, J., and Eddy, S. R. (2011). HMMER web server: interactive sequence similarity searching. *Nucleic Acids Res.* 39, 29–37. doi: 10.1093/nar/gkr367

Gao, L., Yuan, Z., Li, Y., and Ma, Z. (2022). Genome-wide comparative analysis of DNAJ genes and their co-expression patterns with HSP70s in aestivation of the sea cucumber *Apostichopus japonicus*. *Funct. Integr. Genomics* 22, 317–330. doi: 10.1007/s10142-022-00830-2

Gao, L., Yuan, Z., Yu, S., Yang, Y., Li, Y., and He, C. (2018). Genome-wide identification of HSP70/110 genes in sea cucumber *Apostichopus japonicus* and comparative analysis of their involvement in aestivation. *Comp. Biochem. Physiol. Part. D. Genomics Proteomics* 28, 162–171. doi: 10.1016/j.cbd.2018.07.005

Gasteiger, E., Hoogland, C., Gattiker, A., Duvaud, S., Wilkins, M., Appel, R., et al. (2005). "Protein identification and analysis tools on the exPASy server," in *The Proteomics Protocols Handbook*. Ed. J. M. Walker (Totowa, NJ, USA: Humana Press). doi: 10.1385/1-59259-890-0:571

González-Guillén, A. (2021). *Polymita Uncoiled* (Collierville, Memphis: Fundcraft Publishing).

Guindon, S., Dufayard, J.-F., Lefort, V., Anisimova, M., Hordijk, W., and Gascuel, O. (2010). New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst. Biol.* 59, 307–321. doi: 10.1093/sysbio/syq010

Gundappa, M. K., Peñaloza, C., Regan, T., Boutet, I., Tanguy, A., Houston, R. D., et al. (2022). Chromosome-level reference genome for European flat oyster (*Ostrea edulis* L.). *Evol. Appl.* 15, 1713–1729. doi: 10.1111/eva.13460

Han, Z., Truong, Q. A., Park, S., and Breslow, J. L. (2003). Two Hsp70 family members expressed in atherosclerotic lesions. *Proc. Natl. Acad. Sci.* 100, 1256–1261. doi: 10.1073/pnas.252764399

Hasnain, P., and Kaneko, G. (2022). Phylogenetic annotation of *Caenorhabditis elegans* heat shock protein 70 genes. *microPublication Biol.* 2022. doi: 10.17912/micropub.biology.000633

He, J., Han, X., Sun, S., Jin, S., Liu, M., and Han, Z. (2024). Genome-wide identification and transcriptome analysis of the hsp70 gene family in *monodonta labio* reveals its role in response to nanoplastics stress. *Genes* 15, 291. doi: 10.3390/genes15030291

Hoang, D. T., Chernomor, O., Von Haeseler, A., Minh, B. Q., and Vinh, L. S. (2018). UFBoot2: improving the ultrafast bootstrap approximation. *Mol. Biol. Evol.* 35, 518–522. doi: 10.1093/molbev/msx281

Hu, Z., Song, H., Feng, J., Zhou, C., Yang, M.-J., Shi, P., et al. (2022). Massive heat shock protein 70 genes expansion and transcriptional signatures uncover hard clam adaptations to heat and hypoxia. *Front. Mar. Sci.* 9, 898669. doi: 10.3389/fmars.2022.898669

Ikai, A. (1980). Thermostability and aliphatic index of globular proteins. *J. Biochem.* 88, 1895–1898. doi: 10.1093/oxfordjournals.jbchem.a133168

Jahan, K., Nie, H., and Yan, X. (2023). Revealing the potential regulatory relationship between HSP70, HSP90 and HSF genes under temperature stress. *Fish. Shellfish Immunol.* 134, 108607. doi: 10.1016/j.fsi.2023.108607

Kalyaanamoorthy, S., Minh, B. Q., Wong, T. K. F., Von Haeseler, A., and Jermiin, L. S. (2017). ModelFinder: fast model selection for accurate phylogenetic estimates. *Nat. Methods* 14, 587–589. doi: 10.1038/nmeth.4285

Kaneko, G. (2022). Phylogenetic annotation of *Drosophila melanogaster* heat shock protein 70 genes. *microPublication Biol.* 2022. doi: 10.17912/micropub.biology.000558

Katoh, K., and Standley, D. (2013). MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol.* 30, 772–780. doi: 10.1093/molbev/mst010

Kim, J., Kim, H. J., Choi, E., Cho, M., Choi, S., Jeon, M. A., et al. (2024). Expansion of the HSP70 gene family in *Tegillarca granosa* and expression profiles in response to zinc toxicity. *Cell. Stress Chaperones* 29, 97–112. doi: 10.1016/j.cstres.2024.01.004

Kuzmin, E., Taylor, J. S., and Boone, C. (2022). Retention of duplicated genes in evolution. *Trends Genet. TIG* 38, 59–72. doi: 10.1016/j.tig.2021.06.016

Lewis, A. M., Reyes-Tur, B., and Davison, A. (2025). The mitochondrial genome and phylogenetic placement of the world's most beautiful snail, *Polymita picta*. *J. Molluscan Stud.* 91, eyaf006. doi: 10.1093/mollus/eyaf006

Lu, H., Liu, C., Yang, C., He, Z., Wang, L., and Song, L. (2024). Genome-wide identification of the HSP70 genes in Pacific oyster *Magallana gigas* and their response to heat stress. *Cell. Stress Chaperones* 29, 589–602. doi: 10.1016/j.cstres.2024.06.002

Lu, S., Wang, J., Chitsaz, F., Derbyshire, M. K., Geer, R. C., Gonzales, N. R., et al. (2020). CDD/SPARCLE: the conserved domain database in 2020. *Nucleic Acids Res.* 48, 265–268. doi: 10.1093/nar/gkz991

Ma, F., and Luo, L. (2020). Genome-wide identification of Hsp70/110 genes in rainbow trout and their regulated expression in response to heat stress. *PeerJ* 8, e10022. doi: 10.7717/peerj.10022

Maceira-Filgueira, D. (2016). *Polymita*. Eds. M. E. J. Hidalgo-Gato and R. Rodríguez-León (La Habana: Libro rojo de invertebrados terrestres de Cuba).

Melikov, A., and Novak, P. (2024). Heat shock protein network: the mode of action, the role in protein folding and human pathologies. *Folia Biologica (00155500)* 70, 152–165. doi: 10.14712/fb2024070030152

Minh, B. Q., Schmidt, H. A., Chernomor, O., Schrempf, D., Woodhams, M. D., Von Haeseler, A., et al. (2020). IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. *Mol. Biol. Evol.* 37, 1530–1534. doi: 10.1093/molbev/msaa015

Nguyen, L.-T., Schmidt, H. A., Von Haeseler, A., and Minh, B. Q. (2015). IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol. Biol. Evol.* 32, 268–274. doi: 10.1093/molbev/msu300

Nicolai, A., and Ansart, A. (2017). Conservation at a slow pace: terrestrial gastropods facing fast-changing climate. *Conserv. Physiol.* 5, cox007. doi: 10.1093/conphys/cox007

Noble, D. W., Kar, F., Bush, A., Seebacher, F., and Nakagawa, S. (2025). Limited plasticity but increased variance in physiological rates across ectotherm populations under climate change. *Funct. Ecol.* 39, 1176–1193. doi: 10.1111/1365-2435.70031

Pagès, H., Aboyoun, P., Gentleman, R., and DebRoy, S. (2026). Biostrings: Efficient manipulation of biological strings. *R. Package version 2.80.1*. doi: 10.18129/B9.bioc.Biostrings

Quinlan, A. R., and Hall, I. M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* 26, 841–842. doi: 10.1093/bioinformatics/btq033

R Core Team. (2022). *R: a language and environment for statistical computing* Vienna, Austria: R Foundation for Statistical Computing.

Reyes-Tur, B., Chen, Z., Gordillo-Pérez, M. J., Hamadou, A. B., Gerheim, C., Greve, C., et al. (2025). Draft genome of the Cuban painted land snail *Polymita picta*, International Mollusc of the Year 2022. *BMC Genomic Data* 26, 63. doi: 10.21203/rs.3.rs-6464041/v1

Roberts, S., and Feder, M. (2000). Changing fitness consequences of hsp70 copy number in transgenic *Drosophila* larvae undergoing natural thermal stress. *Funct. Ecol.* 14, 353–357. doi: 10.1046/j.1365-2435.2000.00429.x

Rosenzweig, R., Nillegoda, N. B., Mayer, M. P., and Bukau, B. (2019). The Hsp70 chaperone network. *Nat. Rev. Mol. Cell Biol.* 20, 665–680. doi: 10.1038/s41580-019-0133-3

Scheil, A. E., Gärtner, U., and Köhler, H.-R. (2012). Colour polymorphism and thermal capacities in *Theba pisana* (OF Müller 1774). *J. Therm. Biol.* 37, 462–467. doi: 10.1016/j.jtherbio.2012.03.006

Schweizer, M., Triebskorn, R., and Köhler, H. R. (2019). Snails in the sun: Strategies of terrestrial gastropods to cope with hot and dry conditions. *Ecol. Evol.* 9, 12940–12960. doi: 10.1002/ece3.5607

Shao, H., Taguwa, S., Gilbert, L., Shkedi, A., Sannino, S., Guerriero, C. J., et al. (2022). A campaign targeting a conserved Hsp70 binding site uncovers how subcellular localization is linked to distinct biological activities. *Cell Chem. Biol.* 29, 1303–1316.e3. doi: 10.1016/j.chembiol.2022.06.006

Singh, M. K., Han, S., Ju, S., Ranbhise, J. S., Ha, J., Yeo, S. G., et al. (2025). Hsp70: a multifunctional chaperone in maintaining proteostasis and its implications in human disease. *Cells* 14, 509. doi: 10.3390/cells14070509

Staikou, A., Sagonas, K., Spanoudi, O., Savvidou, K., Nazli, Z., Feidantsis, K., et al. (2024). Activities of antioxidant enzymes and Hsp levels in response to elevated temperature in land snail species with varied latitudinal distribution. *Comp. Biochem. Physiol. Part. B. Biochem. Mol. Biol.* 269, 110908. doi: 10.1016/j.cbpb.2023.110908

Suyama, M., Torrents, D., and Bork, P. (2006). PAL2NAL: robust conversion of protein sequence alignments into the corresponding codon alignments. *Nucleic Acids Res.* 34, W609–W612. doi: 10.1093/nar/gkl315

Thorvaldsdóttir, H., Robinson, J. T., and Mesirov, J. P. (2013). Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Briefings Bioinf.* 14, 178–192. doi: 10.1093/bib/bbs017

Thumuluri, V., Almagro Armenteros, J. J., Johansen, A. R., Nielsen, H., and Winther, O. (2022). DeepLoc 2.0: multi-label subcellular localization prediction using protein language models. *Nucleic Acids Res.* 50, 228–234. doi: 10.1093/nar/gkac278

Tull, T., Schmidt, D., and Köhler, H.-R. (2020). Role of ambient wavelength, shell size and pigmentation intensity in the heating of Mediterranean *Theba pisana* (Eupulmonata: Helicidae). *J. Molluscan Stud.* 86, 249–253. doi: 10.1093/mollus/eyaa006

Wang, G., Chen, L., Pan, H., Pang, A., Shi, Y., Chen, S., et al. (2025). The high-quality genome of *Bellamya aeruginosa* reveals the molecular mechanisms underlying its remarkable stress resistance. *BMC Biol.* 23, 285. doi: 10.1186/s12915-025-02401-w

Wang, W., Liu, Q., Liu, Q., and Hendrickson, W. A. (2021). Conformational equilibria in allosteric control of Hsp70 chaperones. *Mol. Cell* 81, 3919–3933, e7. doi: 10.1016/j.molcel.2021.07.039

Wentink, A., Rosenzweig, R., Kampinga, H., and Bukau, B. (2026). Mechanisms and regulation of the Hsp70 chaperone network. *Nat. Rev. Mol. Cell Biol.* 27, 110–128. doi: 10.1038/s41580-025-00890-9

Wu, X., Yang, W., Chen, R., Chen, S., Sun, Q., Ji, Q., et al. (2025). Comprehensive genomic profiling and expression dynamics of the HSP gene family in *Cipangopaludina chinensis* under temperature stress. *Genomics Commun.* 2, e009. doi: 10.48130/gcomm-0025-0009

Xu, K., Xu, H., and Han, Z. (2018). Genome-wide identification of Hsp70 genes in the large yellow croaker (*Larimichthys crocea*) and their regulated expression under cold and heat stress. *Genes* 9, 590. doi: 10.3390/genes9120590

Yanai, I. (2022). Quantifying gene duplication. *Nat. Rev. Genet.* 23, 196–197. doi: 10.1038/s41576-022-00457-w

Yang, S., Zhang, X., Yue, J.-X., Tian, D., and Chen, J.-Q. (2008). Recent duplications dominate NBS-encoding gene expansion in two woody species. *Mol. Genet. Genomics* 280, 187–198. doi: 10.1007/s00438-008-0355-0

Yu, E., Yoshinaga, T., Jalufka, F. L., Ehsan, H., Mark, D. B., and Kaneko, G. (2021). The complex evolution of the metazoan HSP70 gene family. *Sci. Rep.* 11, 17794. doi: 10.1038/s41598-021-97192-9

Zhang, Z. (2022). KaKs_Calculator 3.0: calculating selective pressure on coding and non-coding sequences. *Genomics Proteomics Bioinf.* 20, 536–540. doi: 10.1016/j.gpb.2021.12.002

Zhang, P., Leu, J. I., Murphy, M. E., George, D. L., and Marmorstein, R. (2014). Crystal structure of the stress-inducible human heat shock protein 70 substrate-binding domain in complex with peptide substrate. *PloS One* 9, e103518. doi: 10.1371/journal.pone.0103518

Zheng, W., Xu, X., Chen, Y., Wang, J., Zhang, T., E, Z., et al. (2023). Genome-wide identification, molecular characterization, and involvement in response to Abiotic and Biotic stresses of the HSP70 gene family in Turbot (*Scophthalmus maximus*). *Int. J. Mol. Sci.* 24, 6025. doi: 10.3390/ijms24076025