Journal of Applied Biology & Biotechnology 2026. Article in Press
Available online at http://www.jabonline.in
DOI: 10.7324/jabb.2026.311545

In silico characterization of a metatranscriptome-derived stress-responsive putative DnaJ/HSP40-domain-containing fragment associated with heavy metal homeostasis

Priyanka Sharma , Bhupendra Narayan Singh Yadav , Shristy Maurya , Rajiv Kumar Yadav*
Molecular Biology and Genetic Engineering Laboratory, Department of Botany, Faculty of Science, University of Allahabad, Prayagraj, India.

ARTICLE INFO

Article history:
Received on: 23 February 2026
Accepted on: 26 June 2026
Available Online: XX

Key words:
RNA-Seq, soil metatranscriptome, heavy metal, *in silico* characterization, heat shock proteins, molecular chaperones

ABSTRACT

Heavy metal (HM)-contaminated soils impose severe stress on the environment, yet certain soil microbes exhibit remarkable tolerance, revealing adaptive mechanisms in stress-adapted communities. Micro-eukaryotic communities in such environments exhibit substantial transcriptional reprogramming to mitigate stress-induced cellular damage. This study explores the molecular basis of HM stress adaptation based on transcript-level evidence using RNA-Seq-based metatranscriptomics combined with comprehensive *in silico* characterization. Exp_525919, a highly expressed transcript, was annotated against the UniProt proteome database using BLAST and was predicted as an uncharacterized protein from *Achlya hypogyna (ACHHYP_04605),* showing 57% sequence identity and a significant e-value (7.6e-24). Integrative analyses, including functional annotation, structural modeling, domain architecture, and evolutionary conservation, suggest that Exp_525919 is a putative DnaJ/heat shock protein 40 (HSP40)-domain-containing fragment, potentially associated with cellular protein quality control processes under HM stress conditions. Secondary structure revealed a predominance of coil regions with interspersed β-strands and a central α-helix. GO annotation predicted potential associations with transport, nucleic acid binding, and ribosomal association. Structural modeling generated a high-confidence model (96% coverage, 99.5% confidence), consistent with predicted chaperone-like proteins. Protein–protein interaction analysis revealed that the predicted uncharacterized protein ACHHYP_04605 is closely associated with major molecular chaperones, including HSP70, DnaJ/HSP40, and HSP90, suggesting its involvement in HSP70-centered stress-responsive networks. Network clustering revealed a molecular chaperone architecture in which core protein folding components were identified as central network hubs potentially associated with stress-adaptive responses. Overall, RNA-Seq-based metatranscriptomic analysis identified transcriptionally responsive potential genes and pathways underlying HM tolerance in micro-eukaryotic communities and provided a robust computational foundation for future experimental validation.

1. INTRODUCTION

Heavy metals (HMs) constitute a class of persistent xenobiotics of significant concern due to their non-biodegradable nature and long-term ecological impacts [1]. Rapid industrialization has increased HM contamination in soils through mining, electroplating, leather tanning, textile processing, and improper waste disposal, leading to heightened toxicity and disruption of soil functions [2]. Elevated HM levels disturb cellular homeostasis, suppress growth, and trigger oxidative stress across biological systems. Despite these adverse effects, certain microorganisms demonstrate remarkable tolerance, enabling them to survive and persist in heavily contaminated environments [3].

Microbial tolerance to HMs is mediated by diverse adaptive mechanisms, including extracellular metal chelation, restricted cellular uptake, intracellular sequestration, efflux transport systems, and detoxification of reactive oxygen species [4]. At the molecular level, stress-responsive proteins, particularly heat shock proteins (HSPs), play a vital role in maintaining protein stability, preventing aggregation, and promoting cellular recovery under metal-induced stress [5]. Hexavalent chromium [Cr(VI)] is recognized as a priority environmental pollutant due to its high solubility, mobility, and carcinogenic nature [6]. Industrial effluents enriched with Cr(VI) readily contaminate soils and groundwater, posing serious threats to soil microbial communities and overall ecosystem functioning [7]. Therefore, understanding microbial adaptive responses to chromium stress is crucial for developing effective and sustainable bioremediation strategies.

Extensive studies have been conducted on prokaryotic systems, whereas micro-eukaryotes, an ecologically significant yet understudied component of soil microbiomes, remain poorly characterized, particularly in contaminated environments [8].

*Corresponding Author
Rajiv Kumar Yadav, Molecular Biology and Genetic Engineering Laboratory, Department of Botany, Faculty of Science, University of Allahabad, Prayagraj, India. E-mail: rky@allduniv.ac.in; rajiv02m@gmail.com

Their role in trophic interactions, nutrient cycling, and maintaining ecosystem stability highlights the importance of understanding their responses to HM stress [9]. However, methodological constraints have historically limited their investigation. Advances in high-throughput sequencing have revolutionized soil microbiome research, enabling comprehensive analyses of both taxonomic composition and functional activity [8,10,11]. In particular, metatranscriptomics provides direct insight into actively expressed genes, offering a powerful approach for elucidating stress-response pathways that cannot be captured through DNA-based methods alone [12,13].

By selectively targeting eukaryotic mRNA, RNA sequencing (RNA-Seq)-based metatranscriptomics facilitates the identification of novel stress-responsive genes with potential biotechnological applications [14]. Biocatalysts derived from metal-contaminated environments, possessing unique functional traits, can act as biomarkers or be utilized in bioremediation strategies, thereby contributing to the development of metal-tolerant crop plants [15]. In this study, an uncharacterized protein identified through RNA-Seq-based metatranscriptomic analysis of Cr-contaminated soil was analyzed *in silico* to elucidate its potential role in heavy-metal homeostasis. Particular emphasis was placed on its involvement in stress-response pathways, including HSPs, to better understand the molecular mechanisms underlying microbial adaptation to metal-induced stress.

2. MATERIAL AND METHODS

2.1. Soil Sampling, Chromium Treatment, RNA Isolation

Soil samples were obtained from agroforestry fields in Kanpur, Uttar Pradesh, India. To improve field representativeness for composite sampling, four independent sites were selected. At each site, twenty soil cores (0–10 cm depth) were randomly collected and combined in equal proportions to form a composite sub-sample (CSS). The collected samples were sieved to remove roots and debris, transported to the laboratory at 4°C, and designated as CSS1–CSS4. Physicochemical properties of the soil were assessed by measuring moisture content, pH, electrical conductivity, organic carbon [16], total nitrogen [17], phosphorous [18], and potassium [19]. Micronutrients such as Copper (Cu), Iron (Fe), Manganese (Mn), and Zinc (Zn) were quantified using flame atomic absorption spectrometry (AAS; PinAAcle 500, PerkinElmer, USA), whereas HMs such as Cadmium (Cd), Lead (Pb), and Chromium (Cr) were analyzed by inductively coupled plasma mass spectrometry (ICP-MS; Agilent 7900, USA). Cr(VI) was determined using the USEPA 3060A alkaline digestion method [20].

Equal amounts of CSS1–CSS4 were pooled to obtain a composite soil sample, which was subsequently divided into control (KC) and chromium-treated (KM) groups. For treatment, 30 g of soil was amended with 10 ml of 0.85% NaCl containing 2-mM potassium dichromate ($K_2Cr_2O_7$), whereas control samples received NaCl alone. The samples were incubated at room temperature for 30 minutes to induce transcriptional responses [21]. Total RNA was extracted using the RNeasy PowerSoil Total RNA Isolation Kit (Qiagen, Germany), followed by DNase I treatment. To enhance transcript recovery and reduce technical bias, each sample was extracted in triplicate and subsequently pooled. RNA quality and quantity were evaluated using agarose gel electrophoresis, NanoDrop ND-200, Qubit RNA HS assay, and Agilent TapeStation 2200 before metatranscriptomics sequencing.

2.2. mRNA Enrichment, Library Preparation, and Metatranscriptome Sequencing

High-quality total RNA samples were subjected to mRNA enrichment using the NEBNext Poly(A) mRNA Magnetic Isolation Module (New England Biolabs, USA), following the manufacturer's instructions. Briefly, 1000 ng of total RNA was subjected to poly(A)+ mRNA enrichment using the NEBNext Poly (A) mRNA Magnetic Isolation Module. The enriched mRNA was processed directly according to the manufacturer's protocol, proceeding through fragmentation and first-strand cDNA synthesis without intermediate quantification. Libraries were prepared using the NEBNext® Ultra™ II Directional RNA Library Prep Kit (New England Biolabs, USA), PCR-amplified, purified, and quantified with a Qubit fluorometer using the Qubit dsDNA HS assay kit. Briefly, first- and second-strand cDNA synthesis were performed via reverse transcription, and the resulting double-stranded cDNA was purified using NEBNext Sample Purification Beads. End repair, 3′ adenylation, and ligation of Illumina-compatible adapters were carried out as per the manufacturer's protocol. The libraries were prepared using a dual-index barcoding strategy. Sample KC was assigned Index 1 (i7): TTCCAGGT; Index 2 (i5): CAGTGCTT, corresponding to barcodes S736 and S543, respectively. Sample KM was assigned Index 1 (i7): TACGGTCT; Index 2 (i5): TCCATTGC, corresponding to barcodes S709 and S575, respectively. Final libraries were purified, quantified using a Qubit fluorometer, and assessed for fragment size distribution using the Agilent 2200 TapeStation.

Paired-end metatranscriptome sequencing (2 × 150 bp) was performed on the Illumina NovaSeq 6000 platform (S1 flow cell), generating approximately 25 million reads per sample. Raw reads were quality-checked using FastQC v0.11.8 [22], and adapter and low-quality sequences were removed using Trim Galore v0.4.01 [23]. Only reads with Phred scores ≥ Q30 were retained. High-quality reads were assembled *de novo* using MetaSPAdes v3.15.2 [24], and contigs ≤ 200 bp were filtered using PRINSEQ v0.20.4 [25]. The raw sequencing data have been deposited in the NCBI Sequence Read Archive under BioProject accession number PRJNA1177600.

2.3. Comparative Transcript-Abundance Analysis

Assembled contigs were clustered into representative transcripts using CD-HIT v4.6 with a sequence similarity threshold of ≥ 0.95 to minimize redundancy [26]. The clustered sequences from each sample were subsequently merged to generate a final non-redundant transcriptome assembly, which was further used as a reference for read alignment and expression quantification. Processed reads from each sample were aligned to this reference assembly using Bowtie2 [27], and gene-level read counts were obtained from the aligned datasets. Comparative assessment of gene expression was performed using the DESeq R package [28], with read counts normalized by the library-size-based method implemented in DESeq. Transcripts with $\log_2$ fold change ($\log_2$FC) ≥ 1.5 were classified as comparatively upregulated in the treated sample, whereas those with $\log_2$FC ≤ −1.5 were classified as comparatively downregulated, and transcripts with $\log_2$FC values between −1.5 and 1.5 as relatively stable transcript abundance. The $\log_2$FC threshold (±1.5) was used as an arbitrary exploratory cutoff for comparative transcript-abundance analysis.

Upregulated uncharacterized transcripts were shortlisted for further functional characterization. Contigs ≥ 200 bp were retained for downstream analysis and subjected to gene prediction using MetaGeneMark v3.25 [29]. The predicted gene sequences were annotated against the UniProt Proteome database using the DIAMOND BLAST algorithm [30], and top-scoring hits were used to assign UniProt identifiers. The identifiers were subsequently used to retrieve associated metadata, including gene names, protein names, source organisms, pathway information, and Gene Ontology (GO) terms such as molecular function, biological process, and cellular component.

2.4. *In silico* Characterization of the Selected Sequence

In particular, one predicted uncharacterized protein was selected for *in silico* characterization from the pool of upregulated uncharacterized proteins having a putative role in HM homeostasis. The selection was based on its predicted molecular functions, including ATP binding, HSP70 protein binding, metal ion binding, and unfolded protein binding. Open reading frames (ORFs) were identified from the uncharacterized protein sequence (Exp_525919) using the NCBI ORF finder tool (https://www.ncbi.nlm.nih.gov/orffinder) [31]. Among the predicted ORFs, the longest ORF located on the positive-sense strand was selected for downstream sequence annotation and analysis. Conserved motif identification was performed using MotifFinder (https://www.genome.jp/tools/motif/) [32] to detect functionally relevant sequence patterns. The selected ORF was further analyzed for secondary structure prediction, membrane topology, and functional features of the corresponding protein using the PSIPRED server (https://bioinf.cs.ucl.ac.uk/psipred/) [33]. Structural modeling and homology/analogy-based fold recognition were carried out using the Phyre2.2 server (https://www.sbg.bio.ic.ac.uk/~phyre2/html/page.cgi?id=index) [34] and AlphaFold (https://alphafold.com/) [35]. Protein family classification and conserved domain analysis were conducted using InterProScan (https://www.ebi.ac.uk/interpro/about/interproscan/) [36], which identified the InterPro entry IPR036869. To assess evolutionary conservation, the ORF was queried against the OrthoDB v12.2 (https://www.orthodb.org/) [37] database using *Achlya hypogyna* as the reference organism. The hypothetical protein ACHHYP_04605 was selected for subsequent analyses, including group hierarchy assessment and functional prediction. Finally, protein–protein interaction analysis of the selected ORF from the organism *A. hypogyna* (ACHHYP_04605) was performed using the STRING database v12.0 (https://string-db.org/) [75] to explore potential functional associations and was visualized using Cytoscape v3.10.4 (https://cytoscape.org/) [38].

3. RESULTS AND DISCUSSION

Microbial-mediated reduction is a key strategy for remediating Cr-contaminated sites, converting toxic Cr(VI) to the less mobile Cr(III) [39]. Cr-tolerant microbes like *Aspergillus flavus*, *Cyberlindnera fabianii, Wickerhamomyces anomalus*, and *Candida tropicalis* [40–42] survive under high Cr(VI) stress and actively contribute to detoxification by enzymatically reducing toxic Cr(VI) to the less toxic Cr(III) form, along with processes such as biosorption and immobilization. Thus, microbial tolerance is a critical functional trait that underpins effective chromium bioremediation.

RNA-Seq provides a powerful approach for comprehensive profiling of actively expressed transcripts in complex microbiomes under defined environmental conditions, thereby enabling functional interpretation of microbial responses to environmental stressors [43]. In the present study, RNA-Seq was employed to investigate transcriptional responses of micro-eukaryotic communities inhabiting chromium-contaminated soil following short-term exposure to toxic Cr(VI), with a focus on the *in-silico* characterization of a metatranscriptome-derived heat shock-like protein.

3.1. Soil Sample Characteristics and Metatranscriptomics Sequencing Overview

Soil samples were collected from Kanpur, Uttar Pradesh, India. Tannery wastewater is a well-documented source of chromium contamination, and long-term percolation of untreated or partially treated effluents has led to the accumulation of HMs in the soil profile [44]. Consistent with this, heavy-metal analysis of the sampling site revealed elevated concentrations of cadmium (28.27 mg/kg), lead (48.36 mg/kg), total chromium (1,074.12 mg/kg), and hexavalent chromium (30.01 mg/kg), indicating substantial heavy-metal contamination of the sampling site [45]. The exceptionally high total chromium concentration strongly implicates tannery-derived effluents as the primary contamination source. Owing to the deteriorated soil quality, land use at the site has shifted from food crops to the cultivation of shrubby ornamental plants [46]. The soil sample was classified as sandy loam, and its physicochemical properties are summarized in Table 1. Elevated levels of toxic metals confirmed the contaminated status of the site and provided the environmental context for evaluating chromium-induced transcriptional responses. High-quality RNA was obtained from both control and Cr(VI)-treated samples, with 260/230 and 260/280 absorbance ratios of ~2.03 and ~1.89, respectively, in both samples. The RNA integrity number equivalent (RINe) values were above 8, indicating suitability for metatranscriptomic analysis.

The final library concentrations were 1.93 ng/µl (KC) and 1.11 ng/µl (KM), yielding a total of 19.3 ng and 11.1 ng library DNA, respectively, in the final 10-µl volume. Equimolar pooling of the libraries was performed, and the pooled library was normalized to 4 nM. Following denaturation and dilution according to the Illumina NovaSeq 6000 protocol, the library was diluted to a final loading concentration of 150 pM and loaded onto the flow cell for paired-end sequencing.

Paired-end Illumina sequencing (2 × 150 bp) of mRNA-derived cDNA libraries produced approximately 27.5 million reads for the KC sample and 21.1 million reads for the KM sample. After quality trimming with TrimGalore using a Phred score cutoff of 30, 26.9 million and 20.7 million high-quality reads were retained for KC and KM, respectively. The percentage of adapter and low-quality base removal was 1.98% in KC and 1.99% in KM. *De novo* assembly using MetaSPAdes generated 3,716,782 contigs for KC and 2,368,668 contigs for KM after removing contigs ≤ 200 bp. Further clustering to remove redundancy resulted in 3,702,106 representative contigs for KC and 2,352,027 contigs for KM.

Table 1. Physicochemical properties of the composite soil sample.

Physicochemical parameter	Compound soil sample [Mean ± SD (n = 4)]
Moisture content (%)	11.34 ± 0.74
pH	7.53 ± 0.66
Electrical conductivity (mS/cm)	1.13 ± 0.85
Organic carbon (%)	0.64 ± 0.07
Nitrogen (%)	0.51 ± 0.03
Phosphorus (mg/kg)	69.00 ± 4.16
Potassium (mg/kg)	36.50 ± 7.19
Copper (mg/kg)	31.95 ± 5.62
Iron (mg/kg)	55.20 ± 7.84
Manganese (mg/kg)	24.70 ± 6.63
Zinc (mg/kg)	111.65 ± 16.68
Cadmium (mg/kg)	28.27 ± 8.65
Lead (mg/kg)	48.36 ± 9.69
Chromium (mg/kg)	1074.12 ± 176.76
Chromium (VI) (mg/kg)	30.01 ± 4.26

3.2. Transcript-Abundance Patterns Under Cr (VI) Exposure

In the present study, comparative assessment of gene expression was employed as an exploratory approach to identify transcriptional trends and putative functional pathways associated with Cr(VI) exposure. As the analysis was based on a single pooled control and a single pooled Cr(VI)-treated library without biological replication, the results represent comparative transcript abundance patterns rather than statistically robust differential expression. The comparative metatranscriptomic analysis, therefore, provides a preliminary snapshot of the overall early transcriptional response of the soil microbial community in response to Cr(VI) exposure, highlighting candidate genes and pathways for future validation. A total of 145,209 transcripts were comparatively upregulated and 208,442 were comparatively downregulated in the treated sample relative to the control, while 645,689 transcripts had relatively stable transcript abundance. The predominance of downregulated transcripts suggests a potential inhibitory effect of chromium toxicity on gene expression, whereas differences in contig numbers may be attributed to assembly-related factors. These inhibitory effects of Cr(VI) are consistent with its strong oxidizing nature and impact on cellular processes.

Conversely, the substantial number of upregulated transcripts indicates activation of adaptive and stress-response mechanisms within the micro-eukaryotic community. Several of the upregulated proteins identified in this study are directly implicated in HM homeostasis. Notably, these include HSPs [47], chromate transporters [48], ATP-binding cassette (ABC) transporters [49], HM tolerance proteins [50], sulfate transporters [51], P-type ATPases [52], metallothioneins [53], aldehyde dehydrogenases [54], superoxide dismutase (SOD) [55], cation diffusion facilitator transporters [56], oxidative stress-associated Svf1-like proteins [57], and the metal transporter NRAMP1 [58,59] (Supplementary Material 1). These observed transcript-abundance patterns may be consistent with the involvement of processes associated with detoxification, stress tolerance, and metal transport, as well as molecular chaperones and cellular protection, enabling survival under chromium-induced stress. Collectively, these findings indicate that short-term Cr(VI) exposure induces a substantial shift in the metatranscriptomic landscape of soil micro-eukaryotes, underscoring their sensitivity to HM contamination and their potential role as bioindicators of soil health and environmental stress.

3.3. Metatranscriptomic Evidence for Chaperone-Driven Stress Adaptation

Molecular chaperones constitute a complex client-regulatory network that is essential for preventing protein misfolding and abnormal aggregation, maintaining protein homeostasis, and safeguarding cells against damage under continuously fluctuating environmental conditions [60]. These molecular chaperones, commonly known as HSPs, are proteins that interact with, stabilize, or assist other proteins in achieving their native functional conformations [61,62]. Under stress conditions, HSPs facilitate protein folding and maturation while coordinating protein quality control and degradation pathways, functions that are vital for maintaining cellular proteostasis [60,63,64].

In this study, transcripts associated with molecular chaperone systems exhibited increased abundance under HM stress. RNA-Seq-based metatranscriptomic analysis revealed the upregulation of 1,069 transcripts associated with HSPs (Supplementary Material 2), indicating a widespread transcriptional activation of chaperone-mediated stress response pathways. Several of these HSPs were of microbial origin. For instance, Exp_735170 corresponds to heat shock 70-kDa protein (TTHERM_00105110) from *Tetrahymena thermophila* (Alveolata, SAR clade) and exhibited a ~20-fold increase in expression. Similarly, Exp_409769, annotated a heat shock protein Hsp90 family protein (*DLAC_04639*) from *Tieghemostelium lacteum* (Amoebozoa) showed a 6-fold increase, while Exp_236413, predicted heat shock protein HSP90 (TTHERM_00080030) from *Tetrahymena thermophila* (Alveolata, SAR clade) showed a 4-fold increase in expression.

The pronounced upregulation of HSPs highlights their protective role in maintaining protein homeostasis under adverse conditions. As molecular chaperones, HSP70 and HSP90 ensure proper protein folding and prevent aggregation during HM, oxidative, and thermal stress [47,65]. Heat shock proteins are conserved across all domains of life and are classified according to molecular size. Their expression is triggered by a variety of stressors, including HMs, where they function not only in normal protein folding and assembly but also in the repair and protection of stress-damaged proteins [66].

The predicted uncharacterized protein-coding genes represent an unexplored component of the metatranscriptome with potential roles in cellular adaptation. A total of 4,721 uncharacterized genes were upregulated (Supplementary Material 3). Several of these genes of micro-eukaryotic origin displayed predicted molecular functions of their corresponding protein, such as metal ion binding, HSP70 binding, unfolded protein binding, and chaperone binding, as listed in Supplementary Material 3. Furthermore, uncharacterized proteins possessing ATP-binding and ATPase activities may contribute to HM homeostasis, as ATP-dependent chaperones and ATPases are central to maintaining cellular proteostasis under stress conditions. HM exposure induces protein misfolding and oxidative damage, thereby activating ATP-dependent protein folding machinery, including molecular chaperones such as HSP70 and HSP90 [67,68]. AAA+ ATPases and related ATP-binding proteins further contribute to stress adaptation by facilitating protein quality control, unfolding, and degradation pathways essential for cellular survival under metal stress [69]. The high abundance and strong responsiveness of these uncharacterized proteins suggest their potential involvement in cellular mechanisms of HM tolerance.

3.4. *In silico* Analysis of Exp_525919

Among the predicted upregulated uncharacterized transcripts, the transcript Exp_525919 was selected for *in silico* characterization because of its potential involvement in HM homeostasis. The annotated Exp_525919 transcript corresponded to the *ACHHYP_04605* gene from *A. hypogyna,* with a relatively high sequence identity (57%) and a significant e-value (7.6e-24). Importantly, functional predictions of the encoded protein indicated its involvement in key stress-associated molecular functions, including ATP binding, HSP70 protein binding, metal ion binding, and unfolded protein binding, all of which are associated with HM homeostasis and chaperone-mediated proteostasis. These features, along with its uncharacterized status, made Exp_525919 a particularly compelling candidate for in-depth analysis among the upregulated genes.

Sequence analysis indicated that Exp_525919 is 274 nucleotides long and contains two open reading frames (ORFs) of 213 and 75 nucleotides. The longer ORF (213 nucleotides), encoding a 70 amino acid polypeptide, was chosen for subsequent analyses due to its higher coding potential. Domain analysis of this ORF using the Pfam database identified a conserved DnaJ C-terminal domain (PF01556). The DnaJ_C domain was located between amino acid positions 2 and 67, with a highly significant independent *E*-value of 3.4×10^{-10},

strongly supporting its classification as a member of the HSP40 (DnaJ) co-chaperone family. Due to its short length, Exp_525919 is likely to represent a partial domain or a divergent minimal co-chaperone. Its size is more consistent with a DnaJ C-terminal/peptide-binding domain-containing fragment. Nevertheless, small J-domain-containing proteins or domain fragments have been reported to retain partial functional relevance, although their precise roles often differ from canonical HSP40 proteins [70].

3.5. Secondary Structure Prediction and Functional Annotation

PSIPRED-based analysis of the selected ORF of the uncharacterized transcript (Exp_525919) showed that the predicted protein is predominantly composed of coil regions, interspersed with multiple β-strands and a distinct α-helical segment located in the central region of the sequence, as illustrated in the sequence plot and cartoon model (Fig. 1). Functional prediction using FFPred associated the protein with specific GO terms supported by support vector machine (SVM) reliability scores, offering insights into its potential biological roles, molecular functions, and cellular localization. The analysis identified transport as the most likely biological process, nucleic acid binding as the principal molecular function, and ribosome association as the predominant cellular component. A graphical overview of these predictions is shown in Figure 2, while the complete list of predicted functions along with their corresponding SVM reliability scores is provided in Supplementary Table 1.

3.6. Structural Modeling and Three-Dimensional Structure Prediction

To further investigate the structural features of the Exp_525919 encoded protein, homology- and analogy-based fold recognition was performed using the Phyre2.2 server. The analysis successfully modeled 68 residues (96% of the sequence) with 99.5% confidence (Fig. 3), based on the highest-scoring template c7zhsA_, which showed 43% sequence identity. The template corresponds to a chaperone protein annotated as ubiquitin-like protein SMT3, DnaJ homolog subfamily A member 2, derived from a three-dimensional reconstruction of the cylindrical assembly of DNAJA2 (ΔG/F) imposing D5 symmetry. The predicted structural model exhibited dimensions of 45.451 Å × 27.162 Å × 32.338 Å, supporting a compact and ordered fold consistent with putative chaperone proteins.

For structural interpretation using AlphaFold, a homologous uncharacterized protein from *A. hypogyna* (414 aa) was selected as a structural proxy based on its significant alignment metrics, including a high HSP (high-scoring segment pair score: 190), low E-value (5.88412e-15), 59% sequence identity, 70% positives (45/64), and absence of gaps (0%) and also because this hit was consistent with our previous findings. The alignment was restricted to a ~ 60–70 aa region corresponding to Exp_525919. Therefore, only a partial alignment was used for comparative structural inference. The 3D structure was analyzed using AlphaFold-based predictions and visualized per-residue confidence scores (pLDDT) and predicted

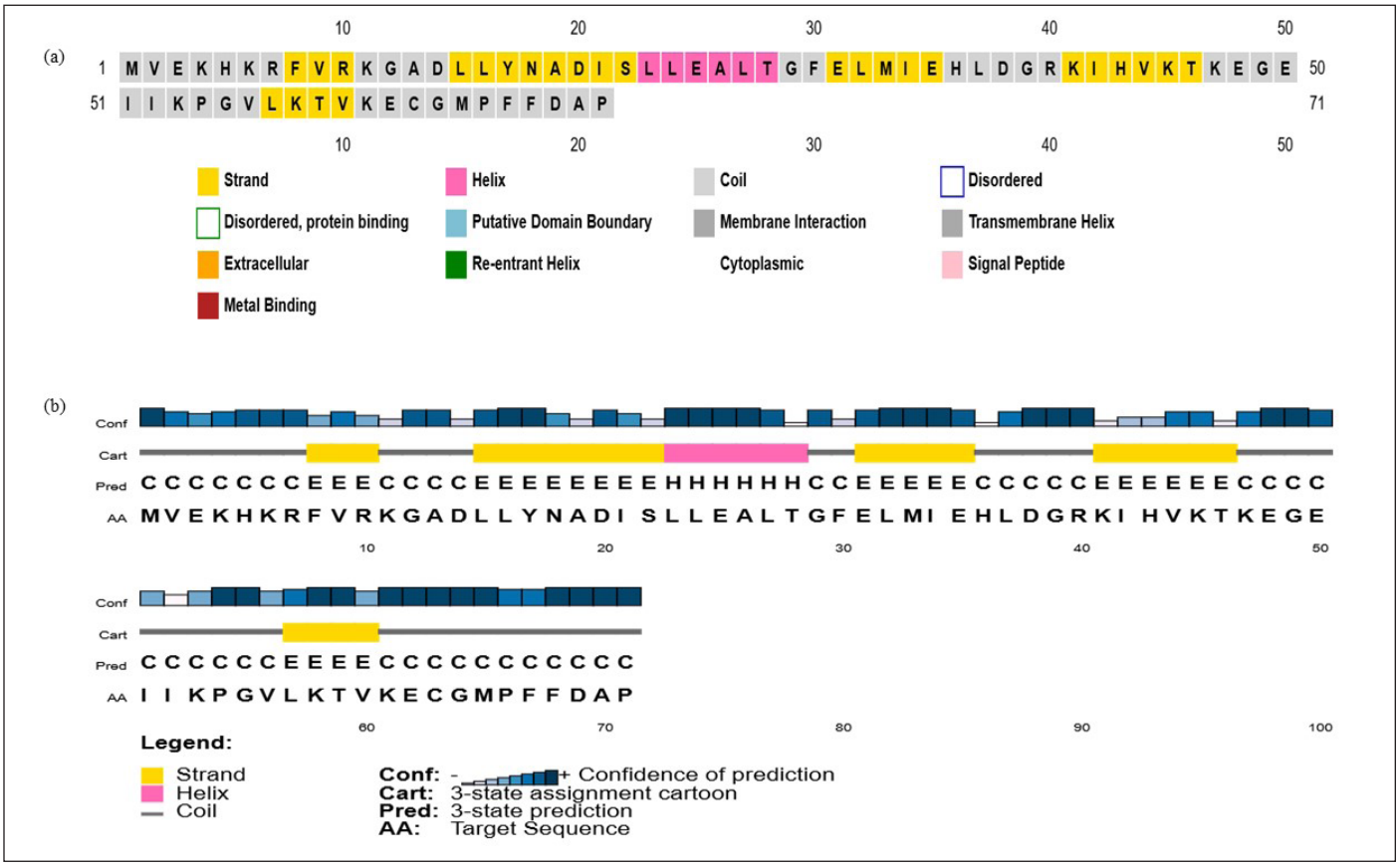

Figure 1. Secondary structure prediction of Exp_525919 protein (longest ORF) using PSIPRED 4.0. (a) The sequence plot illustrates the amino acid sequence annotated with predicted secondary structural elements, including β-strands (yellow), α-helices (pink), and coil regions (grey). This visualization provides a residue-level overview of the structural organization along the protein sequence. (b) The PSIPRED cartoon shows confidence scores and three-state secondary structure assignment along the protein sequence. The prediction indicates a predominance of coil regions interspersed with multiple β-strands and a distinct α-helical segment located in the central region of the sequence.

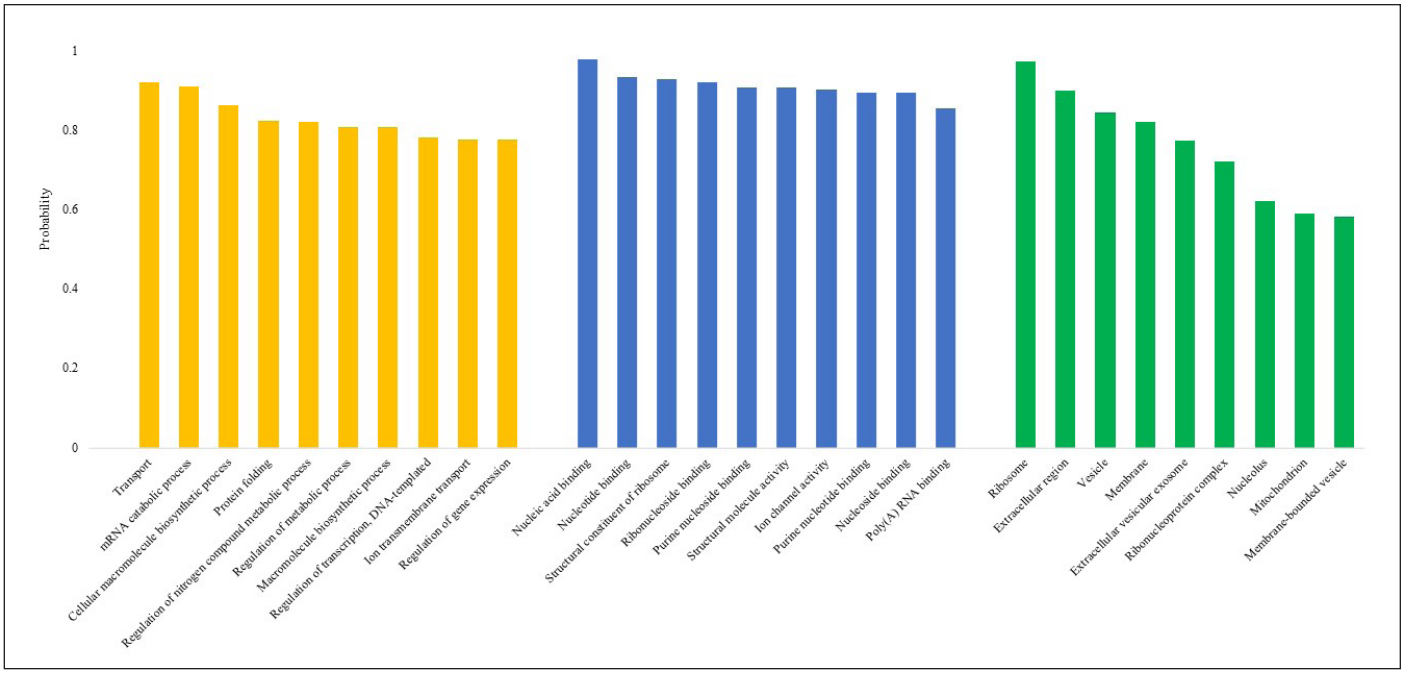

Figure 2. Functional prediction (GO) of Exp_525919 protein (longest ORF) using FFPred.
The bar plot illustrates the predicted Gene Ontology (GO) terms associated with the Exp_525919 protein, categorized into three major functional groups: biological process (yellow), molecular function (blue), and cellular component (green). Transport was identified as the most probable biological process, nucleic acid binding as the primary molecular function, and the ribosome as the predominant cellular component.

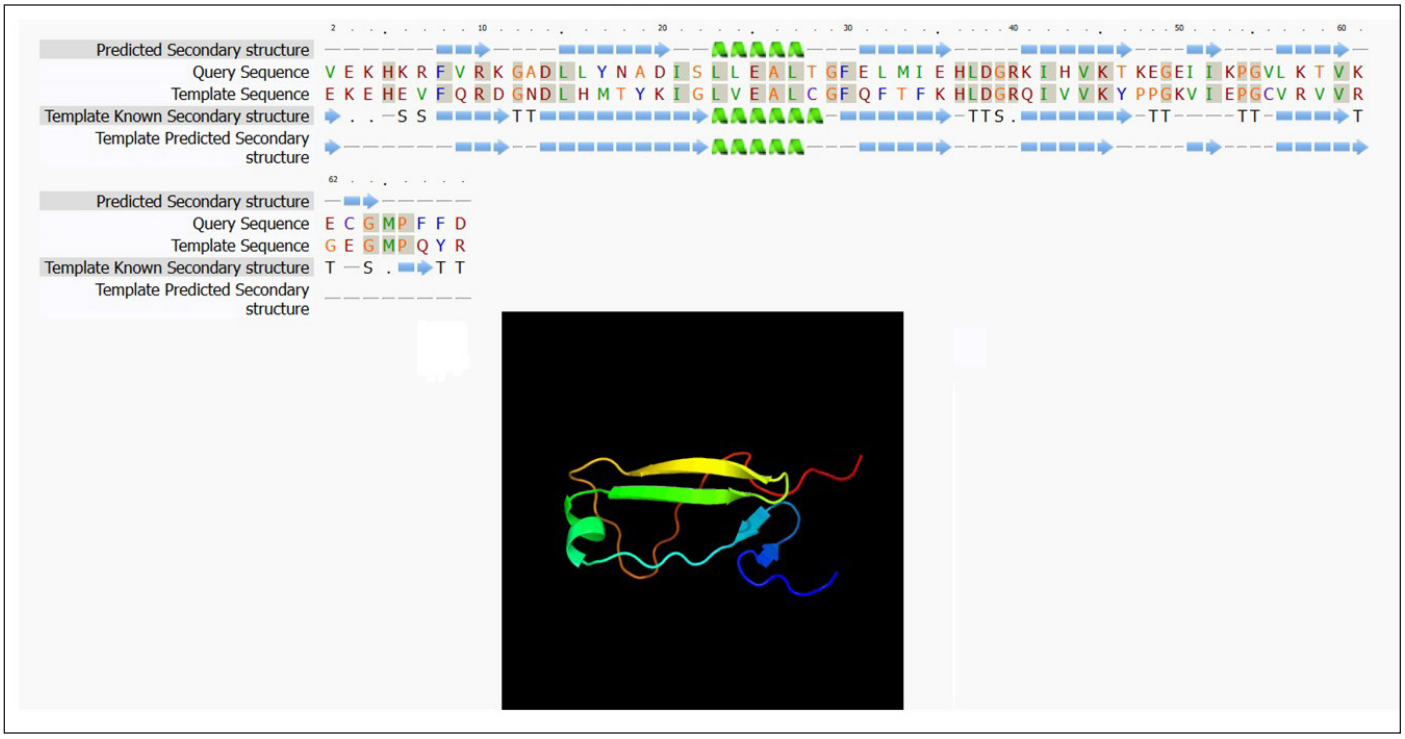

Figure 3. Structural modeling and homology/analogy-based fold recognition using Phyre2.2 server. The predicted 3D structure was generated using the Phyre2.2 server, which employs homology and fold recognition approaches to model protein structures based on known templates. The predicted model covers 68 residues (96% of the sequence) with 99.5% confidence based on the template c7zhsA_ (43% sequence identity). Sequence-structure alignment and the resulting 3D model indicate a compact, ordered fold consistent with chaperone proteins.

aligned error (PAE) matrix (Fig. 4). The 3D structure of the same homologous protein is colored by pLDDT, ranging from high (blue) to low (red). The region corresponding to the Exp_525919 alignment shows comparatively higher confidence, whereas the remaining regions represent portions of the homologous protein not covered by the short query sequence. The PAE heatmap of the homologous

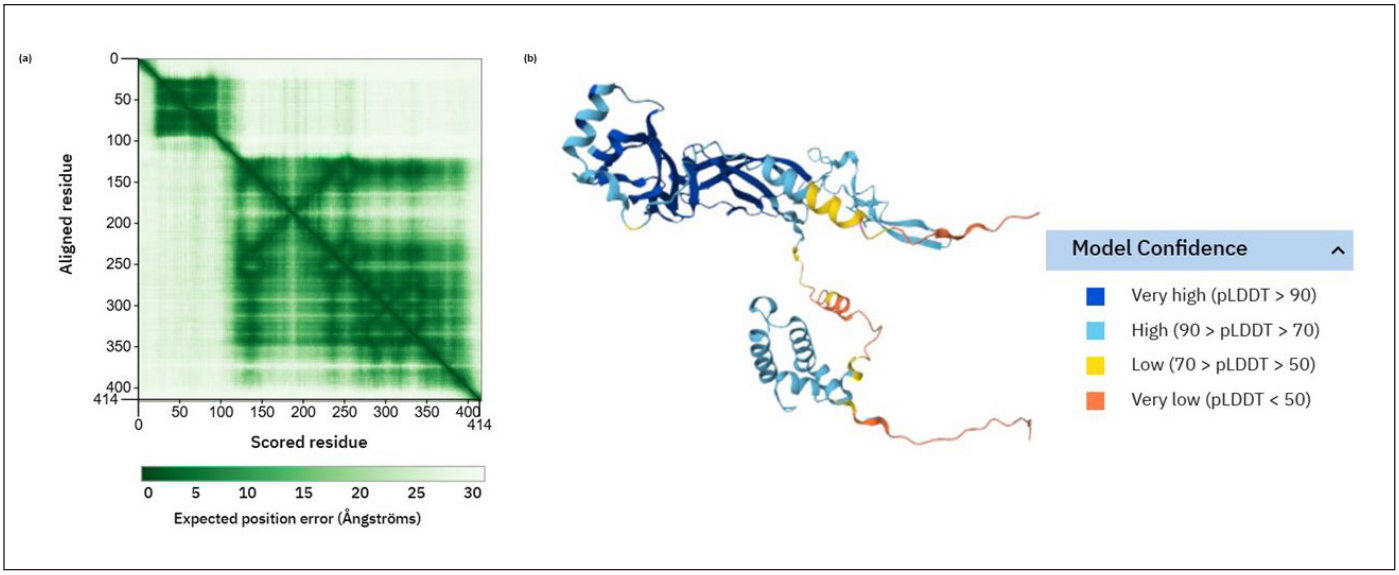

Figure 4. AlphaFold-based structural interpretation of a homologous protein used as a proxy for Exp_525919. (a) The PAE heatmap of the homologous *A. hypogyna* uncharacterized protein (414 aa) shows well-defined low-error regions corresponding to locally aligned segments with the Exp_525919 peptide (~ 60–70 aa). The full matrix reflects the confidence of inter-residue positioning across the homologous full-length protein. (b) The 3D structure of the same homologous protein colored by per-residue confidence scores (pLDDT), ranging from high (blue) to low (red). The region corresponding to the Exp_525919 alignment shows comparatively higher confidence, whereas the remaining regions represent portions of the homologous protein not covered by the short query sequence.

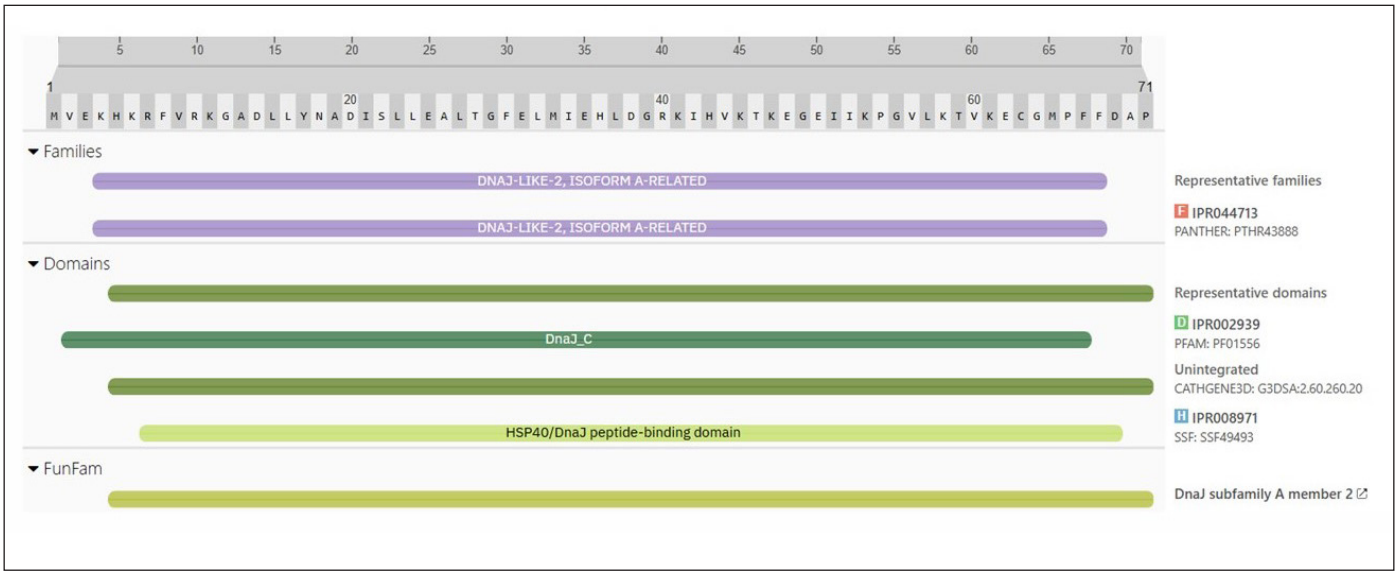

Figure 5. Protein family classification and domain architecture using InterProScan.
The protein is classified within the chaperone J-domain superfamily and assigned to the DnaJ homolog subfamily A member 1/2-like family. The detected DnaJ C-terminal/peptide-binding domain suggests limited similarity to DnaJ/HSP40-related proteins and possible association with chaperone-related pathways; however, co-chaperone activity cannot be inferred and requires experimental validation.

A. hypogyna uncharacterized protein (414 aa) shows well-defined low-error regions corresponding to locally aligned segments with the Exp_525919 peptide (~ 60–70 aa). The full matrix reflects the confidence of inter-residue positioning across the homologous full-length protein.

Collectively, these predictions suggest that Exp_525919, a putative DnaJ C-terminal/peptide-binding domain-containing fragment, may function as a regulatory or protective protein, potentially contributing to cellular homeostasis under environmental stress. Canonical HSP40 (DnaJ) family proteins regulate HSP70 function by stimulating its ATPase activity and facilitating its interaction with nucleotide exchange factors, thereby controlling the ATP-dependent chaperone cycle [71]. However, because Exp_525919 contains only a predicted DnaJ C-terminal/peptide-binding domain-containing fragment and lacks evidence of a canonical J-domain, its direct involvement in HSP70 ATPase regulation cannot be inferred. Nevertheless, the presence of this domain suggests a potential association with chaperone-mediated stress response processes and protein homeostasis under Cr(VI) stress.

3.7. Domain Architecture and Evolutionary Conservation of Exp_525919 Encoded Protein

Protein family classification and conserved domain analysis using InterProScan identified the protein as belonging to the chaperone J-domain superfamily (IPR036869). More specifically, it was classified as DnaJ homolog subfamily A member 1/2-like (IPR044713), with family assignment as DNAJ-LIKE-2, isoform A-related (Fig. 5). Domain architecture analysis revealed the presence of a DnaJ C-terminal domain (Pfam PF01556) and an HSP40/DnaJ peptide-binding domain (SUPERFAMILY SSF49493). Functional family annotation further classified the protein as DnaJ subfamily A member 2 (FUNFAM G3DSA:2.60.260.20:FF:000003), reinforcing its role as a co-chaperone within the HSP70 system.

Initially, Exp_525919 was annotated using UniProt BLAST and was identified as an uncharacterized protein from *A. hypogyna* corresponding to the *ACHHYP_04605* gene with 57% identity and 7.6e-24 e-value. This relatively high level of identity, compared to other hits, justified the selection of *A. hypogyna* as the most appropriate reference organism for subsequent analyses. To further assess evolutionary conservation, the ORF was searched against the OrthoDB database using *A. hypogyna* as the reference organism. This analysis identified several orthologous hypothetical proteins, including ACHHYP_10603, ACHHYP_02934, ACHHYP_04605, and ACHHYP_04602. Among these, ACHHYP_04605 was selected for detailed analysis based on its more complete annotation and higher sequence coverage. Functional predictions for this ortholog suggested molecular functions such as ATP binding, HSP70 protein binding, heat shock protein binding, metal ion binding, unfolded protein binding, and protein-folding chaperone binding. The predicted biological role was protein refolding, and the protein was inferred to localize to both the nucleus and cytosol.

Consistent with these observations, InterPro entry IPR036869 confirmed its classification within the chaperone J-domain superfamily, whose members serve as key regulators of the HSP70 chaperone machinery. HSP70 systems are involved in a wide range of cellular processes, including folding of nascent polypeptides, translocation across organelle membranes, coordination of stress responses, and targeting of proteins for degradation. Members of the HSP40 (DnaJ) family function as co-chaperones by interacting with HSP70 (DnaK) and stimulating its ATPase activity, thereby stabilizing substrate-bound conformations required for efficient protein folding [62,72].

Structurally, canonical DnaJ proteins comprise an N-terminal J-domain, a glycine-rich region, a zinc-finger-like cysteine repeat domain, and a C-terminal substrate-binding region. The J-domain, formed by four α-helices, mediates interaction with the ATPase domain of HSP70 and is essential for chaperone activity [73,74]. The identification of a DnaJ C-terminal/peptide-binding domain-containing fragment in Exp_525919 suggests limited structural similarity to components of the DnaJ/HSP40 protein family and a possible association with chaperone-related stress response functions. However, its functional relationship to canonical DnaJ co-chaperones remains to be established.

3.8. Protein–Protein Interaction Analysis

The protein–protein interaction (PPI) network shows that the predicted uncharacterized protein ACHHYP_04605 has strong predicted associations with multiple molecular chaperones, predominantly members of the HSP70 family, along with DnaJ/HSP40 and HSP90, as listed in Table 2. The high interaction confidence scores (≈0.81−0.90) indicate a biologically meaningful functional linkage [75]. Such clustering with well-characterized chaperones suggests a functional association, as proteins involved in similar biological processes tend to form highly interconnected modules within PPI networks [76,77]. The network was visualized in Cytoscape (Fig. 6).

Notably, ACHHYP_04605 shows interactions with several HSP70-like proteins and HSP70 cognates, indicating its likely involvement in protein folding, refolding, or stabilization pathways. Its association

Table 2. Predicted protein–protein interaction partners of ACHHYP_04605.

Predicted functional partners	Score
ACHHYP_02356 HSP70-like protein	0.896
ACHHYP_10400 Heat shock 70 KDa protein; Belongs to the heat shock protein 70 family	0.845
ACHHYP_11646 HSP70-like protein	0.844
ACHHYP_17465 Heat shock 70 KDa protein; Belongs to the heat shock protein 70 family	0.842
ACHHYP_04460 HSP70-like protein; Belongs to the heat shock protein 70 family	0.842
ACHHYP_16337 Heat shock 70 KDa protein; Belongs to the heat shock protein 70 family	0.841
ACHHYP_14762 Heat shock cognate 70KDa protein	0.837
ACHHYP_14585 Dna J subfamily B protein	0.835
ACHHYP_14729 Heat shock cognate 70KDa protein	0.816
ACHHYP_05781 Heat shock protein 90	0.813

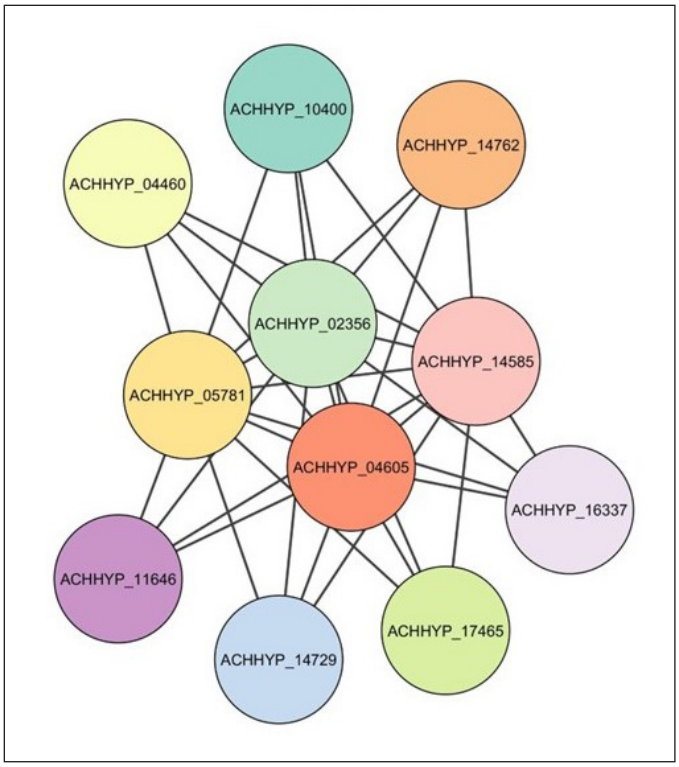

Figure 6. PPI analysis using the STRING database and visualized on Cytoscape. STRING-based PPI network showing strong predicted associations of ACHHYP_04605 with molecular chaperones, predominantly HSP70, DnaJ/HSP40, and HSP90 family proteins, indicating functional linkage within protein folding and stress-response pathways. The high interaction confidence scores (≈0.81−0.90) indicate a biologically meaningful functional linkage.

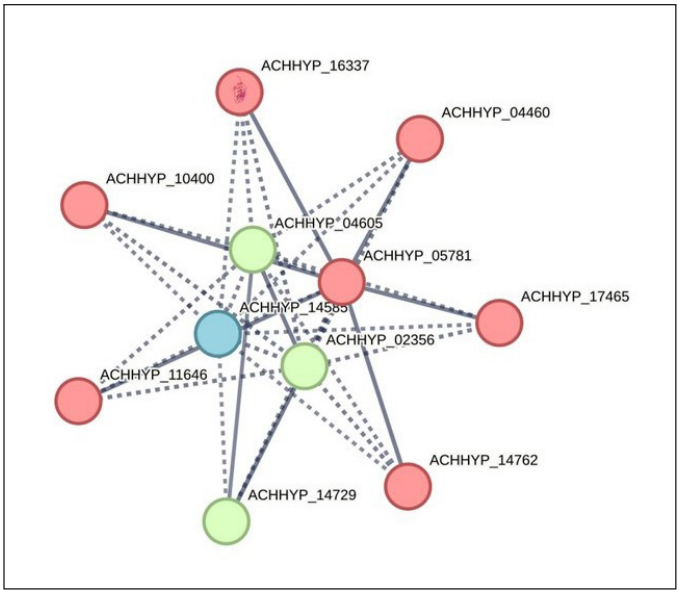

Figure 7. K-means clustering of the PPI network using the STRING database. K-means clustering of the PPI network reveals distinct functional modules, with core chaperone proteins forming densely connected clusters and specialized co-chaperones occupying peripheral positions, supporting a coordinated chaperone machinery.

with multiple HSP70 homologs and DnaJ proteins suggests that it may function as a chaperone client or accessory component within the HSP70-centered stress response network. Furthermore, its interaction with HSP90, a central regulator of multichaperone complexes, supports a potential role in protein maturation and stabilization under stress conditions [78].

Network-based functional inference approaches demonstrate that uncharacterized proteins embedded within chaperone-rich modules may suggest involvement in protein quality control and proteostasis-related stress response [79,80]. Therefore, the interaction profile of ACHHYP_04605 strongly supports its classification as a stress-associated protein linked to cellular proteostasis, likely contributing indirectly to tolerance against environmental stressors. Efficient modulation of protein homeostasis by HSPs requires the coordinated action of co-chaperones. Specific co-chaperones, including HSP40/DNAJ proteins and HSP110-family proteins, modulate HSP70 substrate handling and proteostasis [81,82]. Protein–protein interactions between HSPs and their co-chaperones precisely control the quality-control machinery of client proteins, including protein phosphorylation, ubiquitination, and other post-translational modifications [83,84].

K-means clustering analysis revealed a structured organization of chaperone and folding-related genes into distinct clusters (Fig. 7). Nodes are grouped based on shared functional associations, with edge density reflecting interaction strength. Cluster 1 (red) contained multiple genes broadly associated with protein folding, forming a densely connected subnetwork. Cluster 2 (green) consisted of genes encoding core chaperone components, including HSP70 and HSP40/DnaJ-related proteins, which showed strong connectivity with both cluster 1 and other nodes in the network. Cluster 3 (blue) was represented by a single gene, ACHHYP_14585, annotated as a DnaJ subfamily B protein, positioned at the periphery yet maintaining links with central chaperone nodes. The overall network topology indicates functional diversification within the chaperone machinery, with core protein-folding components acting as central hubs, while specialized co-chaperones contribute to adaptive responses under HM-induced stress.

4. CONCLUSION

This study provides insights into the molecular basis of HM stress adaptation in micro-eukaryotic communities based on transcript-level evidence derived from RNA-Seq-based metatranscriptomics with comprehensive *in silico* characterization. The identification and functional annotation of the predicted upregulated uncharacterized transcript Exp_525919, encoding a putative DnaJ/HSP40-domain-containing fragment, highlights that it may contribute to ATP-dependent protein quality-control systems that maintain cellular homeostasis under chromium stress, as supported by the integration of functional predictions, structural modeling, domain architecture, and evolutionary conservation.

Furthermore, this work expands the functional aspect of a predicted uncharacterized protein and demonstrates how computational approaches can uncover novel components of stress-adaptation pathways. However, due to the absence of biological replicates, the observed expression patterns represent preliminary trends that require validation through replicated transcriptomic approaches.

Protein–protein interaction analysis, inferred from homologous protein networks, identified the predicted uncharacterized protein ACHHYP_04605 as an interactor of major molecular chaperones, including HSP70, DnaJ/HSP40, and HSP90, indicating a likely role in protein folding and quality control. Its association with multiple HSP70 homologs suggests involvement in the HSP70-centered stress response, potentially functioning as a chaperone client or accessory factor, while its interaction with HSP90 supports a role in protein maturation under stress conditions. Network-based inference further classifies ACHHYP_04605 as a stress-associated element linked to cellular proteostasis. K-means clustering revealed a modular organization within the chaperone network, where core folding components act as central hubs and specialized co-chaperones contribute to adaptive stress responses. The model and associated confidence metrics represent computational predictions and require experimental validation. Collectively, these findings advance our understanding of micro-eukaryotic resilience in contaminated environments and provide a foundation for future experimental validation and potential applications in bioremediation strategies.

Overall, the observed transcript-abundance patterns of predicted chaperones and uncharacterized stress-responsive proteins highlight the potential adaptive transcriptional response of micro-eukaryotic communities to challenging environmental conditions. This transcriptional reprogramming reflects adjustments in cellular and metabolic processes that promote survival under altered environments. RNA sequencing thus offers valuable insights into micro-eukaryotic functional responses, revealing novel pathways and enzymes with potential biotechnological applications while advancing our understanding of their adaptive strategies under environmental stress.

5. AUTHORS' CONTRIBUTIONS

All authors made substantial contributions to conception and design, acquisition of data, or analysis and interpretation of data; took part in drafting the article or revising it critically for important intellectual content; agreed to submit to the current journal; gave final approval of the version to be published; and agreed to be accountable for all aspects of the work. All the authors are eligible to be authors as per the International Committee of Medical Journal Editors (ICMJE) requirements/guidelines.

6. ACKNOWLEDGMENTS

The First author sincerely acknowledges the financial support received through the Junior Research Fellowship under the Joint CSIR-UGC NET Fellowship program. This study was funded by the Science and Engineering Research Board, New Delhi (Project No. YSS/2015/000218).

7. CONFLICTS OF INTEREST

The authors report no financial or any other conflicts of interest in this work.

8. ETHICAL APPROVALS

This study does not involve experiments on animals or human subjects.

9. DATA AVAILABILITY

All the data are available with the authors and shall be provided upon request.

10. PUBLISHER'S NOTE

All claims expressed in this article are solely those of the authors and do not necessarily represent those of the publisher, the editors, or the reviewers. This journal remains neutral with regard to jurisdictional claims in published institutional affiliation.

11. DISCLOSURE OF USE OF GENERATIVE AI TOOLS

The authors used ChatGPT (GPT-5.5, OpenAI) solely for language editing, grammar correction, and improvement of manuscript clarity. The AI tool was not used to generate, fabricate, or alter research data, images, statistical analyses, experimental design, results, references, interpretations, or scientific conclusions. The authors reviewed, verified, and edited all AI-assisted content and take full responsibility for the accuracy, originality, integrity, and final content of the manuscript. The AI tool has not been listed as an author, and no AI-generated content was included without human verification.

13. SUPPLEMENTARY MATERIAL

The supplementary material can be accessed at the journal's website: https://jabonline.in/admin/php/uploadss/1508_pdf.pdf

REFERENCES

1. Wuana RA, Okieimen FE. Heavy metals in contaminated soils: a review of sources, chemistry, risks and best available strategies for remediation. ISRN Ecol. 2011; 1–20; doi: https://doi.org/10.5402/2011/402647
2. Kokkinos E, Zouboulis A. Hydrometallurgical recovery of CR(III) from tannery waste: optimization and selectivity investigation. Water (Switzerland) 2020;12(3):1–10; doi: https://doi.org/10.3390/w12030719
3. Bellion M, Courbot M, Jacob C, Guinet F, Blaudez D, Chalot M. Metal induction of a *Paxillus involutus* metallothionein and its heterologous expression in *Hebeloma cylindrosporum*. New Phytologist 2007;174(1):151–8; doi: https://doi.org/10.1111/j.1469-8137.2007.01973.x
4. Ruotolo R, Marchini G, Ottonello S. Membrane transporters and protein traffic networks differentially affecting metal tolerance: a genomic phenotyping study in yeast. Genome Biol 2008;9(4): R67; doi: https://doi.org/10.1186/gb-2008-9-4-r67
5. Pócsi I. Toxic metal/metalloid tolerance in fungi: a biotechnology-oriented approach. In: Banfalvi, G. (ed.). Cellular effects of heavy metals. Springer, Dordrecht, Netherlands, p 31, 2011; doi: https://doi.org/10.1007/978-94-007-0428-2_2
6. Cohen MD, Kargacin B, Klein CB, Costa M. Mechanisms of chromium carcinogenicity and toxicity. Crit Rev Toxicol 1993;23(3):255–81; doi: https://doi.org/10.3109/10408449309105012
7. Joutey NT, Sayel H, Bahafid W, Ghachtouli N El. Mechanisms of hexavalent chromium resistance and removal by microorganisms. Rev Environ Contam Toxicol 2015;233:45–69; doi: https://doi.org/10.1007/978-3-319-10479-9_2
8. Yadav BNS, Sharma P, Maurya S, Yadav RK. Metagenomics and metatranscriptomics as potential driving forces for the exploration of diversity and functions of micro-eukaryotes in soil. 3 Biotech 2023;13(12): 423; doi: https://doi.org/10.1007/s13205-023-03841-3
9. Oliverio AM, Geisen S, Delgado-Baquerizo M, Maestre FT, Turner BL, Fierer N. The global-scale distributions of soil protists and their contributions to belowground systems. Sci Adv 2020;6:eaax8787; doi: https://doi.org/10.1126/sciadv.aax8787
10. Schloter M, Nannipieri P, Sørensen SJ, Van Elsas JD. Microbial indicators for soil quality. Biol Fertil Soils 2018;54(1):1–10; doi: https://doi.org/10.1007/s00374-017-1248-3
11. Huang S, Lentendu G, Fujinuma J, Shiono T, Kubota Y, Mitchell EAD. Soil micro-eukaryotic diversity patterns along elevation gradient are best estimated by increasing the number of elevation steps rather than within elevation band replication. Microb Ecol 2023;86(4):2606–17; doi: https://doi.org/10.1007/s00248-023-02259-x
12. Urich T, Lanzén A, Qi J, Huson DH, Schleper C, Schuster SC. Simultaneous assessment of soil microbial community structure and function through analysis of the meta-transcriptome. PLoS One 2008;3(6): e2527; doi: https://doi.org/10.1371/journal.pone.0002527
13. Stark R, Grzelak M, Hadfield J. RNA sequencing: the teenage years. Nat Rev Genet 2019;20(11):631–56; doi: https://doi.org/10.1038/s41576-019-0150-2
14. Mukherjee A, Reddy MS. Metatranscriptomics: an approach for retrieving novel eukaryotic genes from polluted and related environments. 3 Biotech 2020;10(2): 71; doi: https://doi.org/10.1007/s13205-020-2057-1
15. Tang H, Xiang G, Xiao W, Yang Z, Zhao B. Microbial mediated remediation of heavy metals toxicity: mechanisms and future prospects. Front Plant Sci 2024;15:1420408; doi: https://doi.org/10.3389/fpls.2024.1420408
16. Walkley AJ, Black IA. Estimation of soil organic carbon by the chromic acid titration method. Soil Sci 1934;37:29–38.
17. Piper CS. Soil and plant analysis: a laboratory manual of methods for the examination of soils and the determination of the inorganic constituents of plants. 7th ed. The University of Adelaide, Adelaide, Australia, 1996.
18. Kitson RE, Mellon MG. Colorimetric determination of phosphorous as molybdivanadophosphoric acid. Ind Eng Chem Anal 1944;16:379–83; doi: https://doi.org/10.1021/i560130a017
19. Jackson ML. Soil chemical analysis. Prentice Hall of India Pvt Ltd, New Delhi, India, p 498, 1973.
20. USEPA. Method 3060A, Alkaline Digestion for Hexavalent Chromium. Tests Methods for Evaluating Solid Waste, Physical/Chemical Methods SW 846. Washington DC: US Government Printing Office (GPO); 1996.
21. Yu Z, Pei Y, Zhao S, Kakade A, Khan A, Sharma M, *et al.* Metatranscriptomic analysis reveals active microbes and genes responded to short-term Cr(VI) stress. Ecotoxicology 2021;30(8):1527–37; doi: https://doi.org/10.1007/s10646-020-02290-5
22. Andrews S. FastQC: a quality control tool for high throughput sequence data; 2017. Available from: https://www.bioinformatics.babraham.ac.uk/projects/fastqc/
23. Krueger F. Trim Galore: a wrapper around Cutadapt and FastQC to consistently apply adapter and quality trimming to FastQ files, with extra functionality for RRBS data. Babraham bioinformatics;

2019. Available from: https://github.com/FelixKrueger/TrimGalore
24. Nurk S, Meleshko D, Korobeynikov A, Pevzner PA. MetaSPAdes: a new versatile metagenomics assembler. Genome Res 2017;27(5):824–34; doi: https://doi.org/10.1101/gr.213959.116
25. Schmieder R, Edwards R. Quality control and preprocessing of metagenomic datasets. Bioinformatics 2011;27(6):863–4; doi: https://doi.org/10.1093/bioinformatics/btr026
26. Fu L, Niu B, Zhu Z, Wu S, Li W. CD-HIT: accelerated for clustering the next-generation sequencing data. Bioinformatics 2012;28(23):3150–2; doi: https://doi.org/10.1093/bioinformatics/bts565
27. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. Nat Methods 2012;9(4):357–9; doi: https://doi.org/10.1038/nmeth.1923
28. Anders S, Huber W. Differential expression analysis for sequence count data. Genome Biol 2010;11(10):R106; doi: https://doi.org/10.1186/gb-2010-11-10-r106
29. Zhu W, Lomsadze A, Borodovsky M. Ab initio gene identification in metagenomic sequences. Nucleic Acids Res 2010;38(12): e132; doi: https://doi.org/10.1093/nar/gkq275
30. Buchfink B, Xie C, Huson DH. Fast and sensitive protein alignment using DIAMOND. Nat Methods 2015;12(1):59–60; doi: https://doi.org/10.1038/nmeth.3176
31. National Center for Biotechnology Information (NCBI). ORFfinder: Open Reading Frame Finder. Available from: https://www.ncbi.nlm.nih.gov/orffinder/
32. GenomeNet. MOTIF: searching protein sequence motifs. Available at: https://www.genome.jp/tools/motif/
33. Buchan DWA, Moffat L, Lau A, Kandathil SM, Jones DT. Deep learning for the PSIPRED protein analysis workbench. Nucleic Acids Res 2024;52(W1):W287–93; doi: https://doi.org/10.1093/nar/gkae328
34. Powell HR, Islam SA, David A, Sternberg MJE. Phyre2.2: a community resource for template-based protein structure prediction. J Mol Biol 2025;437(15):168960; doi: https://doi.org/10.1016/j.jmb.2025.168960
35. Bertoni D, Tsenkov M, Magana P, Nair S, Pidruchna I, Querino Lima Afonso M, *et al.* AlphaFold protein structure database 2025: a redesigned interface and updated structural coverage. Nucleic Acids Res 2026;54(1):D358–62; doi: https://doi.org/10.1093/nar/gkaf1226
36. Blum M, Andreeva A, Florentino LC, Chuguransky SR, Grego T, Hobbs E, *et al.* InterPro: the protein sequence classification resource in 2025. Nucleic Acids Res 2025;53(D1):D444–56; doi: https://doi.org/10.1093/nar/gkae1082
37. Tegenfeldt F, Kuznetsov D, Manni M, Berkeley M, Zdobnov EM, Kriventseva EV. OrthoDB and BUSCO update: annotation of orthologs with wider sampling of genomes. Nucleic Acids Res 2025;53(D1):D516–22; doi: https://doi.org/10.1093/nar/gkae987
38. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, *et al.* Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res 2003;13(11):2498–504; doi: https://doi.org/10.1101/gr.1239303
39. Zha S, Yu A, Wang Z, Shi Q, Cheng X, Liu C, *et al.* Microbial strategies for effective hexavalent chromium removal: a comprehensive review. Chem Eng J 2024;489:151457; doi: https://doi.org/10.1016/j.cej.2024.151457
40. Kumar V, Dwivedi SK. Hexavalent chromium reduction ability and bioremediation potential of *Aspergillus flavus* CR500 isolated from electroplating wastewater. Chemosphere 2019;237: 124567; doi: https://doi.org/10.1016/j.chemosphere.2019.124567
41. Bennett RM, Cordero PRF, Bautista GS, Dedeles GR. Reduction of hexavalent chromium using fungi and bacteria isolated from contaminated soil and water samples. Chem Ecol 2013;29(4):320–8; doi: https://doi.org/10.1080/02757540.2013.770478
42. Bahafid W, Tahri Joutey N, Sayel H, Ghachtouli N EL. Mechanism of hexavalent chromium detoxification using *Cyberlindnera fabianii* yeast isolated from contaminated site in Fez (Morocco). J Mater Environ Sci 2013;4(6):840–7.
43. Shakya M, Lo CC, Chain PSG. Advances and challenges in metatranscriptomic analysis. Front Genet 2019; 10:904; doi: https://doi.org/10.3389/fgene.2019.00904
44. Dotaniya ML, Meena VD, Rajendiran S, Coumar MV, Saha JK, Kundu S. Geo-accumulation indices of heavy metals in soil and groundwater of kanpur, India under long term irrigation of tannery effluent. Bull Environ Contam Toxicol 2017;98(5):706–11; doi: https://doi.org/10.1007/s00128-016-1983-4
45. WHO. Permissible limits of heavy metals in soil and plants. Geneva: World Health Organization; 1996.
46. Yadav BNS, Sharma P, Maurya S, Yadav A, Tiwari N, Reddy MS, *et al.* Spatial and seasonal variations of active micro-eukaryotic community structure in heavy metal-contaminated soils. 3 Biotech 2026;16(1):6; doi: https://doi.org/10.1007/s13205-025-04605-x
47. Szebesczyk A, Słowik J. Heat shock proteins and metal ions – Reaction or interaction?. Comput Struct Biotechnol J 2023;21:3103–8; doi: https://doi.org/10.1016/j.csbj.2023.05.024
48. Díaz-Pérez C, Cervantes C, Campos-García J, Julián-Sánchez A, Riveros-Rosas H. Phylogenetic analysis of the chromate ion transporter (CHR) superfamily. FEBS J 2007;274(23):6215–27; doi: https://doi.org/10.1111/j.1742-4658.2007.06141.x
49. Wang X, Wang C, Sheng H, Wang Y, Zeng J, Kang H, *et al.* Transcriptome-wide identification and expression analyses of ABC transporters in dwarf polish wheat under metal stresses. Biol Plant 2017;61(2):293–304; doi: https://doi.org/10.1007/s10535-016-0697-0
50. Gao Y, Yang F, Liu J, Xie W, Zhang L, Chen Z, *et al.* Genome-wide identification of metal tolerance protein genes in *Populus trichocarpa* and their roles in response to various heavy metal stresses. Int J Mol Sci 2020;21(5): 1680; doi: https://doi.org/10.3390/ijms21051680
51. Ferrari M, Cozza R, Marieschi M, Torelli A. Role of sulfate transporters in chromium tolerance in *Scenedesmus acutus* M. (Sphaeropleales). Plants 2022;11(2): 223; doi: https://doi.org/10.3390/plants11020223
52. Smith AT, Smith KP, Rosenzweig AC. Diversity of the metal-transporting P1B-type ATPases. J Biol Inorganic Chem 2014;19(6):947–60; doi: https://doi.org/10.1007/s00775-014-1129-2
53. Yu XZ, Lin YJ, Zhang Q. Metallothioneins enhance chromium detoxification through scavenging ROS and stimulating metal chelation in *Oryza sativa*. Chemosphere 2019;220:300–13; doi: https://doi.org/10.1016/j.chemosphere.2018.12.119
54. Mukherjee A, Yadav R, Marmeisse R, Fraissinet-Tachet L, Reddy MS. Heavy metal hypertolerant eukaryotic aldehyde dehydrogenase isolated from metal contaminated soil by metatranscriptomics approach. Biochimie 2019;160:183–92; doi: https://doi.org/10.1016/j.biochi.2019.03.010
55. Singh V, Singh N, Verma M, Kamal R, Tiwari R, Sanjay Chivate M, *et al.* Hexavalent-Chromium-Induced Oxidative Stress and the Protective Role of Antioxidants against Cellular Toxicity. Antioxidants 2022;11(12): 2375; doi: https://doi.org/10.3390/antiox11122375
56. Legatzki A, Grass G, Anton A, Rensing C, Nies DH. Interplay of the Czc system and two P-type ATPases in conferring metal resistance to *Ralstonia metallidurans*. J Bacteriol 2003;185(15):4354–61; doi: https://doi.org/10.1128/JB.185.15.4354-4361.2003
57. Brace JL, Lester RL, Dickson RC, Rudin CM. SVF1 regulates cell survival by affecting sphingolipid metabolism in *Saccharomyces cerevisiae*. Genetics 2007;175(1):65–76; doi: https://doi.org/10.1534/genetics.106.064527
58. Yan L, Jin H, Raza A, Huang Y, Gu DP, Zou X. Natural resistance-associated macrophage proteins (NRAMPs) are involved in cadmium enrichment in peanut (*Arachis hypogaea* L.) under cadmium

stress. Plant Growth Regul 2024;102(3):619–32; doi: https://doi.org/10.1007/s10725-023-01091-0
59. Yadav BNS, Sharma P, Maurya S, Yadav RK. Functional metatranscriptomics reveals a novel multi-metal tolerant gene of nramp-family, isolated from metal-contaminated soil. Appl Biol Res 2024;26(2):186–98; doi: https://doi.org/10.48165/abr.2024.26.01.22
60. Gu J, He Y, He C, Zhang Q, Huang Q, Bai S, *et al.* Advances in the structures, mechanisms and targeting of molecular chaperones. Sig Transduct Target Ther 2025;10:84; doi: https://doi.org/10.1038/s41392-025-02166-2
61. Hartl FU, Hayer-Hartl M. Converging concepts of protein folding *in vitro* and *in vivo*. Nat Struct Mol Biol 2009;16:574–81; doi: https://doi.org/10.1038/nsmb.1591
62. Frydman J. Folding of newly translated proteins *in vivo*: the role of molecular chaperones. Annu Rev Biochem 2001;70:603–47; doi: https://doi.org/10.1146/annurev.biochem.70.1.603
63. Wu J, Liu T, Rios Z, Mei Q, Lin X, Cao S. Heat Shock Proteins and Cancer. Trends Pharmacol Sci 2017;38(3):226–56; doi: https://doi.org/10.1016/j.tips.2016.11.009
64. Lubkowska A, Pluta W, Strońska A, Lalko A. Role of Heat Shock Proteins (HSP70 and HSP90) in Viral Infection. Int J Mol Sci 2021;22(17):9366; doi: https://doi.org/10.3390/ijms22179366
65. Keller A. A review on heat shock proteins (HSP70 and HSP90) in protein folding and stress response. Int J Mol Biol Biochem 2025;7(1):57–62; doi: https://doi.org/10.33545/26646501.2025.v7.i1a.85
66. Vierling E. The Roles of Heat Shock Proteins in Plants. Annu Rev Plant Biol 1991;42:579–620; doi: https://doi.org/10.1146/annurev.pp.42.060191.003051
67. Mayer MP, Bukau B. Hsp70 chaperones: cellular functions and molecular mechanism. Cell Mol Life Sci 2005;62(6):670–84; doi: https://doi.org/10.1007/s00018-004-4464-6
68. Sharma SS, Dietz KJ. The relationship between metal toxicity and cellular redox imbalance. Trends Plant Sci 2009;14(1):43–50; doi: https://doi.org/10.1016/j.tplants.2008.10.007
69. Ogura T, Wilkinson AJ. AAA+ superfamily ATPases: common structure--diverse function. Genes Cells 2001;6(7):575–97; doi: https://doi.org/10.1046/j.1365-2443.2001.00447.x
70. Hennessy F, Nicoll WS, Zimmermann R, Cheetham ME, Blatch GL. Not all J domains are created equal: implications for the specificity of Hsp40-Hsp70 interactions. Protein Sci 2005;14(7):1697–709; doi: https://doi.org/10.1110/ps.051406805
71. Kampinga HH, Craig EA. The HSP70 chaperone machinery: j proteins as drivers of functional specificity. Nat Rev Mol Cell Biol 2010;11(8):579–92; doi: https://doi.org/10.1038/nrm2941
72. Walsh P, Bursać D, Law YC, Cyr D, Lithgow T. The J-protein family: modulating protein assembly, disassembly and translocation. EMBO Rep 2004;5(6):567–71; doi: https://doi.org/10.1038/sj.embor.7400172
73. Pellecchia M, Szyperski T, Wall D, Georgopoulos C, Wüthrich K. NMR structure of the J-domain and the Gly/Phe-rich region of the *Escherichia coli* DnaJ chaperone. J Mol Biol 1996;260(2):236–50; doi: https://doi.org/10.1006/jmbi.1996.0395
74. Genevaux P, Schwager F, Georgopoulos C, Kelley WL. Scanning mutagenesis identifies amino acid residues essential for the *in vivo* activity of the *Escherichia coli* DnaJ (HSP40) J-domain. Genetics 2002;162(3):1045–53; doi: https://doi.org/10.1093/genetics/162.3.1045
75. Szklarczyk D, Kirsch R, Koutrouli M, Nastou K, Mehryary F, Hachilif R, *et al.* The STRING database in 2023: protein–protein association networks and functional enrichment analyses for any sequenced genome of interest. Nucleic Acids Res 2023;51(D1):D638–46; doi: https://doi.org/10.1093/nar/gkac1000
76. Han JDJ, Bertin N, Hao T, Goldberg DS, Berriz GF, Zhang LV, *et al.* Evidence for dynamically organized modularity in the yeast protein–protein interaction network. Nature 2004;430:88–93; doi: https://doi.org/10.1038/nature02555
77. Barabási AL, Oltvai ZN. Network biology: understanding the cell's functional organization. Nat Rev Genet 2004;5:101–13; doi: https://doi.org/10.1038/nrg1272
78. Taipale M, Jarosz DF, Lindquist S. HSP90 at the hub of protein homeostasis: emerging mechanistic insights. Nat Rev Mol Cell Biol 2010;11(7):515–28; doi: https://doi.org/10.1038/nrm2918
79. Powers ET, Morimoto RI, Dillin A, Kelly JW, Balch WE. Biological and Chemical Approaches to Diseases of Proteostasis Deficiency. Annu Rev Biochem 2009;78:959–91; doi: https://doi.org/10.1146/annurev.biochem.052308.114844
80. Kim YE, Hipp MS, Bracher A, Hayer-Hartl M, Ulrich Hartl F. Molecular Chaperone Functions in Protein Folding and Proteostasis. Annu Rev Biochem 2013;82:323–55; doi: https://doi.org/10.1146/annurev-biochem-060208-092442
81. Velasco-Carneros L, Bernardo-Seisdedos G, Maréchal JD, Millet O, Moro F, Muga A. Pseudophosphorylation of single residues of the J domain of DNAJA2 regulates the holding/folding balance of the Hsc70 system. Protein Sci 2024;33(8):e5105; doi: https://doi.org/10.1002/pro.5105
82. Kandasamy G, Andréasson C. HSP70-HSP110 chaperones deliver ubiquitin-dependent and -independent substrates to the 26S proteasome for proteolysis in yeast. J Cell Sci 2018;131(6):jcs210948; doi: https://doi.org/10.1242/jcs.210948
83. Havalová H, Ondrovičová G, Keresztesová B, Bauer JA, Pevala V, Kutejová E, *et al.* Mitochondrial HSP70 chaperone system—the influence of post-translational modifications and involvement in human diseases. Int J Mol Sci 2021;22:8077; doi: https://doi.org/10.3390/ijms22158077
84. Piva F. Gaining new insights on the HSP90 regulatory network. Bioinformation 2020;16:17–20; doi: https://doi.org/10.6026/97320630016017

How to cite this article:
Sharma P, Singh Yadav BN, Maurya S, Yadav RK. *In silico* characterization of a metatranscriptome-derived stress-responsive putative DnaJ/HSP40-domain-containing fragment associated with heavy metal homeostasis. J Appl Biol Biotech 2026. Article in Press. http://doi.org/10.7324/jabb.2026.311545