Research Square

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

Genome of the tardigrade Cucumibius annulatus: functional annotation and specific proteins

Daria Makarova
ITMO University
Denis Tumanov
St Petersburg University
Elena Nassonova
Institute of Cytology
Mikhail Rayko
mike.rayko@gmail.com

ITMO University

Research Article

Keywords: genome, tardigrade, stress response, comparative genomics

Posted Date: March 6th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-8958382/v1

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

Additional Declarations: No competing interests reported.

Version of Record: A version of this preprint was published at Functional & Integrative Genomics on September 21st, 2026. See the published version at https://doi.org/10.1007/s10142-026-02055-z.

Abstract

Tardigrades, or water bears, are microscopic invertebrates renowned for their ability to survive extreme environmental conditions through a process called cryptobiosis. The genomic underpinnings of this remarkable resilience are a subject of intense scientific interest. In this study, we present the first genome assembly and annotation of a tardigrade species *Cucumibius annulatus*, recently redescribed from North-West Russia. The draft genome was assembled de novo using DNBSEQ sequencing data (PE300 reads), followed by decontamination and annotation. Final assembly has a size of 73 Mb and a BUSCO completeness score of 87.1%. Comparative genomic analysis against three other tardigrade species (*Hypsibius exemplaris*, *Ramazzottius varieornatus*, and *Paramacrobiotus metropolitanus*) revealed unique orthologous gene groups in *C. annulatus.* We identified significant expansions in protein families associated with stress response, including antioxidant enzymes (e.g., Superoxide Dismutase), chaperones (HSP70), and components of the DNA repair and ubiquitin-proteasome systems. Notably, *C. annulatus* exhibits a pronounced species-specific expansion of signaling receptors, particularly receptor guanylate cyclases and G-protein coupled receptors, as well as extracellular matrix proteins like metalloproteases. These findings suggest that adaptations in signaling pathways and structural proteins, in addition to typical stress-response genes, may play a crucial role in the survival strategies.

Introduction

Tardigrades (phylum Tardigrada) are microscopic invertebrates well known for surviving environmental extremes that are lethal to most animals, including severe desiccation, temperature shocks, and ionizing radiation. A key biological state enabling such endurance is cryptobiosis—most prominently anhydrobiosis—during which metabolism is strongly reduced and organisms can persist in this state for extended periods. The molecular basis of these features is encoded in tardigrade genomes and reflects both conserved eukaryotic stress-response systems and lineage-specific genetic features. [1, 2, 3]

Early genome-wide studies highlighted an unexpected range of protective strategies and introduced the idea that tardigrade-unique proteins can directly contribute to cellular protection. A canonical example is the tardigrade damage suppressor protein (Dsup), first characterized in *Ramazzottius varieornatus* and shown to improve radiotolerance when expressed in human cultured cells. [5] Later structural work emphasized the intrinsically disordered nature of Dsup and its ability to form a complex with DNA, supporting a model of direct DNA shielding. [6]

Comparative genomics indicates that extremotolerance is not implemented in exactly the same way across all tardigrades. Instead, the sets and copy numbers of tardigrade-specific families (including CAHS/SAHS/MAHS and related abundant heat-soluble proteins) can differ substantially between species, suggesting that closely related phenotypes may rely on partly different molecular “toolkits”. [7, 8] Recent evolutionary analyses of desiccation- and temperature-related protein families further support a complex history of stress-associated systems in Tardigrada. [9] In addition, anhydrobiosis-linked metabolism is not uniform: trehalose production machinery (TPS–TPP) can be absent in some anhydrobiotic lineages due to repeated gene loss and horizontal transfer, implying that trehalose-based strategies are not universal. [10]

Ionizing radiation tolerance provides another clear example of interspecies differences. In *Hypsibius exemplaris*, radiation tolerance has been linked to strong transcriptional activation of DNA repair pathways (notably BER and NHEJ), and functional experiments implicate at least one core repair factor (XRCC5/Ku80) in survival under irradiation. [11] Comparative transcriptomics has also identified additional tardigrade-specific irradiation-responsive candidates (e.g., TDR1), supporting the idea that different lineages may rely on partly non-overlapping protective components. [12]

An integrative taxonomic study reclassified *Grevenius annulatus* as *Cucumibius annulatus* and established the new genus *Cucumibius*, supported by detailed morphological evidence including new data on cuticular structure. [13] As a newly recognized lineage within Isohypsibioidea, *C. annulatus* provides an opportunity to expand comparative tardigrade genomics and to evaluate how stress-response and cuticle-associated protein repertoires vary across species. In particular, genome-wide analysis allows testing whether commonly discussed extremotolerance-associated families (e.g., CAHS/SAHS/LEA-like protectants, antioxidant enzymes such as SOD, and DNA repair factors) are conserved, expanded, or potentially absent in *C. annulatus*, and whether other lineage-specific expansions dominate its unique gene space. [2, 3, 7, 8]

Here, we present a de novo genome assembly and structural/functional annotation of *Cucumibius annulatus*, followed by a comparative proteomic analysis with three reference tardigrades (*Ramazzottius varieornatus*, *Paramacrobiotus metropolitanus*, and *Hypsibius exemplaris*). We focus on orthogroups unique to *C. annulatus* and describe their functional landscape using COG and Pfam annotations, complemented by targeted curation of candidates linked to stress response/DNA repair and to cuticle- and chitin-associated biology. This study provides a genomic resource for a newly established genus and a comparative framework to interpret lineage-specific features that may contribute to stress adaptation and cuticle biology in tardigrades. [8, 13]

Methods

Genome Sequencing and Assembly

Specimens of *Cucumibius annulatus* were collected from the lake Figurnoe, Leningrad region, North-West Russia. DNA isolation was performed with Arcturus PicoPure DNA Extraction Kit (Thermo Fisher Scientific, Massachusetts). The purified DNA was amplified by multiple displacement amplification (MDA) using the REPLI-g Single Cell amplification kit (Qiagen; Thermo Fisher Scientific), according to the manufacturer’s protocol. DNA sequencing was performed at the Core Facility Centre “Biobank” of the Research Park of St. Petersburg University using the MGI DNBSEQ G400 platform (paired-end reads, 300 bp). The resulting reads (56,3 mln) were assembled de novo using SPAdes (v3.15.2) [14].

Binning and Decontamination

Based on the results of the morphological study, the obtained samples were infected with parasites—most likely representatives of Oomycetes. To separate them from the host genome, binning was performed using MaxBin (version 2.2.7), followed by selection of the most plausible bins. The initial assembly was screened for contaminants using BlobToolKit (v2.7.4) [8]. Contigs identified as originating from bacteria, viruses, or plants based on sequence homology and GC content were removed. Only contigs assigned to Tardigrada, Arthropoda, or those with no hits were retained for the final assembly. The quality and completeness of the final genome assembly were assessed using QUAST (v5.0.2) [16] and BUSCO (v5.2.2) [17] against the Eukaryota odb10 database.

Gene Prediction and Structural Annotation

Gene prediction was performed on the cleaned genome assembly using BRAKER2 (v2.1.6) [18], which integrates evidence from protein homology and ab initio models from GeneMark-ES/ET [19] and AUGUSTUS [20]. Protein evidence for BRAKER was constructed by combining reference proteomes of other tardigrades downloaded from NCBI, including *Ramazzottius varieornatus* (GCA_001949185.1, Rvar_4.0), *Paramacrobiotus metropolitanus* (GCF_019649055.1, Prichtersi_v1.0), and *Hypsibius exemplaris* (GCA_002082055.1, nHd_3.1).

Comparative Genomics and Orthogroup Inference

To identify orthologous gene families, the predicted proteome of *C. annulatus* was compared with the proteomes of three other tardigrade species obtained from NCBI: *Hypsibius exemplaris* (GCA_002082055.1), *Ramazzottius varieornatus* (GCA_001949185.1), and *Paramacrobiotus metropolitanus* (GCA_019640555.1). OrthoFinder (v3.1.0) [22] was used to infer orthogroups (OGs) based on all-vs-all sequence similarity searches performed with DIAMOND [16]. The resulting orthogroups were analyzed to identify unique, expanded, and conserved gene families in *C. annulatus.*

Functional annotation and COG profiling

Functional annotation of the *C. annulatus* protein set was performed using eggNOG-mapper (emapper-2.1.3) with the eggNOG v5 database (downloaded as emapperdb-5.0.2). [23] Predicted functions were summarized by COG categories and complemented with KEGG KO and Pfam annotations for downstream enrichment-style interpretation of lineage-specific protein groups.

Data processing and visualization

Targeted homology searches for trehalose biosynthesis enzymes were performed using DIAMOND (v. 2.1.22; blastp) against the predicted *C. annulatus* proteome, using multiple reference TPS/TPP sequences (*Saccharomyces cerevisiae* Tps1/Tps2; *Caenorhabditis elegans* TPS-1 and GOB-1; *Drosophila melanogaster* Tps1; and bacterial OtsA/OtsB from *Escherichia coli*). Downstream processing of orthogroup lists and functional annotations (COG/Pfam) as well as plotting figures and interpretation of figures (e.g., Venn diagrams, COG barplots, top-Pfam plots) were performed in R (version used locally: R 4.5.2).

Results

Analysis of Symbionts and Decontamination

We identified nine distinct rRNA sequences corresponding to potential contaminants or symbionts. These included *Ferruginibacter* spp. (Chitinophagaceae, 98.50% identity for OY970254.1), one Burkholderiaceae species (92.44% for LR585335.1), and *Corynebacterium kroppenstedtii* (99.89% for CP069509.1), as well as an alphaproteobacterium closely related to endosymbiotic Candidatus *Tisiphia* (Rickettsiaceae, 99.78% for OZ032159.1). Additionally, we found sequence matching oomycete parasite *Haptoglossa zoospora*, (98.24% for KT257318.1, Fig. S1). The binning procedure successfully separated these non-host sequences from the tardigrade genome, allowing us to focus the subsequent analysis on the host genomic content. We also found SSU rRNA of *C. annulatus*, and used it for the verification of the taxonomic position of our sample (99.94% for PQ069999.1).

Genome Assembly

The final genome assembly of *Cucumibius annulatus* spans 73 Mb, with an N50 of 56 kb and a GC-content of 39.1%. The BUSCO assessment indicated a high level of completeness, with 87.1% of the core eukaryotic genes being present (72.1% for Metazoa dataset). Taxonomic, GC- and coverage-based analyses by BlobToolKit show no signs of contamination. Contig distribution revealed one single “blob” assigned to Tardigrada, Arthropoda and a set of no-hit contigs (Fig. 1).

The *Cucumibius annulatus* assembly is within the range of comparable tardigrade genomes in terms of total length (73.0 Mb versus 55.8–170.5 Mb, see Table 1), but according to some metrics (N50 = 96.4 kb; L50 = 197), it is more fragmented than the *R. varieornatus* and *P. metropolitanus* assemblies. In terms of completeness, our genome assembly is comparable to other tardigrades, but has a lower BUSCO score (Metazoa C = 72.7%), while none of the compared assemblies exceeds 80% (Table 1).

Table 1
Comparative assembly quality metrics QUAST and BUSCO scores (metazoa_odb10 and eukaryota_odb10) for four tardigrade genome assemblies

Data comparison metrics	*Ramazzottius varieornatus* (rvar, GCA_001949185.1)	*Paramacrobiotus metropolitanus* (parmetro, GCA_019649055.1)	*Hypsibius exemplaris* (hypexem, GCA_002082055.1)	*Cucumibius annulatus* (cucumibius, this study)
Sequencing technologies	Sanger: ABI 3730xl DNA Analyzer; Illumina: Genome Analyzer IIx	Illumina HiSeq 4000; Oxford Nanopore MinION	PacBio; Illumina HiSeq	DNBSEQ
Contig/scaffold number (≥ 1000 bp)	200	674	1421	1881
Total length (bp, scaffolds ≥ 1000 bp)	55842812	170501563	104154999	72559506
Average contig/scaffold length (bp)	279214	252970	73297	38575
Longest contig/scaffold length (bp)	9333084	4484719	2115976	785594
N50 (bp) (L50)	4740345 (4)	1033953 (50)	342180 (85)	96445 (197)
N90 (bp) (L90)	1295620 (15)	150116 (202)	65573 (343)	21771 (790)
auN (bp)	4838420	1319310	467267	146080
GC (%)	47.51	43.46	45.46	39.17
N's per 100 kbp	750.86	0.06	2058.95	8.70
Total N's	419303	109	2144499	6310
Percent gaps (%) (N/total)	0.751	0.000	2.059	0.009
BUSCO metazoa (n = 954)	C:76.9% [S:75.5%,D:0.6%],F:25.6%	C:76.5% [S:73.3%,D:3.2%],F:1.6%	C:76.2% [S:74.6%,D:1.6%],F:1.8%	C:72.7% [S:71.0%,D:1.8%],F:2.8%
BUSCO eukaryota (n = 255)	C:91.0%[S:90.2%,D:0.8%],F:2.0%	C:90.2% [S:85.5%,D:4.7%],F:2.4%,	C:89.4% [S:88.2%,D:1.2%],F:2.4%	C:87.1% [S:86.3%,D:0.8%],F:3.5%

Comparative Genomics and Orthogroups

The BRAKER pipeline predicted a total of 15,875 protein-coding genes in the *C. annulatus* genome. The OrthoFinder analysis clustered 58,355 proteins from the four tardigrade species into 13,419 orthogroups. A Venn diagram illustrating the distribution of shared and unique orthogroups is shown in Fig. 2 (see detailed list of orthogroups in Table S1). A key outcome is the identification of 318 orthogroups unique to *C. annulatus*. These unique orthogroups collectively contain 1441 *C. annulatus* proteins, of which 719 received eggNOG functional annotations (COG/Pfam/KO fields available) and were used for quantitative functional profiling (see functional annotation of *C. annulatus* proteins in Table S2).

Analysis of specific proteins of C. annulatus-specific groups by COG categories

To characterize specific proteins functions, we compared the COG category composition of eggNOG-annotated proteins from *C. annulatus*-unique orthogroups (n = 719) against the full set of eggNOG-annotated proteins in the *C. annulatus* proteome (n = 12,042). In the unique set, the most represented categories were T (signal transduction mechanisms; 259 prot., 36.0%), O (posttranslational modification, protein turnover, chaperones; 119, 16.6%) and S (function unknown; 110 prot., 15.3%). Additional contributions were observed for G (carbohydrate transport and metabolism; 39), P (inorganic ion transport and metabolism; 32) and C (energy production and conversion; 29), whereas core "housekeeping" functions, such as translation-related J, were underrepresented among unique proteins (Fig. 3). Overall, *C. annulatus*-specific orthogroups show a pronounced shift toward regulatory/sensory functions (COG T) and protein homeostasis modules (COG O), together with a substantial fraction of proteins that remain functionally uncharacterized (COG S).

Pfam domain enrichment among C. annulatus-unique orthogroup proteins

Pfam profiling of proteins from unique orthogroups revealed clear domain-level signatures consistent with the COG-based shift toward signaling and extracellular functions (Fig. 4). The most frequent domains among unique proteins were: Pkinase (185), Pkinase_Tyr (131), ANF_receptor (120), Guanylate_cyc (116), HNOBA (109). This combination is characteristic of receptor-type guanylate cyclase architectures and strongly suggests a species-specific expansion of cGMP-linked receptor signaling in *C. annulatus*. Additional enriched domains included 7tm_1 (43) (GPCR-like receptors), and multiple extracellular/ECM-associated modules: Astacin (68), TSP_1 (53), EGF (48), Trypsin (45), ShK (42), CUB (40), as well as Lectin_C (35) and BTB (29). Collectively, these data indicate that the unique orthogroup fraction is dominated by coherent expansions of sensory signaling receptors and extracellular remodeling proteins.

To further inspect lineage-specific proteins lacking functional assignments, we focused on the subset classified as COG category S (Function unknown) within the *C. annulatus*-unique orthogroups. Among these 110 corresponding proteins, all contained identifiable Pfam domains, indicating that they represent structurally defined proteins that included recurrent signatures associated with membrane-associated signaling (e.g., 7tm_1 and XK-related domains), chromatin and transcriptional regulation (e.g., Chromo, SET, PHD and zinc-finger domains), extracellular interactions (e.g., EGF, Lectin_C and CAP domains). Together, these results suggest that the "function unknown" proteins likely represent species-specific regulatory and signaling components contributing to *C. annulatus*-specific adaptations.

Stress-response and repair-related candidates

Several unique orthogroups contain proteins classically implicated in stress tolerance and cellular repair (Table 2): antioxidant enzymes (SOD-like), chaperones (HSP70-like), selective autophagy receptors (NBR1-like), ubiquitin-mediated proteostasis regulators (BTB/Kelch adaptors), and DDR-associated ubiquitin ligases (e.g., HUWE1-like). Notably, HSP70-family chaperones are generally expected to be conserved; therefore, their assignment to a *C. annulatus*-specific orthogroup likely reflects orthogroup splitting and/or divergence of particular paralogs rather than true lineage-specificity. Together, these proteins form a coordinated set of mechanisms that support oxidative stress control, proteome maintenance, and DNA damage response pathways.

Table 2
Cucumibius annulatus-specific orthogroups with stress-response and repair-related proteins

Type	Orthogroup (#of proteins)	Functional relevance
Antioxidant protection (SOD1 (Superoxide dismutase Cu-Zn) and sod)	**OG0010046 (3)**	KO: K04565; Pfam: **Sod_Cu**. Canonical antioxidant enzyme implicated in reactive oxygen species (ROS) detoxification and commonly discussed in the context of stress tolerance.
	OG0009925 (3)	
HSP70-like putative paralogs	**OG0011604 (2)**	HSP70 is a central molecular chaperone that maintains proteostasis, particularly under stress, by binding unfolded or misfolded proteins, promoting their proper folding, and preventing aggregation.
Selective autophagy / protein quality control	**OG0001225 (9)**	Presence of NBR1 (selective autophagy receptor; "next to BRCA1 gene"), with domain features consistent with PB1/ZZ/UBA modules. Supports a role in stress-associated clearance of damaged proteins/organelles.
Ubiquitination / BTB–Kelch (regulators of protein degradation)	**OG0000257 (24)**	Enriched in BTB/BACK/Kelch proteins, which often act as CUL3 E3 ligase adaptors. Expansion may indicate lineage-specific remodeling of ubiquitin-mediated proteostasis and stress signaling.
E3 ubiquitin ligases (stress/DDR regulation)	**OG0010267 (3)**	RFFL/FEM1-like E3 ubiquitin ligases with ankyrin repeats; regulatory proteins potentially involved in stress and DNA damage response (DDR) pathways
DNA repair	**OG0011414 (2)**	One protein is annotated as HUWE1 (HECT E3 ubiquitin ligase), which has reported links to ubiquitin-mediated regulation of DNA damage response and/or repair-associated processes

Chitin/cuticle-related proteins potentially involved in cuticle remodeling

A second set captures candidates potentially involved in tardigrade cuticle organization and remodeling (Table 3), including a canonical GH18 chitinase with CBM14 (OG0011702), multiple CBM14 chitin-binding proteins (Peritrophin-A-like signals), multidomain chitin-associated proteins (e.g., CBM14 + LysM + polysaccharide deacetylase-like combinations), and extracellular proteases with chitin-binding modules. This repertoire is consistent with the observed enrichment of extracellular domains (TSP_1/EGF/CUB/Trypsin/Astacin) and supports the hypothesis of a dynamically regulated and lineage-specific extracellular/cuticular toolkit in *C. annulatus*. [13]

Table 3
Chitin-related proteins potentially involved in cuticle organization and remodeling in *Cucumibius annulatus*

Type	Orthogroup (#of proteins)	Pfam / domains	K a c
Chitinases (chitin degradation / remodeling)	**OG0011702 (2)**	Glyco_hydro_18 + CBM_14	K K (
Chitinase-like (atypical domain architecture; requires validation)	**OG0010212 (1)**	Glyco_hydro_19	C c
	OG0011447 (1)		
Chitin-binding proteins (structural/extracellular components)	**OG0001235 (9)**	CBM_14; sometimes Chitin_bind_4 (e.g., g15380.t1; g1810.t1)	P A b
Chitin binding + chitin metabolic process (multidomain)	**OG0011703 (2)**	CBM_14, LysM, Polysacc_deac_1 (+ Cu-binding_MopE)	c b c m p
Proteases with chitin-binding modules (extracellular remodeling)	**OG0000353 (19)**	Chitin_bind_4, NIDO, Somatomedin_B, Trypsin	p li
Protease with CBM_14 (candidate cuticle remodeling factor)	**OG0011230 (1)**	CBM_14, Trypsin (+ PAP2, RED_C/RED_N, RRM_1)	p li
Chitin synthase	**OG0007650 (1)**	Chitin_synth_1,Chitin_synth_1N,Chitin_synth_2,EGF,F5_F8_type_C,FYVE_2,Glyco_hydro_47,Glyco_trans_2_3,I-set,JAB,Myosin_head,OTU,PKD_channel,Rab_eff_C,SAM_1,SAM_2,SRCR,TSP_1,UQ_con	k
	OG0003428 (2)		

Discussion

Lineage-specific expansions and molecular adaptations

Extensive expansion of receptor-associated signaling proteins in *C. annulatus* genome suggest that it invests heavily in membrane-associated sensing and regulatory networks, potentially enhancing the detection of environmental changes and coordination of physiological transitions associated with dehydration and cryptobiosis. A second major lineage-specific theme is the expansion of extracellular and matrix-associated proteins. Together with the presence of chitin- and cuticle-related proteins (including GH18 chitinases with CBM14 and multiple chitin-binding proteins), these findings suggest that *C. annulatus*

possesses an expanded toolkit for cuticle remodeling and extracellular structural maintenance, processes that are critical during molting and may contribute to mechanical stability and barrier function during desiccation.

Cucumibius annulatus is the first species of Isohypsibioidea for which a genome has been assembled. Based on rRNA tree topology, this superfamily is a basal lineage within Parachela [13]. *C. annulatus* is strictly freshwater, does not occur in terrestrial habitats, and most likely cannot tolerate drying out. Phylogenetic placement raises the possibility that the prominent expansions we observe—especially in receptor-like signaling proteins and chitin/cuticle-associated genes—reflect either retained ancestral features or lineage-specific adaptations linked to the unusually complex cuticle morphology of *C. annulatus*.

We additionally identified a lineage-specific expansion of globin-domain proteins (OG0000357; 19 proteins). Beyond classical roles in oxygen transport, globins in small invertebrates are often implicated in redox balance and stress physiology. In the context of cryptobiotic cycles and fluctuating oxygen availability, an expanded globin repertoire may contribute to oxygen homeostasis and mitigation of oxidative stress during entry into and recovery from dormant states.

Conserved maintenance systems versus lineage-specific solutions

Despite the pronounced lineage-specific expansions, *C. annulatus* retains a conserved core of stress-response and maintenance pathways typical for tardigrades. Comparative analysis revealed the presence of key DNA repair enzymes (including XPF-, MUS81-, FEN1-, DNA2-, and EXO1-like proteins), as well as a well-developed RNA surveillance and RNAi machinery (exosome components, mRNA decapping and degradation factors, and Dicer- and MUT-7-like proteins), supporting an active model of genome and transcriptome maintenance under stress.

Importantly, *C. annulatus* shares canonical tardigrade heat-soluble protein families with *R. varieornatus*, including 16 CAHS, 13 SAHS and 3 LEA proteins, indicating that the classical stress-protection toolkit is conserved and may support basal stress tolerance. At the same time, the retention of CAHS, SAHS (and LEA) may be biologically relevant beyond anhydrobiosis, as CAHS proteins have been implicated in protection against multiple stress types, including cold stress/cryobiosis. This is consistent with the distribution of *C. annulatus* in boreal and mountainous regions (so called Arctic-alpine areal type) where freshwater habitats can freeze extensively. In contrast, we did not detect convincing Dsup-like homologs in the predicted *C. annulatus* proteome using similarity searches with Dsup from *R. varieornatus*, suggesting that a Dsup-based DNA-shielding mechanism is either absent or too divergent to be recovered by homology-based approaches [5, 6].

In addition, we detect convincing candidates for the trehalose biosynthesis machinery (TPS–TPP) in the predicted *C. annulatus* proteome – 9 proteins with trehalose phosphatase domains (Trehalose_PPase). This number is consistent with the amount of trehalose biosynthesis pathway members among other tardigrade species in the study. Trehalose can act as a cryoprotectant and chemical chaperone. Overwintering under ice can impose hypoxic conditions. The lineage-specific expansion of globin-domain proteins in *C. annulatus* therefore may be interpreted not only in terms of oxidative stress control, but also as a potential adaptation supporting oxygen homeostasis during seasonal oxygen limitation.

However, in contrast to *R. varieornatus*, the dominant lineage-specific trend in *C. annulatus* is not the continued expansion of the heat-soluble proteins, but rather large-scale diversification of signaling receptors and extracellular proteins. These results suggest that *C. annulatus* achieves extremotolerance through a combination of conserved protective mechanisms and lineage-specific regulatory and structural adaptations. In particular, the expansion of signaling receptors and extracellular/structural protein families distinguishes *C. annulatus* from other sequenced tardigrades and may represent an important, previously underappreciated component of tardigrade adaptation.

Conclusion

In this study, we provide the first genome assembly, structural annotation and comparative orthogroup-based analysis for the recently established tardigrade genus *Cucumibius* and its species *Cucumibius annulatus*. The genome-wide comparison against three reference tardigrades indicates 318 orthogroups unique to *C. annulatus* species, with a functional profile strongly shifted toward signal transduction and protein turnover/proteostasis. This pattern suggests that *C. annulatus* combines a conserved “core” of stress-response mechanisms with pronounced lineage-specific expansions that likely affect how environmental cues are sensed and how extracellular structures are maintained.

Declarations

Ethical Approval

Not applicable

Clinical trial number

Not applicable

Consent to participate

Not applicable

Consent to publish

All authors agreed with the content and that all gave explicit consent to submit. They obtained consent from the responsible authorities at the institute where the work has been carried out

Data Availability Statement

Genome assembly was deposited as NCBI BioSample, accession number SAMN55414403

Authors Contributions

All authors contributed to the study conception and design. Sample collection and material preparation were performed by Denis Tumanov. Preparation of material for library construction and sequencing was performed by Elena Nassonova. Data collection and analysis were performed by Daria Makarova and Mikhail Rayko. The first draft of the manuscript was written by Daria Makarova and all authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.

Funding

This study was supported by the Russian Science Foundation: project no. 25-74-20033 (sample collection, morphological description, and species identification) and project no. 23–74-00071 (DNA extraction, whole genome amplification, sequencing, metagenome assembly, analysis of parasite and host genomes, bioinformatic data processing).

Competing Interests

The authors declare no conflicts of interest.

Acknowledgements

This study utilised equipment of the Core Facility Centres ‘Biobank’, ‘Development of Molecular and Cell Technologies’ and ‘Culturing of microorganisms’ of the Research Park of Saint Petersburg University

References

1. Guidetti R, Jönsson KI (2002) Long-term anhydrobiotic survival in semi-terrestrial tardigrades. J Zool 257(2):181–187
2. Møbjerg N, Halberg KA, Jørgensen A, Persson D, Bjørn M, Ramløv H, Kristensen RM (2011) Survival in extreme environments by tardigrades. J Exp Biol 214(Pt 1):1–11
3. Arakawa K (2022) Examples of Extreme Survival: Tardigrade Genomics and Molecular Anhydrobiology. Annu Rev Anim Biosci 10:17–37. 10.1146/annurev-animal-021419-083711
4. Carrero D, Pérez-Silva JG, Quesada V, López-Otín C (2019) Differential mechanisms of tolerance to extreme environmental conditions in tardigrades. Sci Rep 9:14938. 10.1038/s41598-019-51471-8
5. Hashimoto T, Horikawa DD, Saito Y et al (2016) Extremotolerant tardigrade genome and improved radiotolerance of human cultured cells by tardigrade-unique protein. Nat Commun 7:12808. 10.1038/ncomms12808
6. Zarubin M, Murugova T, Ryzhykau Y, Ivankov O, Uversky VN, Kravchenko E (2024) Structural study of the intrinsically disordered tardigrade damage suppressor protein (Dsup) and its complex with DNA. Sci Rep 14:74335. 10.1038/s41598-024-74335-2
7. Boothby TC, Tapia H, Brozena AH et al (2017) Tardigrades use intrinsically disordered proteins to survive desiccation. Mol Cell 65(6):975–984
8. Yoshida Y, Koutsovoulos G, Laetsch DR et al (2017) Comparative genomics of the tardigrades *Hypsibius dujardini* and *Ramazzottius varieornatus*. PLoS Biol 15(7):e2002266. 10.1371/journal.pbio.2002266
9. Fleming JF, Pisani D, Arakawa K (2024) The Evolution of Temperature and Desiccation-Related Protein Families in Tardigrada Reveals a Complex Acquisition of Extremotolerance. *Genome Biology and Evolution*, 16(1), evad217. 10.1093/gbe/evad217. (Erratum: 16(3) evae043; 17(2) evaf018.)
10. Hara Y, Shibahara R, Kondo K, Abe W, Kunieda T (2021) Parallel evolution of trehalose production machinery in anhydrobiotic animals via recurrent gene loss and horizontal transfer. Open Biology 11(7):200413. 10.1098/rsob.200413
11. Clark-Hachtel CM, Hibshman JD, De Buysscher T et al (2024) The tardigrade *Hypsibius exemplaris* dramatically upregulates DNA repair pathway genes in response to ionizing radiation. Curr Biol 34(9):1819–1830e6. 10.1016/j.cub.2024.03.019
12. Anoud M, Delagoutte E, Helleu Q et al (2024) Comparative transcriptomics reveal a novel tardigrade-specific DNA-binding protein induced in response to ionizing radiation. eLife 13:RP92621. 10.7554/eLife.92621
13. Tumanov DV, Shunatova NN, Fedyuk KA (2025) Integrative description of *Grevenius annulatus* … leads to the institution of a new genus. Zoolog Scr 54(2):232–252. 10.1111/zsc.12703
14. Bankevich A, Nurk S, Antipov D et al (2012) SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J Comput Biol 19(5):455–477
15. Challis R, Richards E, Rajan J, Cochrane G, Blaxter M (2020) BlobToolKit—Interactive quality assessment of genome assemblies. G3: Genes Genomes Genet 10(4):1361–1374
16. Gurevich A, Saveliev V, Vyahhi N, Tesler G (2013) QUAST: quality assessment tool for genome assemblies. Bioinformatics 29(8):1072–1075
17. Manni M, Berkeley MR, Seppey M, Simão FA, Zdobnov EM (2021) BUSCO update… Molecular Biology Evolution 38(10):4647–4654

18. Brůna T, Hoff KJ, Lomsadze A, Stanke M, Borodovsky M (2021) BRAKER2: automatic eukaryotic genome annotation with GeneMark-EP + and AUGUSTUS supported by a protein database. NAR Genomics Bioinf 3(1):lqaa108. https://doi.org/10.1093/nargab/lqaa108
19. Lomsadze A, Ter-Hovhannisyan V, Chernoff YO, Borodovsky M (2005) Gene identification in novel eukaryotic genomes by self-training algorithm. Nucleic Acids Res 33(20):6494–6506. https://doi.org/10.1093/nar/gki937
20. Stanke M, Keller O, Gunduz I, Hayes A, Waack S, Morgenstern B (2006) AUGUSTUS: ab initio prediction of alternative transcripts. Nucleic Acids Res 34(suppl2):W435–W439. https://doi.org/10.1093/nar/gkl200
21. Emms DM, Kelly S (2019) OrthoFinder: phylogenetic orthology inference for comparative genomics. Genome Biol 20:238. https://doi.org/10.1186/s13059-019-1832-y
22. Buchfink B, Xie C, Huson DH (2015) Fast and sensitive protein alignment using DIAMOND. Nat Methods 12(1):59–60
23. Huerta-Cepas J, Forslund K, Coelho LP, Szklarczyk D, Jensen LJ, von Mering C, Bork P (2017) Fast genome-wide functional annotation through orthology assignment by eggNOG-mapper. Mol Biol Evol 34(8):2115–2122. https://doi.org/10.1093/molbev/msx148

Figures

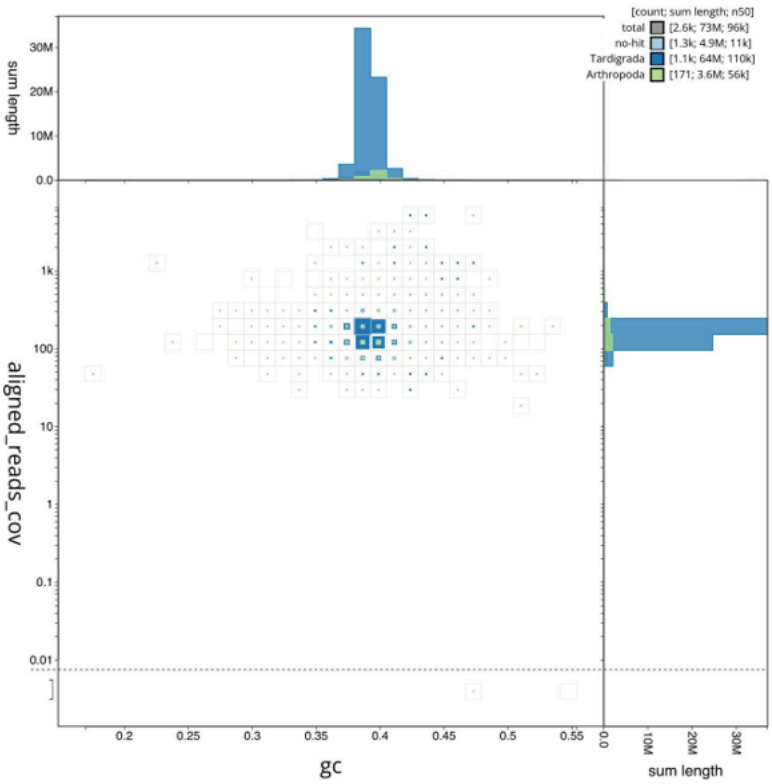


Figure 1

BlobToolKit GC–coverage plot of assembled contigs with taxonomic assignments. The dominant cluster corresponds to the target Tardigrada bin; non-target contigs were removed during decontamination.

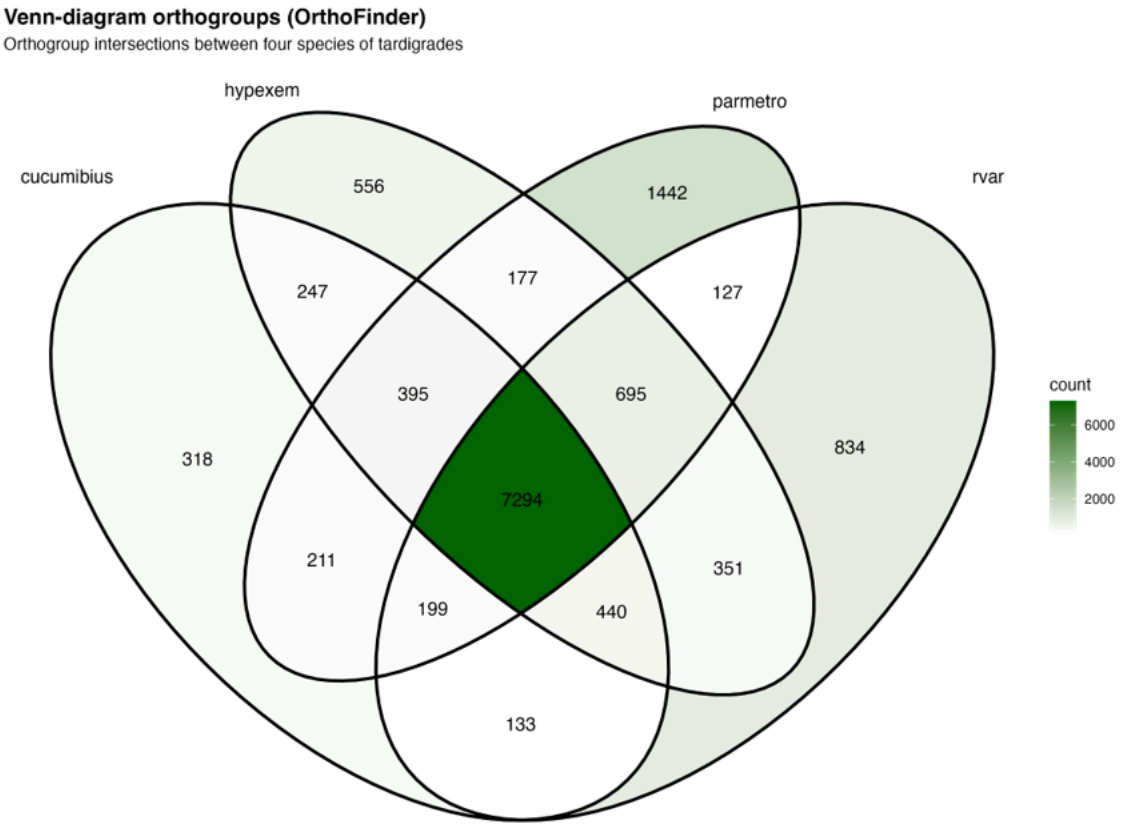
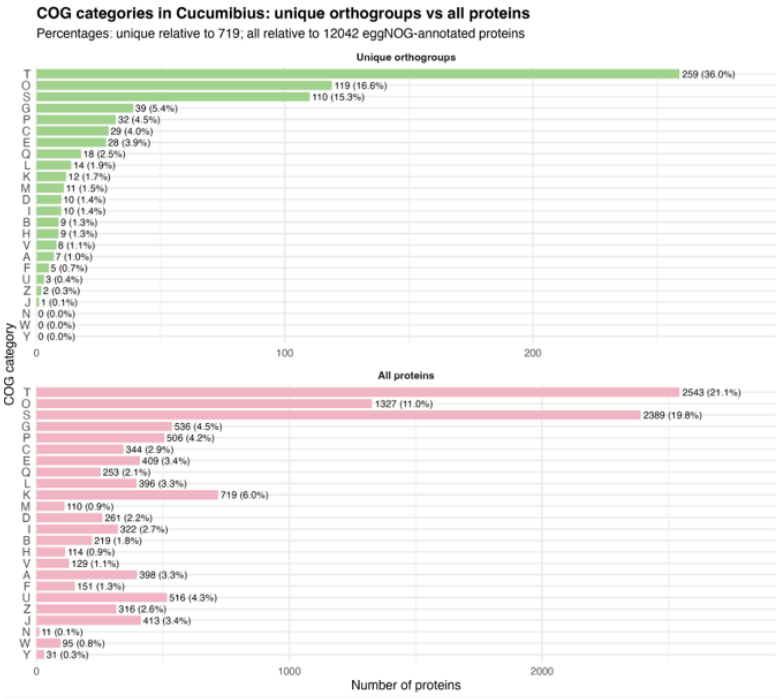


Figure 2

Venn diagram showing the intersection of orthogroups between four species of tardigrades: *Ramazzottius varieornatus* (rvar), *Paramacrobiotus metropolitanus* (parmetro), *Hypsibius exemplaris* (hypexem), and *Cucumibius annulatu*s (cucumibius). The numbers in the sectors correspond to the number of orthogroups that are common or unique to the respective species.



unique & all Cucumibius COG

COG_cat	COG_description	Unique	All
A	RNA processing and modification	7	398
B	Chromatin structure and dynamics	9	219
C	Energy production and conversion	29	344
D	Cell cycle control, cell division, chromosome partitioning	10	261
E	Amino acid transport and metabolism	28	409
F	Nucleotide transport and metabolism	5	151
G	Carbohydrate transport and metabolism	39	536
H	Coenzyme transport and metabolism	9	114
I	Lipid transport and metabolism	10	322
J	Translation, ribosomal structure and biogenesis	1	413
K	Transcription	12	719
L	Replication, recombination and repair	14	396
M	Cell wall/membrane/envelope biogenesis	11	110
N	Cell motility	0	11
O	Posttranslational modification, protein turnover, chaperones	119	1327
P	Inorganic ion transport and metabolism	32	506
Q	Secondary metabolites biosynthesis, transport and catabolism	18	253
R	General function prediction only	0	0
S	Function unknown	110	2389
T	Signal transduction mechanisms	259	2543
U	Intracellular trafficking, secretion, and vesicular transport	3	516
V	Defense mechanisms	8	129
W	Extracellular structures	0	95
Y	Nuclear structure	0	31
Z	Cytoskeleton	2	316

Figure 3

COG category distribution in *Cucumibius annulatus* proteome: unique orthogroup proteins vs all eggNOG-annotated proteins.

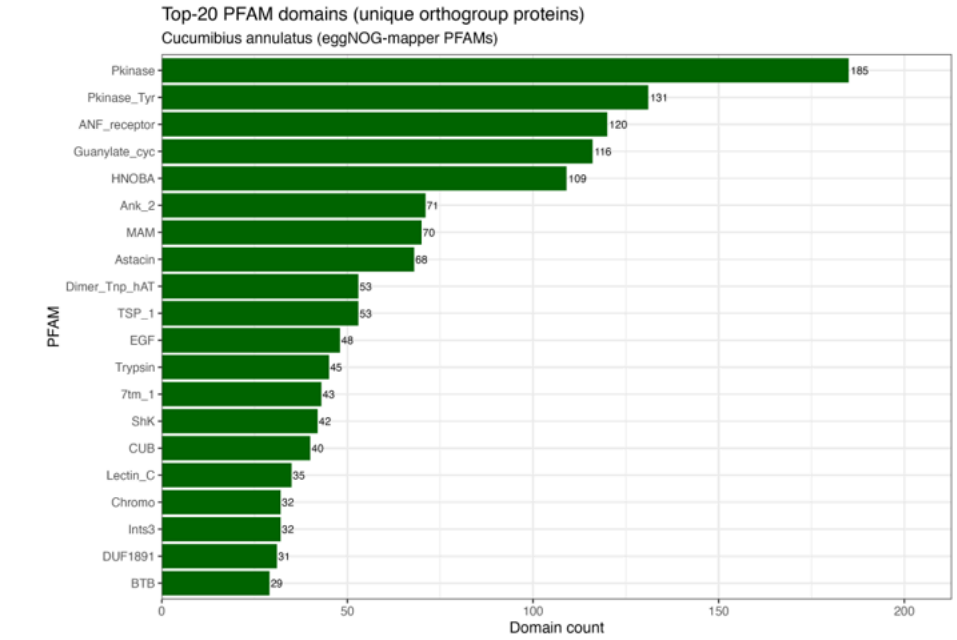


Figure 4

Top-20 Pfam domains among proteins from *Cucumibius annulatus*-specific orthogroups (eggNOG-mapper Pfam assignments). Numbers indicate domain counts.

Supplementary Files

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

- TableS1.xlsx
- TableS2.xlsx
- SupplementalMaterials.docx