Nucleic Acids Research, 2025, gkaf1096
https://doi.org/10.1093/nar/gkaf1096
Database issue

DOO: integrated multi-omics resources for deep ocean organisms

Jiajie She[1,2], **Pei-yuan Qian**[1,2,*], **Longjun Wu**[2,3,*]

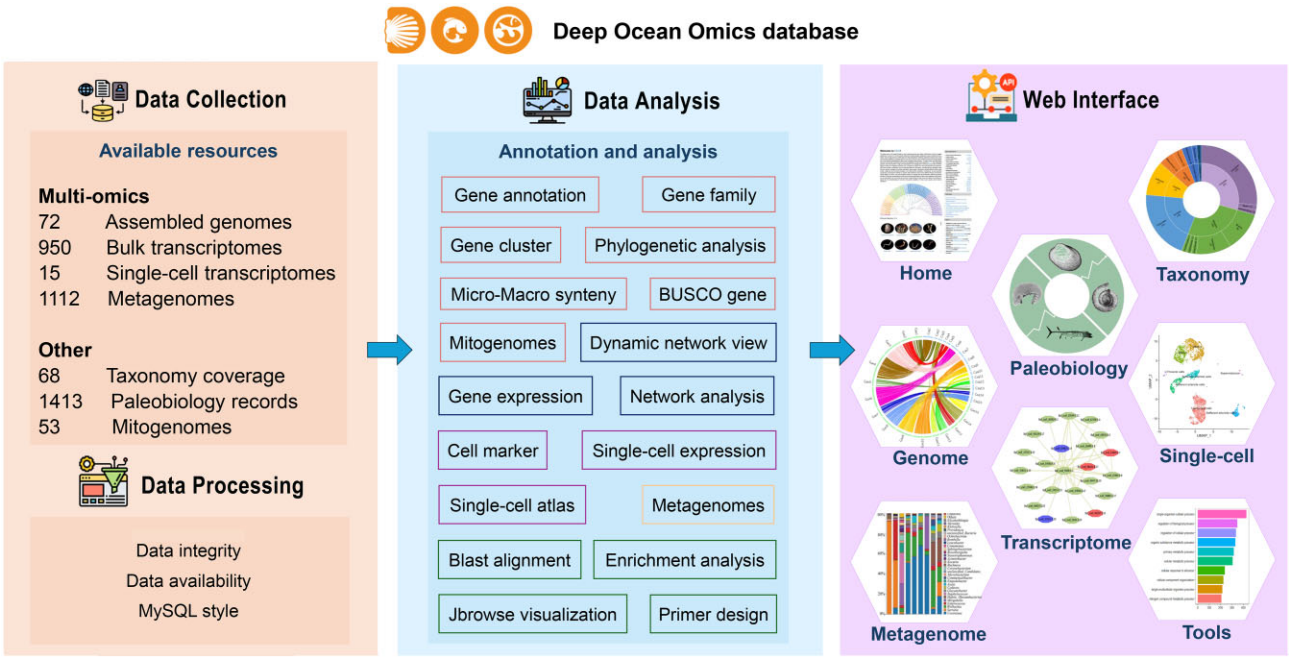
[1]Southern Marine Science and Engineering Guangdong Laboratory (Guangzhou), Nansha, Guangzhou 510000, China
[2]Department of Ocean Science and Otto Poon Center for Climate Resilience and Sustainability, The Hong Kong University of Science and Technology, Hong Kong SAR 999077, China
[3]Hong Kong Branch of the Southern Marine Science and Engineering Guangdong Laboratory (Guangzhou), The Hong Kong University of Science and Technology, Hong Kong SAR, China

*To whom correspondence should be addressed. Email: longjunwu@ust.hk
Correspondence may also be addressed to Pei-yuan Qian. Email: boqianpy@ust.hk

Abstract

The deep ocean is one of Earth's most vast and least explored frontiers, characterized by extreme conditions such as high pressure, limited light, and nutrient scarcity. These environments pose unparalleled challenges to life, making them invaluable for studying genetic and molecular adaptations to extreme conditions. Emerging omics resources have recently provided significant insights into the advanced understanding of deep ocean ecosystems and evolution. However, a centralized resource for deep ocean multi-omics data remains lacking. To bridge this gap, the Deep Ocean Omics (DOO, https://DeepOceanOmics.org) database, a multi-omics atlas for deep ocean organisms, is presented. DOO integrates diverse omics resources from 68 species across seven phyla and 16 classes, encompassing 72 genomes, 950 bulk transcriptomes, 15 single-cell transcriptomes, and 1112 metagenomes, alongside functional support toolkits for functional and comparative analysis. DOO provides a systematic view of genomic information, including genome assembly, phylogeny, gene annotation, BUSCO genes, transcription factors/ubiquitin family, gene clusters, symbiont and mitochondrial genomes, and fossil records. Moreover, DOO offers co-expression networks with expression views across different tissues, and developmental stages and micro- and macrosynteny analyses to elucidate the pan-evolutionary features of genome structure. As the first comprehensive multi-omics resource dedicated to deep ocean organisms, DOO serves as a pivotal platform for uncovering multi-omics underpinnings of deep ocean organisms and offering insights into the understanding of deep ocean biodiversity, evolution, and genetic adaptation under extreme conditions.

Graphical abstract



Received: August 15, 2025. **Revised:** September 21, 2025. **Accepted:** October 2, 2025

Introduction

The deep ocean, as the largest ecosystem on the planet, represents one of the final unexplored frontiers on Earth, harbouring unique ecosystems and extraordinary biodiversity [1–3]. The synergistic interplay of abyssal stressors, including high pressure, oxygen deficiency, perpetual darkness, low temperature, and nutrient limitation, imposes profound physiological constraints on deep ocean organisms [4, 5]. Consequently, deep ocean was long considered almost devoid of organisms akin to a desert [6]. It was only until the 1970s, with the help of HOV Alvin [7], that the scientific community discovered deep ocean habitats with high organismal abundance such as hydrothermal vents and methane seeps. These areas turned out to be biodiversity hotspots that support chemosynthesis-based ecosystems and a diverse array of deep ocean organisms [8, 9]. Organisms living in these deep ocean ecosystems exhibit extreme adaptations and serve as vital components driving critical ecological processes, including chemosynthetic carbon fixation and cross-trophic energy transfer [10, 11]. These deep ocean organisms, which are defined as living beyond the mesopelagic zone (below 1000 m) hereby, represent unparalleled models for elucidating the adaptation mechanisms under hostile conditions, providing deeper understanding of the genetic and molecular basis of evolutionary changes. Due to the extreme remoteness and multifaceted environmental constraints in the deep ocean, studying these organisms is substantially challenging, which severely limits the understanding of their extreme adaptations [12, 13].

Recently, the rapid advancement of multi-omics technologies has opened up a new avenue to study deep ocean organisms with new insights such as the genetic, proteomic, metabolic, and other functional genomic characteristics [14]. These multi-omics approaches provide powerful analytical capabilities for deciphering the molecular adaptations of deep ocean organisms [15–17], uncovering conserved metabolic pathways, stress responses, and adaptations to extreme environments [18, 19]. Indeed, a substantial number of recent studies have reported the genetic and molecular basis of deep ocean adaptation by using multi-omics approaches [20–27]. Despite the increasing application of omics technologies in deep ocean research and the growing number of omics resources, no centralized repository is available for multi-omics data on deep ocean organisms [28–30].

A comprehensive multi-omics database for deep ocean organisms is crucial, offering substantial scientific and practical benefits. It could allow researchers to investigate complex biological questions in a systematic manner, such as adaptations to extreme environments, trophic and community dynamics, life history, and evolution. In this study, Deep Ocean Omics (DOO, https://DeepOceanOmics.org) database, the first comprehensive multi-omics repository dedicated to deep ocean organisms, was established. DOO integrates multi-omics resources (genomes, bulk and single-cell transcriptomes, and metagenomes) and provides convenient toolkits for multidimensional functional and comparative analysis. DOO offers a systematic overview of functional genomic information and multi-omics datasets. It provides valuable customized datasets such as gene co-expression networks and pan-evolutionary features. This database could enhance collaboration and communication, accelerate the utilization of biological resources, and offer new insights into deep ocean biodiversity and evolution.

Materials and methods

Genomic resources and processing

All available genomic resources for deep ocean organisms across Annelida [31–36], Arthropoda [20, 37], Chordata [24, 38–42], Cnidaria [43–49], Echinodermata [50–52], Mollusca [23, 53–61], and Porifera [62] were comprehensively collected from the National Center for Biotechnology Information (NCBI) [63], figshare [64], DRYAD [65], GigaDB [66], and ScienceDB [67]. The collection encompassed genomic sequences, coding sequences, transcript sequences, and protein sequences. Additionally, gene annotations in GFF format were also incorporated. The genomic resources for deep ocean organisms were compiled, incorporating taxonomy, assembly details, habitat information, references, and associated data. As for those organisms with multiple genomic versions, high-quality and full-annotated genomes were selected for subsequent analysis and integration into the database.

Genome annotation and synteny analysis

Protein-coding genes in deep ocean organisms were annotated through the Pfam database [68], the UniProt database [69], the PANTHER database [70], the Gene Ontology (GO) database [68], and the Kyoto Encyclopedia of Genes and Genomes (KEGG) database [71]. Additionally, the Animal Transcription Factor Database (version 4.0) [72] and the iUUCD database [73] were utilized to annotate transcription factors (TFs) and ubiquitin families, respectively. The BUSCO genes were identified at the metazoan level for each organism to support phylogenetic studies [74]. Phylotree.js [75] was employed to visualize three species trees, including deep ocean organisms with full-annotated genomes, deep ocean organisms with chromosome-level genome assemblies, and deep ocean and shallow-water organisms. Micro- and macrosynteny blocks were identified using Pansyn toolkit [76] to illustrate collinearity comparisons. Mitochondrial genomes were included and visualized by OGDRAW [77].

Bulk and single-cell transcriptomic data resources and processing

Publicly available transcriptomic data from the Gene Expression Omnibus (GEO), Genome Sequence Archive, and National Genomics Data Center databases were retrieved. For bulk transcriptomic data, initial quality control was conducted using FastQC (version 0.12.1) [78] to ensure sequencing quality. Adapter and low-quality sequences were trimmed using Trimmomatic (version 0.39) [79] with the following parameters: SLIDINGWINDOW:4:15 MINLEN:36. The cleaned reads were then aligned to the reference genome by using HISAT2 (version 2.2.1) [80]. Gene expression quantification was performed with featureCounts (version 2.0.1) [81] based on the genome annotation file and StringTie (version 2.2.1) [82]. Transcripts per million values were merged into a unified expression matrix for subsequent heatmap visualization and co-expression network construction. For single-cell transcriptomic data, FASTQ files were processed using the CellRanger (version 8.0.1) pipeline [83] with default parameters to perform sampling demultiplexing, barcode processing, and alignment to the reference genome. The generated unique molecular identifier (UMI) count matrices were imported into R (version 4.3.2) and analysed through the R package Seurat (version 4.4.0) [84]. Low-quality cells and ambient RNA con-

tamination were removed based on the following criteria: cells with total RNA counts (nCount_RNA) below 600 or above 5000 were excluded, and cells with mitochondrial gene expression percentage (Percent.mt) outside the 30%–80% range were filtered out. After quality control, the data were normalized and scaled using the 'NormalizeData' and 'ScaleData' functions. Dimensionality reduction was performed using the 'RunUMAP' function and clustering was carried out using the 'FindClusters' function with a resolution parameter set to 0.15. Differentially expressed genes (DEGs) between clusters were identified using the 'FindAllMarkers' function with thresholds of min.pct = 0.25 and logfc.threshold = 0.25. Gene expression patterns were visualized using 'FeaturePlot' and 'VInplot' functions.

Metagenomic data resources and processing

Publicly available metagenomic data from the GEO, Genome Taxonomy Database (GTDB) [85], and published papers were collected. The collection consists of information on microorganisms associated with deep ocean organisms, including genome features, NCBI metadata, and gene structural annotations. Gene functional annotations for assembled microorganism genomes were performed by eggNOG-mapper [86] tool according to the EggNOG, CAZy, GO, and KEGG databases.

Co-expression network construction

Pearson correlation coefficients (PCCs) were calculated between gene pairs to identify high-confidence co-expressed gene pairs, denoted as mutual rank (MR). A higher PCC value means more similar gene expression patterns between genes. The MR algorithm ranks the PCC values, ensuring that low-confidence co-expressed gene pairs are filtered out. The combined PCC and MR values were utilized to construct the gene co-expression network [87, 88] and visualized using Cytoscape (version 3.10.3) [89].

Phylogenetic tree construction

The orthologues and paralogues of deep ocean organisms were identified using OrthoFinder (version 2.5.5) [90]. Protein sequences of one-to-one orthologues were extracted from each orthogroup using an in-house Python script (Supplementary Software S1). Multiple alignments for each orthogroup were performed through MAFFT (version 7.526) [91] with the '–anysymbol' parameter. The aligned sequences were then trimmed using BMGE (version 1.12) tool [92] under default parameters. Phylogenetic trees were constructed using IQ-TREE (version 2.3.5) [93] with '-st AA -m MFP -B 1000 –alrt 1000 –bnni' parameter using maximum likelihood models.

Database construction

The DOO database was constructed under the LAMP (Linux + Apache + MySQL + PHP) environment utilizing established methodologies [94, 95]. The interactive interface was implemented using HTML, CSS, JavaScript, and PHP languages on Ubuntu Linux powered by an Apache server to facilitate users accessing functional genomic and multi-omics information, ensuring smooth data extraction, visualization, and presentation. Data visualization, including charts and graphs, was implemented using Highcharts and the D3 library. DOO is designed with reference to MolluscDB 2.0 database [30] and is now freely available online without registration or login requirements for access.

Results

Overview of DOO database

The DOO database provides a systematic overview of multi-omics resources in deep ocean organisms, serving as a user-friendly platform to explore deep ocean biodiversity and evolution. DOO represents the first comprehensive integration of deep ocean multi-omics datasets, including 72 assembled genomes, 950 bulk transcriptomes, 15 single-cell transcriptomes, and 1112 metagenomes (Table 1). The integration also retrieves taxonomy data from the Global Biodiversity Information Facility (GBIF, https://www.gbif.org), the World Register of Marine Species (WoRMS, https://www.marinespecies.org), and the NCBI Taxonomy database [63], achieving extensive taxonomic coverage across seven phyla (Annelida, Arthropoda, Chordata, Cnidaria, Echinodermata, Mollusca, and Porifera), 16 classes, 35 orders, 42 families, and 68 species (Fig. 1A). DOO provides various genomic information, including assembly details, fossil records, genome phylogeny, BUSCO genes, gene structure, gene annotation, expression profiles, gene families, TFs, and ubiquitin families. Additionally, DOO integrates species-featured resources for gene co-expression networks, single-cell transcriptomes, symbiont genomes, and micro- and macrosynteny analysis for inferring karyotype evolution. Moreover, DOO offers convenient toolkits for gene search, functional enrichment analysis, BLAST alignments, PCR primer design and Jbrowse visualization (Fig. 1B and Table 1).

Species and multi-omics resources

At the genome level, DOO integrates 46 high-quality genomes with full-annotated gene information (Table 2) and provides a phylogenetic tree to elucidate the evolutionary relationship among deep ocean organisms on the main page. Users can select species or click on names in the 'Taxonomy' section to access comprehensive genomic resources, which include species information, taxonomic classification, and an external link to GBIF, WoRMS, and the NCBI Taxonomy database for further details (Fig. 2A and B). Fossils serve as pivotal evidence for reconstructingdeep ocean phylogenetics and evolutionary trajectories [96, 97]. DOO provides 1413 fossil records for deep ocean taxon collected from the Paleobiology Database (PBDB) [98] and the World Register of Deep-Sea Species in the 'Paleobiology' part (Fig. 2C). Users can click on a taxon to access the PBDB database, which provides linked paleobiology information, including taxonomic history, classification, morphology, age range, and collection location of fossils.

DOO summarizes multi-omics data resources across several modules. The 'Genomic Data' module offers detailed information, including genome assembly, BUSCO completeness, gene number, and corresponding published literatures (Fig. 2D). A total of 950 transcriptomic datasets, selected for further expression and network analysis across various developmental stages, tissues, and treatments, are included in the 'Transcriptomic Data' module (Fig. 2E). The 'Single-cell Data' module summarizes single-cell RNA-seq datasets from deep ocean tubeworm *Paraescarpia echinospica*, coupled with single-nucleus RNA-seq datasets from deep ocean mussel *Gigantidas platifrons* (Fig. 2F). The 'Metagenomic Data'

Table 1. Summary of data composition in DOO database

Data	Statistics
Phylum/class/order/family/species	7/16/35/42/68
Assembled/fully annotated genomes	72/46
Tracks in Jbrowse	401
Paleobiological records	1413
Protein-coding genes	1 449 954
GO/KEGG/Pfam/Panther/Interpro annotation	3 606 859/18 539/1 281 713/928 123/3 640 096
TFs/ubiquitin families	60 866/82 844
Gene families/associated genes	212 597/1 385 060
Macrosynteny gene pairs	398 247
Busco gene	16 015
Mitochondrial genomic datasets	53
Transcriptomic data	950
Co-expressed networks/positive/negative gene pairs	6/1 485 344/520 014
Single-cell or single-nucleus datasets	15
Cell types	14
Cell markers	1793
Symbiotic microorganisms	1112
Annotated microbial genes	92 682

module in DOO provides 1112 metagenomic samples from deep ocean symbiotic communities, capturing genomic features of microorganisms associated with deep ocean organisms (Fig. 2G). Deep ocean organisms engage in chemosynthetic symbioses with microorganisms to derive energy. Additionally, certain symbiotic microorganisms produce bioluminescence for predator deterrence and prey attraction. The highly specialized adaptation substantially enhances their survival in extreme habitats [99].

Genome annotation

Advancements in sequencing techniques provide an enhanced understanding of deep ocean evolution through the ever-expanding genomic datasets [12, 100]. However, the potential of these datasets is often limited by fragmentation across studies [101], underscoring an urgent need for integrated genomic resources to ensure data accessibility, standardized functional annotation, and cross-species comparative analysis [102]. The 'Genome' module in DOO is aimed at providing accurate genomic information, cross-species comparisons, and pan-evolutionary analysis for comparative genomic and phylogenetic study.

Based on the 'Gene Search' module, users can access the structural and functional annotations for 1 449 954 protein-coding genes from 46 deep ocean organisms (Fig. 3A). The 'Mitogenomic Data' module provides mitochondrial genomes and annotations for 53 deep ocean organisms (Fig. 3B), aiding the exploration of evolutionary patterns in these organisms [103]. Gene functional annotations are performed for protein-coding genes using several widely used databases (Pfam, InterPro [104], PANTHER, GO, and KEGG) (Fig. 3C). BUSCO gene completeness and uniqueness can serve as a quantitative assessment of genome assembly and annotation completeness [105]. The 'BUSCO Gene' module identifies 16 015 BUSCO genes at the metazoan level for each species (Fig. 3D). TFs critically regulate fundamental biological processes, including growth trajectories, developmental programs, and environmental stress responses [106, 107]. Ubiquitins function as post-translational modifiers mediating protein degradation, cell cycle control, and DNA repair [108, 109]. Both families are annotated in the 'TFs/Ubs' module, comprising 60 866 TFs and 82 844 ubiquitins (Fig. 3E), respectively. The 'Gene Family' module identifies and annotates 212 597 gene families through clustering analysis, allowing users to access detailed gene family information, including gene family members and PANTHER annotations (Fig. 3F). The 'Symbiotic Microorganism' module integrates 131 assembled symbiotic microorganisms (Fig. 3G) from the NCBI and GTDB [85] databases, including genome characteristics such as the quality of microorganisms' genomes through checkM assessment [110] and NCBI metadata, including assembly information and genome category, to facilitate the study of symbiotic interaction. Additionally, the module provides gene structural and functional annotations for symbiont genomes, with all annotations performed by the eggNOG-mapper [86] tool mainly in accordance with the EggNOG, CAZy [111], GO, and KEGG databases (Fig. 3H).

Bulk and single-cell transcriptomic modules

The transcriptome is continuously responding to dynamic physiological and environmental conditions, and its profiling reflects the physiological and functional state of tissues or cells, inferring insights into the genetic basis of extreme adaptation in a specific tissue or cell [112, 113]. The 'Transcriptome' and 'Single-cell' modules in DOO provide gene expression profiles and co-expression networks to infer potential gene regulatory networks and molecular signatures of adaptation mechanisms in deep ocean organisms.

The 'Network Visualization' module represents gene co-expression networks for deep ocean organisms across various tissues and developmental stages, encompassing 2 005 358 co-expressed gene pairs, and dynamic expression views to highlight the tissue- and stress-specific DEGs (Fig. 4A and B). Users can select specific genes to explore corresponding co-expression networks and perform expression profiling analysis (Fig. 4C). Detailed information about co-expressed genes and their relationships is also provided. The "Cell Atlas" module annotates various cell types across deep ocean organisms, presenting a landscape of all cell types (Fig. 4D) and correlations across different repeats or conditions. The 'Cell Marker' module facilitates access to expression levels of marker genes in specific cell types or clusters (Fig. 4E), along with the top 50 marker genes and corresponding *P*-values. The 'Single-Cell Gene Expression' module visualizes the expression patterns of

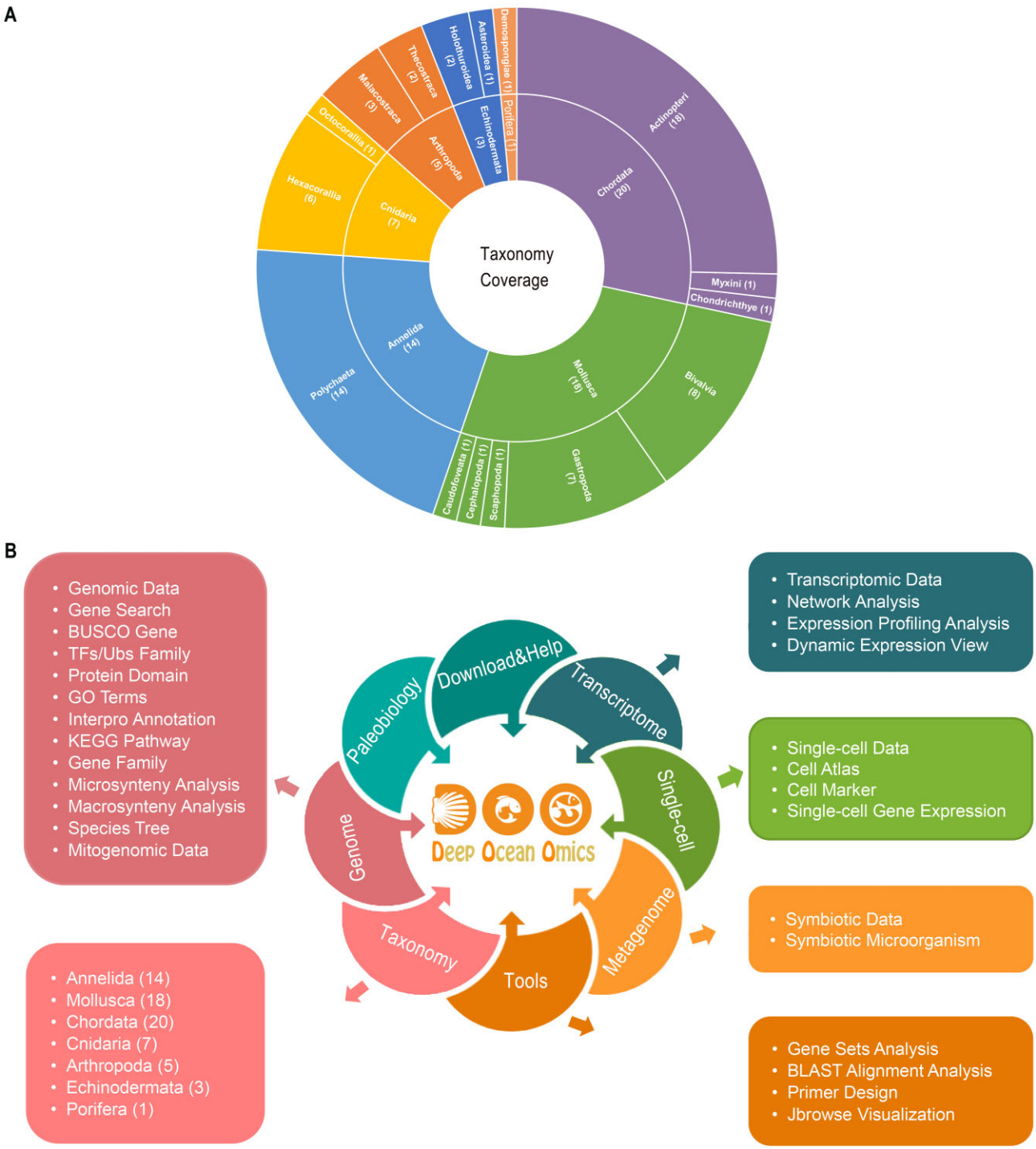


Figure 1. Overview of DOO database. (**A**) Taxonomy coverage of deep ocean organisms across phyla and classes from the inner layer to the outer layer. (**B**) Database architecture and function. Different colours represent different modules in DOO.

protein-coding genes across different cell types by using violin plots (Fig. 4F).

Pan-evolutionary module

Genomic synteny analysis serves as the cornerstone for ancestral karyotype reconstruction, providing critical insights into the mechanisms driving genome diversity, chromosomal rearrangements, and cross-species evolutionary conservation [114, 115]. Several well-known tools for synteny analysis, such as MCScanX [116], SyMAP [117], and DAGchainer [118], primarily focus on microsynteny analysis between closely related species, limiting comparisons in deep phylogeny in scenario such as the deep ocean organisms. The PanSyn tool [76] has been developed to investigate the macroevolution of genome architecture from ancestors to extant species through deep phylogenetic comparisons. In DOO, PanSyn integrated micro- and macrosynteny analyses to reconstruct the ancestral karyotype utilising conserved ancestral linkage groups (ALGs) from *Patinopecten yessoensis*, *Branchios-*

Table 2. Summary of 46 full-annotated genomes from deep ocean organisms

Phylum	Species	Size (Mb)	Assembly level	BUSCO completeness (%)	Scaffold/contig N50 (kb)	GC content (%)	Repeat rate (%)	PubMed
Annelida	*Branchipolynoe longqiensis*	1860	Chromosome	89.30	97 050/566.3	44.35	71.14	[33]
	Riftia pachyptila	560.7	Contig	99.37	2870	40.49	29.99	[34]
	Ridgeia piscesae	658.2	Contig	94.00	365.4	41.00	–	[35]
	Paralvinella palmiformis	601	Contig	96.00	557.2	36.00	–	[35]
	Paraescarpia echinospica	1090	Chromosome	96.40	67 235/253.6	41.00	55.00	[32]
	Osedax frankpressi	285	Scaffold	80.10	426	29.08	29.16	[34]
	Oasisia alvinae	808	Scaffold	96.90	2975	41.04	50.15	[34]
	Osedax fenrisi	699.1	Chromosome	90.10	99 500/1900	27.50	–	–
	Oasisia sp. SBJ-2009	764.2	Chromosome	95.81	54 000/2100	41.5	–	–
	Lamellibrachia satsuma	736	Scaffold	95.50	8074/6600	40.2	–	[36]
	Lamellibrachia luymesi	687.71	Scaffold	95.80	373/24.1	40.16	36.92	[31]
	Lamellibrachia columna	879.7	Chromosome	97.30	472.9/472.5	40.5	–	–
Arthropoda	*Bathynomus jamesi*	5892	Contig	94.80	586.5	37.28	85.32	[37]
	Hirondellea gigas	139 200	Chromosome	89.34	458 700/466.1	38.80	72.75	[20]
Chordata	*Amblyraja radiata*	2600	Chromosome	96.70	62 100/1500	44.50	–	–
	Synaphobranchus kaupii	1500	Chromosome	97.90	76 800/519.8	42.5	–	[38]
	Pseudoliparis swirei	626.4	Chromosome	96.00	25 738/4219	44.00	–	[39]
	Pseudoliparis sp. Yap Trench	731.75	Scaffold	96.17	1260/750	44	53.61	[40]
	Aldrovandia affinis	1230	Scaffold	96.00	3800/2000	43.00	–	[38]
	Chaunax sp.	709.63	Chromosome	97.06	29 421/15 241	40.68	36.14	[41]
	Coryphaenoides rupestris	838.2	Scaffold	95.50	188.2/62.4	47.00	–	[24]
	Ilyophis brunneus	1500	Chromosome	94.30	115 310/2760	42.09	48.41	[42]
Cnidaria	*Actinernus sp. WN-2022*	1400	Scaffold	94.55	71 937/1102	39.14	61.12	[43]
	Actinoscyphia liui	522	Chromosome	94.70	58 422/3400	38	44.06	[44]
	Actinostola sp. Cb2023	424.3	Scaffold	83.20	383.1/372.9	38.70	25.50	[45]
	Alvinactis idsseensis sp. Nov.	456.58	Chromosome	92.80	27 561/1715	38.84	63.27	[46]
	Lophelia pertusa	556.9	Scaffold	88.90	1600/438.6	39.40	57.37	[47]
	Paraphelliactis xishaensis sp. nov.	542.9	Contig	92.90	761.1	37.50	56.60	[48]
	Trachythela sp. YZ-2020	578.2	Contig	90.70	3563	37.50	57.88	[49]
Echinodermata	*Paelopatides sp. Yap*	3700	Contig	89.40	137.1	38.19	76.06	[50]
	Zoroaster sp. YZ-2022	1000	Chromosome	95.91	40 400/376.2	38	63.90	[51]
	Chiridota heheva	1100	Chromosome	94.50	53 240/1220	37.12	70.80	[52]
Mollusca	*Chaetoderma shenloong*	2400	Chromosome	89.52	141 459/785.5	40.93	55.81	[53]
	Siphonodentalium dalli	2 300	Chromosome	95.40	234 870/2110	38	0.76	[54]
	Gigantopelta aegis	1100	Chromosome	94.60	81 591/461.8	37	56.00	[55]
	Bathymodiolus brooksi	2000	Chromosome	94.00	130 400/2900	35	–	–
	Bathymodiolus septemdierum	1400	Chromosome	97.60	99 700/2300	34.5	–	–
	Dracogyra subfusca	1160	Scaffold	32.70	5.9/4.1	37	56.00	[55]
	Chrysomallon squamiferum	404.6	Chromosome	96.60	30 198/1883	29.82	25.22	[56]
	Bathyacmaea lactea	754.3	Contig	94.30	1570	32.67	61.40	[57]
	Architeuthis dux	2700	Scaffold	90.40	4852/5.7	36.10	48.21	[58]
	Placopecten magellanicus	1830	Chromosome	94.10	86 713/6535	38.06	43.04	[59]
	Gigantidas platifrons	1640	Scaffold	96.30	343.3/12.6	34.30	47.93	[23]
	Conchocele bisecta	1900	Chromosome	95.40	106 900/488.6	39.22	66.96	[60]
	Archivesica marissinica	1500	Chromosome	93.50	74 300/79.1	39.00	52.81	[61]
Porifera	*Geodia barretti*	144.8	Scaffold	71.50	48.4	49.30	19.05	[62]

toma floridae, and *Nematostella vectensis*, track syntenic block evolution, classify gene clusters, and elucidate the pan-evolutionary features of genome structure.

The 'Microsynteny Analysis' module represents a correlation heatmap of genome similarities among 21 deep ocean organisms with chromosome-level genome assembly and annotation. Users can click one square to view the distribution and correspondence of microsynteny blocks between two species (Fig. 5A–C). The 'Macrosynteny Analysis' module compares karyotypes between these three ALGs and deep ocean organisms to identify ancestral gene families and macrosynteny blocks. A total of 10 863 ancestral gene families are identified, with each species retaining >5800 of these conserved lineages (Supplementary Table S1). Users can select different reference ALGs to view detailed synteny blocks and their distribution on chromosomes (Fig. 5D–G). Additionally, users can download macrosynteny dot plots, homologous gene pairs, and related gene sequences.

Case study

Studies have highlighted the crucial role and evolutionary importance of microbial symbionts in facilitating deep ocean organisms to adapt to extreme environments [23, 119]. Cathepsins (CTSs), members of the papain family cysteine proteases, are essential for host–symbiont interaction through endopeptidase immune response and cell death regulation [120, 121]. They also play pivotal roles in maintaining bacteriocyte homeostasis and controlling the symbiont population [31, 32]. In this context, the *CTS* gene from deep ocean tubeworm *P. echi-*

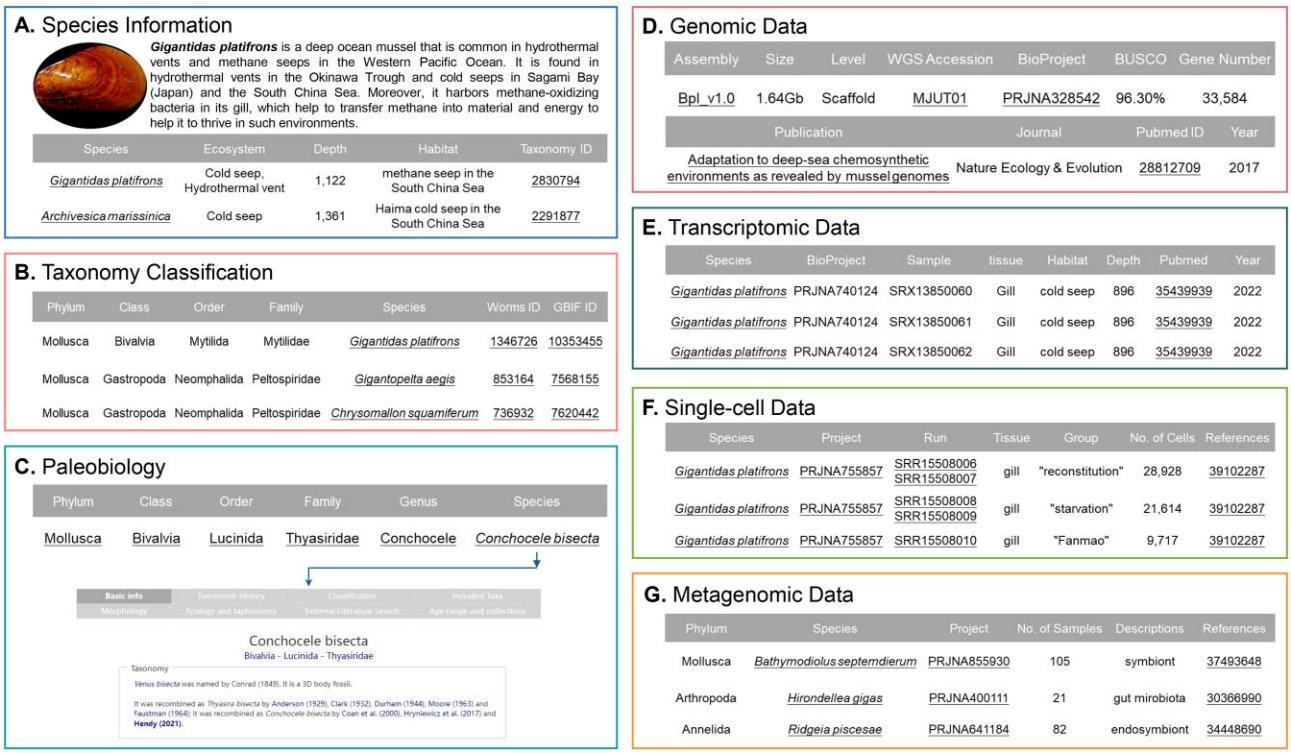


Figure 2. Taxonomy, paleobiology, and multi-omics resources in DOO. (**A**) Species information. (**B**) Taxonomy classification. (**C**) Paleobiology. (**D**) Genomic data. (**E**) Transcriptomic data. (**F**) Single-cell transcriptomic data. (**G**) Metagenomic data. The underlined parts provide a link for users to access detailed information.

nospica was used as an example to explore its potential function at the multi-omics and evolutionary level by using DOO.

The 'Gene Search' module reveals the function of *CTS* gene as a lysosomal cysteine protease (CTS) and its participation in immune response and apoptosis (Fig. 6A). Jbrowse visualization shows that the gene is primarily expressed in trophosome tissue, with RNA-seq peaks enriched in the coding region (Fig. 6B). Trophosome tissue is a specialized symbiotic organ that houses sulphide-oxidising gammaproteobacterial endosymbionts in deep ocean tubeworm [32, 122]. As for the co-expression network of *CTS* (Fig. 6C), positively co-expressed genes were observed to be related to pivotal biological processes such as metabolism regulation (*SAM Mt*), carbon supply for symbionts (*CA*), maintenance of symbiotic homeostasis (*MUC4*), and low oxygen response (*HIF-1α* inhibitor). These co-expressed genes have been reported to be potentially involved in the symbiotic interaction [123–126]. Further analysis of GO enrichment for these co-expressed genes uncovered significant enrichments in specific biological processes such as metabolic processes, oxidoreductase activity, steroid hydroxylase activity, methyltransferase activity, iron ion binding, and oxygen transport and binding (Fig. 6D). These results highlight the processes that are related to the regulation of symbiotic metabolites, the maintenance of symbiotic homeostasis, and chemosynthetic energy exchange. Examination of the single-cell expression levels of *CTS* by using the 'Single Cell' module showed that *CTS* is highly expressed in bacteriocyte-periphery (Bac-P) and bacteriocyte-center (Bac-C) cells (Fig. 6E), and it has been validated as a marker gene of these two bacteriocytes [126]. Bac-P cells host symbionts to facilitate aerobic carbon fixation, symbiont digestion for nutrient harvest, and shuttle nitrogen waste towards the centre. Bac-C cells harbour symbionts to drive anaerobic denitrification for ammonia detoxificationwhile processing host-derived metabolites [126]. Clearly, their strong interaction pattern with symbionts is consistent with the previous analysis results based on the information acquired from DOO. *CTS*s are evolutionarily conserved and used by diverse animal hosts to digest their symbiotic microorganisms [31, 32]. Based on the Jbrowse visualization of RNA-seq peaks in DOO, *CTS* also shows high expression in symbiont-hosting tissues of some other deep ocean organisms. Additionally, comparative genomic analysis for *CTS* gene families was performed on the basis of the data of the 'Gene Family' and 'Protein Domain' modules. The results revealed that most deep ocean organisms in DOO carry this gene, except for *Trachythela sp. YZ-2020*, *Paraphelliactis xishaensis sp. nov.*, and *Ilyophis brunneus*. A significant change was observed in the *CTS* gene family size, with the highest being 37 and the lowest being 0. This finding indicated that *CTS* genes were fast evolving in deep ocean organisms. Additionally, the protein domain of *CTS* genes among deep ocean organisms was explored, and the results showed that most *CTS* genes contain two domains: Peptidase_C1 (papain family cysteine protease) and Inhibitor_I29 [cathepsin propeptide inhibitor domain (I29)]. Moreover, a small number of *CTS* genes contain peptidase_C1 only. Additionally, another type of *CTS* gene containing three domains was discovered among deep ocean molluscs (Fig. 6F). In summary, the case study demonstrates the practicality and reliability of the DOO database, which can provide new insights into deep ocean adaptation and evolution.

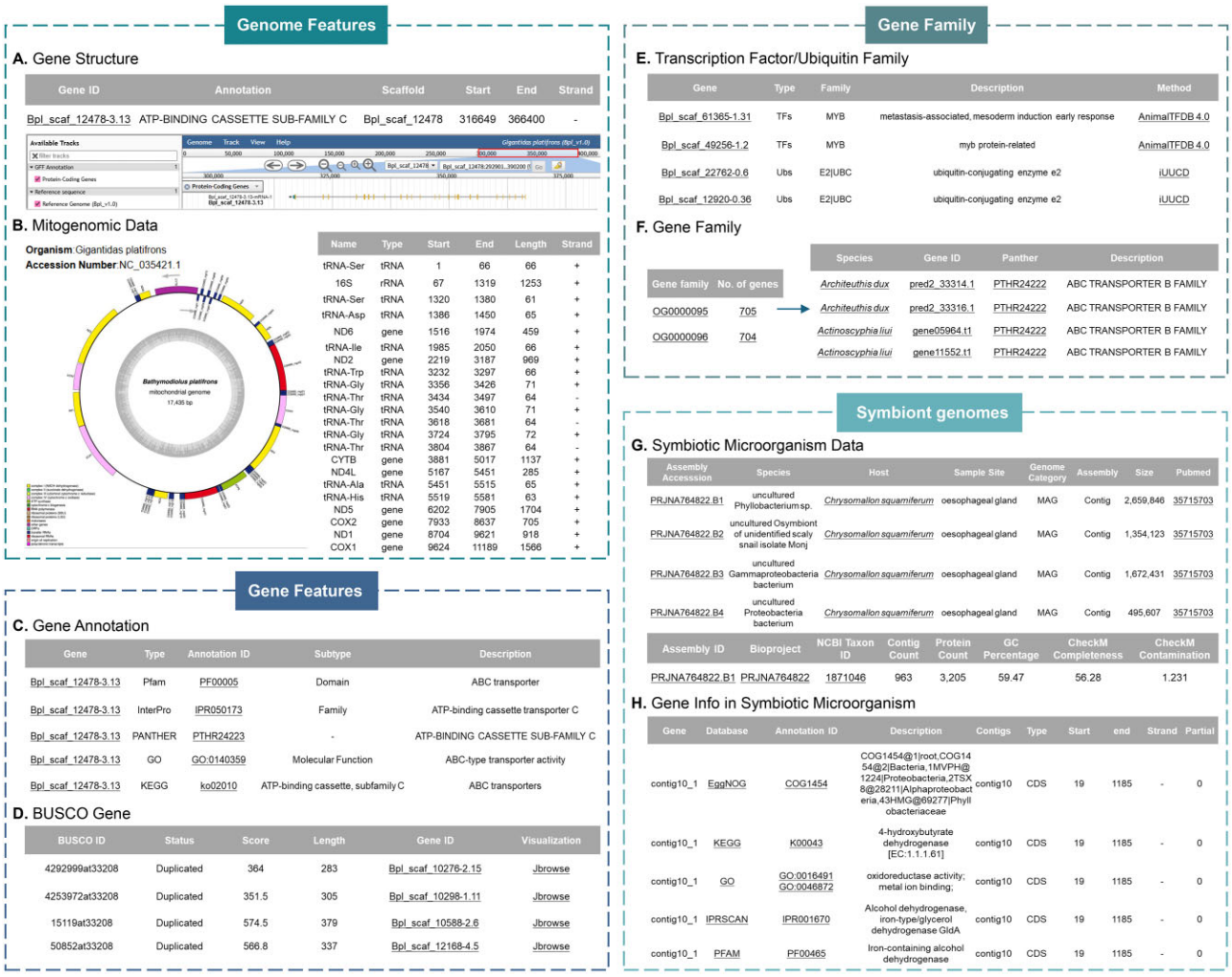


Figure 3. Genome annotation in DOO. (**A**) Gene structure. (**B**) Mitogenomic data. (**C**) Gene annotation. (**D**) BUSCO gene. (**E**) Transcription factors/ubiquitin family. (**F**) Gene family. (**G**) Symbiotic microorganism data. (**H**) Gene info in symbiotic microorganism. The underlined parts provide a link for users to access detailed information.

Future directions

DOO represents the first comprehensive deep ocean multi-omics platform to date, integrating genomic, bulk and single-cell transcriptomic, metagenomic, mitogenomic, and paleobiology resources across seven phyla. Considering that deep ocean adaptation and evolution is a highly active research field, and genomic resources are rapidly expanding, additional high-quality, full-annotated genomes and multi-omics data will be continuously incorporated into DOO. Currently, there are some limitations for our database, mainly including the limited availability of deep ocean genomes, insufficient multi-omics data integration and visualization, and a lack of systematic comparative genomic analysis. To address this gap, we plan to regularly update the database, with a focus on the following improvements: (i) incorporating broader omics data layers (proteome, epigenome, and metabolome) alongside expanded genomic resources; (ii) expanding taxonomy coverage through leveraging ongoing global deep ocean genome initiatives, such as the United Nations Decade of Ocean Science for Sustainable Development, Mariana Trench Environment and Ecology Research, and the Deep Ocean Microbiomes and Ecosystems project, incorporating more species and assembled symbiont genomes; (iii) enhancing comparative genomic insights through systematic comparison between deep ocean organisms and their nearshore relatives; (iv) implementing AI-driven models into deep mining of multi-omics data [127]; and (v) developing user-friendly tools for cross-species multi-omics comparison and visualization.

Acknowledgements

The authors wish to thank all researchers who have generated invaluable deep ocean genomic resources that have been integrated into the DOO database. The authors thank Yi Lan and Ting Xu for their help in identifying deep ocean organisms. The authors also thank Yan Yan and Hongxiu Xiao for their support in single-cell RNA-seq data collection and analysis. Additionally, the authors thank Weizhi Song, Shan Zhang, and Ziyue Xu for their assistance in improving the website.

Author contributions: Longjun Wu (Conceptualization [equal], Funding acquisition [equal], Investigation [equal], Project administration [equal], Resources [equal], Supervision [equal], Writing – review & editing [equal]), Jiajie She (Conceptualization [equal], Data curation [equal], Formal analysis [equal], Funding acquisition [equal], Methodology [equal], Project administration [equal], Resources [equal], Software

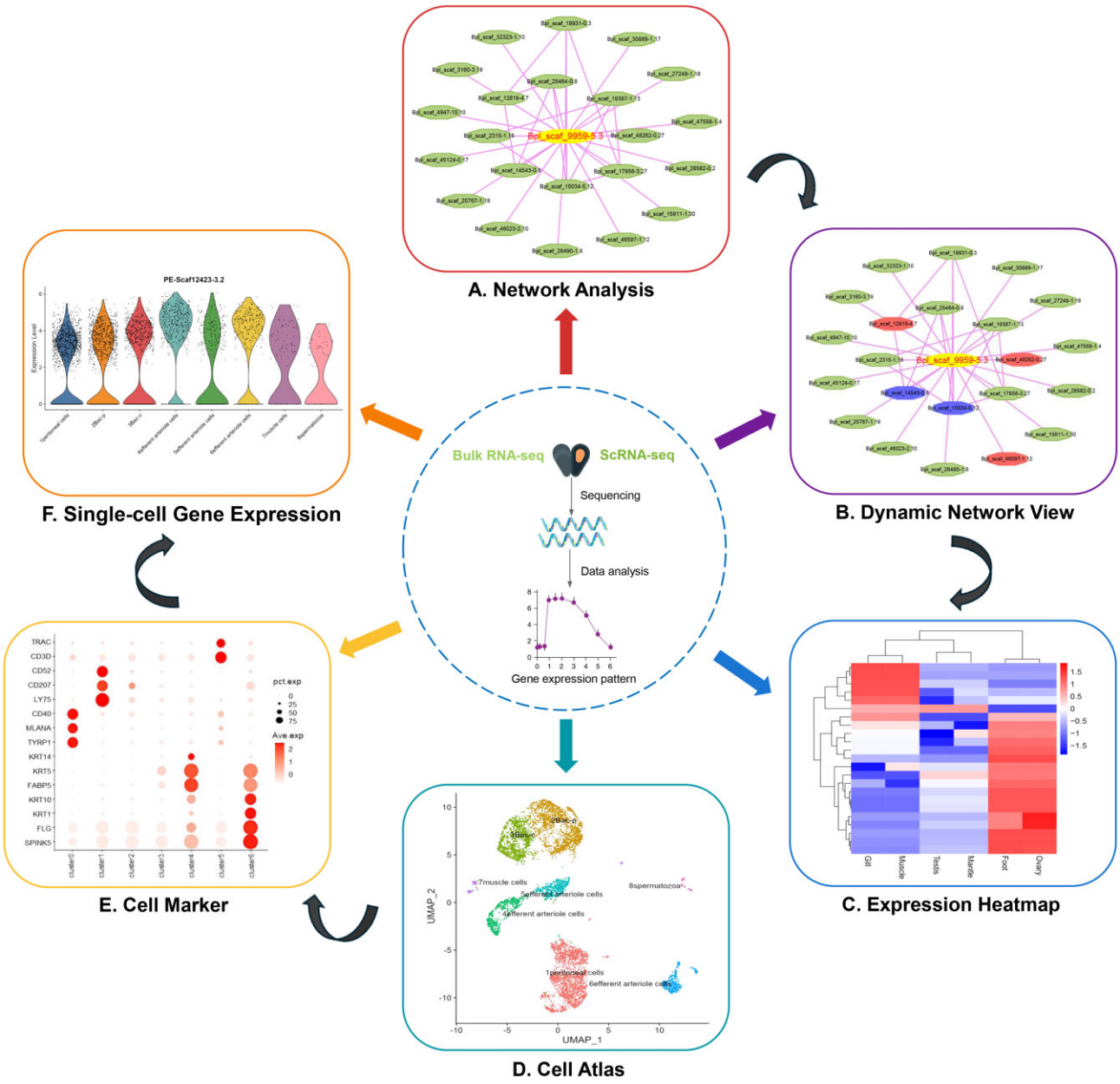


Figure 4. Visualization of bulk and single-cell transcriptomes in DOO. (**A**) Network analysis. Query genes are highlighted in yellow. And the blue line represents a negative co-expression relationship between two genes, and the pink line represents a positive co-expression relationship. (**B**) Dynamic network view. The red and blue hexagons represent upregulated and downregulated genes in particular conditions, respectively. (**C**) Expression heatmap of interested gene lists among different samples. The deeper the colour, the higher the gene expression. (**D**) Cell atlas. Different colours represent different cell types. (**E**) Expression profiles of cell markers in different clusters. The size of dots represents the proportion of gene-expressing cells, and the colour intensity of the dots represents the average level of gene expression. (**F**) Single-cell gene expression. It represents the expression pattern of protein-coding genes on the violin plot across different cell types.

[equal], Visualization [equal], Writing – original draft [equal], Writing – review & editing [equal]), and Pei-Yuan Qian (Conceptualization [equal], Funding acquisition [equal], Resources [equal], Supervision [equal], Writing – review & editing [equal])

Supplementary data

Supplementary data is available at NAR online.

Conflict of interest

None declared.

Funding

This work was funded by grants (2021HJ01, HJRC2022001, and SMSEGL24SC01) of the Southern Marine Science and Engineering Guangdong Laboratory (Guangzhou), and Otto Poon Center for Climate Resilience and Sustainability (CCRS25SC01), Early Career Scheme (26104824) from the Research Grant Council (RGC) of the Hong Kong Special

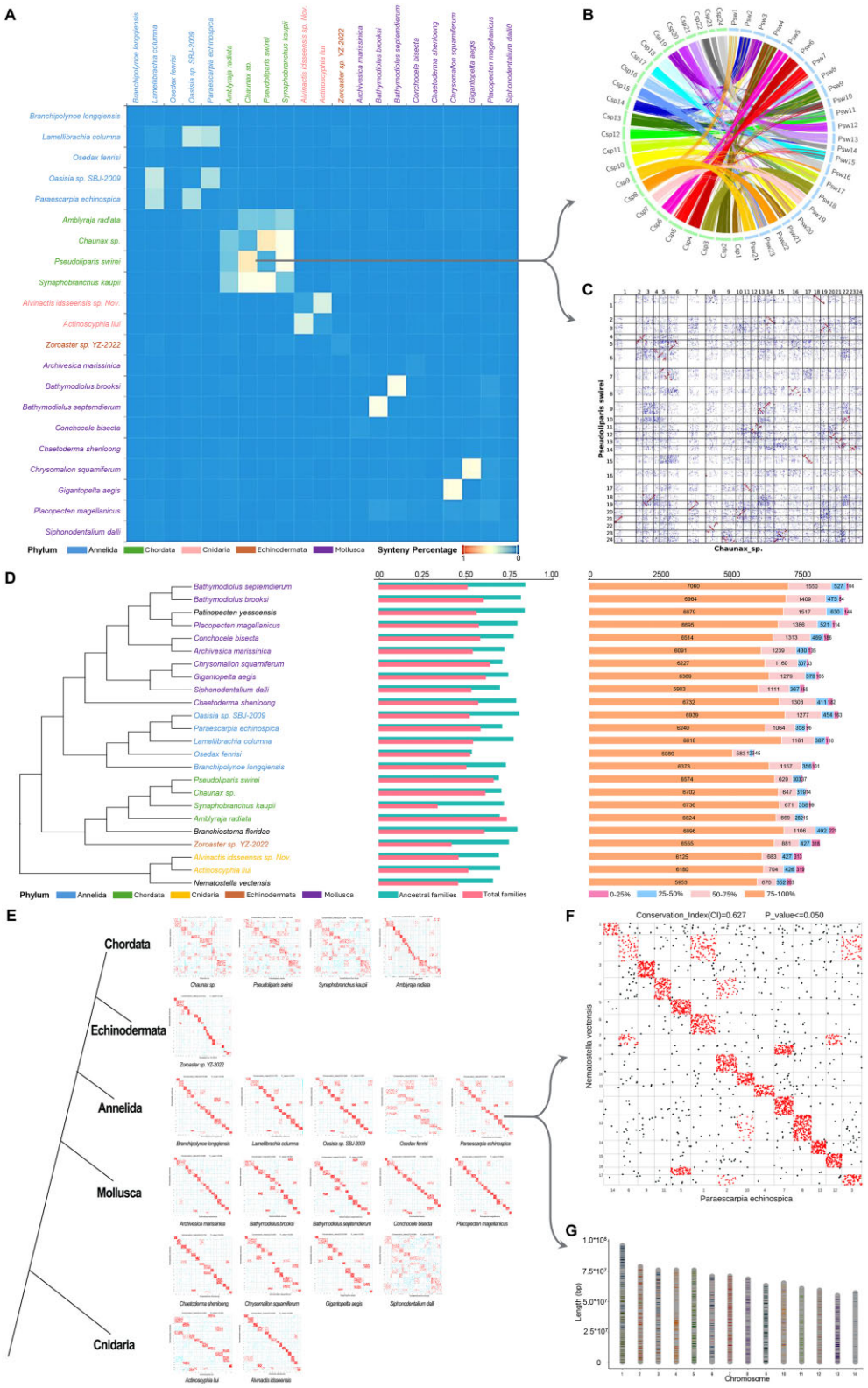


Figure 5. Pan-evolutionary analysis for identifying conserved gene clusters and karyotypes. (**A**–**C**) Microsynteny analysis among deep ocean organisms. (**A**) Correlation squares among deep ocean organisms based on microsynteny blocks. (**B**) Circos plot for distribution and visualization of microsynteny blocks between two deep ocean organisms. (**C**) Dot plot for distribution and visualization of microsynteny blocks between two deep ocean organisms. (**D**–**G**) Macrosynteny analysis between deep ocean organisms with chromosome-level genome assembly and three ancestral linkage groups (ALGs). (**D**) Identification of ancestral gene families across deep ocean organisms and ALGs. The left, middle, and right represent evolutionary relationships, the percentage of ancestral gene families for each species, and the number of ancestral gene families in each species, respectively. (**E**) Macrosynteny analysis for deep ocean organisms across five phyla. (**F**) Analysis and visualization of karyotype evolution between ALGs and deep ocean organisms. (**G**) Visualization of karyotype comparisons between the ancestors and deep ocean organisms.

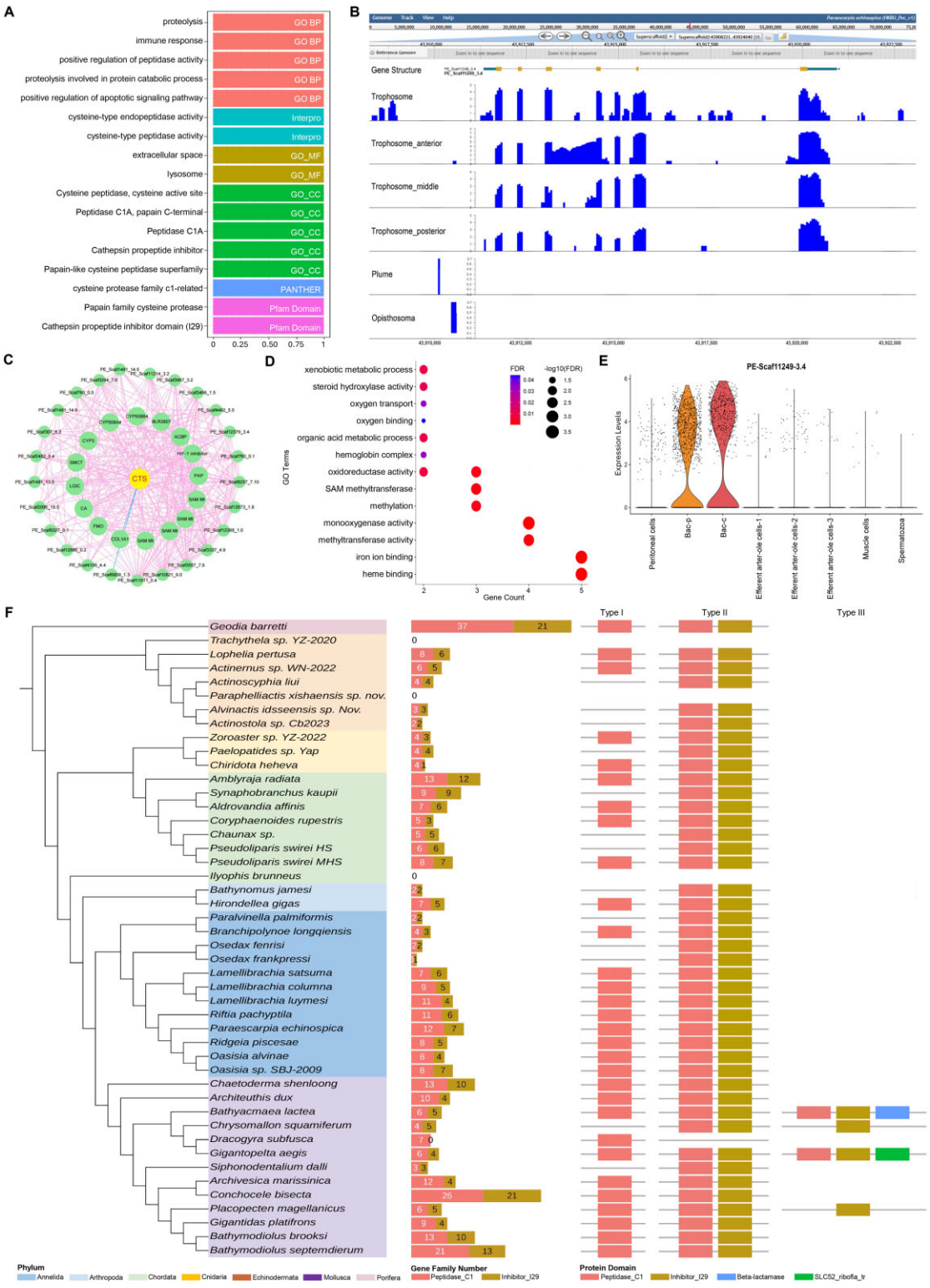


Figure 6. Case study using *CTS* gene. (**A**) Functional annotation for *CTS* gene. Different colours represent different functional types. (**B**) Jbrowse visualization of *CTS* gene with RNA-seq peaks in different tissues. (**C**) Co-expression network of *CTS* gene. *SAM Mt*, SAM-dependent methyltransferase; *CA*, carbonic anhydrase; *MUC4*, mucin protein; *BLR2807*, globin family member; *LGIC*, ligand-gated ion channel protein; *SMCT*, sodium-coupled monocarboxylate transporter, *HIF-1α* inhibitor, hypoxia-inducible factor 1-alpha inhibitor; *ACBP*, acyl-CoA-binding protein; *CYP3* and *CYP508A4*, cytochrome P450 family member; *COL1A1*, alpha-1 collagen protein. (**D**) GO enrichment analysis results for all genes in the co-expression network. (**E**) Violin plot of single-cell expression of *CTS* gene across cell types. (**F**) Phylogenetic tree and protein domain compositions of *CTS* gene families in 46 deep ocean organisms. The left and middle represent the phylogenetic tree of deep ocean organisms and the number of *CTS* gene families with Peptidase_C1 or Inhibitor_I29 domain, respectively. The right represents protein domain compositions of *CTS* gene families, mainly including three types: (i) type I, *CTS* genes with Peptidase_C1 domain only; (ii) type II, *CTS* genes with Peptidase_C1 and Inhibitor_I29 domains; and (iii) type III, *CTS* genes with Inhibitor_I29 or three domains only. Peptidase_C1 (PF00112), papain family cysteine protease; Inhibitor_I29 (PF08246), cathepsin propeptide inhibitor domain (I29); Beta-lactamase (PF00144), beta-lactamase; SLC52_ribofla_tr (PF06237), solute carrier family 52, riboflavin transporter.

Administrative Region, and the Science and Technology Innovation Committee of Shenzhen (JCYJ20220530151207016). Funding to pay the Open Access publication charges for this article was provided by Early Career Scheme (26104824) from Hong Kong RGC.

Data availability

All data and materials in DOO are available to freely access and download at https://DeepOceanOmics.org. The information about database statistics, software, and data submission in this study is shown in the 'Help' module.

References

1. McClain CR, Hardy SM. The dynamics of biogeographic ranges in the deep sea. *Proc Biol Sci* 2010;277:3533–46.
2. Baker M, Ramirez-Llodra E, Billett D. Deep-sea biodiversity and life-history processes preface. *Deep-Sea Res Pt II* 2013;92:1–8.
3. Danovaro R, Fanelli E, Aguzzi J. *et al.* Ecological variables for developing a global deep-ocean monitoring and conservation strategy. *Nat Ecol Evol* 2020;4:181–92. https://doi.org/10.1038/s41559-019-1091-z
4. Levin LA. Ecology of cold seep sediments: interactions of fauna with flow, chemistry and microbes. *Oceanogr Mar Biol* 2005;43:1–46.
5. Campbell KA. The ecology of deep-sea hydrothermal vents. *Science* 2000;289:730–1. https://doi.org/10.1126/science.289.5480.730
6. Felbeck H, Somero GN. Primary production in deep-sea hydrothermal vent organisms—roles of sulfide-oxidizing bacteria. *Trends Biochem Sci* 1982;7:201–4.
7. Rona PA. Deep-diving manned research submersibles. *Mar Technol Soc J* 1999;33:13–25. https://doi.org/10.4031/MTSJ.33.4.3
8. Chapman ASA, Beaulieu SE, Colaco A. *et al.* sFDvent: a global trait database for deep-sea hydrothermal-vent fauna. *Global Ecol Biogeogr* 2019;28:1538–51. https://doi.org/10.1111/geb.12975
9. Pereira OS, Jacobsen M, Carson R. *et al.* Understanding and valuing human connections to deep-sea methane seeps off Costa Rica. *Ecol Econ* 2024;223:108228. https://doi.org/10.1016/j.ecolecon.2024.108228
10. Baldrighi E, Giovannelli D, D'Errico G. *et al.* Exploring the relationship between macrofaunal biodiversity and ecosystem functioning in the deep sea. *Front Mar Sci* 2017;4:198. https://doi.org/10.3389/fmars.2017.00198
11. Käss M, Chikina M, Vedenin A. *et al.* Traits and drivers: functioning of macrobenthic communities across the deep Fram Strait (Arctic Ocean). *Ecol Indic* 2021;123:107324. https://doi.org/10.1016/j.ecolind.2020.107324
12. Feng JC, Liang JZ, Cai YP. *et al.* Deep-sea organisms research oriented by deep-sea technologies development. *Sci Bull* 2022;67:1802–16. https://doi.org/10.1016/j.scib.2022.07.016
13. Liang JZ, Feng JC, Zhang S. *et al.* Role of deep-sea equipment in promoting the forefront of studies on life in extreme environments. *iScience* 2021;24:103299. https://doi.org/10.1016/j.isci.2021.103299
14. Lac L, Leung CK, Hu PZ. Computational frameworks integrating deep learning and statistical models in mining multimodal omics data. *J Biomed Inf* 2024;152:104629. https://doi.org/10.1016/j.jbi.2024.104629
15. Lin YT, Xu T, Chi-Ho J. *et al.* Interactions among deep-sea mussels and their epibiotic and endosymbiotic chemoautotrophic bacteria: insights from multi-omics analysis. *Zool Res* 2023;44:106–25. https://doi.org/10.24272/j.issn.2095-8137.2022.279
16. Wang C, Zheng RK, Sun CM. Multi-omics analyses provide insights into the sulfur metabolism of a novel deep-sea sulfate-reducing bacterium. *iScience* 2024;27:110095.
17. Zheng RK, Wang C, Sun CM. Deep-sea *in situ* and laboratory multi-omics provide insights into the sulfur assimilation of a deep-sea *Chloroflexota* bacterium. *mBio* 2024;15:e0000424. https://doi.org/10.1128/mbio.00004-24
18. Mao N, Shao WT, Cai YF. *et al.* Comparative omics analysis of a deep-sea barnacle species (Cirripedia, Scalpellomorpha) and shallow-water barnacle species provides insights into deep-sea adaptation. *Front Mar Sci* 2024;10:1269411.
19. Dong LH, Zhang Y, Fu BQ. *et al.* Reliable biological and multi-omics research through biometrology. *Anal Bioanal Chem* 2024;416:3645–63. https://doi.org/10.1007/s00216-024-05239-3
20. Zhang H, Sun S, Liu J. *et al.* The amphipod genome reveals population dynamics and adaptations to hadal environment. *Cell* 2025;188:1378–92. https://doi.org/10.1016/j.cell.2025.01.030
21. Xu H, Fang C, Xu W. *et al.* Evolution and genetic adaptation of fishes to the deep sea. *Cell* 2025;188:1393–408. https://doi.org/10.1016/j.cell.2025.01.002
22. Xiao X, Zhao W, Song Z. *et al.* Microbial ecosystems and ecological driving forces in the deepest ocean sediments. *Cell* 2025;188:1363–77. https://doi.org/10.1016/j.cell.2024.12.036
23. Sun J, Zhang Y, Xu T. *et al.* Adaptation to deep-sea chemosynthetic environments as revealed by mussel genomes. *Nat Ecol Evol* 2017;1:0121. https://doi.org/10.1038/s41559-017-0121
24. Gaither MR, Gkafas GA, de Jong M. *et al.* Genomics of habitat choice and adaptive evolution in a deep-sea fish. *Nat Ecol Evol* 2018;2:680–7. https://doi.org/10.1038/s41559-018-0482-x
25. Falcucci G, Amati G, Fanelli P. *et al.* Extreme flow simulations reveal skeletal adaptations of deep-sea sponges. *Nature* 2021;595:537–41. https://doi.org/10.1038/s41586-021-03658-1
26. Han JR, Li S, Li WJ. *et al.* Mining microbial and metabolic dark matter in extreme environments: a roadmap for harnessing the power of multi-omics data. *Adv Biotechnol* 2024;2:26. https://doi.org/10.1007/s44307-024-00034-8
27. Musilova Z, Cortesi F. An (omics) perspective on the evolution of vision in deep-sea fishes reveals exceptional adaptations to life in the extreme. *Funct Ecol* 2025;00:1–10.
28. Levin LA, Bett BJ, Gates AR. *et al.* Global observing needs in the deep ocean. *Front Mar Sci* 2019;6:241. https://doi.org/10.3389/fmars.2019.00241
29. Wei J, Liu P, Liu F. *et al.* EDomics: a comprehensive and comparative multi-omics database for animal evo-devo. *Nucleic Acids Res* 2023;51:D913–23. https://doi.org/10.1093/nar/gkac944
30. Liu FY, Cai BC, Lian SS. *et al.* MolluscDB 2.0: a comprehensive functional and evolutionary genomics database for over 1400 molluscan species. *Nucleic Acids Res* 2024;53:D1075–86. https://doi.org/10.1093/nar/gkae1026
31. Li Y, Tassia MG, Waits DS. *et al.* Genomic adaptations to chemosymbiosis in the deep-sea seep-dwelling tubeworm *Lamellibrachia luymesi*. *BMC Biol* 2019;17:91. https://doi.org/10.1186/s12915-019-0713-x
32. Sun YA, Sun J, Yang Y. *et al.* Genomic signatures supporting the symbiosis and formation of chitinous tube in the deep-sea tubeworm *Paraescarpia echinospica*. *Mol Biol Evol* 2021;38:4116–34. https://doi.org/10.1093/molbev/msab203
33. He X, Wang H, Xu T. *et al.* Genomic analysis of a scale worm provides insights into its adaptation to deep-sea hydrothermal vents. *Genome Biol Evol* 2023;15:evad125. https://doi.org/10.1093/gbe/evad125
34. Moggioli G, Panossian B, Sun YA. *et al.* Distinct genomic routes underlie transitions to specialised symbiotic lifestyles in deep-sea annelid worms. *Nat Commun* 2023;14:2814. https://doi.org/10.1038/s41467-023-38521-6

35. Perez M, Aroh O, Sun YA. *et al.* Third-generation sequencing reveals the adaptive role of the epigenome in three deep-sea polychaetes. *Mol Biol Evol* 2023;40:msad172.
36. Uchida T, Yoshioka Y, Yoshida Y. *et al.* Genomic and transcriptomic analyses illuminate the molecular basis of the unique lifestyle of a tubeworm, *Lamellibrachia satsuma*. *DNA Res* 2023;30:dsad014.
37. Yuan JB, Zhang XJ, Kou Q. *et al.* Genome of a giant isopod, *Bathynomus jamesi*, provides insights into body size evolution and adaptation to deep-sea environment. *BMC Biol* 2022;20:113.
38. Parey E, Louis A, Montfort J. *et al.* Genome structures resolve the early diversification of teleost fishes. *Science* 2023;379:572–5. https://doi.org/10.1126/science.abq4257
39. Xu WJ, Zhu CL, Gao XL. *et al.* Chromosome-level genome assembly of hadal snailfish reveals mechanisms of deep-sea adaptation in vertebrates. *eLife* 2023;12:RP87198. https://doi.org/10.7554/eLife.87198.3
40. Mu YN, Bian C, Liu RY. *et al.* Whole genome sequencing of a snailfish from the Yap Trench (~7,000 m) clarifies the molecular mechanisms underlying adaptation to the deep sea. *PLoS Genet* 2021;17:e1009530. https://doi.org/10.1371/journal.pgen.1009530
41. Zhan ZF, Liang YS, Wei JH. *et al.* Chromosome-level genome assembly and annotation of a sea toad (*Chaunax sp*). *Sci Data* 2024;11:1397.
42. Chen J, Zeng HH, Lv WQ. *et al.* Pseudo-chromosome-length genome assembly for a deep-sea eel *Ilyophis brunneus* sheds light on the deep-sea adaptation. *Sci China Life Sci* 2023;66:1379–91.
43. Law STS, Yu YF, Nong WY. *et al.* The genome of the deep-sea anemone *Actinernus sp.* contains a mega-array of ANTP-class homeobox genes. *Proc Biol Sci* 2023;290:20231563.
44. Li JY, Zhan ZF, Li Y. *et al.* Chromosome-level genome assembly of a deep-sea venus flytrap sea anemone sheds light upon adaptations to an extremely oligotrophic environment. *Mol Ecol* 2024;33:e17504. https://doi.org/10.1111/mec.17504
45. Liu C, Bian C, Gao Q. *et al.* Whole genome sequencing of a novel sea anemone (*Actinostola sp.*) from a deep-sea hydrothermal vent. *Sci Data* 2024;11:102. https://doi.org/10.1038/s41597-024-02944-7
46. Zhou Y, Liu HL, Feng CG. *et al.* Genetic adaptations of sea anemone to hydrothermal environment. *Sci Adv* 2023;9:eadh0474.
47. Herrera S, Cordes EE. Genome assembly of the deep-sea coral *Lophelia pertusa*. *GigaByte* 2023;2023:1–12.
48. Feng CG, Liu RY, Xu WJ. *et al.* The genome of a new anemone species (Actiniaria: hormathiidae) provides insights into deep-sea adaptation. *Deep-Sea Res Pt I* 2021;170:103492.
49. Zhou Y, Feng C, Pu Y. *et al.* The first draft genome of a cold-water coral *Trachythela sp.* (Alcyonacea: stolonifera: clavulariidae). *Genome Biol Evol* 2021;13:evaa265. https://doi.org/10.1093/gbe/evaa265
50. Shao GM, He TL, Mu YN. *et al.* The genome of a hadal sea cucumber reveals novel adaptive strategies to deep-sea environments. *iScience* 2022;25:105545. https://doi.org/10.1016/j.isci.2022.105545
51. Liu J, Zhou Y, Pu YJ. *et al.* A chromosome-level genome assembly of a deep-sea starfish (*Zoroaster cf. ophiactis*). *Sci Data* 2023;10:506. https://doi.org/10.1038/s41597-023-02397-4
52. Pu Y, Zhou Y, Liu J. *et al.* A high-quality chromosomal genome assembly of the sea cucumber *Chiridota heheva* and its hydrothermal adaptation. *GigaScience* 2024;13:giad107. https://doi.org/10.1093/gigascience/giad107
53. Wang Y, Wang MX, Li J. *et al.* A chromosome-level genome assembly of a deep-sea symbiotic Aplacophora mollusc *Chaetoderma sp*. *Sci Data* 2024;11:133. https://doi.org/10.1038/s41597-024-02940-x
54. Song H, Wang YA, Shao HJ. *et al.* Scaphopoda is the sister taxon to Bivalvia: evidence of ancient incomplete lineage sorting. *Proc Natl Acad Sci USA* 2023;120:e2302361120.
55. Lan Y, Sun J, Chen C. *et al.* Hologenome analysis reveals dual symbiosis in the deep-sea hydrothermal vent snail *Gigantopelta aegis*. *Nat Commun* 2021;12:1165.
56. Sun J, Chen C, Miyamoto N. *et al.* The scaly-foot snail genome and implications for the origins of biomineralised armour. *Nat Commun* 2020;11:1657. https://doi.org/10.1038/s41467-020-15522-3
57. Liu RY, Wang K, Liu J. *et al.* *De novo* genome assembly of limpet *Bathyacmaea lactea* (Gastropoda: pectinodontidae): the first reference genome of a deep-sea gastropod endemic to cold seeps. *Genome Biol Evol* 2020;12:905–10. https://doi.org/10.1093/gbe/evaa100
58. da Fonseca RR, Couto A, Machado AM. *et al.* A draft genome sequence of the elusive giant squid, *Architeuthis dux*. *Gigascience* 2020;9:giz152. https://doi.org/10.1093/gigascience/giz152
59. Xie X, Teng WM, Yu ZA. *et al.* Chromosome-level genome assembly of sea scallop *Placopecten magellanicus* provides insights into the genetic characteristics and adaptive evolution of large scallops. *Genomics* 2023;115:110747. https://doi.org/10.1016/j.ygeno.2023.110747
60. Guo Y, Meng LF, Wang MX. *et al.* Hologenome analysis reveals independent evolution to chemosymbiosis by deep-sea bivalves. *BMC Biol* 2023;21:51. https://doi.org/10.1186/s12915-023-01551-z
61. Ip JCH, Xu T, Sun J. *et al.* Host-endosymbiont genome integration in a deep-sea chemosymbiotic clam. *Mol Biol Evol* 2021;38:502–18. https://doi.org/10.1093/molbev/msaa241
62. Steffen K, Proux-Wera E, Soler L. *et al.* Whole genome sequence of the deep-sea sponge *Geodia barretti* (Metazoa, Porifera, Demospongiae). *G3* 2023;13:jkad192. https://doi.org/10.1093/g3journal/jkad192
63. Federhen S. The NCBI Taxonomy database. *Nucleic Acids Research* 2012;40:D136–43. https://doi.org/10.1093/nar/gkr1178
64. Thelwall M, Kousha K. Figshare: a universal repository for academic resource sharing?. *Online Inf Rev* 2016;40:333–46. https://doi.org/10.1108/OIR-06-2015-0190
65. He L, Nahar V. Reuse of scientific data in academic publications: an investigation of dryad digital repository. *ASLIB J Inf Manag* 2016;68:478–94. https://doi.org/10.1108/AJIM-01-2016-0008
66. Armit C, Tuli MA, Hunter C. A decade of GigaScience: gigaDB and the open data movement. *Gigascience* 2022;11:giac053. https://doi.org/10.1093/gigascience/giac053
67. Li C, Hou Y, Li J *et al.* ScienceDB: a public multidisciplinary research data repository for eScience. *In: 2017 IEEE 13th International Conference on e-Science (e-Science)*. 2017, 248–55.
68. Paysan-Lafosse T, Andreeva A, Blum M. *et al.* The Pfam protein families database: embracing AI/ML. *Nucleic Acids Res* 2024;53:D523–34. https://doi.org/10.1093/nar/gkae997
69. Bateman A, Martin MJ, Orchard S. *et al.* UniProt: the universal protein knowledgebase in 2025. *Nucleic Acids Res* 2024;52:D609–17.
70. Thomas PD, Ebert D, Muruganujan A. *et al.* PANTHER: making genome-scale phylogenetics accessible to all. *Protein Sci* 2022;31:8–22. https://doi.org/10.1002/pro.4218
71. Kanehisa M, Goto S. KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res* 2000;28:27–30. https://doi.org/10.1093/nar/28.1.27
72. Shen WK, Chen SY, Gan ZQ. *et al.* AnimalTFDB 4.0: a comprehensive animal transcription factor database updated with variation and expression annotations. *Nucleic Acids Res* 2023;51:D39–45. https://doi.org/10.1093/nar/gkac907
73. Zhou JQ, Xu Y, Lin SF. *et al.* iUUCD 2.0: an update with rich annotations for ubiquitin and ubiquitin-like conjugations. *Nucleic Acids Res* 2018;46:D447–53. https://doi.org/10.1093/nar/gkx1041
74. Tegenfeldt F, Kuznetsov D, Manni M. *et al.* OrthoDB and BUSCO update: annotation of orthologs with wider sampling of

genomes. *Nucleic Acids Res* 2024;53:D516–22. https://doi.org/10.1093/nar/gkae987
75. Shank SD, Weaver S, Pond SL.K. phylotree.js—a JavaScript library for application development and interactive data visualization in phylogenetics. *BMC Bioinf* 2018;19:276. https://doi.org/10.1186/s12859-018-2283-2
76. Yu H, Li Y, Han W. *et al.* Pan-evolutionary and regulatory genome architecture delineated by an integrated macro- and microsynteny approach. *Nat Protoc* 2024;19:1623–78. https://doi.org/10.1038/s41596-024-00966-4
77. Greiner S, Lehwark P, Bock R. OrganellarGenomeDRAW (OGDRAW) version 1.3.1: expanded toolkit for the graphical visualization of organellar genomes. *Nucleic Acids Res* 2019;47:W59–64. https://doi.org/10.1093/nar/gkz238
78. Brown J, Pirrung M, McCue LA. FQC Dashboard: integrates FastQC results into a web-based, interactive, and extensible FASTQ quality control tool. *Bioinformatics* 2017;33:3137–9. https://doi.org/10.1093/bioinformatics/btx373
79. Sewe SO, Silva G, Sicat P. *et al.* Trimming and validation of Illumina short reads using trimmomatic, trinity assembly, and assessment of RNA-seq data. *Methods Mol Biol* 2022;2443:211–32.
80. Thakur V. RNA-seq data analysis for differential gene expression using HISAT2-stringtie-ballgown pipeline. *Methods Mol Biol* 2024;2812:101–13.
81. Liao Y, Smyth GK, Shi W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* 2014;30:923–30. https://doi.org/10.1093/bioinformatics/btt656
82. Pertea M, Pertea GM, Antonescu CM. *et al.* StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nat Biotechnol* 2015;33:290–5. https://doi.org/10.1038/nbt.3122
83. Cao C, Lemaire LA, Wang W. *et al.* Comprehensive single-cell transcriptome lineages of a proto-vertebrate. *Nature* 2019;571:349–54. https://doi.org/10.1038/s41586-019-1385-y
84. Hao Y, Hao S, Andersen-Nissen E. *et al.* Integrated analysis of multimodal single-cell data. *Cell* 2021;184:3573–87. https://doi.org/10.1016/j.cell.2021.04.048
85. Parks DH, Chuvochina M, Rinke C. *et al.* GTDB: an ongoing census of bacterial and archaeal diversity through a phylogenetically consistent, rank normalized and complete genome-based taxonomy. *Nucleic Acids Res* 2022;50:D785–94. https://doi.org/10.1093/nar/gkab776
86. Cantalapiedra CP, Hernández-Plaza A, Letunic I. *et al.* eggNOG-mapper v2: functional annotation, orthology assignments, and domain prediction at the metagenomic scale. *Mol Biol Evol* 2021;38:5825–9. https://doi.org/10.1093/molbev/msab293
87. Zhang LK, Chong YP, Yang XL. *et al.* Comparative transcriptome reveals conserved gene expression in reproductive organs in Solanaceae. *Int J Mol Sci* 2025;26:3568. https://doi.org/10.3390/ijms26083568
88. Yang PT, Yuan Y, Yan C. *et al.* AlliumDB: a central portal for comparative and functional genomics in *Allium*. *Hortic Res* 2024;11:uhad285. https://doi.org/10.1093/hr/uhad285
89. Franz M, Lopes CT, Fong D. *et al.* Cytoscape.js 2023 update: a graph theory library for visualization and analysis. *Bioinformatics* 2023;39:btad031. https://doi.org/10.1093/bioinformatics/btad031
90. Emms DM, Kelly S. OrthoFinder: phylogenetic orthology inference for comparative genomics. *Genome Biol* 2019;20:238. https://doi.org/10.1186/s13059-019-1832-y
91. Katoh K, Misawa K, Kuma K. *et al.* MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res* 2002;30:3059–66. https://doi.org/10.1093/nar/gkf436
92. Criscuolo A, Gribaldo S. BMGE (Block Mapping and Gathering with Entropy): a new software for selection of phylogenetic informative regions from multiple sequence alignments. *BMC Evol Biol* 2010;10:210. https://doi.org/10.1186/1471-2148-10-210
93. Minh BQ, Schmidt HA, Chernomor O. *et al.* IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. *Mol Biol Evol* 2020;37:1530–4. https://doi.org/10.1093/molbev/msaa015
94. Ma X, Yan H, Yang J. *et al.* PlantGSAD: a comprehensive gene set annotation database for plant species. *Nucleic Acids Res* 2022;50:D1456–67. https://doi.org/10.1093/nar/gkab794
95. Ma X, Cheng S, Ding R. *et al.* ipaQTL-atlas: an atlas of intronic polyadenylation quantitative trait loci across human tissues. *Nucleic Acids Res* 2023;51:D1046–52. https://doi.org/10.1093/nar/gkac736
96. Slater GJ, Harmon LJ. Unifying fossils and phylogenies for comparative analyses of diversification and trait evolution. *Methods Ecol Evol* 2013;4:699–702. https://doi.org/10.1111/2041-210X.12091
97. Koch NM, Garwood RJ, Parry LA. Fossils improve phylogenetic analyses of morphological characters. *Proc Biol Sci* 2021;288:20210044.
98. Peters SE, McClennen M. The paleobiology database application programming interface. *Paleobiology* 2016;42:1–7. https://doi.org/10.1017/pab.2015.39
99. Osman EO, Weinnig AM. Microbiomes and obligate symbiosis of deep-sea animals. *Annu Rev Anim Biosci* 2022;10:151–76. https://doi.org/10.1146/annurev-animal-081621-112021
100. Danovaro R, Snelgrove PVR, Tyler P. Challenging the paradigms of deep-sea ecology. *Trends Ecol Evol* 2014;29:465–75. https://doi.org/10.1016/j.tree.2014.06.002
101. Leigh DM, Vandergast AG, Hunter ME. *et al.* Best practices for genetic and genomic data archiving. *Nat Ecol Evol* 2024;8:1224–32. https://doi.org/10.1038/s41559-024-02423-7
102. Grossman RL. Data lakes, clouds, and commons: a review of platforms for analyzing and sharing genomic data. *Trends Genet* 2019;35:223–34. https://doi.org/10.1016/j.tig.2018.12.006
103. Castellana S, Vicario S, Saccone C. Evolutionary patterns of the mitochondrial genome in metazoa: exploring the role of mutation and selection in mitochondrial protein-coding genes. *Genome Biol Evol* 2011;3:1067–79. https://doi.org/10.1093/gbe/evr040
104. Blum M, Chang H Y, Florentino LC *et al.*, InterPro: the protein sequence classification resource in 2025. *Nucleic acids research* 2025;53(D1):D444–56.https://doi.org/10.1093/nar/gkae1082
105. Manni M, Berkeley MR, Seppey M, *et al.* BUSCO Update: novel and streamlined workflows along with broader and deeper phylogenetic coverage for scoring of eukaryotic, prokaryotic, and viral genomes. *Mol Biol Evol* 2021;38:4647–54. https://doi.org/10.1093/molbev/msab199
106. Cheatle Jarvela AM., Hinman VF. Evolution of transcription factor function as a mechanism for changing metazoan developmental gene regulatory networks. *Evodevo* 2015;6:3. https://doi.org/10.1186/2041-9139-6-3
107. Hsia CC, McGinnis W. Evolution of transcription factor function. *Curr Opin Genet Dev* 2003;13:199–206. https://doi.org/10.1016/S0959-437X(03)00017-0
108. Dougherty SE, Maduka AO, Inada T. *et al.* Expanding role of ubiquitin in translational control. *Int J Mol Sci* 2020;21:1151. https://doi.org/10.3390/ijms21031151
109. Kwon YT, Ciechanover A. The ubiquitin code in the ubiquitin-proteasome system and autophagy. *Trends Biochem Sci* 2017;42:873–86. https://doi.org/10.1016/j.tibs.2017.09.002
110. Parks DH, Imelfort M, Skennerton CT, *et al.* CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Res* 2015;25:1043–55. https://doi.org/10.1101/gr.186072.114
111. Drula E Garron M L, Dogan L *et al.*, The carbohydrate-active enzyme database: functions and literature. *Nucleic acids research* 2022;50:D571–7. https://doi.org/10.1093/nar/gkab1045

112. Bump P, Lubeck L. Marine invertebrates one cell at a time: insights from single-cell analysis. *Integr Comp Biol* 2023;63:999–1009. https://doi.org/10.1093/icb/icad034
113. Zhang YJ, Sun J, Chen C. *et al.* Adaptation and evolution of deep-sea scale worms (Annelida: polynoidae): insights from transcriptome comparison with a shallow-water species. *Sci Rep* 2017;7:46205. https://doi.org/10.1038/srep46205
114. Damas J, Corbo M, Kim J. *et al.* Evolution of the ancestral mammalian karyotype and syntenic regions. *Proc Natl Acad Sci USA* 2022;119:e2209139119. https://doi.org/10.1073/pnas.2209139119
115. Wang S, Zhang JB, Jiao WQ. *et al.* Scallop genome provides insights into evolution of bilaterian karyotype and development. *Nat Ecol Evol* 2017;1:0120. https://doi.org/10.1038/s41559-017-0120
116. Wang YP, Tang HB, DeBarry JD. *et al.* MCScanX: a toolkit for detection and evolutionary analysis of gene synteny and collinearity. *Nucleic Acids Research* 2012;40:e49. https://doi.org/10.1093/nar/gkr1293
117. Soderlund C, Bomhoff M, Nelson WM. SyMAP v3.4: a turnkey synteny system with application to plant genomes. *Nucleic Acids Res* 2011;39:e68. https://doi.org/10.1093/nar/gkr123
118. Haas BJ, Delcher AL, Wortman JR. *et al.* DAGchainer: a tool for mining segmental genome duplications and synteny. *Bioinformatics* 2004;20:3643–6. https://doi.org/10.1093/bioinformatics/bth397
119. Osvatic JT, Yuen B, Kunert M. *et al.* Gene loss and symbiont switching during adaptation to the deep sea in a globally distributed symbiosis. *ISME J* 2023;17:453–66. https://doi.org/10.1038/s41396-022-01355-z
120. Hinzke T, Kleiner M, Breusing C. *et al.* Host–microbe interactions in the chemosynthetic *Riftia pachyptila* symbiosis. *mBio* 2019;10:10–1128. https://doi.org/10.1128/mBio.02243-19
121. Lan Y, Sun J, Zhang WP. *et al.* Host–symbiont interactions in deep-sea chemosymbiotic vesicomyid clams: insights from transcriptome sequencing. *Front Mar Sci* 2019;6:680. https://doi.org/10.3389/fmars.2019.00680
122. Yang Y, Sun J, Sun YN. *et al.* Genomic, transcriptomic, and proteomic insights into the symbiosis of deep-sea tubeworm holobionts. *ISME J* 2020;14:135–50. https://doi.org/10.1038/s41396-019-0520-y
123. Nelson DR, Goldstone JV, Stegeman JJ. The cytochrome P450 genesis locus: the origin and evolution of animal cytochrome P450s. *Philos Trans R Soc B* 2013;368:20120474. https://doi.org/10.1098/rstb.2012.0474
124. Porras MAG, Assié A, Tietjen M. *et al.* An intranuclear bacterial parasite of deep-sea mussels expresses apoptosis inhibitors acquired from its host. *Nat Microbiol* 2024;9:2877–91. https://doi.org/10.1038/s41564-024-01808-5
125. Wang H, He K, Zhang H. *et al.* Deciphering deep-sea chemosynthetic symbiosis by single-nucleus RNA-sequencing. *eLife* 2024;12:RP88294. https://doi.org/10.7554/eLife.88294.4
126. Wang H, Xiao HX, Feng BH. *et al.* Single-cell RNA-seq reveals distinct metabolic "microniches" and close host-symbiont interactions in deep-sea chemosynthetic tubeworm. *Sci Adv* 2024;10:eadn3053. https://doi.org/10.1126/sciadv.adn3053
127. Eraslan G, Avsec Z, Gagneur J. *et al.* Deep learning: new computational modelling techniques for genomics. *Nat Rev Genet* 2019;20:389–403. https://doi.org/10.1038/s41576-019-0122-6

Received: August 15, 2025. **Revised:** September 21, 2025. **Accepted:** October 2, 2025