Published online 1 November 2022 *Nucleic Acids Research, 2023, Vol. 51, Database issue* ***D913–D923***
https://doi.org/10.1093/nar/gkac944

EDomics: a comprehensive and comparative multi-omics database for animal evo-devo

Jiankai Wei[1,2,†], Penghui Liu[1,†], Fuyun Liu[1,†], An Jiang[1], Jinghan Qiao[1], Zhongqi Pu[1], Bingrou Wang[1], Jin Zhang[1], Dongning Jia[2], Yuli Li[1,2,*], Shi Wang[1,2,3,*] and Bo Dong[1,2,4,*]

[1]Sars-Fang Centre, MOE Key Laboratory of Marine Genetics and Breeding, College of Marine Life Sciences, Ocean University of China, Qingdao 266003, China, [2]Laboratory for Marine Biology and Biotechnology, Qingdao National Laboratory for Marine Science and Technology, Qingdao 266237, China, [3]Key Laboratory of Tropical Aquatic Germplasm of Hainan Province, Sanya Oceanographic Institution, Ocean University of China, Sanya 572000, China and [4]Institute of Evolution & Marine Biodiversity, Ocean University of China, Qingdao 266003, China

Received August 09, 2022; Revised September 26, 2022; Editorial Decision October 07, 2022; Accepted October 11, 2022

ABSTRACT

Evolutionary developmental biology (evo-devo) has been among the most fascinating interdisciplinary fields for decades, which aims to elucidate the origin and evolution of diverse developmental processes. The rapid accumulation of omics data provides unprecedented opportunities to answer many interesting but unresolved evo-devo questions. However, the access and utilization of these resources are hindered by challenges particularly in non-model animals. Here, we establish a comparative multi-omics database for animal evo-devo (EDomics, http://edomics.qnlm.ac) containing comprehensive genomes, bulk transcriptomes, and single-cell data across 40 representative species, many of which are generally used as model organisms for animal evo-devo study. EDomics provides a systematic view of genomic/transcriptomic information from various aspects, including genome assembly statistics, gene features and families, transcription factors, transposable elements, and gene expressional profiles/networks. It also exhibits spatiotemporal gene expression profiles at a single-cell level, such as cell atlas, cell markers, and spatial-map information. Moreover, EDomics provides highly valuable, customized datasets/resources for evo-devo research, including gene family expansion/contraction, inferred core gene repertoires, macrosynteny analysis for karyotype evolution, and cell type evolution analysis. EDomics presents a comprehensive and comparative multi-omics platform for animal evo-devo community to decipher the whole history of developmental evolution across the tree of life.

INTRODUCTION

Evolutionary developmental biology (evo-devo) is a rapidly emerging discipline that compares the developmental processes of different organisms to understand how such processes evolved (1). This discipline aims to address fundamental questions that are unresolvable either by traditional evolutionary biology or developmental biology alone. Extensive studies that used well-established model organisms (e.g. *Drosophila melanogaster*, *Caenorhabditis elegans*, *Danio rerio* and *Mus musculus*) have obtained novelties over the past decades (2). However, the limited selection of traditional model organisms indicates a large bias in the role of development in evolution (3,4). In filling this great knowledge gap, the use of emerging new model organisms with key phylogeny positions and full-spectrum coverage shows great application potential by revolutionizing the evo-devo field to depict the whole history of developmental evolution across the tree of life (5).

The evo-devo field has been revolutionized by high-throughput sequencing and other advanced technologies (6). Genomics, transcriptomics, and single-cell technologies have sped up the development of many traditional non-model organisms into new model organisms (e.g. ctenophore *Mnemiopsis leidyi*, placozoan *Trichoplax adhaerens*, ascidian *Ciona robusta* [*Ciona intestinalis* type A] and mollusc *Patinopecten yessoensis*) (7,8). Recently, the use of non-model organisms in evo-devo studies has gener-

*To whom correspondence should be addressed. Tel: +86 532 85906578; Fax: +86 532 85906578; Email: bodong@ouc.edu.cn
Correspondence may also be addressed to Shi Wang. Tel: +86 532 85906589; Fax: +86 532 85906589; Email: swang@ouc.edu.cn
Correspondence may also be addressed to Yuli Li. Tel: +86 532 85906587; Fax: +86 532 85906587; Email: liyuli@ouc.edu.cn
†The authors wish it to be known that, in their opinion, the first three authors should be regarded as Joint First Authors.

ated many breakthrough findings or new theories, such as Hox subcluster-mediated body plan diversification (9,10), the convergent evolution of bilaterian nerve cords (11), the origin of animal multicellularity with pluripotency (12), the single intercalation origin of metazoan larvae (13), and the evolution of ancient homomorphic sex chromosomes (14). Although large amounts of multi-omics resources have been accumulated, which are rapidly increasing in diverse animal groups, integrative access and utilization of these scatteredly deposited genomic resources pose a great challenge for the animal evo-devo research community.

To date, genomic databases have been established only for specific animal groups (e.g. ANISEED for tunicates (15), Echinobase for echinoderm (16), and MolluscDB for Mollusca (17)), with huge scarcity for many other animal groups. Moreover, even existing databases cannot acquire comparative information among animal groups. Therefore, establishing an animal evo-devo multi-omics platform or database by integrating extensive genomic and developmental transcriptomic resources from diverse representative animal groups and developing convenient tools for comprehensive analysis of these data is necessary; however, such platform or database has not been established. In this study, we construct the first comparative multi-omics database for animal evo-devo study (namely, EDomics, http://edomics.qnlm.ac) by integrating several current genomic resources from traditional and emerging model organisms.

ANIMAL PHYLOGENY-WIDE REPRESENTATION OF EDOMICS

EDomics aims to understand the evolutionary and developmental processes across the animal kingdom; thus, animal phylogeny-wide selection of representative species (of evo-devo importance) is essential. Apart from several traditional model organisms, the currently increasing genomic data make it possible to extend evo-devo studies to many non-model organisms. Therefore, existing model animals, such as zebrafish, fruit fly, nematode worm, and mouse, and emerging nonstandard model organisms, including axolotl, lungfish, lamprey, amphioxus, ascidian, scallop, coral, and sponge, are included in EDomics (Table 1). A total of 40 species are selected, and they are distributed among 21 phyla belonging to non-bilaterians, Xenacoelomorpha, Protostomia, and Deuterostomia (Figure 1). The animal species that are selected involve the major developmental and evolutionary events, including the presence of different germ layers, formation of radial and bilateral symmetry, and protostome-to-deuterostome, water-to-landing and invertebrate-to-vertebrate evolutionary transition (18). These representative species are also displayed under the tree of life from evogeneao (https://www.evogeneao.com) at the homepage of EDomics. Thus, evo-devo studies integrating model and non-model organisms will provide comprehensive understanding of genetic and developmetnal regulation mechanisms across the animal kingdom.

DATABASE COMPOSITION AND ARCHITECTURE

EDomics represents the comprehensive and integrative multi-omics resources for representative species of diverse groups spanning the whole animal kingdom, including 40 high-quality assembled genomes, 33 mitogenomes, and 1010 bulk transcriptomes derived from major animal phyla (21 in total, Figure 1). EDomics also collects 77 single-cell transcriptomes covering 478 cell types from 12 well-studied species and 28 spatial transcriptome slices (Figure 1). Furthermore, expanded/contracted gene families, core/dispensable gene families for 67 animal groups, genome macrosynteny for 40 species to three ancestral linkage group (ALG) referenced genomes, and developmental correlations for 325 metazoan species pairs are acquired on the basis of cross-species analysis. Then, these datasets are integrated into six sections, including taxonomy, genome, transcriptome, single-cell, spatial-map, and evo-devo in EDomics for searching and inquiry.

EDomics provides a systematic view of genomic information from various aspects, including genome assembly statistics, gene sequence, structure and functional annotations, gene families, transcription factors (TFs), transposable elements (TEs) and gene expressional profiles/networks derived from bulk transcriptomes. It also exhibits spatiotemporal gene expression profiles at single-cell levels, such as cell atlas, cell markers, gene expression, and gene spatial-map information. Convenient visualization of genomic information is integrated into a customized genome browser. Moreover, EDomics provides highly valuable, featured customized datasets or resources particularly for evo-devo research, including gene family expansion/contraction among metazoans, inferred core gene repertoires for metazoans and their decedent ancestors, genome-by-genome macrosynteny analysis for karyotype evolution, origin and diversification of cell type during animal evolution, and developmental correlation between any two metazoan species. EDomics is implemented on the basis of the Linux operating system, using J2EE as the framework, MySQL as the back-end database, and Apache Tomcat as the server. Web user interfaces are developed on the basis of JavaServer pages (JSP), HTML5 and CSS3.

GENOMIC MODULE

A genome is a complete set of DNA sequence of an organism, including its genes and hierarchically functional and structural configuration. A genomic study aims to collectively characterize and quantify interrelations and influence of genes on the organism (19). Advances in genomics have led to the development of discovery-based research and systems biology to facilitate understanding of complex biological systems and evolutionary innovations and diversifications (20,21). At the genomic level, 40 high-quality genomes with well-annotated gene information (e.g. gene structure and function) are presented in EDomics. The genomic characterizations are displayed in the drop-down list of genomic modules, including JBrowse, genomic features, gene search, gene annotation, mitogenomic data, TEs, TFs and gene family clustering.

The genomic features module shows a brief summary of 40 genome assemblies in the database and provides quick links to relevant literature. The gene search module integrates basic gene information from multiple aspects

Table 1. Forty representative species in EDomics and their evo-devo characteristics

Species	Common names	Phylum	Genome size* (Mb)	Gene number*	Scaffold N50 (kb)*	Omics resources**	Evo-devo characteristics
Amphimedon queenslandica	Sponge	Porifera	167	20 776	120	Genome (2); bulk RNA-seq (28); scRNA-seq (2)	Basal metazoan with indirect development, simplest body plan that lacks nerve, muscle, and gut; origin of multicellular metazoan model
Mnemiopsis leidyi	Comb jelly/sea walnut	Ctenophora	156	16 548	187	Genome (2); bulk RNA-seq (56); scRNA-seq (1)	Basal metazoan with direct development; regeneration, axial patterning and bioluminescence model
Trichoplax adhaerens	Placozoan	Placozoa	106	11 520	5979	Genome (2); scRNA-seq (1)	Basal metazoan with compact genome, simplest body plan that lacks organs or internal structures; origin of cellular complexity model
Clytia hemisphaerica	Hydrozoan jellyfish	Cnidaria	421	19 132	1183	Genome (1); bulk RNA-seq (22)	Diploblastic animal with radial symmetry, triphasic life cycle with a medusae stage; developmental and cell biology jellyfish model
Hydra vulgaris	Swiftwater hydra	Cnidaria	852	20 045	96	Genome (2); bulk RNA-seq (26); scRNA-seq (1)	Diploblastic animal with radial symmetry, immortal life and the simplest nervous system; regeneration and stem cell biology cnidarian model
Nematostella vectensis	Starlet sea anemone	Cnidaria	357	24 773	473	Genome (1); bulk RNA-seq (26)	Diploblastic animal with radial symmetry and oral–aboral axis; comparative genomics and developmental biology cnidarian model
Stylophora pistillata	Stony coral	Cnidaria	400	24 833	457	Genome (2); scRNA-seq (3)	Diploblastic animal with radial symmetry; biomineralization and cnidarian-dinoflagellate symbiosis model
Hofstenia miamia	Acoel worm	Acoelomorpha	950	22 454	1045	Genome (1); bulk RNA-seq (24)	Possibly basal triploblastic bilaterian, whole-body regeneration; regeneration and evo-devo model
Lingula anatina	Brachiopod	Brachiopoda	425	34 105	294	Genome (2); bulk RNA-seq (17)	Lophotrochozoan, calcium phosphate shell, radial cleavage and enterocoelic coelom formation that resemble basal deuterostomes
Phoronis australis	Horseshoe worm	Phoronida	498	20 473	655	Genome (1); bulk RNA-seq (5)	Lophotrochozoan, spiral cleavage, body plan with lophophores and U-shaped gut, deuterostomic characteristics of nervous system
Bugula neritina	Brown bryozoan	Bryozoa	215	25 318	94	Genome (2); bulk RNA-seq (3)	Lophotrochozoan, colonial animal that consists of zooids, biradial cleavage but retains traits of spiral development
Notospermus geniculatus	Ribbon worm	Nemertea	859	43 294	239	Genome (2); bulk RNA-seq (13)	Lophotrochozoan, spiral cleavage, unsegmented worm with a broad diversity of larval forms and life cycles
Patinopecten yessoensis	Yesso scallop	Mollusca	1001	24 738	827	Genome (2); bulk RNA-seq (41)	Lophotrochozoan, spiral cleavage, slow-evolving genome with ancient karyotype, subcluster temporal co-linearity expression of *Hox* genes, evolution of noncephalic eyes
Nautilus pompilius	Chambered nautilus	Mollusca	731	17 710	1096	Genome (2); bulk RNA-seq (5)	Lophotrochozoan, cephalopod with minimalist genome, ancestral features of an external chambered shell and pinhole eye
Octopus bimaculoides	Two-spot octopus	Mollusca	2338	33 609	475	Genome (2); bulk RNA-seq (12)	Lophotrochozoan, highly derived body plan with sophisticated nervous system, innovation of camera-like eyes, prehensile arms, and adaptive coloration
Capitella teleta	Polychaete worm	Annelida	333	32 175	188	Genome (1)	Lophotrochozoan, segmented body plan with typical spiralian development, spiralian model for development and regeneration
Intoshia linei	Orthonectid	Orthonectida	42	8724	26	Genome (2)	Lophotrochozoan, highly simplified spiralian with the smallest genome among metazoans, the simplest model for the development of core bilaterian features
Schmidtea mediterranea	Freshwater planarian	Platyhelminthes	787	23 657	80	Genome (2); bulk RNA-seq (30); scRNA-seq (1)	Lophotrochozoan, simple body plan with abundant adult pluripotent stem cells, animal model for stem cell pluripotency and regeneration
Brachionus plicatilis	Wheel animalcule	Rotifera	109	52 286	20	Genome (2)	Lophotrochozoan, parthenogenesis and developmental dormancy, frequent horizontal gene transfer
Priapulus caudatus	Cactus worm	Priapulida	512	15 088	210	Genome (2)	Basal ecdysozoan with ancestral characters of holoblastic radial cleavage and deuterostomic development
Caenorhabditis elegans	Roundworm	Nematoda	100	20 366	17 494	Genome (2); bulk RNA-seq (20)	Ecdysozoan, simplistic body plan and mosaic development, classical model for animal development, aging and behavior

Table 1. Continued

Species	Common names	Phylum	Genome size* (Mb)	Gene number*	Scaffold N50 (kb)*	Omics resources**	Evo-devo characteristics
Loa loa	Eye worm	Nematoda	91	14 908	174	Genome (2)	Ecdysozoan, complex parasitic development and life cycle; evolution of parasite biology and parasitism
Hypsibius dujardini	Water bear	Tardigrada	104	19 939	342	Genome (2); bulk RNA-seq (35)	Ecdysozoan with rapid life cycle and compact genome; cryptobiosis and resistance of extreme temperatures and pressures
Drosophila melanogaster	Fruit fly	Arthropoda	144	13 948	25 287	Genome (2); bulk RNA-seq (51)	Short life cycle, rich genetic and genome resources; classical genetics and development biology model
Strongylocentrotus purpuratus	Sea urchin	Echinodermata	991	27 746	420	Genome (2); bulk RNA-seq (22); scRNA-seq (8)	Deuterostome with radial symmetry body plan; GRN, embryogenesis, and cell cycle study model
Apostichopus japonicus	Sea cucumber	Echinodermata	983	29 445	191	Genome (2); bulk RNA-seq (35)	Deuterostome, saponin biosynthesis, aestivation and visceral regeneration capacity;
Anneissia japonica	Feather star	Echinodermata	590	21 084	623	Genome (1); bulk RNA-seq (27)	Deuterostome with pentameral symmetry body plan, planktonic larva, with limb regeneration capacity; fertilization and regeneration model
Saccoglossus kowalevskii	Acorn worm	Hemichordata	776	20 943	246	Genome (2); bulk RNA-seq (20)	Hemichordate, bilateral symmetry and gill slits, basal deuterostome, with notochord-like structure; evo-devo model
Branchiostoma floridae	Florida lancelet	Chordata	513	26 676	25 441	Genome (2); bulk RNA-seq (38)	Cephalochordate, basal chordate; origin of vertebrate organs study model
Oikopleura dioica	Pelagic tunicate	Chordata	70	17 152	395	Genome (1)	Small pelagic urochordate, basal branch of urochordates, with notochord throughout life, with a miniature genome, live in the cellulose 'house'
Styela clava	Leathery sea squirt	Chordata	340	18 685	20 775	Genome (2); bulk RNA-seq (22)	Urochordate, retrogressive metamorphosis; cell lineage, invasive, and fouling model
Ciona robusta (*Ciona intestinalis* type A)	Ascidian	Chordata	115	16 658	5153	Genome (2); bulk RNA-seq (11); scRNA-seq (10)	Urochordate, mosaic gene expression pattern; cell lineage tracing, embryogenesis, origin of neural crest model
Petromyzon marinus	Sea lamprey	Chordata	1089	17 567	12 998	Genome (2); bulk RNA-seq (11)	Basal vertebrate, jawless fish, origin of vertebrate feature organs and systems model
Danio rerio	Zebrafish	Chordata	1372	25 729	53 345	Genome (2); bulk RNA-seq (47); scRNA-seq (7); spatial RNA-seq (1)	Rich genetic resources, molecular mechanism study of embryogenesis and organogenesis, fish specific genome duplication; origin of vertebrate and human disease model
Protopterus annectens	West African lungfish	Chordata	40 054	22 026	1 974 114	Genome (2); bulk RNA-seq (27)	Osteichthyes, origin of lung and limbs, large genome size; evolutionary transition from water-to-land study model
Xenopus laevis	African clawed frog	Chordata	2765	45 099	136 613	Genome (2); bulk RNA-seq (31)	Amphibian, mosaic expression, origin of placode, genome duplication; embryonic induction, transplantation, and sex-determination model
Ambystoma mexicanum	Mexican axolotl	Chordata	28 207	46 583	1 205 707	Genome (2); bulk RNA-seq (67)	Amphibian, remarkable regeneration ability, large genome size; regeneration vertebrate model
Gallus gallus	Red junglefowl	Chordata	1065	16 766	91 315	Genome (2); bulk RNA-seq (64); scRNA-seq (3); spatial RNA-seq (4)	Aves, ovipara with cleidoic egg; embryonic development and flight evolution model
Mus musculus	House mouse	Chordata	2728	22 173	130 531	Genome (2); bulk RNA-seq (48); scRNA-seq (8); spatial RNA-seq (14)	Mammal, rich genetic resources, mammal reproduction and breeding, and human disease model
Homo sapiens	Human	Chordata	3272	22 360	145 139	Genome (2); bulk RNA-seq (114); scRNA-seq (20); spatial RNA-seq (9)	Primate, origin of intelligent and consciousness, evolution of memory, mind and emotion

*Genome size refers to the assembled size. Protein-coding gene number and scaffoldN50 are calculated on the basis of the downloaded sequence and annotation files.
**Genome accounts for both nuclear genome and mitochondrial genome.

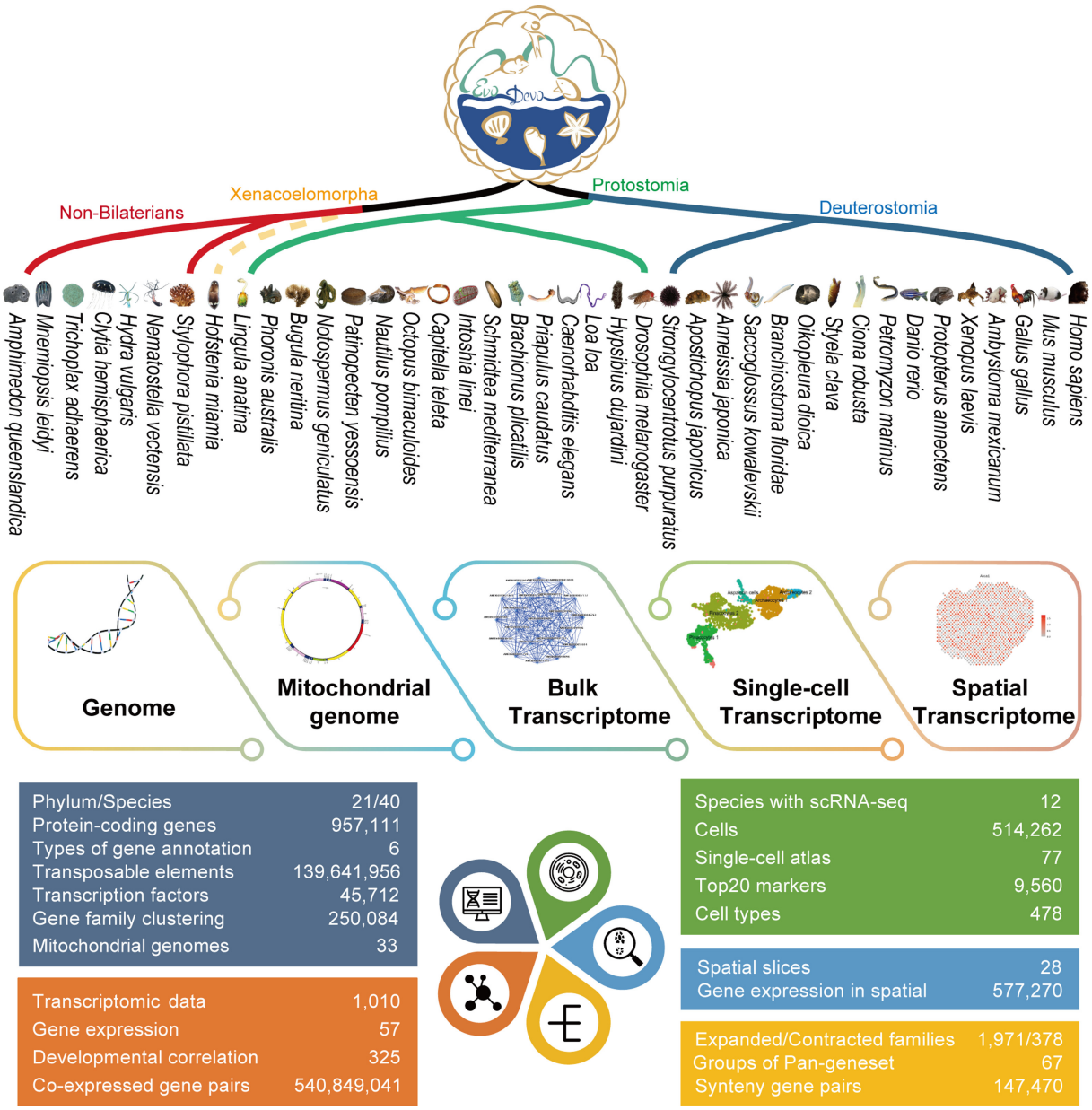

Figure 1. Overview of EDomics and summary of data composition. The 40 species in EDomics are divided into non-bilaterians, Xenacoelomorpha (represented by dashed line due to its debated phylogeny affinity), Protostomia, and Deuterostomia. The genome, mitochondrial genome, bulk transcriptome, single-cell transcriptome, and spatial transcriptome data for each species are collected and deposited in the database. Statistical data are also listed in EDomics.

for all protein-coding genes (Figure 2A). Three types of keywords, including genomic region, gene ID (identity), and gene name, can be selected to obtain relevant information such as gene location, gene size, gene structure, genomic/CDS/protein sequences, functional annotations, homologous genes, and single-cell expression. The gene annotation module (Figure 2A) integrates functional annotation information from six types of public databases, including SWISSPROT (22), KEGG (23), GO (24,25), PFAM (26), IPRSCAN (27) and PANTHER (28). A total of 791 775 genes were annotated with at least one type of annotation. After selecting the target species and searching the database, the user can enter the gene ID to view related annotation information. In the mitogenomic data module, the mitochondrial genome and annotations are presented. For each species, a circos graph showing mitochondrial gene information and an associated table with detailed genomic positions are presented in this module.

TEs are major components of eukaryotic genomes, which have significant effects on genomic function and evolution (29). In the TE module, TEs are correlated with genes based on genomic position information for further gene analysis, which shows that 139 641 956 TEs are associated with 1 058 064 genes. Three alternative search methods, such as a certain genomic interval, the transposon subfamily type, and gene ID, are provided to obtain TE information, including

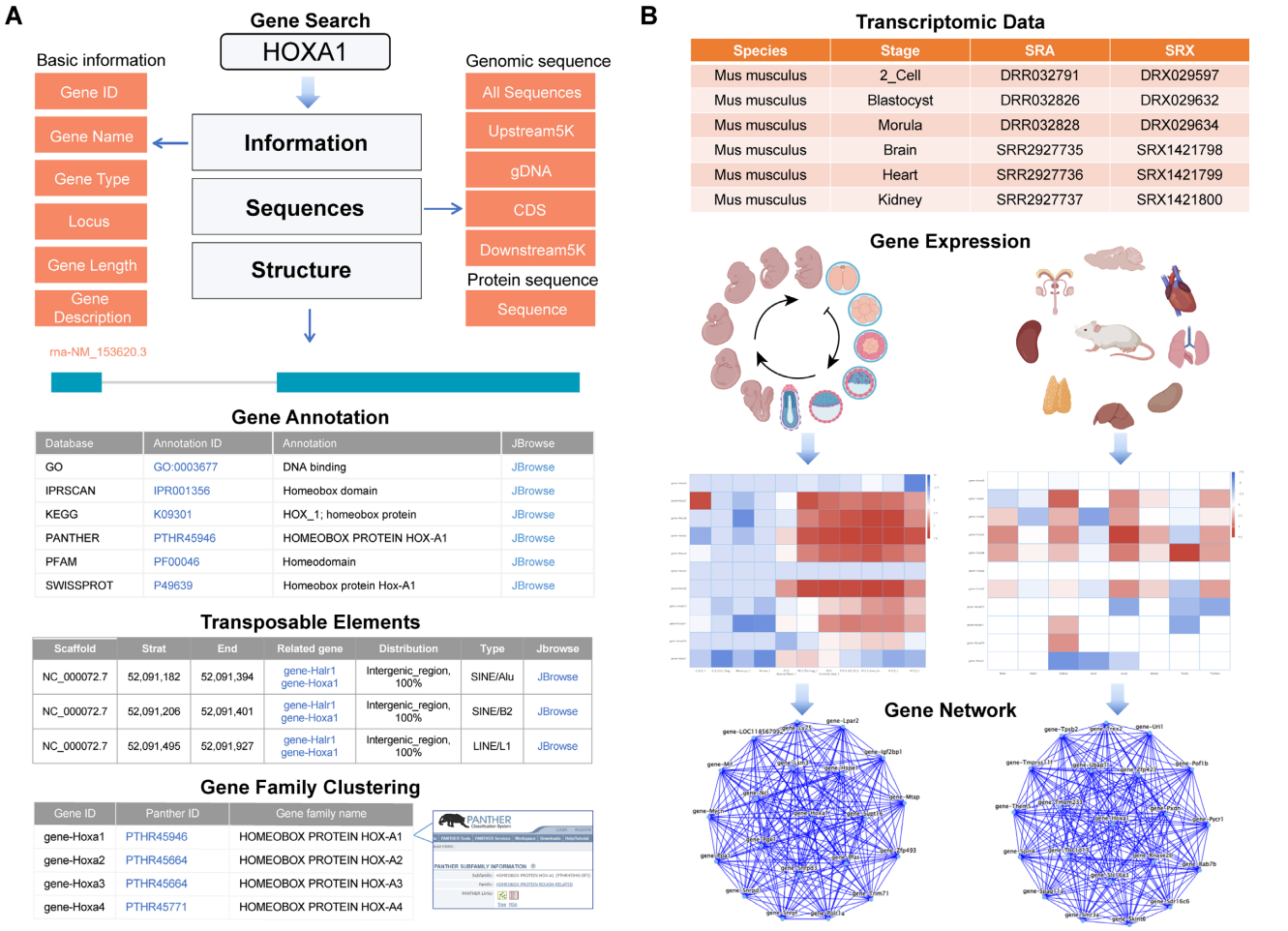

Figure 2. Information of the genomic and transcriptomic module in EDomics. (**A**) Interested genes can be searched through the gene search module to acquire their sequence and structure in the basic gene information module, gene functional annotation information in the gene annotation module, transposon annotation information in the transposon element module, and homologous gene information in the gene family clustering module. (**B**) Bulk transcriptomes of various developmental stages and multiple tissues/organs of 26 species are collected and calculated. Gene expression profiles and co-expression networks of interested genes can be acquired through the gene expression and gene network modules, respectively.

their genomic location, type, related genes and their distribution in genome (Figure 2A). TFs, which serve as master regulators and selector gene, exert control over biological processes that regulate growth, development, and response to the external environment (30,31). In the TF module, TF families of all species in the database are characterized, and a total of 45 712 TF genes are obtained from the 40 metazoan genomes, which are classified into 72 gene families. Users can select a specific TF family name to search the TF family of a species, class, or even all 40 species to obtain gene numbers and gene IDs of TFs. Therefore, the identification and comparison of gene families are critical to understand animal evolution and adaptation (32). In the gene family clustering module, the gene families of 40 species are clustered and annotated, resulting in 250 084 gene clusters containing 916 651 genes. The number of genes in each gene family cluster is displayed. Users can click on the number of gene family clusters to obtain genes with annotation information in each species and gene IDs of interested gene families by clicking the download button (Figure 2A). In addition, clicking on the panther ID will direct to the PANTHER database for the comparison and analysis of gene families with other species. The JBrowse (33,34) module provides a dynamic web platform for genomic visualization and analysis.

TRANSCRIPTOMIC MODULE

The transcriptome is the complete set of transcripts in a cell and their quantity at a specific developmental stage or physiological condition. Understanding the transcriptome is essential to identify the molecular constituents of cells and tissues and investigate gene regulation relationships during development, growth, or disease (35). At the transcriptomic level, 1010 reference genome-profiled transcriptomes derived from the 40 species are selected for further expression and network analysis in EDomics. The datasets are selected on the basis of different developmental stages and different tissues/organs. The transcriptomic features are displayed in the drop-down list of the transcriptomic module, including transcriptomic data, gene expression, and gene network.

The transcriptomic data module (Figure 2B) provides the statistics of reference genome-profiled transcriptomes in

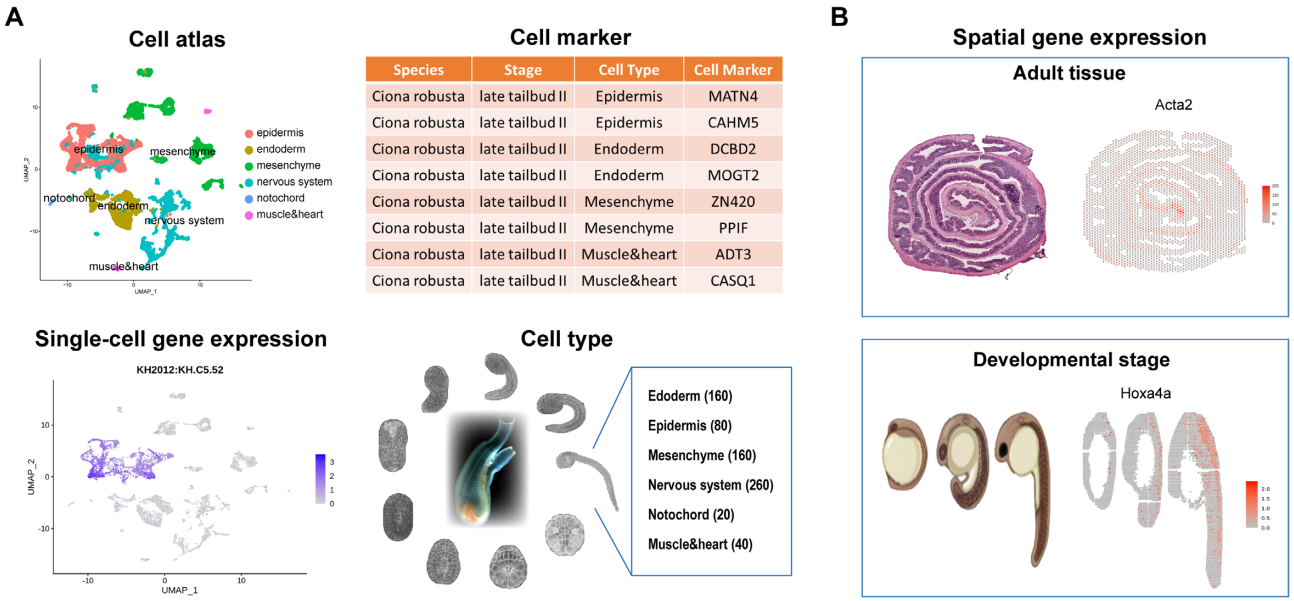

Figure 3. Information of the single-cell and spatial-map module in EDomics. (**A**) The cell atlas is displayed by UMAP plot with annotation of different cell types. The cell marker contains the information of marker genes, including species, tissue or developmental stages, and cell type. The single-cell gene expression provides gene expression heatmap visualized in UMAP plots. The cell type includes the cell type and cell number for each sample. (**B**) The spatial gene expression provides the staining and expression heatmap for each tissue section or developmental stage obtained by spatial transcriptomics.

the database, including the information of sequence read archive (SRA) numbers and links, as well as the origins of adult tissues/organs. We retrieved 1010 bulk RNA-seq datasets belonging to 33 metazoan species from the NCBI SRA database to calculate gene expression profiles in the gene expression module. This module (Figure 2B) provides the expression profile of the query gene list under different developmental stages or adult tissues/organs of the specified species. Three types of keywords, including genomic region, gene ID, and gene name, can be selected to obtain information of the gene expression level. Users can click the download option to obtain the sequence and structure information of the searched genes. Co-expressed genes, reflecting potential relationships in the expression regulatory network and important gene interaction, can be displayed in the format of the gene co-expression network based on the similarity of gene expression patterns (36). In the gene network module, gene co-expression network is constructed using R package of weighted correlation network analysis (37). We obtained 540 849 041 highly correlated co-expressed gene pairs. The target gene with the highest correlation value of the query gene can be visualized by entering the gene ID and setting the number of most correlated genes. Corresponding co-expression networks can be viewed by clicking any co-expressed genes in the network. In addition, a summary table of all co-expressed genes and their functional annotations is provided below the network. Links to basic information of genes are created for each target gene in the summary table in EDomics.

SINGLE-CELL AND SPATIAL-MAP MODULES

Single-cell sequencing examines the sequence information from individual cells using optimized next-generation sequencing, providing a high resolution of cellular differences and a comprehensive understanding of the function of an individual cell with regard to its micro-environment (38). Spatial transcriptomics with single-cell resolution is another research hotspot inherited from single-cell analysis, which allows visualizable and quantitative analysis of the transcriptome with spatial resolution in individual tissue section (39). At the transcriptomic level, 77 single-cell transcriptomic samples covering 478 cell types from 12 well-studied species and 28 spatial transcriptomic samples are listed. The single-cell RNA sequencing features are displayed in the drop-down list of the single-cell module, including single-cell datasets, cell atlas, cell marker, and single-cell gene expression. The spatial expression features are displayed in the spatial-map module.

The single-cell dataset module shows a brief summary of 12 single-cell RNA sequencing projects in the database and provides links to relevant literature. The cell atlas module is constructed on the basis of single-cell transcriptomes in the database, including information on cell type composition and marker genes for each cell cluster. Different cell types are labeled with different colors (Figure 3A). Users can click the marker list and select the clusters, and the marker genes will then be displayed. The cell marker module integrates basic gene information from multiple aspects of marker genes, including species name, tissue or developmental stages, and cell type for each sample (Figure 3A). Top 10 markers of each sample and the marker gene list can be viewed. The single-cell gene expression module (Figure 3A) provides gene expression information of all protein-coding genes at a single-cell level. Gene expression in each sample can be visualized in uniform manifold approximation and projection plots, and the corresponding cell types are also listed.

The spatial-map module (Figure 3B) provides information about visualization and analysis of gene expression in

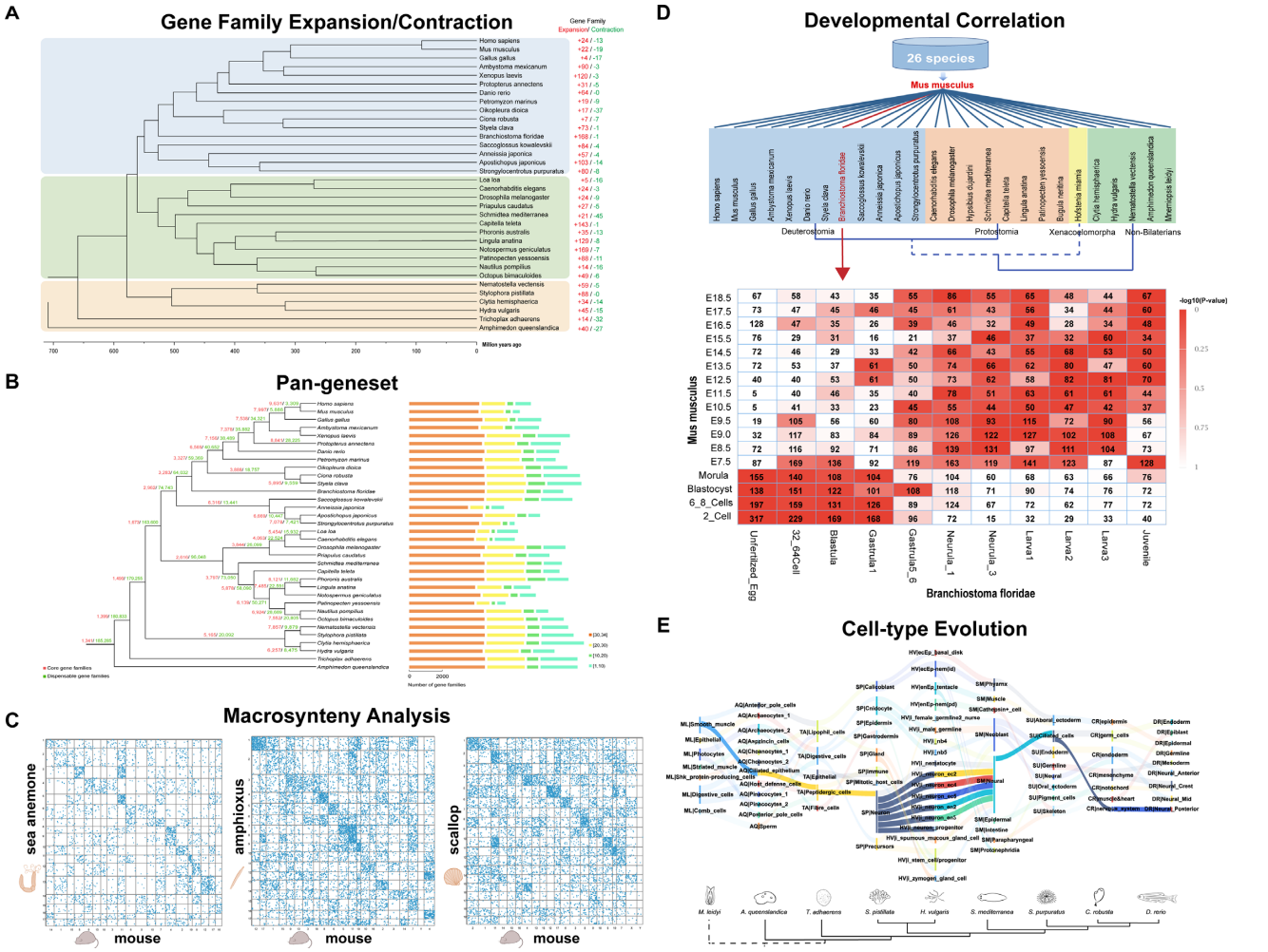

Figure 4. Evo-devo analysis in EDomics. (**A, B**) Gene family evolution for extant species and core/dispensable gene sets for metazoan and their decedent ancestors are shown in the gene family expansion/contraction module and pan-geneset module, respectively. (**C**) Karyotype evolution of the 40 species is explored by comparing with bilaterian ALGs represented by three ancient animal genomes, which are displayed in the macrosynteny analysis module. (**D**) Developmental similarity between any two species with developmental transcriptomic data is displayed in the developmental correlation module. (**E**) Cross-species evolutionary hierarchy of the major cell types is shown in the cell type evolution module.

tissue sections or at developmental stages obtained by spatial transcriptomics. The color of dots in each sample is labeled in accordance with the expression level. Users can search the gene expression patterns through the gene name or gene ID. A total of four species and 28 spatial slices are included in this module.

EVO-DEVO MODULE

In an evo-devo study, cross-species analysis is an important step to discover the conserved key components and effective factors and unravel complicated evolutionary and developmental processes. Cross-species analysis in EDomics can be conducted at genomic, transcriptomic and single-cell transcriptomic levels. They are displayed in the drop-down list of the evo-devo module, including gene family expansion/contraction, pan-geneset, macrosynteny analysis, developmental correlation, and cell type evolution.

Comparative genomic analysis among the 40 species is conducted, and the results are displayed in the drop-down list of the evo-devo module. For the gene family expansion/contraction module (Figure 4A), a number of gene families undergoing expansion and contraction for each lineage are shown in red and green, respectively. In the pan-geneset module (Figure 4B), core/dispensable gene families at different taxonomy levels are displayed in the phylogenetic tree. For a given taxonomy group, pan-geneset represents a core/dispensable geneset that is highly preserved or variable in that group. For each species, all protein-coding genes are classified on the basis of the conservation levels. Users can click certain bar graphs to download the gene IDs of corresponding gene sets. In the macrosynteny analysis module (Figure 4C), the genomes of 40 species are analyzed by comparing with bilaterian ALGs represented by three ancient animal genomes (scallop, *Patinopecten yessoensis*; lancelet, *Branchiostoma floridae*; and sea anemone, *Nematostella vectensis*). The karyotype conservation level of 40 species based on their comparison with specified ALGs can be systematically viewed. A download option is provided for users to obtain corresponding

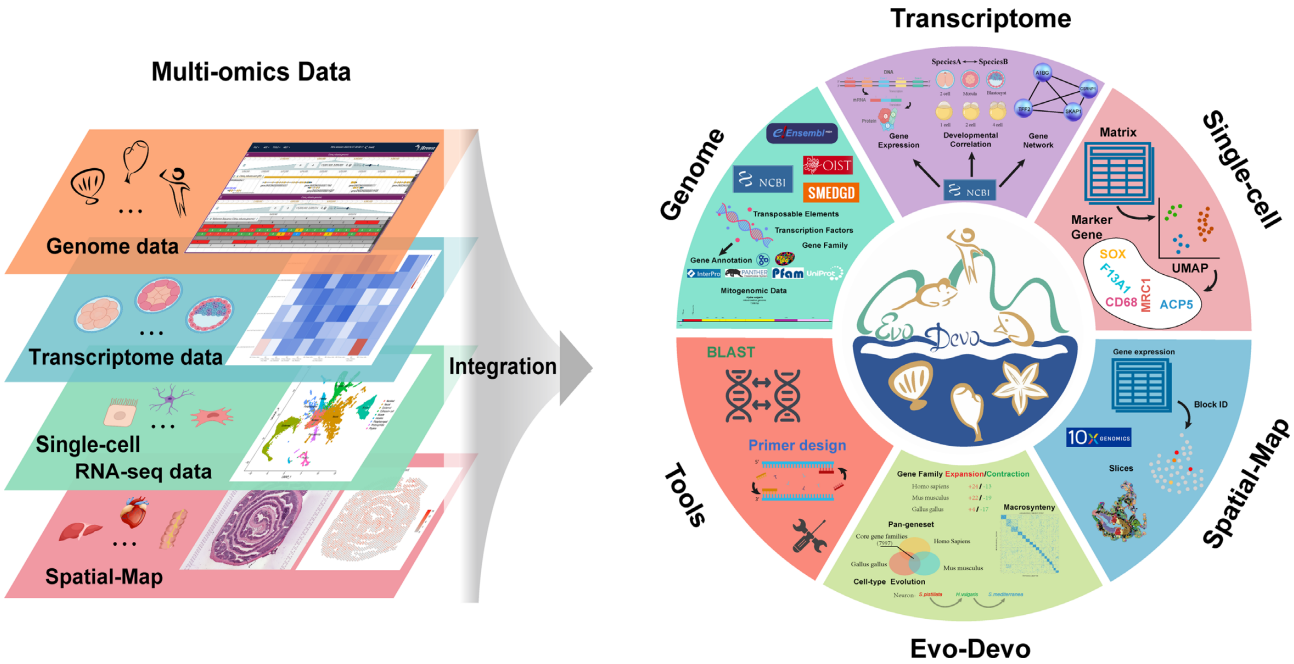

Figure 5. Multi-omics data and cross-species analysis are integrated into six main modules in EDomics. The dataset in EDomics includes not only different levels of omics data (genome, transcriptome, single-cell RNA-seq, and spatial-map), but also the representative species, tissues, developmental stages, and cell types. The integrated data in EDomics are exhibited in six main modules, including genome, transcriptome, single-cell, spatial-map, evo-devo, and tools.

macrosynteny dotplot, homologous gene pairs and related gene sequences.

Comparative transcriptomic analysis is also displayed in the drop-down list of the evo-devo module. The developmental correlation module (Figure 4D) provides analysis of developmental transcriptome correlations between two species. Users can select two species of interest to obtain the correlation among the developmental periods indicated by the number of periods of highly expressed genes homologous to the selected species. The corresponding gene list for every block can be acquired through the download option. Comparative analysis of nine species with single-cell RNA sequencing data is also conducted on the basis of the cell types. The cell type evolution module (Figure 4E) provides the evolutionary hierarchy for the major cell types using MetaNeighbor analysis method (40). The cell types used in this module are divided and annotated according to their original literatures. In the cross-species hierarchy, the links between cell-type pairs are displayed according to their AUROC score (the mean area under the receiver operator characteristic curve), which is acquired from MetaNeighbor. For example, our results reveal the possible evolution roadmaps of neuron cells from comb jelly to zebrafish, and among the six neuronal subtypes in hydra, the ectodermal neuron cells possess the strongest links with neuron cells in coral, which indicates the evolutionary characteristics of neuronal cells (Figure 4E). The heatmap of AUROC scores between adjacent species are also provided for users to quantify the similarity between cell-type pairs.

SUMMARY AND PROSPECTIVE

With the integration of multi-omics data and cross-species analysis, our database enables a systematic evo-devo view of genomic and transcriptomic information across animal kingdom and provides a highly valuable, unique customized platform to illustrate the changes in major development-derived transitions and innovation during organismal evolution (Figure 5). To date, EDomics represents the most comprehensive and integrative multi-omics resources for representative species of diverse groups spanning the whole animal kingdom, including 40 high-quality assembled genomes and 1010 bulk transcriptomes derived from all major animal phyla. It also enables systematic browsing of single-cell RNA-seq data from various aspects (such as cell clusters, cell markers, gene expression, and spatial information). Moreover, key TFs and cellular lineage tracing can be conveniently analyzed and compared across species using the provided tools in our database. EDomics provides not only high-quality multi-omics information for each animal group, but also powerful bioinformatic tools, by which the comparative information could be extracted and analyzed among animal groups (Figure 5).

At present, EDomics is focusing on representative metazoan species, and the species coverage will increase with the progression of multiple large-scale genomic sequencing projects (e.g. EBP (41) and VGP (42)). In the future, we will continuously update EDomics with the emergence of new metazoan genomic and omics data. We will also supply more annotation and functionalities to the database, such as (i) increasing the number of representative species and single-cell/spatial transcriptomic data with the accumulation of newly sequenced animals; (ii) expanding omics resource dimensions (e.g. epigenome, proteome, and metabolome) and integrating multidimensional omics data by deep learning analysis; (iii) systematically collecting and uploading developmental biology imaging data such as traditional embryo development observation, *in situ* hybridization and knockout phenotype of key genes and (iv) adding

online design or analysis tools such as CRISPR gene editing design (43), AlphaFold prediction (44), 3D/4D genomic display (45) and dynamic display of single-cell data.

DATA AVAILABILITY

All data generated or analyzed during this study are included in the manuscript and website.

ACKNOWLEDGEMENTS

We thank lab members in Bo Dong and Shi Wang laboratory for discussion and critical comments on the manuscript; Dr Qi Liu and his team from Wuhan Onemore-tech Co., Ltd for technique support on database construction; Dr Qun Liu and Dr. Guangyi Fan from BGI Research-Qingdao for help on cell type evolution analysis.

FUNDING

National Key Research and Development Program of China [2019YFE0190900, 2022YFC2601302]; Marine S&T Fund of Shandong Province for National Laboratory for Marine Science and Technology (Qingdao) [2022QNLM050101, 2022QNLM030005]; National Natural Science Foundation of China [32130107, 31871499, 32222085]; Key R&D Project of Shandong Province [2021ZLGX03]; Taishan Scholar Program of Shandong Province, China.
Conflict of interest statement. None declared.

REFERENCES

1. Raff,R.A. (2000) Evo-devo: the evolution of a new discipline. *Nat. Rev. Genet.*, **1**, 74–79.
2. Jenner,R.A. and Wills,M.A. (2007) The choice of model organisms in evo-devo. *Nat. Rev. Genet.*, **8**, 311–319.
3. Hall,B.K. (1997) Phylotypic stage or phantom: is there a highly conserved embryonic stage in vertebrates?*Trends Ecol. Evol.*, **12**, 461–463.
4. Stracke,K. and Hejnol,A. (2022) Marine animal evolutionary developmental biology – advances through technology development. *Evol. Appl.*, https://doi.org/10.1111/eva.13456.
5. Zhao,L., Gao,F., Gao,S., Liang,Y., Long,H., Lv,Z., Su,Y., Ye,N., Zhang,L., Zhao,C. *et al.* (2021) Biodiversity-based development and evolution: the emerging research systems in model and non-model organisms. *Sci. China-Life Sci.*, **64**, 1236–1280.
6. Liu,T., Yu,L., Liu,L., Li,H. and Li,Y. (2015) Comparative transcriptomes and evo-devo studies depending on next generation sequencing. *Comput. Math. Method Med.*, **2015**, 896176.
7. Yang,Z., Zhang,L., Hu,J., Wang,J., Bao,Z. and Wang,S. (2020) The evo-devo of molluscs: insights from a genomic perspective. *Evol. Dev.*, **22**, 409–424.
8. Satoh,N. and Levine,M. (2005) Surfing with the tunicates into the post-genome era. *Genes Dev.*, **19**, 2407–2411.
9. Wang,S., Zhang,J., Jiao,W., Li,J., Xun,X., Sun,Y., Guo,X., Huan,P., Dong,B., Zhang,L. *et al.* (2017) Scallop genome provides insights into evolution of bilaterian karyotype and development. *Nat. Ecol. Evol.*, **1**, 0120.
10. Wei,M., Qin,Z., Kong,D., Liu,D., Zheng,Q., Bai,S., Zhang,Z. and Ma,Y. (2022) Echiuran Hox genes provide new insights into the correspondence between Hox subcluster organization and collinearity pattern. *Proc. R. Soc. B*, **289**, 20220705.
11. Martin-Duran,J.M., Pang,K., Borve,A., Le,H.S., Furu,A., Cannon,J.T., Jondelius,U. and Hejnol,A. (2018) Convergent evolution of bilaterian nerve cords. *Nature*, **553**, 45–50.
12. Sogabe,S., Hatleberg,W.L., Kocot,K.M., Say,T.E., Stoupin,D., Roper,K.E., Fernandez-Valverde,S.L., Degnan,S.M. and Degnan,B.M. (2019) Pluripotency and the origin of animal multicellularity. *Nature*, **570**, 519–522.
13. Wang,J., Zhang,L., Lian,S., Qin,Z., Zhu,X., Dai,X., Huang,Z., Ke,C., Zhou,Z., Wei,J. *et al.* (2020) Evolutionary transcriptomics of metazoan biphasic life cycle supports a single intercalation origin of metazoan larvae. *Nat. Ecol. Evol.*, **4**, 725–736.
14. Han,W., Liu,L., Wang,J., Wei,H., Li,Y., Zhang,L., Guo,Z., Li,Y., Liu,T., Zeng,Q. *et al.* (2022) Ancient homomorphy of molluscan sex chromosomes sustained by reversible sex-biased genes and sex determiner translocation. *Nat. Ecol. Evol.*, https://www.nature.com/articles/s41559-022-01898-6.
15. Dardaillon,J., Dauga,D., Simion,P., Faure,E., Onuma,T.A., DeBiasse,M.B., Louis,A., Nitta,K.R., Naville,M., Besnardeau,L. *et al.* (2020) ANISEED 2019: 4D exploration of genetic data for an extended range of tunicates. *Nucleic Acids Res.*, **48**, D668–D675.
16. Arshinoff,B.I., Cary,G.A., Karimi,K., Foley,S., Agalakov,S., Delgado,F., Lotay,V.S., Ku,C.J., Pells,T.J., Beatman,T.R. *et al.* (2022) Echinobase: leveraging an extant model organism database to build a knowledgebase supporting research on the genomics and biology of echinoderms. *Nucleic Acids Res.*, **50**, D970–D979.
17. Liu,F., Li,Y., Yu,H., Zhang,L., Hu,J., Bao,Z. and Wang,S. (2021) MolluscDB: an integrated functional and evolutionary genomics database for the hyper-diverse animal phylum Mollusca. *Nucleic Acids Res.*, **49**, D988–D997.
18. Simakov,O. and Kawashima,T. (2017) Independent evolution of genomic characters during major metazoan transitions. *Dev. Biol.*, **427**, 179–192.
19. Satzinger,H. (2008) Theodor and Marcella Boveri: chromosomes and cytoplasm in heredity and development. *Nat. Rev. Genet.*, **9**, 231–238.
20. Hawkins,R.D., Hon,G.C. and Ren,B. (2010) Next-generation genomics: an integrative approach. *Nat. Rev. Genet.*, **11**, 476–486.
21. Stapley,J., Reger,J., Feulner,P.G.D., Smadja,C., Galindo,J., Ekblom,R., Bennison,C., Ball,A.D., Beckerman,A.P. and Slate,J. (2010) Adaptation genomics: the next generation. *Trends Ecol. Evol.*, **25**, 705–712.
22. Bateman,A., Martin,M.J., Orchard,S., Magrane,M., Agivetova,R., Ahmad,S., Alpi,E., Bowler-Barnett,E.H., Britto,R., Bursteinas,B. *et al.* (2021) UniProt: the universal protein knowledgebase in 2021. *Nucleic Acids Res.*, **49**, D480–D489.
23. Kanehisa,M. and Goto,S. (2000) KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.*, **28**, 27–30.
24. Ashburner,M., Ball,C.A., Blake,J.A., Botstein,D., Butler,H., Cherry,J.M., Davis,A.P., Dolinski,K., Dwight,S.S., Eppig,J.T. *et al.* (2000) Gene ontology: tool for the unification of biology. *Nat. Genet.*, **25**, 25–29.
25. Carbon,S., Douglass,E., Dunn,N., Good,B., Harris,N.L., Lewis,S.E., Mungall,C.J., Basu,S., Chisholm,R.L., Dodson,R.J. *et al.* (2019) The gene ontology resource: 20 years and still GOing strong. *Nucleic Acids Res.*, **47**, D330–D338.
26. Mistry,J., Chuguransky,S., Williams,L., Qureshi,M., Salazar,G.A., Sonnhammer,E.L.L., Tosatto,S.C.E., Paladin,L., Raj,S., Richardson,L.J. *et al.* (2021) Pfam: the protein families database in 2021. *Nucleic Acids Res.*, **49**, D412–D419.
27. Jones,P., Binns,D., Chang,H.Y., Fraser,M., Li,W.Z., McAnulla,C., McWilliam,H., Maslen,J., Mitchell,A., Nuka,G. *et al.* (2014) InterProScan 5: genome-scale protein function classification. *Bioinformatics*, **30**, 1236–1240.
28. Mi,H.Y., Ebert,D., Muruganujan,A., Mills,C., Albou,L.P., Mushayamaha,T. and Thomas,P.D. (2021) PANTHER version 16: a revised family classification, tree-based classification tool, enhancer regions and extensive API. *Nucleic Acids Res.*, **49**, D394–D403.
29. Bourque,G., Burns,K.H., Gehring,M., Gorbunova,V., Seluanov,A., Hammell,M., Imbeault,M., Izsvak,Z., Levin,H.L., Macfarlan,T.S. *et al.* (2018) Ten things you should know about transposable elements. *Genome Biol.*, **19**, 199.
30. Hsia,C.C. and McGinnis,W. (2003) Evolution of transcription factor function. *Curr. Opin. Genet. Dev.*, **13**, 199–206.
31. Lambert,S.A., Jolma,A., Campitelli,L.F., Das,P.K., Yin,Y.M., Albu,M., Chen,X.T., Taipale,J., Hughes,T.R. and Weirauch,M.T. (2018) The human transcription factors. *Cell*, **172**, 650–665.
32. Raghupathy,N. and Durand,D. (2009) Gene cluster statistics with gene families. *Mol. Biol. Evol.*, **26**, 957–968.

33. Skinner,M.E., Uzilov,A.V., Stein,L.D., Mungall,C.J. and Holmes,I.H. (2009) JBrowse: a next-generation genome browser. *Genome Res.*, **19**, 1630–1638.
34. Buels,R., Yao,E., Diesh,C.M., Hayes,R.D., Munoz-Torres,M., Helt,G., Goodstein,D.M., Elsik,C.G., Lewis,S.E., Stein,L. *et al.* (2016) JBrowse: a dynamic web platform for genome visualization and analysis. *Genome Biol.*, **17**, 66.
35. Wang,Z., Gerstein,M. and Snyder,M. (2009) RNA-Seq: a revolutionary tool for transcriptomics. *Nat. Rev. Genet.*, **10**, 57–63.
36. Stuart,J.M., Segal,E., Koller,D. and Kim,S.K. (2003) A gene-coexpression network for global discovery of conserved genetic modules. *Science*, **302**, 249–255.
37. Langfelder,P. and Horvath,S. (2008) WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics*, **9**, 559.
38. Eberwine,J., Sul,J.-Y., Bartfai,T. and Kim,J. (2014) The promise of single-cell sequencing. *Nat. Methods*, **11**, 25–27.
39. Stahl,P.L., Salmen,F., Vickovic,S., Lundmark,A., Navarro,J.F., Magnusson,J., Giacomello,S., Asp,M., Westholm,J.O., Huss,M. *et al.* (2016) Visualization and analysis of gene expression in tissue sections by spatial transcriptomics. *Science*, **353**, 78–82.
40. Crow,M., Paul,A., Ballouz,S., Huang,Z.J. and Gillis,J. (2018) Characterizing the replicability of cell types defined by single cell RNA-sequencing data using MetaNeighbor. *Nat. Commun.*, **9**, 884.
41. Lewin,H.A., Robinson,G.E., Kress,W.J., Baker,W.J., Coddington,J., Crandall,K.A., Durbin,R., Edwards,S.V., Forest,F., Gilbert,M.T.P. *et al.* (2018) Earth biogenome project: sequencing life for the future of life. *Proc. Natl. Acad. Sci. U.S.A.*, **115**, 4325–4333.
42. Rhie,A., McCarthy,S.A., Fedrigo,O., Damas,J., Formenti,G., Koren,S., Uliano-Silva,M., Chow,W., Fungtammasan,A., Kim,J. *et al.* (2021) Towards complete and error-free genome assemblies of all vertebrate species. *Nature*, **592**, 737–746.
43. Chen,Y., Hu,Y., Wang,X., Luo,S., Yang,N., Chen,Y., Li,Z., Zhou,Q. and Li,W. (2022) Synergistic engineering of CRISPR-Cas nucleases enables robust mammalian genome editing. *Innovation*, **3**, 100264.
44. Senior,A.W., Evans,R., Jumper,J., Kirkpatrick,J., Sifre,L., Green,T., Qin,C., Zidek,A., Nelson,A.W.R., Bridgland,A. *et al.* (2020) Improved protein structure prediction using potentials from deep learning. *Nature*, **577**, 706–710.
45. Tang,B., X.,Li., G.,Li., D.,Tian., F.,Li. and Z.,Zhang. (2021) Delta.AR: an augmented reality-based visualization platform for 3D genome. *Innovation*, **2**, 100149.