Chou *et al. BMC Genomics* (2018) 19:618
https://doi.org/10.1186/s12864-018-4987-0

BMC Genomics

DATABASE Open Access

PdumBase: a transcriptome database and research tool for Platynereis dumerilii and early development of other metazoans

Hsien-Chao Chou[1,2†], Natalia Acevedo-Luna[1†], Julie A. Kuhlman[1] and Stephan Q. Schneider[1*]

Abstract

Background: The marine polychaete annelid *Platynereis dumerilii* has recently emerged as a prominent organism for the study of development, evolution, stem cells, regeneration, marine ecology, chronobiology and neurobiology within metazoans. Its phylogenetic position within the spiralian/ lophotrochozoan clade, the comparatively high conservation of ancestral features in the *Platynereis* genome, and experimental access to any stage within its life cycle, make *Platynereis* an important model for elucidating the complex regulatory and functional molecular mechanisms governing early development, later organogenesis, and various features of its larval and adult life. High resolution RNA-seq gene expression data obtained from specific developmental stages can be used to dissect early developmental mechanisms. However, the potential for discovery of these mechanisms relies on tools to search, retrieve, and compare genome-wide information within *Platynereis*, and across other metazoan taxa.

Results: To facilitate exploration and discovery by the broader scientific community, we have developed a web-based, searchable online research tool, PdumBase, featuring the first comprehensive transcriptome database for *Platynereis dumerilii* during early stages of development (2 h ~ 14 h). Our database also includes additional stages over the *P. dumerilii* life cycle and provides access to the expression data of 17,213 genes (31,806 transcripts) along with annotation information sourced from Swiss-Prot, Gene Ontology, KEGG pathways, Pfam domains, TmHMM, SingleP, and EggNOG orthology. Expression data for each gene includes the stage, the normalized FPKM, the raw read counts, and information that can be leveraged for statistical analyses of differential gene expression and the construction of genome-wide co-expression networks. In addition, PdumBase offers early stage transcriptome expression data from five further species as a valuable resource for investigators interested in comparing early development in different organisms. To understand conservation of *Platynereis* gene models and to validate gene annotation, most *Platynereis* gene models include a comprehensive phylogenetic analysis across 18 species representing diverse metazoan taxa.

Conclusions: PdumBase represents the first online resource for the early developmental transcriptome of *Platynereis dumerilii*. It serves as a research platform for discovery and exploration of gene expression during early stages, throughout the *Platynereis* life cycle, and enables comparison to other model organisms. PdumBase is freely available at http://pdumbase.gdcb.iastate.edu.

Keywords: *Platynereis dumerilii*, Early development, Spiralian, Life cycle, Expression profile, Database. Comparative genomics, Evo-devo

* Correspondence: sqs@iastate.edu
†Hsien-Chao Chou and Natalia Acevedo-Luna contributed equally to this work.
[1] Department of Genetics, Developmental and Cell Biology, Iowa State University, 503 Science Hall II, Ames, IA 50011, USA
Full list of author information is available at the end of the article

Background

The annelid *Platynereis dumerilii* is an ideal organism for developmental and comparative studies due to its phylogenetic position as a spiralian/lophotrochozoan, a species-rich but less known third branch of bilateral symmetrical organisms, whose members are instrumental for inferring ancestral bilaterian or urbilaterian features [1–5]. However, in comparison to the other two major bilaterian branches, the deuterostomes that include vertebrates and humans, and the ecdysozoans that include nematodes and arthropods like *C. elegans* and the fruit fly *Drosophila*, respectively, substantially less genetic and molecular studies have been performed in spiralian species. As a spiralian, *Platynereis dumerilii* exhibits intriguing developmental features that include a stereotypic cleavage pattern and invariant cell lineages with predictable cell fates [6–8]. Having invariant cell lineages makes it possible to link gene expression in distinct embryonic cells to later cell progeny and organs in larval and adult stages. Knowing this allows to make predictions, and test hypothesis of how molecular processes regulate the cellular origin and composition of organs and body parts [9, 10]. To date, this powerful property has only been exploited in a limited number of studies [11, 12] but, as better descriptions of later cell lineages and access to stage-specific and cell-specific gene expression data becomes available, its use is expected to increase substantially [13–16]. Additionally, the morphological and genomic attributes exhibited by *Platynereis* have been useful for inferring ancestral gene structures and ancient cell types representative of ancestral bilaterian species [17, 18]. These inferences are based on the findings that *Platynereis* shares many common features with vertebrates, including similar gene expression profiles during the development of the brain, the central nervous system, the eye, appendages, and muscles. These features have been instrumental in developing and testing hypotheses about the urbilaterian body plan and the origin of complex organ systems during animal evolution [3, 4, 19–22]. For these reasons, over the last decade, *Platynereis* has emerged as a powerful spiralian model for comparative genomic analyses, evolutionary developmental biology (evo-devo) studies, stem cells, regeneration, marine ecology, chronobiology and neurobiology [23–27]. *Platynereis* as a laboratory organism boasts an ever-increasing experimental toolkit, expanding the usefulness of this species in identifying and dissecting gene function: during early and late development; during organogenesis; in various cell types; and in neuronal circuits that dictate the circadian and lunar rhythm-controlled swimming behaviors [27–35].

The on average 3 months life cycle of *Platynereis dumerilii* including sexual maturation and mating synchronized by a monthly lunar cycle is well established under laboratory conditions, and contributes to the attractiveness of this annelid spiralian species as a powerful experimental organism [36]. Mating leads to 'instant' external fertilization of thousands of eggs, allowing experimental access to thousands of highly synchronously developing embryonic, larval, juvenile, and adult stages that can be used for large scale biochemical and a variety of Omics studies [37].

Early development in *Platynereis* is characterized by a spiral cleavage mode, a series of invariant, stereotypic asymmetric cell divisions stages, generating a spiral arrangement of embryonic cells characterized by vastly different cell sizes and distinct cell fates [6–9]. This early embryonic spiral cleavage phase transitions after six rounds of cell divisions [about 7 h post fertilization (hpf) at 18 °C], to bilateral symmetrical oriented asymmetric cell divisions. The time of development depends drastically on temperature. Features described at each stage refer to embryos developing at controlled temperature of 18 °C. By 14hpf, the continued rapid cell divisions have given rise to a distinctly organized embryo of ~ 330 cells, a pre-trochophore larval stage that emerges from the jelly coat that surrounds the egg, and begins to rotate via a ring of differentiated multi-ciliated cells, the prototroch, and a developing apical organ, a sensory organ at the animal pole [37].

Within the next few days of development *Platynereis* transitions through several free-swimming larval stages (a primitive trochophore at 24hpf, a more elaborate metatrochophore at 48hpf, and a nectochaeta larvae at 72hpf) [37]. These stages are morphologically characterized by the additions of distinct ciliary structures, an elaboration of the head and trunk region including the establishment of a complex nervous system, larval and adult eyes, and the formation of trunk segments that bear the first pairs of appendages. After 1 week, the now juvenile three segmented young worm switches from a free swimming to a benthic lifestyle. A growth zone at the posterior end continues to add segments and serial appendages throughout juvenile and adult stages while growing to its adult size of ~ 2 in. in length and sexual maturation within 3 months.

The above attributes make *Platynereis* a favorable subject for various high-throughput sequencing (HTS) techniques including but not limited to gene expression profiling though the quantitative analysis of transcriptomes (RNA-seq). RNA-seq provides an unbiased approach to determine the transcriptional inventory for a process, and, by capturing the dynamic temporal expression profiles through the identification of differentially expressed genes between developmental states, enables investigators to gain system-level insights into organismal development [10].

The massive amount of data produced by modern HTS experiments is both a challenge and an opportunity for countless biological discoveries. However, it can only be effectively utilized as a scientific resource in conjunction with dedicated computational pipelines to preprocess, analyze, store, and visualize this information. Well-established algorithms for preprocessing of raw sequencing reads in terms of quality control [38, 39] and de-multiplexing [40, 41] as well as robust approaches for transcriptome assembly [42–44], differential expression analysis [45–47] and functional enrichment studies have been developed. However, as the generated HTS datasets capture a genome-wide transcriptional response, the generated data richness and complexity is often magnitudes higher than the focused interest of the target study. Consequently, much of the data remains unexplored or is not easily accessible to the scientific community. Thus, computational and bioinformatics strategies have to be developed that enable broad, fast, and long-term accessibility of the data, as well as the intuitive ability to easily search, extract, and/or to graphically visualize molecular processes, pathways, and structural composition.

Ideally, such a research tool opens the discovery space and should be capable of efficiently retrieving the information of interest while scaling well with large data sizes. This in turn requires the processed data to be stored in a standardized, queryable, and platform-independent manner such that an application-specific subset of data can efficiently be served to an algorithmic solution built on top of such system. Notably, centralized information systems, such as relational databases, fulfill all of the above requirements by providing storage solutions combined with powerful querying language such as Structured Query Languages (SQL) [48]. These systems can be further combined with modern web technologies to form a graphical user interface capable of representing vast amounts of data in a concise, interactive and platform agnostic manner. Indeed, web-based database solutions have previously been leveraged to represent a large array of biological data over a wide range of organisms [49–51], as well as metadata such as functional annotations [52] and pathway information [53].

For *Platynereis*, several transcriptomic datasets have recently been reported [15, 34, 54], covering various developmental, larval, juvenile, and adult stages during its 3 month long life cycle. However, public accessibility, as well as ease of data analysis is currently limited and requires a priori knowledge of a transcript/gene to be able to retrieve sequence and expression data. This prior knowledge however is often not available, motivating the need for the development of databases, which enable investigators to query transcriptome and expression data based on functional annotation as well as other features. In addition, only a limited number of attempts have been made to link data to other important developmental model species in order to facilitate comparisons of developmental and cell biological modules, a prerequisite for building testable hypothesis in the field of evolutionary and developmental biology.

To close this gap, we developed PdumBase, a comprehensive, stand-alone, and user-friendly web-based resource, offering researchers a sophisticated platform to study molecular function, expression patterns, and relationships between genes at early embryonic stages in *Platynereis dumerilii.* PdumBase comprises transcriptomic data from the first 14 h of development of *Platynereis* capturing the transition from a single fertilized egg to a ~ 330 cells hatched early protrochophore stage. The data in PdumBase represents the first *Platynereis* transcriptome obtained from seven time points of early embryonic development, recently reported by our group [55]. PdumBase also displays data from a novel analysis to construct genome-wide co-expression networks to identify hub genes and genes sharing similar expression patterns in early embryonic stages. Furthermore, PdumBase incorporates transcriptomic data from Conzelmann M. et al. [34] that includes selected time points of later developmental stages including larval, juvenile, and adult stages covering the entire life cycle, as well as sexually mature male and female worms. To facilitate comparative and evo-devo studies PdumBase includes searchable expression data from distinct developmental stages of six other species that were available at the time of database construction: *Strongylocentrotus purpuratus* [56], *Xenopus tropicalis* [57], *Danio rerio* [58], *Ascaris suum* [59], *Nematostella vectensis* [60], and human [61]. The selection of species for comparative analysis of early development, includes a deuterostome invertebrate and vertebrates, an ecdysozoan, a cnidarian, and transcriptional 'germ layer' states of early differentiated human stem cells, respectively, an arbitrary yet informative choice for this 'pilot' feature of PdumBase that will be expanded in future releases. Finally, to understand how the *Platynereis* gene models are related to genes in other species, and to provide additional validation for the previous gene annotations, a comprehensive phylogenetic analysis that includes 18 selected species, representing different metazoan taxa of various evolutionary distances, including key spiralian species like the annelid *Capitella teleta* and the mollusk *Crassostrea gigas,* is available for each *Platynereis* gene model. Thus, this research tool enables exploration, discovery and hypothesis building to the community at large by providing intuitive and interactive access to genome-wide developmental expression data of early development in *Platynereis.*

Construction and content

PdumBase is a user-friendly database that provides a platform to easily access and compare expression profiles of *Platynereis* transcripts during early stages of development (2 to 14 hpf), with the option of displaying later stages of development (24 hpf to 3 months, and sexually mature male and female). For each transcript, PdumBase contains detailed annotation, a preliminary phylogeny tree based on orthologous groups, and co-expression networks. Co-expression data can be obtained based on differential gene expression between consecutive time points throughout early development. In addition, the database integrates transcriptomic information of six additional species, giving the option of examining expression data of orthologous genes. Fig. 1 summarizes the content and the organization of data available in PdumBase.

Transcriptome assembly and annotation pipeline

To create the early reference transcriptome for *Platynereis dumerilii*, we combined the mRNA data of all sequenced libraries (~ 785 million paired-end reads) at 2, 4, 6, 8, 10, 12, and 14 hpf and performed de novo transcriptome assembly using Trinity [62] resulting in a total of 273,087 non-redundant contigs (N50 size: 1466 bp). Detailed methods, characterization and validation on the transcriptome de novo assembly, and the annotation pipeline for gene models have been reported previously [55].

To annotate the assembly, we first identified 28,580 genes (51,260 transcripts) with a predicted open reading frame (ORF) of more than 100 amino acids. These potential protein coding genes were consequently aligned to the Swiss-Prot database [63] and only retained, if the

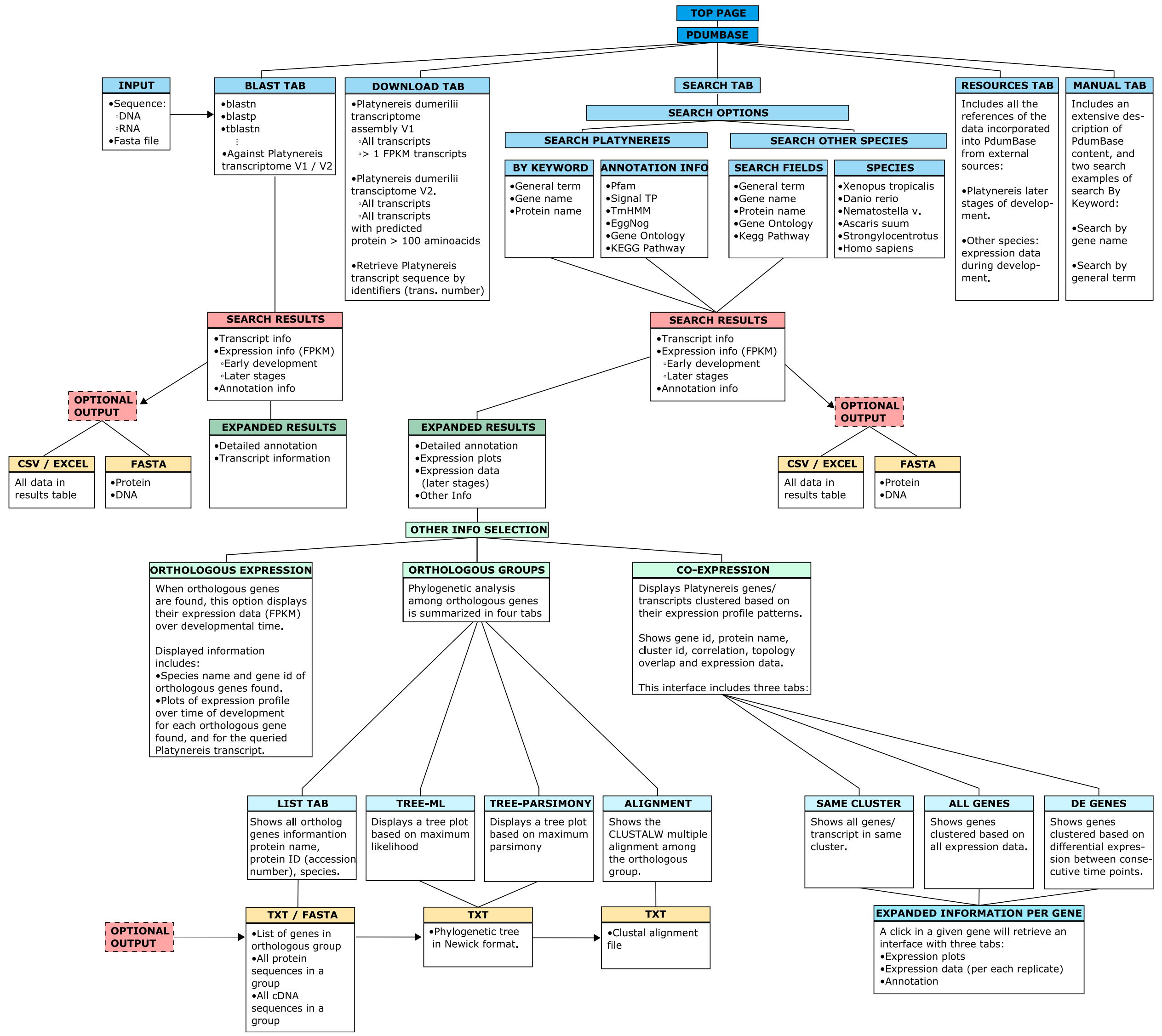

TOP PAGE
PDUMBASE
INPUT
•Sequence:
◦DNA
◦RNA
•Fasta file
BLAST TAB
•blastn
•blastp
•tblastn
⋮
•Against Platynereis transcriptome V1 / V2
DOWNLOAD TAB
•Platynereis dumerilii transcriptome assembly V1
◦All transcripts
◦> 1 FPKM transcripts
•Platynereis dumerilii transciptome V2.
◦All transcripts
◦All transcripts with predicted protein > 100 aminoacids
•Retrieve Platynereis transcript sequence by identifiers (trans. number)
SEARCH TAB
SEARCH OPTIONS
SEARCH PLATYNEREIS
SEARCH OTHER SPECIES
BY KEYWORD
•General term
•Gene name
•Protein name
ANNOTATION INFO
•Pfam
•Signal TP
•TmHMM
•EggNog
•Gene Ontology
•KEGG Pathway
SEARCH FIELDS
•General term
•Gene name
•Protein name
•Gene Ontology
•Kegg Pathway
SPECIES
•Xenopus tropicalis
•Danio rerio
•Nematostella v.
•Ascaris suum
•Strongylocentrotus
•Homo sapiens
RESOURCES TAB
Includes all the references of the data incorporated into PdumBase from external sources:
•Platynereis later stages of development.
•Other species: expression data during development.
MANUAL TAB
Includes an extensive description of PdumBase content, and two search examples of search By Keyword:
•Search by gene name
•Search by general term
SEARCH RESULTS
•Transcript info
•Expression info (FPKM)
◦Early development
◦Later stages
•Annotation info
OPTIONAL OUTPUT
CSV / EXCEL
All data in results table
FASTA
•Protein
•DNA
EXPANDED RESULTS
•Detailed annotation
•Transcript information
SEARCH RESULTS
•Transcript info
•Expression info (FPKM)
◦Early development
◦Later stages
•Annotation info
EXPANDED RESULTS
•Detailed annotation
•Expression plots
•Expression data (later stages)
•Other Info
OPTIONAL OUTPUT
CSV / EXCEL
All data in results table
FASTA
•Protein
•DNA
OTHER INFO SELECTION
ORTHOLOGOUS EXPRESSION
When orthologous genes are found, this option displays their expression data (FPKM) over developmental time.
Displayed information includes:
•Species name and gene id of orthologous genes found.
•Plots of expression profile over time of development for each orthologous gene found, and for the queried Platynereis transcript.
ORTHOLOGOUS GROUPS
Phylogenetic analysis among orthologous genes is summarized in four tabs
CO-EXPRESSION
Displays Platynereis genes/ transcripts clustered based on their expression profile patterns.
Shows gene id, protein name, cluster id, correlation, topology overlap and expression data.
This interface includes three tabs:
LIST TAB
Shows all ortholog genes informantion protein name, protein ID (accession number), species.
TREE-ML
Displays a tree plot based on maximum likelihood
TREE-PARSIMONY
Displays a tree plot based on maximum parsimony
ALIGNMENT
Shows the CLUSTALW multiple alignment among the orthologous group.
SAME CLUSTER
Shows all genes/ transcript in same cluster.
ALL GENES
Shows genes clustered based on all expression data.
DE GENES
Shows genes clustered based on differential expression between consecutive time points.
OPTIONAL OUTPUT
TXT / FASTA
•List of genes in orthologous group
•All protein sequences in a group
•All cDNA sequences in a group
TXT
•Phylogenetic tree in Newick format.
TXT
•Clustal alignment file
EXPANDED INFORMATION PER GENE
A click in a given gene will retrieve an interface with three tabs:
•Expression plots
•Expression data (per each replicate)
•Annotation

Fig. 1 Schematic illustration of the PdumBase sitemap. The content for each tabulator of the PdumBase home interface is summarized in "Tab Boxes" (blue) at the top of the figure."Search Results" (pink), and "Expanded Results" (dark green) outline data retrieved from different search options. The result interface includes the "Show other info" option; a summary of its content is shown in "Other info selection" box (light green). Expandable results option for. Orthologous groups and Co-expression are also shown (light blue). Optional output summarizes downloadable content as well as file format options (light yellow)

corresponding E-value of the alignment had a value of at least 10e-10 or lower. Using this procedure, a total of 17,213 genes (31,806 transcripts) were matched [55]. Pfam domains, signal peptide cleavage sites and transmembrane helices were also identified and integrated into the database. These were predicted using HMMER [64], signal [65] and tmHMM [66] respectively with default parameters. Finally, Gene Ontology (GO) and eggNOG information were obtained from the Swiss-Prot database [63], whereas associated KEGG pathways were identified by a BLASTP search with an E-value cutoff 10e-10.

Expression analysis

To build a comprehensive transcriptional profile of early *Platynereis* development, we extracted mRNA from *Platynereis dumerilii* 2, 4, 6, 8, 10, 12, and 14 h post fertilization (hpf), each with two biological replicas, as described [55]. The resulting reads from these samples were subsequently mapped to the assembled transcripts using Bowtie [67]. The read count for each transcript was estimated by RSEM [68]. The trimmed mean of M-values (TMM) [42], normalized FPKM was calculated using the R package "edgeR" [45]. A transcript was considered as "expressed" if its FPKM is larger than 1. This represents an empirically chosen inclusionary cutoff based on our current methods of validation (1) by whole mount in situ hybridization, which is able to detect transcripts above a level of ~ 5 FPKM, and (2) successful amplification of transcripts by PCR from stages with an FPKM of 1 or higher. Whole mount in situ hybridization was chosen as a method of validation to enable us to determine spatial expression of transcripts, and as a method that was sensitive enough to visualize the onset of gene transcription in a single cell, and on two genomic loci within each nucleus [55].

Thus, for expression analysis, 20,977 transcripts were included that fulfilled the criteria of being expressed in at least one stage at an FPKM larger than 1. These transcripts were clustered based on their expression pattern into 32 groups. We obtained the expression dendrogram using the hierarchical clustering algorithm (R function hclust). Since the determination of number of clusters is context-dependent, we first used an automatic tree-cutting algorithm to classify all genes into 32 groups (R function cutree, k = 32). Pairs of groups were then merged as moving up the hierarchy. Each merge was validated by visual inspection of their average expression patterns of all stages. We repeated this step until no qualified merges were found. This process leads to 15 distinct clusters demonstrating distinct expression patterns. For successive time stages, the differentially expressed (DE) genes were determined using edgeR and considered as differentially expressed if the adjusted *p*-value was smaller or equal to 0.001. A co-expression network was also constructed using WGCNA [69], an R package for weighted correlation network analysis. The co-expression network can provide not only the correlation between two genes, but also their topological similarity. This similarity captures how closely related two genes are by examining if they share similar connectivity with other "third party" genes (Additional file 1: Figure S1). In addition to early stage expression data (2 to 14 hpf), we also incorporated 310 million reads of later stage RNA sequencing datasets ranging from 24 h to 3 months, and sexually mature male and female provided by Conzelmann, et al. [34], that were first mapped to our gene models and displayed within the database. The inclusion of this information provides a more comprehensive time series of gene expression profiles at various times throughout the *Platynereis dumerilii* life cycle.

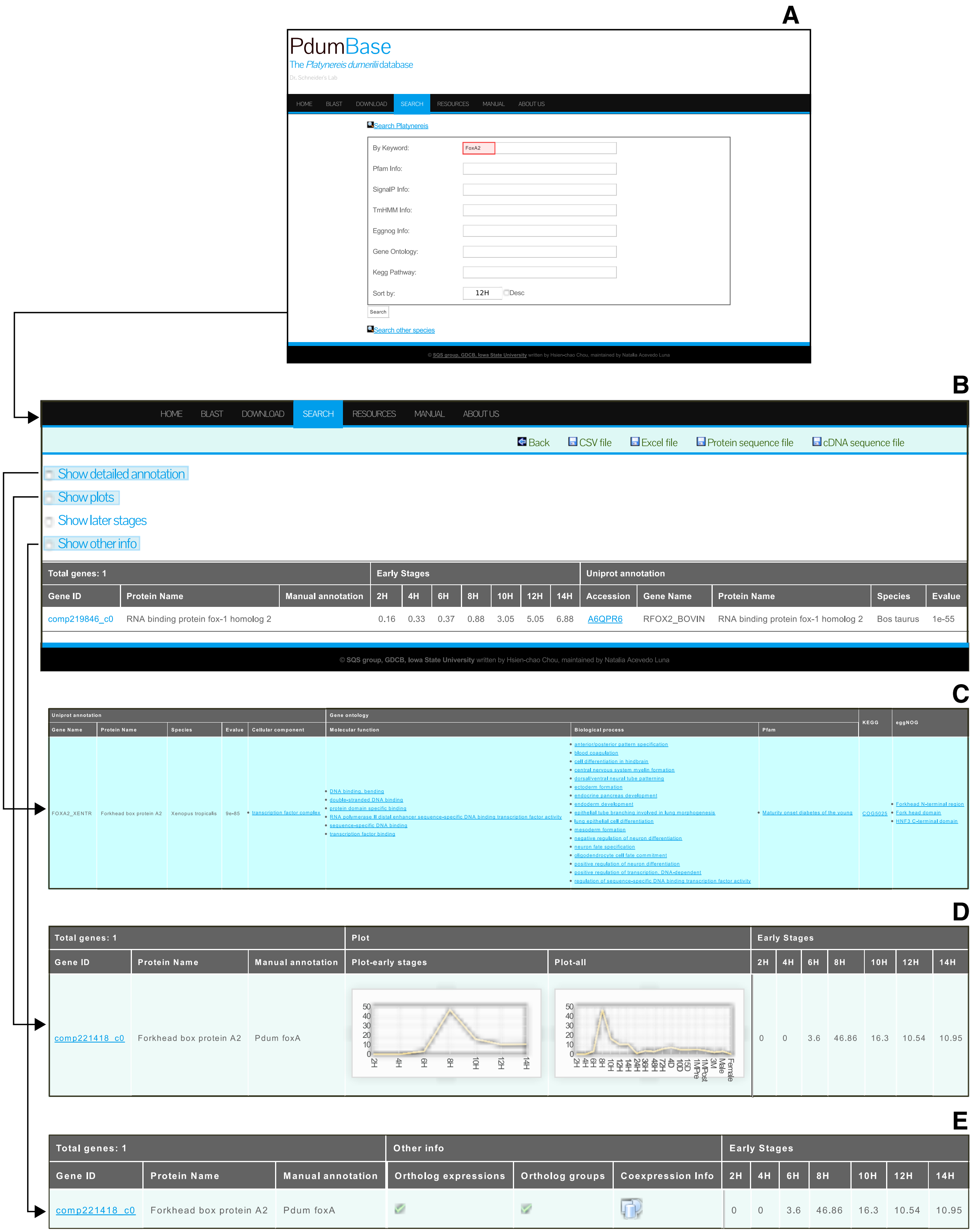
Comparative transcriptome

To enable comparative studies of gene expression between model organisms during early embryogenesis, we incorporated the expression data from six additional species including *Danio rerio* [58], *Xenopus tropicalis* [57], *Homo sapiens* [61], *Strongylocentrotus purpuratus* [56], *Nematostella vectensis* [60] and *Ascaris suum* [59]. Orthologous groups were identified using OrthoMCL [70] with reciprocal BLASTP search at protein level.

To further assist comparative analyses and to provide annotation validation, we constructed a large scale evolutionary comparison between our gene models with 17 other species in phylogenetically informative positions including *Capitella teleta* [71], *Helobdella robusta* [71], *Lottia gigantea* [71] *Crassostrea gigas* [72], *Daphnia pulex* [71], *Tribolium castaneum* [71], *Drosophila melanogaster* [50], *Strongylocentrotus purpuratus* [51], *Saccoglossus kowalevskii* [71], *Branchiostoma floridae* [71], *Danio rerio* [73], *Xenopus tropicalis* [71], *Homo sapiens* [73], *Nematostella vectensis* [71], *Amphimedon queenslandica* [71], *Trichoplax adherens* [71], and *Monosiga brevicollis* [71]. A custom OrthoMCL pipeline was used to identify a total of 40,206 homologous groups out of which 32,482 groups have at least two species. A pairwise BLASP search with cutoff 1e-05 and 50% identity was performed. The orthologous relationship was established only if genes are the reciprocal best hits for any two species. For each *Platynereis dumerilii* gene having orthologous genes, a phylogenetic tree was created using RAxML [74]. Both, maximum likelihood and parsimony trees can be accessed in PdumBase. The multiple alignment for each orthologous group is constructed using CLUSTALW [75].

Utility and discussion

PdumBase provides a comprehensive, versatile online tool to investigate stage specific transcriptional inputs during embryogenesis and throughout the life cycle of the annelid *Platynereis dumerilii* and other selected species. Users can search the genes of interest and examine their functional



PdumBase

The *Platynereis dumerilii* database

Dr. Schneider's Lab

Search Platynereis

By Keyword:	FoxA2
Pfam Info:	
SignalP Info:	
TmHMM Info:	
Eggnog Info:	
Gene Ontology:	
Kegg Pathway:	
Sort by:	12H □ Desc

Search other species

© SQS group, GDCB, Iowa State University written by Hsien-chao Chou, maintained by Natalia Acevedo Luna

□ Show detailed annotation

□ Show plots

□ Show later stages

□ Show other info

Total genes: 1			Early Stages							Uniprot annotation				
Gene ID	Protein Name	Manual annotation	2H	4H	6H	8H	10H	12H	14H	Accession	Gene Name	Protein Name	Species	Evalue
comp219846_c0	RNA binding protein fox-1 homolog 2		0.16	0.33	0.37	0.88	3.05	5.05	6.88	A6QPR6	RFOX2_BOVIN	RNA binding protein fox-1 homolog 2	Bos taurus	1e-55

© SQS group, GDCB, Iowa State University written by Hsien-chao Chou, maintained by Natalia Acevedo Luna

Uniprot annotation					Gene ontology			KEGG	eggNOG
Gene Name	Protein Name	Species	Evalue	Cellular component	Molecular function	Biological process	Pfam		
FOXA2_XENTR	Forkhead box protein A2	Xenopus tropicalis	9e-85	• transcription factor complex	• DNA binding, bending • double-stranded DNA binding • protein domain specific binding • RNA polymerase II distal enhancer sequence-specific DNA binding transcription factor activity • sequence-specific DNA binding • transcription factor binding	• anterior/posterior pattern specification • blood coagulation • cell differentiation in hindbrain • central nervous system myelin formation • dorsal/ventral neural tube patterning • ectoderm formation • endocrine pancreas development • endoderm development • epithelial tube branching involved in lung morphogenesis • lung epithelial cell differentiation • mesoderm formation • negative regulation of neuron differentiation • neuron fate specification • oligodendrocyte cell fate commitment • positive regulation of neuron differentiation • positive regulation of transcription, DNA-dependent • regulation of sequence-specific DNA binding transcription factor activity	• Maturity onset diabetes of the young	COG5025	• Forkhead N-terminal region • Fork head domain • HNF3 C-terminal domain

Total genes: 1			Plot		Early Stages						
Gene ID	Protein Name	Manual annotation	Plot-early stages	Plot-all	2H	4H	6H	8H	10H	12H	14H
comp221418_c0	Forkhead box protein A2	Pdum foxA			0	0	3.6	46.86	16.3	10.54	10.95

Total genes: 1			Other info			Early Stages						
Gene ID	Protein Name	Manual annotation	Ortholog expressions	Ortholog groups	Coexpression Info	2H	4H	6H	8H	10H	12H	14H
comp221418_c0	Forkhead box protein A2	Pdum foxA	✓	✓	▣	0	0	3.6	46.86	16.3	10.54	10.95

Fig. 2 (See legend on next page.)

(See figure on previous page.)
Fig. 2 PdumBase search interface and expandable search result options. **a** PdumBase Search interface. An example under Functional Annotation on "Keyword search" field is shown. **b** The PdumBase Results interface displays the expression data at different stages of development for each identified gene. **c** Screen capture of the interface obtained when the option "Show detailed annotation" is selected from the Results screen. Detailed annotation includes Gene Ontology, KEGG, EggNOG, and more. **d** The results interface when the option "Show plots" is selected. This option displays the expression data (in FPKM) plotted over time of development for each gene found. **e** Result interface with the "Show other info" option selected. Clicking on the Green checkmark icon will redirect to a new tab with expanded search results for: "Ortholog Expression", "Ortholog groups", and "Co-expression"

annotation, expression profiles, co-expression networks, and use the genes for comparative analyses. In addition to the search module, PdumBase also offers the option of downloading the queried data and searching the transcriptome using BLAST. Here we describe the interface and functions of these modules.

Annotation search

To facilitate utility of PdumBase for a broad range of research questions, users may want to mine data using a variety of search parameters. This versatile accessibility to retrieve genes of interest is provided through a powerful search engine allowing users to query specific keywords covering a large array of functional annotation categories. These include gene ID, protein name, gene ontology, Pfam, KEGG pathways, eggnog, SignalP and tmHMM information. The search results are displayed as an expression profile of the matched genes as well as their functional annotation information. By default, the results page displays the Swiss-Prot annotation based on protein level BLAST search, and the expression data for early stages (2 to 14 hpf). Users can expand this table to include expression profiles of later stages (by checking the "Show later stages" box), detailed annotation (by selecting the "Show detailed annotation" box) and other advanced information by selecting "Show plots" and "Show other info" in the interface (Fig. 2). More detailed annotation can be obtained from the interface with links to the gene ontology, Pfam, KEGG pathways, EggNOG, SignalP and tmHMM information of the corresponding genes. Links to external data sources are also provided. Where available, the links to additional pages will be shown, including the expression of orthologous genes whereas phylogenetic analysis and co-expression network information can be obtained by clicking "Show other info" (Fig. 2).

Sequence similarity search

In addition to searching using keywords, PdumBase also supports sequence-based queries via the BLAST search page. Users can input either nucleotide or protein sequence. The search options for BLAST, including the E-value, the number of alignments, and the type of search are customizable. The search results return the expression as well as the functional annotation information for the matched sequence.

Additional properties of each gene such as (i) isoform level data and corresponding plots, (ii) replicate information including all biological and technical replicates, (iii) the raw read count information as estimated by RSEM, and (iv) detailed annotation information (Additional file 2: Figure S2) can also be accessed from the gene ID page. While the main results page displays information on the best hit for genes in *Platynereis*, the detailed annotation information provides information regarding the source of the annotation.

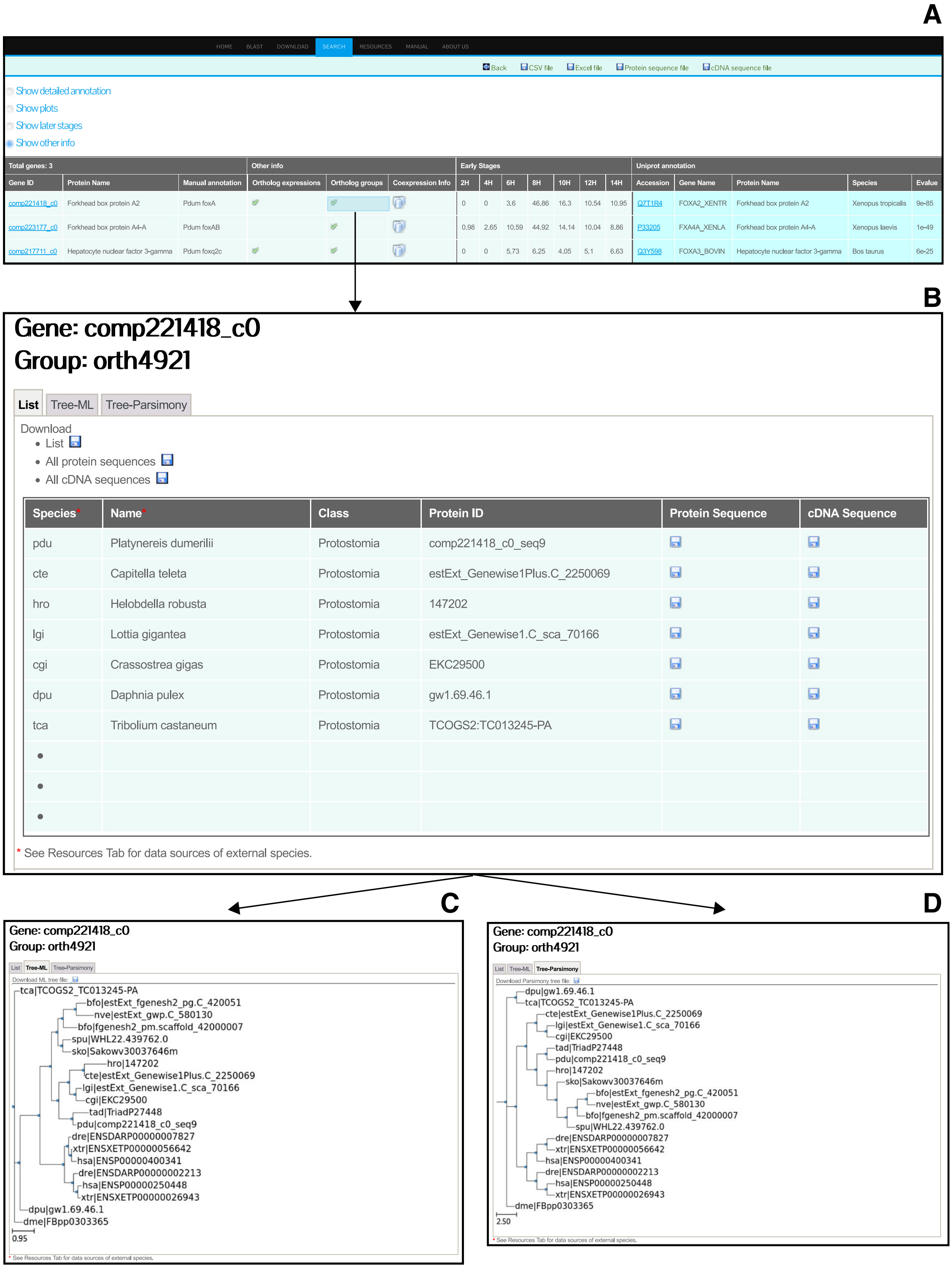
Comparative analysis

For investigators interested in comparative studies, PdumBase also provides a searchable interface for an additional six species including *Danio rerio*, *Xenopus tropicalis*, *Homo sapiens*, *Nematostella vectensis*, *Strongylocentrotus purpuratus*, and *Ascaris suum*.

Through "Show other info" in the search result page, users can link to comparative information provided under the "Ortholog expressions" column if orthologous genes were identified through our automated approach (see above). The developmental expression profiles for all the orthologous genes are also provided in a single window to facilitate easy comparison of orthologous expression patterns. To further facilitate comparative studies and to enable additional validation for the annotations, PdumBase displays the phylogenetic analyses for each *P. dumerilii* gene with homologous counterparts across 18 species. The "Orthologous groups" page includes four tabs: (i) a list of information related to all orthologous genes including the protein name, protein ID (accession number), species, species class and the links for downloading their protein or cDNA sequences, (ii) a tree plot based on maximum likelihood, (iii) a tree plot based on maximum parsimony, and (iv) a CLUSTALW multiple alignment report. The tree files, alignment report and all the protein/cDNA sequences are downloadable (Fig. 3).

Co-expression information

Expression patterns shared by multiple genes may indicate the function in a similar biological process. By clicking on corresponding entry in the "Co-expression Info" column, PdumBase provides co-expression information that may reveal the potential interaction among a set of



Total genes: 3			Other info			Early Stages							Uniprot annotation				
Gene ID	Protein Name	Manual annotation	Ortholog expressions	Ortholog groups	Coexpression Info	2H	4H	6H	8H	10H	12H	14H	Accession	Gene Name	Protein Name	Species	Evalue
comp221418_c0	Forkhead box protein A2	Pdum foxA	✓	✓	🗐	0	0	3.6	46.86	16.3	10.54	10.95	Q7T1R4	FOXA2_XENTR	Forkhead box protein A2	Xenopus tropicalis	9e-85
comp223177_c0	Forkhead box protein A4-A	Pdum foxAB		✓	🗐	0.98	2.65	10.59	44.92	14.14	10.04	8.86	P33205	FXA4A_XENLA	Forkhead box protein A4-A	Xenopus laevis	1e-49
comp217711_c0	Hepatocyte nuclear factor 3-gamma	Pdum foxq2c	✓	✓	🗐	0	0	5.73	6.25	4.05	5.1	6.63	Q3Y598	FOXA3_BOVIN	Hepatocyte nuclear factor 3-gamma	Bos taurus	6e-25

Gene: comp221418_c0
Group: orth4921

Download
- List 💾
- All protein sequences 💾
- All cDNA sequences 💾

Species*	Name*	Class	Protein ID	Protein Sequence	cDNA Sequence
pdu	Platynereis dumerilii	Protostomia	comp221418_c0_seq9	💾	💾
cte	Capitella teleta	Protostomia	estExt_Genewise1Plus.C_2250069	💾	💾
hro	Helobdella robusta	Protostomia	147202	💾	💾
lgi	Lottia gigantea	Protostomia	estExt_Genewise1.C_sca_70166	💾	💾
cgi	Crassostrea gigas	Protostomia	EKC29500	💾	💾
dpu	Daphnia pulex	Protostomia	gw1.69.46.1	💾	💾
tca	Tribolium castaneum	Protostomia	TCOGS2:TC013245-PA	💾	💾
•					
•					
•					

* See Resources Tab for data sources of external species.

Gene: comp221418_c0
Group: orth4921

* See Resources Tab for data sources of external species.

Gene: comp221418_c0
Group: orth4921

* See Resources Tab for data sources of external species.

Fig. 3 PdumBase expandable results option. **a** PdumBase results interface displaying the options under "Show Other Info". Click on the green check mark will open a new tab. **b** Interface retrieved when the green check mark icon under "Ortholog group" is selected. The resulting interface displays for the particular *Platynereis* gene under search, all the species in which an orthologous gene was found. The protein and/or cDNA sequences can be downloaded individually for each orthologous gene, or for the complete set shown in the result table. **c** and **d** Display Phylogenetic trees of the orthologous groups: ML tree (**c**) and Tree- Parsimony (**d**)

genes. This page is tabulated into three categories, namely "The same cluster", "All", and "DE genes". Genes found within the same cluster by clustering analysis are shown in the first tab whereas the second tab includes every developmental gene expression profile independent of the cluster analysis. The co-expression network also provides the topological overlap information, which can be used to examine the similarity of interaction with all other genes in the network. The "DE genes" tab shows the co- expressed genes between each adjacent developmental stage for both, up-regulated and down-regulated genes.

Database implementation

All data content is stored in a MySQL database. The web interface was implemented using PHP and the Smarty template engine. The business logics and presentation are separated by employing the model-view-controller design pattern. A centralized controller is used to coordinate the client requests and generate the corresponding views. All plots are either generated by server-side R scripts or client-side JavaScripts. In addition, a BLAST database was constructed for the transcriptome assembly using NCBI BLAST.

Additional features

Data export

The annotation and BLAST search results are all downloadable in text format (CSV) or as an Excel sheet. In addition, all assembled transcripts; annotation, expression data and differentially expressed genes are also available for download on our website. The raw Illumina reads are available upon request.

Manual

A manual providing a detailed description of all the features and how to access them is available on the PdumBase web page http://pdumbase.gdcb.iastate.edu/platynereis/controller.php?action=manual, as well as in Additional file 3: PDF File 1.

Conclusions

PdumBase offers a user-friendly platform for researchers to study the regulatory landscape for a spiral-cleaving embryo with an emphasis on early developmental stages of *Platynereis dumerilii,* and selected later stages throughout its life cycle. It provides researchers with an online tool for fast, dynamic, retrieval of expression patterns during early stages of development and life cycle. Furthermore, the large-scale data sets and comparative analyses offer valuable information for the studies of molecular dynamics and evolutionary aspects of the *P. dumerilii* transcriptome in comparison to other model systems. It represents one of the first attempts to integrate and harness developmental expression profiles from various species in combination with phylogeny based annotations offering a versatile online research tool to discover and investigate various aspects of animal evolution and development.

Future expansions for PdumBase may include genome-wide expression profiles after experimental manipulations, single cell transcriptomic data for early stages of development, and genomic information of regulatory regions to provide further entry points for promoter and network analysis. Further database subdivisions could display more detailed and/or manually curated aspect of early development e.g. asymmetric cell division, distinct pathways, or the emergence of distinct cell lineages and cell types. Additionally, we consider the inclusion of images of gene expression data as a possible and valuable future expansion. Further PdumBase expansions may include an interactive visualization of clustered genes based on expression profile. Such tool would enable the user to fine tune the cluster parameters while simultaneously visualizing the change in cluster composition allowing to user to examine different sets of genes at a time.

As such, PdumBase can be seen as a prototype for an online research tool to make any large-scale genome-wide data set quickly accessible to researchers without requiring prior expertise in bioinformatics, showcasing how valuable and extensive transcriptional data set can be made accessible for community wide data mining. The versatility and variety of search options enable a wider range of research questions to be investigated both within a single laboratory and across the scientific community. The creation of such a database opens the genome-wide discovery space of data sets to the entire greater scientific community.

Additional files

Additional file 1: Figure S1. Expandable results from PdumBase Result Interface: Co-expression **A**. The Result Interface with the option "Show other info" selected. **B**. Co-expression information interface under the tabulator "The same cluster" displays all the transcripts/genes in the same cluster of a given component. Shown are protein name, correlation, topology overlap, and expression data. **C**. Co-expression information interface under the tabulator "All" displays all genes sorted by the ranking according to correlation score. **D**. Co-expression information interface under the tabulator "DE genes" displays differentially expressed genes between consecutive time points. (PDF 397 kb)

Additional file 2: Figure S2. Expandable results from PdumBase Result Interface: Gene models **A**. The Result Interface allows to click on each gene found, expanding its expression and annotation information. **B**. Shows the interface obtained when clicking on a Gene ID from the result table. Plots of expression data over time of development are shown (early stages and all stages are plotted) Expression plot includes data of all isoforms found for a gene. **C**. Expression Data tab. This tab displays FPKM and Raw counts for a given gene and its possible isoforms. **D**. Annotation Tab. This interface retrieves detailed annotation information. (PDF 502 kb)

Additional file 3: PDF File 1. PdumBase Manual. A description of PdumBase content with detailed instructions on how to use all its features. (PDF 4093 kb)

Abbreviations

BLAST: Basic Local Alignment Search Tool; FPKM: Fragments Per Kilobase per Million mapped reads; GO: Gene ontology; hpf: Hours post fertilization; KEGG: Kyoto Encyclopedia of Genes and Genomes; ORF: Open reading frame; TMM: Trimmed Mean of M-values

Acknowledgements

The authors wish to thank Levi Baber for setting up and maintaining the production server of PdumBase.

Funding

Funding for this work was provided by the Roy J. Carver Charitable Trust and the National Science Foundation (Award ID 1455185) to SQS. The funding bodies had no role in the design of the study or the collection, analysis or interpretation of the data or the writing of the manuscript.

Availability of data and materials

PdumBase is freely at http://pdumbase.gdcb.iastate.edu. The transcriptome assembly and functional annotation is available in the main menu download tab. Raw, preprocessed reads, and BAM files are available upon requests.

Authors' contributions

SQS and HCC conceived the study. HCC processed the data and developed the database. NAL curated the public version of the database, and prepared the manual and figures. HCC and NAL wrote the manuscript. SQS (invertebrate data sets), and JAK (vertebrate data sets) designed and coordinated the study, and revised the manuscript. All authors read, corrected and approved the final manuscript.

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare they have no competing interests.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Author details

[1]Department of Genetics, Developmental and Cell Biology, Iowa State University, 503 Science Hall II, Ames, IA 50011, USA. [2]Present address: Center for Cancer Research, National Institutes of Health, Rockville, MD 20894, USA.

Received: 6 April 2018 Accepted: 31 July 2018
Published online: 16 August 2018

References

1. Raible F, Tessmar-Raible K. Platynereis dumerilii. Curr Biol. 2014;24:R676–7.
2. Lauri A, Brunet T, Handberg-Thorsager M, Fischer AHL, Simakov O, Steinmetz PRH, et al. Development of the Annelid Axochord: Insights into notochord evolution. Science. 2014;345:1365–8.
3. Tessmar-Raible K, Raible F, Christodoulou F, Guy K, Rembold M, Hausen H, et al. Conserved sensory-neurosecretory cell types in annelid and fish forebrain: insights into hypothalamus evolution. Cell. 2007;129:1389–400.
4. Denes AS, Jékely G, Steinmetz PRH, Raible F, Snyman H, Prud'homme B, et al. Molecular architecture of annelid nerve cord supports common origin of nervous system centralization in Bilateria. Cell. 2007;129:277–88.
5. Arendt D, Technau U, Wittbrodt J. Evolution of the bilaterian larval foregut. Nature. 2001;409:81–5.
6. Wilson EB. The cell-lineage of Nereis. A contribution to the cytogeny of the annelid body. J Morphol. 1892;6:361–480.
7. Dorresteijn AWC. Quantitative analysis of cellular differentiation during early embryogenesis of Platynereis dumerilii. Roux's Arch Dev Biol. 1990;199:14–30.
8. Schneider S, Fischer A, Dorresteijn AWC. A morphometric comparison of dissimilar early development in sibling species of Platynereis (Annelida, Polychaeta). Roux's Arch Dev Biol. 1992;201:243–56.
9. Ackermann C, Dorresteijn A, Fischer A. Clonal domains in postlarval Platynereis dumerilii (Annelida: Polychaeta). J Morphol. 2005;266:258–80.
10. Williams EA, Jékely G. Towards a systems-level understanding of development in the marine annelid Platynereis dumerilii. Curr Opin Genet Dev. 2016;39:175–81.
11. Schneider SQ, Bowerman B. β-Catenin Asymmetries after All Animal/Vegetal- Oriented Cell Divisions in *Platynereis dumerilii* Embryos Mediate Binary Cell-Fate Specification. Dev Cell. 2007;13:73–86.
12. Achim K, Eling N, Vergara HM, Bertucci PY, Musser J, Vopalensky P, et al. Whole-body single-cell sequencing reveals transcriptional domains in the annelid larval body. Mol Biol Evol. 2018;35(5):1047–62. Available from: http://academic.oup.com/mbe/advance-article/doi/10.1093/molbev/msx336/4823215
13. Özpolat BD, Handberg-Thorsager M, Vervoort M, Balavoine G. Cell lineage and cell cycling analyses of the 4d micromere using live imaging in the marine annelid platynereis dumerilii. Elife. 2017;6:e30463.
14. Vergara HM, Bertucci PY, Hantz P, Tosches MA, Achim K, Vopalensky P, et al. Whole-organism cellular gene-expression atlas reveals conserved cell types in the ventral nerve cord of *Platynereis dumerilii*. Proc Natl Acad Sci. 2017;114:5878–85. Available from: http://www.pnas.org/lookup/doi/10.1073/pnas.1610602114
15. Achim K, Pettit J-B, Saraiva LR, Gavriouchkina D, Larsson T, Arendt D, et al. High-throughput spatial mapping of single-cell RNA-seq data to tissue of origin. Nat Biotechnol [Internet]. 2015;33:503–9. Available from: https://doi.org/10.1038/nbt.3209
16. Fischer AHL, Arendt D. Mesoteloblast-like mesodermal stem cells in the polychaete annelid platynereis dumerilii (nereididae). J Exp Zool Part B Mol Dev Evol. 2013;320:94–104.
17. Raible F. Vertebrate-Type Intron-Rich Genes in the Marine Annelid *Platynereis dumerilii*. Science. 2005;310:1325–6. Available from: http://www.sciencemag.org/cgi/doi/10.1126/science.1119089
18. Arendt D, Tessmar-Raible K, Snyman H, Dorresteijn AW, Wittbrodf J. Ciliary photoreceptors with a vertebrate-type opsin in an invertebrate brain. Science. 2004;306:869–71.
19. Arendt D, Tessmar K, de Campos-Baptista M-IM, Dorresteijn A, Wittbrodt J. Development of pigment-cup eyes in the polychaete Platynereis dumerilii and evolutionary conservation of larval eyes in Bilateria. Development. 2002;129:1143–54. Available from: http://www.ncbi.nlm.nih.gov/pubmed/11874910
20. Dray N, Tessmar-Raible K, Le GM, Vibert L, Christodoulou F, Schipany K, et al. Hedgehog signaling regulates segment formation in the annelid Platynereis. Science. 2010;329:339–42.
21. Brunet T, Fischer AHL, Steinmetz PRH, Lauri A, Bertucci P, Arendt D. The evolutionary origin of bilaterian smooth and striated myocytes. Elife. 2016;5:e19607.
22. Grimmel J, Dorresteijn AWC, Fröbius AC. Formation of body appendages during caudal regeneration in Platynereis dumerilii: adaptation of conserved molecular toolsets. EvoDevo. 2016;7
23. Gazave E, Béhague J, Laplane L, Guillou A, Préau L, Demilly A, et al. Posterior elongation in the annelid Platynereis dumerilii involves stem cells molecularly related to primordial germ cells. Dev Biol. 2013;382:246–67.
24. Rebscher N, Lidke AK, Ackermann CF. Hidden in the crowd: primordial germ cells and somatic stem cells in the mesodermal posterior growth zone of the polychaete Platynereis dumerillii are two distinct cell populations. EvoDevo. 2012;3
25. Pfeifer K, Dorresteijn AWC, Fröbius AC. Activation of Hox genes during caudal regeneration of the polychaete annelid Platynereis dumerilii. Dev Genes Evol. 2012;222:165–79.
26. Yang M, Liu J, Zhang X, Richardson SD. Comparative toxicity of chlorinated saline and freshwater wastewater effluents to marine organisms. Environ Sci Technol. 2015;49:14475–83.
27. Zantke J, Ishikawa-Fujiwara T, Arboleda E, Lohs C, Schipany K, Hallay N, et al. Circadian and circalunar clock interactions in a marine annelid. Cell Rep. 2013;5:99–113.
28. Backfisch B, Kozin VV, Kirchmaier S, Tessmar-Raible K, Raible F. Tools for gene-regulatory analyses in the marine annelid *Platynereis dumerilii*. PLoS One. 2014;9:e93076.
29. Bannister S, Antonova O, Polo A, Lohs C, Hallay N, Valinciute A, et al. TALENs mediate efficient and heritable mutation of endogenous genes in the marine annelid Platynereis dumerilii. Genetics. 2014;197:77–89.
30. Zantke J, Bannister S, Veedin Rajan VB, Raible F, Tessmar-Raible K. Genetic and genomic tools for the marine annelid Platynereis dumerilii. Genetics. 2014;197:19–31.
31. Randel N, Asadulina A, Bezares-Calderón LA, Verasztó C, Williams EA, Conzelmann M, et al. Neuronal connectome of a sensory-motor circuit for visual navigation. Elife. 2014;3:e02730.

32. Tosches MA, Bucher D, Vopalensky P, Arendt D. Melatonin signaling controls circadian swimming behavior in marine zooplankton. Cell. 2014;159:46–57.
33. Veedin-Rajan VB, Fischer RM, Raible F, Tessmar-Raible K. Conditional and specific cell ablation in the marine annelid *Platynereis dumerilii*. PLoS One. 2013;8:e75811.
34. Conzelmann M, Williams EA, Krug K, Franz-Wachtel M, Macek B, Jékely G. The neuropeptide complement of the marine annelid *Platynereis dumerilii*. BMC Genomics. 2013;14:906. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3890597&tool=pmcentrez&rendertype=abstract
35. Conzelmann M, Offenburger S-L, Asadulina A, Keller T, Munch TA, Jekely G. Neuropeptides regulate swimming depth of Platynereis larvae. Proc Natl Acad Sci [Internet]. 2011;108:E1174–83. Available from: http://www.pnas.org/cgi/doi/10.1073/pnas.1109085108
36. Fischer A, Dorresteijn A. The polychaete Platynereis dumerilii (Annelida): a laboratory animal with spiralian cleavage, lifelong segment proliferation and a mixed benthic/pelagic life cycle. BioEssays. 2004;26:314–25.
37. Fischer AH, Henrich T, Arendt D. The normal development of *Platynereis dumerilii* (Nereididae, Annelida). Front Zool. 2010;7:31. Available from: http://www.frontiersinzoology.com/content/7/1/31
38. O'Connell J, Schulz-Trieglaff O, Carlson E, Hims MM, Gormley NA, Cox AJ. NxTrim: optimized trimming of Illumina mate pair reads. Bioinformatics. 2015;31:2035–7.
39. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics. 2014;30:2114–20.
40. Martin M. Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet.journal. 2011;17:10. Available from: http://journal.embnet.org/index.php/embnetjournal/article/view/200/479
41. Schubert M, Lindgreen S, Orlando L. AdapterRemoval v2: rapid adapter trimming, identification, and read merging. BMC Res Notes. 2016;9:88.
42. Haas BJ, Papanicolaou A, Yassour M, Grabherr M, Blood PD, Bowden J, et al. De novo transcript sequence reconstruction from RNA-seq using the trinity platform for reference generation and analysis. Nat Protoc [Internet]. 2013;8:1494–512. Available from: https://doi.org/10.1038/nprot.2013.084
43. Chevreux B, Pfisterer T, Drescher B, Driesel AJ, Müller WEG, Wetter T, et al. Using the miraEST assembler for reliable and automated mRNA transcript assembly and SNP detection in sequenced ESTs. Genome Res. 2004;14:1147–59. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=419793&tool=pmcentrez&rendertype=abstract
44. Li B, Fillmore N, Bai Y, Collins M, Thomson JA, Stewart R, et al. Evaluation of de novo transcriptome assemblies from RNA-Seq data. Genome Biol. 2014;15:553. Available from: http://genomebiology.biomedcentral.com/articles/10.1186/s13059-014-0553-5
45. Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics. 2010;26:139–40. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2796818&tool=pmcentrez&rendertype=abstract
46. Trapnell C, Pachter L, Salzberg SL. TopHat: discovering splice junctions with RNA-Seq. Bioinformatics. 2009;25:1105–11.
47. Wang L, Feng Z, Wang X, Wang X, Zhang X. DEGseq: an R package for identifying differentially expressed genes from RNA-seq data. Bioinformatics. 2009;26:136–8.
48. Codd EF. A relational model of data for large shared data banks. Commun ACM. 1970;26:64–9.
49. Kitchen SA, Crowder CM, Poole AZ, Weis VM, Meyer E. De Novo Assembly and Characterization of Four Anthozoan (Phylum Cnidaria) Transcriptomes. G3 (Bethesda). 2015;5:2441–52. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4632063/.
50. Attrill H, Falls K, Goodman JL, Millburn GH, Antonazzo G, Rey AJ, et al. FlyBase: establishing a Gene Group resource for Drosophila melanogaster. Nucleic Acids Res [Internet]. 2016;44:D786–92. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4702782/.
51. Cameron RA, Samanta M, Yuan A, He D, Davidson E. SpBase: the sea urchin genome database and web site. Nucleic Acids Res. 2009;37:D750–4.
52. Curwen V, Eyras E, Andrews TD, Clarke L, Mongin E, Searle SMJ, et al. The Ensembl automatic gene annotation system. Genome Res. 2004;14:942–50.
53. Okuda S, Yamada T, Hamajima M, Itoh M, Katayama T, Bork P, et al. KEGG atlas mapping for global analysis of metabolic pathways. Nucleic Acids Res. 2008;36:W423–6.
54. Levin M, Anavy L, Cole AG, Winter E, Mostov N, Khair S, et al. The mid-developmental transition and the evolution of animal body plans. Nature [Internet]. 2016;531:637–41. Available from: http://www.nature.com/doifinder/10.1038/nature16994
55. Chou H-C, Pruitt MM, Bastin BR, Schneider SQ. A transcriptional blueprint for a spiral-cleaving embryo. BMC Genomics. 2016;17:552.
56. Tu Q, Cameron AR, Davidson EH. Quantitative developmental transcriptomes of the sea urchin Strongylocentrotus purpuratus. Developmental Biology. 2014; 385(2):160–167.
57. Tan MH, Au KF, Yablonovitch AL, Wills AE, Chuang J, Baker JC, Wong WH, Li JB. RNA sequencing reveals a diverse and dynamic repertoire of the Xenopus tropicalis transcriptome over development. Genome Research. 2013;23(1):201–216.
58. Harvey SA, Sealy I, Kettleborough R, Fenyes F, White R, Stemple D, Smith JC. Identification of the zebrafish maternal and paternal transcriptomes. Development. 2013;140(13):2703-2710.
59. Wang J, Garrey J, Davis RE. Transcription in Pronuclei and One- to Four-Cell Embryos Drives Early Development in a Nematode. Current Biology. 2014; 24(2):124–133.
60. Helm R, Siebert S, Tulin S, Smith J, Dunn C. Characterization of differential transcript abundance through time during Nematostella vectensis development. BMC Genomics. 2013;14(1):266.
61. Gifford CA, Ziller MJ, Gu H, Trapnell C, Donaghey J, Tsankov A, Shalek AK, Kelley DR, Shishkin AA, Issner R, Zhang X, Coyne M, Fostel JL, Holmes L, Meldrim J, Guttman M, Epstein C, Park H, Kohlbacher O, Rinn J, Gnirke A, Lander ES, Bernstein BE, Meissner A. Transcriptional and Epigenetic Dynamics during Specification of Human Embryonic Stem Cells. Cell. 2013; 153(5):1149-1163.
62. Tulin S, Aguiar D, Istrail S, Smith J. A quantitative reference transcriptome for *Nematostella vectensis* early embryonic development: a pipeline for de novo assembly in emerging model systems. Evodevo. 2013;4:16. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3748831&tool=pmcentrez&rendertype=abstract
63. Jung E, Veuthey AL, Gasteiger E, Bairoch A. Annotation of glycoproteins in the SWISS-PROT database. Proteomics. 2001;1:262–8.
64. Yu DS, Lee DH, Kim SK, Lee CH, Song JY, Kong EB, et al. Algorithm for predicting functionally equivalent proteins from BLAST and HMMER searches. J Microbiol Biotechnol. 2012;22:1054–8.
65. Petersen TN, Brunak S, von Heijne G, Nielsen H. SignalP 4.0: discriminating signal peptides from transmembrane regions. Nat Methods. 2011;8:785–6. Available from: http://dx.doi.org/10.1038/nmeth.1701%5Cn http://www.nature.com/doifinder/10.1038/nmeth.1701
66. Krogh A, Larsson B, von Heijne G, Sonnhammer E. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. J Mol Biol. 2001;305:567–80.
67. Langmead B, Trapnell C, Pop M, Salzberg SL. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol. 2009;10:R25.
68. Li B, Dewey CN, Wang Z, Gerstein M, Snyder M, Katz Y, et al. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. BMC Bioinformatics [Internet]. BioMed Central. 2011;12:323. Available from: http://bmcbioinformatics.biomedcentral.com/articles/10.1186/1471-2105-12-323
69. Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. BMC Bioinformatics. 2008;9:559.
70. Li L, Stoeckert CJJ, Roos DS. OrthoMCL: Identification of Ortholog Groups for Eukaryotic Genomes -- Li et al. 13 (9): 2178 -- Genome Research. Genome Res. 2003;13:2178–2189. Available from: http://genome.cshlp.org/cgi/content/full/13/9/2178
71. Nordberg H, Cantor M, Dusheyko S, Hua S, Poliakov A, Shabalov I, et al. The genome portal of the Department of Energy Joint Genome Institute: 2014 updates. Nucleic Acids Res. 2014;42:D26–31.
72. Zhang G, Fang X, Guo X, Li L, Luo R, Xu F, et al. The oyster genome reveals stress adaptation and complexity of shell formation. Nature [Internet]. 2012; 490:49–54. Available from: https://doi.org/10.1038/nature11413
73. Flicek P, Amode MR, Barrell D, Beal K, Billis K, Brent S, et al. Ensembl 2014. Nucleic Acids Res. 2014;42(Database issue):D749–55. https://doi.org/10.1093/nar/gkt1196.
74. Stamatakis A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics. 2014;30:1312–3.
75. Thompson JD, Gibson TJ, Higgins DG. Multiple sequence alignment using ClustalW and ClustalX. Curr Protoc Bioinformatics. 00: 2.3.1-2.3.22. Available from: https://doi.org/10.1002/0471250953.bi0203s00.