

PdumBase

The *Platynereis dumerilii* Database

Spiralian embryogenesis at your finger tips

An introductory tutorial into its content and features.

Image: Immunofluorescent photomicrograph of 49-cell stage *Platynereis dumerilii* embryo. Animal pole view. Stain shows β -catenin accumulated in nuclei (red) and microtubules (green).

The Schneider lab *Platynereis dumerilii* online database: PdumBase, provides a comprehensive, versatile online tool to investigate stage specific transcriptional inputs during embryogenesis and during the life cycle of the annelid *Platynereis dumerilii* and other selected species (e.g. *Danio rerio*, *Xenopus tropicalis*, *Nematostella vectensis*, *Strongylocentrotus purpuratus*).

This document provides a brief description of the database content and a detailed guide on how to browse its data thorough exemplary searches. The tutorial is intended as a motivational introduction while exploring and trying out the features PdumBase has to offer as an online resource to integrate and visualize our data and findings.

Database Content

In the following, the database content as well as its structure is explained. First, the details of the raw RNA-Seq data sets are highlighted, followed by an introduction into their corresponding expression data and associated annotation profiles. Furthermore the gene expression profiling features of this software are introduced, followed by an introduction into *Platynereis* specific coexpression networks as well as their comparative transcriptome data.

1. RNA-Seq Data sets

Platynereis dumerilii Normal Development Data Set

The Normal Development Data Set is comprised of two sources of information each of which corresponding to different stages of development *Platynereis dumerilii* at specific time points.

- **Early stages data set:** RNA-seq data generated by Schneider lab

Description: This data corresponds to the first comprehensive transcriptome draft during early development in *Platynereis dumerilii* using the de novo assembly strategy. We performed mRNA deep sequencing of distinct stages using the Illumina HiSeq sequencing system with read lengths of 75bp to 100bp.

Time points: 0, 2, 4, 6, 8, 10, 12, 14 hours post fertilization (hpf). Each stage has two biological replicates. The depth of these libraries ranges from 40 to 120 million paired-end reads (see Table 1).

Table 1: Time points from Early Stages data set

Time (hpf)	Description	Time (hpf)	Description
0	Unfertilized egg	8	~ 80-cell
2	Zygote	10	~ 140-cell
4	~ 8-cell	12	~ 220-cell
6	~ 30-cell	14	~ 330-cell

Assembly: All the biological replicates, which contain about 1.5 billion reads, were assembled into 357,961 transcripts in a genome independent manner. Due to alternative splicing events, out of the total transcripts assembled, 193,310 belong to genes.

Time points: This data set consist of 10 time points from 24 hours post fertilization to 3 months old adults. This set also includes female and male RNA-seq samples. There are no biological replicates (Table 2).

Table 2: Late Stages included in data set. Time points are shown in hours post fertilization (hpf), days (d) and months (M).

Time	Description
24 (hpf)	Early trochophore larvae
36 (hpf)	Mid trochophore larvae
48 (hpf)	Early metatrochophore larvae
72 (hpf)	Early nectochaete larvae
4 d	Mid nectochaete larvae
10 d	Errant juvenile
15 d	3-segmented errant juvenile
1 Mpre	
1 Mpost	
3 M	Adult
Male	Sexually mature adult
Female	Sexually mature adult

2. Expression data

The *Platynereis dumerilii* database web interface displays the mean FPKM (fragments per kilobase per million reads mapped) as the default measurement of gene expression. The FPKM for each replicate was obtained by normalizing the total number of mappable reads with the corresponding transcript length. A transcript or gene is considered as expressed if its FPKM is ≥ 1 . Furthermore, the FPKM for each stage was obtained by combining the replicates into a single set.

The result search page displays the mean FPKM values as the default measurement of gene expression (see Figure 1). However, FPKM values from individual samples, as well as the raw counts of each transcript can also be retrieved by clicking on the "Expression data" tab after selecting a particular transcript of interest (Figure 2). For more information we refer the reader to the *Tutorial Example* Section.

Platynereis dumerilii

Dr. Schneider's lab

Total genes: 40			Early stages								Azakenpaulone treatment					
Gene ID	Protein Name	Manual annotation	0H	2H	4H	6H	8H	10H	12H	14H	A06	C06	A12	C12	A06/C06	A12/C12
comp226107_c0	Cytochrome P450 3A4	<input type="text"/> Update	7.9	2.58	1.88	3.7	6.25	5.38	5.27	9.7	2.7	3.34	6.74	7.22	0.81	0.93
comp206017_c0	Dimethyl sulfoxide reductase DmsA	<input type="text"/> Update	0.41	0.07	0.04	0.03	0	0	0.02	0.01	0.02	0	0.01	0.01	0	1
comp206017_c1	Dimethyl sulfoxide reductase DmsA	<input type="text"/> Update	0.37	0.1	0.07	0.05	0.01	0.01	0	0.01	0.01	0.01	0	0.01	1	0
comp221418_c0	Forkhead box protein A2	<input type="text"/> Update	0	0	0	3.6	46.86	16.3	10.54	10.95	1.65	4.88	25.75	8.89	0.34	2.9
comp223177_c0	Forkhead box protein A4-A	<input type="text"/> Update	4.49	0.98	2.65	10.59	44.92	14.14	10.04	8.86	2.14	24.37	4.76	20.47	0.09	0.23

Figure 1: *Platynereis dumerilii* web database. Search result interface displays mean FPKM as measurement of absolute expression

3. Annotation

This section is concerned with describing the different annotations, how these were sourced from external databases for convenient browsing and data exploration specific to *Platynereis dumerilii*.

Gene: comp224014_c0

Plot Expression data Annotation

Show later stages

Pooled:

Type	ID	Early stages								Azakempauflone treatment					
		0-HR	2-HR	4-HR	6-HR	8-HR	10-HR	12-HR	14-HR	A06	C06	A12	C12	A06/C06	A12/C12
FPKM	comp224014_c0	0.74	6.18	11.79	24.42	42.18	56.22	43.98	24.91	26.27	27.09	87	45.59	0.97	1.91
Raw count	comp224014_c0	38	514.29	1766.32	4093.46	6456.45	10216.08	7570.51	5400.53	3076.33	4041.73	19517.91	11667.71	0.76	1.67
FPKM	comp224014_c0_seq12	0.07	0.04	0.1	0.22	2.54	0.23	0.22	0.16	0.13	0.08	4.89	1.31	1.63	3.73
FPKM	comp224014_c0_seq14	0	0	0.8	3.68	2.58	0.92	0.73	0.03	7.7	5.49	1.77	0.58	1.4	3.05
FPKM	comp224014_c0_seq16	0.29	2.76	4.16	2.77	1.48	2.33	2.61	0.75	5.21	4.05	1.19	2.3	1.29	0.52
FPKM	comp224014_c0_seq4	0	0.07	1.41	9.24	19.75	27.45	20.71	12.14	11.06	13.07	56.4	28.14	0.85	2
FPKM	comp224014_c0_seq7	0.34	0	0.34	0.2	0.13	0.06	0.1	0.17	0.44	0.09	0.11	0.05	4.89	2.2
Raw count	comp224014_c0_seq12	2	2.03	8.77	21.43	221.44	23.44	21.26	19.77	6.89	5.12	489.37	149.24	1.35	3.28
Raw count	comp224014_c0_seq14	0	0	32.76	170.75	108.79	45.78	34.03	1.76	191.61	176.74	85.61	32.07	1.08	2.67
Raw count	comp224014_c0_seq16	31	501.26	1321.2	997	485.16	900.65	949.64	354.98	1005.93	1012.08	446.53	986.47	0.99	0.45
Raw count	comp224014_c0_seq4	0	11	388.59	2894.29	5635.07	9243.2	6560.58	5013.03	1859.9	2844.79	18490.39	10496.94	0.65	1.76
Raw count	comp224014_c0_seq7	5	0	15	10	6	3	5	11	12	3	6	3	4	2

(a)

Replicate:

Type	ID	Early stages																						
		0-HR		2-HR		4-HR		6-HR		8-HR		10-HR		12-HR		14-HR								
		SS01	SS02	SS21	SS22	SS41	SS41T	SS42	SS42T	SS61	SS61T	SS62	SS62T	SS81	SS81T	SS82	SS82T	SS101	SS102	SS121	SS122	SS141	SS141T	SS142
FPKM	comp224014_c0	1.31	0.36	4.58	12.05	17.09	17.05	7.36	7.56	20.66	19.78	27.96	26.25	32.4	30.89	40.15	39.67	77.1	56.8	55.35	46.27	22.66	22.21	22.05
Raw Count	comp224014_c0	34	4	144.87	369.04	639	634.01	239	253.99	876.68	874.53	1175.19	1167.5	1357.11	1204.36	2055.55	1839.65	6031.42	4184.75	4083.9	3488.31	1432.64	1386	1318.65
FPKM	comp224014_c0_seq12	0.13	0	0.11	0	0	0.1	0.26	0.05	0.3	0.26	0.25	0.06	0.13	0.07	4.01	3.8	1.26	0.07	0.12	1.03	0.08	0	0.41
FPKM	comp224014_c0_seq14	0	0	0	0	1.8	1.31	0	0	4.43	4.53	2.41	2.66	0.11	0.88	3.34	3.77	1.47	0.5	0.05	1.47	0	0	0.03
FPKM	comp224014_c0_seq16	0.45	0.14	1.82	5.06	5.52	5.21	2.76	2.89	2.91	2.67	2.34	2.51	0.97	1	1.47	1.3	1.76	3.2	2.48	2.9	0.59	0.54	0.67
FPKM	comp224014_c0_seq4	0	0	0.11	0.06	2.1	2.42	0.42	0.45	6.55	6.32	11.63	10.4	15.52	14.27	17.52	17.33	35.31	24.75	24.79	19.39	10.62	10.61	10.17
FPKM	comp224014_c0_seq7	0.6	0	0	0	0.26	0.34	0.39	0.38	0.23	0.36	0.16	0	0.08	0	0.13	0.21	0.12	0	0.13	0.08	0.2	0.05	0.1
Raw Count	comp224014_c0_seq12	2	0	2.12	0	0	2.29	5	1	7.57	6.74	6.05	1.56	3.19	1.67	120.28	103.87	61.18	3.3	5.43	46.96	3.15	0	14.99
Raw Count	comp224014_c0_seq14	0	0	0	0	19.49	14.14	0	0	53.81	57.45	28.28	33.48	1.27	10.09	48.54	49.9	34.65	10.94	1.08	32.61	0	0	0.48
Raw Count	comp224014_c0_seq16	27	4	135.75	365.04	462.99	436.71	203	218.99	274.12	263.26	213.95	245.74	90.81	88.4	165.29	133.13	321.09	538.69	419.5	499.02	84.85	77	92.01
Raw Count	comp224014_c0_seq4	0	0	7	4	153.52	176.87	27	30	538.19	542.08	924.92	886.71	1260.84	1104.2	1719.44	1549.75	5611.5	3631.84	3654.89	2907.72	1340.64	1308	1209.17
Raw Count	comp224014_c0_seq7	5	0	0	0	3	4	4	4	3	5	2	0	1	0	2	3	3	0	3	2	4	1	2

(b)

Figure 2: *Platynereis dumerilii* web database. **Expression data tab interface:** (a) The upper frame displays mean FPKM and raw counts data, from samples as a pool. (b) Lower frame displays expression data from individual replicas.

Uniprot annotation

In our *Platynereis dumerilii* database the search results interface retrieves the Uniprot annotation data, displaying the Uniprot accession number, gene name, protein name, the species of annotation origin, and the E-value (see Figure 3). The annotation was performed using BLASTP by aligning the transcripts with predicted open reading frames (ORF) against non-redundant SwissProt databases. A total 31,806 transcripts (17,213 genes) retrieved at least one hit using an E-value cutoff of 10^{-10} . Among the annotated transcripts, 26% aligned to human and 19% to mouse proteins.

Pfam Annotation

We also annotated for potential protein domains by aligning all transcripts against the Pfam database. The Pfam annotation can be accessed in the database web interface by selecting the option "Show detailed annotation" on the search results page, or by clicking on the tab "Annotation" after having selected a particular transcript from the result interface (see Figure 4). Annotation was performed using HMMER. We were able to assign Pfam domains to 32,464 transcripts (18,146 genes), identifying a total of 431,701 Pfam domains. Furthermore, out of the transcripts with domain annotations, 28,326 (15,690) were also present in the Uniprot BLASTP annotation.

Platynereis dumerilii
Dr. Schneider's lab

HOME BLAST DOWNLOAD SEARCH GEROWSE RESOURCES LAB WIKI FTP OLD VERSION ABOUT US

Comparative information

Manual annotation	Early stages							Azakenpaullone treatment						Uniprot annotation					
	0H	2H	4H	6H	8H	10H	12H	14H	A06	C06	A12	C12	A06/C06	A12/C12	Accession	Gene Name	Protein Name	Species	E-value
<input type="text"/> Update	7.9	2.58	1.88	3.7	6.25	5.38	5.27	9.7	2.7	3.34	6.74	7.22	0.81	0.93	P08684	CYP3A4_HUMAN	Cytochrome P450 3A4	Homo sapiens	2e-48
<input type="text"/> Update	0.41	0.07	0.04	0.03	0	0	0.02	0.01	0.02	0	0.01	0	0	1	E18775	DMSA_ECOLI	Dimethyl sulfoxide reductase DmsA	Escherichia coli	4e-59
<input type="text"/> Update	0.37	0.1	0.07	0.05	0.01	0.01	0	0.01	0.01	0	0.01	1	0	0	E18775	DMSA_ECOLI	Dimethyl sulfoxide reductase DmsA	Escherichia coli	4e-128
<input type="text"/> Update	0	0	0	3.6	46.86	16.3	10.54	10.95	1.65	4.88	25.75	8.89	0.34	2.9	Q17181	FOXK2_XENTR	Forkhead box protein A2	Xenopus tropicalis	9e-85
<input type="text"/> Update	4.49	0.98	2.65	10.59	44.92	14.14	10.04	8.86	2.14	24.37	4.76	20.47	0.09	0.23	E33205	FXA4_XENLA	Forkhead box protein A4.A	Xenopus laevis	1e-49
<input type="text"/> Update	0	0.02	0.05	0.02	0.03	0.04	1.08	5.76	0.01	0.04	0.05	0.05	0.25	1	Q64732	FOXK1_MOUSE	Forkhead box protein B1	Mus musculus	2e-65
<input type="text"/> Update	0.03	0.04	7.09	29.62	31.04	1.4	0.24	0.15	49.23	56.51	4.27	0.58	0.87	7.36	Q99853	FOXK1_HUMAN	Forkhead box protein B1	Homo sapiens	8e-19

Figure 3: *Platynereis dumerilii* web database. **Search results interface** displays Uniprot annotation data on the rightmost panel. Annotation data includes accession number, gene name, protein name, species and E-value. Clicking on the accession number will redirect to the UniProt page for that particular protein.

Platynereis dumerilii
Dr. Schneider's lab

HOME BLAST DOWNLOAD SEARCH GEROWSE RESOURCES LAB WIKI FTP

Back CSV file Excel file Protein sequence file Comparative information

Show detailed annotation
 Show plots
 Show later stages
 Show other info

Total genes: 1			Early stages							Azakenpaullone treatment						
Gene ID	Protein Name	Manual annotation	0H	2H	4H	6H	8H	10H	12H	14H	A06	C06	A12	C12	A06/C06	A12/C12
comp224014_c0	Forkhead box protein J1	<input type="text"/> Update	0.74	6.18	11.79	24.42	42.18	56.22	43.98	24.91	26.27	27.09	87	45.59	0.97	1.91

(a)

Gene ontology		KEGG	eggNOG	Protein domain		
Cellular component	Molecular function	Biological process		Pfam	SignalP	TmHMM
	<ul style="list-style-type: none"> DNA binding, binding double-stranded DNA binding protein domain specific binding transcription factor binding transcription factor binding transcription regulatory region sequence-specific DNA binding 	<ul style="list-style-type: none"> actin cytoskeleton organization activation of Rho GTPase activity brain development central tolerance induction clathrin assembly establishment of apical/basal cell polarity glomerular parietal epithelial cell development heart looping humoral immune response light pattern formation leukocyte migration lung epithelium development metastasis; part of ureteric bud development negative regulation of B cell activation negative regulation of germinal center formation negative regulation of humoral immune response mediated by circulating immunoglobulin negative regulation of interleukin-6 biosynthetic process negative regulation of NF-kappaB transcription factor activity negative regulation of T cell differentiation in thymus negative regulation of T cell proliferation negative regulation of transcription from RNA polymerase II promoter positive regulation of central B cell tolerance induction positive regulation of lung ciliated cell differentiation positive regulation of transcription from RNA polymerase II promoter spermatogenesis 	<ul style="list-style-type: none"> Maturity onset diabetes of the young 	<ul style="list-style-type: none"> COG5026 	<ul style="list-style-type: none"> Fork head domain 	

(b)

Figure 4: *Platynereis dumerilii* web database. **Search results interface**. (a) The search result page allows to customize the information displayed by checking one or more options from the left top corner. (b) Selecting the option: "Show detailed annotation" will show detailed gene ontology, KEGG Pathways and protein domain annotation.

KEGG Pathways Annotation

Identifying the active biological pathways in early stages is crucial to decipher the mechanisms involved in the diversification of embryonic cells. The Kyoto Encyclopedia of Genes and Genomes (KEGG) provides well-annotated pathway databases including metabolism, genetic and cellular processing.

Our assembled transcripts were mapped to KEGG pathways. In total, 18,532 transcripts (10,132 genes) are associated with the known KEGG pathways.

In our database, the KEEG annotation is accessible by selecting the option **”Show detailed annotation”** as seen in Figure 4.

Gene Ontology Annotation

The assembled transcripts were also annotated with Gene ontology (GO) terms of homologous genes. A total of 30,287 transcripts (16,498 genes) could be associated with at least one annotated GO term. The GO annotation shows high enrichment in the function associated with transcription and regulation activities in the biological process and molecular functions. The GO terms related to cell differentiation such as “cell transduction”, “cell adhesion”, “cell division” and “cell cycle” are also enriched.

All annotation information for a given transcript is summarized and displayed in the annotation tab interface (see Figure 5).

It is worth pointing out that one important feature of our database is that the search interface allows for the submission of searches under Blast Info, Pfam, Gene Ontology, and Kegg Pathway, making it possible to narrow down a request by a particular annotation of interest.

Gene: comp224014_c0

Plot Expression data **Annotation**

There are 1 annotation(s)

Accession	Q92949
Protein name	Forkhead box protein J1
Manual annotation	
Protein name (short)	HFH-4
Gene name	FOXJ1_HUMAN
Species	Homo sapiens
Evalue	1e-55
GO: Cellular component	transcription factor complex
GO: Molecular function	DNA binding_bending double-stranded DNA binding protein domain specific binding RNA polymerase II distal enhancer sequence-specific DNA binding transcription factor activity transcription factor binding transcription regulatory region sequence-specific DNA binding
GO: Biological process	actin cytoskeleton organization activation of Rho GTPase activity brain development central tolerance induction cilium assembly establishment of apical/basal cell polarity glomerular parietal epithelial cell development heart looping humoral immune response left/right pattern formation leukocyte migration lung epithelium development metanephric part of ureteric bud development negative regulation of B cell activation negative regulation of germinal center formation negative regulation of humoral immune response mediated by circulating immunoglobulin negative regulation of interleukin-6 biosynthetic process negative regulation of NF-kappaB transcription factor activity negative regulation of T cell differentiation in thymus negative regulation of T cell proliferation negative regulation of transcription from RNA polymerase II promoter positive regulation of central B cell tolerance induction positive regulation of lung ciliated cell differentiation positive regulation of transcription from RNA polymerase II promoter spermatogenesis
KEGG	Maturity onset diabetes of the young
eggNOG	COG5025
Pfam	Fork head domain
SignalP	-
tmHMM	-

7

Figure 5: *Platynereis dumerilii* web database: **Annotation tab interface**. This tab is available once an entry from the result page has been selected. It is accessible via result page → clicking on a gene or transcript of interest → clicking on the Annotation tab

4. Gene Expression Profiling

Our *Platynereis dumerilii* database includes a detailed gene expression profiling of the early developmental stages (2 to 14hpf). An expression profile can be interpreted as the changes in the abundance of a transcript over time.

Plots depicting these fluctuations of transcript abundance (FPKM) are shown for each transcript. Accessing this data is available via the option **"Show Plots"** on the search result page (Figure 6), or by clicking on the transcript of interest and selecting the tab labeled **"Plots"** (Figure 7). For the purpose of the expression profiling analysis we filtered out low expression transcripts.

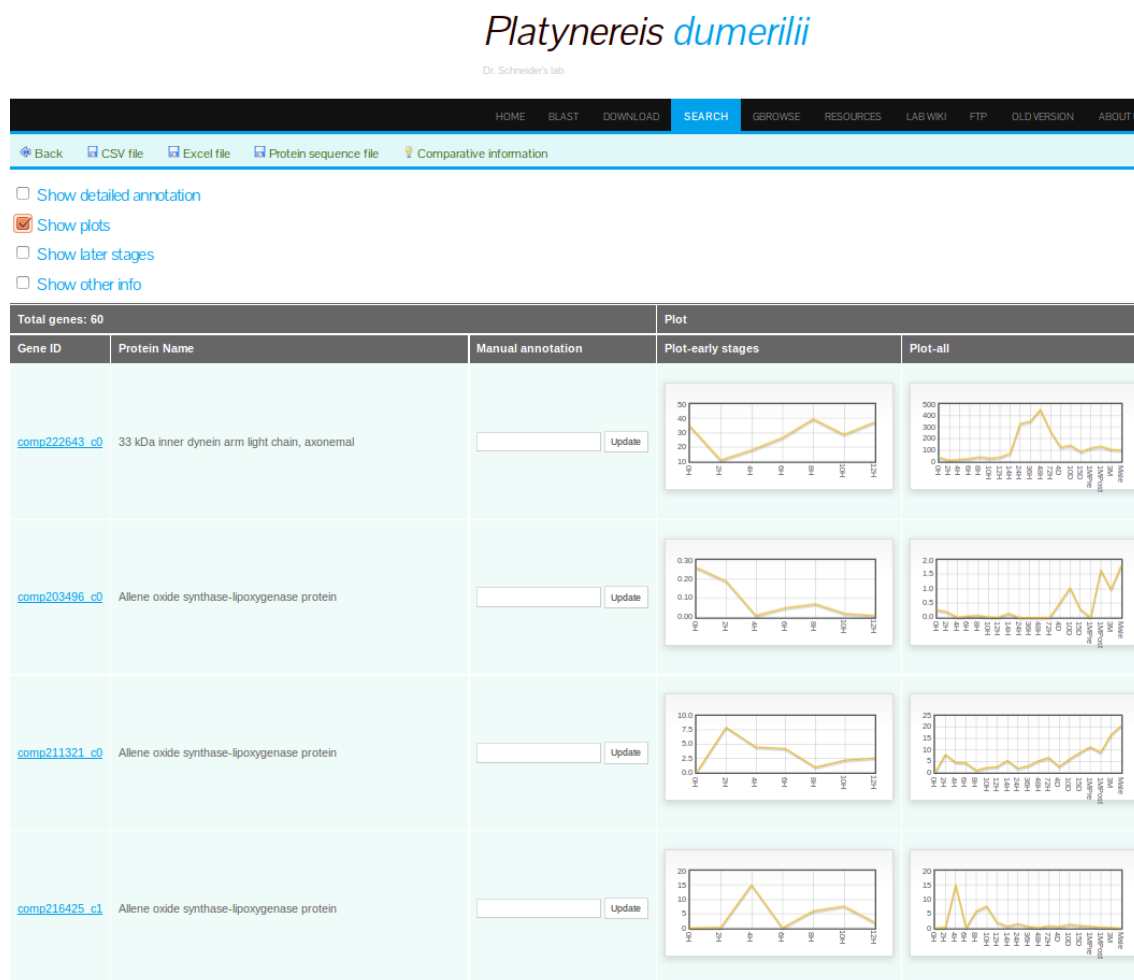


Figure 6: *Platynereis dumerilii* web database: **Search result interface**. Expression profile plots are displayed when the option **"Show plots"** is selected.

Among the assembled transcripts with predicted ORF, 18,940 transcripts and 13,160 genes were found to be expressed in at least one of the 7 stages. After clustering the genes according to their expression profile, we found a total of 15 distinct clusters (see Figure 8).

Clusters 1-4 show the obvious maternal signature with a total of 4,302 genes belonging this group. The clusters 10-15 (5827 genes) correspond to the zygotic genes with slightly different activation time points. The clusters 3 and 11 are the major maternal and zygotic groups respectively showing slow decreased and increased expression patterns. The 6th, 7th, and 8th cluster contain a set of genes whose RNAs were mainly expressed at 4, 6 and 8 hours and degraded after these stages. The 9th cluster is a less dynamic group, showing stable expression throughout all stages.

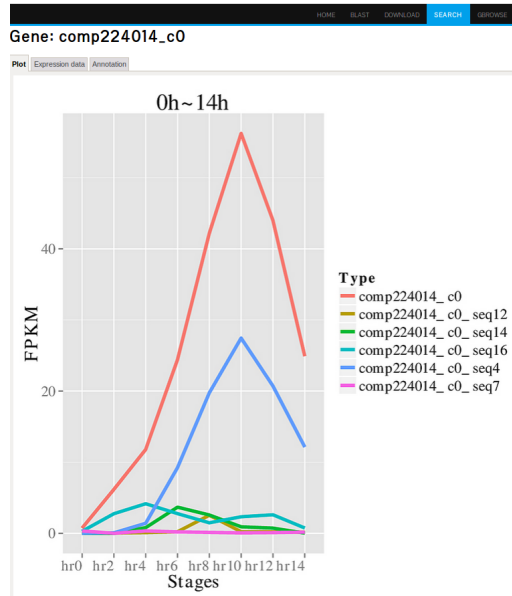


Figure 7: *Platynereis dumerilii* web database: **Plot tab interface**. Shows the expression profile plot for a given transcript.

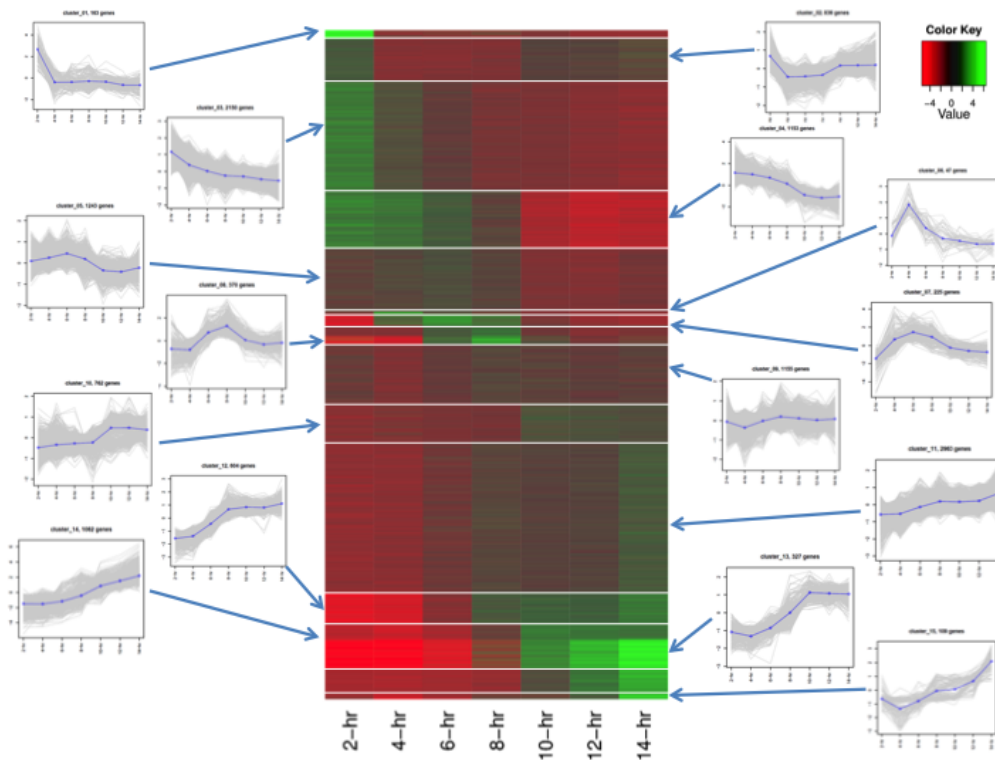


Figure 8: Heat map of 13,160 expressed genes clustered into 15 groups according to the time series patterns.

Access to the cluster information is available by selecting the option **”Show other info”** on the result search page and then, for a selected transcript/gene, clicking on the icon under **”Coexpression info”**. The first tab of the new results page will display all the genes in the same cluster, along with other expression data (see Figure 9).

Coexpression network information: comp224014_c0

The same cluster All Azakenpaulone treatment DE genes

Show later stages

Rank	Gene	Protein name	Cluster	Correlation	Topology overlap
	comp224014_c0	Forkhead box protein J1	cluster_11	1	1
1	comp216113_c4	Transmembrane cell adhesion receptor mua-3	cluster_11	0.76	0.26
	comp216113_c4	Collagen alpha-1(XII) chain	cluster_11	0.76	0.26
2	comp224875_c1	Sperm-associated antigen 1	cluster_11	0.61	0.23
3	comp220840_c0	Dual specificity protein phosphatase 12	cluster_11	0.59	0.25
4	comp216986_c0	Sperm-associated antigen 1	cluster_11	0.57	0.24
5	comp218535_c1	Putative ribosome-binding factor A, mitochondrial	cluster_11	0.56	0.21
6	comp212960_c0	Protein piccolo	cluster_11	0.48	0.21
7	comp222815_c0	Serologically defined colon cancer antigen 3 homolog	cluster_11	0.48	0.19
8	comp201839_c0	Zinc finger and BTB domain-containing protein 16-A	cluster_11	0.42	0.23
	comp201839_c0	GDNF-inducible zinc finger protein 1	cluster_11	0.42	0.23
9	comp223496_c0	STAM-binding protein-like A	cluster_11	0.37	0.13
10	comp220743_c1	Sushi, von Willebrand factor type A, EGF and pentraxin domain-containing protein 1	cluster_11	0.35	0.19
	comp220743_c1	Uncharacterized protein R617	cluster_11	0.35	0.19
11	comp224866_c0	E3 ubiquitin-protein ligase AMFR	cluster_11	0.35	0.20
12	comp218288_c3	Alpha,alpha-trehalose-phosphate synthase [UDP-forming] A	cluster_11	0.33	0.21
	comp218288_c3	Alpha,alpha-trehalose-phosphate synthase [UDP-forming] 1	cluster_11	0.33	0.21
13	comp219901_c0	Ubiquitin-conjugating enzyme E2 G1	cluster_11	0.33	0.23

Figure 9: *Platynereis dumerilii* web database: **Coexpression information interface**. Displays all the transcripts/genes in the same cluster of a given component, shows protein name, correlation and topology overlap.

5. Coexpression Networks

A coexpression network is a correlation network that describes the pairwise correlation patterns of expression data. When a set of genes are highly correlated, they may share similar biological function or be involved in the same biological pathway. A coexpression network can also be used for identifying hub genes which have high connectivity to other genes in a cluster. We used weighted correlation network analysis (WGCNA) to analyze *Platynereis dumerilii* expression profiling data.

For this analysis, we included a total of 13,192 genes whose FPKM was ≥ 1 for at least one sample. Correlation values and topology overlap for the coexpression networks can be found in the database on the **Coexpression information interface**. This page can be reached from the search result interface by selecting the option **"Show other info"** and by clicking on the icon under the column **"Coexpression info"** in the results table. The **Coexpression information interface** is shown in Figure 9.

6. Comparative Transcriptome Data

Ortholog Expression

With the aim of identifying conserved stages of development, we gathered publicly available expression data from five species for which we then identified orthologs w.r.t. *Platynereis dumerilii* (see Tables 3 and 4) and proceeded to establish global comparison expression profiles among the ortholog groups.

The ortholog expression data for a particular *Platynereis dumerilii* transcript, can be found in our database by selecting the option **"Show other info"** and by clicking on the icon under the column *Ortholog Expressions* for the specific transcript of interest. The resulting interface will display the ID number and expression data for the orthologs found for that transcript/gene in the other 5 species (see Figure 10).

[Platynereis_dumerilii](#)

SS00	SS02	SS04	SS06	SS08	SS10	SS12	SS14	H24	H36	H48	H72	D4	D10	D15	M1Pre	M1Post	M3	Male	Female	A06	C06	A12	C12	A06/C06	A12/C12	Protein_name
2.4	58.2	17.94	16.65	16.62	13.94	9.99	8.96	9.54	18.04	13.48	17.15	15.95	15.55	12.92	12.4	11.64	8.34	9.33	32.08	17.74	20.17	17	12.54	0.88	1.36	Transcription factor RFX3

[Ascaris_suum](#)

Gene_id	oocyte	zygote1	zygote2	zygote3	zygote4	24h	46h	64h	96h	116h	7D	L1	L2	Protein_name
ASU_01592	0.5	1.1	2	4.4	6	9	8.4	8.4	7.1	13.1	59.3	20.5	12.6	Abnormal dauer formation protein 19

[Danio_reio](#)

Gene_id	2_cell	64_cell	3.5hpf	6hpf	9hpf	Protein_name
ENSDARG00000014550	3.5353	3.97681	5.66213	2.09896	1.26579	Transcription factor RFX3

[Xenopus_tropicalis](#)

Gene_id	cell2	cell4	cell8	cell16	Stage6	Stage8	Stage9	Stage10	Stage11_12	Stage13_14	Stage15	Stage16_18	Stage19	Stage20_21	Stage22_23	Stage24_26	Stage28	Stage31_32
ENSKETG00000005567	17.2043	19.0805	21.6319	18.7191	18.3343	18.1596	17.5738	12.3135	6.93579	11.2533	16.3181	17.9518	15.4797	11.0376	11.0063	10.407	9.7551	13.9335

[Homo_sapiens](#)

Gene_id	HUES64	Endoderm	Ectoderm	Mesoderm	Hepatoblast like	Protein_name
RFX1	5.42569	5.0263	6.21082	5.92104	12.4656	

[Nematostella_vectensis](#)

Gene_id	2h	7h	12h	24h	5d	10d	Protein_name
e_gw43.110.1	166.029	115.501	134.056	73.32	50.1426	43.2087	Regulatory factor X 3

[Strongylocentrotus_purpuratus](#)

Gene_id	0h	10h	18h	24h	30h	40h	48h	56h	64h	72h	Protein_name
WHL22.19679	89.0913	83.5982	69.7351	28.2143	27.0699	30.8884	38.166	33.884	36.9231	44.9459	Regulatory factor X 3

Figure 10: *Platynereis dumerilii* web database: **Ortholog expression profile interface**. Displays the expression data from the selected *Platynereis dumerilii* gene and the orthologs genes found in the other species along with their expression and annotation data (when available).

Table 3: Species and number of protein sequences for comparative analysis

Species	Number of sequences
<i>Platynereis dumerilii</i>	28,580
<i>Danio rerio</i>	26,241
<i>Xenopus tropicalis</i>	18,442
<i>Homo sapiens</i>	23,393
<i>Nematostella vectensis</i>	27,273
<i>Ascaris suum</i>	15,446

Table 4: Number of orthologs genes between the 6 species

Species	<i>Platynereis dumerilii</i>	<i>Danio rerio</i>	<i>Xenopus tropicalis</i>	<i>Homo sapiens</i>	<i>Nematostella vectensis</i>	<i>Ascaris suum</i>
<i>Platynereis dumerilii</i>		5635	5402	5051	5840	3654
<i>Danio rerio</i>			10784	10246	6731	4307
<i>Xenopus tropicalis</i>				10284	6415	4140
<i>Homo sapiens</i>					6094	3941
<i>Nematostella vectensis</i>						4245
<i>Ascaris suum</i>						

Ortholog Groups

We also identified orthologs genes for 18 selected species (Table 5) using the program OrthoMCL. This program runs all versus all Blastp queries among all the protein sequences from these 18 species and selects the best reciprocal blast hits. Once the orthologs genes were identified, phylogenetic trees were assembled using RaxML.

Table 5: Species and number of genes used to find orthologs groups

Class	Code	Species	Number of genes
Lophotrochozoa	pdu	<i>Platynereis dumerilii</i>	28,580
Lophotrochozoa	cte	<i>Capitella teleta</i>	32,415
Lophotrochozoa	hro	<i>Helobdella robusta</i>	23,423
Lophotrochozoa	lgi	<i>Lottia gigantea</i>	23,851
Lophotrochozoa	cgi	<i>Crassostrea gigas</i>	26,089
Ecdysozoa	dpu	<i>Daphnia pulex</i>	30,907
Ecdysozoa	tca	<i>Tribolium castaneum</i>	16,524
Ecdysozoa	dme	<i>Drosophila melanogaster</i>	13,937
Deuterostomia	spu	<i>Strongylocentrotus purpuratus</i>	20,759
Deuterostomia	sko	<i>Saccoglossus kowalevskii</i>	34,239
Deuterostomia	bfo	<i>Branchiostoma floridae</i>	50,817
Deuterostomia	dre	<i>Danio rerio</i>	26,459
Deuterostomia	xtr	<i>Xenopus tropicalis</i>	18,442
Deuterostomia	hsa	<i>Homo sapiens</i>	23,393
Prebilateria	nve	<i>Nematostella vectensis</i>	27,273
Prebilateria	aqu	<i>Amphimedon queenslandica</i>	29,883
Prebilateria	tad	<i>Trichoplax adhaerens</i>	11,520
Preanimalia	mbr	<i>Monosiga brevicollis</i>	9,196

To access the ortholog genes for a given *Platynereis dumerilii* transcript/gene, select the option "show other info". If ortholog groups are found for that particular transcript, a check-mark will appear under the field "Ortholog groups". Clicking on this icon will open a new interface with four tabs: "List", "Tree-ML", "Tree-Parsimony", and Alignment (see Figures 11, 12, and 13 respectively).

Platynereis dumerilii

Dr. Schneider's lab

HOME
BLAST
DOWNLOAD
SEARCH
GBrowse
RESOURCES
LAB WIKI
FTP
OLD VERSION
ABOUT US

Gene: comp224014_c0
Group: orth5613

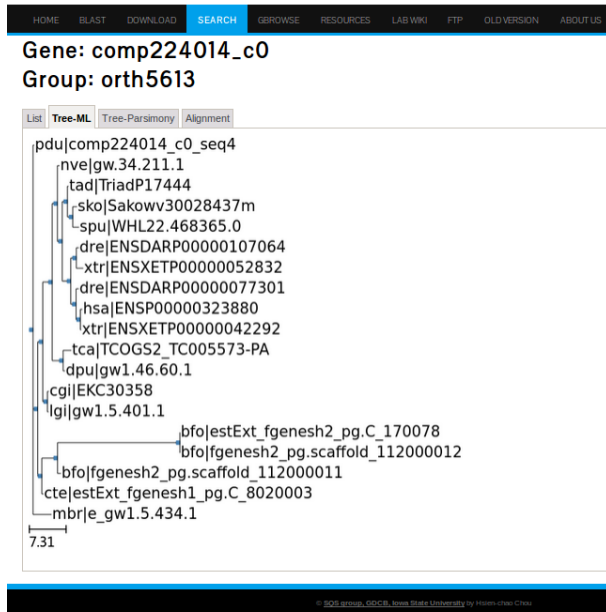
List
Tree-ML
Tree-Parsimony
Alignment

Species	Name	Class	Protein ID	Protein Sequence	cDNA Sequence
pdu	Platynereis dumerilii	Protostomia	comp224014_c0_seq4	↗	↗
cte	Capitella teleta	Protostomia	estExt_fgenes1_pg.C_8020003	↗	↗
lgi	Lottia gigantea	Protostomia	gw1.5.401.1	↗	↗
cgi	Crassostrea gigas	Protostomia	EKC30358	↗	↗
dpu	Daphnia pulex	Protostomia	gw1.46.60.1	↗	↗
tca	Tribolium castaneum	Protostomia	TCOGS2:TC005573-PA	↗	↗
spu	Strongylocentrotus purpuratus	Deuterostomia	WHL22.468365.0	↗	↗
sko	Saccoglossus kowalevskii	Deuterostomia	Sakowv30028437m	↗	↗
bfo	Branchiostoma floridae	Deuterostomia	estExt_fgenes2_pg.C_170078	↗	↗
bfo	Branchiostoma floridae	Deuterostomia	fgenes2_pg.scaffold_112000011	↗	↗
bfo	Branchiostoma floridae	Deuterostomia	fgenes2_pg.scaffold_112000012	↗	↗
dre	Danio rerio	Deuterostomia	ENSDARP00000107064	↗	↗
dre	Danio rerio	Deuterostomia	ENSDARP00000077301	↗	↗
xtr	Xenopus tropicalis	Deuterostomia	ENSXETP00000042292	↗	↗
xtr	Xenopus tropicalis	Deuterostomia	ENSXETP00000052832	↗	↗
hsa	Homo sapiens	Deuterostomia	ENSP00000323880	↗	↗
nve	Nematostella vectensis	Nonbilateria	gw.34.211.1	↗	↗
tad	Trichoplax adherens	Nonbilateria	TriadP17444	↗	↗
mbr	Monosiga brevicollis	Nonbilateria	e_gw1.5.434.1	↗	↗

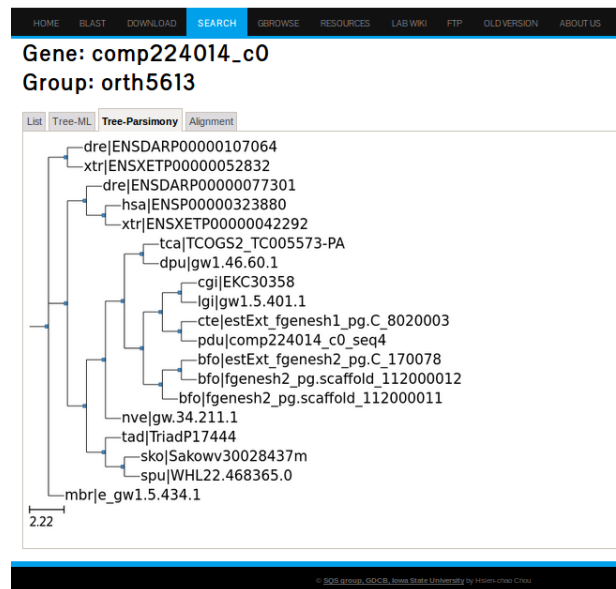
© SDS group, GDCB, Iowa State University by Helen-Chao Chou

13

Figure 11: *Platynereis dumerilii* web database: **List tab interface** under Ortholog groups. Shows the species list, code, name, ortholog protein ID and contains links to access/download the protein and cDNA sequences in Fasta format.



(a)



(b)

Figure 12: *Platynereis dumerilii* web database. **Ortholog groups interface:** (a) Phylogenetic tree among ortholog genes displayed under **Tree-ML** tab (b) Phylogenetic tree displayed under **Tree-Parsimony** tab. Both trees show the species code and the transcript/gene ID.

```
mbr|e_gw1.5.434.1
xtr|ENSXETP00000052832
dre|ENSDARP00000107064
nve|gw.34.211.1
xtr|ENSXETP00000042292
hsa|ENSP00000323880
dre|ENSDARP00000077301
dpu|gw1.46.60.1
tca|TCOGS2_TC005573-PA
spu|WHL22.468365.0
tad|TriadP17444
lgi|gw1.5.401.1
pdu|comp224014_c0_seq4
cte|estExt_fgenesH1_pg.C_802003
cgi|EKC30358
sko|Sakowv30028437m
bfo|estExt_fgenesH2_pg.C_170078
bfo|fgenesH2_pg.scaffold_11200012
bfo|fgenesH2_pg.scaffold_11200011
```

```
LIFMAMQQADKPKLALSEIYDFIVNNFAWYRMADPGWKNSIRHNLSQLKA
LICMAMEASQQRKLTLSAIYSWITQNFCCYFRHADPSWQNSIRHNLSQLKC
LICMAMQASNKTITLSAIYSWITENFCYFRHADPSWQNSIRHNLSQLKC
LICMAMRDTKRKVTITLSAIYKWIENFMFYRVADPTWQNSIRHNLSQLKC
LICMAMQASKKTKITLSAIYKWITDNFCYFRHADPTWQNSIRHNLSQLKC
LICMAMQASKATKITLSAIYKWITDNFCYFRHADPTWQNSIRHNLSQLKC
LICMAMQASKKTKITLSAIYKWITDNFCYFRHADPTWQNSIRHNLSQLKC
LICMAMKS-NKHKMTLSSYKWIENFLYYRNVDPSPWQNSIRHNLSQLKC
LICMAMGK-NGNKMTLSAIYHWIRENFLYYRKAHPSWQNSIRHNLSQLKC
LIWMAMKESKHKITLSSYKWIENFKYYQVADPSWQNSIRHNLSQLKC
LICMAMKESKSKITLSSAIYNWIRENFMYYRIADPSWQNSIRHNLSQLKC
LIGMAMKETQKQKITLSAIYNWITDNFMYYRMADPSWQNSIRHNLSQLKC
LICMAMKETNRNKITLSGIYSWITENYMYRVADPSWQNSIRHNLSQLKC
LICMAMKESKKNKVTLSGIYNWITENFMYYRMADPSWQNSIRHNLSQLKC
---MAMKETKSKITLSAIYNWITDNFMYYRLADPSWQNSIRHNLSQLKC
LICMAMKETKKNKITLSAIYKWIQDNFMYYKVAEPSWQNSIRHNLSQLKC
IDQVLEPSESHLNESTSNILNTNLPPSPSSEGYRHPWEEASELDSIVDMN
IDQVLEPSESHLNESTSNILNTNLPPSPSSEGYRHPWEEASELDSIVDMN
LICMAMKETKSKITLSDIYKIKTNFKYYEMAEPSWQISRQLYSSNKR-
: . : : : : : . . * : : .
```

Figure 13: *Platynereis dumerilii* web database: **Alignment tab interface** under Ortholog groups. Displays CLUSTAL 2.1 multiple sequence alignment.

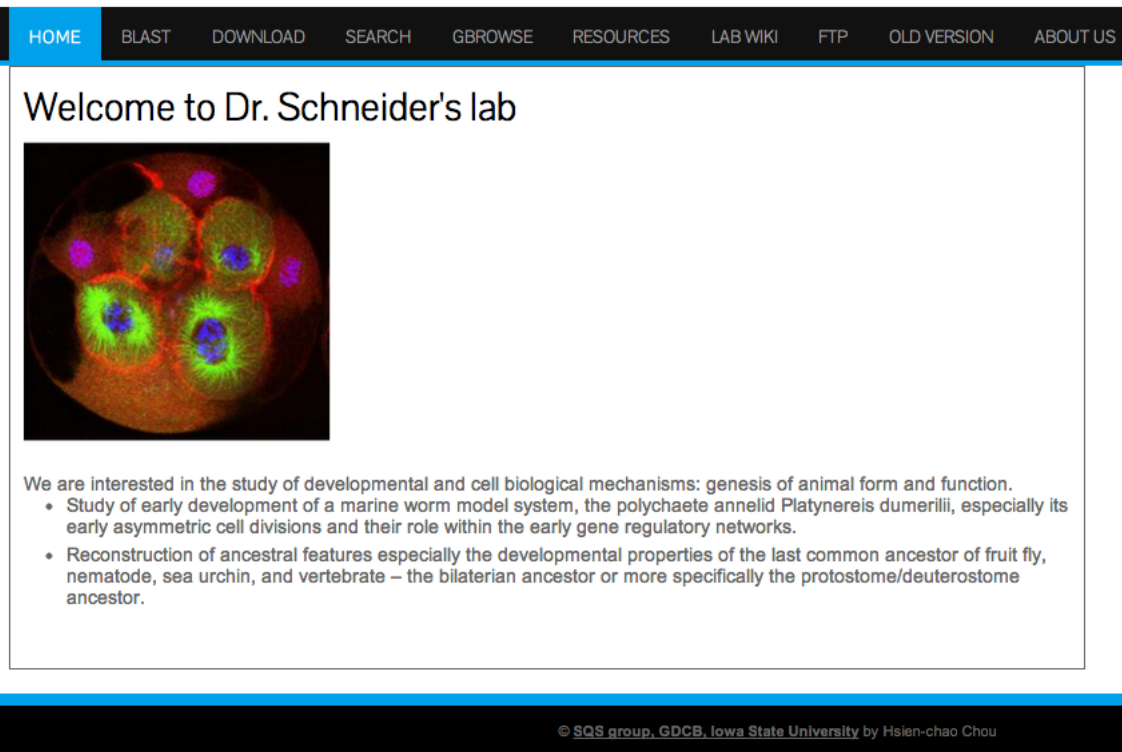
Accessing and browsing the *Platynereis dumerilii* web database

Once connected to Iowa State University server via VPN, access to the *Platynereis dumerilii* web database is provided under the following URL: <http://sqlslab.gdcb.iastate.edu/>. Login and password information is currently required.

When logged in, the system will automatically redirect to the the database web home interface (Figure 14).

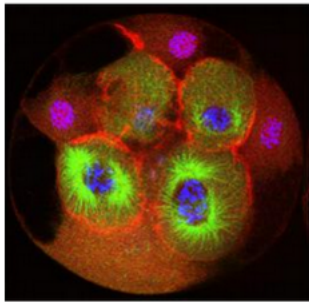
Platynereis dumerilii

Dr. Schneider's lab



HOME BLAST DOWNLOAD SEARCH GBROWSE RESOURCES LAB WIKI FTP OLD VERSION ABOUT US

Welcome to Dr. Schneider's lab



We are interested in the study of developmental and cell biological mechanisms: genesis of animal form and function.

- Study of early development of a marine worm model system, the polychaete annelid *Platynereis dumerilii*, especially its early asymmetric cell divisions and their role within the early gene regulatory networks.
- Reconstruction of ancestral features especially the developmental properties of the last common ancestor of fruit fly, nematode, sea urchin, and vertebrate – the bilaterian ancestor or more specifically the protostome/deuterostome ancestor.

© SQS group, GDCB, Iowa State University by Hsien-chao Chou

Figure 14: *Platynereis dumerilii* web database: **Home interface**

Tutorial Example: Searching under Blast Information

This section will show some of the *Platynereis dumerilii* web database features through exemplary searches using the blast info search function.

Search

The search interface allows to submit searches under different criteria: Blast, Pfam, SignalIP, TmHMM, EggNog, Gene Ontology, and KEEG Pathway (Figure 15). By searching under different or combined fields, the search can be customized according to the user needs.

Platynereis dumerilii

Dr. Schneider's lab

HOME BLAST DOWNLOAD **SEARCH** GBROWSE RESOURCES LAB WIKI FTP OLD VERSION ABOUT US

[Search Platynereis](#)

Blast Info:

Pfam Info:

SignalP Info:

TmHMM Info:

Eggnog Info:

Gene Ontology:

Kegg Pathway:

Sort by: Protein Name Desc

Search

[Search other species](#)

[Search Uniprot genes by GO terms](#)

© SQS group, GDCB, Iowa State University by Hsien-chao Chou

Figure 15: *Platynereis dumerilii* web database: **Search interface**.

In addition, the search interface offers the option of selecting a sorting criteria to retrieve the results according to the expression values from any stage (0 to 14hpf) (Figure 16). This feature can be particularly convenient when searching with terms that might result in a multitude of hits such as "cell cycle" which retrieves more than 1000 genes, or "membrane" with around 500 hits. Therefore, searching for general terms might result in a request which could take more than 60 seconds to load. Please allow time for those general searches to load.

On the other hand, when searching for a particular gene name, for instance the transcription factor **FoxA2** in the field Blast Info, the most likely outcome will be one single hit displaying the *Platynereis dumerilii* transcript/gene with that particular annotation.

HOME BLAST DOWNLOAD **SEARCH** GBROWSE RESOURCES LAB WIKI FTP OLD VERSION

[Search Platynereis](#)

Blast Info:

Pfam Info:

SignalP Info:

TmHMM Info:

Eggnog Info:

Gene Ontology:

Kegg Pathway:

Sort by: Desc

[Search other species](#)

[Search Uniprot genes by GO terms](#)

© SQS group, GDCB, Iowa State University by Hsien-chao Chou

Figure 16: *Platynereis dumerilii* web database: **Search interface**. Searching for FoxA2

Search Results

The resulting search results interface displays by default the transcript or gene model ID, protein name, expression data as mean FPKM from early stages (0 to 14 hpf), expression data in inhibitor experiment, and annotation information (Figure 17).

In addition, the results interface allows to expand the results displayed by selecting from the options on the left upper corner. The user can select one or more options according to his/her particular research needs (see also Expanded search result options Section).

Access to Detailed Information

Clicking on the gene model for FoxA2 "comp21414" will give access to the detailed data results interface. The detailed data result page has three tabs: Plot, Expression data and Annotation, from which different information can be accessed.

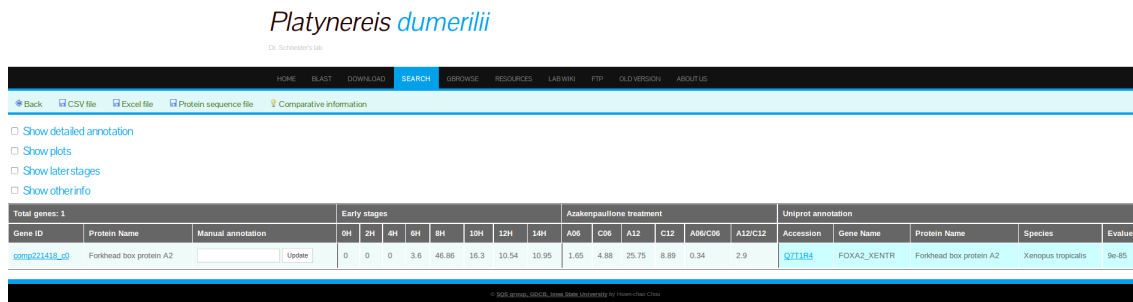


Figure 17: *Platynereis dumerilii* web database: **Search result interface**. shows Gene ID, expression data from early stages and from inhibitor experiment, and annotation information. The data retrieve options are found on the left upper corner.

The Plot Tab

Clicking on the Plot tab will display expression profile data (FPKM values against stages) for early and late states (Figure 18).

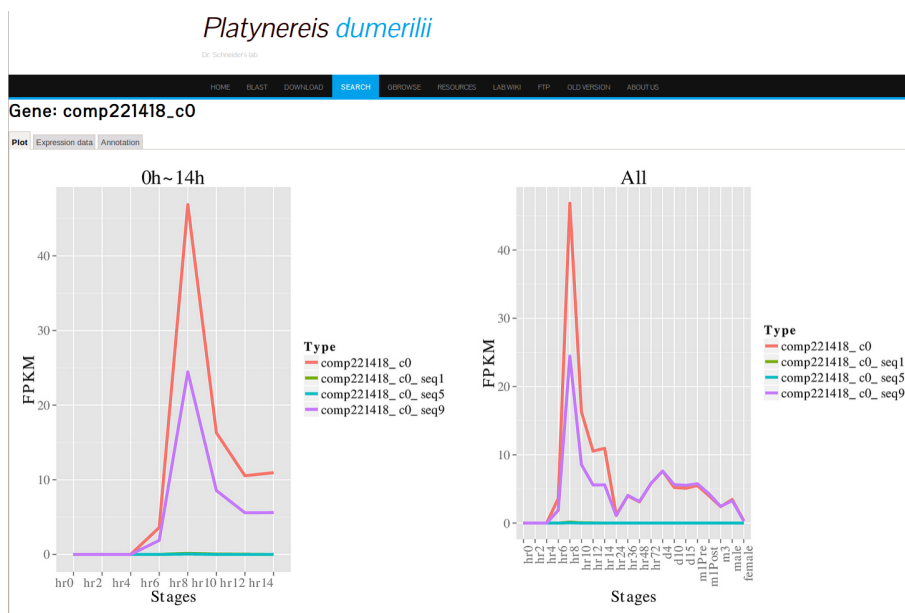


Figure 18: *Platynereis dumerilii* web database: **Plot tab** from Detailed data results interface. Displaying expression profile plots for FoxA2.

The Expression Data Tab

The Expression data tab will show mean and individual sample FPKM values as well as raw counts (Figure 19).

The Annotation Tab

Clicking the annotation tab will retrieve a summary of all annotation related information including: Species from which the annotation was obtained, GO extended annotation, KEGG pathways, EggNog, and Pfam domains (see figure 20).

Gene: **comp221418_c0**

PKM: Expression data Administration

Show later stages

Pooled:

Type	ID	Early stages								
		0-HR	2-HR	4-HR	6-HR	8-HR	10-HR	12-HR	14-HR	
FPKM	comp221418_c0	0	0	0	3.6	46.86	16.3	10.54	10.95	
Raw count	comp221418_c0	0	0	1	1204.49	14313.01	9809.29	3619.71	4737.13	
FPKM	comp221418_c0_seg1	0	0	0	0	0.14	0.05	0.03	0	
FPKM	comp221418_c0_seg5	0	0	0	0	0.04	0	0	0.01	
FPKM	comp221418_c0_seg8	0	0	0	1.88	24.46	8.56	5.98	5.6	
Raw count	comp221418_c0_seg1	0	0	0	0	25.12	9.96	5.04	1	
Raw count	comp221418_c0_seg5	0	0	0	0.05	6.41	0	0	2.01	
Raw count	comp221418_c0_seg8	0	0	1	1203.84	14281.49	9899.34	3614.66	4734.12	

(a)

Azakenpaullone treatment

A06	C06	A12	C12	A06/C06	A12/C12
1.65	4.88	25.75	8.89	0.34	2.9
575.57	2165.94	17180.8	6767.92	0.27	2.54
0	0.03	0.16	0.03	0	5.33
0	0.01	0.01	0.01	0	1
1.67	4.85	25.56	8.85	0.34	2.89
0	4.14	32.5	7.73	0	4.2
0	1.04	1.23	1.39	0	0.88
575.57	2160.77	17147.08	6758.79	0.27	2.54

(b)

Replicate:

Type	ID	Early stages																						
		0HR	2HR	4HR	6HR	8HR	10HR	12HR	14HR	0HR	2HR	4HR	6HR	8HR	10HR	12HR	14HR							
FPKM	comp221418_c0	0	0	0	0	0.81	0	0	2.06	2.1	6.27	6.62	36.62	33.87	60.62	60.73	21.2	15.59	10.96	10.96	9.52	6.86	6.17	
Raw Count	comp221418_c0	0	0	0	0	1	0	0	189	229	381.99	428.87	2094	2095.43	4483.17	4627.98	2627.86	2992.84	2992.99	2757.99	1822.07	1293.09	1127	1029
FPKM	comp221418_c0_seg1	0	0	0	0	0	0	0	0	0	0	0	0.027	0.11	0.26	0.24	0.06	0.02	0.03	0.02	0	0	0	0
FPKM	comp221418_c0_seg5	0	0	0	0	0	0	0	0	0	0	0	0.026	0.02	0	0.026	0	0	0.02	0	0	0	0	0
FPKM	comp221418_c0_seg8	0	0	0	0	0.81	0	0	1.52	1.23	2.26	2.4	18.82	17.7	29.86	29.82	7.82	4.23	3.86	5.1	5.01	4.63	4.36	
Raw Count	comp221418_c0_seg1	0	0	0	0	0	0	0	0	0	0	0	3.64	9.29	9.42	7.05	7.67	3.78	3.82	2.82	1	0	0	0
Raw Count	comp221418_c0_seg5	0	0	0	0	0	0	0	0	0	0	0	6.29	2.45	0	2.76	0	0	9.94	0	0	2.02	0	0
Raw Count	comp221418_c0_seg8	0	0	0	0	1	0	0	162	126	381.99	427.89	2077.81	2082.14	4399.86	4602.05	2625.26	2922.54	2922.94	2752.94	1822.09	1293.12	1127	1029

(c)

Azakenpaullone treatment

A06			C06			A12			C12		
A061	A062	A063	C061	C062	C063	A121	A122	A123	C121	C122	C123
0.37	4.33	0	4.39	8.82	1.47	25.07	23.79	29.01	8.33	8.96	9.82
49	526.42	0	689.42	1243	233.44	5702.93	4483.52	7000.3	2016.4	2175.19	2576.37
0	0	0	0	0.05	0.04	0.09	0.25	0.17	0.08	0	0.02
0	0	0	0	0.03	0	0	0	0.02	0.02	0	0.01
0.37	4.28	0	4.49	8.72	1.47	25.92	23.42	28.48	8.32	8.81	9.75
0	0	0	0	2.08	2.09	6.09	14.01	12.65	6.1	0	1.43
0	0	0	0	1.04	0	0	0	1.17	1.02	0	0.61
49	526.42	0	689.42	1239.87	231.34	5696.84	4469.51	6986.48	2009.28	2175.19	2574.33

(d)

Figure 19: *Platynereis dumerilii* web database: **Expression data tab** from Detailed data results interface. **(a)** Displays expression data (mean FPKM and raw counts) from pooled samples from early stages of normal development. **(b)** Displays expression data from inhibitor experiment from pooled samples. **(c)** and **(d)** show individual replicates expression data for early stages and inhibitor experiment respectively.

HOME BLAST DOWNLOAD **SEARCH** GBrowse RESOURCES LAB WIKI FTP OLD VERSION ABOUT US

Gene: comp221418_c0

Plot Expression data **Annotation**

There are 1 annotation(s)

Accession	Q7T1R4
Protein name	Forkhead box protein A2
Manual annotation	
Protein name (short)	FoxA2
Gene name	FOXA2_XENTR
Species	Xenopus tropicalis
Evalue	9e-85
GO: Cellular component	transcription factor complex
GO: Molecular function	<ul style="list-style-type: none"> 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
GO: Biological process	<ul style="list-style-type: none"> 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
KEGG	Maturity onset diabetes of the young
eggNOG	COG5025
Pfam	<ul style="list-style-type: none"> Forkhead N-terminal region Fork head domain HNF3 C-terminal domain
SignalP	.
tmHMM	

© SOS group, GDCB, Iowa State University by Hsien-chao Chou

Figure 20: *Platynereis dumerilii* web database: **Annotation** tab from Detailed data results interface. Displaying detailed annotation information for FoxA2.

Expanded Search Result Options

The search result default data output can be expanded by selecting the options provided in the search results interface (Figure 21).

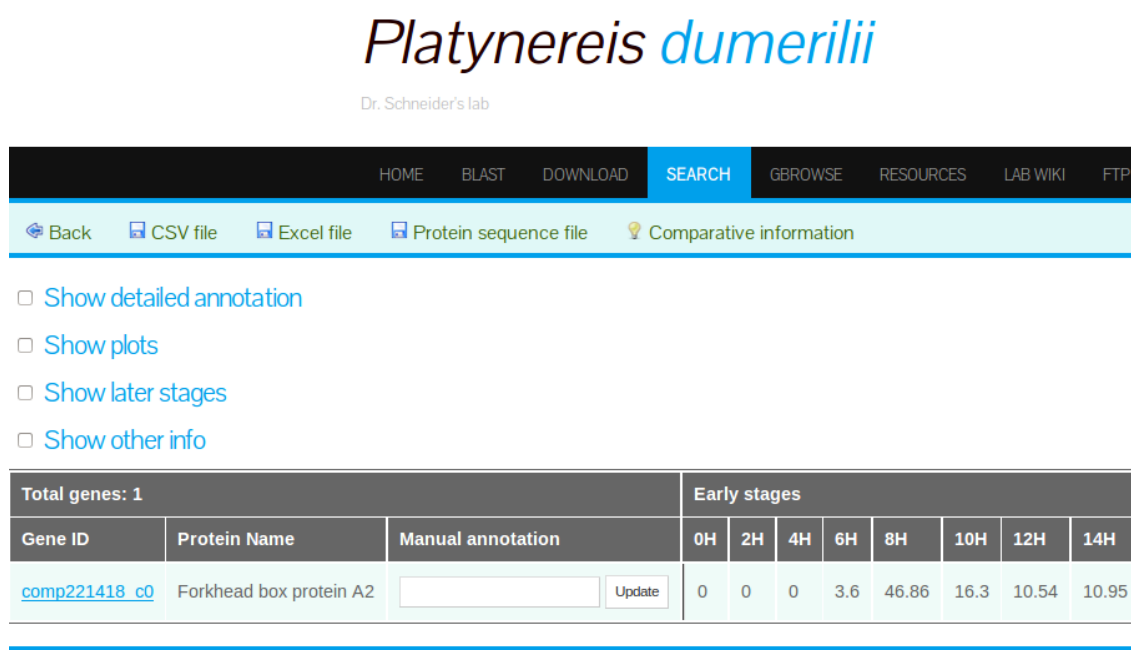


Figure 21: *Platynereis dumerilii* web database: **Search results interface** checking the boxes from the search result options on the left will expand the results displayed.

Selecting "Show plots"

Selecting the "show plots" option will retrieve a visual representation of early and late stage expression profile for all the Gene IDs displayed in the search result interface (Figure 22).

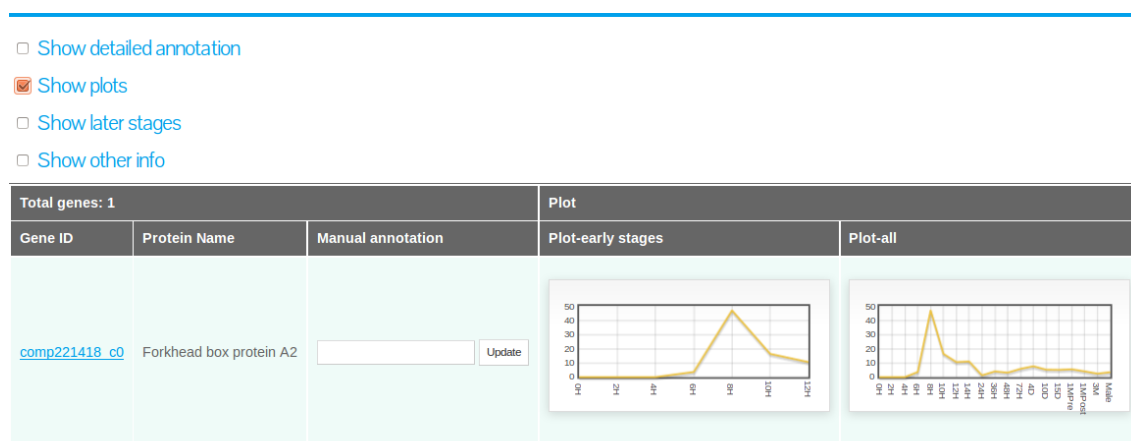


Figure 22: *Platynereis dumerilii* web database: **Search results interface** with the option "Show Plots" selected. Expression plots for both, early and late stages are shown for the gene under search: FoxA2.

Selecting "Show later stages"

To display the mean expression data (FPKM) from later stages of development (24hpf to 3M) it is required to select the option "show later stages" as shown in Figure 23.

Show detailed annotation
 Show plots
 Show later stages
 Show other info

Total genes: 1				Early stages							Later stages												
Gene ID	Protein Name	Manual annotation		0H	2H	4H	6H	8H	10H	12H	14H	24H	36H	48H	72H	4D	10D	15D	1MPre	1MPost	3M	Male	Female
comp221418_c0	Forkhead box protein A2	<input type="text"/> <input type="button" value="Update"/>		0	0	0	3.6	46.86	16.3	10.54	10.95	1.21	3.99	3.1	5.8	7.62	5.21	5.1	5.49	4	2.41	3.46	0.31


© SOS group, GDCB, Iowa State University by Hsien-chao Chou

Figure 23: *Platynereis dumerilii* web database: **Search results interface** with the option "Show later stages" selected. Here the later stages expression data from FOXA2 is displayed.

Selecting "Show other info"

Clicking "Show other info" provides access to additional data on comparative transcriptomics (see Figure 24):

Show detailed annotation
 Show plots
 Show later stages
 Show other info

Total genes: 1				Other info			
Gene ID	Protein Name	Manual annotation		Locus Info	Ortholog expressions	Ortholog groups	Coexpression Info
comp221418_c0	Forkhead box protein A2	<input type="text"/> <input type="button" value="Update"/>		✓	✓	✓	

© SOS group, GDCB, Iowa State University by Hsien-chao Chou

Figure 24: *Platynereis dumerilii* web database: **Search results interface** "Show other info" option selected. Additional information links are displayed.

- Locus Info - if available a green check-mark icon will be displayed.
- Ortholog Expressions - if available a green check-mark icon will be displayed.
- Ortholog groups - if available a green check-mark icon will be displayed.
- Coexpression info - if available a blue icon will be displayed.

It is important to mention that the additional data is not available to all the gene models but only to those transcripts for which orthologs genes were identified. See Table 4 for the estimated numbers of orthologs found.

Locus info link

Clicking on Locus info will show currently available genomic information. e.g. 'potential' promoter sequence upstream of the transcription start site. (See Figure 25).

Coexpression link

Selecting the coexpression link gives access to data about the expression profiling and coexpression. "The same cluster tab" from this interface displays the Gene ID of all genes belonging to the cluster of the gene under search (see Figure 26).

comp221418_c0_seq9-83307888 | comp221418_c0_seq9-CH189M20_supercontig_1_AlIBACs

Alignment

Source	Genome ID	Strand	Start	End	Promoter Start	Promoter End	Score	Coverage
Bac	83307888	+	35637	40861	33637	35636	0.985	1

Promoter sequence (length: 2000)

```
TATAGGTACCACAGGTATATCAGCCATTTTAATAGTTTTATGATTACTTTTCACCAITGAACTGCACCAATACCA
GTTAGCCTTCCGCCTACTTGGCTGATCTGGCCGTGTAAGCAGATCTTGAGATAAGGAAAGAAAGGCAAAA
AGAGAGTTCTCTTTCTAGCAAGCTCGCAACGAATCGCCCAAGAAAGAGATGAGACGAGACCTAAGTGGCTA
CCAAATGTTGAGGGGGCAATTTGATATATTTGCCCTCGAGGAGGGCCGTTTGGTGGGGTAATGAGGGTGT
TTCAAGGGGCGAGTACACTGGGGTGGCCAGCACTGACAATCAAAGGGGTATGAACCTGTATAGAGGGTGGAAAG
GGTAACTAGGATCTGTGATTGTGGAAACGGGGAGAGTCCCAAGTGGTGTACACATTTCAATATCTCAACAAAAGT
CTTAGACCGCAAGTTGAACTGACTATTGTGACCACCATGTACAATTTCTTATGAATGCTGTCATAGTCCCTGCAATG
CTGTCTGGCTACTACTCAAGTAAATAAATGAATTAATAATTTACAAGGATCTGAGAAAGCTTTCCTTGGATACAGT
TGATTTGATGGCCATCAGCAACTCAATCAAATGTTGATTTATAGAAGTATGATGCCCAAATAAGAAAGTCAACT
GAATTAGCATAGCAGGGGGATCACTTTTCTTGGATGTTTCTGACTGAATGCCCTTCAAATGCAAAAAGAGCACT
TTCTAAACAAAGAGTTTATAGATCATGATCTAACATTTCCAGCAAAAGAAATCACATGTTTTCATGAACAGCTCT
TTACCTGGGCAAGGGCATGCCATTTACTTGGCATGTAGCTCTATACAATTTGCTATCATTTACATTTTCATCACT
GAGGCTTAGGTGACATAAGCTTCTCTAAAATTTCTTAAATGCCATATTCAGGATAAATGATGTAAGTTTTGAC
ITAAAGTGTAAAGCTTCCCATCCGTCAACCAAGATTATGATGTTGTATAAAGTCAAGCTTTCATTTCACTGAAATA
CACTGTTCTACATTTTGGAACTTTTAGACTTGCATTGACAGAACTTAGCAGAAATATACATTTGAAATAACCTTG
CTTTTATGTTTTCACATTTACAATTTCAATGCTTGTCTTAACTCCTCAATAGGCACTGGGGCTCTGACTCAACCT
CCCTTTTCAAAAATCAATTAGCAATAACAATGATGAAGTATTTTTTATTAACTAGACAGCTGTATATAGAGATAT
AGTGTGACTTAATGTGGACTGCATAAAAATGACAGATAAAAATGCTAAATAGGCAAAAAGTGAAGGTAGATAAACCC
ATCTGTGGGGTGTATTTGGCTAGTAGTCAATTTGACTCACAATTTGAAATATCAGACATATCAGGGCATGTCAGAGGC
ATATCGAAGGATTTTGAACCTTTATCTCTGTCCAGATTTACATGATCTCCACAGCCACACTTATATAGATGTC
TTTAGGCTTATAGATTACAAGAAGTGTTAGACTTTCAACGGCTTGATTGAGCCAGTAATCACACATCCAGCTTG
ATTACAGCTTGTACTGGACCATTCACGCTGTAAACATGCTGATCCTCAAATAATGACCATAACTGGGCATGCAAAAT
TTAACTCTTGTAGATTATCTCTAATACTACTTATACCATATCTTAATTAATCACTGTATCATGTCACATTTGC
AAAATAATTTATTGCATTTTACTATGTAGACCTAATTTCCCTCAAAGTCTTTCCATGCTCATACCTGACACATT
TTCCACTGAGAGAGATCTTTATGCAATTACATATTTATGCATGGTGAATACTTTATCCCCTGTGAAATATCT
GTACCATTTCCATGCTTTAAACAGGG
```

Figure 25: *Platynereis dumerilii* web database: Search results interface "Show other info" option selected. The Locus info link displays genomic information

Platynereis dumerilii

Dr. Schneider's lab

HOME BLAST DOWNLOAD SEARCH GENOWISE RESOURCES LIB WIKI FTP OLD VERSION ABOUT US

Coexpression network information: comp221418_c0

The same cluster: All Azakenpaullone treatment DE genes

Show later stages

Rank	Gene	Protein name	Cluster	Correlation	Topology overlap	Early stages								Azakenpaullone treatment					
						0-HR	2-HR	4-HR	6-HR	8-HR	10-HR	12-HR	14-HR	A06	C06	A12	C12	A06/C06	A12/C12
	comp221418_c0	Forkhead box protein A2	cluster_12	1	1	0	0	0	3.6	46.86	16.3	10.54	10.95	1.65	4.88	25.75	8.89	0.34	2.9
1	comp221733_c0	Secreted frizzled-related protein 5	cluster_12	0.86	0.38	0.04	0	0.03	2.03	33.34	15.42	8.87	9.4	1.64	3.56	5.93	18.33	0.46	0.32
2	comp225143_c0	Hepatocyte nuclear factor 4-beta	cluster_12	0.75	0.36	0	0.38	0.08	5.64	42.2	22.15	12.06	11.77	2.19	6.98	47.57	13.75	0.31	3.46
3	comp223840_c0	Sodium- and chloride-dependent glycine transporter 2	cluster_12	0.61	0.30	0.19	1.33	1.35	6	61.99	34.97	20.08	23.38	0.78	5.54	21.2	21.36	0.14	0.99
	comp223840_c0	Sodium-dependent proline transporter	cluster_12	0.61	0.30	0.19	1.33	1.35	6	61.99	34.97	20.08	23.38	0.78	5.54	21.2	21.36	0.14	0.99
	comp223840_c0	Sodium- and chloride-dependent betaine transporter	cluster_12	0.61	0.30	0.19	1.33	1.35	6	61.99	34.97	20.08	23.38	0.78	5.54	21.2	21.36	0.14	0.99
4	comp221096_c0	Serine/threonine-protein kinase Kist	cluster_12	0.54	0.31	3.49	0.89	0.48	3.47	13.14	7.17	5.54	5.86	2.06	4.1	16.85	8.71	0.5	1.93
5	comp219534_c1	Synaptotagmin-1	cluster_12	0.52	0.28	0.44	0.46	1.1	1.8	32.94	12.38	9.49	16.14	1.44	5.86	53.04	58.42	0.25	0.91
6	comp225491_c0	Protein KHNYN	cluster_12	0.50	0.26	1.13	0.29	0.22	1.18	11.56	6.4	4.11	5.38	0.72	2.05	9.81	7.44	0.35	1.32
	comp225491_c0	Ribonuclease ZC3H12A	cluster_12	0.50	0.26	1.13	0.29	0.22	1.18	11.56	6.4	4.11	5.38	0.72	2.05	9.81	7.44	0.35	1.32
7	comp214240_c0	CCAAT/enhancer-binding protein beta	cluster_12	0.34	0.22	13.28	2.59	2.61	11.25	57.03	21.97	14.9	32.94	5.92	13.51	55.78	32.58	0.44	1.71

© 2019 Genou, GENO, Iowa State University & Helmholtz PSI

Figure 26: *Platynereis dumerilii* web database: Search results interface "Show other info" option selected. The Coexpression info link displays the list of genes clustered with the gene under study. The expression profile of FoxA2 clusters with 7 other genes belonging to cluster 7.

23

Orthologs groups link

Clicking the Orthologs groups link gives access to an interface with tree tabs: List, Tree-ML and Tree-Parsimony. As mentioned in the section "Comparative transcriptome data", 18 species were selected to assess the ortholog groups. The first tab shows the list of species from which orthologs were found for the searched gene. This interface also allows to download the protein and cDNA sequences of the orthologs in Fasta format (see Figure 27).

The second and third tab under the Orthologs groups link will display phylogenetic trees based on ML and parsimony analysis respectively. Figure 28 shows the tree-ML for the FoxA2 ortholog genes.

Gene: comp221418_c0

Group: orth4921

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		
dme	Drosophila melanogaster	Protostomia	FBpp0303365		
spu	Strongylocentrotus purpuratus	Deuterostomia	WHL22.439762.0		
sko	Saccoglossus kowalevskii	Deuterostomia	Sakowv30037646m		
bfo	Branchiostoma floridae	Deuterostomia	estExt_fgenes2_pg.C_420051		
bfo	Branchiostoma floridae	Deuterostomia	fgenes2_pm.scaffold_42000007		
dre	Danio rerio	Deuterostomia	ENSDARP00000002213		
dre	Danio rerio	Deuterostomia	ENSDARP00000007827		
xtr	Xenopus tropicalis	Deuterostomia	ENSXETP00000056642		
xtr	Xenopus tropicalis	Deuterostomia	ENSXETP00000026943		
hsa	Homo sapiens	Deuterostomia	ENSP00000250448		
hsa	Homo sapiens	Deuterostomia	ENSP00000400341		
nve	Nematostella vectensis	Nonbilateria	estExt_gwp.C_580130		
tad	Trichoplax adherens	Nonbilateria	TriadP27448		

Figure 27: *Platynereis dumerilii* web database: **Search results interface** "Show other info" option selected. The **Ortholog groups** link displays the list species where orthologs were found. For FoxA2, orthologs were found in all of the 18 selected species.

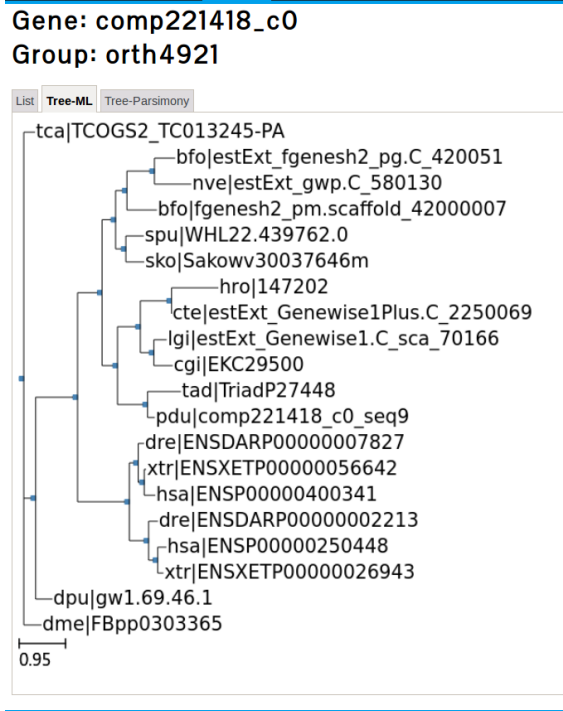


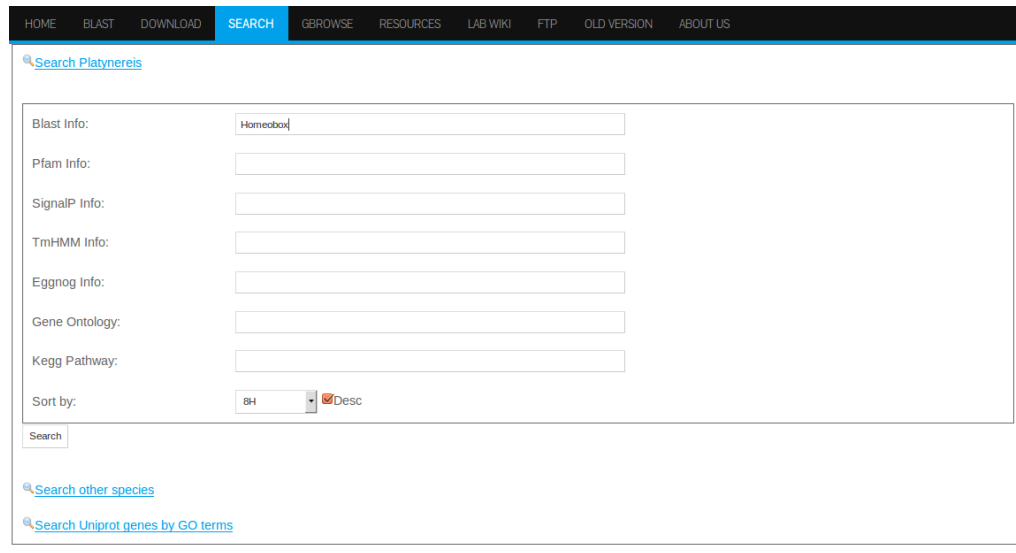
Figure 28: *Platynereis dumerilii* web database: **Search results interface** "Show other info" option selected. The **Ortholog groups** Tree-ML tab displays phylogenetic tree constructed with the ortholog protein sequences. Tree-ML for FoxA2 orthologs among the 18 species.

Example search for ”Homeobox genes”

This final example will show a sample search with multiple results, indicating the options that our web database offers to download the data in case further analysis is required.

Finding homeobox genes that are highest expressed at 8hpf

Searching for homeobox term in the blast field at the search interface will retrieve 114 hits. To find the highest expressed homeobox genes at 8 hpf, it is required to sort the hits by expression values at 8 hpf in descending order (See Figure 29).



The screenshot shows the search interface of the *Platynereis dumerilii* web database. The navigation bar includes links for HOME, BLAST, DOWNLOAD, SEARCH (highlighted), GBrowse, RESOURCES, LAB WIKI, FTP, OLD VERSION, and ABOUT US. Below the navigation bar, there is a search bar with the text "Search Platynereis". The main search area contains several input fields: Blast Info (containing "Homeobox"), Pfam Info, SignalP Info, TmHMM Info, Eggnog Info, Gene Ontology, and Kegg Pathway. Below these fields is a "Sort by:" dropdown menu set to "8H" and a checked "Desc" checkbox. A "Search" button is located at the bottom left of the search area. Below the search area, there are two additional search options: "Search other species" and "Search Uniprot genes by GO terms".

Figure 29: *Platynereis dumerilii* web database: **Search interface** search required to Find homeobox genes that are highest expressed at 8hpf.

Downloading results from *Platynereis dumerilii* web database

One important feature of our web database is that it allows to download the search results in different formats. The search results can be downloaded in both, comma separated value (CSV) format file and Excel file. Furthermore, the protein sequences from the genes displayed in the results can be downloaded in Fasta format. Links to download are found in the upper frame of the search result interface (see Figure 30).

HOME BLAST DOWNLOAD **SEARCH** GBROWSE RESOURCES

Back CSV file Excel file Protein sequence file Comparative information

Show detailed annotation
 Show plots
 Show later stages
 Show other info

Total genes: 116

Gene ID	Protein Name	Manual annotation
comp213046_c0	Homeobox protein otx5	<input type="text"/> Update
comp223956_c0	ALX homeobox protein 1	<input type="text"/> Update
comp220564_c0	Pre-B-cell leukemia transcription factor 1	<input type="text"/> Update
comp216317_c2	Homeobox protein Hmx	<input type="text"/> Update
comp218004_c3	Homeobox protein prophet of Pit-1	<input type="text"/> Update
comp220111_c0	Dorsal root ganglia homeobox protein	<input type="text"/> Update
comp223762_c1	Homeobox protein PKNOX2	<input type="text"/> Update
comp218070_c0	Homeobox protein OTX2-B	<input type="text"/> Update
comp217412_c0	ALX homeobox protein 1	<input type="text"/> Update
comp213198_c0	Homeobox protein TGIF2	<input type="text"/> Update

Figure 30: *Platynereis dumerilii* web database: **Search results interface**. Here the result page is displaying the ten top hits, sorted by expression level at 8 hpf. Links to download data are shown with a floppy disk icon and are found in the upper frame.

Concluding Remarks

Given the here presented features and ease of use that our *Platynereis dumerilii* database offers, we are confident that this work will provide a reliable resource to the community for transcriptome studies due to its extensive content and user friendly design.

List of Figures

1	<i>Platynereis dumerilii</i> web database. Search result interface displays mean FPKM as measurement of absolute expression	3
2	<i>Platynereis dumerilii</i> web database. Expression data tab interface: (a) The upper frame displays mean FPKM and raw counts data, from samples as a pool. (b) Lower frame displays expression data from individual replicas.	4
3	<i>Platynereis dumerilii</i> web database. Search results interface displays Uniprot annotation data on the rightmost panel. Annotation data includes accession number, gene name, protein name, species and E-value. Clicking on the accession number will redirect to the UniProt page for that particular protein.	5
4	<i>Platynereis dumerilii</i> web database. Search results interface. (a) The search result page allows to customize the information displayed by checking one or more options from the left top corner. (b) Selecting the option: "Show detailed annotation" will show detailed gene ontology, KEGG Pathways and protein domain annotation.	5
5	<i>Platynereis dumerilii</i> web database: Annotation tab interface. This tab is available once an entry from the result page has been selected. It is accessible via result page → clicking on a gene or transcript of interest → clicking on the Annotation tab	7
6	<i>Platynereis dumerilii</i> web database: Search result interface. Expression profile plots are displayed when the option "Show plots" is selected.	8
7	<i>Platynereis dumerilii</i> web database: Plot tab interface. Shows the expression profile plot for a given transcript.	9
8	Heat map of 13,160 expressed genes clustered into 15 groups according to the time series patterns.	9
9	<i>Platynereis dumerilii</i> web database: Coexpression information interface. Displays all the transcripts/genes in the same cluster of a given component, shows protein name, correlation and topology overlap.	10
10	<i>Platynereis dumerilii</i> web database: Ortholog expression profile interface. Displays the expression data from the selected <i>Platynereis dumerilii</i> gene and the orthologs genes found in the other species along with their expression and annotation data (when available).	11
11	<i>Platynereis dumerilii</i> web database: List tab interface under Ortholog groups. Shows the species list, code, name, ortholog protein ID and contains links to access/download the protein and cDNA sequences in Fasta format.	13
12	<i>Platynereis dumerilii</i> web database. Ortholog groups interface: (a) Phylogenetic tree among ortholog genes displayed under Tree-ML tab (b) Phylogenetic tree displayed under Tree-Parsimony tab. Both trees show the species code and the transcript/gene ID.	14

13	<i>Platynereis dumerilii</i> web database: Alignment tab interface under Ortholog groups. Displays CLUSTAL 2.1 multiple sequence alignment.	14
14	<i>Platynereis dumerilii</i> web database: Home interface	15
15	<i>Platynereis dumerilii</i> web database: Search interface	16
16	<i>Platynereis dumerilii</i> web database: Search interface . Searching for FoxA2 . . .	17
17	<i>Platynereis dumerilii</i> web database: Search result interface . shows Gene ID, expression data from early stages and from inhibitor experiment, and annotation information. The data retrieve options are found on the left upper corner.	18
18	<i>Platynereis dumerilii</i> web database: Plot tab from Detailed data results interface. Displaying expression profile plots for FoxA2.	18
19	<i>Platynereis dumerilii</i> web database: Expression data tab from Detailed data results interface. (a) Displays expression data (mean FPKM and raw counts) from pooled samples from early stages of normal development. (b) Displays expression data from inhibitor experiment from pooled samples. (c) and (d) show individual replicates expression data for early stages and inhibitor experiment respectively. . .	19
20	<i>Platynereis dumerilii</i> web database: Annotation tab from Detailed data results interface. Displaying detailed annotation information for FoxA2.	20
21	<i>Platynereis dumerilii</i> web database: Search results interface checking the boxes from the search result options on the left will expand the results displayed.	21
22	<i>Platynereis dumerilii</i> web database: Search results interface with the option "Show Plots" selected. Expression plots for both, early and late stages are shown for the gene under search: FoxA2.	21
23	<i>Platynereis dumerilii</i> web database: Search results interface with the option "Show later stages" selected. Here the later stages expression data from FOXA2 is displayed.	22
24	<i>Platynereis dumerilii</i> web database: Search results interface "Show other info" option selected. Additional information links are displayed.	22
25	<i>Platynereis dumerilii</i> web database: Search results interface "Show other info" option selected. The Locus info link displays genomic information	23
26	<i>Platynereis dumerilii</i> web database: Search results interface "Show other info" option selected. The Coexpression info link displays the list of genes clustered with the gene under study. The expression profile of FoxA2 clusters with 7 other genes belonging to cluster 7.	23
27	<i>Platynereis dumerilii</i> web database: Search results interface "Show other info" option selected. The Ortholog groups link displays the list species where orthologs were found. For FoxA2, orthologs were found in all of the 18 selected species. . . .	24
28	<i>Platynereis dumerilii</i> web database: Search results interface "Show other info" option selected. The Ortholog groups Tree-ML tab displays phylogenic tree constructed with the ortholog protein sequences. Tree-ML for FoxA2 orthologs among the 18 species.	25
29	<i>Platynereis dumerilii</i> web database: Search interface search required to Find homeobox genes that are highest expressed at 8hpf.	26
30	<i>Platynereis dumerilii</i> web database: Search results interface . Here the result page is displaying the ten top hits, sorted by expression level at 8 hpf. Links to download data are shown with a floppy disk icon and are found in the upper frame.	27

List of Tables

1	Time points from Early Stages data set	2
2	Late Stages included in data set. Time points are shown in hours post fertilization (hpf), days (d) and months (M).	3
3	Species and number of protein sequences for comparative analysis	12
4	Number of orthologs genes between the 6 species	12
5	Species and number of genes used to find orthologs groups	12