Marine Genomics xxx (2014) xxx–xxx

Contents lists available at ScienceDirect

Marine Genomics

journal homepage: www.elsevier.com/locate/margen

Sequencing and analysis of the transcriptome of the acorn worm *Ptychodera flava*, an indirect developing hemichordate

Shu-Hwa Chen [a,1], Kun-Lin Li [a,b,1], I-Hsuan Lu [a], Yu-Bin Wang [a], Che-Huang Tung [c], Hsiu-Chi Ting [c], Ching-Yi Lin [c,d], Chung-Yen Lin [a,b,e,*], Yi-Hsien Su [c,*], Jr-Kai Yu [c,d,**]

[a] Institute of Information Science, Academia Sinica, 128 Academia Road, Section 2, Nankang, Taipei 11529, Taiwan
[b] Institute of Fisheries Science, National Taiwan University, Taipei, Taiwan
[c] Institute of Cellular and Organismic Biology, Academia Sinica, 128 Academia Road, Section 2, Nankang, Taipei 11529, Taiwan
[d] Institute of Oceanography, National Taiwan University, Taipei, Taiwan
[e] Institute of Population Health Sciences, National Health Research Institutes, 35 Keyan Road, Zhunan, Miaoli County 35053, Taiwan

ARTICLE INFO

Article history:
Received 20 January 2014
Received in revised form 28 April 2014
Accepted 28 April 2014
Available online xxxx

Keywords:
Transcriptome
Next generation sequencing
Development
Evolution
Signaling pathway
Hemichordate

ABSTRACT

Hemichordates are the sister group of echinoderms, and together they are closely related to chordates within the deuterostome lineage. Therefore, hemichordates represent an important animal group for the understanding of both the evolution of developmental mechanisms in deuterostome animals and the origin of chordates. Recently, the majority of studies investigating hemichordates have focused on the direct-developing enteropneust hemichordate *Saccoglossus kowalevskii*; few have focused on the indirect-developing hemichordates, partly because of the lack of extensive genomic resources in these animals. In this study, we report the sequencing and analysis of a transcriptome from an indirect-developing enteropneust hemichordate *Ptychodera flava*. We sequenced a mixed cDNA library from six developmental stages using the Roche GS FLX Titanium System to generate more than 879,000 reads. These reads were assembled into 17,990 contigs with an average length of 1316 bp. We found that 60% of the assembled contigs, along with 28% of the unassembled singleton reads, had significant hits to sequences in the NCBI database by a BLASTx search, and we also annotated these sequences and obtained Gene Ontology (GO) terms for 6744 contigs and 5802 singletons. We further identified candidate *P. flava* transcripts corresponding to genes involved in major developmental signaling pathways, including the Wnt, Notch and TGF-β signaling pathways. Using available genome/transcriptome datasets from the direct-developing hemichordate *S. kowalevskii*, the echinoderm *Strongylocentrotus purpuratus* and the chordate *Branchiostoma floridae*, we found that 90%, 80% and 73% of the annotated protein sequences in these respective species matched our *P. flava* transcriptome in a homology search. We also constructed a database for the *P. flava* transcriptome, and researchers can easily access this dataset online. Our dataset significantly increases the amount of available *P. flava* sequence data and can serve as a reference transcriptome for future studies using this species.

1. Introduction

Hemichordates are a group of marine invertebrates closely related to two other animal groups: the echinoderms and chordates. Recent phylogenomic analyses further confirmed that hemichordates and echinoderms are sister groups and comprise a clade called Ambulacraria (Bourlat et al., 2006; Dunn et al., 2008; Swalla and Smith, 2008). Together with the chordates, these animal groups represent the three major phyla (Hemichordata, Echinodermata and Chordata) of the deuterostome animals. Although hemichordates lack a notochord, excluding them from the phylum Chordata, they exhibit other "chordate characteristics", such as pharyngeal gill slits (Gillis et al., 2012; Gonzalez and Cameron, 2009; Rychel and Swalla, 2007). The developmental mechanism and morphogenic process of the hemichordate collar chord are also reminiscent of the chordate dorsal nerve cord (Kaul and Stach, 2010; Luttrell et al., 2012; Miyamoto and Wada, 2013; Nomaksteinsky et al., 2009). Moreover, recent studies on the genomic structures of *Hox* and *ParaHox* genes in hemichordates have helped to infer the ancestral clustering of these genes in the common ancestor of deuterostome animals (Freeman et al., 2012; Ikuta et al., 2013). Thus, hemichordates represent a key group of model animals for understanding the evolution of early deuterostomes and the origin of chordates.

Within the hemichordates, two distinct adult living modes can be recognized: the solitary vermiform enteropneusts (acorn worms) and the colonial tube-dwelling pterobranchs (Hyman., 1959). Currently,

* Corresponding authors.
** Correspondence to: J-K. Yu, Institute of Cellular and Organismic Biology, Academia Sinica, 128 Academia Road, Section 2, Nankang, Taipei, 11529, Taiwan. Tel.: +886 2 27899517; fax: +886 2 2785 8059.
E-mail addresses: yamatolin@gmail.com (C.-Y. Lin), yhsu@gate.sinica.edu.tw (Y.-H. Su), jkyu@gate.sinica.edu.tw (J.-K. Yu).
[1] These authors contributed equally to this work.

http://dx.doi.org/10.1016/j.margen.2014.04.010
1874-7787/

most of the research on hemichordate development has focused on two species of enteropneusts: *Ptychodera flava* and *Saccoglossus kowalevskii* (Lowe et al., 2004; Luttrell et al., 2012; Rottinger and Lowe, 2012). In recent years, other species of enteropneust worms and pterobranchs have also been investigated in developmental studies (Miyamoto et al., 2010; Miyamoto and Saito, 2007; Sato et al., 2008, 2009; Sato and Holland, 2008). Although *P. flava* and *S. kowalevskii* are both enteropneust worms, they exhibit noticeable differences in their early life cycles. *P. flava* is an indirect developer. After fertilization, the *P. flava* embryo develops into a feeding tornaria larva and remains planktonic in the water column for many months. During this planktonic period the larva passes through six stages of modifications, finally undergoing metamorphosis to become a juvenile (Hadfield, 1975). On the other hand, *S. kowalevskii* bypasses the larval stage and its embryo develops directly into a juvenile about five days after fertilization, and then the juvenile quickly becomes a benthic feeder (Rottinger and Lowe, 2012). It has long been recognized that the tornaria larvae of hemichordates are similar to the auricularia or dipleurula larvae of echinoderms (Holland, 2003; Rottinger and Lowe, 2012), suggesting that indirect development is the ancestral state of hemichordate development and likely can be traced back to the common ancestor of the Ambulacraria (i.e., hemichordates and echinoderms). Over the past decades, developmental studies on echinoderms, notably using indirect-developing sea urchin embryos, have constructed gene regulatory networks (GRNs) for germ layer specification, major axis patterning and tissue/cell type differentiation (Oliveri et al., 2008; Peter and Davidson, 2011; Rafiq et al., 2012; Saudemont et al., 2010; Su et al., 2009). Thus, comparative studies using indirect-developing hemichordate species, such as *P. flava*, on the early embryonic/larval patterning mechanisms will generate important insights into the evolution of body plans within the deuterostome lineage.

Previously, a large-scale EST project was performed on the direct-developer *S. kowalevskii* (Freeman et al., 2008), and that project has provided a tremendous amount of resources for evolutionary and developmental studies on this species (Barrantes et al., 1999; Darras et al., 2011; Gillis et al., 2012; Green et al., 2013; Lemons et al., 2010; Lowe et al., 2003, 2006; Pani et al., 2012). In contrast, a similar resource has not been available for the indirect-developing model hemichordate *P. flava*. In this study, we sequenced and analyzed the transcriptome of *P. flava* by a next generation sequencing (NGS) approach. Using the Roche 454 pyrosequencing platform, we generated approximately 879,000 reads from a mixed-staged cDNA library, and the reads were assembled into 17,990 contigs with an average length of 1316 bp. We have also constructed a searchable database for the assembled contigs and unassembled singleton reads, along with their basic annotation information. This set of *P. flava* transcriptomic data will serve as an important resource for future genomic and developmental studies on this species.

2. Materials and methods

2.1. Obtaining P. flava embryonic samples

Gravid animals of the enteropneust acorn worm *P. flava* were collected at Chito Bay, Penghu Islands, Taiwan (23°38′54.17″ N, 119°36′14.40″ E). Spawning of *P. flava* was induced by heat stimulation following a previously described method (Tagawa et al., 1998). *In vitro* fertilization and the subsequent culturing of embryos/larvae were performed at 23 °C according to published methods (Lowe et al., 2004; Tagawa et al., 1998).

2.2. RNA extraction and cDNA library construction

Total RNA was extracted from unfertilized eggs and embryonic/larval stages including early blastula (12 h post-fertilization (hpf)), late blastula (16 hpf), early gastrula (24 hpf), late gastrula (43 hpf) and tornaria larva (65 hpf) using the reagent TRIzol (Invitrogen). The tornaria larvae were collected before feeding with algae. The RNA from the eggs (20 μg), the late blastula stage (10 μg) and each of the remaining developmental stages (20 μg each) were mixed together. Subsequently, 50 μg of this pooled total RNA sample was used by Mission Biotech (Taipei, Taiwan) to construct the cDNA library. PolyA-RNA was purified using the Poly(A)Purist™ MAG Kit (Ambion) and fragmented by a RNA fragmentation solution. Double-stranded cDNA was synthesized using the cDNA Synthesis System Kit (Roche) with random primers as described in the protocol provided by the manufacturer. Subsequently, cDNA was purified using AMPure beads (Agencourt).

2.3. Sequencing and assembly

Sequencing was performed on one picotiter plate following the protocols for the GS FLX Titanium System (Roche Diagnostic) by Mission Biotech. The resulting raw read data .sff files were converted into fastq format and were cleaned by the SeqClean software (http://www.tigr.org/tdb/tgi/software/) to eliminate adaptors, low complexity sequences and very short sequences (less than 40 bp). After quality trimming, the remaining sequences were assembled by *de novo* assembling with the Newbler program (454 Roche, version 2.7). Other programs including Newbler 2.5, Newbler 2.3, MIRA and Celera assembler have also been applied. Because Newbler 2.7 assembly contains longer contigs, with most mappable 454 raw reads and efficient running time, we performed the following analyses based on this set of data.

2.4. Functional annotation

The assembled isotigs (contigs) and unassembled singlet reads (singletons) were annotated by BLASTx sequence homolog matches in the NCBI non-redundant protein sequence (nr) database (database releasing date: 2013/06/03) (Altschul et al., 1997) (E-value cutoff $1E^{-5}$). The longest putative open reading frame (ORF) of each transcript (contigs + singletons) was detected by TransDecoder (http://transdecoder.sourceforge.net/). ORFs ≥30 amino acid residues were scanned with SignalP (Petersen et al., 2011) and tmHMM (Krogh et al., 2001) to detect transmembrane signatures and with HMMER (http://hmmer.janelia.org/) to identify protein domains based on the PFAM database (Punta et al., 2012) (version 26). Annotations from the Gene Ontology and KEGG gene databases were derived from the GO Consortium repository info and KEGG Automatic Annotation Server (Moriya et al., 2007), respectively, in the single best-hit mode.

2.5. Transcriptome comparison

The putative protein sequences of *P. flava* were derived from the assembled transcriptome (contigs and singletons) using TransDecoder as previously mentioned. The protein sequences of *S. kowalevskii*, *Strongylocentrotus purpuratus* and *Branchiostoma floridae* were downloaded from NCBI, SpBase (https://spbase.org) (Cameron et al., 2009) and the NCBI genome project site (BioProject PRJNA33245, PRJNA20249, genome assembly version 2) (Putnam et al., 2008), respectively. Each set was further filtered using CD-HIT (Li and Godzik, 2006) to eliminate redundant protein sequences. BLASTp searches of the protein sequences of each of the four species were performed against the other three species to identify the presence of matching sequences (E-value cutoff $1E^{-5}$). A bidirectional best hit of two sequences from two species is recognized as a true sequence homolog.

3. Results and discussion

3.1. 454-sequencing and assembly

The one-plate 454 pyrosequencing reaction of the *P. flava* mixed-staged cDNA library using the GS FLX Titanium platform generated

Please cite this article as: Chen, S.-H., et al., Sequencing and analysis of the transcriptome of the acorn worm Ptychodera flava, an indirect developing hemichordate, Mar. Genomics (2014), http://dx.doi.org/10.1016/j.margen.2014.04.010

Table 1
Statistics of the sequencing and assembly results.

Raw data		
Raw reads	Number of raw reads	879,681
	Read length (bp)	
	Average	362
	Max	755
	Min	40
	Total bases (bp)	318,114,873
	%N	0.0261%
Reads for assembly	Number of reads for assembly	832,751
	Average length (bp)	368
Assembly statistics		
Contigs	Number of contigs (≥50 bp)	17,990
	Total bases of contigs (bp)	23,677,180
	Number of mappable reads	628,104
	Bases of total mapped read	231,382,584
	Average depth	9.77
	%N	0.0005%
	Contig length (bp)	
	N50	901
	Average	1,316
	Max	13,112
Singletons	Number of singleton reads (≥50 bp)	125,601
	Total bases of singletons (bp)	42,447,474
	%N	0.0272%
	Read length (bp)	
	Average	338
	Max	642

879,681 raw reads with an average length of 362 bp (Table 1). After removing the adaptors of each sequence, 832,751 reads (94.67% of the raw reads, 307 Mbp in total) passed the quality control filters and were used in the *de novo* assembly. The average length of these high quality reads was 368 bp, and these reads were assembled using Newbler with criteria of the minimum length ≥50 bp. The average length of the 17,990 contigs was 1316 bp (Table 1 and Fig. 1). Additionally, 125,601 singletons were unable to be assembled into contigs. These singletons had an average length of 338 bp and were maintained in the final transcriptome in the following BLAST analyses. The contig number from our assembly is comparable to that obtained from expressed sequencing tag (EST) sequencing project for the hemichordate *S. kowalevskii* (Freeman et al., 2008).

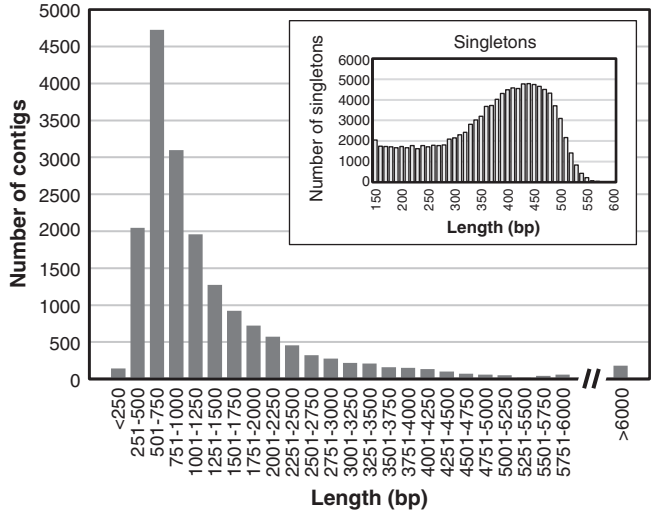


Fig. 1. Overview of the *de novo* assembly of the *P. flava* transcriptome. Distribution of the contig length is shown in the main figure; the inserted panel shows the size distribution of unassembled singletons that are longer than 150 bp.

3.2. Sequence annotation

Contigs and singletons exceeding 150 bp were searched by BLASTx against the NCBI non-redundant protein sequence (nr) database. A total of 41,713 *P. flava* sequences, including 10,705 contigs and 31,008 singletons, showed significant matches in NCBI nr database with an E-value cutoff of $1E^{-5}$ (Fig. 2; Supplementary Table S1). These matches corresponded to 17,050 unique NCBI accession numbers. When the contigs were set as the queries in the BLAST searches, 62.56% of the matched unique IDs were from the direct-developing enteropneust hemichordate *S. kowalevskii* (Supplementary Table S2). This overlap is consistent because *S. kowalevskii* is the only hemichordate species that has been used in a comprehensive EST sequencing project. Compared to the contigs, the short singleton reads have lower chance to cover conserved domain sequences; when singletons correspond to untranslated regions or non-conserved regions of coding sequences, it is unlikely to find matches in the database. Nevertheless, a significant portion of the singletons (27.76%) still can be annotated by a BLAST search (Fig. 2). Therefore, we included both contigs and singletons in the subsequent analyses even though these singletons likely represent partial transcripts. Marine invertebrate cultures are seldom bacteria-free, and we found that the percentage of sequences that match to bacterial sequences is 0.68% for the contig set and 1.40% for the contig + singleton set, respectively. Therefore, bacteria contamination is not a serious problem in the initial sampling procedure.

3.2.1. Gene ontology annotation

Gene ontology annotation was performed on the transcriptome for three independent categories: biological process, cellular component and molecular function. In total, 6744 contigs (37.56%) and 5802 singletons (5.19%) were annotated from the composition of the protein domain. Compared to the contigs, a lower percentage of singletons are annotated with protein domains, possibly because of their short length. Nevertheless, singletons may cover transcript species with lower expression levels and their protein products may function in the developmental process (GO:0032502), regulation of biological processes (GO:0048518, GO:0048519) and activity of receptors (GO: 0004872). The distribution of the GO terms assigned to the transcriptome for the three categories are shown in Fig. 3.

3.2.2. KEGG pathway analysis

To further analyze the transcriptome, we mapped the sequences to KEGG pathways to describe gene functions with an emphasis on biochemical pathways. Orthologous assignment and mapping of the contigs and singletons onto the biological pathways were performed using the KEGG Automatic Annotation Server. The results showed that 15,381 *P. flava* transcripts comprising 4785 contigs and 10,596 singletons were mapped onto 313 different pathways. Table 2 summarizes the top 10 ranked KEGG pathways in terms of the mapping coverage (number of mapped items/number of total items in a pathway). The complete KEGG hierarchical ontology structure and KEGG pathway view of the whole *P. flava* transcriptome can be retrieved in our *P. flava* transcriptome online database (described below in Section 3.5).

3.3. Candidate genes related to developmental signaling pathways

The transcriptome of *P. flava* was examined primarily to identify candidate genes that might be involved in developmental processes. Below we use KEGG pathway maps to illustrate candidate genes for three important signaling pathways that are involved in embryonic development in various metazoan animals (Figs. 4–6).

3.3.1. Wnt signaling pathway

The Wnt signaling pathway is highly conserved in metazoan animals and regulates many developmental processes. The Wnt pathway is activated by the binding of extracellular Wnt ligands to the membrane-

Please cite this article as: Chen, S.-H., et al., Sequencing and analysis of the transcriptome of the acorn worm Ptychodera flava, an indirect developing hemichordate, Mar. Genomics (2014), http://dx.doi.org/10.1016/j.margen.2014.04.010

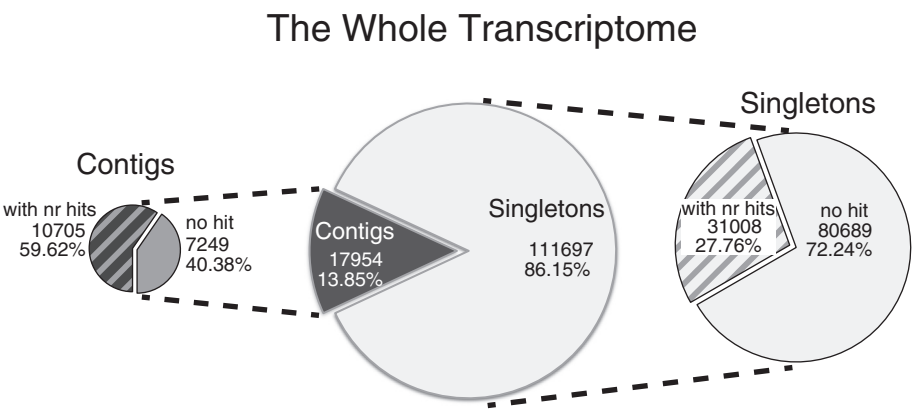
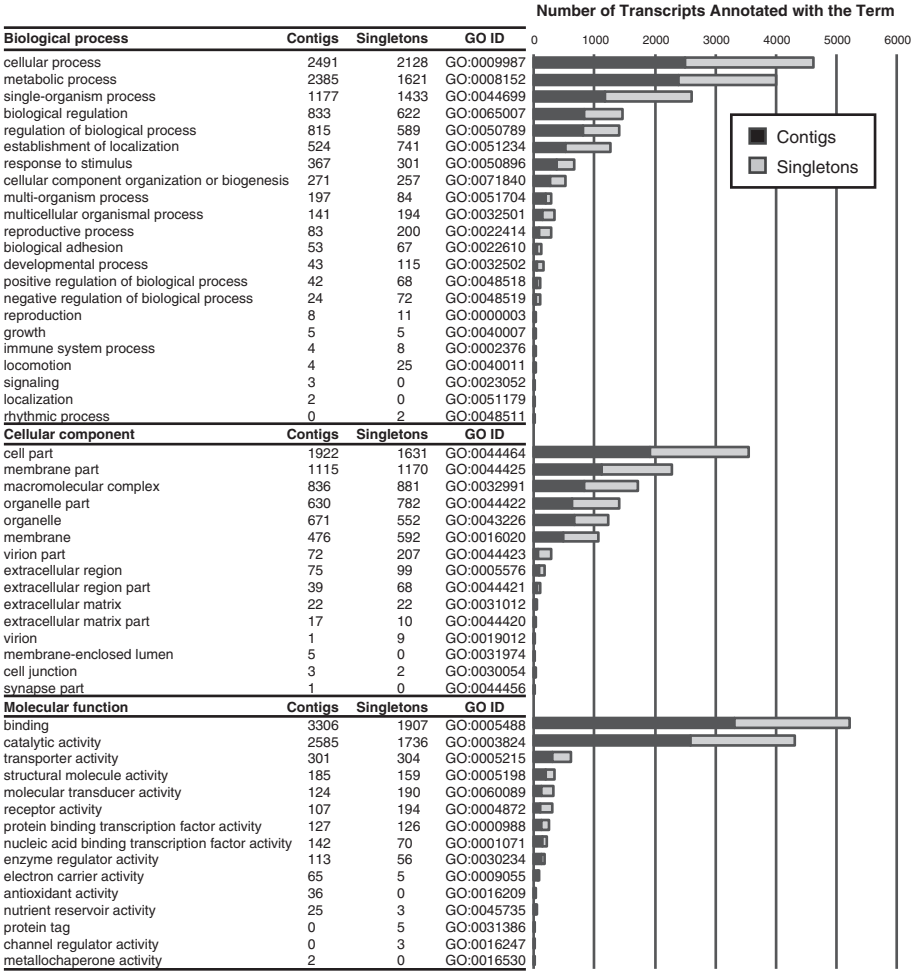


Fig. 2. BLAST annotation results of the *P. flava* transcriptome. The whole transcriptome refers to the combination of assembled contigs and unassembled singletons that are longer than 150 bp.

bound Frizzled receptors; subsequently, the intracellular signaling transduction can be achieved by three distinct mechanisms: the canonical pathway, the planar cell polarity (PCP) pathway and the Wnt/Ca^{2+} pathway (Fig. 4). Activation of the canonical Wnt pathway results in the stabilization and accumulation of β-catenin in the cytoplasm; subsequently, β-catenin enters into the nucleus and cooperates with a TCF/Lef transcription factor to activate transcription of target genes (Valenta et al., 2012). The canonical Wnt pathway functions in embryonic axis determination, endomesoderm specification and neural patterning in diverse metazoan animals (Croce et al., 2006; Darras et al., 2011; Demilly et al., 2013; Imai et al., 2000; Onai et al., 2009; Pani et al., 2012; Wikramanayake et al., 2003, 2004; Zorn et al., 1999). In this study, we identified *P. flava* contigs and singletons matching Wnt ligands, Wnt antagonist FRP, Frizzled receptors, LRP5/6 co-receptor, β-catenin and TCF/Lef in the Wnt KEGG pathway map (Fig. 4; Supplementary Table S3). We also found matches to members of the β-catenin



Biological process	Contigs	Singletons	GO ID
cellular process	2491	2128	GO:0009987
metabolic process	2385	1621	GO:0008152
single-organism process	1177	1433	GO:0044699
biological regulation	833	622	GO:0065007
regulation of biological process	815	589	GO:0050789
establishment of localization	524	741	GO:0051234
response to stimulus	367	301	GO:0050896
cellular component organization or biogenesis	271	257	GO:0071840
multi-organism process	197	84	GO:0051704
multicellular organismal process	141	194	GO:0032501
reproductive process	83	200	GO:0022414
biological adhesion	53	67	GO:0022610
developmental process	43	115	GO:0032502
positive regulation of biological process	42	68	GO:0048518
negative regulation of biological process	24	72	GO:0048519
reproduction	8	11	GO:0000003
growth	5	5	GO:0040007
immune system process	4	8	GO:0002376
locomotion	4	25	GO:0040011
signaling	3	0	GO:0023052
localization	2	0	GO:0051179
rhythmic process	0	2	GO:0048511
Cellular component	**Contigs**	**Singletons**	**GO ID**
cell part	1922	1631	GO:0044464
membrane part	1115	1170	GO:0044425
macromolecular complex	836	881	GO:0032991
organelle part	630	782	GO:0044422
organelle	671	552	GO:0043226
membrane	476	592	GO:0016020
virion part	72	207	GO:0044423
extracellular region	75	99	GO:0005576
extracellular region part	39	68	GO:0044421
extracellular matrix	22	22	GO:0031012
extracellular matrix part	17	10	GO:0044420
virion	1	9	GO:0019012
membrane-enclosed lumen	5	0	GO:0031974
cell junction	3	2	GO:0030054
synapse part	1	0	GO:0044456
Molecular function	**Contigs**	**Singletons**	**GO ID**
binding	3306	1907	GO:0005488
catalytic activity	2585	1736	GO:0003824
transporter activity	301	304	GO:0005215
structural molecule activity	185	159	GO:0005198
molecular transducer activity	124	190	GO:0060089
receptor activity	107	194	GO:0004872
protein binding transcription factor activity	127	126	GO:0000988
nucleic acid binding transcription factor activity	142	70	GO:0001071
enzyme regulator activity	113	56	GO:0030234
electron carrier activity	65	5	GO:0009055
antioxidant activity	36	0	GO:0016209
nutrient reservoir activity	25	3	GO:0045735
protein tag	0	5	GO:0031386
channel regulator activity	0	3	GO:0016247
metallochaperone activity	2	0	GO:0016530

Fig. 3. Summary of the Gene Ontology annotation of the transcriptome. The number of annotated contigs and singletons in each of the selected GO categories is shown on the left; the graph on the right is the distribution of transcripts in each of the selected GO categories.

Please cite this article as: Chen, S.-H., et al., Sequencing and analysis of the transcriptome of the acorn worm Ptychodera flava, an indirect developing hemichordate, Mar. Genomics (2014), http://dx.doi.org/10.1016/j.margen.2014.04.010

Table 2
Top ten KEGG pathways with highest mapping coverage (number of mapped items/number of total items in a pathway).

KEGG pathways		Mapped items	Coverage (mapped/total)
00531	Glycosaminoglycan degradation	13	92.9%
00563	Glycosylphosphatidylinositol(GPI)-anchor biosynthesis	24	92.3%
04146	Peroxisome	65	91.5%
03015	mRNA surveillance pathway	54	91.5%
00953	Drug metabolism—other enzymes	20	90.9%
03040	Spliceosome	107	89.2%
03022	Basal transcription factors	31	88.6%
03013	RNA transport	116	86.6%
04150	mTOR signaling pathway	32	86.5%
00510	N-Glycan biosynthesis	37	84.1%

destruction complex such as GSK-3β, Axin, CK1 and APC. In addition, members of the non-canonical Wnt pathways (the PCP pathway and Wnt/Ca^{2+} pathway) were well represented in the *P. flava* transcriptome with significant sequence similarity (Fig. 4; Supplementary Table S3).

3.3.2. Notch signaling pathway

The Notch signaling pathway regulates numerous cellular decisions during development in diverse metazoan animals (Bray, 2006). Both the Notch receptor and its ligands, Delta and Serrate, are membrane proteins; thus the activation of the pathway involves direct cell contact (Fig. 5). After the binding of either Delta or Serrate, the Notch receptor undergoes a series of cleavage events and releases the Notch intracellular domain (NICD). This domain then enters the nucleus and cooperates with a CSL family transcription factor and its co-activator Mastermind (Mam) to promote transcription of target genes. The Notch signaling pathway is reported from various animal systems to function in different developmental processes, including neurogenesis, cell lineage decision and boundary formation (Barrantes et al., 1999; Bray, 2006; Hudson and Yasuo, 2006; Lu et al., 2012; Materna and Davidson, 2012; Oda et al., 2007; Pasini et al., 2006; Rivera and Weisblat, 2009). Candidates of the Notch signaling pathway members were identified within the *P. flava* transcriptome by sequence similarity, and they are shown in Fig. 5 and Supplementary Table S4.

3.3.3. TGF-β signaling pathway

The TGF-β signaling pathway is used repeatedly during development in many animals to control various processes. The TGF-β superfamily proteins comprise a large group of structurally related ligands; they are synthesized as precursors that are cleaved into mature ligands containing six to nine conserved cysteines near their C-terminus. The mature ligands bind to type-II and type-I receptors of the serine/threonine kinase family to assemble a complex. This binding results in the phosphorylation of the type-I receptor and subsequently activates the Smad family transcription factors to regulate downstream target genes (Massagué, 2012). One of the best-described functions of the TGF-beta signaling pathway is the BMP pathway for dorsoventral patterning in bilaterian animals (De Robertis, 2008), although the dorsoventral axis along the BMP signaling gradient is inverted in the chordate lineage compared to other bilaterian animals (Denes et al., 2007; Lowe et al., 2006; Yu et al., 2007). In this study, we identified candidate TGF-β signaling pathway members in the *P. flava* transcriptome, and the results are listed in Fig. 6 and Supplementary Table S5.

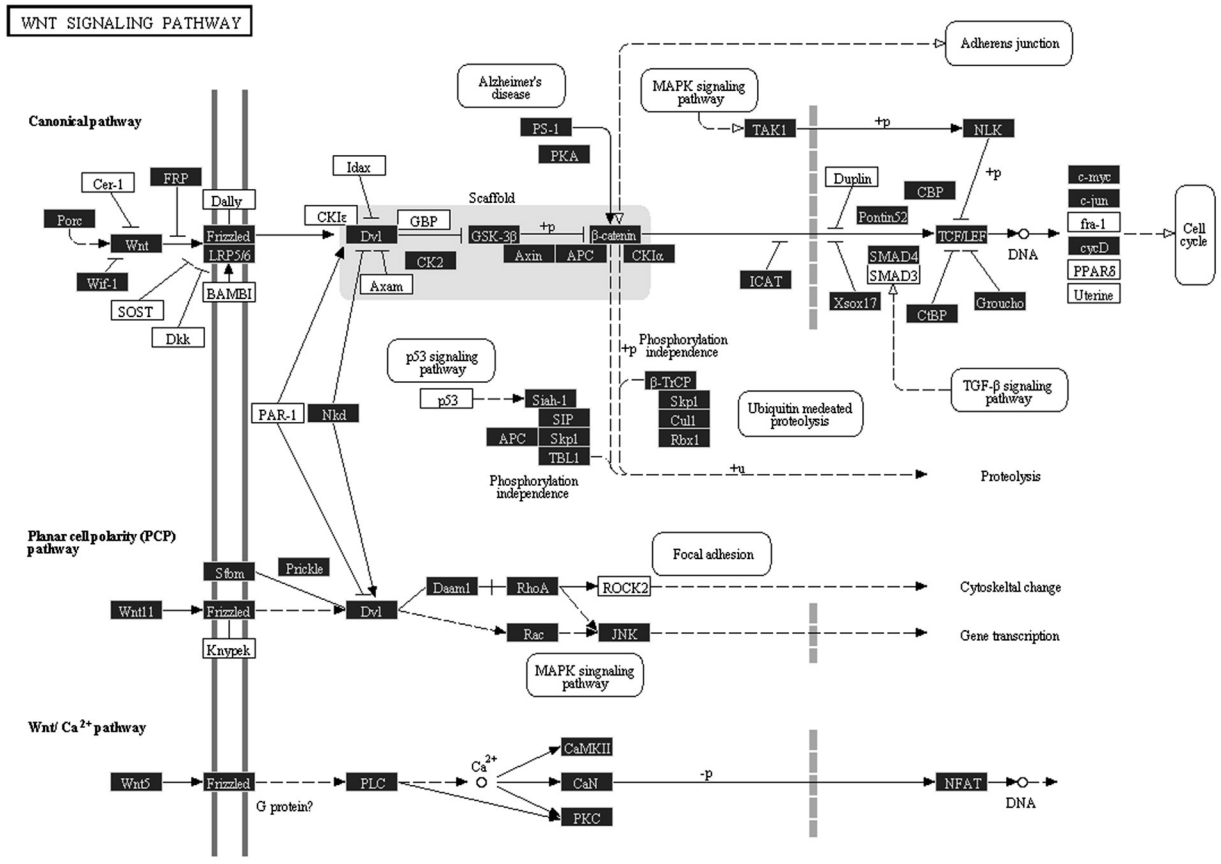


Fig. 4. Three Wnt signaling pathways in the KEGG pathway map. Members of the Wnt signaling pathways that are present in the *P. flava* transcriptome are in black-shaded boxes. Members of the pathways that are not found in the transcriptome are in white boxes.

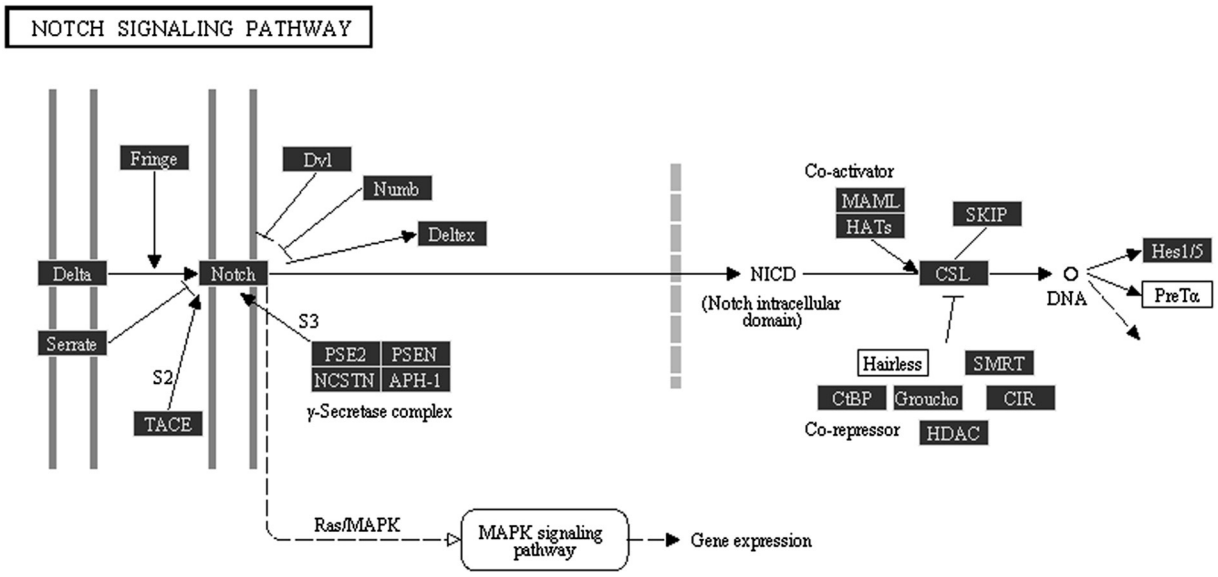


Fig. 5. Delta/Notch signaling pathway in the KEGG pathway map. Members of the pathway that are present in the *P. flava* transcriptome are in black-shaded boxes. Members of the pathways that are not found in the transcriptome are in white boxes.

3.4. Comparative analysis of the P. flava transcriptome

To analyze the coverage of the *P. flava* transcriptome, we generated two predicted protein sets from the assembled contigs and the whole transcriptome with both contigs and singletons and compared them with the protein repertoires of related deuterostome animals. The protein sets from the assembled contigs and the whole transcriptome contain 12,013 and 58,550 putative proteins, respectively. Reciprocal BLAST searches were performed between the *P. flava* protein sets and the predicted proteins from the three related clades represented by the hemichordate *S. kowalevskii*, echinoderm *S. purpuratus* and cephalochordate *B. floridae* (Fig. 7). Approximately two-thirds of the proteins from the assembled contigs and 40% from the whole transcriptome matched the protein repertoires of the three species. Reciprocally, using protein sequences of the three related species as queries to BLAST the *P. flava* protein sets, we observed that the

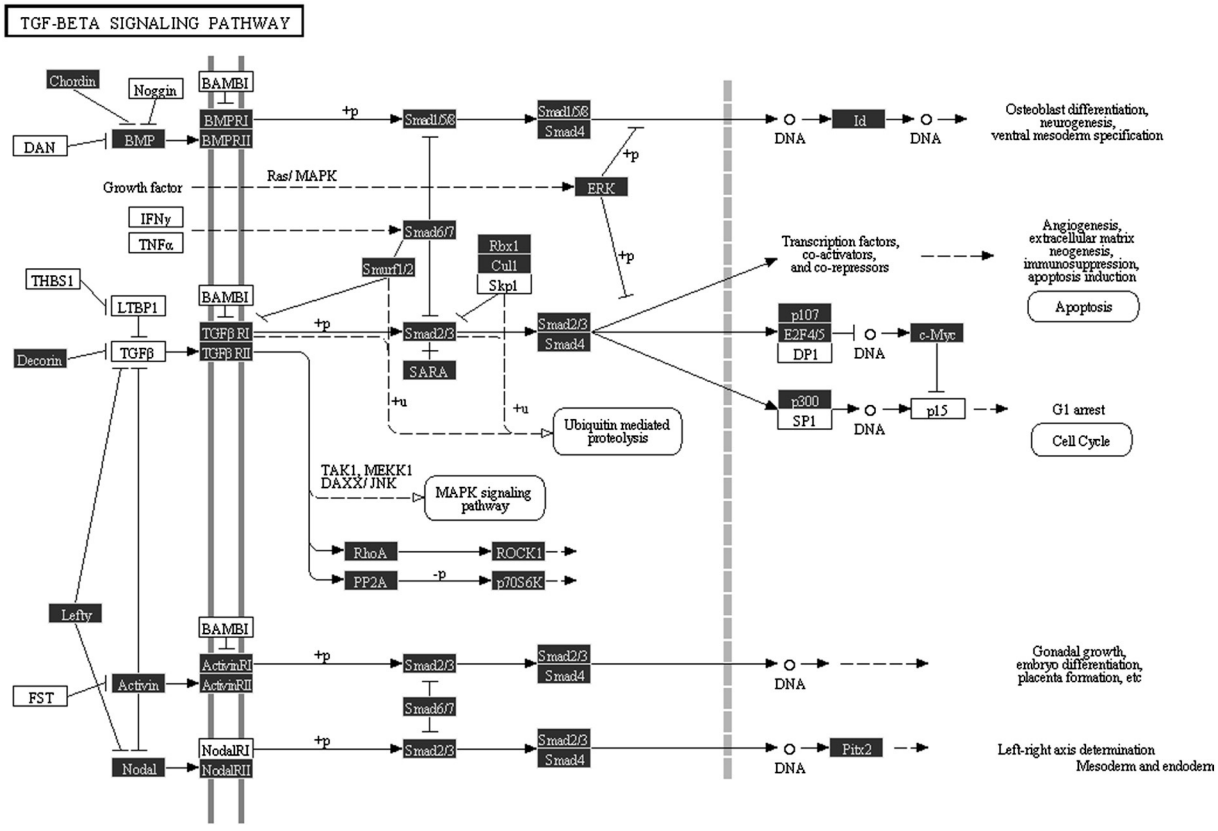


Fig. 6. TGF-β signaling pathway in the KEGG pathway map. Members of the pathway that are present in the *P. flava* transcriptome are in black-shaded boxes. Members of the pathways that are not found in the transcriptome are in white boxes.

Please cite this article as: Chen, S.-H., et al., Sequencing and analysis of the transcriptome of the acorn worm Ptychodera flava, an indirect developing hemichordate, Mar. Genomics (2014), http://dx.doi.org/10.1016/j.margen.2014.04.010

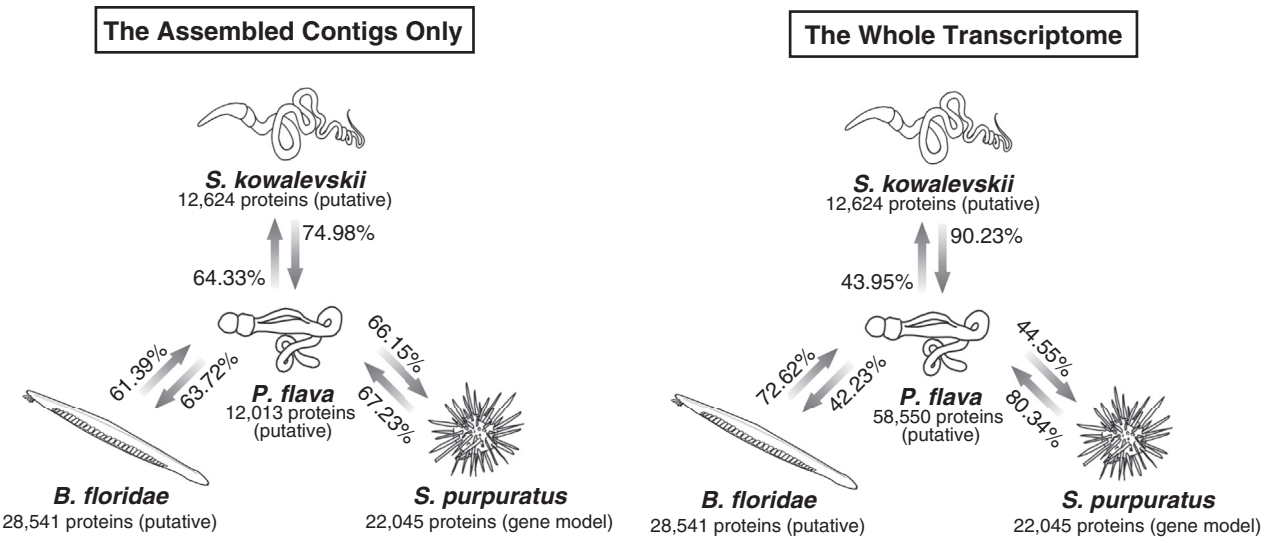


Fig. 7. Comparisons of putative proteins between *P. flava* and direct-developing hemichordate *S. kowalevskii*, sea urchin *S. purpuratus*, and amphioxus *B. floridae*. Arrowheads indicate the direction of the BLAST search, from the query to the target database.

percentages of the three protein repertoires matching *P. flava* proteins were higher in the set from the whole transcriptome than that of the contigs (74.98% to 90.23% for *S. kowalevskii*; 67.23% to 80.34% for *S. purpuratus*; 61.39% to 72.62% for *B. floridae*). Therefore, the whole transcriptome with contigs and singletons provided a more complete collection of transcripts for future studies. Notably, the graded percentages of the matched sequences between the three related species and the *P. flava* protein sets are consistent with their phylogenetic relationships. For example, *S. kowalevskii* is most closely related to *P. flava* and has the highest percentage of proteins matching the *P. flava* protein sets. On the other hand, *B. floridae* is more distant from *P. flava* and has the lowest percentage of proteins matching the *P. flava* protein sets.

3.5. *P. flava* transcriptome database

Although transcriptome sequencing was also performed recently in *P. flava* to identify putative dorsoventral patterning genes (Rottinger

Fig. 8. Screenshots of the *P. flava* transcriptome database. Basic sequence information (the central bottom panel) of the transcriptome can be accessed by a sequence similarity search (BLAST search, the right top panel), keyword search against the annotation table of transcripts (full-text search, the central panel), or by browsing pathways and protein categories (KEGG global view, the left panel). From a user input interface, the website can also perform the computation of enrichment in GO or KEGG items (enrichment analysis, the right bottom panel).

Please cite this article as: Chen, S.-H., et al., Sequencing and analysis of the transcriptome of the acorn worm Ptychodera flava, an indirect developing hemichordate, Mar. Genomics (2014), http://dx.doi.org/10.1016/j.margen.2014.04.010

and Martindale, 2011), the majority of the sequence data generated from that study is not currently available in the public domain. To facilitate the sharing of *P. flava* transcriptomic data, we have deposited our raw sequencing data into the NCBI database (BioProject PRJNA227405). Additionally, we also constructed a searchable database for our *P. flava* transcriptome (http://molas.iis.sinica.edu.tw/hemichordate; Fig. 8). Users can perform sequence-based searches on the assembled transcriptome (contigs only or contigs + singletons) or the derived protein coding sets. Users can access the annotation results for each sequence, including for both the assembled contigs and the singletons. Users can also perform keyword-based searches on the annotations including the descriptions of the nr best match, functional annotation from KEGG and GO terms, and annotations from the protein structure scan. Furthermore, the whole transcriptome can be viewed in KEGG hierarchical ontology structures and pathways. The enrichment analysis provides a batch query by a *P. flava* transcriptome entry ID list, in which the enrichment of GO terms and KEGG pathways can be calculated. Our transcriptome represents transcripts from several embryonic stages. Therefore, sequences not found in the transcriptome database do not necessarily imply that they are missing in *P. flava*. For example, among the eight *Hox* (*PfHox1*, *4*, *5*, *6*, *9/10*, *11/13a*, *11/13b* and *11/13c*) and three *ParaHox* (*PfGsx*, *PfXlox* and *PfCdx*), of which the cDNA sequences are available in the GenBank (Ikuta et al., 2013; Peterson, 2004), only *PfHox9/10*, *PfXlox* and *PfCdx* are found in the transcriptome database. It is currently unknown whether other *Hox* genes are not expressed during embryogenesis. *PfGsx* is known to be expressed during embryogenesis and its low expression level may explain the lack of detection in the transcriptome (Ikuta et al., 2013). It is reported that sea urchin *S. purpuratus* has greatly expanded innate immunity repertoire (Hibino et al., 2006; Sodergren et al., 2006). For example, 222 genes encoding Toll-like receptors (TLR) are identified in the sea urchin genome whereas other animal genomes usually contain one to 20 TLR genes. The *P. flava* transcriptome does not seem to contain a large number of TLR. Only 4 contigs and 1 singleton match to 19 sea urchin TLR group IA and 12 IB genes with complete sequences, respectively (Hibino et al., 2006). This result suggests that *P. flava* may not have expanded their innate immunity repertoire; alternatively, immune-related genes may be activated only after immune challenges and therefore are not detected in our transcripome analysis. It will require further information from *P. flava* genome sequencing project to confirm.

3.6. Conclusions

In this study, we produced a large-scale developmental transcriptomic dataset for the indirect developing Ptychoderidae enteropneust hemichordate *P. flava* using the 454 GS FLX platform. This dataset significantly increases the amount of *P. flava* sequence data in the public database, allowing gene discovery for future developmental studies and comparative analyses into a broad range of fundamental questions regarding the evolution of developmental mechanisms in deuterostome animals. We have also constructed an online database to easily access to this *P. flava* transcriptomic dataset. We anticipate that this dataset will assist future studies on *P. flava*, notably for genome annotation and comparative genomic studies with other organisms.

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.margen.2014.04.010.

Acknowledgments

Research funds were provided by the National Science Council, Taiwan (101-2627-B-001-001 to YHS; 101-2923-B-001-004-MY2 to YHS and JKY) and the Academia Sinica, Taiwan (Career Development Award, AS-98-CDA-L06 to JKY).

References

Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J., Zhang, Z., Miller, W., Lipman, D.J., 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. 25, 3389–3402.

Barrantes, I.B., Elia, A.J., Wunsch, K., Hrabe de Angelis, M.H., Mak, T.W., Rossant, J., Conlon, R.A., Gossler, A., de la Pompa, J.L., 1999. Interaction between Notch signalling and Lunatic fringe during somite boundary formation in the mouse. Curr. Biol. 9, 470–480.

Bourlat, S.J., Juliusdottir, T., Lowe, C.J., Freeman, R., Aronowicz, J., Kirschner, M., Lander, E.S., Thorndyke, M., Nakano, H., Kohn, A.B.., Heyland, A., Moroz, L.L., Copley, R.R., Telford, M.J., 2006. Deuterostome phylogeny reveals monophyletic chordates and the new phylum Xenoturbellida. Nature 444, 85–88.

Bray, S.J., 2006. Notch signalling: a simple pathway becomes complex. Nat. Rev. Mol. Cell Biol. 7, 678–689.

Cameron, R.A., Samanta, M., Yuan, A., He, D., Davidson, E., 2009. SpBase: the sea urchin genome database and web site. Nucleic Acids Res. 37, D750–D754.

Croce, J., Duloquin, L., Lhomond, G., McClay, D.R., Gache, C., 2006. Frizzled5/8 is required in secondary mesenchyme cells to initiate archenteron invagination during sea urchin development. Development 133, 547–557.

Darras, S., Gerhart, J., Terasaki, M., Kirschner, M., Lowe, C.J., 2011. Beta-catenin specifies the endomesoderm and defines the posterior organizer of the hemichordate *Saccoglossus kowalevskii*. Development 138, 959–970.

De Robertis, E.M., 2008. Evo-devo: variations on ancestral themes. Cell 132, 185–195.

Demilly, A., Steinmetz, P., Gazave, E., Marchand, L., Vervoort, M., 2013. Involvement of the Wnt/beta-catenin pathway in neurectoderm architecture in *Platynereis dumerilii*. Nat. Commun. 4, 1915.

Denes, A.S., Jekely, G., Steinmetz, P.R., Raible, F., Snyman, H., Prud'homme, B., Ferrier, D.E., Balavoine, G., Arendt, D., 2007. Molecular architecture of annelid nerve cord supports common origin of nervous system centralization in bilateria. Cell 129, 277–288.

Dunn, C.W., Hejnol, A., Matus, D.Q., Pang, K., Browne, W.E., Smith, S.A., Seaver, E., Rouse, G.W., Obst, M., Edgecombe, G.D., Sorensen, M.V., Haddock, S.H., Schmidt-Rhaesa, A., Okusu, A., Kristensen, R.M., Wheeler, W.C., Martindale, M.Q., Giribet, G., 2008. Broad phylogenomic sampling improves resolution of the animal tree of life. Nature 452, 745–749.

Freeman Jr., R.M., Wu, M., Cordonnier-Pratt, M.M., Pratt, L.H., Gruber, C.E., Smith, M., Lander, E.S., Stange-Thomann, N., Lowe, C.J., Gerhart, J., Kirschner, M., 2008. cDNA sequences for transcription factors and signaling proteins of the hemichordate *Saccoglossus kowalevskii*: efficacy of the expressed sequence tag (EST) approach for evolutionary and developmental studies of a new organism. Biol. Bull. 214, 284–302.

Freeman, R., Ikuta, T., Wu, M., Koyanagi, R., Kawashima, T., Tagawa, K., Humphreys, T., Fang, G.C., Fujiyama, A., Saiga, H., Lowe, C., Worley, K., Jenkins, J., Schmutz, J., Kirschner, M., Rokhsar, D., Satoh, N., Gerhart, J., 2012. Identical genomic organization of two hemichordate hox clusters. Curr. Biol. 22, 2053–2058.

Gillis, J.A., Fritzenwanker, J.H., Lowe, C.J., 2012. A stem-deuterostome origin of the vertebrate pharyngeal transcriptional network. Proc. Biol. Sci. 279, 237–246.

Gonzalez, P., Cameron, C.B., 2009. The gill slits and pre-oral ciliary organ of Protoglossus (Hemichordata: Enteropneusta) are filter-feeding structures. Biol. J. Linn. Soc. 98, 898–906.

Green, S.A., Norris, R.P., Terasaki, M., Lowe, C.J., 2013. FGF signaling induces mesoderm in the hemichordate *Saccoglossus kowalevskii*. Development 140, 1024–1033.

Hadfield, M.G., 1975. Hemichordata. , Academic Press, INC, New York.

Hibino, T., Loza-Coll, M., Messier, C., Majeske, A.J., Cohen, A.H., Terwilliger, D.P., Buckley, K.M., Brockton, V., Nair, S.V., Berney, K., Fugmann, S.D., Anderson, M.K., Pancer, Z., Cameron, R.A., Smith, L.C., Rast, J.P., 2006. The immune gene repertoire encoded in the purple sea urchin genome. Dev. Biol. 300, 349–365.

Holland, N.D., 2003. Early central nervous system evolution: an era of skin brains? Nat. Rev. Neurosci. 4, 617–627.

Hudson, C., Yasuo, H., 2006. A signalling relay involving Nodal and Delta ligands acts during secondary notochord induction in Ciona embryos. Development 133, 2855–2864.

Hyman, L.H., 1959. The Invertebrates: Smaller Coelomate Groups, Volume V. , McGraw-Hill Book Company, Inc, New York.

Ikuta, T., Chen, Y.C., Annunziata, R., Ting, H.C., Tung, C.H., Koyanagi, R., Tagawa, K., Humphreys, T., Fujiyama, A., Saiga, H., Satoh, N., Yu, J.K., Arnone, M.I., Su, Y.H., 2013. Identification of an intact ParaHox cluster with temporal colinearity but altered spatial colinearity in the hemichordate *Ptychodera flava*. BMC Evol. Biol. 13, 129.

Imai, K., Takada, N., Satoh, N., Satou, Y., 2000. (Beta)-catenin mediates the specification of endoderm cells in ascidian embryos. Development 127, 3009–3020.

Kaul, S., Stach, T., 2010. Ontogeny of the collar cord: neurulation in the hemichordate *Saccoglossus kowalevskii*. J. Morphol. 271, 1240–1259.

Krogh, A., Larsson, B., von Heijne, G., Sonnhammer, E.L., 2001. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. J. Mol. Biol. 305, 567–580.

Lemons, D., Fritzenwanker, J.H., Gerhart, J., Lowe, C.J., McGinnis, W., 2010. Co-option of an anteroposterior head axis patterning system for proximodistal patterning of appendages in early bilaterian evolution. Dev. Biol. 344, 358–362.

Li, W., Godzik, A., 2006. Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. Bioinformatics 22, 1658–1659.

Lowe, C.J., Wu, M., Salic, A., Evans, L., Lander, E., Stange-Thomann, N., Gruber, C.E., Gerhart, J., Kirschner, M., 2003. Anteroposterior patterning in hemichordates and the origins of the chordate nervous system. Cell 113, 853–865.

Lowe, C.J., Tagawa, K., Humphreys, T., Kirschner, M., Gerhart, J., 2004. Hemichordate embryos: procurement, culture, and basic methods. Methods Cell Biol. 74, 171–194.

Lowe, C.J., Terasaki, M., Wu, M., Freeman Jr., R.M., Runft, L., Kwan, K., Haigo, S., Aronowicz, J., Lander, E., Gruber, C., Smith, M., Kirschner, M., Gerhart, J., 2006. Dorsoventral patterning in hemichordates: insights into early chordate evolution. PLoS Biol. 4, e291.

Please cite this article as: Chen, S.-H., et al., Sequencing and analysis of the transcriptome of the acorn worm Ptychodera flava, an indirect developing hemichordate, Mar. Genomics (2014), http://dx.doi.org/10.1016/j.margen.2014.04.010

Lu, T.M., Luo, Y.J., Yu, J.K., 2012. BMP and Delta/Notch signaling control the development of amphioxus epidermal sensory neurons: insights into the evolution of the peripheral sensory system. Development 139, 2020–2030.

Luttrell, S., Konikoff, C., Byrne, A., Bengtsson, B., Swalla, B.J., 2012. Ptychoderid hemichordate neurulation without a notochord. Integr. Comp. Biol. 52, 829–834.

Massagué, J., 2012. TGFbeta signalling in context. Nat. Rev. Mol. Cell Biol. 13, 616–630.

Materna, S.C., Davidson, E.H., 2012. A comprehensive analysis of Delta signaling in pre-gastrular sea urchin embryos. Dev. Biol. 364, 77–87.

Miyamoto, N., Saito, Y., 2007. Morphology and development of a new species of Balanoglossus (Hemichordata: Enteropneusta: Ptychoderidae) from Shimoda, Japan. Zool. Sci. 24, 1278–1285.

Miyamoto, N., Wada, H., 2013. Hemichordate neurulation and the origin of the neural tube. Nat. Commun. 4, 2713.

Miyamoto, N., Nakajima, Y., Wada, H., Saito, Y., 2010. Development of the nervous system in the acorn worm *Balanoglossus simodensis*: insights into nervous system evolution. Evol. Dev. 12, 416–424.

Moriya, Y., Itoh, M., Okuda, S., Yoshizawa, A.C., Kanehisa, M., 2007. KAAS: an automatic genome annotation and pathway reconstruction server. Nucleic Acids Res. 35, W182–W185.

Nomaksteinsky, M., Rottinger, E., Dufour, H.D., Chettouh, Z., Lowe, C.J., Martindale, M.Q., Brunet, J.F., 2009. Centralization of the deuterostome nervous system predates chordates. Curr. Biol. 19, 1264–1269.

Oda, H., Nishimura, O., Hirao, Y., Tarui, H., Agata, K., Akiyama-Oda, Y., 2007. Progressive activation of Delta-Notch signaling from around the blastopore is required to set up a functional caudal lobe in the spider *Achaearanea tepidariorum*. Development 134, 2195–2205.

Oliveri, P., Tu, Q., Davidson, E.H., 2008. Global regulatory logic for specification of an embryonic cell lineage. Proc. Natl. Acad. Sci. U. S. A. 105, 5955–5962.

Onai, T., Lin, H.C., Schubert, M., Koop, D., Osborne, P.W., Alvarez, S., Alvarez, R., Holland, N.D., Holland, L.Z., 2009. Retinoic acid and Wnt/beta-catenin have complementary roles in anterior/posterior patterning embryos of the basal chordate amphioxus. Dev. Biol. 332, 223–233.

Pani, A.M., Mullarkey, E.E., Aronowicz, J., Assimacopoulos, S., Grove, E.A., Lowe, C.J., 2012. Ancient deuterostome origins of vertebrate brain signalling centres. Nature 483, 289–294.

Pasini, A., Amiel, A., Rothbacher, U., Roure, A., Lemaire, P., Darras, S., 2006. Formation of the ascidian epidermal sensory neurons: insights into the origin of the chordate peripheral nervous system. PLoS Biol. 4, e225.

Peter, I.S., Davidson, E.H., 2011. A gene regulatory network controlling the embryonic specification of endoderm. Nature 474, 635–639.

Petersen, T.N., Brunak, S., von Heijne, G., Nielsen, H., 2011. SignalP 4.0: discriminating signal peptides from transmembrane regions. Nat. Methods 8, 785–786.

Peterson, K.J., 2004. Isolation of Hox and Parahox genes in the hemichordate *Ptychodera flava* and the evolution of deuterostome Hox genes. Mol. Phylogenet. Evol. 31, 1208–1215.

Punta, M., Coggill, P.C., Eberhardt, R.Y., Mistry, J., Tate, J., Boursnell, C., Pang, N., Forslund, K., Ceric, G., Clements, J., Heger, A., Holm, L., Sonnhammer, E.L., Eddy, S.R., Bateman, A., Finn, R.D., 2012. The Pfam protein families database. Nucleic Acids Res. 40, D290–D301.

Putnam, N.H., Butts, T., Ferrier, D.E., Furlong, R.F., Hellsten, U., Kawashima, T., Robinson-Rechavi, M., Shoguchi, E., Terry, A., Yu, J.K., Benito-Gutierrez, E.L., Dubchak, I., Garcia-Fernandez, J., Gibson-Brown, J.J., Grigoriev, I.V., Horton, A.C., de Jong, P.J., Jurka, J., Kapitonov, V.V., Kohara, Y., Kuroki, Y., Lindquist, E., Lucas, S., Osoegawa, K., Pennacchio, L.A., Salamov, A.A., Satou, Y., Sauka-Spengler, T., Schmutz, J., Shin, I.T., Toyoda, A., Bronner-Fraser, M., Fujiyama, A., Holland, L.Z., Holland, P.W., Satoh, N., Rokhsar, D.S., 2008. The amphioxus genome and the evolution of the chordate karyotype. Nature 453, 1064–1071.

Rafiq, K., Cheers, M.S., Ettensohn, C.A., 2012. The genomic regulatory control of skeletal morphogenesis in the sea urchin. Development 139, 579–590.

Rivera, A.S., Weisblat, D.A., 2009. And Lophotrochozoa makes three: Notch/Hes signaling in annelid segmentation. Dev. Genes Evol. 219, 37–43.

Rottinger, E., Lowe, C.J., 2012. Evolutionary crossroads in developmental biology: hemichordates. Development 139, 2463–2475.

Rottinger, E., Martindale, M.Q., 2011. Ventralization of an indirect developing hemichordate by NiCl(2) suggests a conserved mechanism of dorso-ventral (D/V) patterning in Ambulacraria (hemichordates and echinoderms). Dev. Biol. 354, 173–190.

Rychel, A.L., Swalla, B.J., 2007. Development and evolution of chordate cartilage. J. Exp. Zool. B Mol. Dev. Evol. 308, 325–335.

Sato, A., Holland, P.W., 2008. Asymmetry in a pterobranch hemichordate and the evolution of left-right patterning. Dev. Dyn. 237, 3634–3639.

Sato, A., Bishop, J.D., Holland, P.W., 2008. Developmental biology of pterobranch hemichordates: history and perspectives. Genesis 46, 587–591.

Sato, A., White-Cooper, H., Doggett, K., Holland, P.W., 2009. Degenerate evolution of the hedgehog gene in a hemichordate lineage. Proc. Natl. Acad. Sci. U. S. A. 106, 7491–7494.

Saudemont, A., Haillot, E., Mekpoh, F., Bessodes, N., Quirin, M., Lapraz, F., Duboc, V., Rottinger, E., Range, R., Oisel, A., Besnardeau, L., Wincker, P., Lepage, T., 2010. Ancestral regulatory circuits governing ectoderm patterning downstream of Nodal and BMP2/4 revealed by gene regulatory network analysis in an echinoderm. PLoS Genet. 6, e1001259.

Sodergren, E., Weinstock, G.M., Davidson, E.H., Cameron, R.A., Gibbs, R.A., Angerer, R.C., Angerer, L.M., Arnone, M.I., Burgess, D.R., Burke, R.D., Coffman, J.A., Dean, M., Elphick, M.R., Ettensohn, C.A., Foltz, K.R., Hamdoun, A., Hynes, R.O., Klein, W.H., Marzluff, W., McClay, D.R., Morris, R.L., Mushegian, A., Rast, J.P., Smith, L.C., Thorndyke, M.C., Vacquier, V.D., Wessel, G.M., Wray, G., Zhang, L., Elsik, C.G., Ermolaeva, O., Hlavina, W., Hofmann, G., Kitts, P., Landrum, M.J., Mackey, A.J., Maglott, D., Panopoulou, G., Poustka, A.J., Pruitt, K., Sapojnikov, V., Song, X., Souvorov, A., Solovyev, V., Wei, Z., Whittaker, C.A., Worley, K., Durbin, K.J., Shen, Y., Fedrigo, O., Garfield, D., Haygood, R., Primus, A., Satija, R., Severson, T., Gonzalez-Garay, M.L., Jackson, A.R., Milosavljevic, A., Tong, M., Killian, C.E., Livingston, B.T., Wilt, F.H., Adams, N., Belle, R., Carbonneau, S., Cheung, R., Cormier, P., Cosson, B., Croce, J., Fernandez-Guerra, A., Geneviere, A.M., Goel, M., Kelkar, H., Morales, J., Mulner-Lorillon, O., Robertson, A.J., Goldstone, J.V., Cole, B., Epel, D., Gold, B., Hahn, M.E., Howard-Ashby, M., Scally, M., Stegeman, J.J., Allgood, E.L., Cool, J., Judkins, K.M., McCafferty, S.S., Musante, A.M., Obar, R.A., Rawson, A.P., Rossetti, B.J., Gibbons, I.R., Hoffman, M.P., Leone, A., Istrail, S., Materna, S.C., Samanta, M.P., Stolc, V., Tongprasit, W., Tu, Q., Bergeron, K.F., Brandhorst, B.P., Whittle, J., Berney, K., Bottjer, D.J., Calestani, C., Peterson, K., Chow, E., Yuan, Q.A., Elhaik, E., Graur, D., Reese, J.T., Bosdet, I., Heesun, S., Marra, M.A., Schein, J., Anderson, M.K., Brockton, V., Buckley, K.M., Cohen, A.H., Fugmann, S.D., Hibino, T., Loza-Coll, M., Majeske, A.J., Messier, C., Nair, S.V., Pancer, Z., Terwilliger, D.P., Agca, C., Arboleda, E., Chen, N., Churcher, A.M., Hallbook, F., Humphrey, G.W., Idris, M.M., Kiyama, T., Liang, S., Mellott, D., Mu, X., Murray, G., Olinski, R.P., Raible, F., Rowe, M., Taylor, J.S., Tessmar-Raible, K., Wang, D., Wilson, K.H., Yaguchi, S., Gaasterland, T., Galindo, B.E., Gunaratne, H.J., Juliano, C., Kinukawa, M., Moy, G.W., Neill, A.T., Nomura, M., Raisch, M., Reade, A., Roux, M.M., Song, J.L., Su, Y.H., Townley, I.K., Voronina, E., Wong, J.L., Amore, G., Branno, M., Brown, E.R., Cavalieri, V., Duboc, V., Duloquin, L., Flytzanis, C., Gache, C., Lapraz, F., Lepage, T., Locascio, A., Martinez, P., Matassi, G., Matranga, V., Range, R., Rizzo, F., Rottinger, E., Beane, W., Bradham, C., Byrum, C., Glenn, T., Hussain, S., Manning, G., Miranda, E., Thomason, R., Walton, K., Wikramanayke, A., Wu, S.Y., Xu, R., Brown, C.T., Chen, L., Gray, R.F., Lee, P.Y., Nam, J., Oliveri, P., Smith, J., Muzny, D., Bell, S., Chacko, J., Cree, A., Curry, S., Davis, C., Dinh, H., Dugan-Rocha, S., Fowler, J., Gill, R., Hamilton, C., Hernandez, J., Hines, S., Hume, J., Jackson, L., Jolivet, A., Kovar, C., Lee, S., Lewis, L., Miner, G., Morgan, M., Nazareth, L.V., Okwuonu, G., Parker, D., Pu, L.L., Thorn, R., Wright, R., 2006. The genome of the sea urchin *Strongylocentrotus purpuratus*. Science 314, 941–952.

Su, Y.H., Li, E., Geiss, G.K., Longabaugh, W.J., Kramer, A., Davidson, E.H., 2009. A perturbation model of the gene regulatory network for oral and aboral ectoderm specification in the sea urchin embryo. Dev. Biol. 329, 410–421.

Swalla, B.J., Smith, A.B.., 2008. Deciphering deuterostome phylogeny: molecular, morphological and palaeontological perspectives. Philos. Trans. R. Soc. Lond. B Biol. Sci. 363, 1557–1568.

Tagawa, K., Nishino, A., Humphreys, T., Satoh, N., 1998. The spawning and early development of the Hawaiian acorn worm (Hemichordate), *Ptychodera flava*. Zool. Sci. 15, 85–91.

Valenta, T., Hausmann, G., Basler, K., 2012. The many faces and functions of beta-catenin. EMBO J. 31, 2714–2736.

Wikramanayake, A.H., Hong, M., Lee, P.N., Pang, K., Byrum, C.A., Bince, J.M., Xu, R., Martindale, M.Q., 2003. An ancient role for nuclear beta-catenin in the evolution of axial polarity and germ layer segregation. Nature 426, 446–450.

Wikramanayake, A.H., Peterson, R., Chen, J., Huang, L., Bince, J.M., McClay, D.R., Klein, W.H., 2004. Nuclear beta-catenin-dependent Wnt8 signaling in vegetal cells of the early sea urchin embryo regulates gastrulation and differentiation of endoderm and mesodermal cell lineages. Genesis 39, 194–205.

Yu, J.K., Satou, Y., Holland, N.D., Shin, I.T., Kohara, Y., Satoh, N., Bronner-Fraser, M., Holland, L.Z., 2007. Axial patterning in cephalochordates and the evolution of the organizer. Nature 445, 613–617.

Zorn, A.M., Butler, K., Gurdon, J.B., 1999. Anterior endomesoderm specification in Xenopus by Wnt/beta-catenin and TGF-beta signalling pathways. Dev. Biol. 209, 282–297.

Please cite this article as: Chen, S.-H., et al., Sequencing and analysis of the transcriptome of the acorn worm Ptychodera flava, an indirect developing hemichordate, Mar. Genomics (2014), http://dx.doi.org/10.1016/j.margen.2014.04.010