Okinawa Institute of Science and Technology
Graduate University

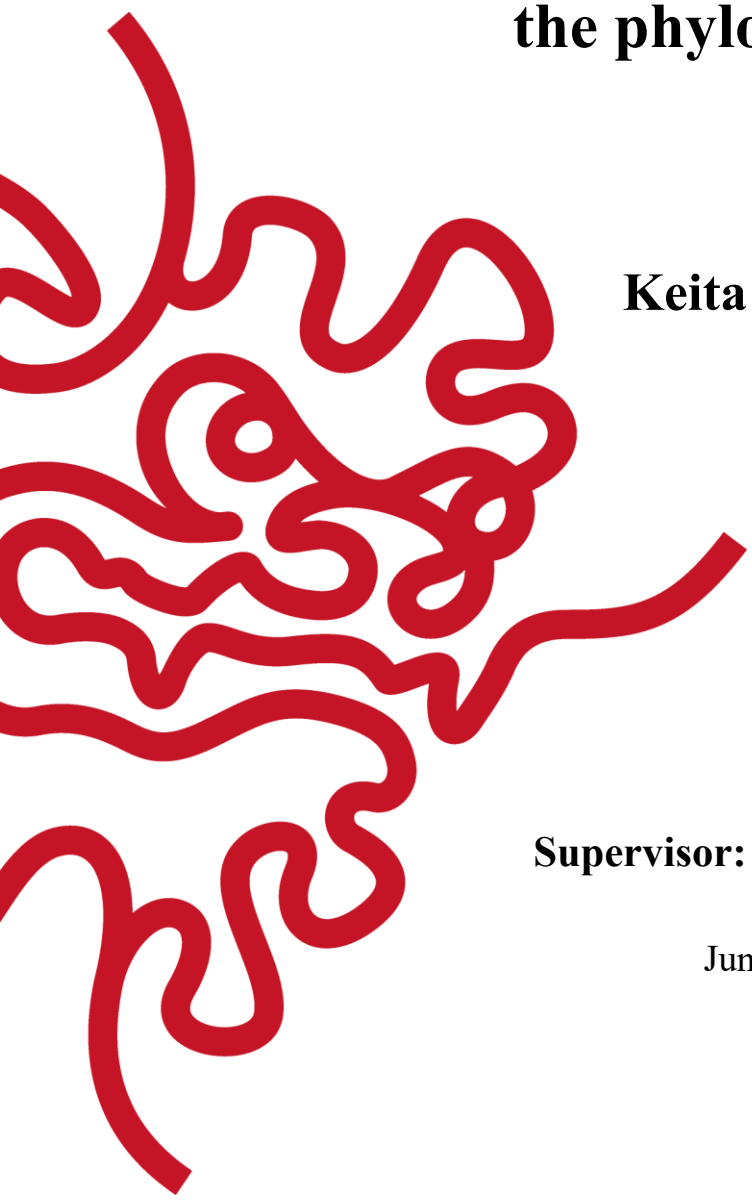
Thesis submitted for the degree

Doctor of Philosophy

Comparative transcriptome analysis of basal deuterostomes and its implications for the phylotypic stage

by

Keita Ikegami



Supervisor: Noriyuki Satoh

June, 2017

Declaration of Original and Sole Authorship

I, Keita Ikegami, declare that this thesis entitled “Comparative transcriptome analysis of basal deuterostomes and its implications for the phylotypic stage” and the data presented in it are original and my own work.

I confirm that:

- No part of this work has previously been submitted for a degree at this or any other university.
- References to the work of others have been clearly acknowledged. Quotations from the work of others have been clearly indicated, and attributed to them.
- In cases where others have contributed to part of this work, such contribution has been clearly acknowledged and distinguished from my own work.
- None of this work has been previously published elsewhere.

Signature:

Date: 20 June, 2017

Abstract

The modification of embryogenesis in animals is an interesting research subject in evolutionary developmental biology. The phylotypic stage was originally described as a developmental period shared across vertebrates, where embryos resemble each other, and is represented by a narrowing waist of the developmental "hourglass" model. Recent studies of comprehensive gene expression have found support for this model in other metazoans, plants, and fungus. However, the existence of the phylotypic stage remains unresolved in deuterostomes. Thus, this study examines whether the phylotypic stage can be found across four deuterostome taxa, namely *Ciona intestinalis* from the Phylum Urochordata, *Branchiostoma floridae* from the Phylum Cephalochordata, *Ptychodera flava* from the Phylum Hemichordata, and *Acanthaster planci* from the Phylum Echinodermata. Comparison of gene expression profiles was carried out by microarray analysis of RNA across predetermined developmental time points. In this thesis, I developed a method of microarray probe design in *A. planci,* and proposed guidelines for the technology for use with other marine invertebrates. Next, I compared the gene expression profile with this microarray method. I found that the overall gene expression somewhat resembled each other in *A. planci, P. flava* and *B. floridae,* while *C. intestinalis* exhibited a unique pattern. The gene expression profiles of *A. planci, P. flava* and *B. floridae* showed narrowing waist-like pattern from the blastula to gastrula stages up to early larval stages; these stages were more conserved during embryogenesis in deuterostome taxa. However, I failed to find evidence for a typical vertebrate-like phylotypic stage in the four deuterostome taxa. (250 words)

Acknowledgments

I acknowledge following persons for their suggestion and comments in preparation of this thesis.

Prof. Noriyuki Satoh: supervisor, and helpful suggestion and comments on the thesis.
Dr. Mayuko Hamada and OIST SQC staff members: teaching me the method of microarray analyses and helpful discussion on the results.
Drs. Jr-Kai Yu, Kunifumi Tagawa, Takeshi Noda, Koki Nishitsuji, Mayuko Hamada, Takeshi Takeuchi, Asuka Arimoto, Kenneth Baughman, and Eiichi Shoguchi: help in collection of samples.
Dr. Takeshi Kawashima and Kanako Hisata: suggestion for bioinformatics tools.
Members of Marine Genomics Unit: generous support for my research.

Contents

Abstract 3

Acknowledgments 4

1. Introduction 7

1.1. Developmentally relevant genes and comprehensive gene expression profiles 7

1.2. Deuterostomes and their phylogeny 7

1.3. Phylotypic stage and hourglass pattern 9

1.4. Objective 13

2. Materials and Methods 15

2.1. Sampling animals 15

2.2. RNA extraction 16

2.3. Microarray 16

2.4. Developmental transcriptome of basal deuterostomes 19

3. Results 22

3.1. Theoretical optimization of custom DNA microarray probes for non-model metazoans 22

3.1.1. Introduction 22

3.1.2. Materials and Methods 23

3.1.3. Results and Discussion 29

3.2. Comprehensive analyses of invertebrate deuterostome transcriptomes during embryogenesis 36

3.2.1. Echinoderms 36

3.2.2. Hemichordates 47

3.2.3. Cephalochordates 57

3.2.4. Urochordates 68

3.3. Comparative analysis of developmental transcriptome in four deuterostome taxa 78

4. Discussion 86

4.1. An improvement of microarray system in non-model animals 87

4.2. Characterization of gene expression profiles during embryogenesis of invertebrate deuterostomes 89

4.3. Comparative analyses of developmental gene expression profiles among deuterostome taxa 91
4.4. Future direction 94
5. Conclusion 95
References 96
Supplementary Data 107

1. Introduction

1.1. Developmentally relevant genes and comprehensive gene expression profiles

The study of comparison between developmental patterns and evolutionary relationship across taxa is evolutionary developmental biology or EvoDevo (Carroll et al., 2004; Davidson, 2006; Gilbert, 2013; Satoh, 2016). Genes encoding transcription factors and cell-cell signaling molecules, play essential roles in embryological development of animals. These genes were characterized as "tool-kit" genes (Carroll et al., 2004). It was discovered that the expression and function of tool-kit genes are shared by various groups of metazoans (Carroll et al., 2004; Gilbert, 2013). Although metazoans exhibit different modes of embryogenesis to form larvae and adults with different morphology, the exploration of the evolution of metazoans by molecular developmental biological methods became possible (EvoDevo).

Technical advances in molecular biology and genome science have contributed to advances in the study of metazoan evolution. In order to study how a global change in gene expression is involved in development, more quantitative analyses of gene expression are essential. Microarray and RNA-seq are methods for examining gene expression profile. These techniques can be used not only to compare global changes of gene expression across embryogenesis of a given organism, but can also be used to compare gene expression profiles between different animals with different modes of embryogenesis.

1.2. Deuterostomes and their phylogeny

The Superphylum Deuterostomia comprises three phyla, Echinodermata, Hemichordata, and Chordata (Fig. 1.1) (Brusca and Brusca, 2003; Ruppert et al., 2004; Nielsen, 2012; Brusca et al., 2016). However, a recent study suggested that deuterostomes comprise two major groups, the Ambulacraria and the Chordata. The former includes two phyla, Echinodermata and Hemichordata, while the latter comprises three phyla: Cephalochordata, Urochordata (Tunicata) and Vertebrata (Satoh et al., 2014a) (Fig. 1.1). Echinodermata contains five extant classes, Crinoidea (sea lilies), Asteroidea (starfish), Ophiuoidea (brittle stars), Echinoidea (sea urchins), and Holothuroidea (sea cucumbers). Hemichordata consists of two extant classes, Pterobranchia and Enteropneusta (acorn worms).

At lower taxonomic levels, the Cephalochordata, or lancelets, comprise about 30 species, represented by *Branchiostoma*. Urochordata (Tunicata) is a distinct phylum of about 3000

described species, consisting of three classes, Ascidiacea, Thaliacea and Appendicularia. Vertebrata consists of two major groups, Agnatha and Gnathostomata; the former includes a single extant class Cyclostomata while the latter includes Chondrichthyes, Osterichthyes, Amphibia, Reptilia, Aves, and Mammalia (Kardong, 2009). Recently, *Xenoturbella* and acoels have been proposed to be a newly recognized phylum assigned to the Deuterostomia (Philippe et al., 2009; Nakano et al., 2013). However, the phylogenetic position of the new grouping proposed as the Xenacoelomorpha is still controversial; therefore, the present study of deuterostome evolution does not include this animal group.

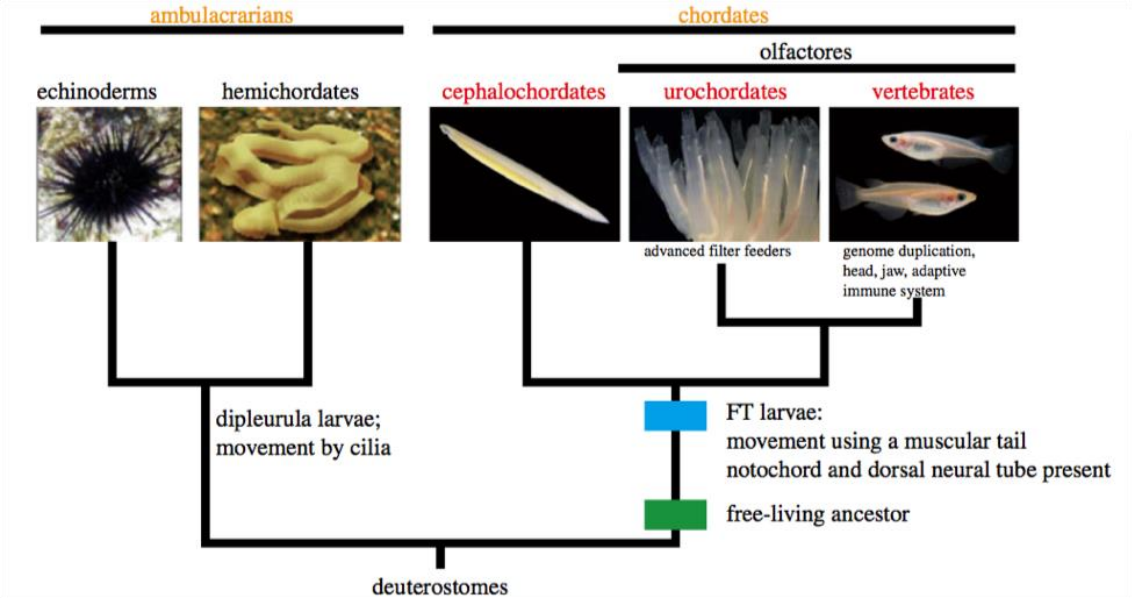


Figure 1.1: Phylogenetic relationships of deuterostomes

Schematic representation of deuterostome groups. Representative developmental events associated with the evolution of deuterostomes are included [modified from Satoh et al. (2014a)].

Although there are several exceptions, ambulacrarians and chordates share so-called deuterostomicity, which includes radial cleavage, a coelomic cavity formed from mesoderm, and pharyngeal gills (Willmer, 1990; Brusca and Brusca, 2003; Nielsen, 2012; Brusca et al., 2016). In addition, echinoderms have a distinctive hard exoskeleton made of calcium-carbonates. Urochordates are only deuterostomes that can synthesize cellulose (Satoh, 2009). Vertebrates have evolved hard bones and an adaptive immune system (Kardong, 2009; Satoh et al., 2014a). In addition, chordates invariably possess a notochord, a dorsal neural tube and somites (Satoh, 2008; Swalla and Smith, 2008; Satoh, 2011; Satoh,

2016).

Traditional metazoan phylogeny is based on modes of embryogenesis and similarities of larval and/or adult morphology, as observed in physiological characters and fossil record (Gee, 1996; Jefferies et al., 1996; Nielsen, 2012). On the other hand, molecular phylogeny is a recent method that can resolve phylogenic relationship of metazoans. Molecular phylogeny was first carried out on the basis of comparisons of 18S rDNA sequences (Wada and Satoh, 1994; Halanych, 1995), and later on protein-coding gene sequences (Dunn et al., 2008; Philippe et al., 2009). Qualitative traits of genes and genomes have given insights into various aspects of deuterostome relationships (Perseke et al., 2013). Recent studies of molecular phylogeny, comparative genomics, and evolutionary developmental biology have demonstrated that echinoderms and hemichordates form a clade, and cephalochordates, urochordates and vertebrates form another distinct clade (Halanych, 1995; Zeng and Swalla, 2005; Delsuc et al., 2006; Bourlat et al., 2006). The former is called the Ambulacraria and the latter, the Chordata. In addition, within the chordate clade, cephalochordates diverged first, and urochordates and vertebrates form a sister group (sometimes called Olfactors) (Gorman et al., 1971; Vandekerckhove and Weber, 1984; Delsuc et al., 2006; Putnam et al., 2008). This recent consensus view of deuterostome phylogeny appears to be robust, as a great variety of data from different disciplines support it. Based on these findings, a recent study has proposed that the Chordata should be recognized as a superphylum such as the Ambulacraria (Satoh et al., 2014a). The Cephalochordata, the Urochordata and the Vertebrata each should be elevated to the phylum level, as are the Echinodermata and Hemichordata.

1.3. Phylotypic stage and hourglass pattern

Charles Darwin proposed the evolution of animals by means of natural selection (Darwin, 1859). A central question of the evolution of deuterostomes is the origin of chordates (Romer, 1967; Tokioka, 1971; Gee, 1996; Kusakabe et al., 1997; Hall, 1999; Cameron et al., 2000; Oda et al., 2002; Gerhart et al., 2005; Lacalli, 2005; Swalla, 2006; Swalla and Smith, 2008; Philippe et al., 2009; Satoh et al., 2012; Satoh et al., 2014b; Satoh, 2016). For the last 100 years, zoologists have framed development with evolution, as morphological alterations become embodied through embryonic developmental processes (Gee, 1996; Hall, 1999). In particular, the discovery of “tool-kit” genes during 1985-1995 has formed basis for the molecular approach to the study of animal evolution (Davidson, 2001; Carroll et al., 2004; Davidson, 2006). The tool-kit genes encode transcription factors or signal transduction

molecules that control cell differentiation and morphogenesis. These tool-kits are reused by animals with different body plans over and over again (Pearson et al., 2005). For example, Slack et al. (1993) proposed the concept of the zootype as an animal archetype in order to not only characterize animals on the basis of morphology, but also to establish that the spatiotemporal gene expression patterns of Hox cluster genes are conserved amongst metazoans. Later, this idea was weakened when a disorganized Hox cluster was found in cnidarians (Martinez et al., 1998). Thus, molecular mechanisms that control animal development and evolution have remained a central focus in EvoDevo (evolutionary developmental biology), as mentioned before.

Kalinka and Tomancak (2012) proposed four models of embryonic conservation in the evolution of early animal embryos (Fig. 1.2):

(a) Early conservation (Fig. 1.2a)

In this model, the earliest developmental stages are considered foundational, and any apparent conservation in later stages is the delayed realization of the conservation of genes and proteins acting early (Richardson, 1999; Comte et al., 2010).

(b) Hourglass model (Fig. 1.2b)

Conservation is considered greatest in mid-embryogenesis and is either the result of the need for coordination between growth and patterning when the body plan is being built (Duboule, 1994), or the result of a global increase in the complexity of interactions between genes and processes during the phylotypic period (Raff, 1996).

(c) Adaptive penetrance (Fig. 1.2c)

This model posits that the most important beneficial mutations are likely to occur during the phylotypic period precisely because this is when the body plan is established (Richardson, 1999).

(d) Ontogenetic adjacency (Fig. 1.2d)

This model posits that small changes are most likely between events that are adjacent in the developmental sequence of events (Poe and Wake, 2004).

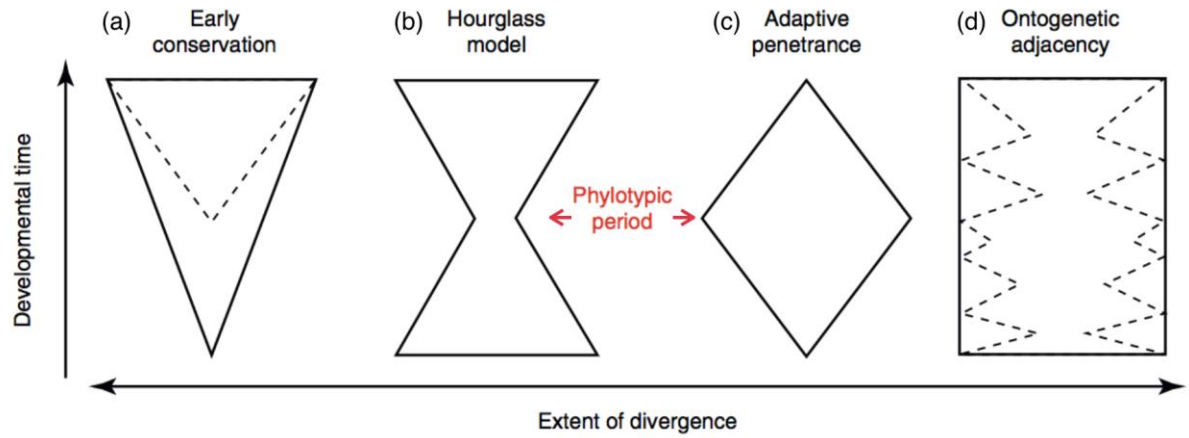


Figure 1.2: Four patters of embryonic conservation

A schematic representation to compare four different models that posit different patterns of conservation during animal embryogenesis. In all models, development from egg to adult is shown on the y-axis, and evolutionary divergence is represented on the x-axis (adopted and modified from Kalinka and Tomancak (2012)).

The question of embryonic conservation pattern has been approached using transcriptome analysis. For example, in vertebrates, one feature of the phylotypic stage is the "developmental hourglass" model, which represents high diversity (*i.e*. low similarity) during early and late embryogenesis and low diversity (high similarity) during mid-embryonic stages (Duboule, 1994; Irie and Kuratani, 2014). Recent transcriptomic analyses have supported this model, in which the pharyngula stage during mid-embryonic stages is suggested to be the phylotypic stage of vertebrates including zebrafish, *Xenopus*, turtle, chick, and mice (Irie and Kuratani, 2011; Wang et al., 2013) (Fig. 1.3). The pharyngula stage is characterized by formation of a head, pharyngeal arches, somites, a neural tube, epidermis, kidney tubules and longitudinal kidney ducts (but no metanephros), a heart with chambers, at least a transient cloaca, no middle ear, no gills on the pharyngeal segments, no tongue, and no penis or uterus (Irie and Kuratani, 2011).

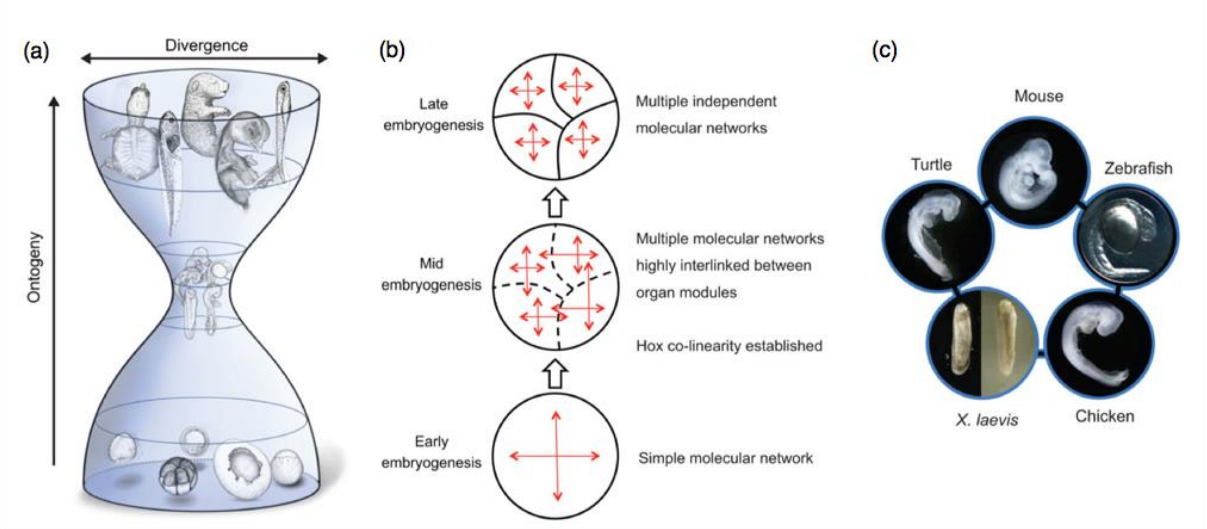


Figure 1.3: Schematic view of the developmental hourglass model

(a) The developmental hourglass model predicts that mid-embryonic organogenesis stages (phylotypic period) represent the period of highest conservation. The phylotypic period is the source of the basic body plan at the phylum level (Irie and Kuratani, 2011). (b) Hourglass-like divergence has been proposed to result from the spatiotemporal co-linearity of Hox cluster genes (Duboule, 1994), from the existence of highly interdependent molecular networks at the phylotypic stage (Raff, 1996). (c) Potential phylotypic period for vertebrates. Two stages of *X. laevis* are shown, as there was no statistically significant difference between these two stages (adopted and modified from Irie and Kuratani (2014)).

The developmental hourglass model has been explored not only in vertebrates, but also in *Drosophila* (Kalinka et al., 2010), *Caenorhabditis* (Levin et al., 2012) and trochozoans (Xu et al., 2016). Moreover, the model has also been investigated in plants (Quint et al., 2012; Drost et al., 2016) and fungi (Cheng et al., 2015).

1.4. Objective

In this thesis, the term "phylotype" is used to refer to a characteristic features of morphology shared by vertebrate embryos. In the developmental hourglass model, this has been proposed to be the pharyngula stage during mid-embryonic stages, which is the formation of a head, pharyngeal arches, somites, a neural tube, epidermis, kidney tubules and other structures (Irie and Kuratani, 2011). The vertebrate hourglass model, or the presence of a phylotypic stage, has been characterized by recent transcriptomic analyses as a waist-like narrowing of gene expression profile or constraint of gene expression (Fig. 1.3). Although the pharyngula stage is not present in bilaterians other than vertebrates, a waist-like narrowing of gene expression profile has been demonstrated in *Drosophila* (Kalinka et al., 2010), *Caenorhabditis* (Levin et al., 2012) and trochozoans (*i.e.* the Pacific oyster, the Pacific abalone and sand worm) (Xu et al. (2016). Constraints of the gene expression profile have been found in several protostomes, supporting the extension of the idea of developmental hourglass model to invertebrates (Kalinka et al., 2010; Levin et al., 2012; Xu et al., 2016).

I am interested in the evolution of deuterostomes and the origin of chordates. Thus, the objective of this study is to examine whether the developmental hourglass model is also applicable to four deuterostome taxa; namely, cephalochordates, urochordates, hemichordates and echinoderms. To this end, I performed transcriptomic analysis by microarray, and data obtained were analyzed by bioinformatics tools to address two questions.

(1) I hypothesize that the developmental hourglass model with a vertebrate-like phylotypic stage is applicable to all deuterostome groups and will be supported by gene expression profiles. The morphological resemblance between cephalochordate and urochordate embryos and vertebrate embryos suggests that the pharyngula-like stage of cephalochordate and urochordate embryos are the late-tailbud stage or early larval stage. If so, does the tailbud stage or early larval stage show a waist-like narrowing in the global gene expression profile? On the other hand, superficially, the vertebrate pharyngula-like stage is not present in ambulacrarian embryos. However, if there is a stage in the gene expression profiles with

waist-like narrowness, is it related to the vertebrate pharyngula-like stage?
(2) I propose another idea in which a stage in the gene expression profiles with waist-like narrowness is not directly associated with the so-called vertebrate pharyngula-like stage. According to this idea, single or multiple occurrences of constraints are expected in all four deuterostome taxa. Are the embryonic stages with the gene expression constraints shared between cephalochordates and urochordates, or between echinoderms and hemichordates, or among the four deuterostome taxa? Or does each taxon show independent constraint patterns? The constraint may suggest which embryological stages contain developmental events related to the evolution of each taxon.

The phylotypic stage has been explained by comparing the similarity of gene expression profiles, and therefore can be examined by developmental transcriptome analyses (Irie and Kuratani, 2011). Computational tools to deduce the phylotypic stage are classified into two methods; namely, ortholog-based and tree-based. Both methods are used on the transcriptomic dataset. The ortholog-based method calculates a similarity in transcriptomes of orthologous genes from multiple organisms (Kalinka et al., 2010; Irie and Kuratani, 2011; Levin et al., 2012; Gerstein et al., 2014; Levin et al., 2016). On the other hand, the tree-based method refers to conserved positions of genes in phylogeny, and thus requires single organisms (Domazet-Loso and Tautz, 2010; Quint et al., 2012; Drost et al., 2016; Xu et al., 2016). In this thesis, I compare evidence for the phylotypic stage from both ortholog-based and tree-based methodologies.

2. Materials and Methods

In order to obtain transcriptomes of embryos and larvae at different developmental stages, microarray experiments were performed for four deuterostome taxa, namely, *Ciona intestinalis* as a representative of urochordates, *Branchiostoma floridae* as a representative of cephalochordates, *Ptychodera flava* as a representative of hemichordates, and *Acanthaster planci* as a representative of echinoderms. DNA microarray experiments included sampling of materials, RNA extraction and quality assessment, fluorophore labeling, DNA microarray probe design, hybridization and microarray data processing. Table 2.1 summarizes the procedures carried out in this study.

Table 2.1 A summary of microarray experiments carried out in this thesis

	Sampling	Staging/ RNA extraction	Microarray design	Hybridization/ Data process
Ciona intestinalis	This study (NBRP)	This study	-	This study
Branchiostoma floridae	-	-	-	This study
Ptychodera flava	This study	This study	This study	This study
Acanthaster planci	This study	This study	This study	This study

2.1. Sampling animals

(a) *Acanthaster planci*: *A. planci* adults were collected from a wild population at Onna, Okinawa, Japan. Maturation of eggs was artificially induced in the laboratory and fertilization was performed following a method described by Shoguchi et al. (2000). Embryos and larva were kept in a Petri dish or in glass beakers, through which seawater was gently passed. Temperature was controlled at 28˚C with a 12-hour light cycle. Before sampling, timing of normal embryonic stages was examined. Observations under stereoscopic microscope were carried out by referring to morphologies and characteristics of embryos and larvae described by previous studies (Henderson, 1969; Hayashi et al., 1973; Lucas, 1982; Chia et al., 1993). Replicates were obtained in the range from 1-cell stage embryos to bipinnaria larvae from a single batch.

(b) *Ptychodera flava:* *P. flava* adults were collected from wild populations at Paiko, Hawaii, U.S.A., and Onna and Motobu, Okinawa, Japan. Spawning was artificially induced

in the laboratory and fertilization was performed following a method described by a previous study (Lowe et al., 2004). Embryos and larva were kept in glass beakers at room temperature (~25˚C). No antibiotics were used for development. Developmental stages were identified based on descriptions of Tagawa et al. (1998). Replicates were obtained in the range from 1-cell stage embryos to tornaria larvae from a single batch.

(c) *Ciona intestinalis:* *C. intestinalis* adults were provided from the National BioResource Project (Sasakura et al., 2009; Yamazaki et al., 2010). They were kept in tanks filled with seawater at 18˚C until used for experiment. Eggs and sperms were surgically extracted from individuals and fertilization was performed according to a method described by Hamada et al. (2011). No antibiotics were used for development. Developmental stages were identified based on description of Hotta et al. (2007). After fertilization, the chorion was removed. Replicates were obtained from single batch in the range from 1-cell stage embryos to pre-juveniles at late-body-axis-rotation stage.

Samples of *Branchiostoma floridae* were kindly provided by Dr. Jr-Kai Yu at Academia Sinica, Taiwan.

2.2. RNA extraction

At each developmental stage, 20, 50, and 100 embryos (or larvae) were collected for *C. intestinalis, P. flava,* and *A. planci,* respectively. Specimens were dissolved into TRIzol Reagent (Thermo Fisher Scientific) and stored at -80˚C until used. Total RNAs were extracted and subsequently cleaned up using RNeasy Micro Kit (Qiagen) together with DNase (Qiagen) treatment. The quality was assessed by 2100 Bioanalyzer (Agilent Technologies) using RNA 6000 Pico kit (Agilent Technologies) and quantified using a NanoDrop microscale spectrophotometer (NanoDrop Technologies).

2.3. Microarray

2.3.1. Probe design

The overall method of the microarray probe design, summarized in Fig. 3.1, included (a) gene model preparation, (b) probe optimization, and (c) custom DNA microarray production. In this study, the probe design for *A. planci,* was improved as a case study (see details in Result 3.1 section). The same methodology was applied to *P. flava* gene models to design optimized probes.

2.3.2. Microarray hybridization and data process

(a) Platforms and gene models

Agilent microarray platform systems used for hybridizing RNA samples of target animals (Table 2.2) were as follows:

C. intestinalis, GPL5576 (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GPL5576);
B. floridae, GPL23316 (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GPL23316);
P. flava, GPL23317 (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GPL23317); and
A. planci, GPL23315 (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GPL23315).

The *B. floridae* gene model is available at the RefSeq (NCBI Accession: PRJNA33245). The *P. flava* and *A. planci* gene model are available at the Genome Projects, OIST Marine Genomics Unit (http://marinegenomics.oist.jp). For *A. planci* microarray, the model obtained from a specimen from the Great Barrier Reef (GBR) in Australia and one from Okinawa (OKI), Japan, were optimally merged together to obtain 28,380 gene models (described in Results 3.1.3). Originally, the GBR model contains 24,747 gene models for mRNA and protein and the OKI model contains 24,323 gene models for mRNA and protein, respectively. Low Input Quick Amp Labeling Kit for one-color (Agilent Technologies) was used for cyanine3 labeling. The hybridization was carried out following the manufacture's protocol. Exceptions to the manufacture's protocol were made in the start concentration of cyanine3-labeled cRNA in *P. flava* and *A. planci* samples (see following paragraph).

Table 2.2: Target sequences in microarrays

	DNA microarray # target sequence	Gene model mRNA/Protein
Ciona intestinalis	19,964	CIYS model*
Branchiostoma floridae	50,694	50,817
Ptychodera flava	34,601	34,647
Acanthaster planci	28,376	28,380

* CIYS model is composed of 19,889 mRNAs and 20,189 proteins

(b) Validations of start concentration of cy3-cRNA for hybridization

In the manufacture's protocol, 1650 ng of cy3-labeled cRNA was recommended for hybridization. However, due to the small amount of total RNA, the yield of cy3-cRNA was lower than that of 1650 ng in some samples of *P. flava* and *A. planci*. The protocol recommended standardizing start cy3-RNA concentration in the same platform. Thus, to find compatible amount of start cy3-cRNA concentration to normal procedure, an experiment was conducted to verify whether 1100 ng and 825 ng of cy3-cRNA were adequate in *P. flava* and *A. planci*, respectively. Raw spot signals were compared by Pearson's correlation and Spearman's rank correlation. In *P. flava*, start cy3-cRNA concentrations of 1650 ng, 825 ng, and 600 ng were compared using samples from the late gastrula stage (batch #16). In *A. planci*, start cy3-cRNA concentration of 1650 ng and 825 ng were compared using samples from the unfertilized egg stage (batch #4) and the late gastrula stage (batch #3).

(c) Normalization of microarray signal intensity

Quantile normalization was applied to probe intensities between samples using the Limma package in R (Ritchie et al., 2015). Normalization was separately performed for each target organism.

(d) Acquisition of full length target sequence for the Ciona intestinalis microarray

In the *C. intestinalis* platform, a 60-mer probe sequence was not assigned to a full-length sequence in public databases. A probe has a unique ID starting from "CIYS" so that this set of gene models used in the microarray platform was named as CIYS model (Table 2.2). In this study, full-length mRNA and protein sequences of the CIYS model was determined. To obtain full length mRNA, 60-mer probes were blasted against all datasets in ANISEED (Tassy et al., 2010). Then, full-length mRNA sequence was blasted against protein sequences of *C. intestinalis* in the RefSeq (O'Leary et al., 2016) and Ensembl (Aken et al., 2016). Those full-length mRNA sequences which were not assigned to any of RefSeq or Ensembl were translated into amino acid sequences using transdecoder (Haas et al., 2013).

(e) Different versions of *Branchiostoma floridae* genome assembly

The *B. floridae* genome assembly has two versions, ver-1 and ver-2. *B. floridae* microarray probes were designed based on the ver-1 assembly. In this thesis, a microarray platform based on the ver-1 was used to perform transcriptomics. However, *B. floridae* genome assembly ver-2 was used as target sequences of probes. Full-length mRNA from ver-1 was blasted against ver-2. Protein sequences were obtained from genome assembly ver-2 gff file.

(f) Preparation of developmental transcriptome

Developmental transcriptome was prepared by following steps: (1) probe signal intensity was filtered by base call from the Agilent Feature Extraction. If the flag "gIsFeatNonUnifOL" showed "1" in the Agilent Feature Extraction then the signal intensity was discarded. (2) Probe replicates were merge into mean of a target gene model. In general, several probes were designed from different positions of a gene model and then the set of probes was replicated several times. Here, probes sharing the same target gene model were averaged. (3) Stage replicates were merged into a single mean of a stage.

2.4. Developmental transcriptome of basal deuterostomes

In order to describe overall characteristics of transcriptomes, hierarchal clustering and principle component analysis (PCA) was performed. Gene expression profiles and sampled developmental stages were clustered based on result of the hierarchal clustering. Statistically significant associations between gene expression profiles and functions were identified using gene ontology (GO) enrichment analysis. Then, cross-comparisons of transcriptome among target organisms were performed to investigate conserved transcriptomic pattern. Tree-based analysis and ortholog-based analysis were used.

(a) Hierarchical clustering

Dynamic expression of genes was grouped into 12 arbitrary gene expression patterns according to hierarchical clustering (*e.g.* Fig. 3.8). Similar gene expression patterns were sorted close to each other, which are believed to share related regulatory system (D'Haeseleer et al., 2000; Wagner et al., 2007). A developmental transcriptome was filtered by relative expression change, defined by the ratio of maximum signal intensity to minimum signal intensity. Note that relative gene expression change was calculated using normalized signal intensity, rather than signal intensity scaled by log2. Gene expression changes >= 2-fold were used for hierarchical clustering. Distance matrix was calculated by the Euclidean distance and clustered by the complete-linkage clustering using R. Heatmap was visualized using the pheatmap package in R. Dendrogram for gene expression profiles and sampled stages were obtained independently.

(b) Gene ontology enrichment analysis

Both Gene ontology (GO) and the Generic GO slim were obtained from the Gene Ontology Consortium (Ashburner et al., 2000; Gene Ontology Consortium., 2015). GO was annotated to a target gene model by Swiss-Prot and TrEMBL in Universal Protein Resource (UniProt)

databases (The UniProt Consortium., 2017). Swiss-Prot is a manually-curated database and TrEMBL is a computationally-annotated database. A target gene model was first blasted against Swiss-Prot (BLASTP, p-value < 1e-1). If a gene model did not match any of the Swiss-Prot sequences, the gene model was subsequently scanned against TrEMBL (BLASTP, p-value < 1e-1). Top hits of the BLAST search in Swiss-Prot or TrEMBL were considered to be homologous protein sequences of a gene model. Then, using annotation of either SwissProt or TrEMBL, a target gene model was linked to GO. GO enrichment and GO slim mapping was performed using Goatools (Tang et al., 2015).

Following this, subset genes were clustered based on gene expression profile and developmental stage, respectively. A subset of genes from gene expression profiles was extracted from the dendrogram by hierarchical clustering. A subset gene for a stage-specific expression was defined by Student's *t*-test. At a target gene model, the averaged whole gene expression was compared to gene expression at each stage. If a single stage was found to be significant among all stages, the significant stage was considered as stage-specific gene expression.

(c) Ortholog-based analysis

A set of single copy orthologs was used as orthologous gene. Ortholog gene families were identified using default parameters of the OrthoMCL (Li et al., 2003). Pairwise cross-comparison was performed between stages of different organisms using ortholog genes. Conservation of transcriptomic pattern was evaluated by the Spearman's rank correlation coefficient *rho*. For higher values of *rho*, both compared stages were considered to conserve similar gene expression profile.

(d) Tree-based analysis

First, genomic phylostratigraphy of target organisms was calculated using software available on the Internet (https://github.com/AlexGa/Phylostratigraphy). Default parameters were used for calculations, and the provided protein sequences ("phyloBlastDB_Drost_Gabel_Grosse_Quint.fa") were used for BLAST database. The developmental transcriptome was then incorporated into genomic phylostratigraphy to obtain a transcriptome age index (TAI) score using myTAI (Drost et al., 2015). TAI was calculated for each stage of a given target organism. A higher score indicated the expression of younger genes while a lower score indicated the expression of older genes, as described by Domazet-Loso and Tautz (2010). For calculation of vertebrate TAI (*i.e.* mouse, chick,

African clawed frog, zebrafish), developmental transcriptome data from Irie and Kuratani (2011) were used. Protein sequences were obtained from RefSeq and UniProt based on the Affymetrix Annotation File for each microarray platform used in each of the target vertebrate groups.

3. Results

3.1. Theoretical optimization of custom DNA microarray probes for non-model metazoans

3.1.1. Introduction

DNA microarray and deep sequencing technologies have made it possible to obtain high throughput quantification of transcriptomes (Morozova et al., 2009; Malone and Oliver, 2011). DNA microarrays and deep sequencing technologies can be characterized as biased and unbiased technologies, respectively. Microarrays are biased because they require prior knowledge of gene sequences. They can detect only known gene models, while sequencing is unbiased because it requires no prior knowledge of gene models and because it can be used to determine unknown sequences of genes and transcripts. Deep sequencing technology dominates genome-wide studies (Goodwin et al., 2016). However, for mid-scale screening targeting defined sequences, DNA microarray remains the method of choice due to economy of experimental costs and simplicity of experimental procedures and data analysis. If numbers of target sequences and specimens are too large to be examined by real-time PCR, but it is not necessary to detect all genes (as in an expression analysis of known gene groups along a time series or under various sets of conditions), microarray analysis is more suitable. In addition to gene expression analysis, microarrays are also useful in medical and environmental surveys examining many biomarkers in a large number of specimens, for example, in detection of SNPs in genotyping, cytogenetic analysis of chromosome aberrations, or in environmental monitoring to identify microbes, etc.

DNA microarrays use single stranded probes, which are hybridized with fluorescently labeled target transcripts to assess the intensity of gene expression. The outcome of microarray analysis is limited by probe quality in terms of specificity of hybridization, sensitivity of detection, uniformity in size and distribution of spots on the microarray, complete coverage of the target gene models, etc. Thus, improvements in probe quality lead to more reliable microarray results. Since microarrays have great capacity in terms of the total number of spots or probes, probe optimization for custom microarray design is recommended. Although DNA microarrays have been optimized for model organisms with well-defined gene models and well-documented alternative splice variants, microarrays for non-model organisms from wild populations require careful design.

Recent advances in deep sequencing technology for both genomic DNA and transcribed mRNAs have allowed the examination of gene expression profiles of non-model metazoans, such as ascidians and sea urchins. However, high levels of polymorphism in the genomes of non-model metazoans from wild populations make it difficult to correctly assemble genome sequences and gene models. Owing to accumulating sequence data, it has become much easier to obtain gene models and transcribed sequences for model and non-model organisms to use in custom DNA microarrays.

Here, a theoretical optimization of custom DNA microarray probes for the crown-of-thorns starfish (COTS) was attempted, as a case study of a non-model organism. COTS genomes from two different wild populations have been decoded (Hall et al., 2017). This study aimed to increase the number of target-specific hybridizing probes and consequently to increase the number of target gene models on the array. For this purpose, an Agilent Custom DNA Microarray based on SurePrint inkjet deposition of *in situ* synthesized 60-mer oligonucleotides using the eArray web interface (https://earray.chem.agilent.com/earray/) was applied to COTS. Subsequently, the method was compared with default methods provided by the Agilent eArray. The procedures discussed below were implemented as scripts and have been deposited in the GitHub repository (https://github.com/ikegamikeita/probe-optimization).

3.1.2. Materials and Methods

The overall method of the microarray probe design is summarized in Fig. 3.1, which includes (a) gene model preparation, (b) probe optimization, and (c) custom DNA microarray production.

(a) Gene model preparation

Two draft genomes of the COTS, *A. planci*, from different populations have been published, one from the Great Barrier Reef (Australia; GBR model) and the other from Okinawa (Japan; OKI model) (http://marinegenomics.oist.jp) (Hall et al., 2017). Since both gene models were from the same species albeit different populations, it was assumed that they would share similar gene families and possess population-specific singletons. In this study, the OKI model was used for an *A. planci* custom microarray, and part of the GBR model was used to supplement the custom microarray in order to fill in gene models that were not predicted in the OKI model.

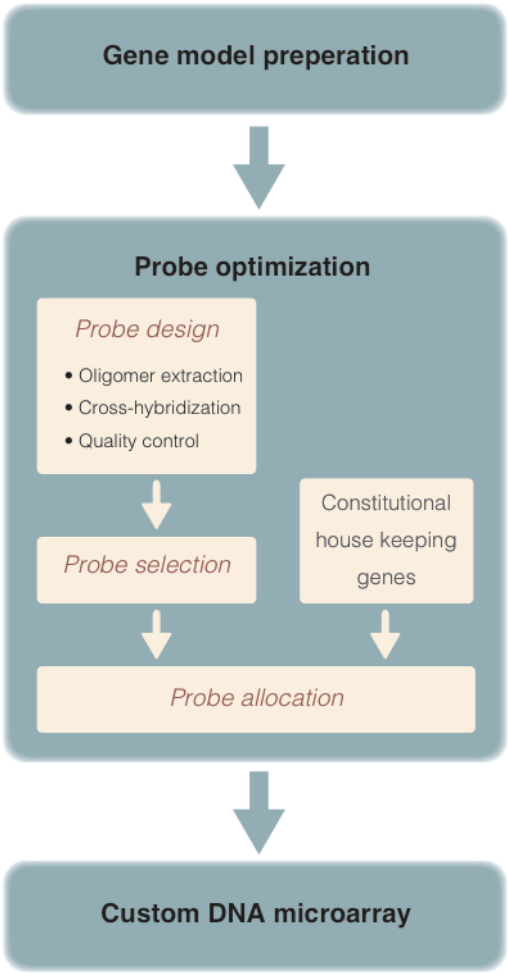


Figure 3.1: Schematic drawing of the custom DNA microarray design

Prior to probe optimization, gene model preparation determines efficient usage of the finite probe space of the custom DNA microarray. Probe optimization comprises three steps: (1) probe design, which includes information of cross-hybridization number and quality control to every oligomer sequence, (2) probe selection, which identifies a representative probe set for every gene model, and (3) probe allocation, which fills available probe space with representative probe sets while considering replicates. In addition, probes for constitutive house-keeping genes were added, in order to meet manufacturer's criteria.

OrthoMCL (version 2.0.4), which provides orthologous gene clusters, was used to identify common gene families shared among populations and population-specific singletons (Li et al., 2003). Default parameters of OrthoMCL were used to remove genetic redundancy from the GBR model by identifying singletons. A gene was judged to be redundant when it was assigned to a common gene family. Thereafter, the entire OKI model and singletons of the GBR model were combined to form the reference gene model of *A. planci* for this study. Completeness of the *A. planci* reference genome was evaluated using Benchmarking Universal Single-Copy Orthologs (BUSCO, version 2.0b4), which identify coverage of highly conserved metazoan orthologs using a "metazoan" dataset (Simao et al., 2015).

(b) Probe design

The objective of probe design is to extract oligonucleotide sequences from gene models and then to perform a cross-hybridization check and quality control for all oligomers. First, to define an oligomer within 1000 nucleotides (nt) of the 3'-end, 60-mer oligonucleotides were extracted one nucleotide at a time (Fig. 3.2a). A 60-mer oligonucleotide that is composed only of A, T, G, and C was kept, and oligomers that contain sequence including N or some other character were discarded.

Second, the cross-hybridization number of all oligomers was counted. All oligomers were mapped back onto the *A. planci* reference gene model sequences using bowtie2 with '-f --very-sensitive --norc --all --threads 12' parameters (Langmead and Salzberg, 2012) . We defined cross-hybridization as an oligomer that aligned to a non-target gene model ≤6 nucleotide mismatches, in which both gap open and gap extension were zero in the Sequence Alignment/Map (SAM) format. In other words, any oligomers that do not fit the cross-hybridization definition were categorized as non-cross-hybridizing oligomers (*i.e.* specific oligomers). This cross-hybridization definition theoretically guarantees >90% hybridization identity using 60-mer probes to target transcripts.

Third, quality control of all oligomer sequences was examined using the Probe Check function in the Agilent eArray, which provides a quality index for every oligomer sequence (*i.e.* BC1, BC2, BC3, BC4 and BC_poor; where BC1 is the highest quality and BC_poor the lowest). Subsequently, principle component analysis (PCA) was performed to characterize quality index according to multi-dimensional result of the quality control. Variables used for PCA were melting temperature, Tm, maximum repeated base number of single nucleotide, PolyX, and base content, GC%, for each probe sequence from the Probe Check function.

(c) Probe selection

The aim of probe selection is to identify a representative set of probes for a target gene model. All information about oligomers with cross-hybridization number and quality index was incorporated into probe selection. That is, the following options were employed for probe selection: (1) the range of sequence lengths from the 3'-end, (2) the size of the windows, (3) the maximum representative probe number from each target gene model, and (4) the maximum cross-hybridization number to allow for representative probe selection.

(1) The range of sequence lengths from the 3'-end specifies the range of sequence in which oligomers were effectively extracted. In regard to the use of polyT primer during cDNA synthesis, 1000 nt is recommended by Agilent.

(2) The size of the window defined the range of window where one of the representative probes was selected (Fig. 3.2b). A candidate probe was selected from a window. Thus, a sufficiently large window size reduced overlapping probe sequences.

(3) The maximum probe number is a limitation of the representative probe number selected from each target gene model. The maximum probe number was decided based upon the number of target gene models and the total number of available spots.

(4) The maximum cross-hybridization number specifies the acceptable number of cross-hybridizations. In this study, representative probes were primarily selected from specific hybridized oligomers. However, if the selected representative probes did not reach the limit of the defined probe number, additional representative probes were selected from cross-hybridizing oligomers on a secondary basis. All oligomers were classified according to the quality control index and the position of the extracted oligomer's 3'-end. For each window, one representative probe was selected based on criteria that prioritized the highest quality index and that was closest to the 3'-end (Fig. 3.2c).

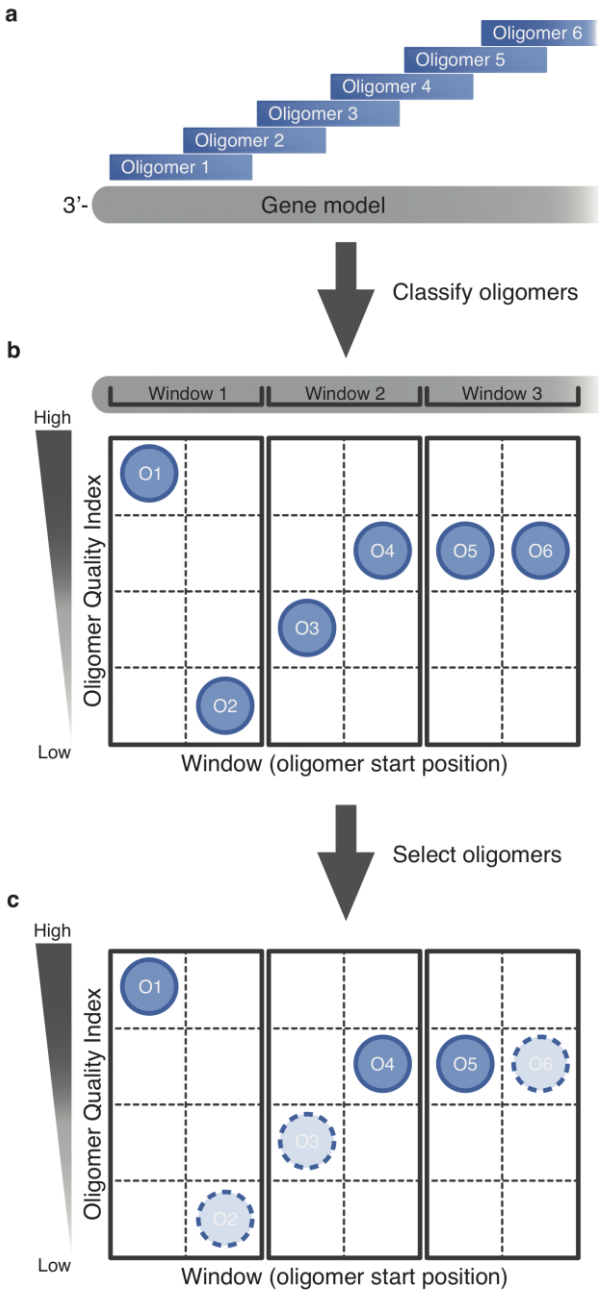


Figure 3.2: An example of how to select optimal probes for custom array design

(a) Gene models were divided into oligomers one nucleotide at a time. (b) An adjustable or arbitrary size window was set. Then, each oligomer was classified according to its oligomer quality indices, BC scores, and window based on the starting position of the oligomer. (c) A single representative oligomer was selected as a probe candidate from each window. Criteria were to choose oligomers with higher quality indices and the greatest proximity to the 3'-end. For example, in Window 1, oligomer 1 was selected as a representative probe because it had the highest quality index. In Window 2, oligomer 4 was selected as a representative probe because the quality index was considered higher priority than position. In Window 3, although both oligomers had equivalent index scores, oligomer 5 was selected as the representative probe because it was closer to the 3'-end of the gene.

(d) Estimation of constitutive house-keeping genes from the COTS reference gene models

The Agilent eArray recommends the inclusion of constitutive house-keeping genes. To this end, a list of reference genes for constitutive house-keeping genes was defined based on EST results from the NCBI UniGene database (Wheeler et al., 2003) on animals phylogenetically related to *A. planci* (*i.e.* deuterostomes), the list of loading genes used in TaqMan Array Human Endogenous Control (Thermo Fisher Scientific), and the recently-reported list of genes expressed continuously through early embryogenesis in *Xenopus tropicalis* (Owens et al., 2016).

Each gene was manually categorized according to KEGG Orthology (KO) (Kanehisa and Goto, 2000; Kanehisa et al., 2016). Corresponding orthologous genes in human, mouse, *Ciona intestinalis*, *Branchiostoma floridae*, *Strongylocentrotus purpuratus* and *Saccoglossus kowalevskii* were used to construct a protein sequence database of house-keeping genes (obtained June 27th, 2016). All *A. planci* reference gene models were searched against the database using BLASTP (Altschul et al., 1990). The following thresholds defined the possible house-keeping genes of *A. planci* that fulfilled either condition: (1) E-value < 1e-40 or (2) E-value < 1e-5 and alignment identity > 50%.

(e) Probe allocation

The aim of this section is to fill the available probe space with representative probe sets while considering replicates. Since the *A. planci* reference transcriptome was assembled *de novo* and its annotation was uncertain, the Custom Design Guidance for the Agilent eArray was modified to fit our experimental design. The modification included two instructions: “replicated non-control probes” and “probe set representing non-differentially expressed genes.” Thus, predicted *A. planci* constitutive house-keeping genes were used to perform ~10 technical replicates, which covered >1% of all probes.

In this study, Agilent’s SurePrint 180K custom microarray was chosen so that approximately 180,000 probes could be defined. Around 5,000 probes were necessary for the Agilent control probe set and more than 1% of all probe space (*i.e.* 1,800 probes) was used for constitutively-expressed genes. Since the *A. planci* genome contains 28,380 reference gene models, three complete replicates of representative probe sets were planned, which would adequately fill the remaining probe space. Agilent control was automatically assigned and

feature layout was randomized also automatically by the Agilent eArray.

(f) Evaluation of the custom-made DNA microarray

The DNA microarray designed by this method was evaluated by comparing it to the default Agilent eArray microarrays. The default eArray has two methods to make microarrays: the best probe methodology that prioritizes probe quality and the best distribution methodology that prioritizes an even distribution of probes along each target sequence. The following comparisons were performed: (1) complete coverage of target gene models, (2) the proportion of probe quality indices, BC scores, (3) space or a distance between two adjacent probes, and (4) proportion of hybridization of all probes. Using the three different methodologies (my method, Agilent's best probe method and the best distribution method), a full set of probes was constructed for probe quality comparison.

3.1.3. Results and Discussion

(a) Quality assessment of the *A. planci* cross-population gene models

The GBR and OKI models had 24,747 and 24,323 gene models, respectively. Using OrthoMCL, 18,584 common gene families, 4,057 singletons for the GBR model and 3,870 singletons for the OKI model were identified (Fig. 3.3a). The complete OKI model (86%) and singletons of the GBR model (14%) were combined to create a 28,380-reference gene model of *A. planci* (100%). This reference genome was assessed with BUSCO, which showed 88% completeness relative to a "metazoan" dataset.

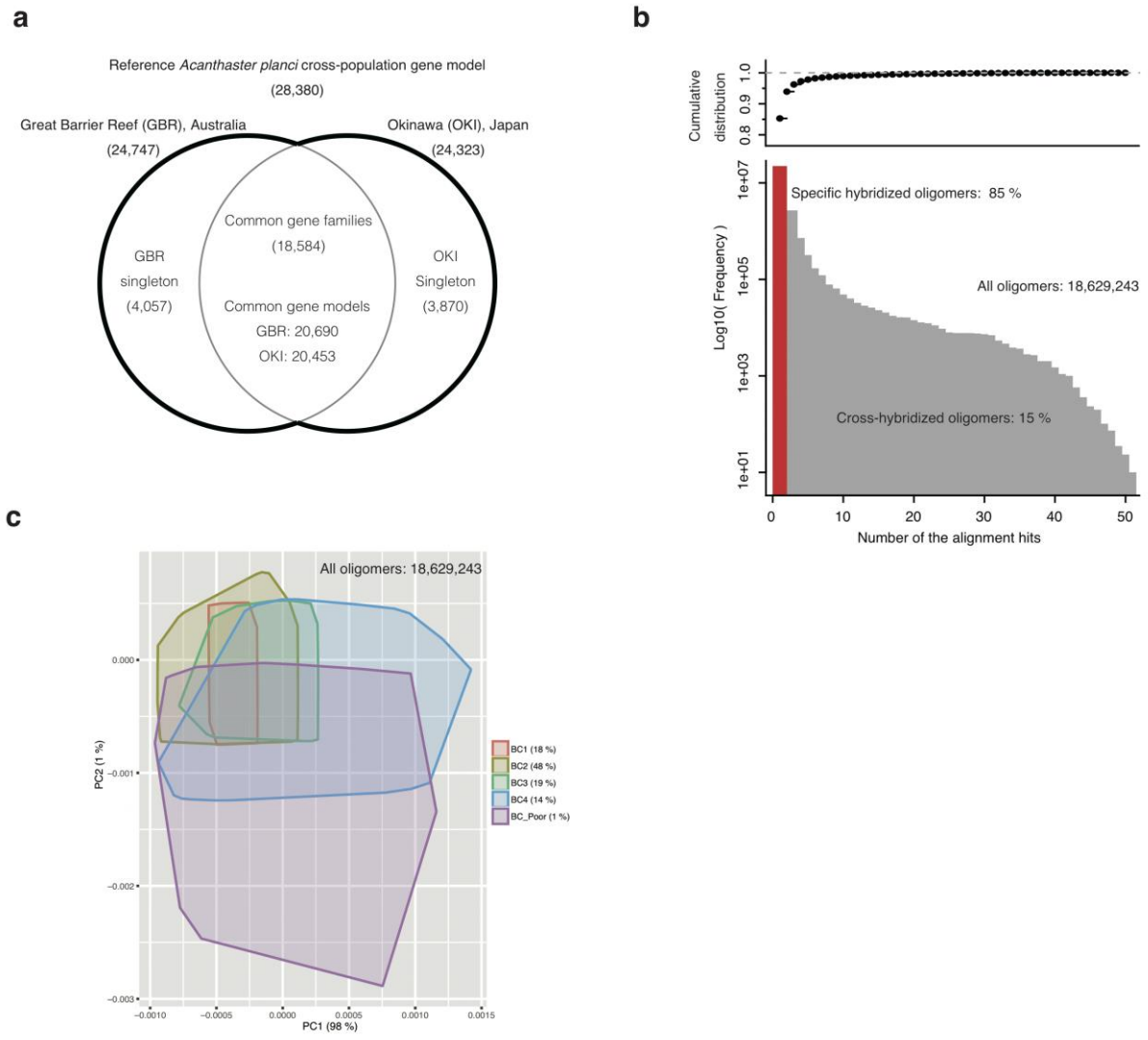


Figure 3.3: **Reference gene models and preparation for probe selection**

(a) Gene models from two different populations of *Acanthaster planci* were combined to form the reference models comprising 28,380 genes. (b) In order to gain oligomers thoroughly extracted from the reference gene model, cross-hybridizations to non-target gene models were checked. The lower panel shows a histogram of cross-hybridized probes with the vertical axis in log-scale. The upper panel shows the cumulative distribution of frequencies, starting from non-alignment (*i.e.* no cross-hybridization or specific hybridization; 85%). (c) the PCA result of quality control for all probes using the Probe Check function in the Agilent eArray. Variables used were melting temperature, Tm, maximum repeated base number of single nucleotides, PolyX, and base content, GC%.

(b) Optimization of a set of probes for *A. planci* reference gene models

Among the *A. planci* cross-population gene models, 18,629,243 oligomers were extracted, excluding oligomer sequences containing non-A, -T, -G or –C. On average, 657 +/- 292 (SD) oligomers were extracted from each gene model. The cross-hybridization number for every oligomer was determined using bowtie2. 85% of all oligomers showed specific hybridization whereas the remaining 15% had cross-hybridization (Fig. 3.3b).

Quality control was carried out using the Probe Check function in the Agilent eArray. The outcome of quality control was analyzed by PCA, which showed that 18% of all oligomers were BC1, the highest quality (Fig. 3.3c). Cumulatively, 66% of all oligomers were composed of BC1 and BC2. 85% were categorized as BC1-BC3, and 99% as BC1-BC4. The proportion of BC_Poor was low (1%). The range of PCA suggests strictness of required conditions of Tm, PolyX, and GC% for each BC score. All information about oligomers with their cross-hybridization numbers and quality indices were incorporated into the probe selection. A total of 84,922 oligomers were selected as representative probes for 28,376 of the 28,380 gene models. All four gene models where probes were not selected had Ns inserted into their sequences, which caused the remaining A, T, G and C sequences to be shorter than 60 nt.

(c) Hypothetical prediction of the constitutive house-keeping genes

All *A. planci* reference gene models were searched against the constitutive house-keeping gene database to obtain 238 potential house-keeping genes (Table 3.1). We could not obtain corresponding gene models in *A. planci* for ATP6, CDKN1B, or EIF2B1. For each predicted *A. planci* house-keeping gene model, 714 probes were assigned from among the optimized probe sets.

Table 3.1: The 37 constitutive genes and their candidate gene models in *A. planci*

KEGG ID	KEGG Name	# of gene models
K06619	ABL1, abelson tyrosine-protein kinase 1 [EC:2.7.10.2]	74
K05692	actin beta/gamma 1	20
K02126	ATPeF0A, MTATP6, ATP6; F-type H+-transporting ATPase subunit a	0
K14323	CASC3, MLN51; protein CASC3	1
K06624	CDKN1B, P27, KIP1; cyclin-dependent kinase inhibitor 1B	0
K11594	DDX3X, bel, ATP-dependent RNA helicase DDX3X [EC:3.6.4.13]	21
K12617	DNA topoisomerase 2-associated protein PAT1	2
K03231	EEF1A, elongation factor 1-alpha	4
K03234	EEF2, elongation factor 2	4
K03239	EIF2B1; translation initiation factor eIF-2B subunit alpha	0
K09428	ELF1_2_4; E74-like factor 1/2/4	6
K04402	GADD45, growth arrest and DNA-damage-inducible protein	1
K00134	GAPDH, gapA, glyceraldehyde 3-phosphate dehydrogenase [EC:1.2.1.12]	3
K11253	H3, histone H3	18
K01749	hemC, HMBS; hydroxymethylbilane synthase [EC:2.5.1.61]	2
K11405	histone deacetylase 8 [EC:3.5.1.98]	3
K00760	hprT, hpt, HPRT1, hypoxanthine phosphoribosyltransferase [EC:2.4.2.8]	1
K14843	PES1, NOP7; pescadillo	1
K00927	PGK, pgk; phosphoglycerate kinase [EC:2.7.2.3]	3
K03538	POP4, RPP29; ribonuclease P protein subunit POP4 [EC:3.1.26.5]	1
K03063	PSMC4, RPT3; 26S proteasome regulatory subunit T3	15
K17943	PUM; pumilio RNA-binding family	2
K08773	RALBP1, RalA-binding protein 1	1
K02884	RP-L19, MRPL19, rplS; large subunit ribosomal protein L19	1
K02907	RP-L30, MRPL30, rpmD; large subunit ribosomal protein L30	1
K02908	RP-L30e, RPL30, large subunit ribosomal protein L30e	1
K02921	RP-L37Ae, RPL37A; large subunit ribosomal protein L37Ae	1
K02941	RP-LP0, RPLP0; large subunit ribosomal protein LP0	1
K02962	RP-S17e, RPS17; small subunit ribosomal protein S17e	1
K02997	RP-S9e, RPS9, small subunit ribosomal protein S9e	2
K03006	RPB1, POLR2A; DNA-directed RNA polymerase II subunit RPB1 [EC:2.7.7.6]	4
K03108	signal recognition particle subunit SRP72	2
K03120	TBP, tbp; transcription initiation factor TFIID TATA-box-binding protein	2
K07375	TUBB, tubulin beta	30
K08770	UBC; ubiquitin C	3
K01195	uidA, GUSB; beta-glucuronidase [EC:3.2.1.31]	1
K16197	YWHAB_Q_Z, 14-3-3 protein beta/theta/zeta	5

(d) Comparison of the custom-made DNA microarray and the default eArray-generated microarray

In the present study, a total of 84,922 probes for reference gene models and 714 probes for constitutive house-keeping genes were selected, respectively. They completed more than 99% of all gene models (Table 3.2). Among the 84,922 probes, 96% of the probes had hybridization specificity. In addition, the quality index proportions were 75.65% for BC1, 20.80% for BC2, 2.29% for BC3, 1.20% for BC4, and 0.06% for BC_Poor. The custom DNA microarrays have been deposited in NCBI's Gene Expression Omnibus (Edgar et al., 2002) and are accessible through GEO Platform accession number GPL23315 (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GPL23315).

Table 3.2: Summary of the *A. planci* custom DNA microarray

	Number of probes	Number of gene models
Selected probes	84,922	28,376
Constitutive house-keeping genes	714	238 (duplication with the selected probes)
Total	85,636	28,376

The advantages and disadvantages in the present optimization of DNA microarrays were compared with those in two methods provided by the eArray ("best quality" and "best distribution") (Fig. 3.4). First, in terms of complete coverage of gene models, the present method is slightly superior to those of eArray (Fig. 3.4a). Second, with regards to the BC score proportion, the eArray best quality looks best among the three methods, but there is no significant difference between them (Fig. 3.4b). Third, as to distance between probes, the eArray best distribution method is best (Fig. 3.4c). Although the score of the best quality method is quite low, the score of this study is comparable to that of best distribution method. Fourth, when comparing the hybridization score, a marker of probe specificity, the present method far surpasses the other two (Fig. 3. 4d).

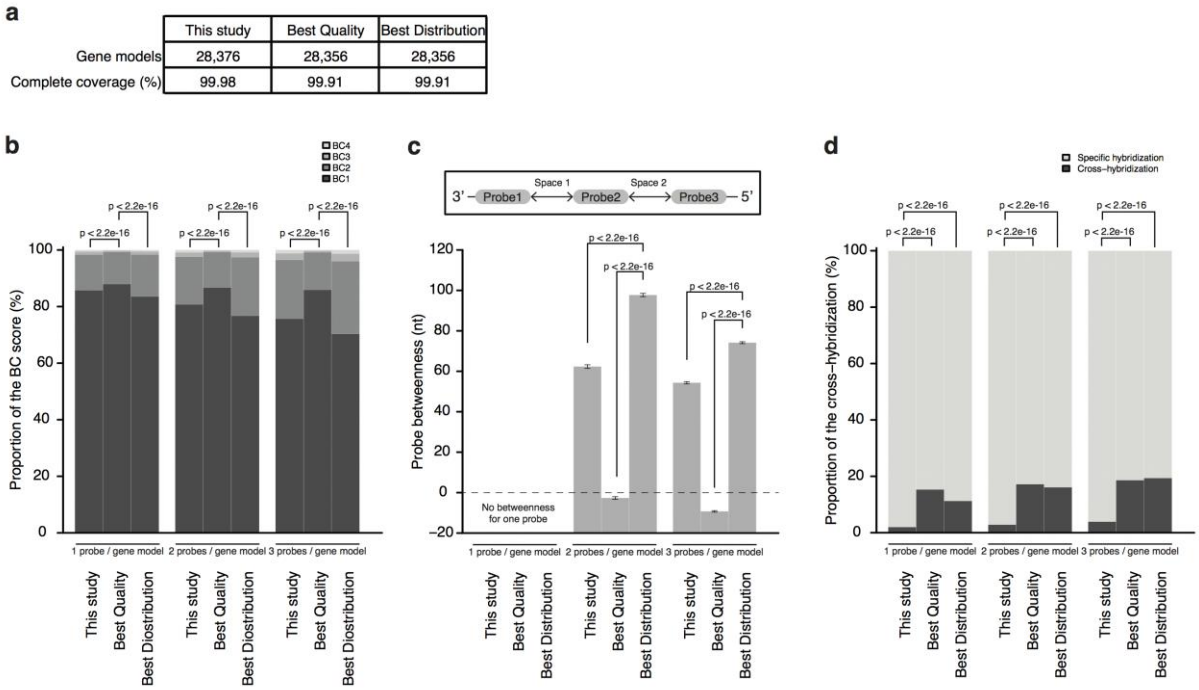


Figure 3.4: Comparison of the *Acanthaster planci* custom DNA microarrays

Custom DNA microarrays designed by three different methodologies: the method developed in this study and two default methods from the Agilent eArray: the best-quality method that prioritizes probe quality, and the best-distribution method that prioritizes distribution of probes. Using each of the three methods, sets of probes designated one, two, and three, respectively, were made for each gene model. (a) With respect to complete coverage of gene models, the developed method was slightly better than those of eArray. (b) Chi-square tests were carried out to evaluate whether the proportion of BC scores by the best-quality methodology was significantly better than the other methods. (c) Distance between two adjacent probes. No distance when single probe was defined from a gene model. *t*-tests were performed to evaluate whether the probe distribution by best-distribution method is significantly better. (d) Proportions of hybridization represented at cross-hybridized and specifically hybridized probes. Chi-square tests were carried out to evaluate whether the specific hybridization by method developed in this study was significantly better than previous studies.

Overall comparison of the three methods indicates that the method of this study is advantageous when probes are designed to suppress cross-hybridization. In addition, when the *A. planci* reference gene model is used, the method of this study covers more gene model. The eArray best quality method indeed produces a set of high quality probes. However, this is at the cost of many probes that overlap one another narrowly within a given gene model. On the other hand, the best distribution is advantageous for producing multiple probes that are well distributed along a gene model. The method developed here yielded the best results when probes were designed for high specificity as well as complete coverage of gene models. Namely, unlike eArray, which is a Web portal supplied by Agilent to assist researchers by offering a simple means to create new microarray designs (SureSelect enrichment libraries, https://earray.chem.agilent.com/earray/helppages/index.htm), the present method is superior because probes are designed to suppress cross-hybridization. By this method, the coverage rate of genes that were not cross-hybridized was much higher than with the other two methods.

The method developed in the present study provides insight into the design of microarrays, and is applicable to most non-model metazoans. The rapidly increasing availability of sequence data for non-model organisms has expanded the options for experimental design of gene expression analysis, including the choice of quantitative RT-PCR, microarray, or high-throughput sequencing, depending on the number of samples and genes. In addition, this method is not limited only to examination of gene expression in one organism, but can also be used for gene identification and microbe detection.

3.2. Comprehensive analyses of invertebrate deuterostome transcriptomes during embryogenesis

3.2.1. Echinoderms

The sea urchin *Strongylocentrotus purpuratus* has been extensively used for quantitative embryological research (Davidson, 2001; Davidson, 2006; McClay, 2011). However, the mode of sea urchin embryogenesis may not always be representative of echinoderms. The sea urchin embryo forms micromeres at the 16-cell stage, and this pattern of cleavage is not seen in the other echinodermata. To overcome this difficulty, the presence of the phylotypic stage in echinodermswas examined instead using the crown-of-thorns starfish (COTS), *Acanthaster planci*. Sampling of *A. planci* adults is rather easy in Okinawa. In addition, the COTS genome has been sequenced with high quality of assembly (Baughman et al., 2014; Hall et al., 2017).

Although the embryogenesis of sea urchins has been most extensively studied, urchin embryos are unique in forming 16-cell micromeres, which play roles in body axis establishment and larval skeleton (reviewed by Davidson, 2006). The starfish embryo is a superior choice for studies of embryogenesis as it shows more of the common features of deuterostome embryogenesis (Henderson, 1969; Hayashi et al., 1973; Lucas, 1982; Chia et al., 1993; Baughman et al., 2014). The cleavage is radially symmetrical, and the embryo develops into a hollow blastula (Fig. 3.5b-d). In the case of *A. planci,* wrinkled blastulae are frequently observed (Fig. 3.5e). The late blastula starts to rotate within the fertilization membrane and hatches out from it. Gastrulation begins by the invagination of archenteron from the vegetal pole side, which ingresses inside the blastocoel (Fig. 3.5f, g). The archenteron ingresses further and bends to make contact with the future stomodeum (Fig. 3.5h, h'). During the mouth opening, a paired coelomic pouch is formed (Fig. 3.5i, j). The embryo is now developed into a bipinnaria, which swims with ciliary bands that are formed at the larval surface (Fig. 3.5j, k). After further development, the larva metamorphoses to a young juvenile.

Developmental temperature was fixed at 28˚C based on a previous study (Keesing et al., 1997). Under the culture conditions used here, soon after observation of *A. planci* embryogenesis, the overall development of *A. planci* was less synchronized than reported in previous studies (Lamare et al., 2014), in particular, around the wrinkled blastula stage

(Hayashi et al., 1973). Due to this, morphological characters based on descriptions from previous studies (Henderson, 1969; Hayashi et al., 1973; Lucas, 1982; Chia et al., 1993) were used instead of sample timing. Timings for embryonic stages before larval formation were recorded when approximately one half of normal embryos showed stage-specific characteristics (Table 3.3). On the other hand, stages of larvae were recorded when the first few individuals showed stage-specific characteristics, since progression from gastrula to larva had individual variances and almost 1/3 ceased development after gastrulation.

As shown in Table 3.3, a total of 18 developmental stages of *A. planci* embryos and larvae were collected in this study. These were unfertilized egg (matured oocyte), early cleavage (2/4-cell stage), mid cleavage (32-cell stage), late cleavage (~100-cell stage), early blastula, mid blastula, late blastula, early gastrula I, early gastrula II, mid gastrula, late gastrula I, late gastrula II, early bipinnaria larva, mid bipinnaria larva, and late bipinnaria larva. On the advice of several starfish researchers, attempts were made to rear bipinnaria larvae to the brachiolaria larvae stage by feeding them algae with a mixture of antibiotics. However, brachiolaria larvae and newly metamorphosed juveniles were not able to be raised.

Total RNAs were extracted from whole embryos and larvae, and the quality of total RNA was examined. To validate the cy3-cRNA concentration for hybridization, positive control was performed by *B. floridae* microarray using 1650 ng of cy3-cRNA sample from adult females. Correlation of technical replicate of two identical samples were compared, and they showed Pearson's correlation coefficient $cor = 0.84$ ($P < 0.001$) and Spearman's rank correlation coefficient $rho = 0.93$ ($P < 0.001$). In *A. planci* microarray, two samples were compared. In one sample, correlation between 1650 ng and 1100 ng showed $cor = 0.99$ ($P < 0.001$) and $rho = 0.97$ ($P < 0.001$). The other sample showed $cor = 0.81$ ($P < 0.001$) and $rho = 0.93$ ($P < 0.001$). Due to limited number of microarrays and samples of all stages, positive controls were not able to performed for *A. planci* microarray and samples. Instead, positive controls were performed using *B. floridae* microarray. Using the positive control as a criterion for validity to reduce the start concentration of cy3-cRNA, 1100 ng in *A. planci* seemed to be compatible to 1650 ng in *B. floridae* microarray. Thus, 1100 ng of cy3-cRNA was used in all samples of *A. planci* for hybridization.

Raw signal intensity from the Agilent Feature Extraction was normalized and scaled to log2. After averaging stage replicates, no missing data was found throughout all stages (Fig. 3.6).

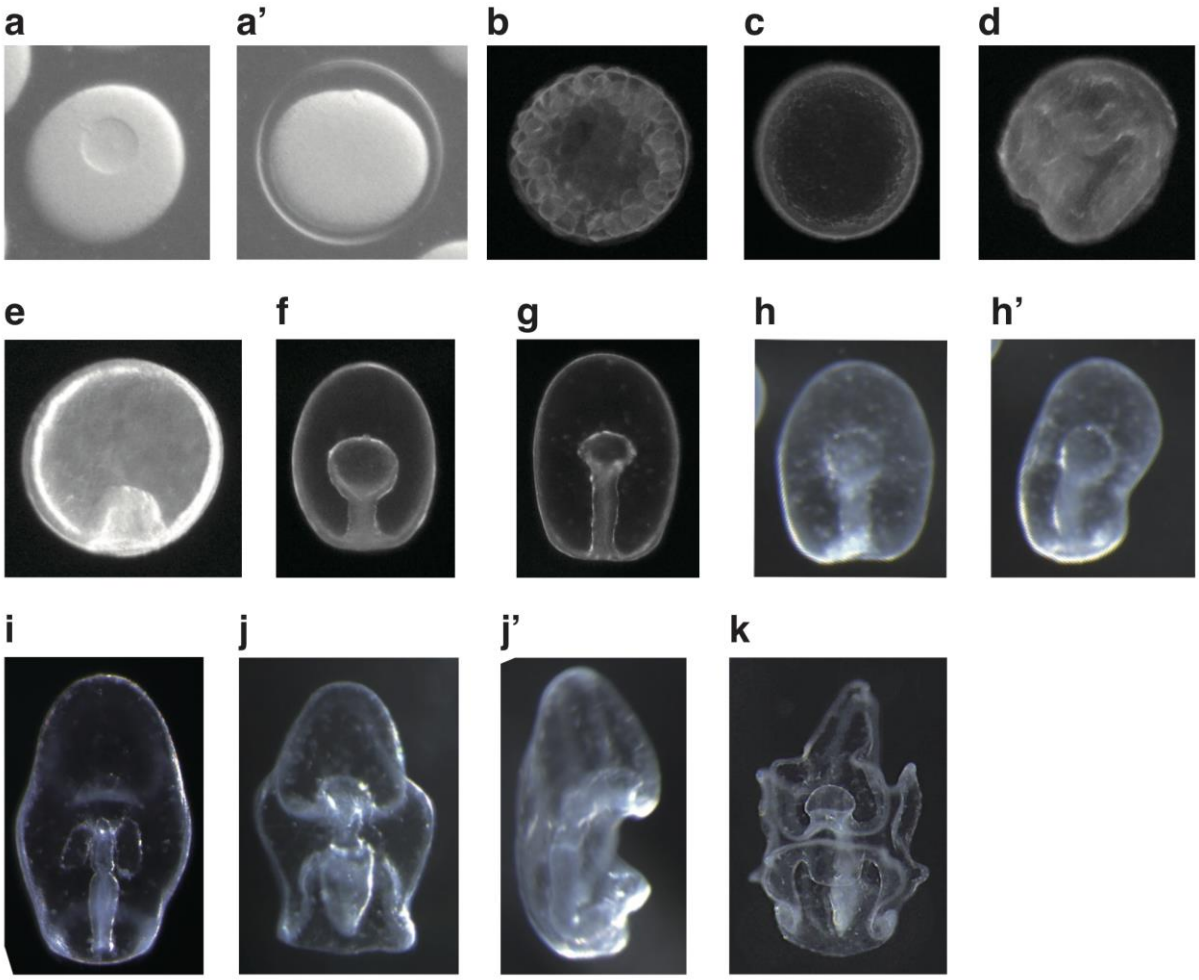


Figure 3.5: Embryos and larvae of *Acanthaster planci* developed at 28˚C

(a-k) Embryos and larvae at different stages. Hatching occurs between mid blastula and late blastula stages.

Table 3.3: Stages of sampling of *A. planci* embryos and larvae

Stage	hpf*	Characteristic features	#R**	References
Unfertilized egg	0	Clear germinal vesicle is present before inducing artificial maturation.	3	Chia et al., 1993
2-cell	2		4	Chia et al., 1993
32-cell	3		4	Chia et al., 1993
~100-cell	4	Surface of the embryo is rough.	3	Chia et al., 1993
Early blastula	8	Surface of the embryo is smooth. Thick blastula wall.	3	Chia et al., 1993
Mid blastula	10	Wrinkled blastula.	2	Chia et al., 1993; Hayashi et al. 1973
Late blastula	12	After hatching. Thin blastula wall. Appearance of the vegetal plate.	2	Chia et al., 1993
Early gastrula I	13.5	Initiation of the invagination.	2	Chia et al., 1993
Early gastrula II	15 - 17	Elongation of the archenteron reaches 1/4 to 1/3 the size of the embryo.	3	Chia et al., 1993
Mid gastrula	19 - 23	Expansion of the tip of archenteron. No mesenchymal cell.	3	Chia et al., 1993
Late gastrula I	26.5 - 35	Developed tip of the archenteron. Existence of the mesenchymal cell.	3	Chia et al., 1993
Late gastrula II	42 - 48	Archenteron bends toward future stomodeum.	2	Chia et al., 1993
Early bipinnaria	51 - 63	After mouth opening and paired coelomic pouch.	3	Henderson, 1969; Lucas, 1982
Mid bipinnaria	74.5 - 91	Upward alimentary canal and anus.	1	Henderson, 1969
Late bipinnaria	141 - 282.5	Advanced bipinnaria arm.	1	Henderson, 1969

* hpf, hours after fertilization

** #R, number of replicates

(a) Developmental path

Embryogenesis of *A. planci* was visualized on a 2D plane by principal component analysis (PCA) through developmental stages for which transcriptomes were available (Fig. 3.6). Plots by principal component-1 (PC-1, 38%) and PC-2 (20%) obtained from PCA results were connected according to developmental time course to form a developmental path. In most cases, the closest stages in both developmental progress and PCA were connected together. However, some stages showed less consistency between developmental progress and distance of plots such as among unfertilized stages and 32-cell stage, and between early gastrula II stage and late gastrula I stage. Morphological differencse in late gastrula II stage (Fig. 3.5h-h') and early bipinnaria larva stage (Fig. 3.5i) were not so distinct and they were highly similar in gene expression. In the same way, mid bipinnaria larva stage and late bipinnaria larva stage were also highly similar in gene expression, although they were different in larval body size and shape.

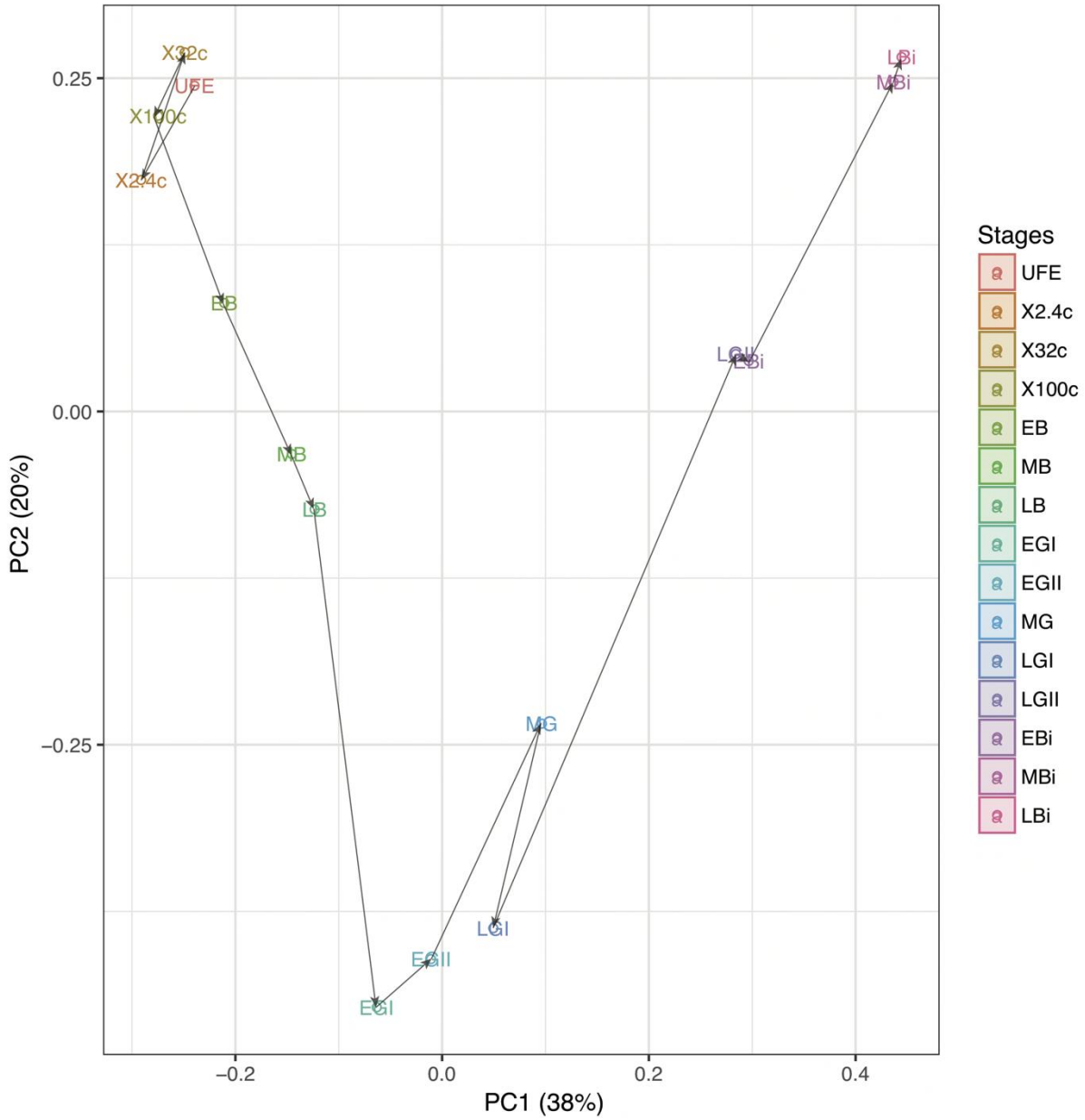


Figure 3.6: Developmental trajectory of *Acanthaster planci* embryogenesis

Labels represent UFE, unfertilized egg; X2.4c, 2/4-cell; X32c, 32-cell; X100c, ~100-cell; EB, early blastula; MB, mid blastula; LB, late blastula; EGI, early gastrula I; EGII, early gastrula II; MG, mid gastrula; LGI, late gastrula I; LGII, late gastrula II; EBi, early bipinnaria; MBi, mid bipinnaria; LBi, late bipinnaria; respectively. Developmental trajectory was analyzed using principal component analysis.

(b) Dendrogram

Dendrogram gave additional indications for developmental path (Fig. 3.7). The result showed that a major distinction is present between periods before and after gastrulation. Clusters of stages before gastrulation period were subclustered into cleavage period (group 1) and blastula period (group 2). The formation of two subclusters suggests the occurrence of maternal-to-zygotic transition (MZT) between ~100-cell stage and early blastula stage. The expression of transcripts stored in mature oocyte was presented in group 1, while the expression of transcript from zygotic genome was in group 2. The cluster of stages after gastrulation period was subclustered into gastrulation period (group 3) and bipinnaria larval period (group 4). The developmental path shown in Fig. 3.6 suggested a similarity between late gastrula II stage and early bipinnaria larva stage. Presumably reflecting this feature, the dendrogram also showed closely related late gastrula II stage and early bipinnaria larva stage.

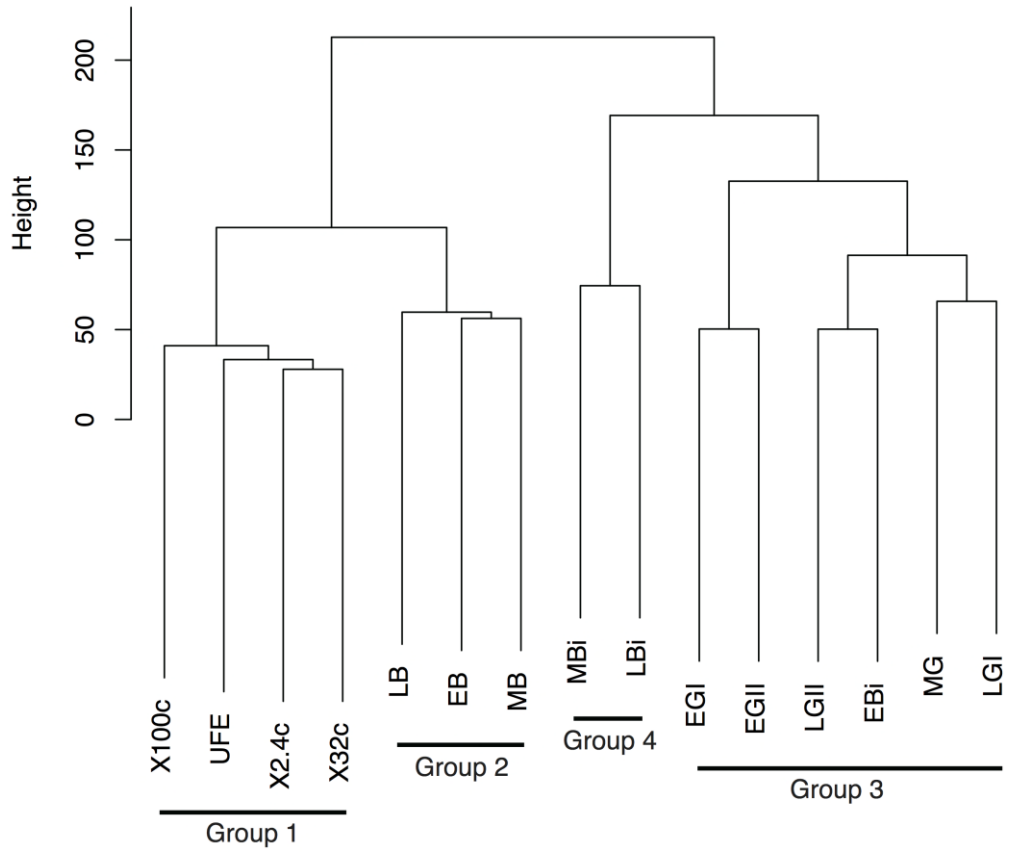


Figure 3.7: Dendrogram of sampled stages in *Acanthaster planci*

Labels represent UFE, unfertilized egg; X2.4c, 2/4-cell; X32c, 32-cell; X100c, ~100-cell; EB, early blastula; MB, mid blastula; LB, late blastula; EGI, early gastrula I; EGII, early gastrula II; MG, mid gastrula; LGI, late gastrula I; LGII, late gastrula II; EBi, early bipinnaria; MBi, mid bipinnaria; LBi, late bipinnaria; respectively.

(c) Gene expression profiles

Of 24,321 genes examined, 12,482 genes showed >= 2-fold change and 11,839 genes showed < 2-fold change in expression level, respectively. Constitutively expressed genes were classified into three groups supported by levels of normalized signal intensity (Supplementary Table 1). In short, genes were annotated to binding protein, cell adhesion, locomotory behavior, transporter activity, catalytic activity, and others. Patterns and/or profiles of dynamic gene expression changes >= 2-fold were shown in Fig. 3.8 and Fig. 3.9. The results indicate the existence of diverse expression patterns, and twelve clusters were extracted to represent gene expression patterns (Figs. 3.8 and 3.9). Group of clustered gene expression were classified roughly into up-regulated clusters, down-regulated clusters, and stable clusters. From the trend of mean and median gene expression at each group, up-regulation was demonstrated in groups 2 (672 genes), 4 (639 genes), 7 (252 genes), 8 (81 genes), 9 (5 genes), 11 (17 genes), and 12 (4 genes). Down-regulation was seen in groups 3 (2,220 genes), 5 (567 genes), 6 (169 genes) and 10 (4 genes). Stable gene expression was observed in group 1, which includes 7,852 genes, the group with the largest number of genes (Figs. 3.8 and 3.9). Most showed different degrees of gradual change from the start stage (unfertilized egg stage) to the final stage (late bipinnaria stage). However, groups 9 and 11 showed a sudden high gene expression after a certain silent expression period (Fig. 3.9), while group 10 showed a sudden decrease after gastrulation, and group 12 showed constant high expression. Since the number of genes in groups 9, 10, 11, and 12 was not that many, future studies should characterize these genes.

Functional associations of various patterns in gene expression were examined by GO enrichment analysis. In this study, special attention was paid on following eight developmentally-related terms in GO slim: nucleic acid binding transcription factor activity (GO:0001071; n=628), DNA binding (GO:0003677; n=1,851), signal transduction (GO:0007165, n=2,828), cell-cell signaling (GO:0007267; n=414), embryo development (GO:0009790; n=174), developmental maturation (GO:0021700; n=81), cell differentiation (GO:0030154; n=1,153), and anatomical structure development (GO:0048856; n=2,249). Note that “n” indicates number of annotated genes. The results showed that DNA binding was enriched at group 9 (P = 0.002), which was further examined to be histone binding (P = 0.042) in GO term. The other development terms were not associated with any of the other 11 groups. Subsequently, the functional association of those genes filtered out by expression change was investigated. Expression profiles were clustered into 3 groups. Since the relative gene expression change was less than 2-fold throughout the sampled stages, genes were

grouped into constitutively low, medium and high expressed clusters. No significantly enriched developmentally-related GO slim terms were found. This suggests that the developmentally-related GO terms appear steadily regardless of expression change or arbitrarily clustered 12 groups in *A. planci*.

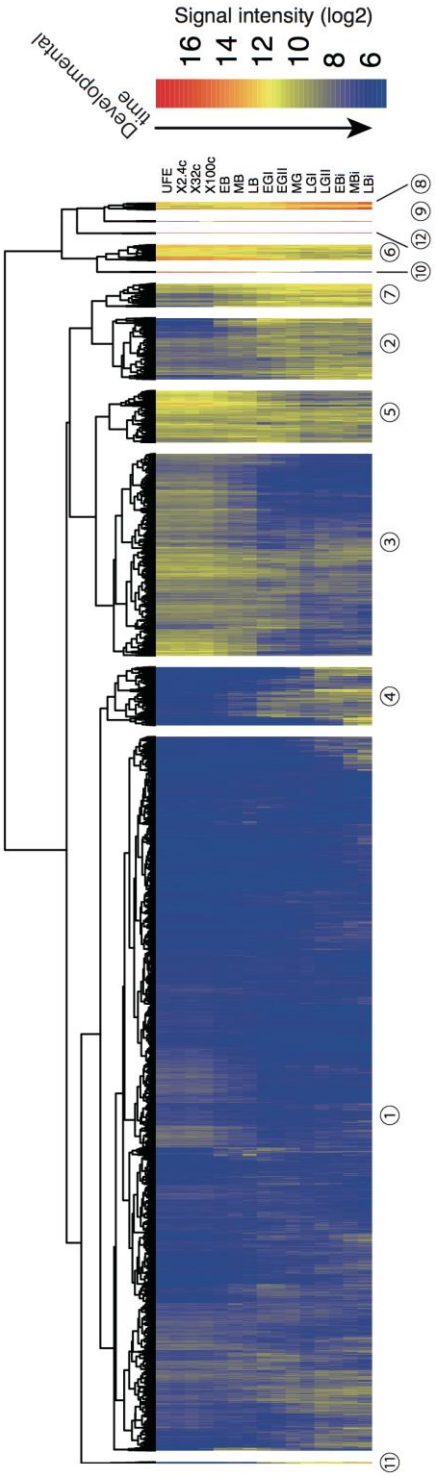


Figure 3.8: Developmental transcriptome in *Acanthaster planci*

Labels represent UFE, unfertilized egg; X2.4c, 2/4-cell; X32c, 32-cell; X100c, ~100-cell; EB, early blastula; MB, mid blastula; LB, late blastula; EGI, early gastrula I; EGII, early gastrula II; MG, mid gastrula; LGI, late gastrula I; LGII, late gastrula II; EBi, early bipinnaria; MBi, mid bipinnaria; LBi, late bipinnaria; respectively. Signal intensities were ordered by hierarchical clustering on the x-axis. Sampled stages were ordered by developmental process on the y-axis. Numbers in the circle below each cluster corresponds to the group name of gene expression profile clusters shown in Fig. 3.7.

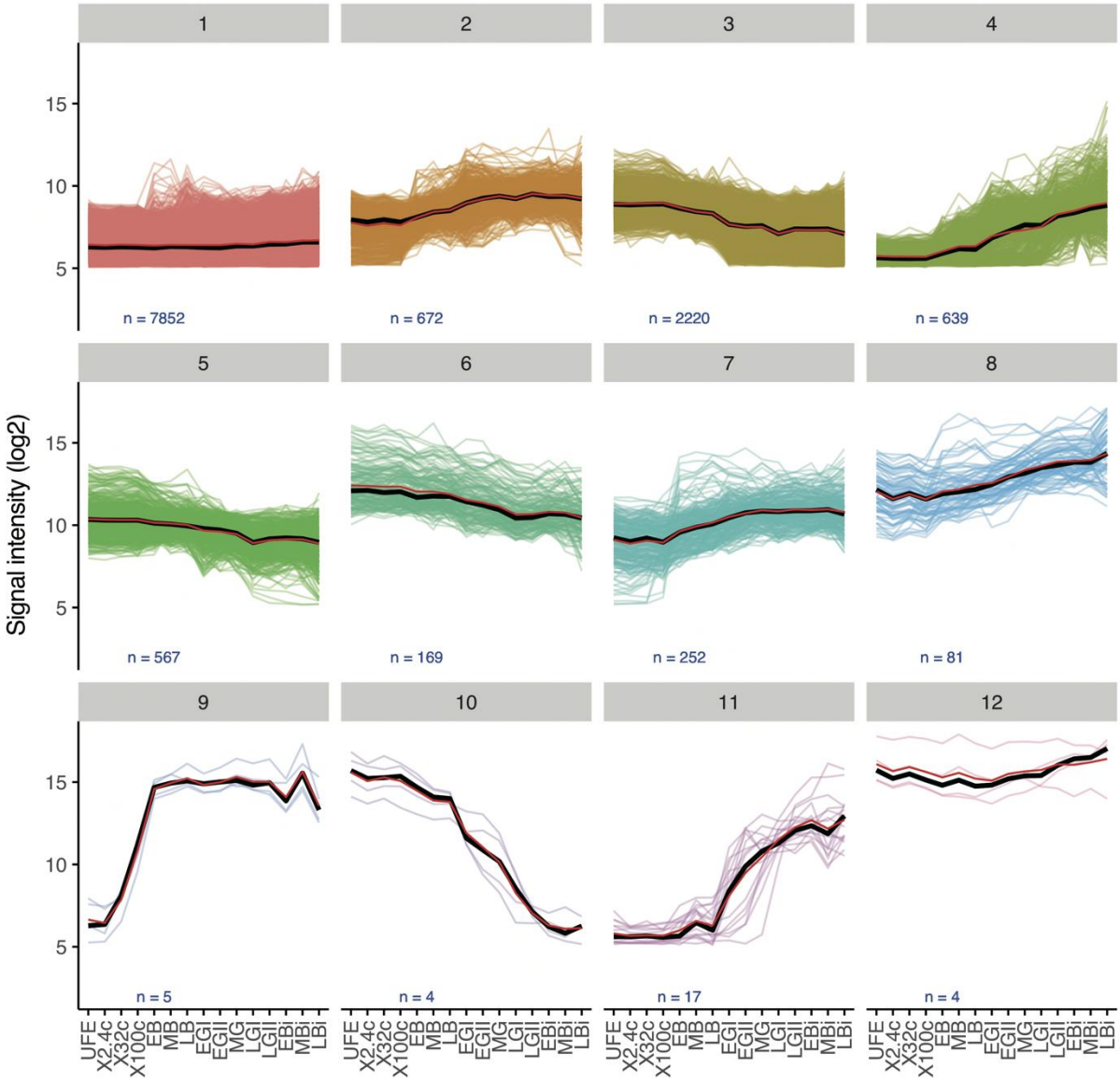


Figure 3.9: Gene expression patterns in developmental transcriptome of *Acanthaster planci*

Labels represent UFE, unfertilized egg; X2.4c, 2/4-cell; X32c, 32-cell; X100c, ~100-cell; EB, early blastula; MB, mid blastula; LB, late blastula; EGI, early gastrula I; EGII, early gastrula II; MG, mid gastrula; LGI, late gastrula I; LGII, late gastrula II; EBi, early bipinnaria; MBi, mid bipinnaria; LBi, late bipinnaria; respectively. The grey title bar in each box representing the name of the group corresponds to the numbers in the circles shown in Fig. 3.8. Developmental time is represented from left to right along the x-axis in each box. Normalized signal intensity is shown along the y-axis of each box. Red line represents mean of gene expression profiles and black line represents median, respectively, in each box.

3.2.2. Hemichordates

Hemichordata consists of two classes, Enteropneusta and Pterobranchia. As sampling of pterobranchs is very difficult and their embryogenesis is not well known (Rottinger and Lowe, 2012), this work focused on enteropneusts (acorn worms) only. Nevertheless, acorn worms still provide an interesting embryological contrast (Rottinger and Lowe, 2012): *Ptychodera flava,* an indirect developer (Tagawa et al., 2001) and *Saccoglossus kowalevskii*, a direct developer (Lowe et al., 2003) exhibit similar form as adults, but with quite different modes of embryogenesis. This study looked only at the indirect developer, *Ptychodera flava.*

The embryogenesis of *Ptychodera flava* has been described in Tagawa et al. (1998) and Lin et al. (2016). Fertilization evokes the fertilization membrane, in which early embryogenesis proceeds. Cleavage is radial and holoblastic, and a blastula with a hollow blastocoel is formed (Fig. 3.10b). Gastrulation occurs as the archenteron invaginates from the vegetal pole towards the stomodeum (Fig. 3.10c). One feature of *Ptychodera* embryogenesis is the formation of a hydrocoel. At the late gastrula stage, the hydrocoel is formed at the aboral side and develops into tornaria larvae with characteristic morphology and ciliary bands (Fig. 3.10d, f, g). The larvae spend several months as plankton before metamorphosis (Fig. 3.10h). Larvae metamorphose into juveniles with a tripartite body plan comprising of anterior proboscis, middle collar region, and posterior trunk (Fig. 3.10i, j).

P. flava adults were collected in Hawaii and Okinawa, and specimens of early embryos were obtained. As shown in Table 3.4, a total of 13 developmental stages of *A. planci* embryos and larvae were examined in this study: fertilized eggs (matured oocyte), cleavage stage embryos (32-cell stage), late blastulae, early gastrulae, mid gastrulae, late gastrulae I, late gastrulae II, and larvae at 2, 3, 4, 5, 6, and 7 days after fertilization (Table 3.4).

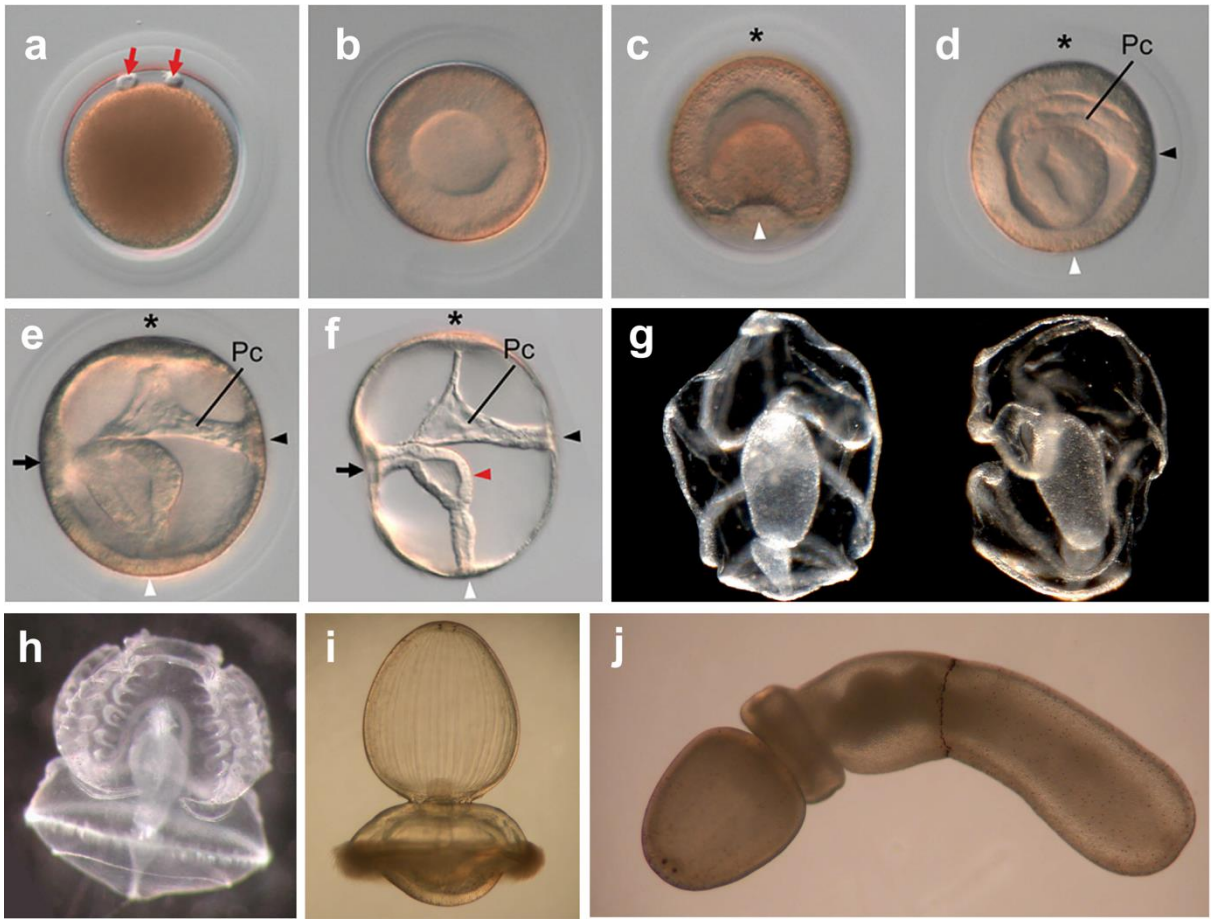


Figure 3.10: Embryogenesis of *Ptychodera flava*.

(a) Fertilized egg, red arrows indicating polar bodies; (b) blastula; (c) early gastrula, white arrow indicating the position of blastopore and asterisk the animal pole; (d) mid gastrula; Pc, hydrocoel, arrowhead indicating the opening of the hydropore; (e) late gastrula, arrow indicating future mouth-opening site; (f) Mullar stage larva, red arrowhead indicating tripartite gut; early tornaria larva, (g) 2-month old larva, (h) 6-month old tornaria larva, (i) larva immediately before metamorphosis, and (j) juvenile 3 days after metamorphosis. (a-f, courtesy of Jr-Kai Yu and g-j, courtesy of Kinifumi Tagawa. From Satoh (2016).)

Table 3.4: *Ptychodera flava*: Developmental stages of embryonic samples and number of replicates

Label	Name	# of replicates
FE	Fertilized egg	7
32c	32-cell	1
LB	Late blastula	1
EG	Early gastrula	1
MG	Mid gastrula	1
LGI	Late gastrula I	2
LGII	Late gastrula II	4
H	2 day (hatching) larva	6
3d	3 day larva	4
4d	4 day larva	4
5d	5 day larva	4
6d	6 day larva	6
7d	7 day larva	4

Validations of start cy3-cRNA concentration for hybridization

The average amount of RNAs obtained from *P. flava* was 825 or 600 ng. Due to limited number of microarrays and samples, positive controls could not be performed on the *P. flava* microarray. Instead, positive control was performed using *B. floridae* microarray and samples using 1650 ng of cy3-cRNA sample from adult females. Comparisons of technical replicates of two identical *B. floridae* samples had Pearson's correlation coefficient *cor* = 0.84 ($P < 0.001$) and Spearman's rank correlation coefficient *rho* = 0.93 ($P < 0.001$). In the *P. flava* microarray, correlation between 1650 ng and 825 ng showed *cor* = 0.97 ($P < 0.001$) and *rho* = 0.90 ($P < 0.001$), and 1650 ng and 600 ng showed *cor* = 0.88 ($P < 0.001$) and *rho* = 0.88 ($P < 0.001$). Using the positive control as a criterion for validity to reduce start concentration of cy3-cRNA, 825 ng in *P. flava* seem to be compatible to 1650 ng using *B. floridae*. Thus, 825 ng of cy3-cRNA was used in all samples of *P. flava* for hybridization. Raw signal intensity from the Agilent Feature Extraction was normalized and scaled to log2. After averaging stage replicates no missing data was found throughout all stages (Fig. 3.6).

(a) Developmental path

Embryogenesis until the 7-day-old larvae of *P. flava* was projected onto the 2D plane by principal component analysis (PCA) through their developmental transcriptome (Fig. 3.11). Plots by principal component-1 (PC-1, 21%) and PC-2 (18%) obtained from PCA results were connected according to developmental time course to form a developmental path. Developmental transcriptomes analyzed by PCA frequently result in bowl-shaped trajectory, which less likely to overlap and/or intersect (Anavy et al., 2014; Owens et al., 2016). The trajectory of *P. flava* showed a rather odd pattern compared to previous studies of various metazoans and in *A. planci* (Fig. 3.6). For example, fertilized eggs, 32-cell embryos, and late blastulae showed a close association at the upper middle region of the plane. Next, early gastrulae were positioned near the lower left corner of the plane. The plots gradually increased and reached a cluster of 3-day-old to 7-day-old larvae (Fig. 3.11). In other words, developmental progress from late blastula to early gastrula stages, and progress from early gastrula to mid gastrula stages clearly overlapped. Similarly, progress from late gastrula I to 7-day larva stages showed some overlap and intersection. In general, the overall trajectory was less likely to follow a bowl-shaped pattern. This was possibly caused by a unique property of the early gastrula, or some artifact of sampling.

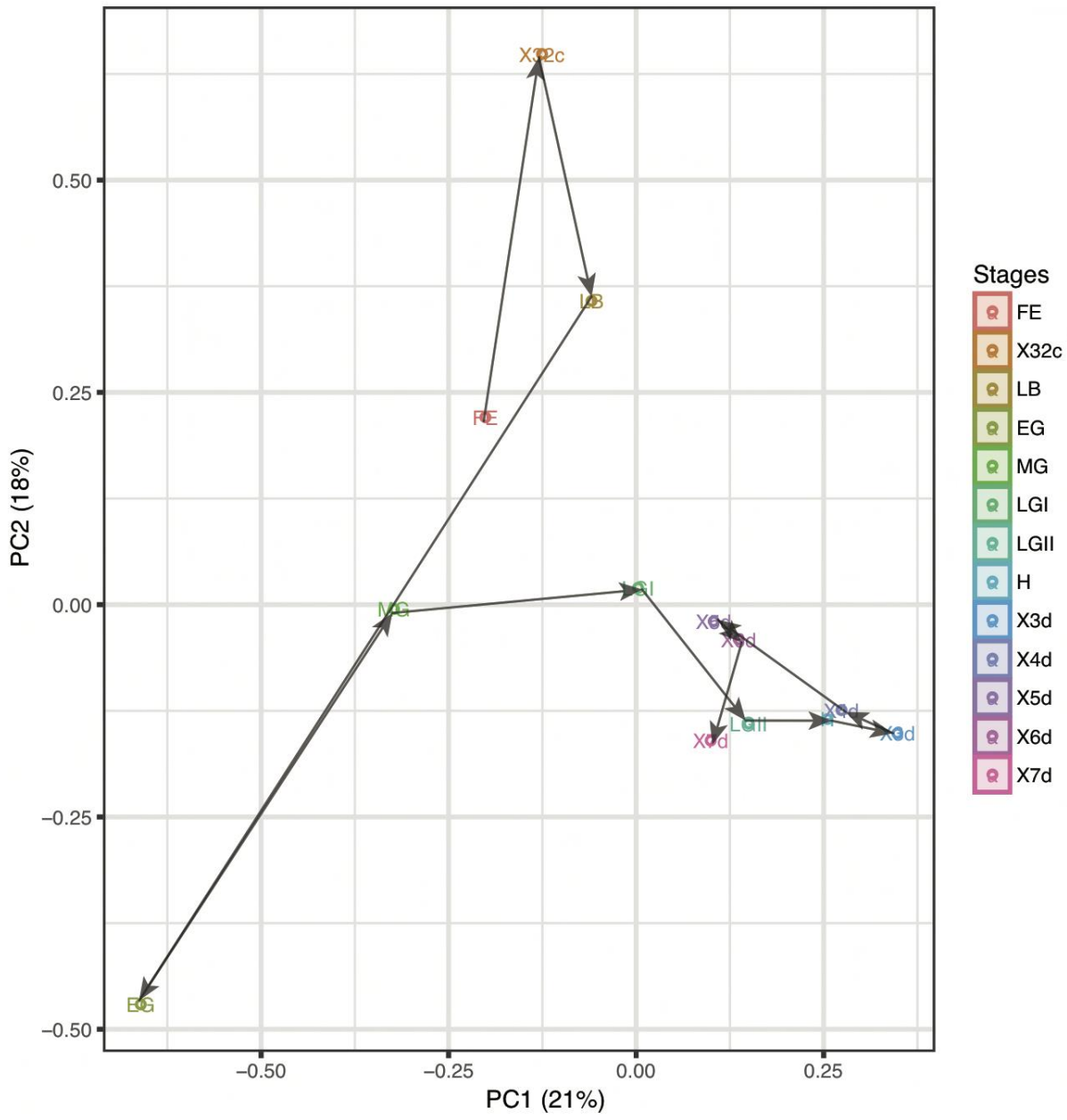


Figure 3.11: Transcriptome of *Ptychodera flava*

Labels represent FE, fertilized egg; X32c, 32-cell; LB, late blastula; EG, early gastrula; MG, mid gastrula; LGI, late gastrula I; LGII, late gastrula II; H, hatching (2 day) larva; X3d, 3 day larva; X4d, 4 day larva; X5d, 5 day larva; X6d, 6 day larva; X7d, 7 day larva; respectively. Developmental trajectory was analyzed by principal component analysis.

(b) Dendrogram

Although the PCA analysis of *P. flava* yielded a complex result, the dendrogram of sampled stages of *P. flava* gave comparatively discrete results (Fig. 3.12). There was a major distinction between the gastrula and later stages. The former cluster consisted of fertilized eggs, blastulae, and early gastrula stages, subclustered into cleavage period (group 1) and blastula/gastrula period (group 2). The late blastula stage was classified together with a series of gastrula stages in the group 2. The latter major cluster consisted of late gastrula II to 7-day larva stages, and could be subclustered to the first half of larval stage (group 3) and the second half of larval stage (group 4). The group 3 may imply a Muller stage in tornaria larva, and similarly the group 4 may imply a Heider stage in tornaria larva.

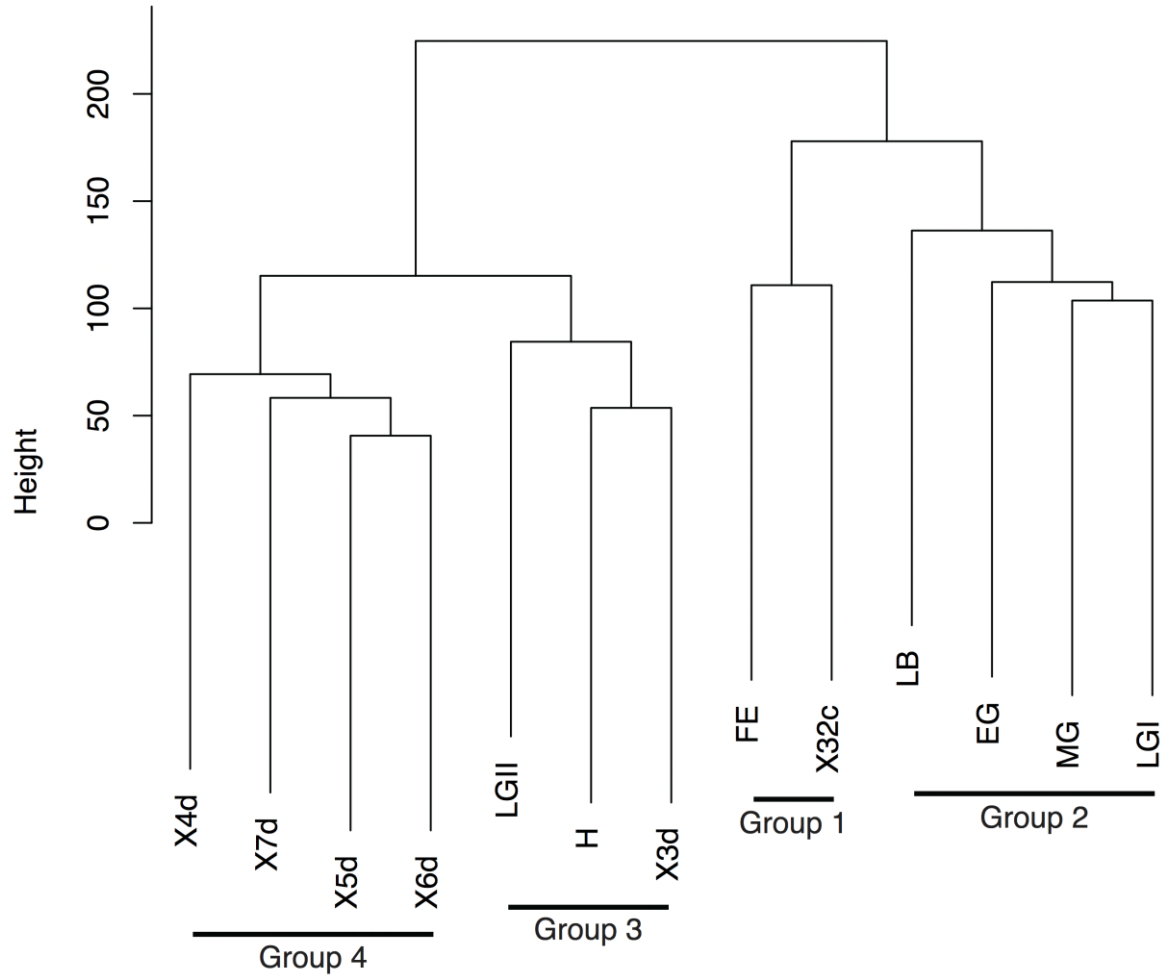


Figure 3.12: Dendrogram of sampled stages in *Ptychodera flava*

Labels represent FE, fertilized egg; X32c, 32-cell; LB, late blastula; EG, early gastrula; MG, mid gastrula; LGI, late gastrula I; LGII, late gastrula II; H, hatching (2 day) larva; X3d, 3 day larva; X4d, 4 day larva; X5d, 5 day larva; X6d, 6 day larva; X7d, 7 day larva; respectively. Dendrogram of developmental stages by hierarchal clustering.

(c) Gene expression profiles

Of 34,601 target genes examined, 11,643 genes showed >= 2-fold change and 22,958 genes showed < 2-fold change in expression, respectively. Constitutively expressed genes were classified into three groups supported by levels of normalized signal intensity (Supplementary Table 2). In short, genes were annotated to ribosomal protein, binding protein, catalytic activity, and other functions. Patterns and/or profiles of dynamic gene expression changes >= 2-fold are shown in Fig. 3.13 and Fig. 3.14. These results indicate the existence of diverse expression patterns, and twelve clusters were extracted to represent gene expression patterns (Figs. 3.13 and 3.14). Group of clustered gene expression were classified roughly into up-regulated clusters, down-regulated clusters, and stable clusters. From the trend of mean and median gene expression at each group, up-regulation was seen in groups 2 (2633 genes), 4 (90 genes), 7 (77 genes), 9 (135 genes), 10 (67 genes), 11 (8 genes), and 12 (29 genes). Down-regulation was observed in groups 3 (555 genes), 5 (265 genes), 6 (145 genes) and 8 (20 genes), while stable gene expression was found in group 1, the largest group, with 7619 genes (Figs. 3.8 and 3.9).

Functional associations of various patterns in gene expression were examined by GO enrichment analysis. Special attention was paid to the following eight developmentally-related terms in GO slim: nucleic acid binding transcription factor activity (GO:0001071; n=683), DNA binding (GO:0003677; n=2,265), signal transduction (GO:0007165, n=4,154), cell-cell signaling (GO:0007267; n=709), embryo development (GO:0009790; n=228), developmental maturation (GO:0021700; n=107), cell differentiation (GO:0030154; n=1,635), and anatomical structure development (GO:0048856; n=3,353). Note that “n” indicates number of annotated genes. These analyses showed that DNA binding was enriched in group 11 (P = 0.038). The other developmentally-related terms were not associated with any of the other 11 groups.

Functional association of those genes filtered out by relative expression change was investigated next. All genes were clustered into 3 groups, which showed almost constant extant expression (relative expression changes less than 2) based on three different intensities, namely, low, medium and high level expression genes. No significantly enriched developmentally-related GO slim term was found. However, in the constitutively highly-expressed gene group (n=4), the following GO terms were significantly enriched: translation (GO:0006412, P=4.19e-5), peptide biosynthetic process (GO:0043043, P=7.50e-5), amide biosynthetic process (GO:0043604, P=5.89e-4), peptide metabolic

process (GO:0006518, *P*=1.02e-3), cellular amide metabolic process (GO:0043603, *P*=7.16e-3), ribosomal subunit (GO:0044391, *P*=9.34e-6), cytosolic part (GO:0044445, *P*=5.06e-4), ribonucleoprotein complex (GO:1990904, *P*=2.03e-2), intracellular ribonucleoprotein complex (GO:0030529, *P*=2.03e-2), structural constituent of ribosome (GO:0003735, *P*=1.33e-5), and structural molecule activity (GO:0005198, *P*=1.24e-2). Developmentally-related GO slim terms were steady regardless of expression change or arbitrarily clustered 12 groups in *P. flava*.

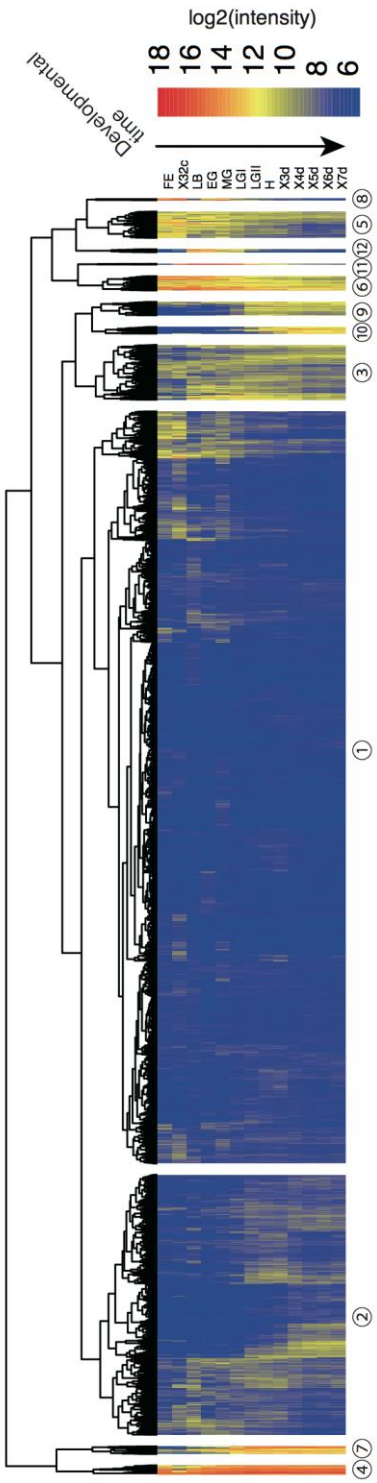


Figure 3.13: Developmental transcriptome of *Ptychodera flava*

Labels represent FE, fertilized egg; X32c, 32-cell; LB, late blastula; EG, early gastrula; MG, mid gastrula; LGI, late gastrula I; LGII, late gastrula I; H, hatching (2 day) larva; X3d, 3 day larva; X4d, 4 day larva; X5d, 5 day larva; X6d, 6 day larva; X7d, 7 day larva; respectively. Genes were ordered by hierarchical clustering on the x-axis. Sampled stages were ordered by developmental process on the y-axis. Numbers in the circle below each cluster correspond to the group name of gene expression profile clusters shown in Fig. 3.12.

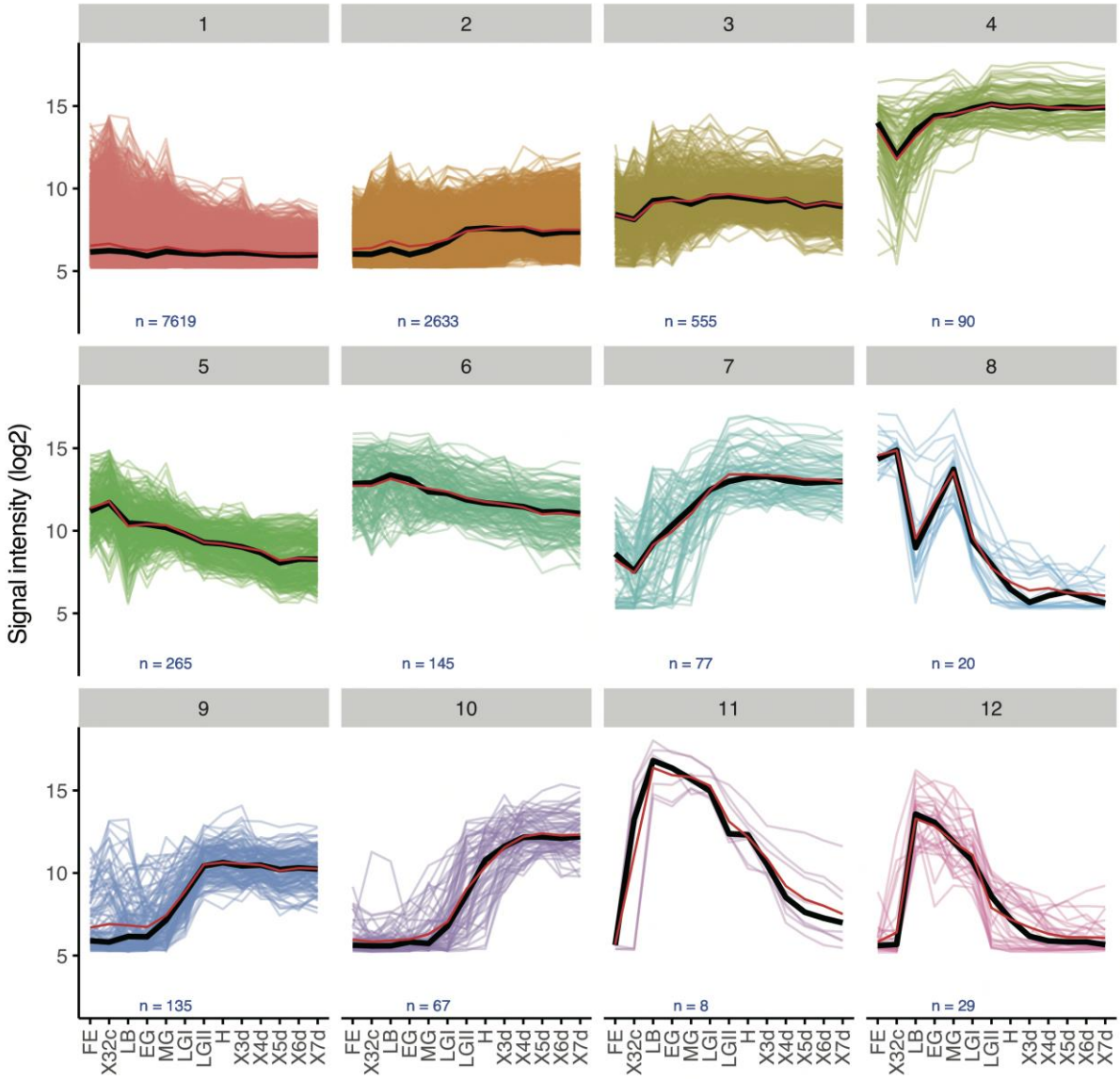


Figure 3.14: Gene expression patterns in developmental transcriptome of *Ptychodera flava*

Labels represent FE, fertilized egg; X32c, 32-cell; LB, late blastula; EG, early gastrula; MG, mid gastrula; LGI, late gastrula I; LGII, late gastrula I; H, hatching (2 day) larva; X3d, 3 day larva; X4d, 4 day larva; X5d, 5 day larva; X6d, 6 day larva; X7d, 7 day larva; respectively. The y-axis represents genes, as clustered into 12 groups, while the x-axis represents sampled developmental stages. The grey title bar in each box representing the name of the group corresponds to the numbers in the circles shown in Fig. 3.13. Developmental time is represented from left to right along the x-axis in each box. Normalized signal intensity is shown along the y-axis of each box. Red line represents mean of gene expression profiles and black line represents median, respectively, in each box.

3.2.3. Cephalochordates

Cephalochordates or lancelets are a group of basal chordates, manifesting an ancestral body plan and embryogenesis (Bertrand and Escriva, 2011; Satoh, 2016). Vertebrates are likely to have evolved from some lancelet-like ancestor by developing the brain, the jaw, and more complex sensory systems such as eyes, noses, and ears (Holland, 2013; Satoh, 2016). The mode of vertebrate embryogenesis is therefore not always identical to that of amphioxus (Hirakow and Kajita, 1990; Bertrand and Escriva, 2011).

Early embryogenesis of the cephalochordate *Branchiostoma* resembles that of ambulacrarians, suggesting a common ancestor (Whittaker, 1997; Bertrand and Escriva, 2011). Fertilization evokes the fertilization membrane, in which cleavage of the embryo proceeds. Cleavage is radial, and the blastula forms with a hollow blastocoel. However, from this point, embryogenesis proceeds differently from ambulacrarians. In contrast to a typical ambulacrarian archenteron invagination from the vegetal plate, gastrulation occurs as the vegetal plate moves towards the animal pole and contacts the ectoderm (Fig. 3.15d-f). The blastocoel disappears in this process. The mode of neurulation is similar to that of urochordates and vertebrates (Fig. 3.15g, h). During neurulation, the body becomes elongated, with gradual appearance of somites along both the left and right sides of the body. Somitogenesis and pharyngogenesis produce a long, thin larva with a body plan like that of the adult (Fig. 3.15i-k). The mouth and first gill slits are located on the left and right sides of the pharynx, respectively. The pelagic larvae resemble juveniles. Adult structures associated with burrowing, including the atrium and oral tentacles are formed during metamorphosis and juvenile development (Whittaker, 1997).

B. floridae specimens were provided by Dr. Jr-Kay Yu at the Institute of Cellular and Organismic Biology, Academia Sinica, Taiwan. As shown in Table 3.5, sampling was carried out on specimens including unfertilized eggs (Fig. 3.15a), 1-4-cell stage embryos, 8-16-cell stage embryos, 32-64-cell stage embryos (Fig. 3.15b), late blastulae (Fig. 3.15c), early gastrulae (Fig. 3.15d), late gastrulae (Fig. 3.15e), early neurulae (Fig. 3.15f), late neurula (Fig. 3.15g), 24-hr larva (knife shape), 48-hr larva (mouth just open), 72-hr larva (two gill slits), 1-cm juveniles, and adult males and females (with mature gonads).

The quality check of total RNA was performed at the laboratory, and followed by cyanine3 labeling and hybridization. Prior to normalizing microarray signal intensities, microarray

target sequences were modified. The *B. floridae* platform system was designed from genome assembly version-1. Target gene models were changed from version-1 to version-2 (Table 3.6). Assembly version-1 was composed of 50,817 gene models, while version-2 was composed of 28,623 gene models (Table 3.6). When comparing the two gene models, 2,312 version-1 gene models were not able to transform into the version-2 models, while all of the remaining 48,505 version-1 gene models were able to transform into the version-2 models. When comparing microarray target sequences, 2,289 target sequences were not able to transform into the version-2 gene models, while the remaining 48,405 target sequences could be transformed. Overall, 99.8% of target gene models were sufficiently changed into version 2 models. In addition, this improved model was used in subsequent analyses such as identifying ortholog gene families and clustering gene expression profiles. After the modification in the target gene models, raw signal intensity from the Agilent Feature Extraction was normalized and scaled to log2. No missing data was found in all stages in subsequent averaging stage replicates (Fig. 3.7).

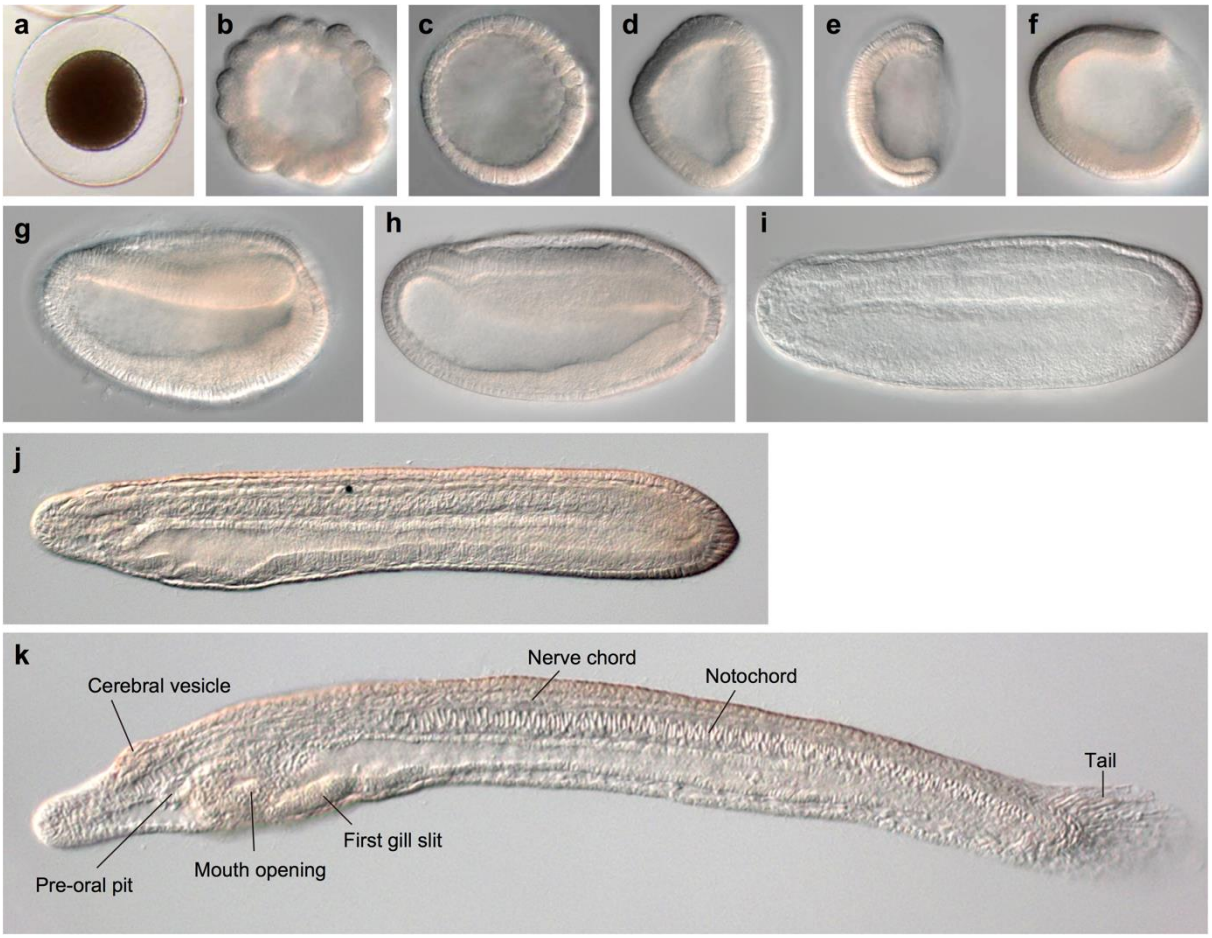


Figure 3.15: Embryogenesis of *Branchiostoma floridae*.

(a) Fertilized egg, (b) 128-cell embryo, (c) late blastula, (d) early gastrula, (e) mid gastrula, (f) late gastrula, (g) early neurula, (h) mid neurula, (i) late neurula, (j) early larva (L1 stage), and (k) 36-hr larva (L2 stage). (Courtesy of Jr-Kai Yu. From Satoh (2016)).

Table 3.5 Developmental stages of *Branchiostoma floridae* for sampling of embryos and adults, and number of replicates

Label	Name	# of replicates
UF	Unfertilized egg	1
1/4c	1- to 4-cell	1
8/16c	8- to 16-cell	1
32/64c	32- to 64-cell	1
LB	~Late blastula	1
G2/3	Early gastrula	1
G5/6	Late gastrula	1
N1	Early neurula	1
N3	Late neurula	1
L1	Early knife-shaped (24 hr) larva	1
L2	Open mouth (48 hr) larva	1
L3	2-gill slit (72 hr) larva	1
Ju	1cm juvenile	1
Af	Adult female with mature gonads	1
Am	Adult male with mature gonads	1

Table 3.6 Target sequences in *Branchiostoma floridae* genome assemblies

	Microarray # target sequence	Gene model mRNA/Protein
Assembly ver. 1	50,694	50,817
Assembly ver. 2	28,560	28,623

(a) Developmental path

Embryonic development of *B. floridae* was projected onto 2D plane by principal component analysis (PCA) through their developmental transcriptomes (Fig. 3.16). Plots by principal component-1 (PC-1, 32%) and PC-2 (17%) obtained from PCA results were connected according to developmental time course to form a developmental path. The overall trajectory clearly described a bowl-shaped pattern (Fig. 3.16), as found in previous studies (Anavy et al., 2014; Owens et al., 2016). Although a notable gene expression change was seen in developmental path between early and late neurula stages, the overall profile was as expected. In order to obtain information of relationship between juveniles and adults, a dotted line was drawn to indicate progress from juvenile stage to stages of adult female and male with mature gonads, respectively. The plot of the adult female stage was closer to the unfertilized egg stage than that of the adult male stage. This is likely caused by the possession of oocytes in the female gonad.

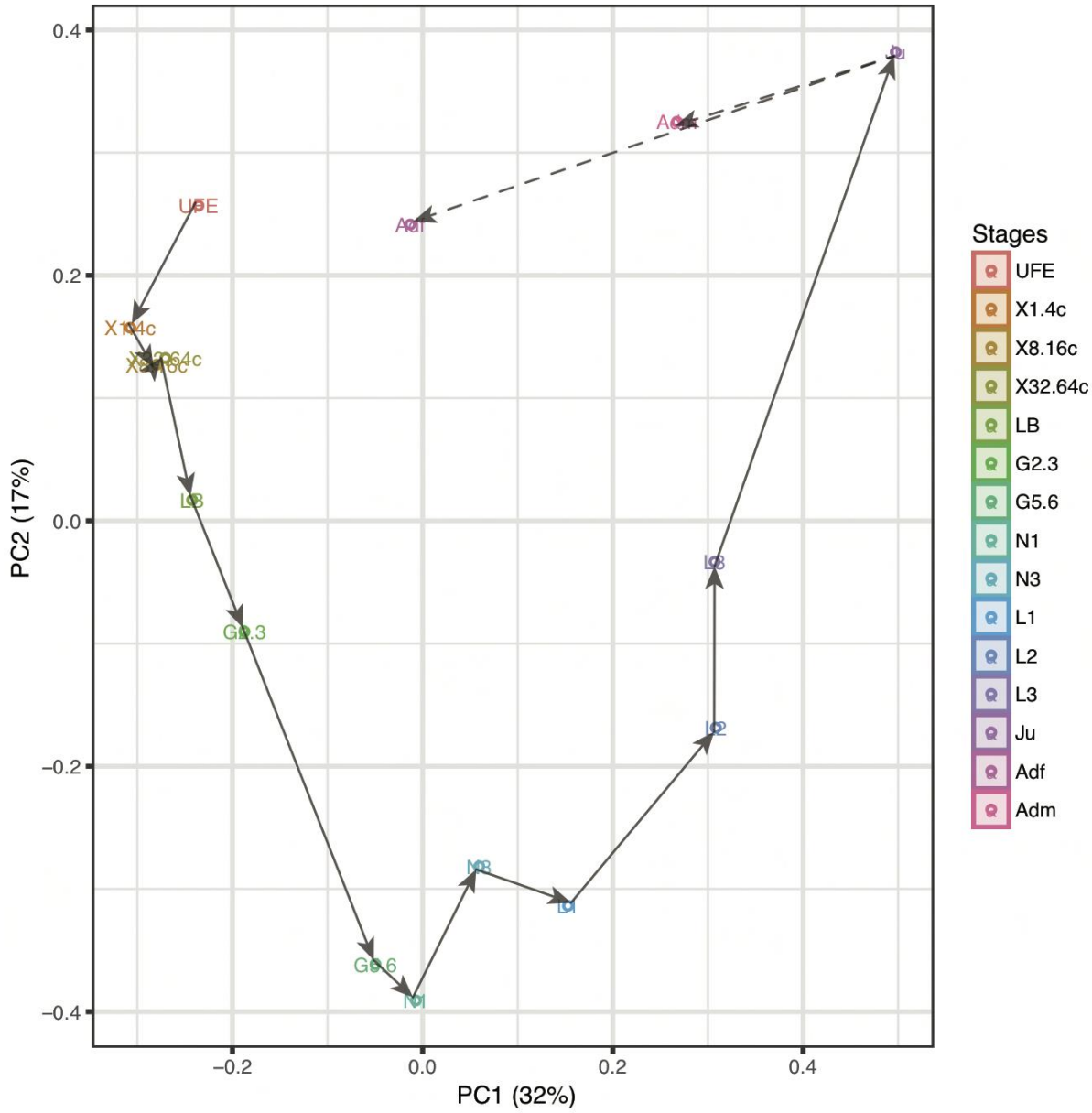


Figure 3.16: Transcriptome of *Branchiostoma floridae*

Labels represent UFE, unfertilized egg; X1.4c, 1/4-cell; X8.16c, 8/16-cell, X32.64c, 32/64-cell; LB, late blastula; G2.3, early gastrula; G5.6, late gastrula; N1, early neurula; N3, late neurula; L1, 24h larva; L2, 48h larva; L3, 72h larva; Ju, 1cm juvenile; Adf, Adult female; Adm, Adult male; respectively. Developmental trajectory was analyzed by principal component analysis. Dotted lines are theoretical independent split from juvenile to adult female and male, respectively.

(b) Dendrogram

As shown in Fig. 3.17, the dendrogram of *B. floridae* sampled stages was comparable to dendrograms of echinoderm (Fig. 3.7) and hemichordate (Fig. 3.12). Stages were clustered into four groups. First cluster consists of juvenile and adult male stages (group 4), which was one of the major cluster. The remaining stages formed three subclusters: the first composed of late neurula and larval stages (group 3), the second consisting of late gastrula and early neurula stages (group 2), and the third subcluster comprising unfertilized eggs, cleavage-stage embryos and early gastrula, and adult female stages (group 1). Transcriptome data of unfertilized egg and adult (female) clustered closely together (see also Fig. 3.16), suggesting a characteristic shaping of developmental transcriptome rather circulating structure through individual life cycle.

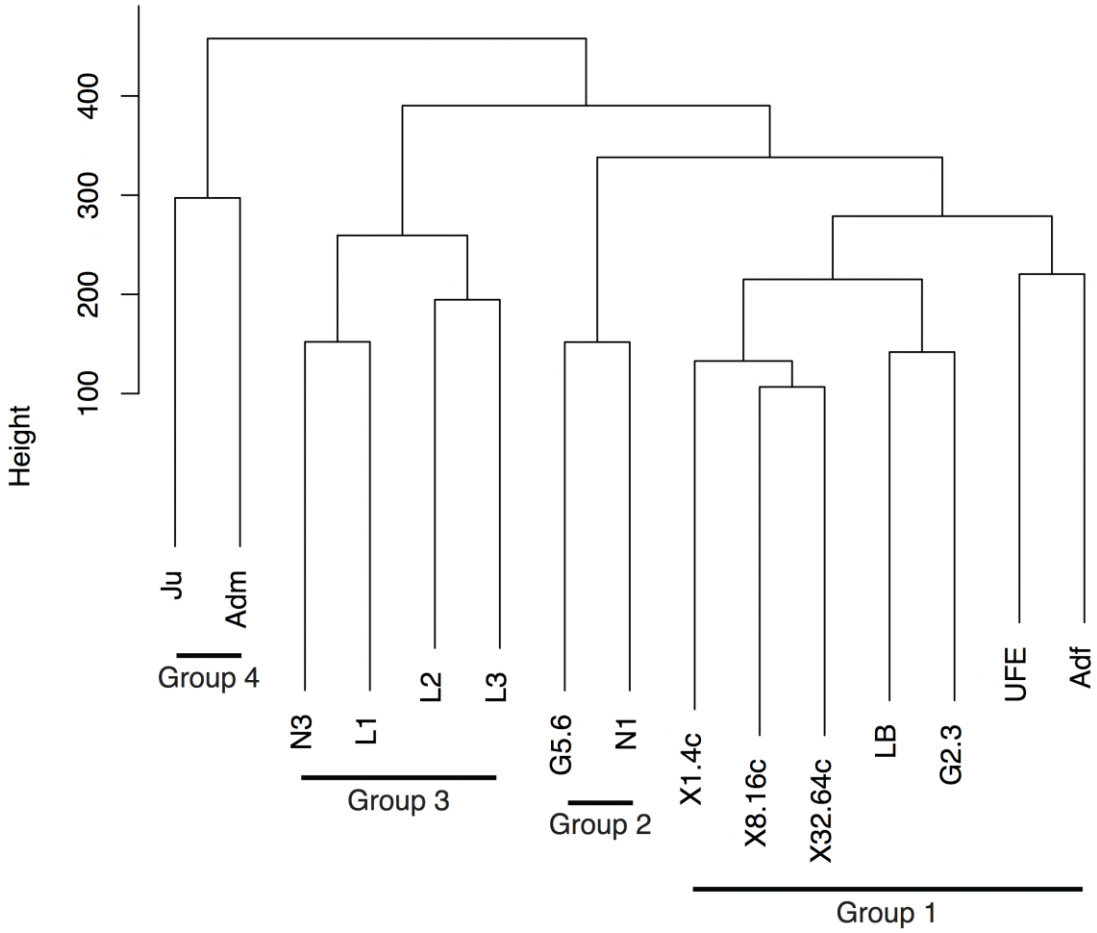


Figure 3.17: Dendrogram of sampled stages in *Branchiostoma floridae*

Labels represent UFE, unfertilized egg; X1.4c, 1/4-cell; X8.16c, 8/16-cell, X32.64c, 32/64-cell; LB, late blastula; G2.3, early gastrula; G5.6, late gastrula; N1, early neurula; N3, late neurula; L1, 24h larva; L2, 48h larva; L3, 72h larva; Ju, 1cm juvenile; Adf, Adult female; Adm, Adult male; respectively. Dendrogram of developmental stages by hierarchal clustering.

(c) Gene expression profiles

Of 28,560 target genes examined, 27,716 genes showed >= 2-fold change in expression and 844 genes showed < 2-fold expression change, respectively. Constitutively expressed genes were classified into three groups supported by levels of normalized signal intensity (Supplementary Table 3). In short, genes were annotated to catalytic activity, transporter activity, binding protein, and others. Patterns and/or profiles where dynamic gene expression more than doubled are shown in Fig. 3.18 and Fig. 3.19. The results indicate the existence of diverse expression patterns, and twelve clusters were extracted to represent gene expression patterns (Figs. 3.18 and 3.19). Group 1 contains 17,720 genes, whose expression was stable for the entire period of sampling, from unfertilized eggs and adult stage. These genes likely have house-keeping functions. On the other hand, the gene expression profiles of groups 3 (1,719 genes) and 7 (237 genes) are also rather constant, although the former showed intermediate level and the latter high level of expression (Figs. 3.18 and 3.19). Up-regulation was found in groups 2 (2,808 genes), 4 (129 genes), 5 (599 genes), 9 (110 genes), 10 (407 genes), and 11 (18 genes). The expression profiles of groups 5, 9, 10, and 11 resembled each other. The level of gene expression increased during gastrula stages and decreased towards later juvenile stages, although they differed from each other in timings of up-regulation and down-regulation. Significant down-regulation was seen in groups 6 (3,128 genes), 8 (829 genes), and 12 (12 genes). A special interest was found in group 4 (129 genes). In this group, maternal gene expression was quickly down-regulated after fertilization, kept silent until the end of neurulation, and then quickly re-activated in the subsequent stages of larvae, juveniles and adults (Figs. 3.18 and 3.19).

Functional associations of various patterns in gene expression were examined by GO enrichment analysis. Special attention was paid to the following 8 development-related terms in GO slim: nucleic acid binding transcription factor activity (GO:0001071; n=860), DNA binding (GO:0003677; n=2,488), signal transduction (GO:0007165, n=3,920), cell-cell signaling (GO:0007267; n=584), embryo development (GO:0009790; n=228), developmental maturation (GO:0021700; n=140), cell differentiation (GO:0030154; n=1,727), and anatomical structure development (GO:0048856; n=3,743). Note that “n” indicates number of annotated genes. In group 5, DNA binding (GO:0003677, 130 genes, *P*=3.35e-8) and nucleic acid binding transcription factor activity (GO:0001071, 46 genes, *P*=9.61e-5) were enriched. In group 6, nucleic acid binding transcription factor activity (GO:0001071, 184 genes, *P*=4.13e-8) and DNA binding (GO:0003677, 594 genes, *P*=6.15e-8) were enriched. In group 7, DNA binding (GO:0003677, 63 genes, *P*=1.78e-8)

was enriched. In group 8, DNA binding (GO:0003677, 165 genes, *P*=3.32e-8) was enriched. In group 10, nucleic acid binding transcription factor activity (GO:0001071, 29 genes, *P*=4.38e-2) was enriched. In group 11, DNA binding (GO:0003677, 8 genes, *P*=2.64e-2) was enriched. Then, functional association of those genes filtered out by level of expression change was investigated. Expression profiles were clustered into 3 groups. Since gene expression changes were less than 2-fold throughout sampled stages, genes were grouped into constitutively low, medium and high expressed clusters. In constitutively low expressed group, cell-cell signaling (GO:0007267, 50 genes, *P*=5.26e-7) was enriched.

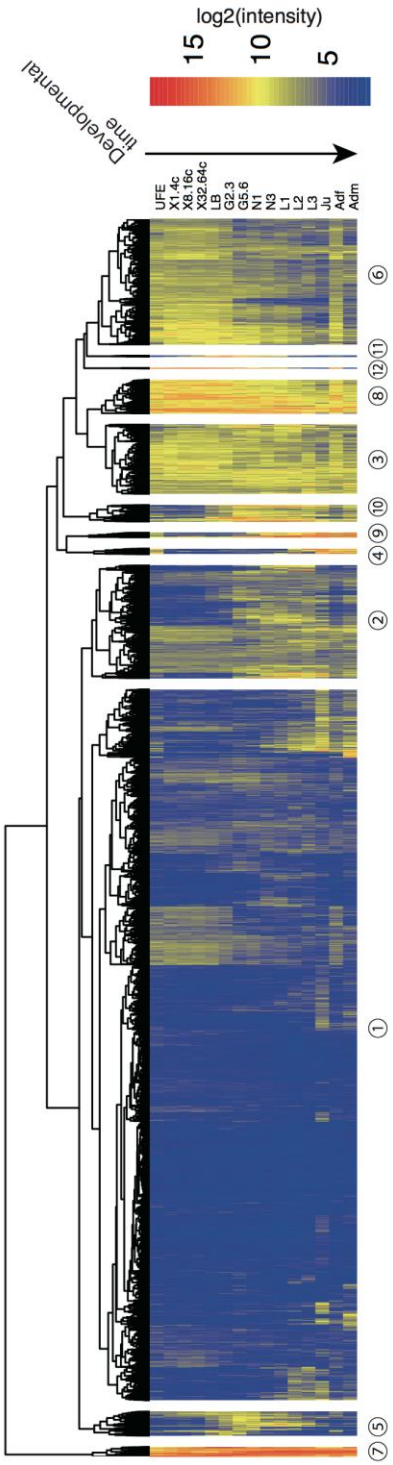


Figure 3.18: Transcriptome of *Branchiostoma floridae*

Labels represent UFE, unfertilized egg; X1.4c, 1/4-cell; X8.16c, 8/16-cell, X32.64c, 32/64-cell; LB, late blastula; G2.3, early gastrula; G5.6, late gastrula; N1, early neurula; N3, late neurula; L1, 24h larva; L2, 48h larva; L3, 72h larva; Ju, 1cm juvenile; Adf, Adult female; Adm, Adult male; respectively. Hierarchal clustering of microarray target gene models. Genes were ordered by hierarchical clustering on the x-axis. Sampled stages were ordered by developmental process on the y-axis. Numbers in the circle below each cluster correspond to the group name of gene expression profile clusters shown in Fig.. 3.19.

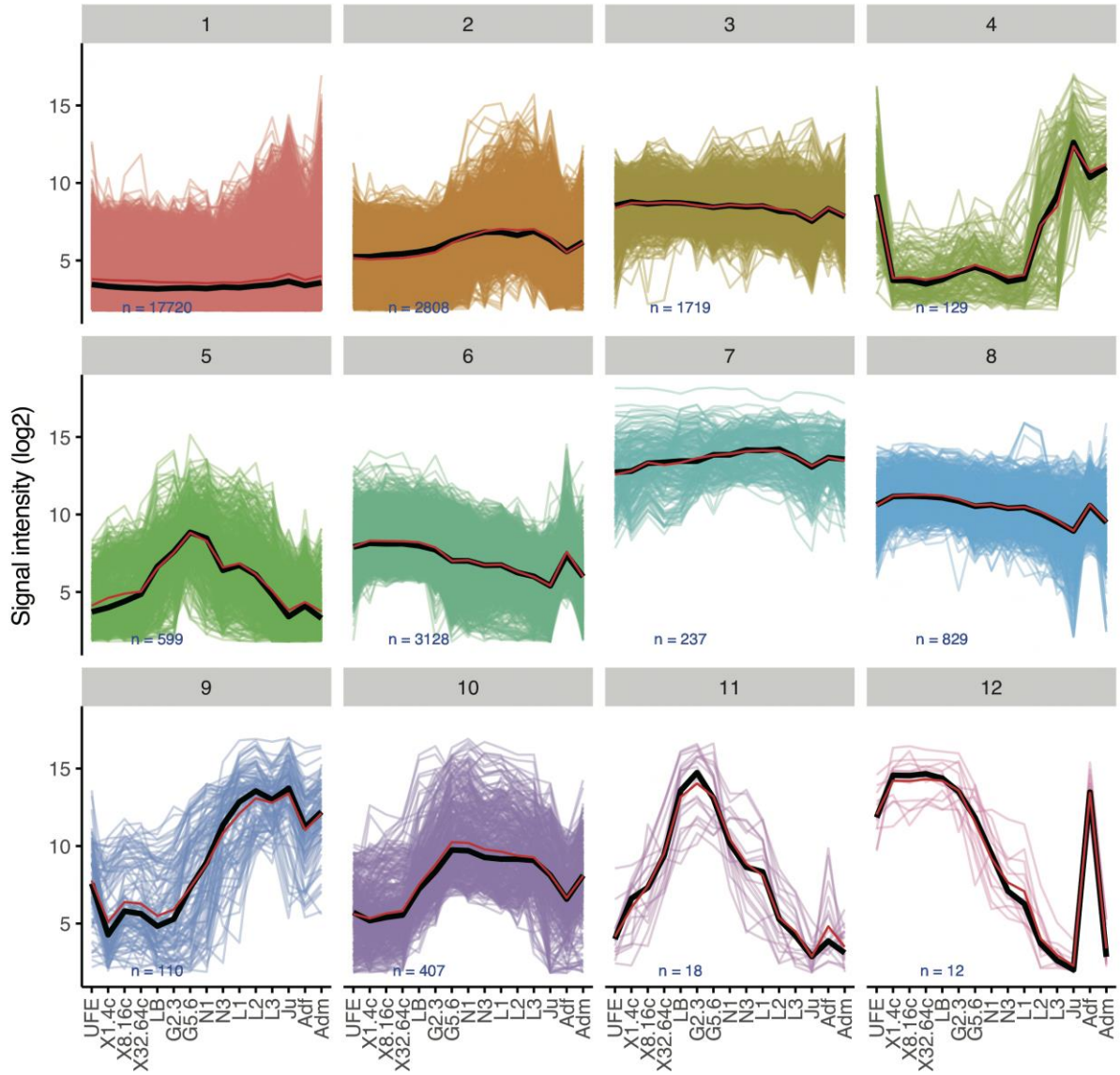


Figure 3.19: Developmental transcriptome of *Branchiostoma floridae*

Labels represent UFE, unfertilized egg; X1.4c, 1/4-cell; X8.16c, 8/16-cell, X32.64c, 32/64-cell; LB, late blastula; G2.3, early gastrula; G5.6, late gastrula; N1, early neurula; N3, late neurula; L1, 24h larva; L2, 48h larva; L3, 72h larva; Ju, 1cm juvenile; Adf, Adult female; Adm, Adult male; respectively. The y-axis represents genes, as clustered into 12 groups, while the x-axis represents sampled developmental stages. The grey title bar in each box representing the name of the group corresponds to the numbers in the circles shown in Fig. 3.18. Developmental time is represented from left to right along the x-axis in each box. Normalized signal intensity is shown along the y-axis of each box. Red line represents mean of gene expression profiles and black line represents median, respectively, in each box.

3.2.4. Urochordates

Among several ascidian species, *Ciona intestinalis* has been used the most frequently for studies of developmental biology (Satoh, 2013). *C. intestinalis* is the seventh animal of which the genome has been decoded (Dehal et al., 2002). Together with the available cDNA resources, *Ciona* has thus become a model organism for studying developmental genomics (Satoh, 2013).

Ascidian embryogenesis is unique as compared to ambulacrarian and cephalochordate development (Fig. 3.20) (Conklin, 1905; Satoh, 2013). Ascidian embryogenesis has been described as a typical "mosaic" type, in which developmental fate of embryonic cells is determined at very early stage of embryogenesis, and if a certain cell is destroyed, its neighboring cells cannot compensate for the cells lost (reviewed in Satoh, 2013). Cleavage is bilaterally symmetrical. A typical blastocoel is not formed. Gastrulation occurs at the 112-cell stage (3.20f), which is followed by the neurulation on the future ventral side (3.20g, h). Neurulation is common to cephalochordates and vertebrates. Soon after neurulation, the trunk region becomes elongated to form a tailbud embryo (3.20i, j). The embryogenesis of *Ciona intestinalis* proceeds approximately 18 hours after fertilization at 18˚C (Satoh, 2014) (Satoh, 2013). A tadpole-type larva hatches out from the chorion (Fig. 3.20k). The larva swims for a couple of days without opening the mouth and then metamorphoses to the juvenile (Fig. 3.20l).

As shown in Table 3.7, a total of 12 developmental stages were sampled, which include fertilized eggs (Fig. 3.20a), 64-cell stage embryos, mid gastrulae (Fig. 3.20f), mid neurulae (Fig. 3.20g), early tailbud embryos (Fig. 3.20i), mid tailbud embryos (Fig. 3.20j), late tailbud embryos, mid swimming larvae (Fig. 3.20k), early body-axis-rotation stage, mid body-axis-rotation stage, and late body-axis-rotation stage.

Total RNA was extracted. The quality of RNA was checked at the laboratory. Raw signal intensity from the Agilent Feature Extraction was normalized and scaled to log2. A comparison of signal intensity and base call showed missing gene expression data in *C. intestinalis* transcriptome in the 47 CIYS models. However, after averaging stage replicates no missing data was found at all target animals through all stages.

In *C. intestinalis* microarray, no target sequences were annotated to public data. Thus, to

define the CIYS model, target sequences were manually linked to public data, and those sequences with no link to public database were manually translated to amino acid sequences. The CIYS model consists of 19,889 mRNAs and 20,189 proteins. After removing redundant sequences, the CIYS model consists of 16,288 mRNAs and 16,633 proteins.

Figure 3.20: Embryogenesis of the ascidian, *Ciona intestinalis*.

Embryos were dechorionated to reveal their outer morphology. (a) Fertilized egg, (b) 2-cell embryo, (c) 4-cell embryo, (d) 16-cell embryo, (e) 32-cell embryo, (f) gastrula (~150 cells), (g, h) neurula, (i, j) tailbud embryos, (k) tadpole larva, and (l) a juvenile a few days after metamorphosis. (From Satoh (2016).)

Table 3.7 *Ciona intestinalis*, Developmental stages of embryonic samples and number of replicates

Label	Name	# of replicates
FE	Fertilized egg	3
64c	64-cell	3
MG	Mid gastrula	3
MN	Mid neurula	3
IT	Initial tailbud	3
ET	Early tailbud	3
MT	Mid tailbud	3
LT	Late tailbud	3
MSL	Mid swimming larva	3
EBAR	Early body-axis-rotation	3
MBAR	Mid body-axis-rotation	3
LBAR	Late body-axis-rotation	3

(a) Developmental path

Embryonic development from unfertilized egg stage to juvenile stage of *C. intestinalis* was projected onto 2D plane by principal component analysis (PCA) through their developmental transcriptomes (Fig. 3.21). Plots by principal component-1 (PC-1, 33%) and PC-2 (28%) obtained from PCA results were connected according to developmental time course to form a developmental path. The trajectory of the developmental path shows the closest stages in both developmental progress and PCA were connected. A clear U-shaped or V-shaped pattern was obtained (Fig. 3.21). At the bottom of the figure lies the mid tailbud stage, while stages of unfertilized egg and juveniles were highest on left and right (Fig. 3.21). This implies a notable gene expression change in the developmental path at the mid tailbud stage.

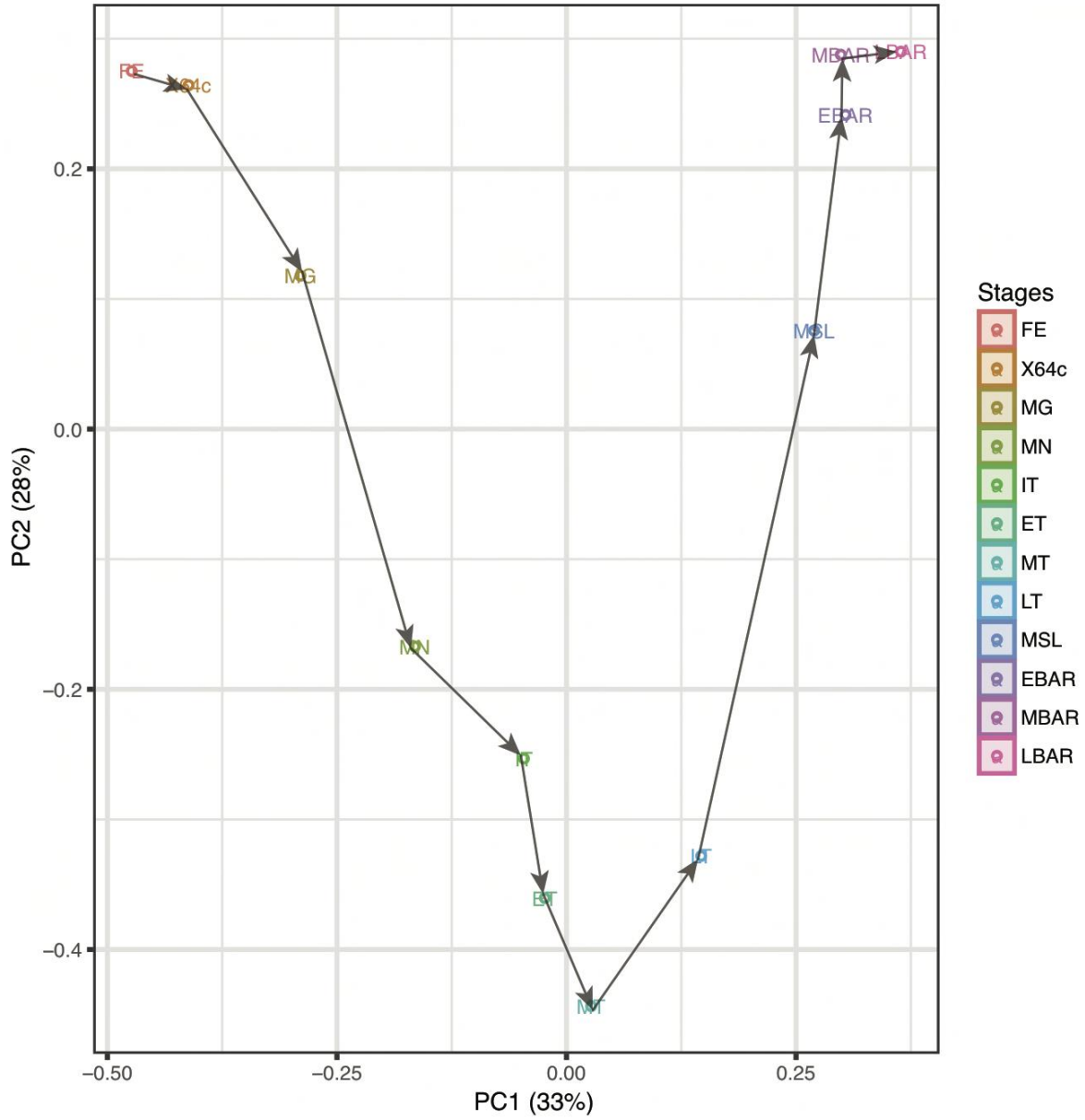


Figure 3.21: Transcriptome of *Ciona intestinalis*

Labels represent FE, fertilized egg; X64c, 64-cell; MG, mid gastrula; MN, mid neurula; IT, initial tailbud; ET, early tailbud; MT, mid tailbud; LT, late tailbud; MSL, mid swimming larva; EBAR, early body-axis-rotation; MBAR, mid body-axis-rotation; LBAR, late body-axis-rotation; respectively. Developmental trajectory was analyzed by principal component analysis.

(b) Dendrogram

The dendrogram gave additional indications for developmental transcriptome (Fig. 3.22), revealing two major clusters. The first included unfertilized eggs, 64-cell embryos, and mid gastrulae, while the second included latter stages of neurulae, tailbud embryos, larvae and several body-axis-rotation stages of juveniles (Fig. 3.22). In more detail, the former was composed of two subgroups: unfertilized eggs (group 1) and 64-cell embryos/mid gastrulae (group 2). The latter was composed of two subgroups: neurulae/tailbud embryos (group 3) and juveniles (postembryonic stages) (group 4). Since the first subgroup was further subdivided into one with earlier stages of tailbud larvae and the other with later stages of tailbud larvae, there might be some changes in developmental transcriptomes here.

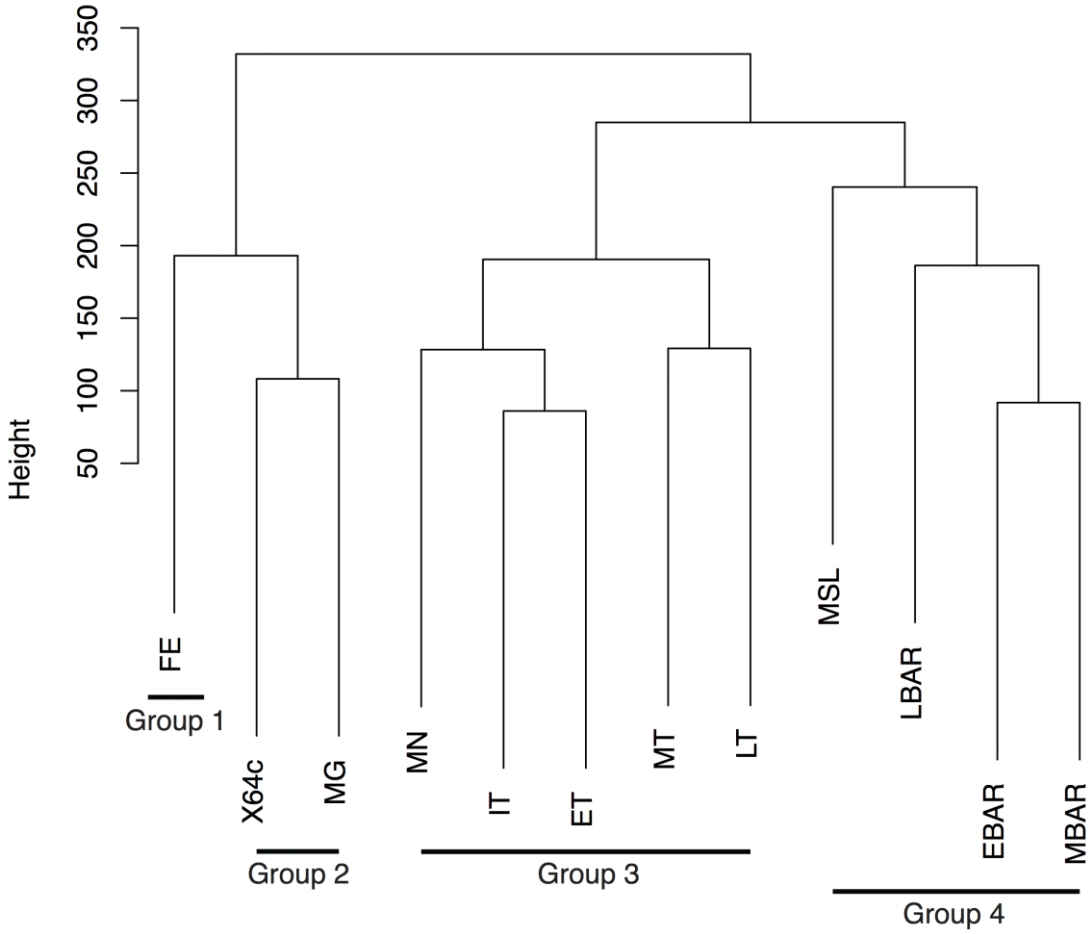


Figure 3.22: Dendrogram of sampled stages in *Ciona intestinalis*

Labels represent FE, fertilized egg; X64c, 64-cell; MG, mid gastrula; MN, mid neurula; IT, initial tailbud; ET, early tailbud; MT, mid tailbud; LT, late tailbud; MSL, mid swimming larva; EBAR, early body-axis-rotation; MBAR, mid body-axis-rotation; LBAR, late body-axis-rotation; respectively. Dendrogram of developmental stages by hierarchal clustering.

(c) Gene expression profiles

Of 19,964 genes examined, 19,406 genes showed >= 2-fold change and 558 genes showed < 2-fold change in expression, respectively. Constitutively expressed genes were classified into three groups, as supported by levels of normalized signal intensity (Supplementary Table 4). In short, genes were annotated to ribosomal protein, catalytic activity, binding protein including transcription factor, transport activity, and others. Patterns and/or profiles in dynamic gene expression changes with >= 2-fold were shown in Fig. 3.23 and Fig. 3.24. The results indicate that *C. intestinalis* also showed the diverse expression patterns, and twelve clusters were extracted to represent gene expression patterns (Figs. 3.23 and 3.24). A similar pattern was noticed in groups 1 (9,310 genes), 2 (3,319 genes), 5 (408 genes), and 6 (1,719 genes), although the signal intensity was lowest in group 1 and highest in group 5 (Figs. 3.23 and 3.24). The up-regulation pattern was found in groups 3 (977 genes), 8 (641 genes), 9 (234 genes) and 12 (111 genes). The down-regulation pattern was seen in group 4 (2,171 genes). The up- and down-regulation pattern was seen in groups 7 (246 genes), 10 (241 genes) and 11 (29 genes). Each of the expression profiles may reflect genes with different functions.

Functional associations of various patterns in gene expression were examined by GO enrichment analysis. Special attention was paid on following eight developmentally-related terms in GO slim: nucleic acid binding transcription factor activity (GO:0001071; n=507), DNA binding (GO:0003677; n=1,262), signal transduction (GO:0007165, n=1,903), cell-cell signaling (GO:0007267; n=169), embryo development (GO:0009790; n=147), developmental maturation (GO:0021700; n=78), cell differentiation (GO:0030154; n=935), and anatomical structure development (GO:0048856; n=1,659). Note that “n” indicates number of annotated genes. There was no developmentally-related GO term enriched found in gene expression profiles of 12 clusters. Then, functional association of those genes that was filtered out by level of expression change was investigated. All genes were clustered into 3 groups. Since the gene expression change was less than 2-fold throughout sampled stages, genes were grouped into constitutively low, medium and high expressed clusters. No significantly enriched developmentally-related GO slim terms were found. This suggests that the developmentally-related GO terms remain constant regardless of expression change or arbitrarily clustered 12 groups in *C. intestinalis*.

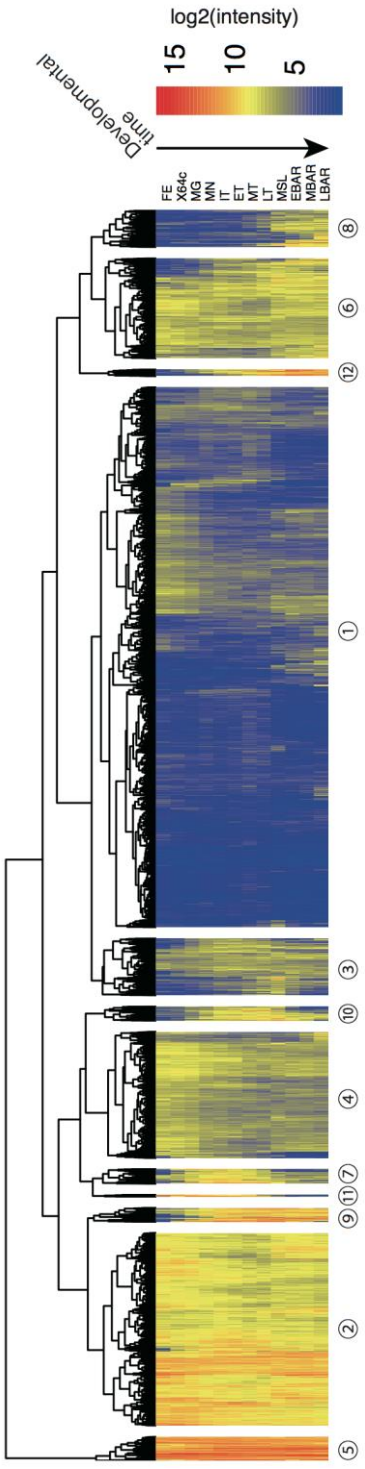


Figure 3.23: Transcriptome of *Ciona intestinalis*

Labels represent FE, fertilized egg; X64c, 64-cell; MG, mid gastrula; MN, mid neurula; IT, initial tailbud; ET, early tailbud; MT, mid tailbud; LT, late tailbud; MSL, mid swimming larva; EBAR, early body-axis-rotation; MBAR, mid body-axis-rotation; LBAR, late body-axis-rotation; respectively. Hierarchal clustering of microarray target gene models. Genes were ordered by hierarchical clustering on the x-axis. Sampled stages were ordered by developmental process on the y-axis. Numbers in the circle below each cluster correspond to the group name of gene expression profile clusters shown in Fig. 3.24.

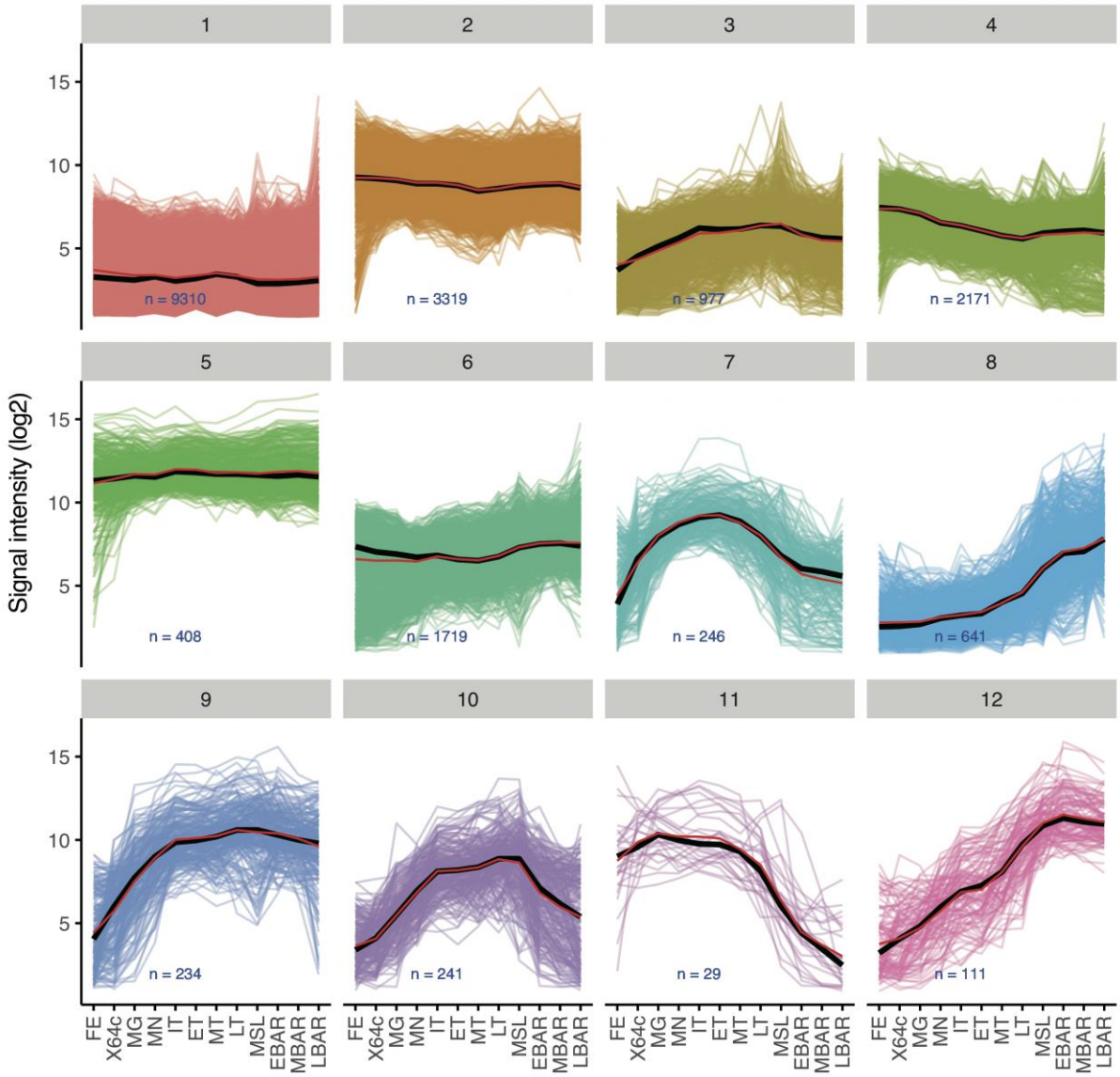


Figure 3.24: Transcriptome of *Ciona intestinalis*

Labels represent FE, fertilized egg; X64c, 64-cell; MG, mid gastrula; MN, mid neurula; IT, initial tailbud; ET, early tailbud; MT, mid tailbud; LT, late tailbud; MSL, mid swimming larva; EBAR, early body-axis-rotation; MBAR, mid body-axis-rotation; LBAR, late body-axis-rotation; respectively. The y-axis represents genes, as clustered into 12 groups, while the x-axis represents sampled developmental stages. The grey title bar in each box representing the name of the group corresponds to the numbers in the circles shown in Fig. 3.23. Developmental time is represented from left to right along the x-axis in each box. Normalized signal intensity is shown along the y-axis of each box. Red line represents mean of gene expression profiles and black line represents median, respectively, in each box.

3.3. Comparative analysis of developmental transcriptome in four deuterostome taxa

An hourglass-like pattern has been observed in the developmental transcriptome by quantifying genome-wide gene expression at different developmental stages or periods. A variety of methods has been used to quantify developmental stages of a wide range of organisms (Domazet-Loso and Tautz, 2010; Kalinka et al., 2010; Irie and Kuratani, 2011; Levin et al., 2012; Quint et al., 2012; Wang et al., 2013; Gerstein et al., 2014; Ninova et al., 2014; Cheng et al., 2015; Drost et al., 2016; Levin et al., 2016; Xu et al., 2016). In particular, two methods have been used to demonstrate this hourglass pattern. Ortholog-based analysis was used for a comparison of gene expression profiles among four groups of vertebrates (fish, amphibian, bird and mammal) (Irie and Kuratani, 2011), and later for a comparison of ten animal taxa (Levin et al., 2016). The other method is tree-based analysis, which was used for comparison of gene expression profiles during zebrafish embryogenesis (Domazet-Loso and Tautz, 2010).

(a) Ortholog-based method: Gene expression profiles observed in developmental transcriptome are a product of gene expressions due to complicated combinations of genes and spatiotemporal patterns. In ortholog-based analysis, an orthologous gene, especially a single copy ortholog, is used for pairwise comparison of developmental stages across species. This owes to an idea that ortholog genes would share homologous features such as regulatory mechanisms and functions. Thus, orthologous genes shared between different organisms are comparable. The expression of ortholog genes was compared to evaluate similarity between different developmental stages. Highly similar stages across species or groups can be regarded as candidate phylotypic stages. If the stages compared are homologous or adequately associated, then conservation of the developmental hourglass model becomes an option.

(b) Tree-based method: Tree-based analysis refers to the combination of developmental transcriptome and phylostratigraphy or related evolutionary indices. Phylostratigraphy, which was introduced by Domazet-Loso et al. (2007), classifies a gene into a relative age or phylostratum according to evolutionary context, to allow the first appearance of a new gene family that may be related to functional novelties to be investigated. Later, genomic phylostratigraphy was combined with developmental transcriptomic data, which showed the relative activity of gene age at a particular developmental stage (Domazet-Loso and Tautz, 2010). This quantification was named transcriptome age index (TAI). By regarding a

phylostratum as an evolutionary index and then changing the type of index, one could obtain a different type of score for each developmental transcriptome. For example, dN/dS ratio that measures adaptive selections of genes and pn/ps ratio that measures sequence diversity were incorporated as alternative options for evolutionary index (Quint et al., 2012; Gossmann et al., 2016). TAI or related indices were calculated for each stage. If plotted scores form "relatively young or less conserved"-"relatively old or conserved"-"relatively young or less conserved" patterns (*e.g.* "high TAI value"-"low TAI value"-"high TAI value" pattern in the case of TAI profile), this was regarded as the developmental hourglass model that has been conserved in the developmental processes of the target organisms.

3.3.1. Ortholog-based analysis

First, single copy orthologs were searched among paired target organisms, and the number of ortholog genes between different pairs of organisms determined (Table 3.7). The similarity of gene expression profiles across developmental stages of target organisms was analyzed (Fig. 3.25). In Fig. 3.25a, target organisms and sampled stages are aligned both horizontally and vertically. The larger square enclosed by bold lines indicates pairwise comparison between two organisms. Smaller squares inside the large indicate comparison between stages. The small squares are colored according to the Spearman's rank correlation coefficient *rho*; red color denotes a high similarity between stages, while blue color indicates a low similarity.

Table 3.7: Number of shared single copy orthologs

Ciona intestinalis vs *Branchiostoma floridae*	4,408
Ciona intestinalis vs *Ptychodera flava*	3,082
Ciona intestinalis vs *Acanthaster planci*	4,629
Branchiostoma floridae vs *Ptychodera flava*	3,985
Branchiostoma floridae vs *Acanthaster planci*	5,575
Ptychodera flava vs *Acanthaster planci*	4,747
Shared orthologs across all four species	2,636

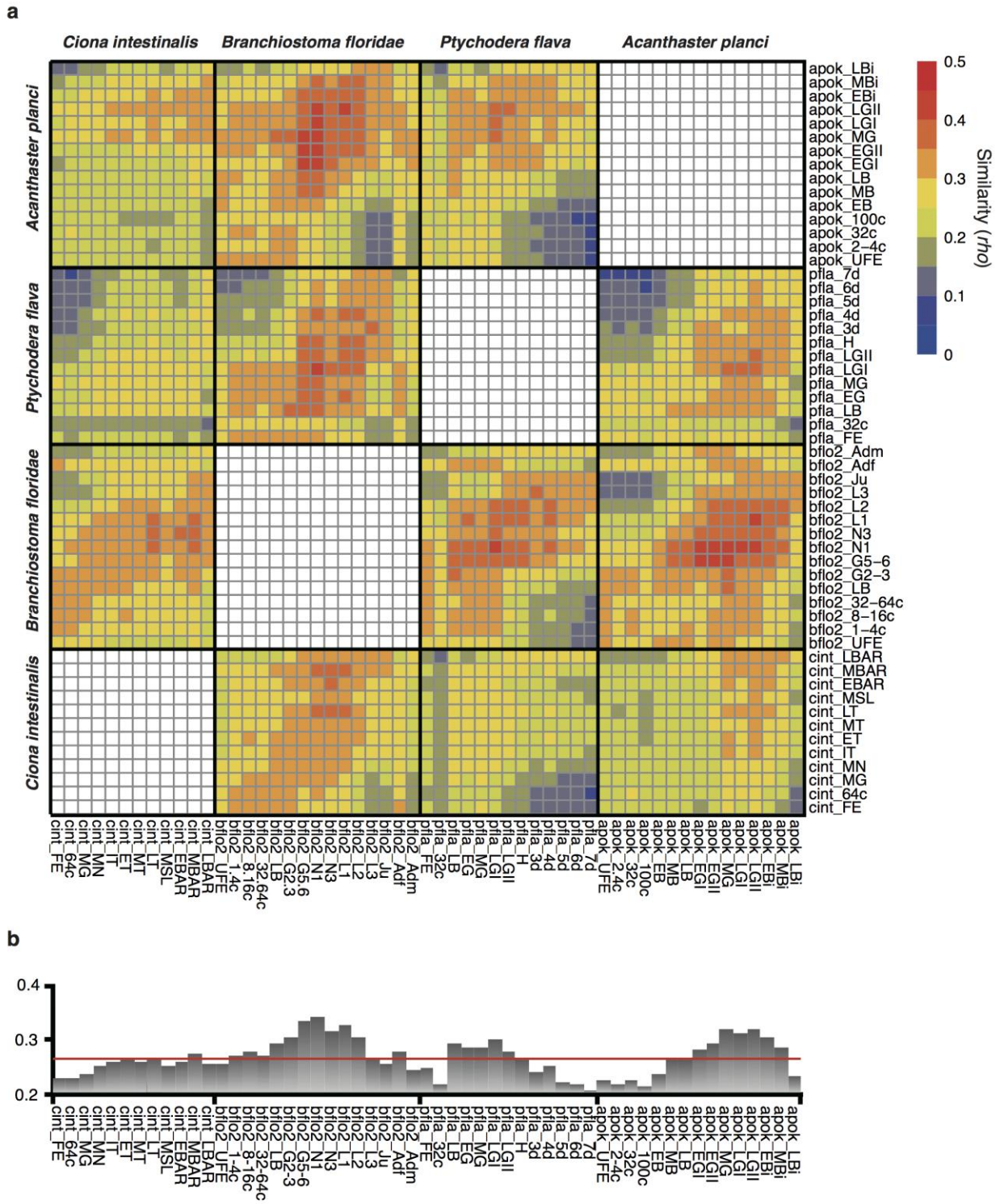


Figure 3.25: Conservation of gene expression similarity across basal deuterostomes

(a) Similarity of transcriptomes among paired developmental stages of organisms. Single copy orthologs shared by all four taxa were used. A color in a grid corresponds to the Spearman's rank correlation coefficient *rho*. (b) Mean correlation coefficient for each developmental stage against all of the other stages. Red line represent median. cint, *Ciona intestinalis*; bflo2, *Branchiostoma floridae*; pfla, *Ptychodera flava*; apok, *Acanthaster planci*.

The results shown in Fig. 3.25 suggest that some patterns exist. When comparing between each pair (*i.e.* between *A. planci* and *B. floridae* or between *P. flava* and *B. floridae*), the highest (red) or higher (orange) similarity tends to be located near the center of each large comparison square, as enclosed by bold black lines. Developmental stages in the center correspond to those of late blastula, gastrula and early larva. The lowest (blue) or lower (purple) similarity scores tend to be distributed around the edges, and correspond to stages of cleavage-stage embryo and late larva. Taken together, this suggests that the developmental stages of blastula, gastrula and early larva exhibit similar gene expression profiles, while the gene expression patterns do not resemble each other at earlier or later stages.

When overall similarity was examined amongst the four pairs, the number of red areas (showing higher similarity) or orange areas (showing moderate similarity) were lower when *Ciona* was compared against *A. planci, P. flava* and *B. floridae* (Fig. 3.25a). This suggests that the gene expression profile of the urochordate *C. intestinalis* differs from those of the three other deuterostome taxa. This is more evident when the mean correlation coefficient was displayed for each developmental stage of the four taxa (Fig. 3.25b). In this figure, the red line represents median of the grade of gene expression. *B. floridae, P. flava* and *A. planci* showed peaks higher than the median line; from late blastula to larva stages for *B. floridae,* from late blastula to newly hatched larva for *P. flava,* and from early gastrula to early larva for *P. flava,* respectively. These stages correspond to those with higher similarity in the pairwise comparisons shown in Fig. 3.25a. On the other hand, *C. intestinalis* showed almost no peaks above the median line, suggesting less similarity of the gene expression profiles in *C. intestinalis.*

3.3.2. Tree-based analysis

First, the number of genes used for the analysis obtained for each phylostratum were:

- *C. intestinalis*: Cellular Organisms (6418); Eukaryota (4874); Opisthokonta (650); Metazoa (1037); Eumetazoa (523); Bilateria (650); Deuterostomia (124); Chordata (351); Tunicata (43); Ascidiacea (7); Ciona (4); and *Ciona intestinalis* (3633).
- *B. floridae*: Cellular Organisms (12669); Eukaryota (6571); Opisthokonta (1101); Metazoa (1857); Eumetazoa (1035); Bilateria (1055); Deuterostomia (284); Chordata (331); Branchiostoma (43); and *Branchiostoma floridae* (3614).
- *P. flava*: Cellular Organisms (9502); Eukaryota (6252); Opisthokonta (1251); Metazoa (2625); Eumetazoa (2022); Bilateria (2074); Deuterostomia (1570); Enteropneusta (1338); and *Ptychodera flava* (7967).
- *A. planci*: Cellular Organisms (7026); Eukaryota (5215); Opisthokonta (778); Metazoa (1734); Eumetazoa (1129); Bilateria (1331); Deuterostomia (820); Eleutherozoa (490); and *Acanthaster planci* (5798).

Then, a set of TAI scores was calculated for each organism (Fig. 3.26). Fig. 3.26 shows the comparison of phylostratigraphy of the echinoderm *Acanthaster planci*, the hemichordate *Ptychodera flava*, the cephalochordate *Branchiostoma floridae*, and the urochordate *Ciona intestinalis*; (a) comparing all four taxa, (b) comparing the two ambulacrarian taxa, and (c) comparing two chordate taxa of *B. floridae* and *C. intestinalis.* TAI score (evolutionary younger genes with higher score while evolutionary old genes with lower score) was shown on the y-axis and developmental stages on the x-axis. In the figure, conserved gene expression profiles appear as peaks of high-low-high pattern.

Fig. 3.26 shows the relative age of expressed genes across (a) all four invertebrate deuterostome taxa, (b) the two ambulacrarian taxa and (c) the two invertebrate chordate taxa. Because the invertebrate chordates (cephalochordates and urochordates) contain the stage of neurulae and tailbud-like larvae (shown in pink), the phylostratigraphies of the four taxa may not always be compared directly. Nevertheless, comparison of the phylostratigraphy demonstrated the following points; first, the echinoderm, the hemichordate and the cephalochordate exhibited five, three and three peaks of high-low-high pattern, respectively (Fig. 3.26a). The five peaks of *A. planci* appear at the 32-cell stage, mid blastula stage, mid gastrula stage, late gastrula stage and mid larva stage (Fig. 3.26a, upper). The three peaks of *P. flava* appear at the 32-cell stage, mid gastrula stage, and newly hatched larva stage (Fig. 3.26a, upper middle). The peaks of *B. floridae* appear at the 32~64-cell stage, neurula stage

and early larva stage (Fig. 3.26a, lower middle). In contrast, *C. intestinalis* showed a different pattern. After fertilization, the score gradually increased in the mid tailbud stage, gradually decreased down larval stages, and then slightly increased again during later juvenile stages (Fig. 3.26a, lower). The results of the gene expression profile examined by tree-based analyses indicate that the *C. intestinalis* profile was distinct among the four invertebrate deuterostome taxa, as consistent with the ortholog-based analyses.

Fig. 3.26b is the comparison of *A. planci* and *P. flava* (ambulacrarians). Although sampled stages were not always comparable, the two taxa showed common peaks at 32-cell stage, mid blastula stage and hatched larva stage. This suggests a similarity in the gene expression profiles during early embryonic development among ambulacrarians. The result coincides with that obtained by the ortholog-based analyses. Comparison of the phylostratigraphy between cephalochordates and urochordates again showed that the two look different to each other (Fig. 3.26c). Together with comparison of the four invertebrate deuterostome taxa, this suggests that urochordates have a unique or derived pattern of gene expression compared to cephalochordates. The highest peak in *C. intestinalis* was at the middle tailbud stage (Fig. 3.26), suggesting a divergent repertoire of genes being expressed.

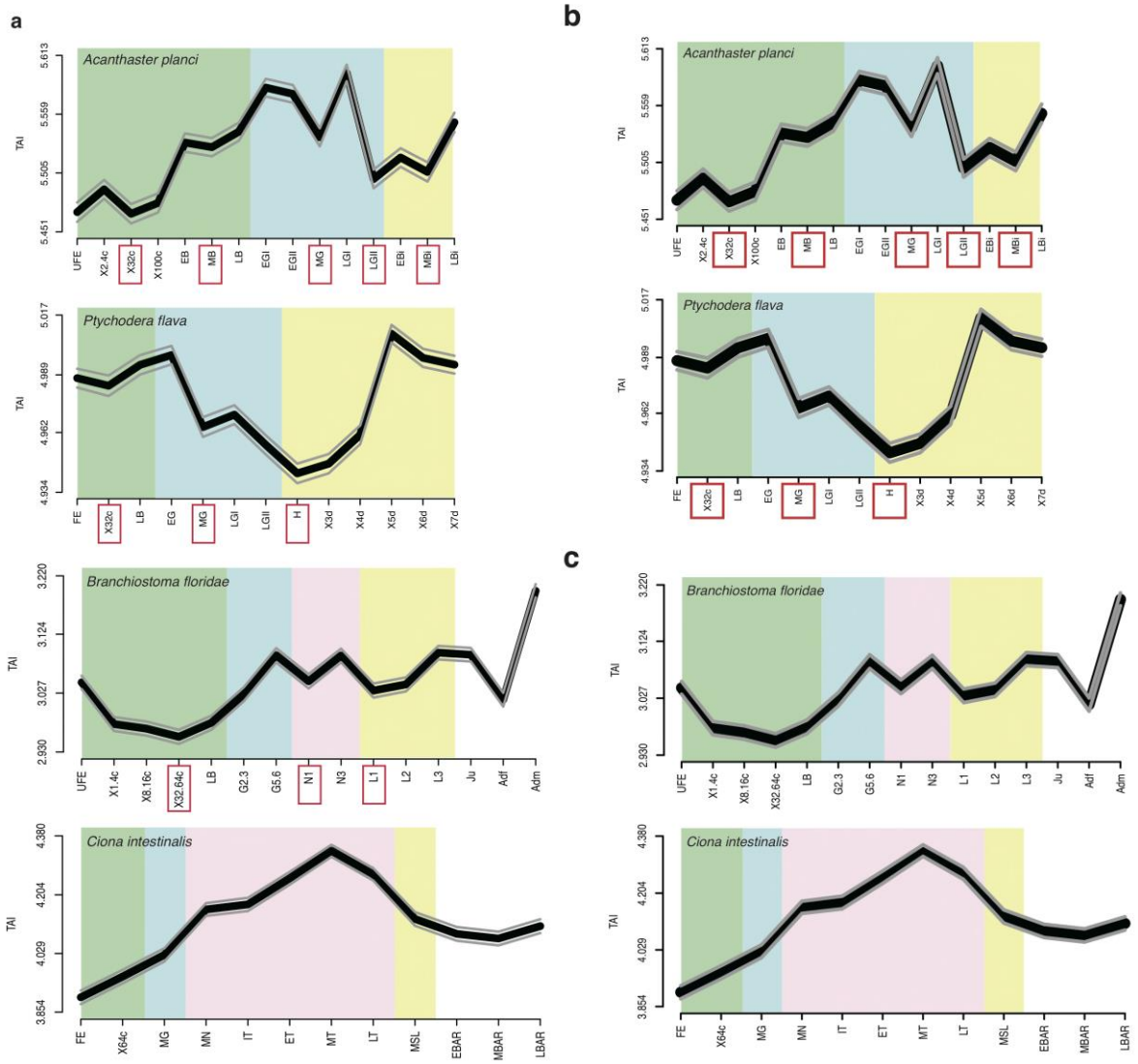


Figure 3.26: Relative age of expressed genes across basal deuterostomes

Transcriptome age index (TAI) applied to *Acanthaster planci*, *Ptychodera flava*, *Branchiostoma floridae*, and *Ciona intestinalis* from top to bottom. X-axis represents sampled stages in time course order. Y-axis represents TAI where a larger value indicates relatively younger transcriptome activity while smaller value indicates relatively old transcriptome activity. Green indicates cleavage-blastula period, blue indicates gastrula period, pink indicates neurula-tailbud period, and yellow indicates larval period. (a) Comparison of all four taxa, (b) comparison of the two ambulacrarian taxa of *A. planci* and *P. flava*, and (c) comparison between two chordate taxa of *B. floridae* and *C. intestinalis*. Red boxes on x-axes represent peaks of high-low-high pattern of TAI.

Lastly, in order to compare whether ortholog- and tree-based analyses correlate to each other, TAI analysis for zebrafish as a representative of vertebrates, was performed. Domazet-Loso and Tautz (2010) carried out TAI analysis and showed that the 24h-embryo exhibits a waist-like narrowness pattern: the phylotypic stage. Approximately 30% of the transcriptome data were used after verification since there were many missing values among the data. The zebrafish TAI profile peaked at almost the same developmental stages as described by Domazet-Loso and Tautz (2010) (Fig. 3.27). Namely, an upward peak at gastrula stage, a first downward peak at around 24h-embryo stage (corresponding to the pharyngula stage), and a second downward peak around the 14-day larva stage; thus the results of the TAI analysis are not inconsistent with those obtained by the ortholog-based analysis.

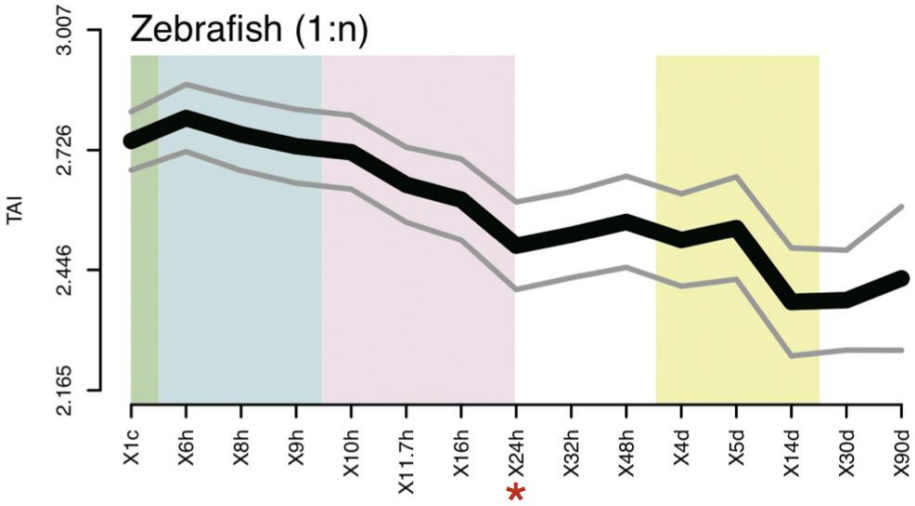


Figure 3.27 Relative age of expressed genes across vertebrates

Transcriptome age index (TAI) was performed on *Danio rerio*. X-axis represents sampled stages in order of developmental process. Y-axis represents TAI where a larger value indicates highly conserved (*i.e.* evolutionary older) transcriptome activity while a smaller value indicates less conserved (*i.e.* evolutionary younger) transcriptome activity. Green represents cleaving period, blue represents gastrula period, pink represent neurula-tailbud period, and yellow represent larval period. Accepted multiple types of phylostratum, depicted as "1:n" is shown. Red asterisk denotes the proposed vertebrate phylotypic stage.

4. Discussion

According to Kalinka and Tomancak (2012), there are four models to explain patterns of embryonic development in relation to conservation and diversification in forms of embryos, larvae, juveniles and adults (Fig. 1.2). They are (a) early conservation, (b) developmental hourglass model, (c) adaptive penetrance, and (d) ontogenetic adjacency. Of these, recent transcriptome studies of gene expression profiles support the developmental hourglass model (see below). The developmental hourglass model was originally discussed by comparison of egg morphology, mode of embryogenesis, and juvenile forms across vertebrate classes such as fish, amphibians, reptiles, birds and mammals. In the model, the conservation is exhibited by a narrow waist of the developmental hourglass, called the phylotypic stage or period, which corresponds to the pharyngeal stage (Duboule, 1994; Raff, 1996; Irie and Kuratani, 2011). The conservation is considered greatest in mid-embryogenesis and is either the result of the need for coordination between growth and patterning as the body plan is being built (Duboule, 1994), or the result of a global increase in the complexity of interactions between genes and developmental processes during the phylotypic period (Raff, 1996).

Recently, a number of studies have explored evidence for the developmental hourglass model. First, the model is supported by changes in gene expression pattern during vertebrate embryogenesis (Irie and Kuratani, 2011; Wang et al., 2013). The stage that shows a narrow waist for gene expression profile and higher similarity of gene expression corresponds approximately to the pharyngeal stage (Irie and Kuratani, 2011). Second, evidence for the developmental hourglass model has also been found in metazoans outside vertebrates, including *Drosophila* (Kalinka et al., 2010), *Caenorhabditis* (Levin et al., 2012) and trochozoans including the Pacific oyster, the Pacific abalone and sand worm (Xu et al. (2016). Third, evidence for the model has also been proposed in plants (Quint et al., 2012; Drost et al., 2016) and fungi (Cheng et al., 2015). Therefore, the urochordate *Ciona intestinalis,* the cephalochordate *Branchiostoma floridae,* the hemichordate *Ptychodera flava,* and the echinoderm *Acanthaster planci* may provide evidence for additional taxa in which the hourglass model can be observed.

4.1. An improvement of microarray system in non-model animals

The tool used here to examine the gene expression profiles was microarray. DNA microarray and deep sequencing technologies have made it possible to obtain high throughput quantification of transcriptomes (Morozova et al., 2009; Malone and Oliver, 2011). Usually, a custom-made microarray is used for gene expression experiments. However, most of the custom-made microarrays are oriented based on gene information of model organisms, and are not always best for non-model organisms such as those collected from a wild population. Thus, improvements in the microarray probe design for non-model deuterostomes were attempted, using the starfish *Acanthaster planci* as a case study.

DNA microarrays use single stranded probes, which are hybridized with fluorescently labeled target transcripts to assess the intensity of gene expression. The outcome of microarray analysis is limited by probe quality in terms of specificity of hybridization, sensitivity of detection, uniformity in size and distribution of spots on the microarray, and complete coverage of the target gene models. Thus, improvements in probe quality lead to more reliable microarray results. Since microarrays have great capacity in terms of the total number of spots or probes, probe optimization for custom microarray design is recommended. DNA microarrays have been optimized for model organisms with well-defined gene models and well-documented alternative splice variants. On the other hand, microarrays for non-model organisms from wild populations require careful design.

DNA microarrays are biased because they require prior knowledge of gene sequences. They can detect only known gene models. On the hand, sequencing itself is unbiased because it requires no prior knowledge of gene models and because it can be used to determine unknown sequences of genes and transcripts. Deep sequencing technology dominates genome-wide studies (Goodwin et al., 2016). In this study, the echinoderm *Acanthaster planci,* the hemichordate *Ptychodera flava,* the cephalochordate *Branchiostoma floridae,* and the urochordate *Ciona intestinalis* were selected. Genomes of all these species have been decoded: *Ciona intestinalis* in 2002 (Dehal et al., 2002), *Branchiostoma floridae* in 2008 (Putnam et al., 2008), *Ptychodera flava* in 2015 (Simakov et al., 2015), and *Acanthaster planci* in 2017 (Hall et al., 2017), indicating that all of the four specimens have well-defined gene models. Nevertheless, the gene models of these marine invertebrates are not always thoroughly characterized. For example, the *B. floridae* genome assembly has two versions, the older version-1 and the newer version-2. *B. floridae* microarray probes were

designed based on the version-1 assembly. The work in this thesis improved the method so that *B. floridae* genome assembly version-2 was used as target sequences of probes. Full-length mRNA from version-1 was searched against version-2 using BLAST. Protein sequences were obtained from genome assembly version-2 gff file. Another example is *A. planci* microarray. There are two gene models, one obtained from a specimen of Great Barrier Reef (GBR), Australia and the other from Okinawa (OKI), Japan. Originally, the BGR and OKI models contain 24,747 and 24,323 gene models for mRNAs and proteins, respectively. In this study, the two models were optimally merged together to obtain 28,380 model genes.

Together, these efforts improved the probe design for *A. planci.* Details of methods and results were described in the Result section 3.1. The DNA microarray designed by this method was evaluated by comparing it to the default Agilent eArray microarrays. Overall comparison of the three methods indicates that the method of this study is advantageous when probes are designed to suppress cross-hybridization. In addition, when the *A. planci* reference gene model is used, the method of this study covers more gene model. The eArray best quality method produces a set of high quality probes. However, the best quality method makes many probes that overlap one another narrowly within a given gene model. In contrast, the best distribution is advantageous for producing multiple probes that are well distributed along a gene model. The method developed here yields the best results when probes were designed for high specificity as well as complete coverage of gene models.

The new method developed in the present study provides insight into the design of microarrays and is applicable to most non-model metazoans. The rapidly increasing availability of sequence data for non-model organisms has expanded the options for experimental design of gene expression analysis, including the choice of quantitative RT-PCR, microarray, or high-throughput sequencing, depending on the number of samples and genes. In addition, this method is not limited only to examination of gene expression in one organism, but can also be used for gene identification and microbe detection.

4.2. Characterization of gene expression profiles during embryogenesis of invertebrate deuterostomes

Several methods were applied for the gene expression profiles during embryogenesis of each of the four deuterostome groups. First, the developmental trajectory was determined by connecting PCA and developmental time (Fig. 3.6 for *A. planci*, Fig. 3.11 for *P. flava*, Fig. 3.16 for *B. floridae,* and Fig. 3.21 for *C. intestinalis*). Developmental transcriptome analyzed by PCA result in a bowl-shaped trajectory, which implies less overlap and/or intersection of components (Anavy et al., 2014; Owens et al., 2016). This pattern was confirmed in *A. planci* (Fig. 3.6), *B. floridae* (Fig. 16)*,* and *C. intestinali*s (Fig. 3.21). However, such pattern was not obtained in *P. flava* (Fig. 3.11). *P. flava* showed a complex profile with several overlapping components. This was possibly caused by mixture of specimens at different developmental stages, obtained from batches with less synchronous development. Future studies should collect embryos and larvae with fewer stage contaminations, and transcriptome analyses should be repeated using appropriate sampling.

Small-scale differences were present among the four taxa examined by dendrogram analyses (Fig. 3.7 for *A. planci*, Fig. 3.12 for *P. flava*, Fig. 3.17 for *B. floridae,* and Fig. 3.22 for *C. intestinalis*). In most cases, the dendrogram is comprised of two clades; one includes early embryonic stages up to gastrulae, and the other includes stages of gastrulae, larvae, and juveniles. This suggests the significance of the mid-blastula transition; namely, transition in gene expression profiles from maternally controlled ones to zygotically controlled ones. Furthermore, the latter clades were further subdivided into several smaller groups, suggesting different gene expression profiles depend on differentiation of various cell types in different stages of larvae and juveniles.

Classification of gene expression profiles showed several interesting profiles. Gene expression patterns were arbitrarily divided into twelve groups depending on their similarity (Fig. 3.8 for *A. planci*, Fig. 3.13 for *P. flava*, Fig. 3.18 for *B. floridae,* and Fig. 3.23 for *C. intestinalis*). In ambulacrarians *A. planci* (Fig. 3.8) and *P. flava* (Fig. 3.13), group 1 occupies a large proportion of the figure. These genes are constantly expressed at a lower level from unfertilized egg stage to late larval stage (Figs. 3.8 and 3.13). Genes included in this group are likely to have a function as house-keeping genes. These genes were annotated, as shown in Supplementary Tables 1-4. In chordate deuterostomes (Figs. 3.18 and 3.23), the proportion of group 1 genes was as large when compared to ambulacrarian groups. Instead,

groups with genes with different expression patterns increased. At present, it is not clear whether this overall difference is associated with the presence of neurula and tailbud stages in chordate embryogenesis, which is absent in ambulacrarians. This should be examined in future studies.

In the developmental gene expression profile analyses (Fig. 3.9 for *A. planci*, Fig. 3.14 for *P. flava*, Fig. 3.19 for *B. floridae,* and Fig. 3.25 for *C. intestinalis*), the presence of groups of genes with specific expression profiles were observed. Especially, groups 9, 10 and 11 in *A. planci*, groups 8, 11 and 12 in *P. flava,* groups 4, 9, 11 and 12 in *B. floridae,* and groups 11 and 13 in *C. intestinalis* appeared to be interesting. For example, in groups 11 and 12 in *P. flava* and groups 11 and 12 in *B. floridae,* the signal intensity of certain genes increased dramatically towards gastrula stages and decreased dramatically towards larval or juvenile stages. Since the number of such genes is not always large, future studies may identify the genes, whose function should then be examined in relation to such dynamic expression pattern.

Genes of the *B. floridae* group 4 are of particular interest, as the intensity or gene expression was comparatively high in unfertilized eggs, then suddenly decreased after fertilization, and increased suddenly again after larval stage. It is likely that genes in this group are involved in the formation of gametes. However, it should be mentioned that the present analysis provided overall changes of genes involved in a similar process and function. Such information cannot be obtained in research that analyzes independent genes one by one.

4.3. Comparative analyses of developmental gene expression profiles among deuterostome taxa

Comparison of developmental gene expression profiles among the four deuterostome taxa was done using two methods, frequently used for comparative analyses of developmental gene expression profiles. One is ortholog-based analysis and the other is tree-based analysis. The ortholog-based method calculates similarity of the expression profile of orthologous genes from multiple organisms (Kalinka et al., 2010; Irie and Kuratani, 2011; Levin et al., 2012; Levin et al., 2016). The tree-based method calculates conserved positions of gene expression in phylogeny, which requires single organisms (Domazet-Loso and Tautz, 2010; Quint et al., 2012; Drost et al., 2016; Xu et al., 2016).

Data for later stages are missing in the echinoderm *Acanthaster planci* and the hemichordate *Ptychodera flava* and chordate deuterostomes (cephalochordates and urochordates) contain the stage of neurulae- and tailbud-like larvae. Therefore, the phylostratigraphy obtained by the tree-based analyses of the four taxa may not be directly compared. Nevertheless, the echinoderm, the hemichordate and the cephalochordate exhibited a similar pattern. Namely, peaks appear at stages of cleaving embryos, gastrulae and early larvae. If the phylotypic stage is defined as the stage with the lowest score, these phylotypic stages of echinoderms, hemichordates and cephalochordates correspond to the blastula stage and early larval stage. In contrast, in the urochordate, the score gradually increased during the mid tailbud stage, gradually decreased down larval stages, and then increased during later juvenile stages.

The ortholog-based analyses provided further evidence for the similarity and dissimilarity of developmental gene expression profiles. Comparison of the four taxa as a whole suggests that the phylotypic period of *A. planci*, *P. flava*, and *B. floridae* occurs from gastrula to early larval stages (Fig. 3.25). On the other hand, the phylotypic stage of *C. intestinalis* is likely to be that of tailbud embryos and juveniles.

This thesis hypothesized that the developmental hourglass model with a vertebrate-like phylotypic stage is applicable to all deuterostome groups. Neither ortholog-based analysis (Fig. 3.25) nor tree-based analysis (Fig. 3.26) supported this hypothesis. Namely, a narrow waist of gene expression profile that corresponds to the vertebrate-type phylotypic stage was not found in urochordates, cephalochordates, hemichordates and echinoderms. The hourglass-model-like pattern of gene expression is specific to vertebrates but not shared

among all the deuterostome taxa, or even among the three chordate taxa. In other words, perhaps the pharyngeal phylotypic stage evolved after divergence of the vertebrate lineage, and may be recognized as a common feature of vertebrate groups.

It was proposed that gene expression profiles of cephalochordates and urochordates would be similar, based on morphological similarities in embryonic stages such as neurulae, tailbud embryos and fish- or tadpole-like larvae. However, this idea may also be incomplete, judging from results shown in Figs. 3.25 and 3.26. The urochordate shows a single peak during the mid tailbud stage, while echinoderms, hemichordates and cephalochordates show several sharp peaks (Fig. 3.26a). These two profiles appear quite different from each other. It should be emphasized that the mode of gene expression profiles in *Ciona* (urochordates) differs from that of cephalochordates (Fig. 3.26c) and from the two ambulacrarian taxa too (Fig. 3.26a). This may be related to previous proposals of urochordate specificity, such as an advanced filter-feeder hypothesis (Satoh, 2009). Urochordates have evolved as a specialist filter-feeder, in which embryogenesis proceeds very promptly compared to other deuterostomes. Bilateral cleavage, distinct lineage, early determination of developmental fates, gastrulation around 110-cell stage, newly-hatched larva with only 2600 constituent cells, larvae without feeding stage, and metamorphosis of larvae to juveniles within a few days, are all features specific to urochordates. The relationship in the gene expression profiles between ambulacrarian groups and vertebrates was missed, because the vertebrate pharyngula-like stage is not present in ambulacrarian embryos. No similarity in the mode of gene expression profiles between these two groups, suggesting an independency of the two groups in the mode of embryogenesis, was found in the present study.

The presence and absence of the vertebrate-type phylotypic stage among the four other groups of deuterostomes, gene expression profile has been discussed in relation to evolutionary conservation and/or constraints. Conservation may be recognized as a narrowness or peak of gene expression profiles. The tree-based analyses of the present study provide some clues about the evolution of the mode of embryogenesis, and the global pattern of gene expression. As shown in Fig. 3.26, the three taxa of deuterostomes, echinoderms, hemichordates and cephalochordates showed three common peaks, while urochordates exhibited no peaks, unique to this taxon. The common peaks appear at stages of cleaving embryos, gastrulae and early larvae. The distinct peak at cleaving embryos corresponds to mid-blastula transition, during which maternally controlled gene expression is replaced by zygotically controlled gene expression (Newport and Kirschner, 1982). A

recent study of Levin et al. (2016) compared gene expression profiles among ten animal groups including sponges, comb jellies, cnidarians and seven bilaterian groups. They found a similar peak at blastula stage as a constraint of gene expression profile, which they proposed as phylotypic stage.

The two other common peaks of the three deuterostome taxa were at the gastrula stage and early larva stage (Fig. 3.26). Gastrulation includes morphogenesis and forms two- or three-germ-layered embryos and formation of embryonic axes, antero-posterior axis and dorso-ventral axis (Stern, 2004; Wolpert, 2011). Genes involved in the gastrulation have been shown to conserve by most animals (Davidson, 2006), and therefore resulted in a sharp peak in the global gene expression profiles. Larval formation is also a significant event during embryogenesis. It is highly likely that transition from immobile embryos to mobile larvae requires expression of many genes, which may be shared by almost all animal taxa.

Again, the lack of the high-low-high pattern in the TAI profile during *Ciona* embryogenesis is an exceptional case, in which cleavages finish within a very short period of time (approximately seven-times cleavage by 4 hours after fertilization). This was also found in the ortholog-based analysis shown in Fig. 2.5b. The mean *rho* score of urochordates was lower than in cephalochordates, hemichordates and echinoderms, suggesting that the mode of urochordate cleavage is unique among deuterostomes. On the other hand, vertebrates have these three peaks, which might be lower than the peak corresponding to pharyngula stage, and thus not under-appreciated by previous studies. The pharyngula stage might be that stage at which gene expression is more conserved than other stages. In the present study, the two methods, ortholog-based and tree-based, did not always provide comparison of similar results. Although a common result using zebrafish transcriptome data was generated, the usage of the methods may require more careful identification of orthologs.

4.4. Future direction

The aim of the present study was to compare gene expression profiles among non-vertebrate deuterostomes and to determine if the phylotypic stage exists among the taxa or in each taxon. The original idea of the developmental hourglass model has been found from comparison of embryogenesis among different classes of vertebrates, namely, fish, frogs, reptiles, birds and mammals (Duboule, 1994; Raff, 1996). By comparing the developmental gene expression profiles among upper clades rather than classes of a given phylum, namely between phyla, the developmental stages that showed conserved profile of gene expression between echinoderms, hemichordates, and cephalochordates but not in urochordates were identified. One suggestion for future studies is that the method used in the present study does not always fit to the analysis of comparison among taxa. As the original comparison was among classes of vertebrates, one would also need to compare the developmental gene expression profile among classes of a given taxon. For example, five different classes of echinoderms (crinoids, asteroids, ophiuroids, echinoids, and holothuroids) could be compared to each other as in the case of vertebrates (McClay, 2011). The similarity of larval form has been discussed between ophiuroids and echinoids as pluteus-type larvae, and between crinoids and holothiuroids as dorioraria larvae (Brusca et al., 2016). The comparison of the gene expression profiles among echinoderm classes provides us insights in evolution of echinoderms.

Another interesting comparison may be between two hemichordates, *Saccoglossus kowalevskii* and *Ptychodera flava*. Since the former is a direct developer while the latter is an indirect developer, a comparison of the two may shed light on molecular changes of embryogenesis.

5. Conclusion

Based on the microarray analyses, this thesis analyzed the comprehensive gene expression patterns during embryogenesis of four deuterostome taxa, *Acanthaster planci* of the Phylum Echinodermata, *Ptychodera flava* of the Phylum Hemichordata, *Branchiostoma floridae* of the Phylum Cephalochordata, and *Ciona intestinalis* of the Phylum Urochordata. Several methods were employed to examine the developmental gene expression pattern. *A. planci, P. flava* and *B. floridae* exhibit similar profiles. The tree-based analysis indicated that the peaks appear at stages of cleaving embryos, gastrulae and early larvae. However, it is impossible to conclude that these peaks correspond to a vertebrate-type phylotypic stage. The urochordate showed a different pattern, in which, after fertilization, the score gradually increased towards the mid tailbud stage, and then gradually decreased towards larval stages and then rose again towards later juvenile stage. The profile might be modified by the "precocious" mode of urochordate embryogenesis. Although this thesis attempted to characterize the developmental gene expression profiles in the four deuterostome taxa, further methodological improvement is essential to complete such studies.

References

Aken, B. L., Ayling, S., Barrell, D., Clarke, L., Curwen, V., Fairley, S., Fernandez Banet, J., Billis, K., Garcia Giron, C., Hourlier, T., Howe, K., Kahari, A., Kokocinski, F., Martin, F. J., Murphy, D. N., Nag, R., Ruffier, M., Schuster, M., Tang, Y. A., Vogel, J. H., White, S., Zadissa, A., Flicek, P. & Searle, S. M. 2016. The Ensembl gene annotation system. *Database,* 2016**,** 1-19.

Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. 1990. Basic local alignment search tool. *J Mol Biol,* 215**,** 403-410.

Anavy, L., Levin, M., Khair, S., Nakanishi, N., Fernandez-Valverde, S. L., Degnan, B. M. & Yanai, I. 2014. BLIND ordering of large-scale transcriptomic developmental timecourses. *Development,* 141**,** 1161-1166.

Ashburner, M., Ball, C. A., Blake, J. A., Botstein, D., Butler, H., Cherry, J. M., Davis, A. P., Dolinski, K., Dwight, S. S., Eppig, J. T., Harris, M. A., Hill, D. P., Issel-Tarver, L., Kasarskis, A., Lewis, S., Matese, J. C., Richardson, J. E., Ringwald, M., Rubin, G. M. & Sherlock, G. 2000. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat Genet,* 25**,** 25-29.

Baughman, K. W., McDougall, C., Cummins, S. F., Hall, M., Degnan, B. M., Satoh, N. & Shoguchi, E. 2014. Genomic organization of Hox and ParaHox clusters in the echinoderm, Acanthaster planci. *Genesis,* 52**,** 952-958.

Bertrand, S. & Escriva, H. 2011. Evolutionary crossroads in developmental biology: amphioxus. *Development,* 138**,** 4819-4830.

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.

Brusca, R., Moore, W. & Shuster, S. 2016. *Invertebrates*, Sinaeur Associates.

Brusca, R. C. & Brusca, G. J. 2003. *Invertebrates*, Sinauer Associates.

Cameron, C. B., Garey, J. R. & Swalla, B. J. 2000. Evolution of the chordate body plan: new insights from phylogenetic analyses of deuterostome phyla. *Proc Natl Acad Sci U S A,* 97**,** 4469-4474.

Carroll, S., Grenier, J. & Weatherbee, S. 2004. *From DNA to diversity: molecular genetics and the evolution of animal design*, Wiley-Blackwell.

Cheng, X., Hui, J. H., Lee, Y. Y., Wan Law, P. T. & Kwan, H. S. 2015. A "developmental

hourglass" in fungi. *Mol Biol Evol,* 32**,** 1556-1566.

Chia, F., Oguro, C. & Komatsu, M. 1993. Sea-star (asteroid) development. *Oceanogr Mar Biol Annu Rev,* 31**,** 223-257.

Comte, A., Roux, J. & Robinson-Rechavi, M. 2010. Molecular signaling in zebrafish development and the vertebrate phylotypic period. *Evol Dev,* 12**,** 144-156.

Conklin, E. G. 1905. Mosaic development in ascidian eggs. *J Exp Zool A Ecol Genet Physiol,* 2**,** 145-223.

D'Haeseleer, P., Liang, S. & Somogyi, R. 2000. Genetic network inference: from co-expression clustering to reverse engineering. *Bioinformatics,* 16**,** 707-726.

Darwin, C. 1859. *On the origin of the species*, John Murray.

Davidson, E. H. 2001. *Genomic regulatory systems: in development and evolution*, Academic Press.

Davidson, E. H. 2006. *The regulatory genome: gene regulatory networks in development and evolution*, Academic press.

Dehal, P., Satou, Y., Campbell, R. K., Chapman, J., Degnan, B., De Tomaso, A., Davidson, B., Di Gregorio, A., Gelpke, M., Goodstein, D. M., Harafuji, N., Hastings, K. E., Ho, I., Hotta, K., Huang, W., Kawashima, T., Lemaire, P., Martinez, D., Meinertzhagen, I. A., Necula, S., Nonaka, M., Putnam, N., Rash, S., Saiga, H., Satake, M., Terry, A., Yamada, L., Wang, H. G., Awazu, S., Azumi, K., Boore, J., Branno, M., Chin-Bow, S., DeSantis, R., Doyle, S., Francino, P., Keys, D. N., Haga, S., Hayashi, H., Hino, K., Imai, K. S., Inaba, K., Kano, S., Kobayashi, K., Kobayashi, M., Lee, B. I., Makabe, K. W., Manohar, C., Matassi, G., Medina, M., Mochizuki, Y., Mount, S., Morishita, T., Miura, S., Nakayama, A., Nishizaka, S., Nomoto, H., Ohta, F., Oishi, K., Rigoutsos, I., Sano, M., Sasaki, A., Sasakura, Y., Shoguchi, E., Shin-i, T., Spagnuolo, A., Stainier, D., Suzuki, M. M., Tassy, O., Takatori, N., Tokuoka, M., Yagi, K., Yoshizaki, F., Wada, S., Zhang, C., Hyatt, P. D., Larimer, F., Detter, C., Doggett, N., Glavina, T., Hawkins, T., Richardson, P., Lucas, S., Kohara, Y., Levine, M., Satoh, N. & Rokhsar, D. S. 2002. The draft genome of Ciona intestinalis: insights into chordate and vertebrate origins. *Science,* 298**,** 2157-2167.

Delsuc, F., Brinkmann, H., Chourrout, D. & Philippe, H. 2006. Tunicates and not cephalochordates are the closest living relatives of vertebrates. *Nature,* 439**,** 965-968.

Domazet-Loso, T., Brajkovic, J. & Tautz, D. 2007. A phylostratigraphy approach to uncover the genomic history of major adaptations in metazoan lineages. *Trends Genet,* 23**,** 533-539.

Domazet-Loso, T. & Tautz, D. 2010. A phylogenetically based transcriptome age index

mirrors ontogenetic divergence patterns. *Nature,* 468**,** 815-818.

Drost, H. G., Bellstadt, J., O'Maoileidigh, D. S., Silva, A. T., Gabel, A., Weinholdt, C., Ryan, P. T., Dekkers, B. J., Bentsink, L., Hilhorst, H. W., Ligterink, W., Wellmer, F., Grosse, I. & Quint, M. 2016. Post-embryonic Hourglass Patterns Mark Ontogenetic Transitions in Plant Development. *Mol Biol Evol,* 33**,** 1158-1163.

Drost, H. G., Gabel, A., Grosse, I. & Quint, M. 2015. Evidence for active maintenance of phylotranscriptomic hourglass patterns in animal and plant embryogenesis. *Mol Biol Evol,* 32**,** 1221-1231.

Duboule, D. 1994. Temporal colinearity and the phylotypic progression: a basis for the stability of a vertebrate Bauplan and the evolution of morphologies through heterochrony. *Dev Suppl***,** 135-142.

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.

Edgar, R., Domrachev, M. & Lash, A. E. 2002. Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. *Nucleic Acids Res,* 30**,** 207-210.

Gee, H. 1996. *Before the Backbone: Views on the origin of the vertebrates*, Springer.

Gene Ontology Consortium. 2015. Gene Ontology Consortium: going forward. *Nucleic Acids Res,* 43**,** D1049-1056.

Gerhart, J., Lowe, C. & Kirschner, M. 2005. Hemichordates and the origin of chordates. *Curr Opin Genet Dev,* 15**,** 461-467.

Gerstein, M. B., Rozowsky, J., Yan, K. K., Wang, D., Cheng, C., Brown, J. B., Davis, C. A., Hillier, L., Sisu, C., Li, J. J., Pei, B., Harmanci, A. O., Duff, M. O., Djebali, S., Alexander, R. P., Alver, B. H., Auerbach, R., Bell, K., Bickel, P. J., Boeck, M. E., Boley, N. P., Booth, B. W., Cherbas, L., Cherbas, P., Di, C., Dobin, A., Drenkow, J., Ewing, B., Fang, G., Fastuca, M., Feingold, E. A., Frankish, A., Gao, G., Good, P. J., Guigo, R., Hammonds, A., Harrow, J., Hoskins, R. A., Howald, C., Hu, L., Huang, H., Hubbard, T. J., Huynh, C., Jha, S., Kasper, D., Kato, M., Kaufman, T. C., Kitchen, R. R., Ladewig, E., Lagarde, J., Lai, E., Leng, J., Lu, Z., MacCoss, M., May, G., McWhirter, R., Merrihew, G., Miller, D. M., Mortazavi, A., Murad, R., Oliver, B., Olson, S., Park, P. J., Pazin, M. J., Perrimon, N., Pervouchine, D., Reinke, V., Reymond, A., Robinson, G., Samsonova, A., Saunders, G. I., Schlesinger, F., Sethi, A., Slack, F. J., Spencer, W. C., Stoiber, M. H., Strasbourger, P., Tanzer, A.,

Thompson, O. A., Wan, K. H., Wang, G., Wang, H., Watkins, K. L., Wen, J., Wen, K., Xue, C., Yang, L., Yip, K., Zaleski, C., Zhang, Y., Zheng, H., Brenner, S. E., Graveley, B. R., Celniker, S. E., Gingeras, T. R. & Waterston, R. 2014. Comparative analysis of the transcriptome across distant species. *Nature,* 512**,** 445-448.

Gilbert, S. F. 2013. *Developmental Biology*, Sinauer Associates Incorporated.

Goodwin, S., McPherson, J. D. & McCombie, W. R. 2016. Coming of age: ten years of next-generation sequencing technologies. *Nat Rev Genet,* 17**,** 333-351.

Gorman, A. L., McReynolds, J. S. & Barnes, S. N. 1971. Photoreceptors in primitive chordates: fine structure, hyperpolarizing receptor potentials, and evolution. *Science,* 172**,** 1052-1054.

Gossmann, T. I., Saleh, D., Schmid, M. W., Spence, M. A. & Schmid, K. J. 2016. Transcriptomes of plant gametophytes have a higher proportion of rapidly evolving and young genes than sporophytes. *Mol Biol Evol,* 33**,** 1669-1678.

Haas, B. J., Papanicolaou, A., Yassour, M., Grabherr, M., Blood, P. D., Bowden, J., Couger, M. B., Eccles, D., Li, B., Lieber, M., Macmanes, M. D., Ott, M., Orvis, J., Pochet, N., Strozzi, F., Weeks, N., Westerman, R., William, T., Dewey, C. N., Henschel, R., Leduc, R. D., Friedman, N. & Regev, A. 2013. De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nat Protoc,* 8**,** 1494-1512.

Halanych, K. M. 1995. The phylogenetic position of the pterobranch hemichordates based on 18S rDNA sequence data. *Mol Phylogenet Evol,* 4**,** 72-76.

Hall, B. 1999. *Evolutionary developmental biology*, Springer.

Hall, M. R., Kocot, K. M., Baughman, K. W., Fernandez-Valverde, S. L., Gauthier, M. E. A., Hatleberg, W. L., Krishnan, A., McDougall, C., Motti, C. A., Shoguchi, E., Wang, T., Xiang, X., Zhao, M., Bose, U., Shinzato, C., Hisata, K., Fujie, M., Kanda, M., Cummins, S. F., Satoh, N., Degnan, S. M. & Degnan, B. M. 2017. The crown-of-thorns starfish genome as a guide for biocontrol of this coral reef pest. *Nature,* 544**,** 231-234.

Hamada, M., Shimozono, N., Ohta, N., Satou, Y., Horie, T., Kawada, T., Satake, H., Sasakura, Y. & Satoh, N. 2011. Expression of neuropeptide- and hormone-encoding genes in the Ciona intestinalis larval brain. *Dev Biol,* 352**,** 202-214.

Hayashi, R., Komatsu, M. & Oguro, C. 1973. Wrinkled blastula of the sea star, Acanthaster planci (Linnaeus). *Proc Japanese Soc Syst Zool,* 9**,** 59-61.

Henderson, J. 1969. Preliminary observations on the rearing and development of Acanthaster planci (L)(Asteroidea) larvae. *Fisheries notes,* 3**,** 69-75.

Hirakow, R. & Kajita, N. 1990. An electron microscopic study of the development of amphioxus, Branchiostoma belcheri tsingtauense: cleavage. *Journal of Morphology,* 203**,** 331-344.

Holland, L. Z. 2013. Evolution of new characters after whole genome duplications: insights from amphioxus. *Semin Cell Dev Biol,* 24**,** 101-109.

Hotta, K., Mitsuhara, K., Takahashi, H., Inaba, K., Oka, K., Gojobori, T. & Ikeo, K. 2007. A web-based interactive developmental table for the ascidian Ciona intestinalis, including 3D real-image embryo reconstructions: I. From fertilized egg to hatching larva. *Dev Dyn,* 236**,** 1790-1805.

Irie, N. & Kuratani, S. 2011. Comparative transcriptome analysis reveals vertebrate phylotypic period during organogenesis. *Nat Commun,* 2**,** 248.

Irie, N. & Kuratani, S. 2014. The developmental hourglass model: a predictor of the basic body plan? *Development,* 141**,** 4649-4655.

Jefferies, R. P., Brown, N. A. & Daley, P. E. 1996. The early phylogeny of chordates and echinoderms and the origin of chordate left–right asymmetry and bilateral symmetry. *Acta Zoologica,* 77**,** 101-122.

Kalinka, A. T. & Tomancak, P. 2012. The evolution of early animal embryos: conservation or divergence? *Trends Ecol Evol,* 27**,** 385-393.

Kalinka, A. T., Varga, K. M., Gerrard, D. T., Preibisch, S., Corcoran, D. L., Jarrells, J., Ohler, U., Bergman, C. M. & Tomancak, P. 2010. Gene expression divergence recapitulates the developmental hourglass model. *Nature,* 468**,** 811-814.

Kanehisa, M. & Goto, S. 2000. KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res,* 28**,** 27-30.

Kanehisa, M., Sato, Y., Kawashima, M., Furumichi, M. & Tanabe, M. 2016. KEGG as a reference resource for gene and protein annotation. *Nucleic Acids Res,* 44**,** D457-462.

Kardong, K. V. 2009. *Vertebrates: comparative anatomy, function, evolution*, McGraw-Hill.

Keesing, J. K., Halford, A. R., Hall, K. C. & Cartwright, C. M. 1997. Large-scale laboratory culture of the crown-of-thorns starfish Acanthaster planci (L.)(Echinodermata: Asteroidea). *Aquaculture,* 157**,** 215-226.

Kusakabe, T., Araki, I., Satoh, N. & Jeffery, W. R. 1997. Evolution of chordate actin genes: evidence from genomic organization and amino acid sequences. *J Mol Evol,* 44**,** 289-298.

Lacalli, T. C. 2005. Protochordate body plan and the evolutionary role of larvae: old controversies resolved? *Can J Zool,* 83**,** 216-224.

Lamare, M., Pecorino, D., Hardy, N., Liddy, M., Byrne, M. & Uthicke, S. 2014. The thermal

tolerance of crown-of-thorns (Acanthaster planci) embryos and bipinnaria larvae: implications for spatial and temporal variation in adult populations. *Coral Reefs,* 33**,** 207-219.

Langmead, B. & Salzberg, S. L. 2012. Fast gapped-read alignment with Bowtie 2. *Nat Methods,* 9**,** 357-359.

Levin, M., Anavy, L., Cole, A. G., Winter, E., Mostov, N., Khair, S., Senderovich, N., Kovalev, E., Silver, D. H., Feder, M., Fernandez-Valverde, S. L., Nakanishi, N., Simmons, D., Simakov, O., Larsson, T., Liu, S. Y., Jerafi-Vider, A., Yaniv, K., Ryan, J. F., Martindale, M. Q., Rink, J. C., Arendt, D., Degnan, S. M., Degnan, B. M., Hashimshony, T. & Yanai, I. 2016. The mid-developmental transition and the evolution of animal body plans. *Nature,* 531**,** 637-641.

Levin, M., Hashimshony, T., Wagner, F. & Yanai, I. 2012. Developmental milestones punctuate gene expression in the Caenorhabditis embryo. *Dev Cell,* 22**,** 1101-1108.

Li, L., Stoeckert, C. J., Jr. & Roos, D. S. 2003. OrthoMCL: identification of ortholog groups for eukaryotic genomes. *Genome Res,* 13**,** 2178-2189.

Lin, C. Y., Tung, C. H., Yu, J. K. & Su, Y. H. 2016. Reproductive periodicity, spawning induction, and larval metamorphosis of the hemichordate acorn worm Ptychodera flava. *J Exp Zool B Mol Dev Evol,* 326**,** 47-60.

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., 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.

Lucas, J. 1982. Quantitative studies of feeding and nutrition during larval development of the coral reef asteroid Acanthaster planci (L.). *J Exp Mar Bio Ecol,* 65**,** 173-193.

Malone, J. H. & Oliver, B. 2011. Microarrays, deep sequencing and the true measure of the transcriptome. *BMC Biol,* 9**,** 34.

Martinez, D. E., Bridge, D., Masuda-Nakagawa, L. M. & Cartwright, P. 1998. Cnidarian homeoboxes and the zootype. *Nature,* 393**,** 748-749.

McClay, D. R. 2011. Evolutionary crossroads in developmental biology: sea urchins. *Development,* 138**,** 2639-2648.

Morozova, O., Hirst, M. & Marra, M. A. 2009. Applications of new sequencing technologies for transcriptome analysis. *Annu Rev Genomics Hum Genet,* 10**,** 135-151.

Nakano, H., Lundin, K., Bourlat, S. J., Telford, M. J., Funch, P., Nyengaard, J. R., Obst, M. & Thorndyke, M. C. 2013. Xenoturbella bocki exhibits direct development with

similarities to Acoelomorpha. *Nat Commun,* 4**,** 1537.

Newport, J. & Kirschner, M. 1982. A major developmental transition in early Xenopus embryos: I. characterization and timing of cellular changes at the midblastula stage. *Cell,* 30**,** 675-86.

Nielsen, C. 2012. *Animal evolution: interrelationships of the living phyla*, Oxford University Press.

Ninova, M., Ronshaugen, M. & Griffiths-Jones, S. 2014. Conserved temporal patterns of microRNA expression in Drosophila support a developmental hourglass model. *Genome Biol Evol,* 6**,** 2459-2467.

O'Leary, N. A., Wright, M. W., Brister, J. R., Ciufo, S., Haddad, D., McVeigh, R., Rajput, B., Robbertse, B., Smith-White, B., Ako-Adjei, D., Astashyn, A., Badretdin, A., Bao, Y., Blinkova, O., Brover, V., Chetvernin, V., Choi, J., Cox, E., Ermolaeva, O., Farrell, C. M., Goldfarb, T., Gupta, T., Haft, D., Hatcher, E., Hlavina, W., Joardar, V. S., Kodali, V. K., Li, W., Maglott, D., Masterson, P., McGarvey, K. M., Murphy, M. R., O'Neill, K., Pujar, S., Rangwala, S. H., Rausch, D., Riddick, L. D., Schoch, C., Shkeda, A., Storz, S. S., Sun, H., Thibaud-Nissen, F., Tolstoy, I., Tully, R. E., Vatsan, A. R., Wallin, C., Webb, D., Wu, W., Landrum, M. J., Kimchi, A., Tatusova, T., DiCuccio, M., Kitts, P., Murphy, T. D. & Pruitt, K. D. 2016. Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. *Nucleic Acids Res,* 44**,** D733-745.

Oda, H., Wada, H., Tagawa, K., Akiyama-Oda, Y., Satoh, N., Humphreys, T., Zhang, S. & Tsukita, S. 2002. A novel amphioxus cadherin that localizes to epithelial adherens junctions has an unusual domain organization with implications for chordate phylogeny. *Evol Dev,* 4**,** 426-434.

Owens, N. D., Blitz, I. L., Lane, M. A., Patrushev, I., Overton, J. D., Gilchrist, M. J., Cho, K. W. & Khokha, M. K. 2016. Measuring Absolute RNA Copy Numbers at High Temporal Resolution Reveals Transcriptome Kinetics in Development. *Cell Rep,* 14**,** 632-647.

Pearson, J. C., Lemons, D. & McGinnis, W. 2005. Modulating Hox gene functions during animal body patterning. *Nat Rev Genet,* 6**,** 893-904.

Perseke, M., Golombek, A., Schlegel, M. & Struck, T. H. 2013. The impact of mitochondrial genome analyses on the understanding of deuterostome phylogeny. *Mol Phylogenet Evol,* 66**,** 898-905.

Philippe, H., Derelle, R., Lopez, P., Pick, K., Borchiellini, C., Boury-Esnault, N., Vacelet, J., Renard, E., Houliston, E., Queinnec, E., Da Silva, C., Wincker, P., Le Guyader, H.,

Leys, S., Jackson, D. J., Schreiber, F., Erpenbeck, D., Morgenstern, B., Worheide, G. & Manuel, M. 2009. Phylogenomics revives traditional views on deep animal relationships. *Curr Biol,* 19**,** 706-712.

Poe, S. & Wake, M. H. 2004. Quantitative tests of general models for the evolution of development. *Am Nat,* 164**,** 415-422.

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.

Quint, M., Drost, H. G., Gabel, A., Ullrich, K. K., Bonn, M. & Grosse, I. 2012. A transcriptomic hourglass in plant embryogenesis. *Nature,* 490**,** 98-101.

Raff, R. A. 1996. *The Shape of Life: Genes, Development, and the Evolution of Animal Form*, University of Chicago Press.

Richardson, M. K. 1999. Vertebrate evolution: the developmental origins of adult variation. *Bioessays,* 21**,** 604-613.

Ritchie, M. E., Phipson, B., Wu, D., Hu, Y., Law, C. W., Shi, W. & Smyth, G. K. 2015. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res,* 43**,** e47.

Romer, A. S. 1967. Major steps in vertebrate evolution. *Science,* 158**,** 1629-1637.

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

Ruppert, E., Fox, R. & Barnes, R. 2004. *Invertebrate Zoology, A functional evolutionary approach*, Brooks/Cole.

Sasakura, Y., Inaba, K., Satoh, N., Kondo, M. & Akasaka, K. 2009. Ciona intestinalis and Oxycomanthus japonicus, representatives of marine invertebrates. *Exp Anim,* 58**,** 459-469.

Satoh, N. 2008. An aboral-dorsalization hypothesis for chordate origin. *Genesis,* 46**,** 614-622.

Satoh, N. 2009. An advanced filter-feeder hypothesis for urochordate evolution. *Zoolog Sci,* 26**,** 97-111.

Satoh, N. 2011. Tunicate Embryos and Cell Specification. *eLS*.

Satoh, N. 2013. *Developmental Genomics of Ascidians*, Wiley-Blackwell.

Satoh, N. 2016. *Chordate Origins and Evolution: The Molecular Evolutionary Road to Vertebrates*, Academic Press.

Satoh, N., Rokhsar, D. & Nishikawa, T. 2014a. Chordate evolution and the three-phylum system. *Proc Biol Sci,* 281**,** 20141729.

Satoh, N., Tagawa, K., Lowe, C. J., Yu, J. K., Kawashima, T., Takahashi, H., Ogasawara, M., Kirschner, M., Hisata, K., Su, Y. H. & Gerhart, J. 2014b. On a possible evolutionary link of the stomochord of hemichordates to pharyngeal organs of chordates. *Genesis,* 52**,** 925-934.

Satoh, N., Tagawa, K. & Takahashi, H. 2012. How was the notochord born? *Evol Dev,* 14**,** 56-75.

Shoguchi, E., Satoh, N. & Maruyama, Y. K. 2000. A starfish homolog of mouse T-brain-1 is expressed in the archenteron of Asterina pectinifera embryos: possible involvement of two T-box genes in starfish gastrulation. *Dev Growth Differ,* 42**,** 61-68.

Simakov, O., Kawashima, T., Marletaz, F., Jenkins, J., Koyanagi, R., Mitros, T., Hisata, K., Bredeson, J., Shoguchi, E., Gyoja, F., Yue, J. X., Chen, Y. C., Freeman, R. M., Jr., Sasaki, A., Hikosaka-Katayama, T., Sato, A., Fujie, M., Baughman, K. W., Levine, J., Gonzalez, P., Cameron, C., Fritzenwanker, J. H., Pani, A. M., Goto, H., Kanda, M., Arakaki, N., Yamasaki, S., Qu, J., Cree, A., Ding, Y., Dinh, H. H., Dugan, S., Holder, M., Jhangiani, S. N., Kovar, C. L., Lee, S. L., Lewis, L. R., Morton, D., Nazareth, L. V., Okwuonu, G., Santibanez, J., Chen, R., Richards, S., Muzny, D. M., Gillis, A., Peshkin, L., Wu, M., Humphreys, T., Su, Y. H., Putnam, N. H., Schmutz, J., Fujiyama, A., Yu, J. K., Tagawa, K., Worley, K. C., Gibbs, R. A., Kirschner, M. W., Lowe, C. J., Satoh, N., Rokhsar, D. S. & Gerhart, J. 2015. Hemichordate genomes and deuterostome origins. *Nature,* 527**,** 459-465.

Simao, F. A., Waterhouse, R. M., Ioannidis, P., Kriventseva, E. V. & Zdobnov, E. M. 2015. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics,* 31**,** 3210-3212.

Slack, J. M., Holland, P. W. & Graham, C. F. 1993. The zootype and the phylotypic stage. *Nature,* 361**,** 490-492.

Stern, C. D. 2004. *Gastrulation: From Cells to Embryo*, Cold Spring Harbor Laboratory Press.

Swalla, B. J. 2006. Building divergent body plans with similar genetic pathways. *Heredity (Edinb),* 97**,** 235-243.

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. *Zoolog Sci,* 15**,** 85-91.

Tagawa, K., Satoh, N. & Humphreys, T. 2001. Molecular studies of hemichordate development: a key to understanding the evolution of bilateral animals and chordates. *Evol Dev,* 3**,** 443-454.

Tang, H., Klopfenstein, D., Pedersen, B., Flick, P., Sato, K., Ramirez, F., Yunes, J. & Mungall, C. 2015. GOATOOLS: tools for gene ontology. *Zenodo.*

Tassy, O., Dauga, D., Daian, F., Sobral, D., Robin, F., Khoueiry, P., Salgado, D., Fox, V., Caillol, D., Schiappa, R., Laporte, B., Rios, A., Luxardi, G., Kusakabe, T., Joly, J. S., Darras, S., Christiaen, L., Contensin, M., Auger, H., Lamy, C., Hudson, C., Rothbacher, U., Gilchrist, M. J., Makabe, K. W., Hotta, K., Fujiwara, S., Satoh, N., Satou, Y. & Lemaire, P. 2010. The ANISEED database: digital representation, formalization, and elucidation of a chordate developmental program. *Genome Res,* 20**,** 1459-1468.

The UniProt Consortium. 2017. UniProt: the universal protein knowledgebase. *Nucleic Acids Res,* 45**,** D158-169.

Tokioka, T. 1971. Phylogenetic Speculation Of The Tunicata. *Publications Of The Seto Marine Biological Laboratory,* 19**,** 43-63.

Vandekerckhove, J. & Weber, K. 1984. Chordate muscle actins differ distinctly from invertebrate muscle actins. The evolution of the different vertebrate muscle actins. *J Mol Biol,* 179**,** 391-413.

Wada, H. & Satoh, N. 1994. Details of the evolutionary history from invertebrates to vertebrates, as deduced from the sequences of 18S rDNA. *Proc Natl Acad Sci U S A,* 91**,** 1801-1804.

Wagner, G. P., Pavlicev, M. & Cheverud, J. M. 2007. The road to modularity. *Nat Rev Genet,* 8**,** 921-931.

Wang, Z., Pascual-Anaya, J., Zadissa, A., Li, W., Niimura, Y., Huang, Z., Li, C., White, S., Xiong, Z., Fang, D., Wang, B., Ming, Y., Chen, Y., Zheng, Y., Kuraku, S., Pignatelli, M., Herrero, J., Beal, K., Nozawa, M., Li, Q., Wang, J., Zhang, H., Yu, L., Shigenobu, S., Wang, J., Liu, J., Flicek, P., Searle, S., Wang, J., Kuratani, S., Yin, Y., Aken, B., Zhang, G. & Irie, N. 2013. The draft genomes of soft-shell turtle and green sea turtle

yield insights into the development and evolution of the turtle-specific body plan. *Nat Genet,* 45**,** 701-706.

Wheeler, D. L., Church, D. M., Federhen, S., Lash, A. E., Madden, T. L., Pontius, J. U., Schuler, G. D., Schriml, L. M., Sequeira, E., Tatusova, T. A. & Wagner, L. 2003. Database resources of the National Center for Biotechnology. *Nucleic Acids Res,* 31**,** 28-33.

Whittaker, J. 1997. Cephalochordates, the lancelets. *Embryology: Constructing the Organism***,** 365-381.

Willmer, P. 1990. *Invertebrate relationships: patterns in animal evolution*, Cambridge University Press.

Wolpert, L. 2011. *Principles of Development*, Oxford University Press.

Xu, F., Domazet-Loso, T., Fan, D., Dunwell, T. L., Li, L., Fang, X. & Zhang, G. 2016. High expression of new genes in trochophore enlightening the ontogeny and evolution of trochozoans. *Sci Rep,* 6**,** 34664.

Yamazaki, Y., Akashi, R., Banno, Y., Endo, T., Ezura, H., Fukami-Kobayashi, K., Inaba, K., Isa, T., Kamei, K., Kasai, F., Kobayashi, M., Kurata, N., Kusaba, M., Matuzawa, T., Mitani, S., Nakamura, T., Nakamura, Y., Nakatsuji, N., Naruse, K., Niki, H., Nitasaka, E., Obata, Y., Okamoto, H., Okuma, M., Sato, K., Serikawa, T., Shiroishi, T., Sugawara, H., Urushibara, H., Yamamoto, M., Yaoita, Y., Yoshiki, A. & Kohara, Y. 2010. NBRP databases: databases of biological resources in Japan. *Nucleic Acids Res,* 38**,** D26-32.

Zeng, L. & Swalla, B. J. 2005. Molecular phylogeny of the protochordates: chordate evolution. *Can J Zool,* 83**,** 24-33.

Supplementary Data

Supplementary Table 1: Top 10 constitutive genes in *Acanthaster planci*

Gene model	UniProt ID	UniProt definition	E-value
High intensity cluster			
oki.110.2.t1	*		
oki.57.59.t1	P27449	V-type proton ATPase 16 kDa proteolipid subunit	1E-78
oki.294.4.t1	*		
oki.14.125.t1	Q6NUX8	Rho-related GTP-binding protein RhoA-A	4E-124
oki.10.71.t1	P62196	26S protease regulatory subunit 8	0.0
oki.89.62.t1	*		
oki.164.38.t1	Q9BRT6	Protein LLP homolog	1E-8
oki.117.19.t1	Q9Z0U1	Tight junction protein ZO-2	2E-91
oki.39.74.t1	Q9BXS5	AP-1 complex subunit mu-1	0.0
oki.130.13.t1	*		
Mid intensity cluster			
oki.124.11.t1	P21399	Cytoplasmic aconitate hydratase	0.0
oki.3.143.t1	O00560	Syntenin-1	1E-95
oki.199.25.t1	P36776	Lon protease homolog, mitochondrial	0.0
oki.76.20.t1	Q9HB07	UPF0160 protein MYG1, mitochondrial	4E-135
oki.236.13.t1	O76536	Hyalin (Fragment)	1E-108
oki.68.94.t1	P05300	Lysosome-associated membrane glycoprotein 1	7E-29
oki.96.121.t1	P19404	NADH dehydrogenase [ubiquinone] flavoprotein 2, mitochondrial	9E-70
oki.17.2.t1	P50579	Methionine aminopeptidase 2	0.0
oki.21.89.t1	P40616	ADP-ribosylation factor-like protein 1	2E-94
oki.59.44.t1	Q5ZLJ4	Transmembrane protein 70, mitochondrial	2-E7
Low intensity cluster			
oki.411.6.t1	Q6IMM9	Putative gag protein	1E-42
oki.198.25.t1	*		
oki.145.6.t1	P32007	ADP/ATP translocase 3	8E-69
oki.249.34.t1	*		
oki.9.44.t1	*		
oki.17.10.t1	A0A1S3JP20	ras association domain-containing protein 1-like isoform X3	7E-6
oki.189.40.t1	*		
oki.39.60.t1	P50135	Histamine N-methyltransferase	6E-21
oki.8.232.t1	Q8N594	MPN domain-containing protein	2E-69
oki.119.37.t1	Q15773	Myeloid leukemia factor 2	2E-55

*, Not available in both Swiss-Prot and TrEMBL

Supplementary Table 2: Top 10 constitutive genes in *Ptychodera flava*

Gene model	UniProt ID	UniProt definition	E-value
High intensity cluster			
pfl_40v0_9_20150316_1g9384.t1	P60866	40S ribosomal protein S20	2E-69
pfl_40v0_9_20150316_1g27807.t1	P50914	60S ribosomal protein L14	4E-48
pfl_40v0_9_20150316_1g31098.t1	P49207	60S ribosomal protein L34	8E-46
pfl_40v0_9_20150316_1g22239.t1	Q71UM5	40S ribosomal protein S27-like	6E-45
Mid intensity cluster			
pfl_40v0_9_20150316_1g6492.t1	Q9UQ90	Paraplegin	0.0
Low intensity cluster			
pfl_40v0_9_20150316_1g12111.t1	Q9JHI5	Isovaleryl-CoA dehydrogenase,	4E-86
pfl_40v0_9_20150316_1g30650.t1	Q6IPT4	NADH-cytochrome b5 reductase-like	6E-32
pfl_40v0_9_20150316_1g8299.t1	H0YL14	Protein C9orf69	2E-16
pfl_40v0_9_20150316_1g12445.t1	Q9NQM4	Protein PIH1D3	2E-33
pfl_40v0_9_20150316_1g17492.t1	O14638	Ectonucleotide	6E-83
pfl_40v0_9_20150316_1g12421.t1	A0A1S3IF71	flavin reductase (NADPH)-like	2E-19
pfl_40v0_9_20150316_1g16658.t1	Q9Y4W6	AFG3-like protein 2	0.0
pfl_40v0_9_20150316_1g28994.t1	O73885	Heat shock cognate 71 kDa protein	2E-80
pfl_40v0_9_20150316_1g19280.t1	Q9NRP2	COX assembly mitochondrial protein	4E-17
pfl_40v0_9_20150316_1g15793.t1	Q9UK73	Protein fem-1 homolog B	0.0

*, Not available in both Swiss-Prot and TrEMBL

Supplementary Table 3: Top 10 constitutive genes in *Branchiostoma floridae*

Gene model	UniProt ID	UniProt definition	E-value
High intensity cluster			
XP_002603630.1	C3YJ82	Putative uncharacterized protein	0.0
XP_002586419.1	C3ZYF6	Putative uncharacterized protein	4E-90
XP_002586341.1	C3ZYK5	Putative uncharacterized protein	5E-92
XP_002590018.1	C3ZN66	Putative uncharacterized protein	2E-159
Mid intensity cluster			
XP_002608642.1	P10606	Cytochrome c oxidase subunit 5B, mitochondrial	2E-30
XP_002587265.1	Q6KC79	Nipped-B-like protein	0.0
XP_002609523.1	Q9BSF8	BTB/POZ domain-containing protein 10	7E-149
XP_002605586.1	Q8IVB5	LIX1-like protein	1E-152
Low intensity cluster			
XP_002610914.1	Q8NDX5	Polyhomeotic-like protein 3	4E-61
XP_002604143.1	Q13099	Intraflagellar transport protein 88 homolog	0.0
XP_002598845.1	A2AJ76	Hemicentin-2	8E-28
XP_002605025.1	C3YF33	Putative uncharacterized protein	0.0
XP_002611594.1	Q02153	Guanylate cyclase soluble subunit beta-1	2E-106
XP_002589654.1	Q86VW1	Solute carrier family 22 member 16	2E-18
XP_002589890.1	Q01484	Ankyrin-2	1E-9
XP_002588186.1	C3N775	Methionine synthase	8E-32
XP_002591432.1	Q01581	Hydroxymethylglutaryl-CoA synthase, cytoplasmic	0.0
XP_002592193.1	O75596	C-type lectin domain family 3 member A	9E-7

Supplementary Table 4: Top 10 constitutive genes in *Ciona intestinalis*

Gene model	UniProt ID	UniProt definition	E-value
High intensity cluster			
CIYS18885	P84323	60S ribosomal protein L32	3E-60
CIYS7839	P84323	60S ribosomal protein L32	3E-60
CIYS14475	P62900	60S ribosomal protein L31	5E-61
CIYS16389	P62820	Ras-related protein Rab-1A	5E-131
CIYS19127	Q15008	26S proteasome non-ATPase regulatory subunit 6	0.0
CIYS16261	Q07955	Serine/arginine-rich splicing factor 1	2E-24
CIYS8561	P62258	14-3-3 protein epsilon	4E-130
CIYS11634	Q2NKU6	Protein dpy-30 homolog	3E-26
CIYS19551	P67871	Casein kinase II subunit beta	6E-147
CIYS13082	P49427	Ubiquitin-conjugating enzyme E2 R1	1E-101
Mid intensity cluster			
CIYS15148	P43307	Translocon-associated protein subunit alpha	4E-57
CIYS16810	Q00169	Phosphatidylinositol transfer protein alpha isoform	1E-109
CIYS16399	Q6NYC1	Bifunctional arginine demethylase and lysyl-hydroxylase JMJD6	0.0
CIYS11924	O75940	Survival of motor neuron-related-splicing factor 30	6E-58
CIYS15325	Q15904	V-type proton ATPase subunit S1	5E-12
CIYS15778	Q9Y3A3	MOB-like protein phocein	9E-128
CIYS5541	Q15370	Elongin-B	2E-27
CIYS16163	Q9M099	Serine carboxypeptidase 24	4E-113
CIYS6327	Q9BX10	GTP-binding protein 2	0.0
CIYS6342	P61026	Ras-related protein Rab-10	2E-112
Low intensity cluster			
CIYS1567	P23229	Integrin alpha-6	2E-136
CIYS11084	Q15678	Tyrosine-protein phosphatase non-receptor type 14	2E-64
CIYS21396	O08863	Baculoviral IAP repeat-containing protein 3	3E-80
CIYS15626	F6W5V6	Uncharacterized protein	1E-105
CIYS7905	Q9HCD6	Protein TANC2	0.0
CIYS845	O76082	Solute carrier family 22 member 5	2E-52
CIYS770	Q6ZQQ6	WD repeat-containing protein 87	4E-20
CIYS4389	Q9UL62	Short transient receptor potential channel 5	1E-78
CIYS20649	*		
CIYS17556	P52945	Pancreas/duodenum homeobox protein 1	5E-37

*, Not available in both Swiss-Prot and TrEMBL