Head transcriptome profiling of glossiphoniid leech (Helobdella austinensis) reveals clues about proboscis development

Hee-Jin Kwak, Sung-Gwon Lee, Soon Cheol Park, Jung-Hyeuk Kim, David A. Weisblat, Chungoo Park and Sung-Jin Cho

1Department of Biological Sciences and Biotechnology, College of Natural Sciences, Chungbuk National University, Cheongju, Chungbuk 28644, Republic of Korea
2Department of Ecology, Evolution and Behavior, Alexander Silberman Institute of Life Sciences, Faculty of Science, Hebrew University of Jerusalem, Jerusalem 9190401, Israel
3School of Biological Sciences and Technology, Chonnam National University, Gwangju 61186, Republic of Korea
4Department of Life Science, Chung-Ang University, Seoul 06974, Republic of Korea
5Wildlife Disease Response Team, National Institute of Wildlife Disease Control and Prevention, Incheon 22689, Republic of Korea
6Department of Molecular and Cell Biology, University of California, 385 Weill Hall, Berkeley, CA 94720-3200, USA

H-JK, 0000-0003-2494-2564; S-GL, 0000-0002-0928-9950; SCP, 0000-0001-7450-9959; J-HK, 0000-0003-4427-9544; DAW, 0000-0002-9617-4130; CP, 0000-0002-9545-6654; S-JC, 0000-0001-6126-6310

Cephalization refers to the evolutionary trend towards the concentration of neural tissues, sensory organs, mouth and associated structures at the front end of bilaterian animals. Comprehensive studies on gene expression related to the anterior formation in invertebrate models are currently lacking. In this study, we performed de novo transcriptional profiling on a proboscis-bearing leech (Helobdella austinensis) to identify differentially expressed genes (DEGs) in the anterior versus other parts of the body, in particular to find clues as to the development of the proboscis. Between the head and the body, 132 head-specific DEGs were identified, of which we chose 11 to investigate their developmental function during embryogenesis. Analysis of the spatial expression of these genes using in situ hybridization showed that they were characteristically expressed in the anterior region of the developing embryo, including the proboscis. Our results provide information on the genes related to head formation and insights into the function of proboscis-related genes during organogenesis with the potential roles of genes not yet characterized.

1. Introduction

In the development of bilaterian animals, a broadly conserved genetic toolkit underlies the specification of positional identity along the body axis, usually including the establishment of a centralized nerve system, concentration of sensory organs and development of specialized mouth parts at the most anterior region [1–3]. The mouth parts of animals have undergone tremendous specialization and diversification as they have evolved to capture and ingest food according to their diverse life styles and ecosystems [4–7], in large part by modifying the regulation and deployment of the broadly conserved toolkit. Invertebrates especially show a wide variety of mouthpart structures such as tooth, mandible/maxilla and proboscis, depending on diverse food sources, at the end of the head part [7–12]. Among various mouthpart structures in invertebrates, the organ called the proboscis has been known as a specialized apparatus for piercing and holding the host’s gut wall [13], capturing prey...
through injecting toxin [14,15] and fluid sucking [16,17], and the proboscis as a feeding apparatus has a tube for sucking up food such as nectar, body fluid and cellular constituents [18,19]. Fluid sucking of insects is generated in the canalized proboscis by suction pressure produced from the proboscis–cranial sucking pump complex with capillary activity which directs the flow of fluid into the oesophagus [19–22]. On the other hand, the feeding tube of the leech, another representative proboscis-bearing animal, consists only of a muscle complex. The force generated by this complex induces peristalsis for moving food in an aboral direction [11,16].

Cell lineage experiments performed with a well-studied leech model, Helobdella (family Glossiphoniidae), have provided information on the development of the leech proboscis [23,24]. The formation of germ layer-specific tissues in the proboscis is achieved through orchestrated proliferation and interdigitation of clones arising from cells in the early lineage of the mesodermal proteoblast DM‘ (4d cell’ in spinal nomenclature), which mainly contribute to the formation of non-segmental mesoderm that can develop into the digestive tract during organogenesis [23]. At stage 10 (220 h to 245 h after zygotic development), the proboscis develops in an everted. During stage 11, the proboscis gradually inverts to its resting adult position within the foregut region, by which time it consists of a complex array of longitudinal, radial and circular muscles [23,25]. These radical and dynamic developmental changes raise the question of whether essential factors are actively expressed at stage 10 for the formation of the head or specific structures such as the proboscis.

Over the past few years, many studies have been performed to understand the developmental involvement of various transcription factors [26–28] and signalling molecules [29–31], including physiological substances in the anterior region of the leech [32]. Previous studies have suggested that anterior development involves a combination of numerous signalling pathways and transcription factors that and the anterior head is a vast hub of gene expression. However, large-scale studies on expressed genes have not been reported. Towards this end, we used stage 10 leech embryos and conducted RNA sequencing (RNA-seq), a successful method for understanding differential gene expression in specific tissues or under different conditions in a wide variety of animals and plants [33–35]. In this study, we provide detailed information on the stage 10 leech embryo transcriptome and present proboscis-specific genes identified via differentially expressed gene (DEG) analysis between the head and body. Our results will be used as comparative data for the evolutionary relevance of the proboscis and also fill the gap in transcriptional information about the missing stage during leech embryo development.

2. Results and discussion

2.1. Internal structure development of the proboscis during organogenesis

The harmonious proliferation and differentiation of the progeny of lineage-specific blast cells and micromeres give rise to the development of epidermis, ganglia and musculature with the formation of primary mouthparts that can develop into a future digestive tract [23–25,36]. After epiboly, the anterior germinal plate develops into the precursor of the mouthpart (stomodeum), surrounded by primary nerves and muscle fibres (figure 1a–d) [23]. At stage 10, the proboscis gradually protrudes, covered with epithelial tissues which develop into the proboscis sheath. This process involves the active development of a complex of innervated muscle layers in the proboscis structure and oesophagus (figure 1a, e–h). By stage 11, the proboscis invaginates into the mouth pore located in the foregut region with placement of a presumptive circumferential muscle layer and spanning of the radial muscle along with development of longitudinal and oesophageal muscle (figure 1a, i–l). The timing of the morphological differentiation of the proboscis thus suggests that stage 10 might be a crucial step in the development of the internal structure and that proboscis-related genes might be highly expressed during this dynamic event.

2.2. De novo transcriptome assembly of a stage 10 leech embryo

We used RNA-seq to find specific genes in the proboscis compared with the body region at stage 10 (figure 2a and electronic supplementary material, figure S1). More than 941 714 raw reads of the body and 97 667 raw reads of the proboscis were obtained, yielding 84 million and 78 million filtered reads, respectively, with high-quality values (filtered-out ratio greater than 98%; Q50 > 98%; for details, see electronic supplementary material, table S1). Subsequently, we annotated the assembled transcripts using BLASTp (e-value ≤ 1 × 10−50) based on Uniprot/SwissProt and UniRef50 databases to obtain complete sequence coverage. We obtained 13 735 annotated transcripts, including 8003 genes annotated based on the Uniprot/SwissProt database and 5732 genes annotated based on the UniRef50 database (electronic supplementary material, tables S1 and S2). To examine the completeness of this Helobdella austenensis embryo transcriptome, we assessed annotated protein-coding sequences. A total of 82.5% (250/303 genes) and 78.4% (767/978 genes) of the eukaryote and metazoan single-copy orthologues were identified, respectively (figure 2b).

2.3. Identification of differentially expressed genes and gene ontology analysis

Of the 13 735 annotated transcripts, 263 were DEGs in the head and the body (false discovery rate [FDR] < 0.05) (figure 2c). Of these 263 DEGs, 132 and 131 were specifically upregulated in the head and the body, respectively (electronic supplementary material, tables S1 and S3). Characteristically, more than half of these DEGs were uncharacterized genes covered by Uniprot reference clusters [37] (70 of 132), suggesting that anterior development involved a larger proportion of putatively leech-specific genes during anterior formation.

Gene ontology (GO) terms associated with DEGs were used to identify differences within representative classification (molecular function, biological process, protein class) (figure 3). Genes related to binding (GO:0005488), cellular process (GO:0009987) and gene-specific transcription factor (PC00264) accounted for the highest at 43.5% (27/62),
51.6% (32/62) and 22.6% (14/62), respectively. Interestingly, among these genes, homeodomain transcription factors (OTX1, SIX3, SIX6, Nkx1–2, Hmx) were the most common, followed by basic helix–loop–helix transcription factors (SIM1, SIM2, PTF1A, FER3). These results indicate that the expression of common anterior developmental factors might be conserved in the anterior of leech embryos [3,38]. They also suggest that various transcription factors can activate crucial signalling pathways and give rise to organization of the nerve system in the anterior region [3,39–43].

For further investigation, we focused on genes that were both highly and specifically expressed in the head, assuming that highly expressed genes and those most differentially expressed would mainly contribute to head formation, including proboscis development. We selected carefully based on criteria (highly expressed genes annotated, head FPKM > 100; highly expressed genes uncharacterized, head FPKM > 10 and body FPKM <1), and 11 transcripts including three uncharacterized genes were identified (table 1). Among these head-

**Figure 1.** Proboscis formation and internal structure differentiation. The formation and internal structure differentiation of the leech proboscis. All embryos are oriented anterior to the left and dorsal to the top of the panel. White dotted rectangles indicate the magnified region in (e) and (i). White dotted lines indicate the sectional regions. (a) Schematic of organogenesis stages. Between stages 10 and 11, the proboscis invaginates (red arrows) into the mouth pore and locates in the foregut region. (b–d) At stage 9, the oral precursor stomodeum (white dotted circles) develops in the anterior region. (e,f) Lateral view of the stage 11 embryo showing the everted proboscis. (g–h) Sectioned views showing the primordium of innervated muscle fibres (white arrowheads). (i–j) Lateral view of the stage 11 embryo showing ganglia and the invaginated proboscis located in the foregut region. (k–l) Sectioned views showing differentiated muscles and nerve fibres (blue arrowheads). Scale bar, 200 µm in panels b, e, i; 20 µm in panels g and k. as, anterior sucker; AZD, after zygotic development; cc, coelomic cavity; cm, circular muscle; ep, everted proboscis; gp, germinial plate; l, lumen; lm, longitudinal muscle; o, oesophagus; om, oesophageal muscle; p, proboscis; pc, proboscis chamber; ps, proboscis sheath; rm, radial muscle; s, stomodeum; vg, ventral ganglion; vlm, ventral longitudinal muscle; y, yolk.
related DEGs, genes encoding secreted frizzled-related protein (sFRP) homologues were highly and specifically expressed based on both criteria. These findings suggest that a variety of genes are related to the development of the anterior region. Among them, the high and characteristic expression of sFRPs might be correlated with anterior formation [3,38,44].

In addition to transcriptional expression analysis, we also investigated spatial expression patterns of these 11 transcripts to examine their embryological contribution during organogenesis.

2.4. Widespread expression of the hs71l and scot1 homologue

In stage 9, a homologue of the heat shock protein (HSP) hs71l was expressed in the stomodeum and ventral germinal plate (figure 4a). Its expression in the mouthpart persisted. Its expression was also detected in the rostral ganglia, salivary gland and visceral muscle, including the precursor of the posterior sucker (figure 4b). After proboscis invagination, it was expressed ubiquitously in the whole body with discriminate expression in the rostral ganglia, salivary gland and hindgut (figure 4c–c'). HSPs are multifunctional proteins related to protein maturation, protein refolding, protein import and translocation. HSPs facilitate proteolytic degradation of unstable proteins by targeting proteins to lysosomes or proteasomes under normal and stress conditions [45–47]. Although some HSPs have been shown to have embryological functions such as neurodevelopment in previous studies [48,49], comparable expression patterns in lophotrochozoans have not to our knowledge been previously described. Our results for the first time suggest that hs71l plays multifunctional roles in tissues or organs undergoing dynamic structural changes during leech embryogenesis. Succinyl-CoA-3-oxaloacid CoA transferase 1 (SCOT1) is related to ketone metabolism, which catalyses acetoacetate to acetoacyl–coenzyme A (acetoacyl-CoA) in the mitochondria [50,51]. At stage 9, the scot1 homologue was expressed around the stomodeum and in the adhesive zone, by which the embryo attaches to the maternal venter during the period of development after hatching from the fertilization membrane and cocoon and before the rear sucker becomes functional (figure 4d). At stage 10, scot1 was expressed in the proboscis, salivary gland and visceral muscle (figure 4e). Even after proboscis invagination, its overall expression pattern in the proboscis persisted along with the developing head, salivary glands and hindgut (figure 4f and electronic supplementary material, figure S3). We believe that, during embryogenesis, scot1 is expressed in structures and developing regions in which ketone metabolism might be required [52–54]. These results suggest that the stage of organogenesis might be a period in which tissues and organs undergo dynamic changes with organ-specific metabolic process of ketone.

2.5. Genes expressed in specific tissue layers in the proboscis

Previous lineage analyses [23,24,36] of leech proboscis development and our morphostructural analysis revealed that the
development of the proboscis is marked by orderly expansion of radial muscles and differentiation of longitudinal muscles. We hypothesized that characteristic core signalling genes might contribute to this process.

Calmodulin (CaM) and calmodulin-like (CML) proteins known as major regulators of Ca\(^2^+\)-dependent signalling are key regulators of various mechanisms such as cell proliferation, motility and cell cycle progression [55–57]. In addition, CaM is an upstream regulator of calmodulin kinase (CaMK), which activates CaMK signalling to induce the development of the central nervous system [58,59]. At stage 9, transcripts of cml23 homologue were expressed in the furrow of the stomodeum and the adhesion site for attachment to the maternal venter (figure 4g). The adhesion site develops into the future epithelium of the anterior sucker. Transcripts of the cml23 homologue were continuously expressed in the region, including nuclei of radial muscle and ventral ganglia (figure 4h–i). Our results indicate that the cml23 homologue is related to the cellular process of the developing proboscis and central nerve system during organogenesis.

sFRPs are key modulators of the Wnt signalling system, which determines the anteroposterior axis according to the distribution of Wnt ligands and their frizzled receptors; Wnt signalling is essential for bilaterian head formation [3,44]. In this study, sfrp orthologues showed an anterior-specific expression pattern in embryos during organogenesis (figures 4i–f and 5a–c'). At stage 9, these homologues were expressed around the stomodeum. At stages 10 and 11, these two paralogues showed different expression patterns. Transcripts of sfrp1/2/5a were expressed in the radial muscle precursor in the proboscis but weakly expressed in the posterior body (figure 4k–l). However, sfrp1/2/5b was expressed in the posteriormost proboscis sheath and the anterior intrinsic muscle [60] (figure 5b–c). Based on these spatial evidences, sfrp orthologues might play a role as conserved anterior morphogens with proboscis differentiation factors [3,44,61].

Zinc finger CCHC-type containing protein 24 (ZCH24) is one of the zinc finger domain proteins [62] known to be related to the development of somitogenic mesoderm in vertebrates [63], although it is largely unknown in invertebrates. In developing leech embryos, the zch24 homologue was only expressed inside the stomodeum at stage 9 (figure 5d) and in radial muscle precursors from the anterior to the posterior of the proboscis by stage 11 (figure 5e–f). These restricted expressions of zch24 transcripts in proboscis indicate that zch24 is a possible proboscis-specific gene in H. austeniensis as well as related to the radial muscle differentiation of the developing proboscis.

2.6. Nervous system-specific expression of the Dmrt93B homologue

At stage 9, Dmrt93B was expressed around and inside the stomodeum and in the ectodermal region of the germinal plate (figure 5g). At stage 10, its transcripts were mainly expressed in the rostral segment of ganglia and the proboscis including ventral ganglia (figure 5h). After proboscis invagination, Dmrt93B transcripts were expressed in the rostral segments, epidermis, ventral ganglion, nerve canal in the proboscis cavity [16] and the posterior sucker (figure 5i,f). Dmrt genes are well known to be mainly sex dimorphic factors [64,65]. They have been actively studied in various fields such as neuronal differentiation and brain development, in addition to differentiation of animal sex determinants and reproductive organs [66,67]. Furthermore, in Panarthropoda, Dmrt93B is expressed in the nervous system including mouth development [68]. Taken together, these results indicate that this highly conserved gene also plays a role as a transcription factor for the development of the mouth and nervous system during leech embryogenesis.

2.7. Developmental relevance of the defence-related gene

Defence protein is characterized by a reeler domain. It is an innate immune substance that protects against pathogens...
Table 1. Genes of interest selected based on the three criteria. Asterisks indicate previously studied genes (for detailed phylogeny, see electronic supplementary material, figure S2).

| gene                      | expression level (FPKM) | q-value  | source                                 |
|---------------------------|-------------------------|----------|----------------------------------------|
| highly expressed genes annotated (head FPKM >100) |                        |          |                                        |
| SFRP2 (sFRP1/2/5a*)       | 6.36                    | 124.08   | 6.79 × 10⁻⁶                            |
| CML23                     | 5.26                    | 101.45   | 6.83 × 10⁻⁶                            |
| HS71L                     | 29.7                    | 401.82   | 7.88 × 10⁻⁵                            |
| SCOT1                     | 15.97                   | 116.86   | 4.61 × 10⁻¹                            |
| highly expressed genes uncharacterized (head FPKM >100) |                        |          |                                        |
| UniRef50_T1FEH8           | 1.05                    | 119.36   | 8.26 × 10⁻¹¹                          |
| UniRef50_A7TSG4           | 16.24                   | 182.23   | 3.04 × 10⁻⁴                           |
| UniRef50_T1F354           | 51.5                    | 246.87   | 4.90 × 10⁻²                           |
| differentially expressed genes annotated (head FPKM >10 and body <1) |                      |          |                                        |
| ZCH24                     | 0.03                    | 28.61    | 8.44 × 10⁻¹⁴                          |
| DFP3                      | 0.21                    | 19.42    | 3.92 × 10⁻⁹                           |
| DMTA2 (DMRT93B*)          | 0.41                    | 22.09    | 1.17 × 10⁻⁸                           |
| SFRP1 (sFRP1/2/5b*)       | 0.68                    | 21.69    | 5.22 × 10⁻⁸                           |

[69,70]. In this study, dfp3 homologue was not expressed at stage 9 (figure S5). However, it began to be expressed in the epithelium region of the proboscis at stage 10 (figure 5k). Its expression in the epithelium surrounding the proboscis was maintained even after proboscis invagination (figure 5l). We suggest that this expression pattern is related to the development process of leech embryos. Helobdella embryos are surrounded by a vitelline membrane until stage 9. After hatching, they attach to the maternal venter and take off from it [25]. Based on the developmental evidence of spatial expression, the innate immune system in the leech might begin its protective function as the leech breaks through the vitelline membrane and becomes exposed to the environment.

2.8. Developmental contribution of uncharacterized genes during organogenesis

At stage 9, UniRef50_T1FEH8 transcripts were expressed in the periphery of the stomodeum along with overall expression in the germinal band. Then, transcripts were expressed in the epithelium of the everted proboscis and rostral ganglia at stage 10. After proboscis invagination, UniRef50_T1FEH8 transcripts were expressed in the proboscis chamber, proboscis sheath and ventral ganglion (figure 6c–f). From these results, we found that UniRef50_T1FEH8 is related to development of the epithelium of the proboscis and ganglia formation.

UniRef50_A7TSG4 transcripts were expressed in the foregut precursor region and around the stomodeum at stage 9 and in the proboscis epithelium, oesophageal funnel and overall gut boundaries at stage 10. At stage 11, the expression at stage 10 was maintained, in addition to the expression in the proboscis chamber, proboscis sheath, oesophagus and the developing hindgut (figure 6d–f). Therefore, we found that UniRef50_A7TSG4 is related to the epithelium of the proboscis along with overall gut development.

UniRef50_T1F354 transcripts were expressed in the stomodeum, coelomic cavity and visceral muscle precursors at stage 9 (figure 6g). At stage 10, these transcripts were expressed in the radial muscle precursor in the anterior proboscis, fibres of the proboscis sheath, body and visceral muscles (figure 6h). The expression was maintained in the same region at stage 11 (figure 6i). Therefore, UniRef50_T1F354 is related to the overall mesodermal tissue development during organogenesis. Their intense expression in the proboscis with major organs of these uncharacterized or maybe leech-specific genes indicates that these unique genes are not only important for embryogenesis but also are specifically related to the proboscis development.

3. Conclusion

The feeding organs of invertebrates have undergone extensive diversification and specialization over the course of evolution. We performed a comprehensive study to expand our knowledge about proboscis evolution and to discover genes related to the development of this specialized organ. Through transcriptome profiling, DEG analysis and spatial analysis, we identified specific genes related to proboscis development including highly conserved anterior formation factors. This not only provides novel information that has not been previously elucidated but also suggests that leeches also undergo anterior formation by conserved morphogens along with diverse transcription factors and that the developing proboscis is a vast hub of signal molecules. Furthermore, the presence of numerous uncharacterized genes in the anterior formation suggests a possible involvement of specific genes in proboscis development, including head formation in leeches, raising questions about the evolution of mouth formation in other proboscis-bearing animals. Taken together, our comprehensive survey will be used for comparative
analysis of feeding organ diversity and proboscis formation in invertebrates.

4. Material and methods

4.1. Animals

Adult *H. austinensis* specimens were bred in the Laboratory of Cellular and Development Biology (Department of Biology, Chungbuk National University, Republic of Korea). They were bred in bowls containing artificial freshwater placed in a biological oxygen demand (BOD) incubator at 22°C. Embryos were placed in Petri dishes with lids containing *Helobdella triserialis* (HTR) saline medium (4.8 mM NaCl, 1.2 mM KCl, 2 mM MgCl$_2$, 8 mM CaCl$_2$ and 1 mM maleic acid) according to the protocol for handling leech embryos.

4.2. RNA isolation and RNA-seq library preparation

Embryos were cultured in clean HTR medium until stage 10 (220–245 h after zygotic development). To obtain a sufficient amount of RNA, about 400 embryos were used in the experiment. The head region was dissected using insect pins...
(Shiga, no. 2) under a Leica ZOOM 2000 stereomicroscope (Leica, Wetzlar, Germany). Next, the proboscis region was transferred to a 1.7 ml tube containing 200 µl of TRIzol reagent (Invitrogen, Carlsbad, CA, USA). RNA was isolated from each sample using TRIzol reagent according to the manufacturer’s instructions. The purity and integrity of the total RNA isolated from dissected embryo samples were examined using a NanoDrop 2000C spectrophotometer (Thermo Scientific, Waltham, MA, USA) and a Bioanalyzer 2100 (Agilent Technologies, Palo Alto, CA, USA). Then, total RNA concentration was calculated using Quant-IT RiboGreen (Invitrogen, R11490). To assess the integrity of the total RNA, samples were run on a TapeStation RNA screentape (Agilent, 5067–5576). Only high-quality RNA (RNA integrity number greater than 7.0) was used to construct an RNA library. A library was independently prepared with 1 µg of total RNA for each sample using Illumina TruSeq Stranded Total RNA Sample Prep Kit v. 2 (Illumina, Inc., San Diego, CA, USA). Ribosomal RNA in total RNA was depleted using a Ribo-Zero kit (Epicentre Biotechnologies, Madison, WI, USA). After the rRNA depletion, the remaining RNA was purified, fragmented and primed for cDNA synthesis. Cleaved RNA fragments were copied into first-strand cDNAs using reverse transcriptase and random hexamers followed by second-strand cDNA synthesis using DNA polymerase I, RNase H and dUTP. Next, cDNA fragments underwent end repair, addition of a single A base

Figure 5. Spatial expression of DEGs (annotated) during organogenesis. Whole-mount in situ hybridization results of differentially expressed genes (annotated). All embryos are oriented anterior to the left and dorsal to the top of the panel. The rightmost column shows a cross-sectional view of a stage 11 embryo. White arrows indicate expression in the developing mouth part. Red arrows point to nervous expression of candidate genes. Yellow arrow indicates expression in the intrinsic muscle. White dotted lines indicate the sectioned region seen in the rightmost column. (a) By stage 10, sfrp1/2/5b is expressed in and around the stomodeum. (b) At stage 10, sfrp1/2/5b is expressed in the posterior epithelial tissues of the everted proboscis. (c, c’) After proboscis invagination, sfrp1/2/5b remains expressed in the proboscis sheath and the intrinsic muscle. (d) At stage 9, zch24 is expressed inside the stomodeum. (e) After proboscis eversion, zch24 is expressed in the radial muscle precursor. (f, f’) After proboscis invagination, zch24 is expressed in the radial muscle layer (white arrowheads). (g) Dmrt93B is expressed in the ectodermal lineage in the germinal plate and the stomodeum. (h) Dmrt93B is expressed in the proboscis, ganglia and the visceral muscle region. (i) At stage 11, Dmrt93B expression remains in the rostral ganglia, proboscis and posterior sucker. (j) In a cross-sectional view, Dmrt93B is detected in the epithelial tissue of the body, ventral ganglia (red arrowhead) and proboscis innervation (white arrowheads). (j) No dfp3 expression is visible at stage 9. (k) At stage 10, dfp3 transcripts start to be detected in the everted proboscis. (l, l’) By stage 11, dfp3 is expressed in the longitudinal muscle layer (white arrowhead). Scale bar, 200 µm in panels a–c; 100 µm in panel c. AZD, after zygotic development; sfrp1/2/5b, secreted frizzled protein 1/2/5b; zch24, zinc finger CCHC domain-containing protein 24; Dmrt93B, doublesex-and male-abnormal-3-related transcription factor 93B; dfp3, defence protein 3.
and adapter ligation. Products were then purified and enriched with polymerase chain reaction (PCR) to create the final cDNA library. All libraries were quantified using quantitative PCR according to the qPCR Quantification Protocol Guide (KAPA Library Quantification kits for Illumina Sequencing platforms) and qualified using a TapeStation D1000 ScreenTape (Agilent Technologies, Waldbronn, Germany). Finally, library samples were sequenced using an Illumina HiSeq 2000 system (Illumina, Inc., San Diego, CA, USA) to generate paired-end reads.

4.3. De novo assembly, differential expression and gene ontology analyses

Transcriptome assembly was performed using Trinity v. 2.1.1 [71]. After assembly, protein-coding sequences (CDSs) within assembled transcripts were predicted using the Uniprot/Swiss-Prot database (http://www.uniprot.org), BLASTP v. 2.2.31+ and TransDecoder v. 3.0.0 (http://transdecoder.sourceforge.net) included in the Trinity software. Redundant CDSs with identity greater than 99% were clustered and the longest CDS was left using CD-HIT v. 4.6.5 [72]. Finally, a set of non-redundant coding sequences (NRCDS) was generated. To quantify each transcript’s expression level, RNA-seq reads for each sample were aligned against the transcriptome sequences using Bowtie v. 2.2.6 [73] and abundances of reads according to transcripts were estimated with RSEM v. 1.2.26 [74]. For annotated NRCDS, homology search was performed against the Uniprot/SwissProt database using BLASTP with an E-value cut-off of $10^{-10}$ and the best blast hit. For one-to-one protein and gene mapping, the one with the highest expression level was selected if multiple NRCDS were annotated in a single gene. To minimize the loss of unannotated NRCDS, it was reannotated with the UniRef50 database, an automatically annotated database and clustered sets of sequences from UniProt Knowledgebase and selected UniParc records [37]. Differentially expressed genes between the head and body were identified using the edgeR program with an FDR cut-off of less than 0.05. GO terms of differentially expressed genes were identified using PANTHER (Protein Analysis Through Evolutionary Relationships, http://pantherdb.org/) [75].
4. Gene cloning and probe synthesis

Specific primers (electronic supplementary material, table S4) for genes of interest were designed based on transcriptome sequences. Genes were amplified from cDNA of H. Austinessis at stage 10 using a Takara ExTaq kit (Takara Bio Inc., Kusatsu, Japan) according to the manufacturer’s instructions with the following cycling conditions: pre-denaturation at 94°C for 5 min, denaturation at 94°C for 30 s, variable annealing temperature for 30 s, variable extension time at 72°C and post-extension at 72°C for 5 min. These amplified fragments were cloned into pGEM T vector (Promega, Madison, WI, USA). RNA probes labelled with digoxigenin were constructed using a MEGAscript kit (Ambion, Austin, TX, USA) and a DIG RNA Labeling Mix (Roche, Basel, Switzerland) according to each manufacturer’s instruction.

4.5. Whole mount in situ hybridization

Fluorescence in situ hybridization (FISH) was performed as previously described [29,32] as follows: embryos were treated with protease from Streptomyces griseus (Sigma-Aldrich, St. Louis, MO, USA) in 1x phosphate-buffered saline (PBS), rinsed three times with glycine dissolved in PBS at room temperature (RT) for 5 min, fixed with 4% paraformaldehyde at RT for 40 min and rinsed with PBS (1× PBS + 0.1% Tween-20) three times. Next, prehybridization was performed at 64.7°C for one day in hybridization buffer (50% formamide, 5× saline-sodium citrate (SSC), 1x Denhardt’s solution, 0.1% 3-[(3-cholamidopropyl)dimethylammonio]-1-propanesulfonate (CHAPS), 100 mg ml⁻¹ of heparin, 0.1% Tween 20 and 100 mg ml⁻¹ of transfer RNA (tRNA)]. After incubation, the prehybridized buffer was replaced with fresh hybridization buffer containing 2 ng ml⁻¹ of corresponding probe. Embryos were then hybridized at 64.7°C for 2 days. After washing embryos with PBS, they were preincubated with maleic acid buffer (100 mM maleic acid, 150 mM NaCl, pH 7.5) at RT for 15 min, blocked with 1% blocking reagent for nucleic acid hybridization and detection (Roche, Basel, Switzerland) at RT for 2 h and incubated with anti-DIG/POD (1:1000 dilution) in 1% blocking reagent at 4°C for 16 h. After incubation, embryos were rinsed twice with TNT buffer (0.1 M Tris–HCl (pH 7.5), 0.15 M NaCl, 0.1% Tween 20) at RT and washed with amplification solution provided by the NEN Tris solution, 0.1% Tween 20 and labelled with DAPI in PBT (1:100) at RT for 20 min. Fluorescence-stained embryos were imaged using a LSM 710 confocal microscope (Carl Zeiss, Oberkochen, BW, Germany). The obtained images were edited using ZEN software (Carl Zeiss). Fluorescence-labelled embryos were then dehydrated in ethanol series (70%, 90% and 100% diluted in 1 × PBS) and propylene oxide, followed by infiltration with plastic embedding solution (PolyBed). Embryos were cut with a microtome blade (Leica 818; Leica, Wetzlar, Germany) under an Olympus SZ-STS microscope (Olympus, Tokyo, Japan). These sections were then mounted with Fluoromount-G (SouthernBiotech, Birmingham, AL, USA). Stained embryos and slide samples were imaged using a LEICA DM6 B with a LEICA DFC450 C camera (Leica, Wetzlar, Germany). Obtained images were edited using a Las X software (Leica), and prepared as figure plates using Adobe Illustrator CS6 (Adobe, San Jose, CA, USA).

Data accessibility. Sequences generated in this study are deposited in GenBank (electronic supplementary material, table S4). Raw sequencing data are deposited in the NCBI Sequence Read Archive (SRA) database with accession numbers of SRR15013283 and SRR15013284 under BioProject PRJNA742851.

The data are provided in electronic supplementary material [76].

Authors’ contributions. H.-J.K.: conceptualization, data curation, formal analysis, funding acquisition, investigation, methodology, project administration, resources, validation, visualization, writing—original draft, writing—review and editing; S.-G.L.: data curation, methodology, resources, software, validation, visualization, writing—original draft; writing—review and editing; S.C.P.: conceptualization, data curation, formal analysis, methodology, resources, software, validation, visualization, writing—original draft, writing—review and editing; S.R.: conceptualization, data curation, funding acquisition, project administration, supervision, writing—original draft, writing—review and editing; J.-H.: formal analysis, methodology, writing—original draft, writing—review and editing; D.A.W.: conceptualization, investigation, writing—original draft, writing—review and editing; C.P.: data curation, methodology, software, supervision, validation, visualization, writing—original draft, writing—review and editing; S.-J.C.: conceptualization, data curation, funding acquisition, investigation, methodology, project administration, resources, supervision, validation, visualization, writing—original draft, writing—review and editing.

All authors gave final approval for publication and agreed to be held accountable for the work performed therein.

Competing interests. We declare we have no competing interests.

Funding. This research was supported by the Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (2021R1I1A101041998). This research was supported by the Chung-Ang University Research Grants in 2019. This work was supported by a grant from the National Institute of Biological Resources (NIBR), funded by the Ministry of Environment (MOE) of the Republic of Korea (NIBR202130202).
