Genome and transcriptome of freshwater leech Whitmania pigra reveal key genes related to morphogenesis, signal pathways and neurogenesis during embryonic development

Jiali Liu[1], Jinxin Liu[1], Mingyue Li[2], Lisi Zhou[1], Weijun Kong[1], Hailin Zhang[3], Panpan Jin[1], Fuhua Lu[1], Linchun Shi[1], and Gufa Lin[3]

[1]Chinese Academy of Medical Sciences & Peking Union Medical College Institute of Medicinal Plant Development
[2]Shanghai Jiao Tong University School of Medicine
[3]Tongji University

July 6, 2022

Abstract

Whitmania pigra Whitman (phylum Annelida) is a widely distributed freshwater leech in East Asia, with an annual consumption of nearly 1,000 tons dry bodies for curing blood stasis syndrome. W. Pigra has be seen as a representative model organism of neurodevelopmental studies. Here, we sequenced a Chinese individual by sin-gle-molecule real-time (SMRT) long-read sequencing, and generated a de novo as-sembly of 178.87 Mb with contig N50 2.0 Mb. In addition, we obtained a total of 239.64 Gb transcriptome data of eight representative developmental phases of embryos (from blastocyst stage to maturity). Totally, 21482 genes were annotated, among these, 3114 genes were differentially expressed with phase-specific expression pattern, and mainly in the middle and late development (G, H, I, J phase). Using a comprehensive transcriptome dataset, we demonstrated that 57, 49 and 77 DEGs were respectively related to morphogenesis, signal pathways and neurogenesis. 49 DEGs related to signal pathways included 30 wnt genes, 14 notch genes, and 5 hedgehog genes. In par-ticular, we found a cluster consisting of 7 genes related to signal pathways as well as synapses, which were essential for regulating embryonic development. To some ex-tent, our results are helpful to reveal the whole picture of development mechanism from the perspective of transcriptome and also provide new clues for organogenesis and neurodevelopmental studies of Annelida species.

Article type: Resource Article

Molecular Ecology Resources

Resources Article-Permanent Genetic Resources

Title:

Genome and transcriptome of freshwater leech *Whitmania pigra*reveal key genes related to morphogenesis, signal pathways and neurogenesis during embryonic development

Running Title: Embryonic transcriptome of *Whitmania pigra*

Authors:

Jiali Liu*

Institute of Medicinal Plant Development

Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing 100193, China

Jinxin Liu*

Institute of Medicinal Plant Development

Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing 100193, China

Mingyue Li

Renji Hospital

Shanghai Jiao Tong University School of Medicine, Shanghai 200127, China

Lisi Zhou

Institute of Medicinal Plant Development

Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing 100193, China

Weijun Kong

Institute of Medicinal Plant Development

Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing 100193, China

Hailin Zhang

Key Laboratory of Spine and Spinal Cord Injury Repair and Regeneration of Ministry of Education, Orthopaedic Department of Tongji Hospital, School of Life Sciences and Technology

Tongji University, Shanghai 200065, China

Panpan Jin

Institute of Medicinal Plant Development

Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing 100193, China

Fuhua Lu

Institute of Medicinal Plant Development

Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing 100193, China

Linchun Shi++

Institute of Medicinal Plant Development

Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing 100193, China

Gufa Lin++

Key Laboratory of Spine and Spinal Cord Injury Repair and Regeneration of Ministry of Education, Orthopaedic Department of Tongji Hospital, School of Life Sciences and Technology

Tongji University, Shanghai 200065, China

++ Corresponding authors

* These authors contributed equally to this article.

Corresponding authors:

Gufa Lin,*lingufa@tongji.edu.cn*

Linchun Shi, linchun_shi@163.com, 010-57833206

Abstract: *Whitmania pigra* Whitman (phylum Annelida) is a widely distributed freshwater leech in East Asia, with an annual consumption of nearly 1,000 tons dry bodies for curing blood stasis syndrome. *W. Pigra* has be seen as a representative model organism of neurodevelopmental studies. Here, we sequenced a Chinese individual by single-molecule real-time (SMRT) long-read sequencing, and generated a de novo assembly of 178.87 Mb with contig N50 2.0 Mb. In addition, we obtained a total of 239.64 Gb transcriptome data of eight representative developmental phases of embryos (from blastocyst stage to maturity). Totally, 21482 genes were annotated, among these, 3114 genes were differentially expressed with phase-specific expression pattern, and mainly in the middle and late development (G, H, I, J phase). Using a comprehensive transcriptome dataset, we demonstrated that 57, 49 and 77 DEGs were respectively related to morphogenesis, signal pathways and neurogenesis. 49 DEGs related to signal pathways included 30 *wnt* genes, 14 *notch* genes, and 5 *hedgehog* genes. In particular, we found a cluster consisting of 7 genes related to signal pathways as well as synapses, which were essential for regulating embryonic development. To some extent, our results are helpful to reveal the whole picture of development mechanism from the perspective of transcriptome and also provide new clues for organogenesis and neurodevelopmental studies of Annelida species.

Keywords: *Whitmania pigra* Whitman; genome; transcriptional dynamics; morphogenesis; signal pathway; neurogenesis

1. Introduction

Whitmania pigra Whitman (abbreviated as *W. pigra*) is a species of freshwater leech in the family of Haemopidae, which belongs to the class Hirudinea. As a protozoan species of medicinal leech specified in the Chinese Pharmacopoeia, the dried whole body of the leech *W. pigra* possesses many therapeutic properties for the treatment of the hematencephalon and other thrombosis-related diseases(Dong et al. 2016), such as hypertension(Srivastava and Sharma 2010), cerebral infarction(Li et al. 2021), ischemic stroke(Xu et al. 2016) and atherosclerosis(Hu et al. 2020). More than 100 kinds of Chinese patented preparations, containing single leeches or its compatibility form with other herbs, were officially approved to the medical market. Actually, the most common species added in Chinese patent medicine and circulated in the medicinal market is *W. pigra* . Current findings have shown that nearly thirty effective proteins and peptides have been isolated and clearly characterized from leeches(Iwama et al. 2022), and less in *W. pigra* (Khan et al. 2019b). It is accessible to make tree-based orthology determination and structure-based alignment with publicly available transcriptomes for rapidly exploration of bioactive ingredients(Kvist et al. 2020; Müller et al. 2022). However, the lack of available genomic and transcriptomic data of *W. pigra* has limited on the efficient exploration of anticoagulants and effective utilization of leech resources.

Adult leech (such as *Hirudo medicinalis*) has been used as a research model on neurobiology for its regeneration capability in the central nervous system, widely used to solve electrophysiological characteristics, behavior, treatment and development. The involvement of microglial cells in the neuronal regeneration was first shown in leech(Sieger and Peri 2013). *W. pigra* also has experimentally accessible nervous system with ordered structure and a relatively small number of neurons. Its central nervous system consists of 34 ganglia, including 21 ganglia lined up in the middle of the whole body, and 6 head ganglia plus 7 tail ganglia at both ends of the body. Each ganglion has less than 200 pairs of neurons(Nicholls and Baylor 1968). Identifiable by electrical characteristics and function, each neuron plays a specific and unique role(Nicholls and Baylor 1968). The differentiation of these nerve cells started at stage 6a in leech embryos(Kuo et al. 2020), but the genes involved in the leech neural differentiation are not clear. Thus, elucidating the molecular mechanism of embryonic development has been a research hotspot in the scientific community.

Dramatically different body plans and cellular processes of development have been evolved in nature, but early cell division patterns are remarkable conserved among taxa. Therefore, it is very necessary and wise to select model organisms representing different taxa for research. Glossiphoniid leeches of the genus *Helobdella* have been used for developmental studies since the 1970s. In 2001, with the clearer observation of the development process, Huang and Weisblat(Weisblat and Huang 2001) standardized the naming of blast cells and summarized the developmental process into 11 stages. In situ hybridization facilitated to display the temporal and spatial expression of genes, has been showed the unique expression patterns of 13 *wnt* genes

on the germinal plate in *Helobdella robusta* (Cho et al. 2010). However, this approach relies on specific fluorescent probes and therefore only a selected small number of genes be characterized, the whole-scale dynamic transcriptome changes during embryonic development are not yet available. Compared with leech *Helobdella robusta* , *W. pigra* 's embryos are more resistant to stress, relatively big and easily to obtain. In May, a *W. pigra* weighing 20g can produce 5.9 cocoons in average and each cocoon can hatch 37.3 leech seedlings(Li et al. 2020). The embryos grow by the nutrition supply of yolk and cocoon fluid, and develop into complete larvae in about one month. Therefore, *W. pigra* is a potential model for developmental biology.

To facilitate understanding of the molecular mechanism underlying developmental events in *W. pigra* , we integrated genomic data of adult leech and transcriptomic data of eight representative developmental stages of embryos (from blastocyst to maturity). Utilizing a comprehensive transcriptome dataset, we screened the key transcripts and pathways in the development process and verified sequencing data by cloning and qPCR technology. Our results provide a better understanding of *W. pigra* development at the molecular level and lay a foundation for further functional research.

2. Materials and Methods

2.1. Sample Collection and Observation

Adult leech *W. pigra* was collected from leech breeding base in Weishan Lake, Shandong Province. In order to obtain fresh cocoons in time, pregnant leeches were reared in an artificial ceramic breeding tube, which comprises a first shell, a second shell and a joint part (Figure 1). The first shell and the second shell cooperate with each other through the joint part to form at least one accommodation cavity for leech to breed. Wet towel covering and regular water spraying are necessary to keep humidity at about 60%. The breeding tube is convenient to open and examine leeches and cocoons in time, so as to take out cocoons from tubes at regular intervals. For embryo collection, a small opening was cut at the top of a clean cocoon, and the embryos were pipetted out together with cocoon fluid for observation and sample collection. Embryos were observed under a stereomicroscope MZ75 (Leica) or M165FC (Leica), and images were taken with digital cameras. The images were prepared with software Picolay for focus stacking.

2.2. Genome Sequencing and RNA-seq

Genomic DNA was extracted from muscle tissue dissected from the body of a single adult *W. pigra* leech using a tissue genomic DNA Extraction Kit (TIANGEN, China). Samples of defined developmental stages were frozen with liquid nitrogen to extract RNA using RNAprep pure Tissue Kit (TIANGEN, China) in accordance with the standard protocol. The quality and quantity of DNA and RNA were tested by Nanodrop 2000 spectrophotometer (Thermo, USA) and 1% agarose gel electrophoresis. Reverse transcription Kit (Promega, USA) was used to synthesize the first strand of cDNA.

A 260 bp paired-end shotgun libraries were prepared for sequencing on an Illumina NovaSeq 4000 platform to generate an initial survey based on a 19-*K-mer* distribution. The software Jellyfish (v2.1.4)(Marçais and Kingsford 2011) was used for counting k-mers, and the software GenomeScope (v1.0)(Vurture et al. 2017) was used for estimating the genome size. Through whole-genome sequencing, a DNA library was constructed using the standard protocol of PacBio. RNA-seq libraries were generated using Vazyme kit and sequenced using NovaSeq 6000 platform (Illumina, USA). Sequencing data have been accessible in the NCBI Sequence Read Archive.

2.3. Genome Assembly, Annotation and RNA-seq Data Analysis

After filtering the low-quality and short clips of the raw fastq data from PacBio sequencing, the filtered data was corrected using Canu(Koren et al. 2017) and assembly with WTDBG(https://github.com/ruanjue/wtdbg). Error correction were performed on the assembly result based on the Illumina sequencing data using Pilon(Worley et al. 2014). We use CEGMA v2.5(Parra et al. 2007) to evaluate the integrity of leech genome assembly through sequence similarity (identity $> 70\%$) alignment. Founded on an integrated strategy including ab initio, homology-based and UniGene prediction, the gene content of the genome of *W. pigra* were analyzed and integrated with EVM(Haas et al. 2008)

software. Using RepeatMasker(Chen 2004) and GeneWise(Birney et al. 2004) software to predict repetitive sequence and pseudogenes, respectively. Based on Rfam(Griffiths-Jones et al. 2005) database, microRNA and rRNA were identified using Blastn for genome-wide alignment, and tRNA were predicted by tRNAscan-SE (Lowe and Eddy 1997). Gene annotation was performed through blasting against GO, COG, KEGG, nr, nt, TrEMBL, KOG and Pfam databases.

Clean data of RNA-Seq was obtained by filtering off adaptor sequences, low-quality reads, duplicated sequences and poly-N from raw reads and then aligned with our reference genome of *W. pigra* using HISAT2(Kim et al. 2015). The aligned reads were assembled and evaluated using StringTie(Pertea et al. 2015) software package. On the basis of the criteria of fold change > 2 and false discovery rate (FDR) < 0.05, differentially expressed genes were chosen. Function information of genes were annotated in GO, COG, KEGG, nr, SwissProt, KOG, eggNOG and Pfam databases.

2.4. Verification Through Cloning and Real-Time Quantitative PCR

The target fragments to be verified was obtained by PCR. After ligating the target fragments with vectors and then transforming to *E. coli* , we picked up the positive colonies to extract the plasmids for sequencing using Sanger's sequencing techniques. The sequencing result can make up the error rate of high-throughput sequencing and act as a role of verification. Quantitative PCR was performed using TB Green kit (Takara, Japan) on CFX96 Real-Time System (BIO-RAD, the U.S.A). The cycling parameters were set up as follows: step 1: 95degC for 30 s; step 2: 95degC for 5 sec, 55degC for 30 sec and 72degC for 60 sec, performed with 40 cycles; step 3: (melt curve) 65-95degC, increase 0.5degC every time and keep it for 5 sec. The PCR reaction system was 25 μL: 12.5 μL of 2× TB Green mix, 2 μL of the forward and reverse primers (2.5 μM), 2 μL of DNA template, made up the volume to 25 μl with sterilized ultrapure water. The specific primers used above were synthesized by SANGON biotech company. The *α-τυβυλιν* gene was chosen as an internal reference gene for normalizing the relative expression content of genes using $2^{-\Delta\Delta^{\tilde{}}T}$ method. All primer sequences are listed in Supplementary Table S6-7.

3. Results

3.1. Embryonic Development of Whitmania pigra

The whole embryonic development process of *W. pigra* from egg laying to complete larvae was be observed. Embryo firstly underwent cleavage stage (A to C in Figure 2). Helical division was happened for two cell cycles. As the teloplasm (yolk-deficient domains of cytoplasm) becomes unevenly divided, the axis of spiral division gradually leaned towards the animal pole. After the cocoon formation for about four days later, the embryo reached the blastocyst stage (D in Figure 2). The blastopore could be observed, and it will develop into the further mouth. After multiple mitoses, blastoderm became thinned gradually and the syncytial yolk cell (SYC) was tightly surrounded in the middle. The embryo underwent severe deformation and the blastopore opened and closed to exchange materials with cocoon fluid (E in Figure 2). The epiboly of germinal band marks gastrulation, after which coelomic cavities arose in a front-to-back progression and the embryo developed a clear dorsal-ventral axis and an anterior-posterior axis (G in Figure 2). Figure 1H to 1K is the stage of organogenesis and refinement of leech. The boundary of crop ceca and intestine became clear distinguishable, as the result of refinement of structure. During these stages, the posterior sucker differentiated into cup shape and extended to the body width. Eleven pairs of eye spots were gradually dark. The nervous system was developed, and the ganglions at the midline of the back can be seen. Figure L shows a leech larva that resemble the adults, with well-defined segments, muscle and organs. Compared to phase K, the most obvious difference of L was the deepening of yellowish longitudinal lines with black spots on the back. This moment, the yolk in the embryo was exhausted and leech larvae can crawl out of cocoon for their first feeding.

3.2. Genome Sequencing and Assembly

The *W. pigra* genome was estimated to be 219.40 Mb in size with a medium level of repetition (33.27%) and heterozygosity (0.61%) based on 19-k -mer analysis. A total of 80.0× coverage of single-molecule sequences

of the PacBio Sequel platform was used for assembly, and 350.0× Illumina data was generated for sequencing error corrections and gap filling. The size of the final assembly was 178.87 Mb, with 483 contigs and a contig N50 of ˜ 2.00 Mb (Supplementary Table S1). The completeness of the genome of *W. pigra* were estimated, and 445 genes (97.16%) and 245 genes (98.79%) were identified when we blast against CEGMA v2.5 database. 19128 high-quality protein-coding genes and 456 pseudogenes due to frameshifts or premature stop codons were predicted. The average length of protein-coding genes was 4673.06 bp, and the mean sequence length of exons was 205.57 bp. Furthermore, noncoding RNA genes were predicted, yielding 738 tRNA, 51 rRNA and 8 miRNA. A total of 69.07 Mb of repetitive elements in the *W. pigra* genome were annotated. Results shows that 90.09% of the genes can be annotated into nr, nt, GO, KEGG, pfam, TrEMBL, KOG, COG databases.

3.3. RNA Sequencing, Mapping and New Transcripts Exploiting

Eight representative stages of the *W. pigra* embryos, covering the starting point of cell differentiation to individual maturation, corresponding to D-L in Figure 2 (see Table 1) were sequenced. After quality control, we obtained a total of 239.64 Gb clean data from 24 samples, with each sample reaching 6.14Gb. The Q30 base was no less than 91.18 % and the mean GC content was 44.37%(Table S2). Results from StringTie showed that the average comparison efficiency between reads and our reference genome was 81.15%, and 84.9% reads were distributed in exon region, indicating that most of reads were located on mature mRNA. 21482 genes were identified in our transcriptome, 19765 were present as known or predicted transcripts in major transcriptome databases, such as Pfam, eggNOG, nr, COG, GO, KEGG. And 2354 new transcripts were not discovered in the genome , among them 77 genes were directly related to development according to GO database. Apparently, “development” “morphogenesis” “cell” and “neurogenesis” etc. were high-frequency words in GO annotation information (Figure 3).

3.4. Differential Expression Analysis of Whitmania pigra Embryonic Transcriptome

To gain insights of the dynamic gene expression profiles in different stages of embryonic development of *W. pigra* , we performed pairwise comparison with fold change [?] 2 and FDR [?] 0.05 as cutoff values for differentially expressed genes (DEGs). In total, we identified 15161 DEGs in the comparison of 28 groups, of which 12671 were up-regulated and 2490 down-regulated. After removing duplications, we obtained 3114 DEGs, which accounted for approximately 13% of the leech transcriptome. The number of DEGs was the least in E-F period (early gastrula stage), J phase contained the largest number of DEGs, followed by L phase (Figure 4A). It suggested that the differentiation begins at the blastocyst stage and reaches the peak at the time of organ differentiation. The heatmap of DEGs in eight stages shows obvious phase-specific expression pattern both vertically and horizontally. D-F phases represent gastrula stage, cluster (a) genes were uniquely highly expressed in D-F phases; genes in G-J phases share similar expression patterns, participating in organogenesis and refinement; L stage represents developed larva, genes in cluster (c) were only highly expressed in L phase (Figure 4B).

Functional annotation of DEGs was performed, E vs J group containing most of DEGs. According to the annotation results of GO database, the top5 terms in biological process were negative regulation of peptidase activity, DNA duplex unwinding, organonitrogen compound catabolic process, inductive cell migration and cellular protein metabolic process. It shows that the energy required for embryonic development mainly comes from the decomposition of organic nitrogen. DNA unwinding represents the rapid progress of cell division. The process of cell communication and migration promotes tissue and organ differentiation. Under cellular component category, the first three terms were the gap junction, MCM complex and actin cytoskeleton. Gap junction, a special membrane structure between cells, are critically important in many biological activities, including development and differentiation (Maeda and Tsukihara 2011). MCM2-7 complex is essential to DNA duplication, assembled on chromatin in G1 phase of mitosis and separated during S phase (Quan et al. 2015). Molecular function was mainly related to extracellular region, serine-type endopeptidase inhibitor activity and catalytic activity. EggNOG enrichment analyses revealed that, the function of J phase was mainly related to signal transduction mechanisms, posttranslational modification, protein turnover, chaperones and cytoskeleton. KEGG annotation showed “global and overview maps”, “translation”, “signal transduction”, transport and catabolism” and “endocrine system” led the rankings. These results indicate

that the development of *W. pigra* is closely associated with metabolism, cell structure and cell proliferation and differentiation (Figure 5).

To identify species specificity, we searched all of the assembled unigenes against the NR database (Figure S1). Glossiphoniid leech*Helobdella robusta* share the most of homologous protein with*W. pigra* (56.68%), followed by Annelida *Capitella teleta*(11.35%) and *Ceratitis capitata* (4.15%).

3.5. Genes Involved In Morphogenesis During Embryonic Development

The above pairwise comparison highlights the dynamic gene expression pattern in *W. pigra*embryonic development. From the perspective of the GO terms related to "morphogenesis" at the eight developmental stages of *W. pigra* , we identified 57 significantly DEGs (Supplemental Table S3), and selected 14 representative genes for further analysis (Figure 6). These genes were expressed at very low level across early developmental periods while highly expressed in G-J stages. From the annotation results of the GO database, gene functions include dorsal/ventral pattern formation, eye development, dendrite morphogenesis, dorsal closure, cell morphogenesis involved in neuron differentiation and so on.

The formation of dorsal-ventral axis and anteroposterior axis is the most obvious morphological feature during *W. pigra* embryonic development. It has been shown that *EVM0012139* , encoding Fascin (80.4% similar with *Helobdella robusta*), is ubiquitous in early embryonic development, contributing to dorsal/ventral pattern formation and the growth of muscle, dendrite and other cells. It was highly expressed in stage G and peaked in stage L. Similarly, gene*EVM0013532* and *EVM0001864* was annotated to participate in the formation of anterior/posterior pattern. *EVM0001864* has low identity with characterized protein, the similarity with evolutionarily conserved C2H2-type zinc finger protein is about 30% and it was annotated as transcriptional activator in SwissProt database. Therefore, we suggested it plays a role in the development and differentiation of tissues/organs as a transcription factor. It maintained uniform and stable expression during eight stages of development. Matrix metalloproteinase translated by *EVM0005181*has proteolytic activity and effect on signal factors, so its high expression is closely related to embryonic development (Verslegers et al, 2013). The occurrence of salivary gland is of great value for the medicinal use of leech. *EVM0003169* was annotated as a matrix metalloproteinase 1 isoform X3 promoting the morphogenesis of salivary gland. *EVM0014112* was highly expressed in G stage, mainly involved in the development of dorsal closure through the protein interaction mediated by FERM domain. *EVM0001642* had to do with the differentiation of photoreceptor cell, expressed in G-J stage. This indicates that the photosensitive ability of *W. pigra* forms before maturity. The dynamic expression profile of *EVM0001787*indicated that the cell morphogenesis involved in neuron differentiation mainly occurred during G to L stages. In addition, gene*EVM0007438* and *EVM0014293* were related to immunoglobulin domains. These two genes were both highly expressed in the late period of leech, especially *EVM0014293* , with the highest FPKM value reached 133.90, indicating their important roles in immune system.

3.6. Signal Pathways in W. pigra Embryonic Development

Among all the DEGs, 49 genes were related to Wnt pathway, Hedgehog pathway and Notch pathway (Supplemental Table S4). We selected fifteen genes for further detailed analysis. This included 10 *wnt* genes, 3 *notch* genes and 2 *hedgehog* genes (Figure 7). Instead of having a high expression in specific periods like morphogenesis genes, genes related to signal transduction were highly expressed in decentralized period. This suggested that these genes undertake different developmental tasks throughout the whole developmental process.

Wnt and *hedgehog* genes were all expressed at a low level in E-F stage, followed by a significant growth in stage G, indicating Wnt cooperated with Hedgehog participated in blastoderm differentiation and organogenesis. Here, we explicitly annotate six Wnt family members, including Wnt1, Wnt2, Wnt4, Wnt5, Wnt16 and Wnt16B (Wnt16 and Wnt16B are shown in Supplemental Table S5). From the perspective of gene expression, Wnt1, Wnt2 and Wnt16 were widely involved in the development of G-J stage, and still play a great role after leech maturity. Wnt4 and Wnt5 mainly regulate embryonic development, and the expression level decreased after maturation.

In addition, a 21.7-Kb-long sequence was found related to signal pathway (Figure 8A), including seven genes (*EVM0018524* ,*EVM0015683* , *EVM0007677* , *EVM0018061* ,*EVM0002178* , *EVM0014014* and *EVM0004456*). The gene interval between these genes was no more than 3979 bp. Totally five kinase domains and five transferase domains were predicted (Figure 8B), playing a role in a multitude of cellular processes. Therefore, these genes can be seen as a gene cluster which potentially perform the same function. Furthermore, gene *EVM0018524* , *EVM0015683* and*EVM0004456* also mapped to five kinds of synaptic pathways, including glutamatergic, cholinergic, serotonergic, GABAergic and dopaminergic synapse. To some extent, this result proved that signal transduction was accompanied with neurogenesis.

Gene *EVM0001864* and *EVM0009838* are related to Hedgehog signaling. *EVM0001864* has conserved C2H2-type double zinc finger domain, playing a role in transmitting Hedgehog signals and morphogenesis. *EVM0009838* was highly expressed in G stage. Its encoding protein includes four kind of Immunoglobulin domain (IgV, IgC1, IgC2, and IgI) and fibronectin type III domain.

The expression of three notch genes (*EVM0016927, EVM0005841* and*EVM0007457*) was higher in the middle stage and lower in the early and late stage. *EVM0016927* was involved in dorso-ventral axis formation pathway (ko04320), playing an essential role in pattern formation. *EVM0005841* encodes Calcium-binding EGF domain, which usually imparts specific functions to proteins in coagulation cascade, here was annotated as a processed neurogenic locus Notch protein.*EVM0007457* negatively regulated the Notch signaling pathway, with similar change trend of gene expression horizontal as *EVM0005841* . It suggested the activation and inhibition of notch pathway are regulated simultaneously to regulate the normal growth of embryos.

3.7. Neural Development In W. pigra

Our annotation results of all DEGs showed that 77 genes were assigned to "neurogenesis", taking part in the development of the central nervous system, peripheral nervous system or neural structure of leeches (Supplemental Table S5). The annotation from KOG database showed that gene function can be classified into the following categories: transcription, ion channel, cytoskeleton, and function unknown. Here, 10 representative genes and their expression data throughout eight developmental stages are listed in Figure 9.

EVM0003050 was expressed at its highest level in L stage, its coding protein Homeobox influence transcription process to play a role in central and peripheral nervous system development. *EVM0002561*encode calcium-activated potassium channel protein, determining neuron projection morphogenesis. Similar functions include calcium ion binding (*EVM0000129*), neurotransmitter-gated ion-channel ligand binding domain (*EVM0005725*), calcium-activated potassium channel (*EVM0016702*) and so on. Tyrosine phosphatase genes usually participate in the post-translational modification for neurogenesis and motor neuron axon guidance. *EVM0007698* encodes protein-tyrosine phosphatase, it was highly expressed in D-F stage, suggesting its role in the early-stage regulation of synaptic transmission.*EVM0006869* serves as a neurogenic locus protein with a low expression level. It contributes to the development of central nervous system and peripheral nervous system. Significantly, eight genes encode Innexin (such as *EVM0006839*), functioning in gap junction and synapse. *EVM0007438* is related to Immunoglobulin domain, its encoding protein fasciclin-2 is a neural cell adhesion molecule, widely involved in the development of the nervous system. *Whitmania_pigra_newGene_2185* is related to muscleblind-like protein, contributing to the development of peripheral nervous system. Protein singed and glass were coded by *EVM0012139* and*Whitmania_pigra_newGene_5855* respectively, playing an important role in neurogenesis and central nervous system development.

3.8. Gene Verification By Gene Cloning and qPCR

To validate our RNA sequence data, we successfully cloned twenty-nine genes into plasmids and sequenced, including 15 genes related to signal pathways, 10 genes related to signal transduction mechanisms, one gene related to dorsal-ventral pattern and three genes related to myogenic, TGF-beta propeptide and Jak-STAT represent respectively. In addition, we randomly verified three genes (related to morphogenesis and neurogenesis) by qPCR experiment. The results showed that the expression pattern determined by qPCR was basically consistent with the results of RNA sequencing (Figure S2). Taken together, the transcriptome analyses were very reliable.

4. Discussion

Whitmania pigra is a classical traditional Chinese medicine for promoting blood circulation and removing blood clot. It is also a model animal on neurobiology representing annelids and a potential research model of embryology. However, the dynamic transcriptome during the process of development has not been studied yet. Therefore, it's necessary to study the embryonic development of *W. pigra* . We divided the late development process of *W. pigra* into 8 stages according to the morphological characteristics. Our staging criteria might have discrepancies with standard 11 stages. But we have paid more attention to the gross morphological changes, rather than precise cell lineage, which is so far difficult to pursue. The staging system we developed is adequate for the purpose of sample collection, as verified by Pearson's correlation coefficient (r) analysis(Ahlgren et al. 2003) of our RNA sequencing data. Results showed that r^2value was close to 1.

The combination of genome and transcriptome facilitates the analysis of gene clusters with specific functions. We have compiled a complete set of analysis procedures to screen out genes with a distance of less than 3000 on the same KEGG pathway. It is speculated that these genes cooperate to complete the same function, which is of great reference significance to the study of gene function. *EVM0018524* and*EVM0015683* encode cAMP-dependent protein kinase catalytic subunit beta (PRKACB), active PRKACB regulated cell proliferation(Faherty et al. 2007), differentiation(Azeloglu et al. 2014), cell cycle, migration(O'Connor and Mercurio 2001) and other processes.*EVM0007677* (PKC1) has been implicated to regulate peptide neurosecretion in *Caenorhabditis elegans* , however there is no experimental evidence of a developmental relevance. *EVM0018061*(GRK-3) encode G protein-coupled receptor kinase. Kin-1 (*EVM0002178* , *EVM0004456)* is essential for larval development. Polypeptides resulting from alternative splicing from kin-1 has been found to have substantial differences in embryonic and adult nematodes(Bowen et al. 2006). *EVM0014014 (SGK-1)* is the crucial kinase for the control of development, stress response, and longevity(Hertweck et al. 2004). Only one gene cluster related to signal pathways was shown here, and other gene clusters will be further explored later.

Using a comprehensive transcriptome dataset, we obtained the dynamic transcriptome changes that occurred from blastocyst stage to maturity. Genes related to dorsal/ventral pattern and anterior/posterior formation were highly expressed in G-J stage, while genes related to salivary gland differentiation and photoreceptor differentiation were dynamic expressed in H stage. Morphological characteristics are realized through induction mediated by signal transduction and organogenesis regulated by complex signal networks(Rishikaysh et al. 2014). Studies have shown that Wnt signaling cascade with Notch, and Hedgehog signaling to regulate the balance of embryonic cell(Katoh 2008). Our results are in consistence with this conclusion. Most genes were involved in the regulation of more than one signal pathways, such as *EVM000120* and*EVM0013532* , indicating that these signal pathways play a coordinated role in embryonic development. Notably, our results suggest that wnt gene *EVM0003169* plays an important role in the regulation of salivary gland development. Studies have been conducted to compare the transcriptome levels of salivary glands before and after leech blood sucking, and identified the top 20 KEGG pathway include cAMP, MAPK and PI3K-Akt signal pathway except Wnt pathway(Khan et al. 2019a). Therefore, we speculate that Wnt signal pathway is of great significance for the development of salivary glands but not for production of active substances in saliva.

In neural development of leech, Homeobox were the most abundant genes (11 genes), and there are studies showing that Homeobox gene families exhibit a regional distribution near the anterior posterior body axis in leeches, and the main function is about morphogenesis(Aisemberg et al. 1993; Meriaux et al. 2011). It is tempting to speculate, based on our results, that the Homeobox genes are involved in the development of the*W. pigra* nervous system. The phosphorylation process of proteins is the final step to transmit information between nerve cells. Through a review of the literature, it was found that phosphorylases and kinases can also affect the growth and development of nerve cells.*EVM0007698* encoding tyrosine phosphatase protein was highly expressed in D-F stages, likely participating in the construction of nerve cells. Meanwhile, molecular studies on the development of the leech nervous system also found many substances containing Ig domain(Hattori et al. 2007), which further supporting the necessity of this domain. Secondly, studies

in mice have shown that Fascin protein is expressed in sciatic nerve and hippocampal nerve, and the actin binding domain of Fascin protein can induce axonal extension(Nagel et al. 2012).*EVM0007438* showed similar functional annotation, mainly promote the development of synaptic part of central nerve during G-I stages.

Acknowledgements: Thanks to Linchun Shi and Gufa Lin for helpful guidance and suggestions. This work was supported by Beijing Municipal Natural Science Foundation (7202136), and Major Research, Development Projects of Sichuan Science and Technology Plan Projects (2019YFS0017), and CAMS Innovation Fund for Medical Sciences (CIFMS)(2021-1-I2M-029).

References

Ahlgren P, Jarneving B, Rousseau R. 2003. Requirements for a cocitation similarity measure, with special reference to Pearson's correlation coefficient. *Journal of the American Society for Information Science and Technology* **54** : 550-560.

Aisemberg GO, Wysocka-Diller J, Wong VY, Macagno ER. 1993. Antennapedia-Class homeobox genes define diverse neuronal sets in the embryonic CNS of the leech. *Journal of Neurobiology* **24** : 1423-1432.

Azeloglu EU, Hardy SV, Eungdamrong NJ, Chen Y, Jayaraman G, Chuang PY, Fang W, Xiong H, Neves SR, Jain MR. 2014. Interconnected network motifs control podocyte morphology and kidney function. *Science signaling* **7** : ra12-ra12.

Birney E, Clamp M, Durbin R. 2004. GeneWise and genomewise. *Genome research* **14** : 988-995.

Bowen LC, Bicknell AV, Tabish M, Clegg RA, Rees HH, Fisher MJ. 2006. Expression of multiple isoforms of the cAMP-dependent protein kinase (PK-A) catalytic subunit in the nematode, Caenorhabditis elegans.*Cellular signalling* **18** : 2230-2237.

Carini M, Shi L, Chamberlain TW, Liu M, Valenti G, Melle-Franco M, Paolucci F, Khlobystov AN, Pichler T, Mateo-Alonso A. 2019. Wall- and Hybridisation-Selective Synthesis of Nitrogen-Doped Double-Walled Carbon Nanotubes. *Angew Chem Int Ed Engl* **58** : 10276-10280.

Chen N. 2004. Using Repeat Masker to identify repetitive elements in genomic sequences. *Current protocols in bioinformatics***5** : 4.10.11-14.10.14.

Cho S-J, Valles Y, Giani Jr VC, Seaver EC, Weisblat DA. 2010. Evolutionary dynamics of the wnt gene family: a lophotrochozoan perspective. *MOLECULAR BIOLOGY AND EVOLUTION* **27** : 1645-1658.

Dong H, Ren J-X, Wang J-J, Ding L-S, Zhao J-J, Liu S-Y, Gao H-M. 2016. Chinese medicinal leech: ethnopharmacology, phytochemistry, and pharmacological activities. *Evidence-based complementary alternative medicine* **2016** .

Faherty S, Fitzgerald A, Keohan M, Quinlan L. 2007. Self-renewal and differentiation of mouse embryonic stem cells as measured by Oct4 expression: the role of the cAMP/PKA pathway. *In Vitro Cellular Developmental Biology-Animal* **43** : 37-47.

Griffiths-Jones S, Moxon S, Marshall M, Khanna A, Eddy SR, Bateman A. 2005. Rfam: annotating noncoding RNAs in complete genomes.*Nucleic acids research* **33** : D121-D124.

Haas BJ, Salzberg SL, Zhu W, Pertea M, Allen JE, Orvis J, White O, Buell CR, Wortman JR. 2008. Automated eukaryotic gene structure annotation using EVidenceModeler and the Program to Assemble Spliced Alignments.*Genome biology* **9** : R7-R7.

Hattori D, Demir E, Kim HW, Viragh E, Zipursky SL, Dickson BJ. 2007. Dscam diversity is essential for neuronal wiring and self-recognition.*Nature* **449** : 223-227.

Hertweck M, Gobel C, Baumeister R. 2004. C. elegans SGK-1 is the critical component in the Akt/PKB kinase complex to control stress response and life span. *Developmental cell* **6** : 577-588.

Hu B, Xu L, Li Y, Bai X, Xing M, Cao Q, Liang H, Song S, Ji A. 2020. A peptide inhibitor of macrophage migration in atherosclerosis purified from the leech Whitmania pigra. *Journal of ethnopharmacology***254** : 112723.

Iwama RE, Tessler M, Kvist S. 2022. Leech anticoagulants are ancestral and likely to be multifunctional. *Zoological Journal of the Linnean Society* .

Katoh M. 2008. WNT signaling in stem cell biology and regenerative medicine. *Current drug targets* **9** : 565-570.

Khan MS, Guan D, Kvist S, Ma L, Xie J, Xu S. 2019a. Transcriptomics and differential gene expression in Whitmania pigra (Annelida: Clitellata: Hirudinida: Hirudinidae): Contrasting feeding and fasting modes.*Ecology evolution* **9** : 4706-4719.

Khan MS, Guan DL, Kvist S, Ma LB, Xie JY, Xu SQ. 2019b. Transcriptomics and differential gene expression in Whitmania pigra (Annelida: Clitellata: Hirudinida: Hirudinidae): Contrasting feeding and fasting modes. *Ecology and evolution* **9** : 4706-4719.

Kim D, Langmead B, Salzberg SL. 2015. HISAT: a fast spliced aligner with low memory requirements. *Nature methods* **12** : 357-360.

Koren S, Walenz BP, Berlin K, Miller JR, Bergman NH, Phillippy AM. 2017. Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. *Genome research* **27** : 722-736.

Kuo D-H, De-Miguel FF, Heath-Heckman EAC, Szczupak L, Todd K, Weisblat DA, Winchell C. 2020. A tale of two leeches: Toward the understanding of the evolution and development of behavioral neural circuits.*Evolution Development* **22** : 471-493.

Kvist S, Manzano-Marin A, de Carle D, Trontelj P, Siddall ME. 2020. Draft genome of the European medicinal leech Hirudo medicinalis (Annelida, Clitellata, Hirudiniformes) with emphasis on anticoagulants.*Scientific reports* **10** : 1-11.

Li M, Liu J, Wang M, Huang Q, Shi L, Ma L. 2020. Population investigation of *Whitmania pigra* in Weishan Lake area and quality evaluation of its medicinal materials after cocoon production.*Modern Chinese Medicine* **22** : 849-855.

Li T, Sun Q, Wang B, Liu W, Wang S, Li M, Gao Y. 2021. Efficacy and safety of Maixuekang Capsule in treatment of acute ischemic stroke: A systematic evaluation and GRADE profile. *Chinese Traditional and Herbal Drugs* **52** : 483-494.

Lowe TM, Eddy SR. 1997. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic acids research***25** : 955-964.

Maeda S, Tsukihara T. 2011. Structure of the gap junction channel and its implications for its biological functions. *Cellular Molecular Life Sciences* **68** : 1115-1129.

Marcais G, Kingsford C. 2011. A fast, lock-free approach for efficient parallel counting of occurrences of k-mers. *Bioinformatics***27** : 764-770.

Meriaux C, Arafah K, Tasiemski A, Wisztorski M, Bruand J, Boidin-Wichlacz C, Desmons A, Debois D, Laprevote O, Brunelle A et al. 2011. Multiple changes in peptide and lipid expression associated with regeneration in the nervous system of the medicinal leech. *PLoS One* **6** : e18359.

Muller C, Wang Z, Hamann M, Sponholz D, Hildebrandt JP. 2022. Life without blood: Molecular and functional analysis of hirudins and hirudin-like factors of the Asian non-hematophagous leech Whitmania pigra. *Journal of Thrombosis Haemostasis* .

Nagel J, Delandre C, Zhang Y, Forstner F, Moore AW, Tavosanis G. 2012. Fascin controls neuronal class-specific dendrite arbor morphology.*Development* **139** : 2999-3009.

Nicholls Ja, Baylor D. 1968. Specific modalities and receptive fields of sensory neurons in CNS of the leech. *Journal of Neurophysiology***31** : 740-756.

O'Connor KL, Mercurio AM. 2001. Protein kinase A regulates Rac and is required for the growth factor-stimulated migration of carcinoma cells.*Journal of Biological Chemistry* **276** : 47895-47900.

Parra G, Bradnam K, Korf I. 2007. CEGMA: a pipeline to accurately annotate core genes in eukaryotic genomes. *Bioinformatics***23** : 1061-1067.

Pertea M, Pertea GM, Antonescu CM, Chang T-C, Mendell JT, Salzberg SL. 2015. StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nature biotechnology* **33** : 290-295.

Quan Y, Xia Y, Liu L, Cui J, Li Z, Cao Q, Chen XS, Campbell JL, Lou H. 2015. Cell-cycle-regulated interaction between Mcm10 and double hexameric Mcm2-7 is required for helicase splitting and activation during S phase. *Cell reports* **13** : 2576-2586.

Rishikaysh P, Dev K, Diaz D, Qureshi WMS, Filip S, Mokry J. 2014. Signaling involved in hair follicle morphogenesis and development.*International journal of molecular sciences* **15** : 1647-1670.

Sieger D, Peri F. 2013. Animal models for studying microglia: the first, the popular, and the new. *Glia* **61** : 3-9.

Srivastava A, Sharma R. 2010. A brief review on applications of leech therapy. *Arch Appl Sci Res* **2** : 271-274.

Vurture GW, Sedlazeck FJ, Nattestad M, Underwood CJ, Fang H, Gurtowski J, Schatz MC. 2017. GenomeScope: fast reference-free genome profiling from short reads. *Bioinformatics* **33** : 2202-2204.

Weisblat DA, Huang FZ. 2001. An overview of glossiphoniid leech development. *Canadian Journal of Zoology* **79** : 218-232.

Worley KC, English AC, Richards S, Ross-Ibarra J, Han Y, Hughes D, Deiros DR, Vee V, Wang M, Boerwinkle E. 2014. Improving genomes using long reads and Pbjelly 2. in *International Plant and Animal Genome Conference Xxii San Diego, CA* , p. 15.

Xu H, Shi Y, Zhang Y, Jia Q, Li D, Zhang Y, Liu F, Yang H. 2016. Identification of key active constituents of Buchang Naoxintong capsules with therapeutic effects against ischemic stroke by using an integrative pharmacology-based approach. *Molecular BioSystems* **12** : 233-245.

Supplementary Materials: Table S1. Sequencing and assembly statistics of the draft genome of *Whitmania pigra* ; Table S2. RNA-Seq results of eight stages; Table S3. 57 genes related to morphogenesis; Table S4. 49 genes related to signal pathways; Table S5. 77 genes related to neurogenesis; Table S6. Primers used in real-time PCR; Table S7. Primers used for gene cloning and sequencing results. Figure S1. Species distribution analysis in NR database. Figure S2. Verification of three unigenes (*EVM0001864* , *EVM0014293* ,*EVM0006153*) using qPCR (N=3).

Data Accessibility: The high-throughput sequencing datasets presented in this study can be found in the National Center for Biotechnology Information (NCBI) SRA online repository. The accession number of genome is ASM2161333v1. The accession number of the BioProject is PRJNA777652. The accession numbers of the BioSample specimens are SAMN22870527-SAMN22870550. And the SRA accession numbers for the above 24 BioSample specimens are SRR16775025-SRR16775048, respectively.

Author Contributions: Data curation, Hailin Zhang and Panpan Jin; Formal analysis, Mingyue Li; Funding acquisition, Linchun Shi; Methodology, Mingyue Li and Lisi Zhou; Software, Linchun Shi; Supervision, Jinxin Liu, Gufa Lin and Linchun Shi; Visualization, Weijun Kong; Writing – original draft, Jiali Liu and Fuhua Lu; Writing – review & editing, Gufa Lin and Linchun Shi.

Conflicts of Interest: The authors declare no conflicts of interest.

Tables and Figures:

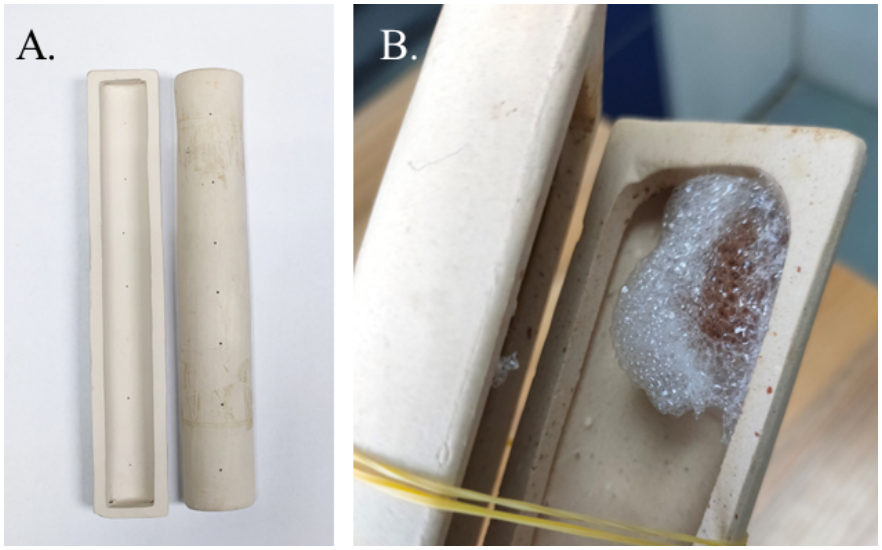


Figure 1. Ceramic breeding tube for leech. *(A)* The panorama of breeding tube, including a first shell, a second shell and a rubber band used to fix shells, there are spiracles at tube wall. *(B)* A leech produced a cocoon in the breeding tube for less than five hours, and the honeycomb-like structure on the surface of the cocoon was still not completely dry yet and the interior is pink.

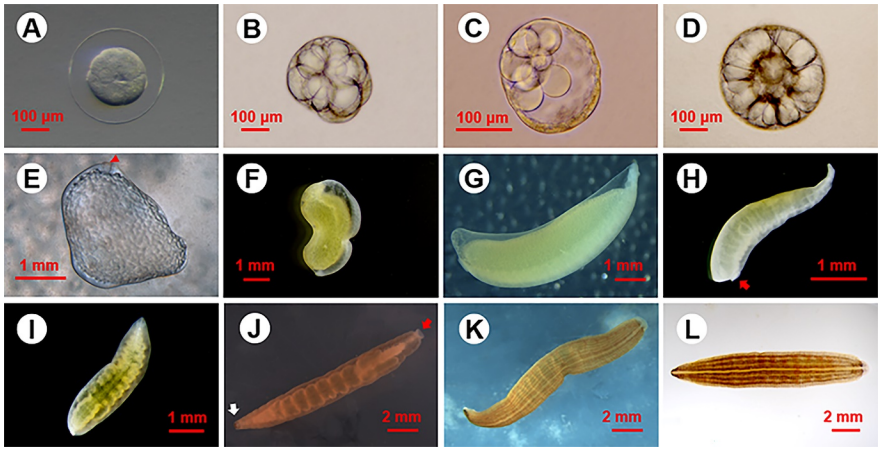


Figure 2. The process of embryonic development of *W. pigra* . *(A-C)*Cleavage stages: embryo undergoes spiral division along with teloplasm unequal distribution; *(D)* Blastocyst stage: SYC wrapped by blastoderm; *(E-F)* Gastrula stage: the embryo has a strong deformation movement, the shape of the blastopore (red triangle) changes; *(G-K)* Organogenesis and refinement. The development events in this stage include the differentiation and refinement of crop ceca and intestine, caudal sucker (red arrow) and eye spots (white arrow); *(L)* Juvenile, yolk depletion in embryos, the color of the five longitudinal body surface lines was darker. Larvae leave the cocoon and begin to feed for the first time.

Table 1: Selected eight representative phases for sequencing (N=3).

Stage	Duplication 1	Duplication 2	Duplication 3	Representative developmental events
D	T1	T2	T18	Blastocyst stage
E	T3	T19	T23	Severe deformation and epiboly
F	T4	T20	T24	Coelomic cavities formation
G	T5	T6	T7	Distinct dorsoventral pattern, anterior-posterior axis
H	T8	T9	T21	The organogenesis and refinement of crops, intestine, eyes and post
I	T10	T11	T12	
J	T13	T14	T15	
L	T16	T17	T22	Body surface lines

Figure 3. A word cloud of 77 potential new genes related to development according to GO database. The minimum frequency was set to 2, and 83 words matched the filtering criteria. The color of words indicates the frequency, black means high and red means low.

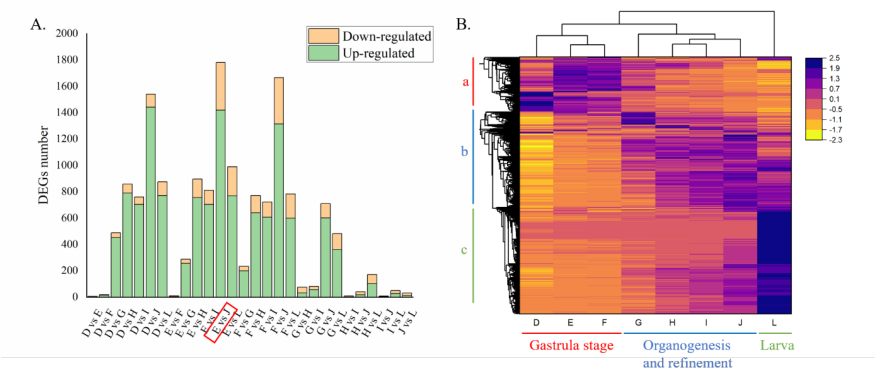


Figure 4. *(A)* Barplot displaying the number of differentially expressed genes in paired comparison. *(B)* Heat maps of DEGs shows that gene expression patterns can be divided into three branches both in vertical and horizontal.

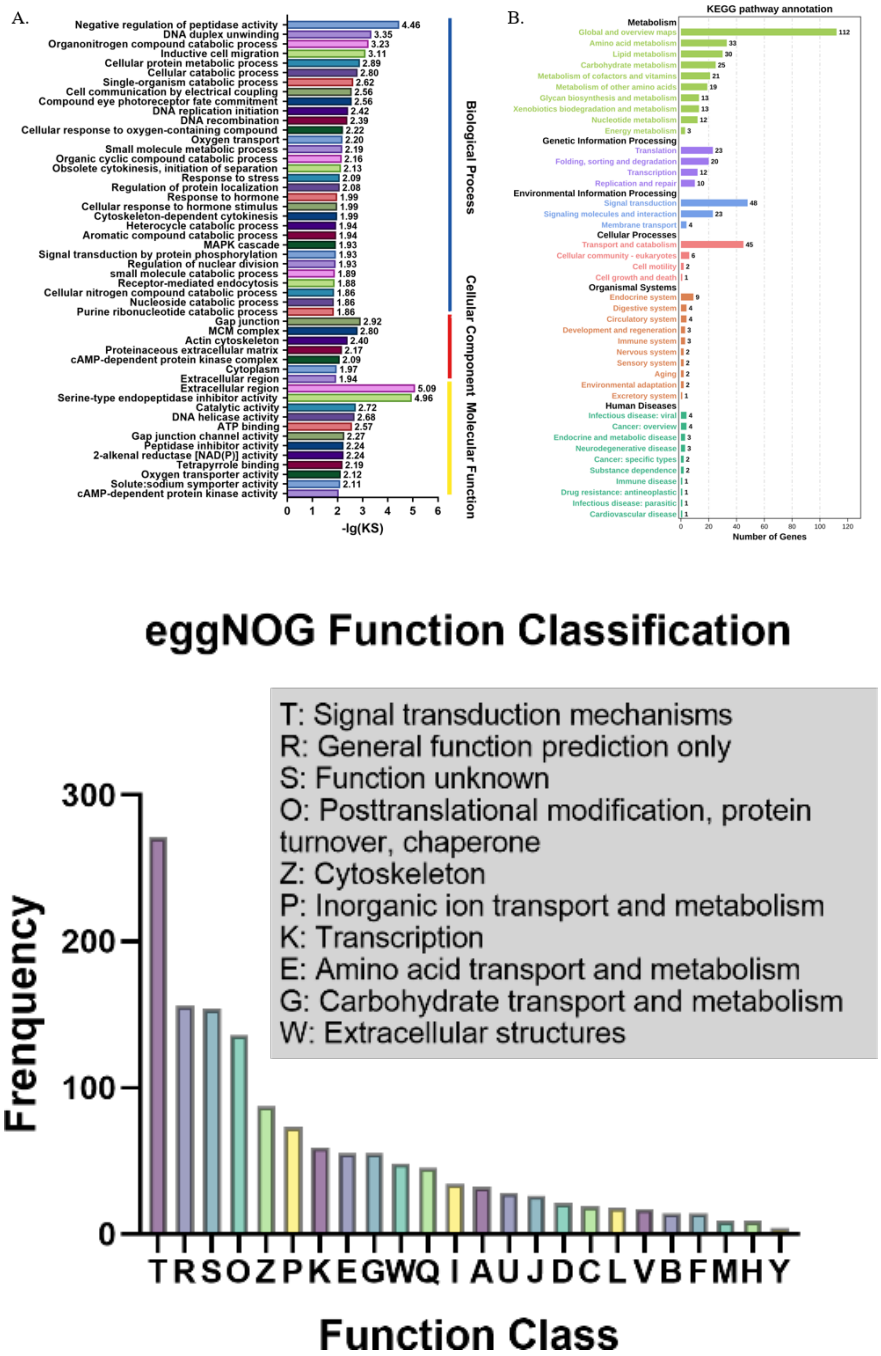


Figure 5. Differential expression analysis of E vs J stage. *(A)* Functional annotation in GO databases. The results show that DEGs were annotated in biological process, cellular component and molecular function. *(B)* Analysis of expression in KEGG database. *(C)* Functional annotation in eggNOG databases. (*Q: Secondary metabolites biosynthesis, transport and catabolism; I: Lipid transport and metabolism; A: RNA processing and modification. U: Intracellular trafficking, secretion, and vesicular transport. J: Translation, ribosomal structure and biogenesis. D: Cell cycle control, cell division, chromosome partitioning; C: Energy production*

and conversion. L: Replication, recombination and repair; V: Defense mechanisms; B: Chromatin structure and dynamics; F: Nucleotide transport and metabolism; M: Cell wall/membrane/envelope biogenesis; H: Coenzyme transport and metabolism; Y: Nuclear structure).

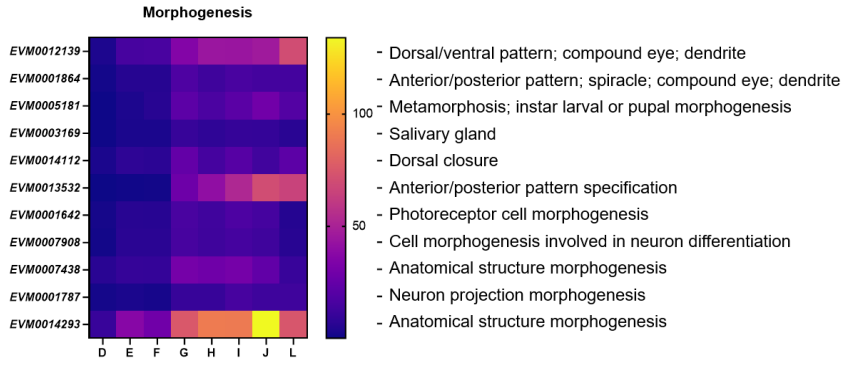


Figure 6. DEGs related to morphogenesis during embryonic development of leech (D-L stage).

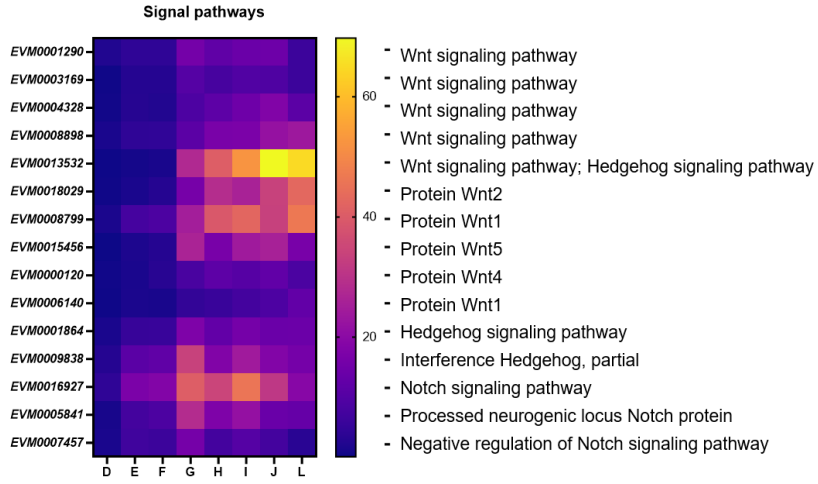


Figure 7. DEGs related to signal pathways during embryonic development of leech (D-L stage).

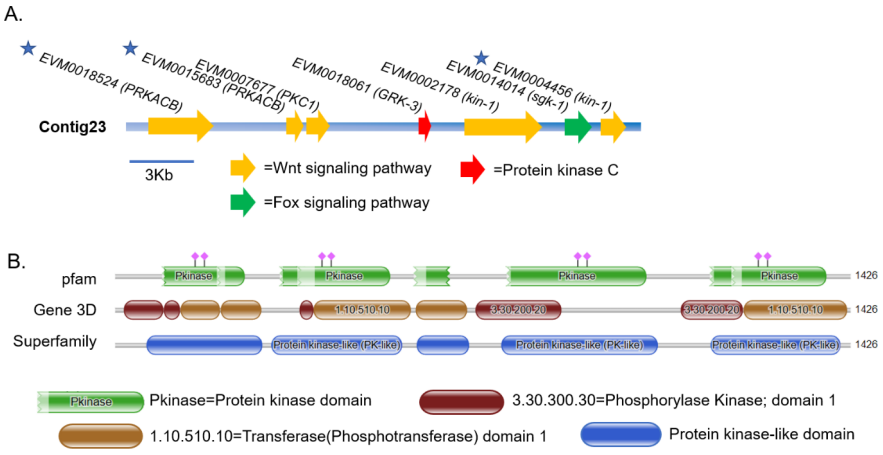


Figure 8. A gene cluster related to Wnt signaling pathway in contig23.*(A)* Seven genes closely distributed in contig23. Asterisks indicated that three genes was involved in synaptic pathways. *(B)*Predicted protein domain arrangement.

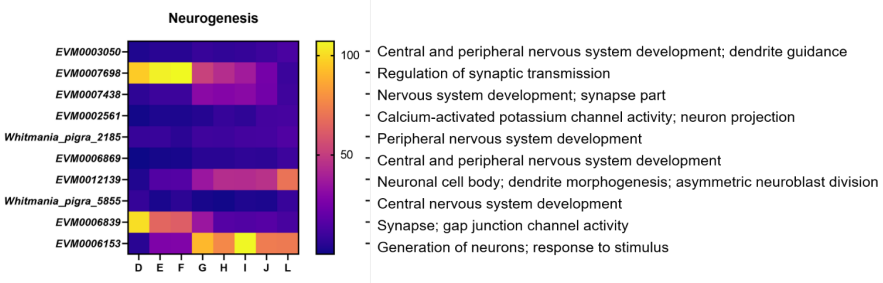
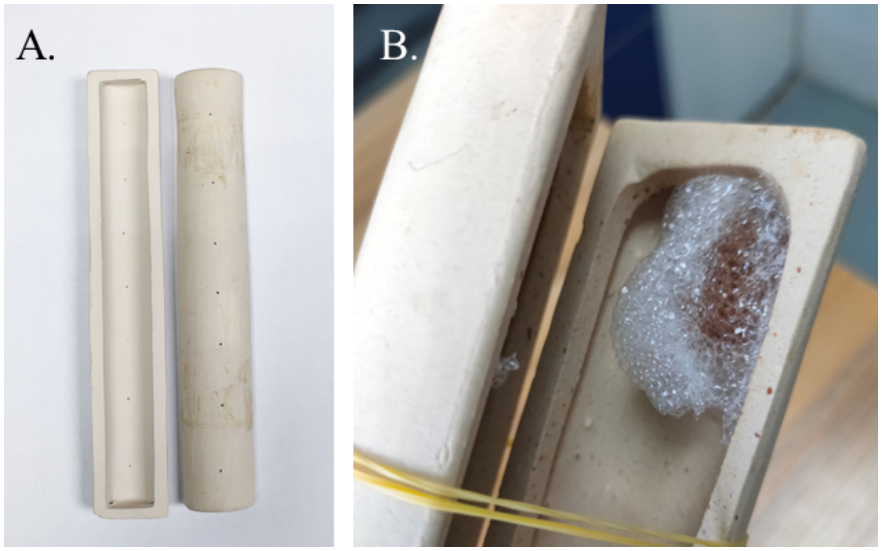
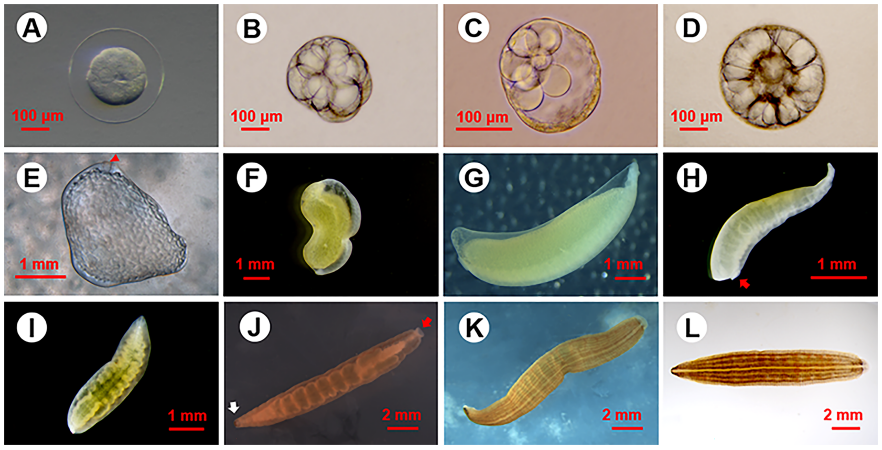


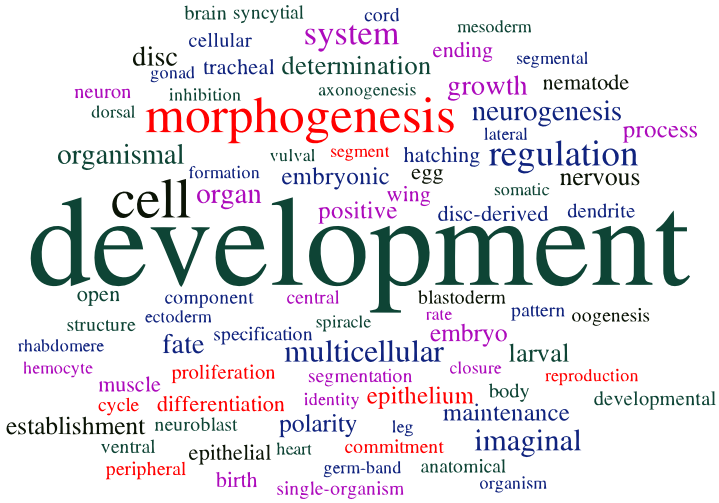
Figure 9. DEGs related to neurodevelopment during D-L stage.

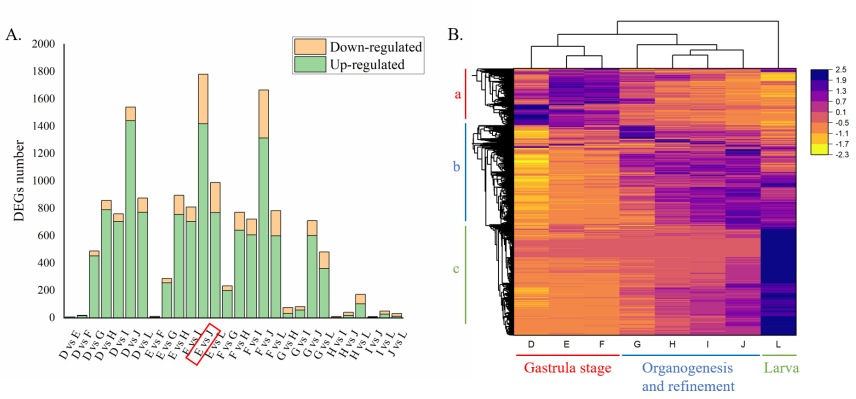
Hosted file

`Table 1.docx` available at https://authorea.com/users/493524/articles/575920-genome-and-transcriptome-of-freshwater-leech-whitmania-pigra-reveal-key-genes-related-to-morphogenesis-signal-pathways-and-neurogenesis-during-embryonic-development









Hosted file

`Annotation.zip` available at https://authorea.com/users/493524/articles/575920-genome-and-transcriptome-of-freshwater-leech-whitmania-pigra-reveal-key-genes-related-to-morphogenesis-signal-pathways-and-neurogenesis-during-embryonic-development

Hosted file

`Mor.svg` available at https://authorea.com/users/493524/articles/575920-genome-and-transcriptome-of-freshwater-leech-whitmania-pigra-reveal-key-genes-related-to-morphogenesis-signal-pathways-and-neurogenesis-during-embryonic-development

Hosted file

`Sig.svg` available at https://authorea.com/users/493524/articles/575920-genome-and-transcriptome-of-freshwater-leech-whitmania-pigra-reveal-key-genes-related-to-morphogenesis-signal-pathways-and-neurogenesis-during-embryonic-development

Hosted file

`wnt.svg` available at https://authorea.com/users/493524/articles/575920-genome-and-transcriptome-of-freshwater-leech-whitmania-pigra-reveal-key-genes-related-to-morphogenesis-signal-pathways-and-neurogenesis-during-embryonic-development

Hosted file

`Neu.svg` available at https://authorea.com/users/493524/articles/575920-genome-and-transcriptome-of-freshwater-leech-whitmania-pigra-reveal-key-genes-related-to-morphogenesis-signal-pathways-and-neurogenesis-during-embryonic-development