

RESEARCH ARTICLE

De novo transcriptome assembly of four organs of *Collichthys lucidus* and identification of genes involved in sex determination and reproduction

Wei Song^{1,2}, YiBing Zhang¹, XiaoJuan Zhang¹, JianFang Gui^{1*}**1** Institute of Hydrobiology, Chinese Academy of Sciences, Wuhan, China, **2** East China Sea fisheries Research Institute, Chinese Academy of Fishery Sciences, Shanghai, China* jfgui@ihb.ac.cn**OPEN ACCESS**

Citation: Song W, Zhang Y, Zhang X, Gui J (2020) De novo transcriptome assembly of four organs of *Collichthys lucidus* and identification of genes involved in sex determination and reproduction. PLoS ONE 15(3): e0230580. <https://doi.org/10.1371/journal.pone.0230580>

Editor: Peng Xu, Xiamen University, CHINA

Received: December 18, 2019

Accepted: March 3, 2020

Published: March 27, 2020

Copyright: © 2020 Song et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: All relevant data are available at the NCBI database (accession PRJNA605916).

Funding: This work was supported by the Basic Research Fund for State-Level Nonprofit Research Institutes of ESCFRI, CAFS(#Dong2018Z01). This funding body has no role in the design of the study, collection, analysis, interpretation of data or in writing the manuscript.

Competing interests: The authors have declared that no competing interests exist.

Abstract

The spinyhead croaker (*Collichthys lucidus*) is a commercially important fish species, which is mainly distributed in the coastal regions of China. However, little is known about the molecular regulatory mechanism underlying reproduction in *C. lucidus*. A *de novo* transcriptome assembly in brain, liver, ovary and testis tissues of *C. lucidus* was performed. Illumina sequencing generated 60,322,004, 57,044,284, 60,867,978 and 57,087,688 clean reads from brain, liver, ovary and testis tissues of *C. lucidus*, respectively. Totally, 131,168 uni-genes with an average length of 644 bp and an N50 value of 1033 bp were assembled. In addition, 1288 genes were differentially expressed between ovary and testis, including 442 up-regulated and 846 down-regulated in ovary. Functional analysis revealed that the differentially expressed genes between ovary and testis were mainly involved in the function of sexual reproduction, sex differentiation, development of primary male sexual characteristics, female gamete generation, and male sex differentiation. A number of genes which might be involved in the regulation of reproduction and sex determination were found, including *HYAL* and *SYCP3* and *BMP15*. Furthermore, 35,476 simple sequence repeats (SSRs) were identified in this transcriptome dataset, which would contribute to further genetic and mechanism researches. *De novo* transcriptome sequencing analysis of four organs of *C. lucidus* provides rich resources for understanding the mechanism of reproductive development of *C. lucidus* and further investigation of the molecular regulation of sex determination and reproduction of *C. lucidus*.

Introduction

Collichthys lucidus, which belongs to Perciformes, is an economically important fish with small somatotype and usually lives in the benthic zone of coastal waters [1]. It's widely distributed in the middle and western Pacific Ocean, including Japan, Korea and China [2]. *C. lucidus* is appreciated by customers because of its good quality of meat, which contributing to fishery

production [3]. Statistics showed that it reached 222.6t per year [4]. In recent years, studies revealed that the resources of *C. lucidus* have been declining due to the overfishing and the damage of environment [5]. Therefore, it is necessary to carry out genetic research and exploit reproductive techniques to increase fish production for resource protection and ecological balance.

The hypothalamic-pituitary-gonadal (HPG) axis plays a key role in the regulation of animal reproduction and development through the secretion of reproductive hormones regulation [6, 7]. Gonadotropin-releasing hormone (GnRH) is a key neuroendocrine regulator of the HPG axis, and its mRNA expression level keeps changing during the gonadal development cycle, which is positively correlated with gonadal development [8]. Studied reported that GnRH mRNA expression level was remarkably different during the ovarian development cycle of fish such as scorpion *Sebastes rastrelliger*, and *Oncorhynchus mykiss* [9,10]. Abnormal HPG axis can lead to defective reproductive performance of animals. As previous article reported, a variety of natural or synthetic compounds, such as monocrotophos pesticide, semicarbazide and thyroid hormones, could lead to significant effect on fish reproduction through HPG [11–13]. For example, Hou et al. investigated the effects of norethindrone on gonad differentiation and sex determination of juvenile zebrafish *Dario renio*, and revealed that the expressions of genes associated with the HPG axis was altered by exposure to norethindrone, thus resulting in affecting sex differentiation [14]. In addition, Liang et al. found that progesterone had effects on reproduction by changing the transcriptional expression of genes related to HPG axes during the early development of zebrafish *Dario renio* [15]. However, the regulation of reproduction through this axis in *C. lucidus* is still unknown.

In the past few years, studies on *C. lucidus* mostly focused on feeding ecology and genetic diversity [16, 17]. However, the molecular mechanisms of *C. lucidus* have rarely been studied. Transcriptome sequencing is an effective method to investigate genes involved in specific biological processes and is also a rapid and effective method to discover molecular markers such as SSRs, single nucleotide polymorphisms (SNPs) [18]. With the rapid development of high-throughput sequencing, the transcriptome sequencing have been widely used. Liu et al. performed *de novo* transcriptome sequencing of muscle tissues of *C. lucidus* and provided rich molecular resources for *C. lucidus* [19]. Chen et al. obtained 78,671 putative SNPs using RNA-seq and demonstrated that 29 SNP loci exhibited bi-allelic polymorphism in genotyping [20]. However, the gonad transcriptome data of *C. lucidus* is not currently available.

The HPG axis plays an important role in fish reproductive development. The liver is the main hormone secreting organ in fish, and the development of fish gonads requires the participation and regulation of various steroid hormones. Therefore, in this study, brain, liver, ovary, and testis tissues of *C. lucidus* were sequenced and *de novo* assembled to obtain the transcriptome data and discover genes that might be involved in the regulation of reproductive development using Illumina NextSeq 500 platform. The analysis of transcriptome sequences will provide valuable resources for understanding the mechanism governing reproductive development in *C. lucidus* and it will be helpful for the further research on reproduction of *C. lucidus*. Furthermore, the identification and analysis of SSR loci and the molecular markers obtained in our study will lay the foundation for the research of population genetics in *C. lucidus*.

Materials and methods

Ethical statement

The collection of samples and the handling of animals in this article were approved by the Chinese Academy of Fishery Sciences Welfare Committee, Shanghai, China. All procedures

conducted with the fishes were performed in accordance with relevant guidelines and regulations. All efforts were made to minimize animal suffering.

Sample collection and RNA extraction

Five wild *C. lucidus* individuals were obtained from Ningde sea area, Fujian province, China. Fishes were euthanized by intraperitoneal injection with excessive pentobarbital (150mg/kg). Brain, liver, ovary and testis tissues of *C. lucidus* weighing about 0.5 grams each were harvested. All samples were saved in RNAlater (TransGen Biotech, Beijing, China) and then stored at -80°C for RNA extraction. Total RNAs of each sample were extracted using Animal Total RNA Isolation Kit (Foregene, Chengdu, China) following the introductions by manufacture. Total RNA concentration was measured by Nanodrop 2000 (Invitrogen, Carlsbad, CA, USA), and the RNA integrity was evaluated by Agilent 2100 Bioanalyzer (Agilent Technologies, CA, USA).

cDNA library preparation and sequencing

Qualified RNA was processed using Genomic Sample Preparation kit (Illumina, Beijing, China) for library construction. Briefly, mRNA was isolated using magnetic beads with Oligo (dT) and the mRNA was then mixed with fragmentation buffer to obtain short fragments of 200-300bp. Next, the fragments were used to synthesize first-strand cDNA with random primers, and first-strand cDNA was transformed into double-strand cDNA by using DNA Polymerase I and RNase H. Then these cDNA fragments were processed by an end-repair and the ligation of adapters according to the manufacturer's protocol (Beckman Coulter, Beverly, USA). The products were purified and enriched with PCR. Finally, the cDNA library qualified with an Agilent Bioanalyzer 2100 system. The resultant cDNA libraries were sequenced on the Illumina NextSeq 500 platform (Illumina, USA).

Sequence data processing and *de novo* assembly

To obtain high-quality clean reads, raw reads with adaptors and poly-N, and low-quality reads containing more than 50% bases with a Q-values <20 were removed. *De novo* assembly with clean reads was performed using Trinity software, by which transcripts and unigenes were generated.

Functional annotation

Gene functions of all the assembled unigenes were annotated to NCBI non-redundant protein sequences (Nr, <http://www.ncbi.nlm.nih.gov/protein/>), Kyoto Encyclopedia of Genes and Genomes (KEGG, <http://www.kegg.jp/>), Gene Ontology (GO, <http://www.geneontology.org/>), evolutionary genealogy of genes: Non-supervised Orthologous Groups (eggNOG, http://eggnoG.embl.de/version_4.0.beta/) and Swiss-Prot (http://web.expasy.org/docs/swiss-prot_guideline.html).

Quantification of gene expression levels and functional enrichment

Clean reads were mapped to assembled transcriptome using RSEM to obtain the read count for each gene and the expression level of each gene was estimated using fragments per kilobase of transcript sequence per millions base pairs sequenced (FPKM). Differentially expressed genes (DEGs) were identified with a standard of fold change and p-value ($|\log_2(\text{fold change})| > 1$ and p-value < 0.05).

After obtaining the DEGs, The function was predicted by GO analysis in three categories: cellular component (CC), molecular function (MF) and biological process (BP). KEGG was performed to identify the cellular signal pathways that DEGs involved in. The threshold of significance for GO and KEGG analyses was defined by p value < 0.05 .

Simple Sequence Repeat (SSR) identification

MISA (Version 1.0, <http://pgrc.ipkgatersleben.de/misa/misa.html>) was used to identify SSR motifs. For di-, tri-, tetra-, penta- and hexa-nucleotides, the lowest number of base repeats was 10, 6, 5, 5, 5 and 5, respectively. In addition, if the distance between two SSRs was less than 100 bp, they would be merged into a composite SSR.

Quantitative real-time PCR (qRT-PCR)

In order to validate the RNA-seq results, eight differentially expressed genes between ovary and testis were chosen for qRT-PCR analysis with the same RNA samples from Illumina sequencing. Gene-specific primers were designed using Primer Premier5.0. B2M was the internal reference gene. First strand cDNA was synthesized using the RevertAid First Strand cDNA synthesis kit (Thermo Fisher Scientific, MA, USA). qRT-PCR was carried out with an ABI Q6 real-time PCR machine (Applied Biosystem Inc., MA, USA) using QuantiFast SYBR Green PCR Kit (Qiagen, Hilden, Germany) in 10 μ L reaction mixture with 1 μ L cDNA as template. The PCR cycling parameters were: 95°C for 10 min; 45 cycles of 95°C for 15 s and 60°C for 60 s followed by a melting curve. Every sample was performed in triplicate. Amplicons were verified through melting-curve analysis. The $2^{-\Delta\Delta CT}$ method was used to determine the relative gene expression. Quantitative data were expressed as $\text{means} \pm \text{SD}$ (standard deviation) representing the relative expression. The significance of differences in expression was determined using one-way ANOVA followed by t -test with a threshold significance level of $P < 0.05$. The primers were displayed in [S1 Table](#).

Investigation of SSR polymorphism

A total of 10 primer pairs ([S2 Table](#)) were synthesized and 210 individuals of *C. lucidus* from 7 different locations, including Lianyungang (LYG), Zhoushan (ZS), Chongming (CM), Wenzhou (WZ), Xiamen (XM), Ningde (ND), Dafeng (DF), were selected for polymorphism investigation with the SSRs. Total DNA was extracted from dorsal muscles using traditional phenol-chloroform extraction protocols as described by Ma et al. [17]. The total volume of PCR reactions were 20 μ L containing 1 μ L template DNA, 1 μ L forward and reverse primers (10 μ M), 1 μ L 2.5 mM dNTPs, 1 μ L EasyTaq DNA polymerase (Beijing Trans Gen Biotech Co., Ltd. China), 2 μ L 10 \times EasyTaq buffer and 13 μ L ddH₂O. Amplification program of PCR was as follows: denaturation for 5 min at 94°C, 35 cycles of 94°C for 30 s, 59°C for 30 s, and 72°C for 1 min, and a final step at 72°C for 10 min. Alleles were detected on a 12% polyacrylamide gel. UPMGA evolution tree was constructed using MEGA 5.1 based on Nei genetic distance.

Statistical analysis

Statistical analysis was performed through the SPSS 13.0 statistical software (Chicago, IL, USA). The data were expressed as $\text{mean} \pm \text{standard deviation (SD)}$. Comparison between two difference independent groups was performed by two-sided Student's test. A p -value < 0.05 was considered to be a statistically significant difference.

Results

Sequencing and *de novo* assembly

Totally, the transcriptome sequencing generated 237,519,998 raw reads (60,954,060, 57,468,802, 61,486,012 and 57,611,124 reads from brain, liver, ovary and testis tissues of *C. lucidus*, respectively). After removing low quantity reads and trimming adaptor reads, 60,322,004 (98.96%), 57,044,284 (99.26%), 60,867,978 (98.99%) and 57,087,688 (99.09%) clean reads from brain, liver, ovary and testis tissues of *C. lucidus* were obtained, respectively (S3 Table). *De novo* assembly of clean reads generated a total of 189,181 transcripts with an average length of 818 bp and a N50 length of 1463 bp. The longest transcript of each gene was selected and yielded 131,168 unigenes. The average length of unigenes was 644 bp and the N50 length was 1033 bp (S4 Table).

Functional annotation

Unigenes were annotated using NR, GO, KEGG and Swissport databases, and a 10^{-5} e-value cut-off value was used. 51,590 (39.33%), 27,708 (21.12%), 12,883 (9.82%) and 44,051 (33.58%) significant hits were produced, respectively. Among these unigenes, 53,200 (40.56%) were annotated at least in one database and 8851 (6.75%) unigenes were annotated in all four databases (S5 Table). The similarity distribution of the top hits showed that 61% of the mapped sequences had similarities more than 80%, while 37% of the hits had similarities ranging from 40% to 80% (Fig 1A). The E-value distribution had a comparable pattern with 51% of the mapped sequences with high homologies ($<1e-60$), whereas 48% of the homological sequences ranged between $1e-5$ to $1e-60$ (Fig 1B). The species distribution of NR BLAST matches showed that the top two species was *Maylandia zebra* (26%) and *Oreochromis niloticus* (21%) (Fig 1C).

GO and KEGG analysis of *C. lucidus* transcriptomes

A total of 27,708 unigenes were clustered into 67 GO terms under three different GO categories containing (biological process (BP), cellular component (CC) and molecular function (MF). The number of unigenes in each GO term were shown in Fig 2A. The top 3 terms ranked by number of unigenes in BP were cellular process (17,477), single-organismal process (16,219) and biological regulation (11,558). For CC, unigenes were mainly involved in cell (16,229), cell part (16,151) and membrane (11,815). For MF, binding (16,123) was the most prevalent, followed by nucleic acid binding catalytic activity (10,504) and signal transducer activity (1,974).

To further explore the biological pathways, all unigenes were mapped in KEGG database and 12,883 unigenes were mapped to 32 categories, among which signal transduction (1,877), endocrine system (849), immune system (820), signaling molecules and interaction (814) and transport and catabolism (668) were the largest five groups (Fig 2B).

Analysis of differentially expressed unigenes

In order to search for genes involved in reproduction, we identified differentially expressed unigenes among brain, liver, testis and ovary tissues. Between brain and liver, 992 genes were upregulated in brain and 264 genes were upregulated in liver. Between brain and ovary, 749 genes were upregulated in brain and 317 genes were upregulated in ovary. Between brain and testis, 685 genes were upregulated in brain and 388 genes were upregulated in testis. Between liver and ovary, 211 genes were upregulated in liver and 114 genes were upregulated in ovary. Between liver and testis, 138 genes were upregulated in liver and 315 genes were upregulated in testis. Between ovary and testis, 442 genes were upregulated in ovary and 846 genes were

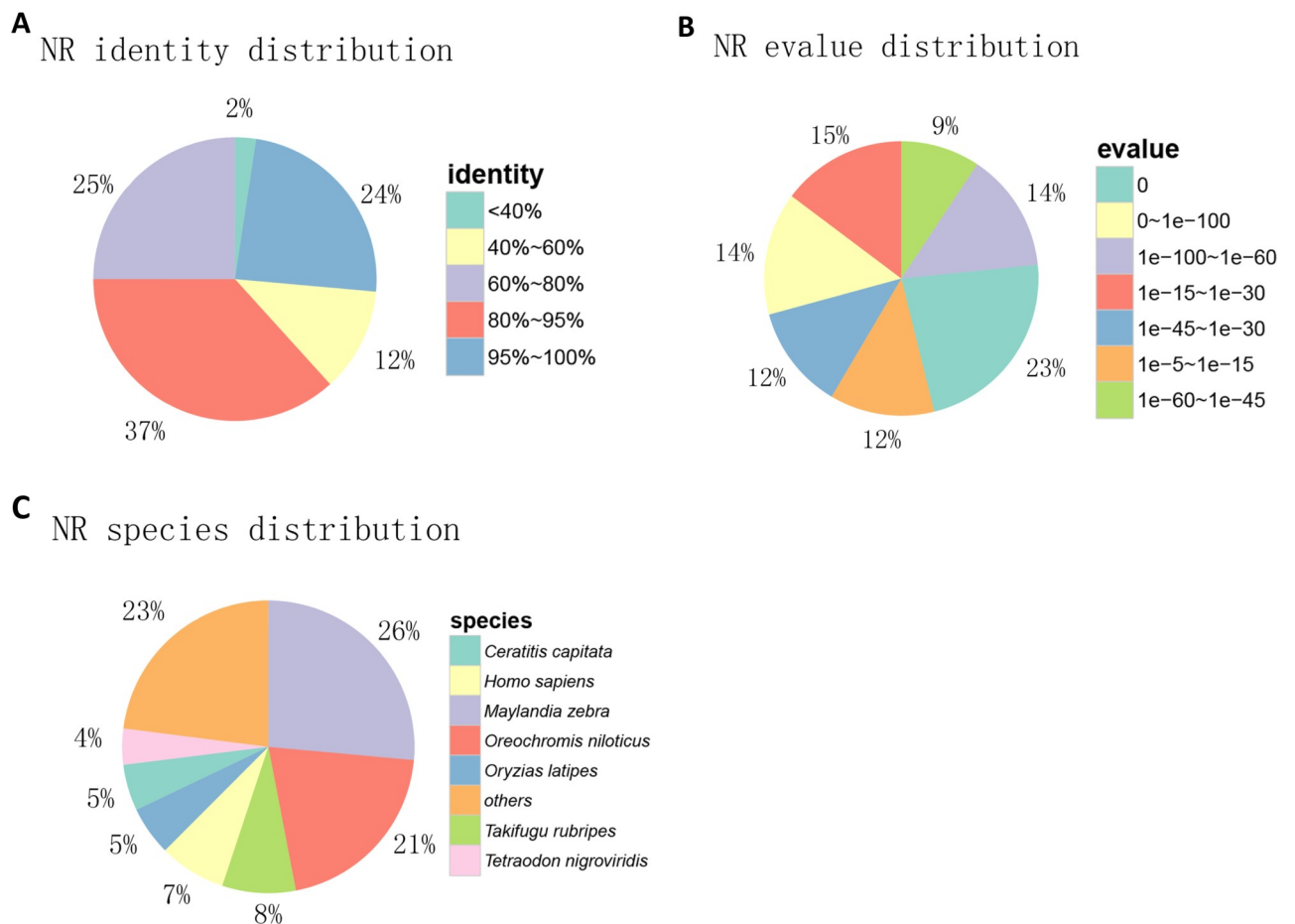


Fig 1. Transcripts BLAST against NR. (A) The similarity of mapped sequences with genes in NR database. (B) E-value distribution. (C) Species distribution.

<https://doi.org/10.1371/journal.pone.0230580.g001>

upregulated in testis (Fig 3, S6 Table). Additionally, cluster heatmap of unigenes revealed that more unigenes were highly expressed in brain, while less unigenes were highly express in liver. A gene was identified as organ-specific when the FPKM was at least 10 in one organ whereas less than 1 in other organs. By this standard, 2,143, 304, 333, 556 unigenes were found specifically expressing in brain, liver, ovary and testis, respectively (Fig 4).

Functional analysis of differentially expressed unigenes

To investigate pathways connected to reproduction and gonad development, we particularly focused on differentially expressed unigenes between ovary and testis. These differentially expressed unigenes were assigned to 165 pathways. Differentially expressed unigenes between ovary and testis were mainly associated with PPAR signaling pathway, ovarian steroidogenesis, oocyte meiosis, steroid hormone biosynthesis, apoptosis-multiple species and proximal tubule bicarbonate reclamation (Fig 5).

Identification and characterization of genes involved in reproduction and sexual development

Based on the *de novo* assembly and annotation, 652 and 650 candidate unigenes were identified for reproduction (GO: 0000003) and reproductive process (GO: 0022414) respectively.

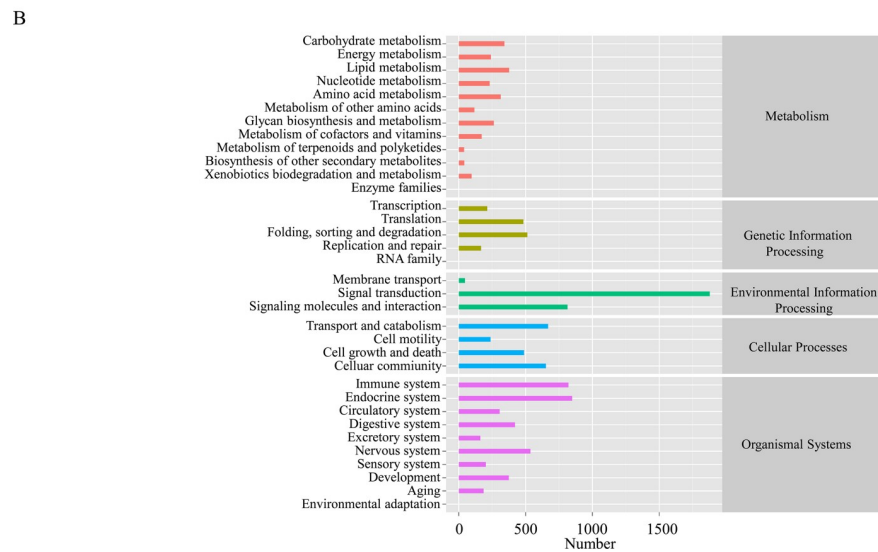
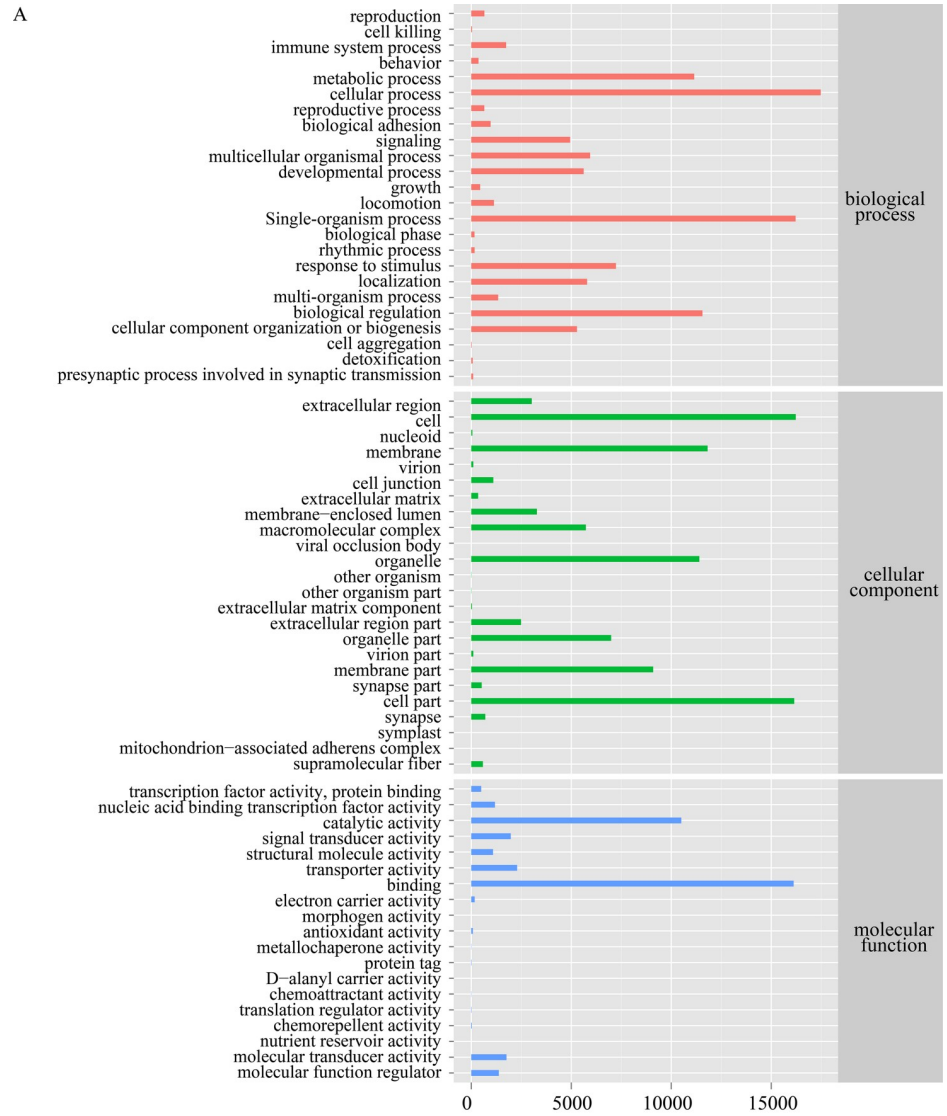


Fig 2. Functional analysis of *C. lucidus* unigenes. (A) GO annotation of unigenes from *C. lucidus* based on biological processes, molecular function, and cellular components. (B) KEGG annotation of unigenes based on metabolism, genetic information processing, environmental information processing, cellular processes, and organismal systems.

<https://doi.org/10.1371/journal.pone.0230580.g002>

Furthermore, 8 unigenes were enriched in both terms of reproduction, sex differentiation and development, implying that they dominated the regulation of reproduction and sex development. Among the 8 unigenes, *HYAL* (hyaluronidase), *KLHL10* (Kelch-like protein 10), *ROPNIL* (rhopilin associated tail protein 1 like), *ODF3L2* (outer dense fiber of sperm tails 3 like 2), *SYCP3* (synaptonemal complex protein 3) and *SRPK3-like* (serine/threonine-protein kinase 3-like) were upregulated in testis, whereas, *BMP15* (bone morphogenetic protein 15) and *RGS14* (regulator of G protein signaling 14) were upregulated in ovary (Table 1).

Identification of SSRs and SSR marker polymorphism

SSRs were predicted using MISA 1.0 (<http://pgrc.ipkgatersleben.de/misa/misa.html>). Totally, we found 35,476 SSRs. Among them, the most affluent repeat motif was mono- and di-nucleotide repeats, accounting for 41.18% (14,608) and 38.23% (13,563), respectively, followed by tri- (18.74%, 6,648), tetra- (1.73%, 614), penta- (0.09%, 31) and hexa-nucleotide (0.03%, 12) (S7 Table).

To study the genetic structure of different geographic populations of *C. lucidus*, 10 primer pairs were designed to validate the amplification and polymorphism in *C. lucidus* collected from 7 different regions. Among these primer pairs, one pair (primer 187) exhibited polymorphisms among the seven colonial *C. lucidus* (S1A, S1B Fig). The polymorphic SSR marker was then used to perform genetic correlation analysis among the *C. lucidus* groups. The UPGMA clustering produced a dendrogram that separated the *C. lucidus* into two main groups, from which Ningde (ND) completely separate from the others. Moreover, Lianyungang (LYG) and Zhoushan (ZS) formed a sub-cluster, while Chongming (CM) and Dafeng (DF) formed a sub-cluster (S1C Fig).

Validation of gene expression using qRT-PCR

To verify the expression of genes involved in reproduction, 8 genes enriched in reproduction terms were validated by qRT-PCR analysis. In accordance with the results of RNA-seq, qRT-PCR showed that *ROPNIL*, *KLHL10*, *ODF3L2*, *SYCP3* and *SRPK3-like* were specific expressed in testis, whereas *BMP15* and *RGS14* were specific expressed in ovary (Fig 6).

Discussion

Researches have reported that *de novo* transcriptome sequencing has been used in many aquatic animals [21]. Meng et al. obtained 128,904,126 clean reads in all from the ovary and testis tissues of *Portunus trituberculatus* by RNA-Seq [22]. Besides, gonadal transcriptomic sequencing of *Pelteobagrus fulvidraco* generated 193,477,064 clean reads by Lu et al. [23]. However, little is known about the transcriptome of *C. lucidus* organs. In our study, we gathered 60,322,004 (98.96%), 57,044,284 (99.26%), 60,867,978 (98.99%) and 57,087,688 (99.09%) paired-end clean reads from brain, liver, ovary and testis tissues of *C. lucidus*, assembled 189,181 transcripts and identified 131,168 unigenes. It can provide more genetic resources for the research of reproduction in *C. lucidus*.

HPG axis is the sex endocrine system of fish and plays an important role in fish reproduction via a cascade of hormones [24]. Transcriptome analysis of HPG axis from *Lateolabrax maculatus* indicated 748, 349 and 319 unigenes were uniquely expressed in the brain, ovary

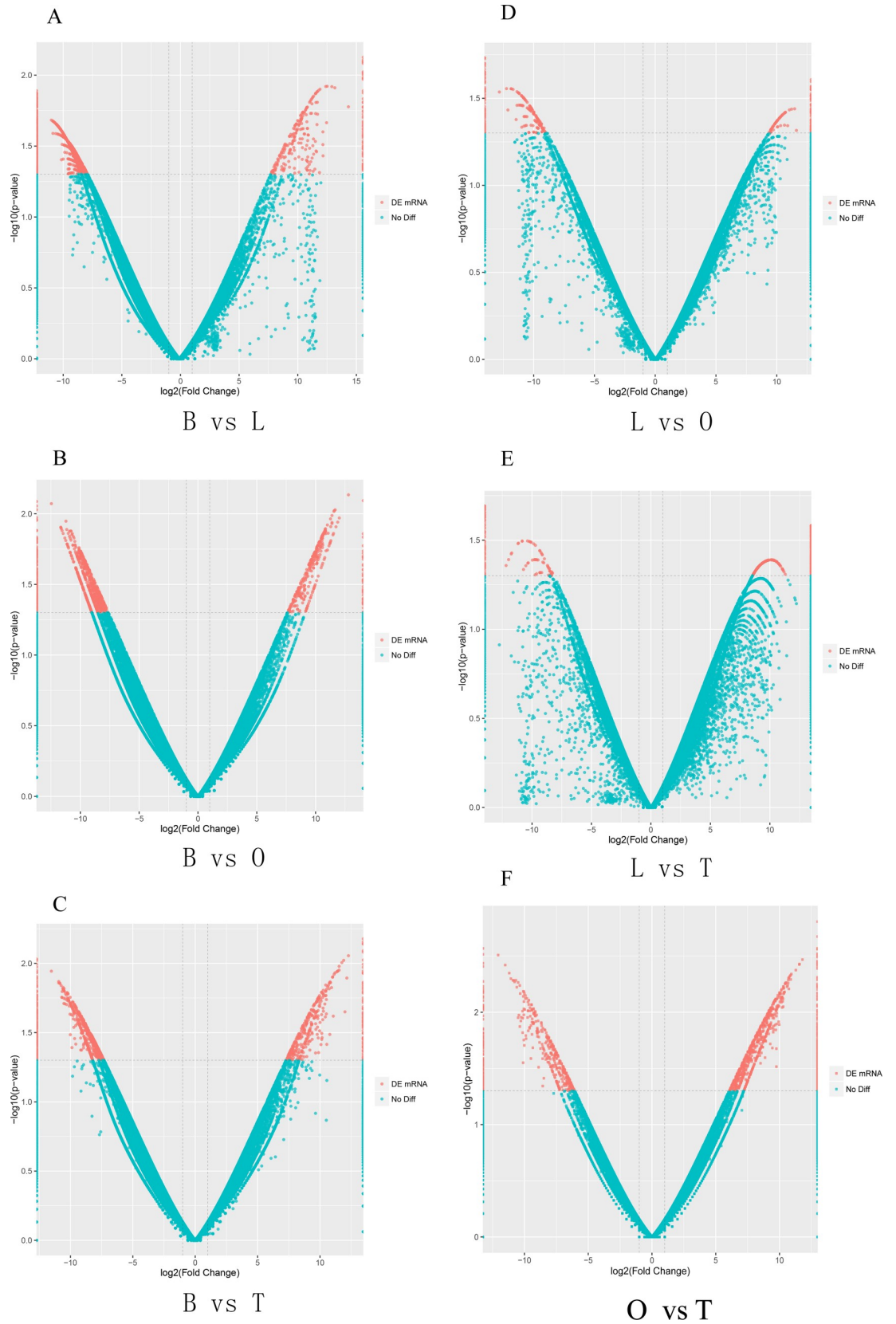


Fig 3. Analysis of differentially expressed unigenes. Volcano plot shows differentially expressed unigenes between each pair of brain (B), liver (L), ovary (O) and testis (T) tissues. (A) Differentially expressed unigenes in B vs L. (B) Differentially expressed unigenes in B vs O. (C) Differentially expressed unigenes in B vs T. (D) Differentially expressed unigenes in L vs O. (E) Differentially expressed unigenes in L vs T. (F) Differentially expressed unigenes in O vs T. Red dots represent differentially expressed unigenes. Blue dots represent not-significant differentially expressed unigenes. The red dots on the left represented the up-regulated unigene expression of the samples on the left, and the red far dots on the right represented the up-regulated unigene expression of the sample on the right.

<https://doi.org/10.1371/journal.pone.0230580.g003>

and testis tissues, respectively, while 26623 unigenes were commonly expressed in all HPG tissues [25]. Xu et al. analyzed the gene expression of HPG axis in *Cyprinus carpio* and suggested that 9, 28, and 212 genes were differentially expressed in hypothalamus, pituitary, and ovary, respectively [26]. Therefore, gene expression of the HPG axis was different in different kinds of fish. In this study, the vast majority of the organ-specific expression was found in brain, containing 2,143 unigenes, whereas, only 304, 333 and 556 in liver, ovary and testis of *C. lucidus*,

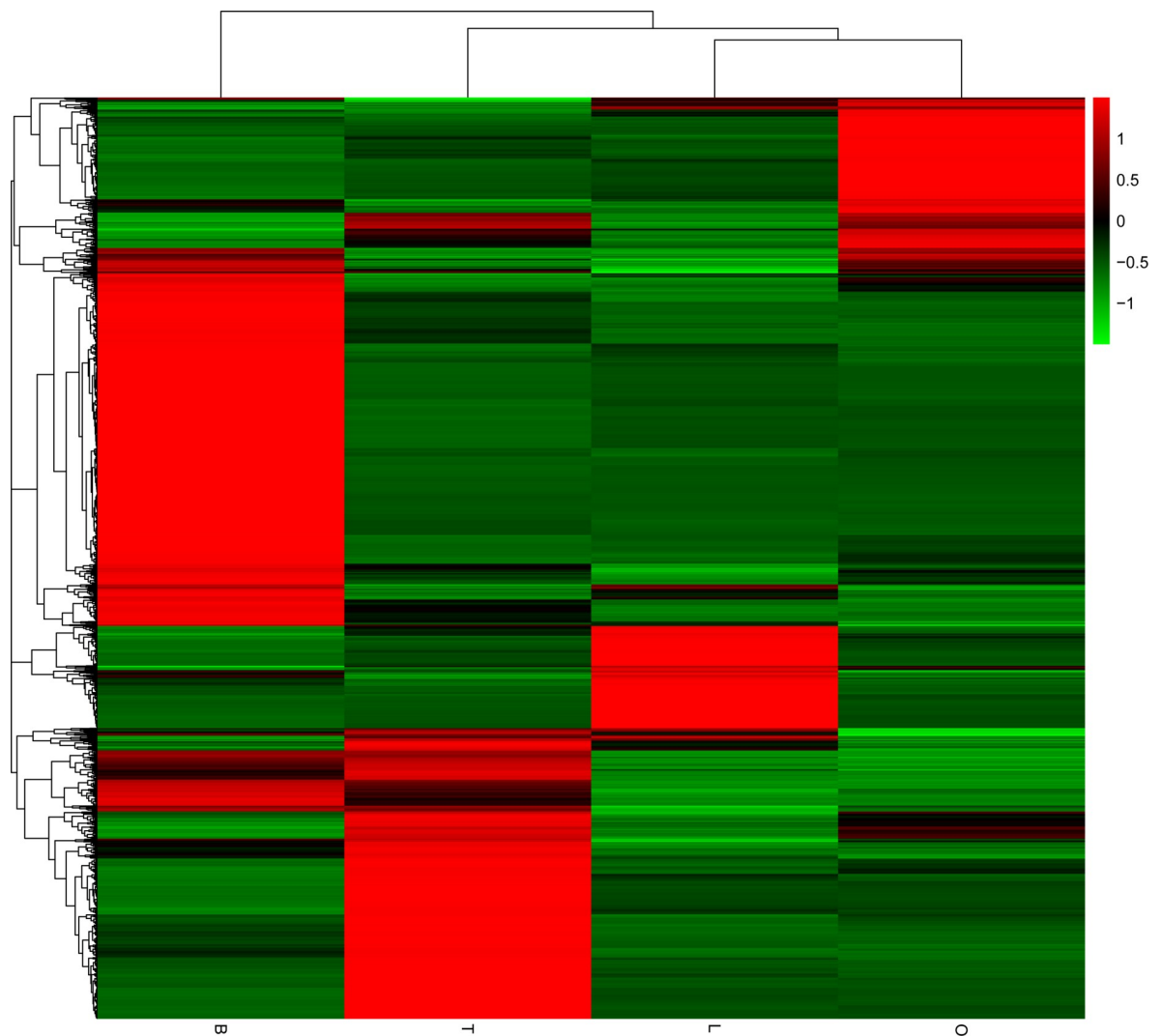


Fig 4. Cluster heatmap of unigenes between any two tissues of B (brain), L (liver), O (ovary) and T (testis). Different rows represent different unigenes. Different colors represent different cluster grouping information, red represents unigenes with high expression, and green represents unigenes with low expression.

<https://doi.org/10.1371/journal.pone.0230580.g004>

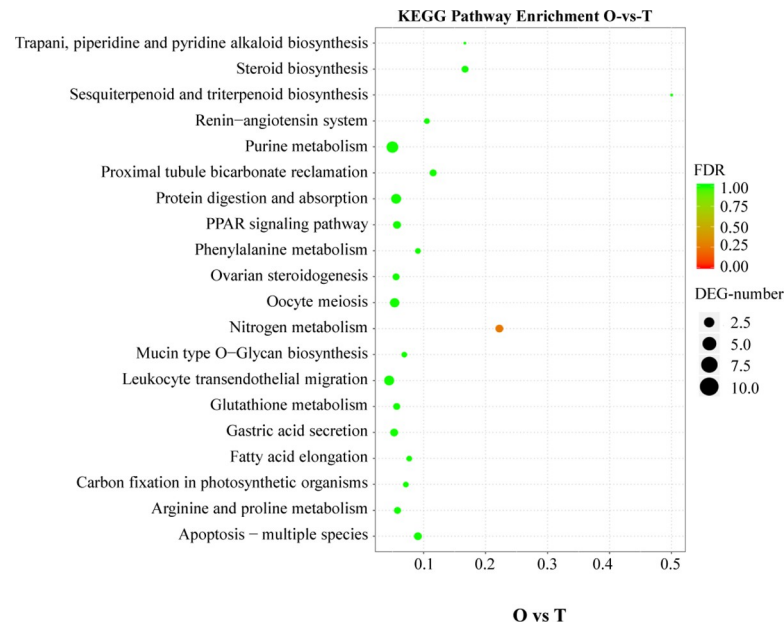


Fig 5. KEGG pathway enrichment analysis of differentially expressed genes between ovary (O) and testis (T) tissues. The size of the dot indicates how much the differentially expressed genes are annotated into the pathway, and the color indicates the significance of the pathway. The figure shows the 20 most significant pathways between O and T tissues.

<https://doi.org/10.1371/journal.pone.0230580.g005>

respectively. In addition, 1288 genes were differentially expressed between ovary and testis. Studies of the expression of genes associated with the HPG axis can promote understanding the role of HPG axis in the regulation of fish reproduction [27]. Wang et al. found that altered expressions of genes in the HPG axis was closely related to the disrupted oogenesis in yellow catfish [28]. Therefore, the expression of genes associated with the HPG axis in the ovary and testis might be altered, resulting in the regulation of the sex determination and reproduction in *C. lucidus*. These data in our study will provide a foundation for further understanding the molecular regulation of gonadal development and reproduction.

GO and KEGG analyses were used for functional categorization of the assembled unigenes, and for acquiring information of gene functions that can contribute to predicting the role of protein interaction networks in cells [29]. In our study, cellular process, cell, binding were the most abundant classifications in biological process, cellular component and molecular function, respectively. This result was similar to the result of Meng et al. who performed *de novo* transcriptome sequencing of *Portunus trituberculatus* ovary and testis [22]. The KEGG

Table 1. Different genes in GO terms related to reproduction.

ID	O	T	foldChange(T/O)	log ₂ FoldChange	pvalue
c35502_g1(HYAL)	0	1950	-	-	0.003
c55252_g1(KLHL10)	2	1765	882.5	9.79	0.008
c55445_g1(BMP15)	9245	20	0.002	-8.85	0.015
c28789_g1(ROPN1L)	15	4292	286.13	8.16	0.019
c47852_g1(ODF3L2)	0	217	-	-	0.023
c65782_g1(SYCP3)	53	20287	382.77	8.58	0.024
c66108_g1(RGS14)	19878	48	0.003	-8.18	0.026
c54644_g1(SRPK3-like)	0	157	-	-	0.028

<https://doi.org/10.1371/journal.pone.0230580.t001>

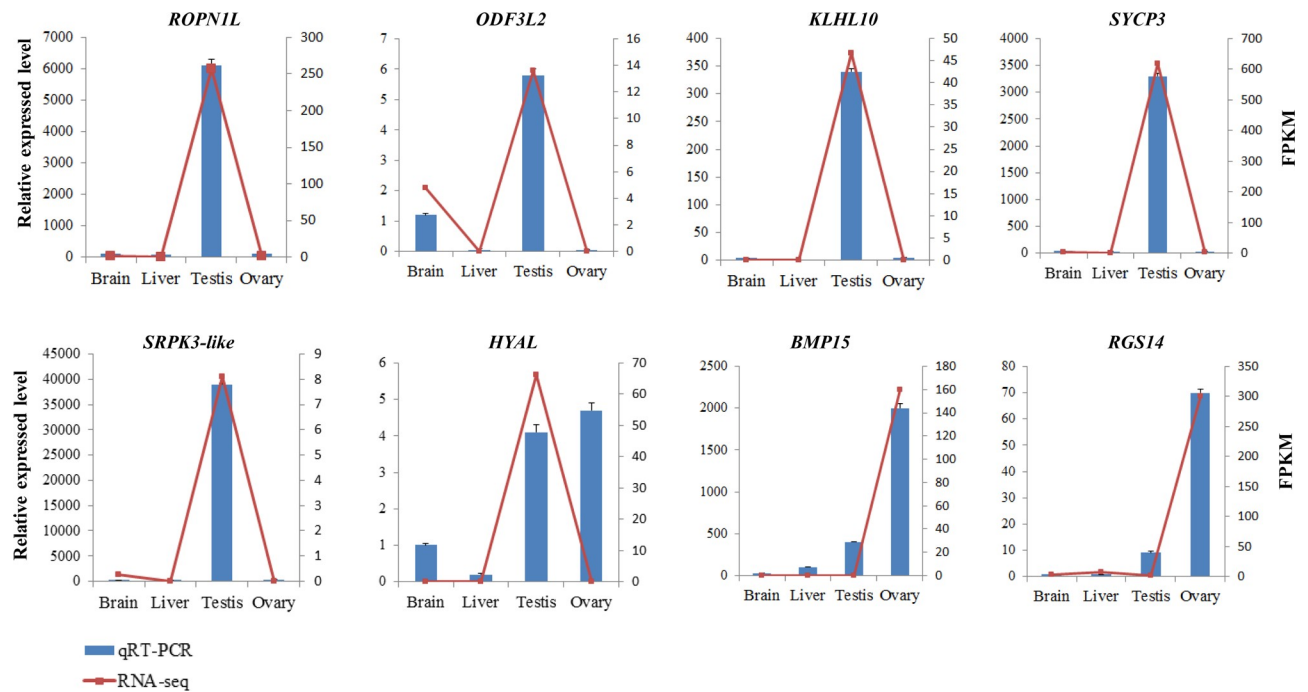


Fig 6. Genes involved in the reproduction were verified by qRT-PCR. *ROPN1L*: rhophilin associated tail protein 1 like, *ODF3L2*: outer dense fiber of sperm tails 3 like 2, *KLHL10*: Kelch-like protein 10, *SYCP3*: synaptonemal complex protein 3, *SRPK3-like*: serine/threonine-protein kinase 3-like, *HYAL*: hyaluronidase, *BMP15*: bone morphogenetic protein, *RGS14*: regulator of G protein signaling 14. * represents significant difference between the two groups ($P < 0.05$).

<https://doi.org/10.1371/journal.pone.0230580.g006>

annotation indicated signal transduction was the largest group that a lot of genes and pathway might be involved. Signal transduction is the process by which physical and chemical signals are transmitted through cells, and this process is crucial for the development and maintenance of normal gametogenesis [30]. In addition, in order to identify pathways participating in reproduction, we carried out KEGG analysis of differentially expressed unigenes between ovary and testis and found a large number of genes which were mainly associated with PPAR signaling pathway, ovarian steroidogenesis, oocyte meiosis, steroid hormone biosynthesis, apoptosis-multiple species and proximal tubule bicarbonate reclamation. Therefore, it can help us further understanding the molecular mechanism of reproduction in *C. lucidus*.

Hyaluronidase was an important component of fertilization in mammals for acrosomal reaction [31]. In this study, *HYAL*, the gene encoding hyaluronidase, was exactly upregulated in testis. It indicated that hyaluronidase may contribute to fertilization in *C. lucidus*. Besides, *SYCP3*, the gene encoding synaptonemal complex protein 3, which closely connect to synapsis in meiosis [32]. In the present study, *SYCP3* was upregulated in testis. In agreement with our result, this gene was examined to reflect the level of meiosis and spermatogenesis in zebrafish, tilapia and catfish [33–35]. In addition, *BMP15* belongs to the subgroup of transforming growth factor β (TGF β) superfamily, and plays important roles in regulating ovarian functions [36]. Dranow et al. found *BMP15* was required for maintaining female sexual fate in zebrafish [37]. In our qRT-PCR, *BMP15* was highly expressed in ovary, suggesting that it might participate in sex determination.

SSRs is one of the most widely used DNA molecular markers [38]. SSR markers are increasingly used as marker systems in molecular genetics studies because of their rich polymorphism, high conservation, and reproducibility [39, 40]. For the conservation of *C. lucidus*,

SSRs can act as effective molecular markers for quantifying genetic diversity. In our study, we screened 35,476 SSRs. The mononucleotide repeats (14608, 41.18%) were predominant followed by dinucleotide (13563, 38.23%) and tricleotide repeats (6648, 18.74%). This was similar to the reported results of Yue et al. who identified 12151 SSRs in which mononucleotide repeat motifs accounted for the first largest group in the *Acipenser sinensis* transcriptome [41]. Moreover, we performed genetic correlation analysis among the *C. lucidus* groups collected from 7 different regions using polymorphic SSR markers and discovered Lianyungang (LYG) and Zhoushan (ZS) formed a sub-cluster, Chongming (CM) and Dafeng (DF) formed a sub-cluster, while Ningde (ND) completely separate from the others. It indicated that different regional population of *C. lucidus* had different genetic relationships. It provides a basis for further study on the genetic structure of different geographic populations of *C. lucidus*.

Conclusion

Our *de novo* transcriptome sequencing generated 131,168 unigenes in all, among which a large number of genes were potentially involved in reproduction and sex determination. This study will enrich genetic information for *C. lucidus* and provide a basis for further research on the molecular regulation of reproduction and sex determination of this species. Additionally, 35,476 SSRs were identified, which will serve as molecular markers for the research of genetic diversity in *C. lucidus*.

Supporting information

S1 Table. PCR primers for the validation of RNA-seq data by qRT-PCR.

(DOCX)

S2 Table. S2 Table 2 Primers used for SSR polymorphic analysis in this study.

(DOCX)

S3 Table. Data analysis of raw reads and clean reads in *C. lucidus* from B (brain), L (liver), O (ovary) and T (testis).

(DOCX)

S4 Table. Statistics of the transcriptome sequencing and *de novo* assembly data in *C. lucidus*.

(DOCX)

S5 Table. Summary of function annotations of unigenes in *C. lucidus*.

(DOCX)

S6 Table. Unigenes differentially expressed in each two organs.

(DOCX)

S7 Table. SSRs identified in *C. Lucidus*.

(DOCX)

S1 Fig. SSR polymorphism analysis of seven *C. lucidus* populations from different locations. (A and B) Polyacrylamide gel electrophoresis shows the SSR polyacrylamide. (C) Dendrogram constructed with UPGMA clustering of *C. Lucidus* from seven different locations. LYG: Lianyungang, ZS: Zhoushan, CM: Chongming; WZ: Wenzhou, XM: Xiamen, ND: Ningde, DF: Dafeng.

(TIF)

Acknowledgments

We show appreciation to Shanghai YingBiotech Company for the dynamic help provided during the analysis of the sequencing data.

Author Contributions

Conceptualization: JianFang Gui.

Data curation: Wei Song, YiBing Zhang, XiaoJuan Zhang.

Funding acquisition: JianFang Gui.

Investigation: Wei Song, YiBing Zhang, XiaoJuan Zhang.

Methodology: Wei Song, JianFang Gui.

Project administration: JianFang Gui.

Resources: JianFang Gui.

Software: JianFang Gui.

Supervision: JianFang Gui.

Validation: Wei Song.

Visualization: JianFang Gui.

Writing – original draft: Wei Song, YiBing Zhang, XiaoJuan Zhang.

Writing – review & editing: Wei Song, YiBing Zhang, XiaoJuan Zhang.

References

1. Li H, Liu D, Li C, Zhou S, Tian D, Xiao D, et al. Exosomes secreted from mutant-HIF-1 α -modified bone-marrow-derived mesenchymal stem cells attenuate early steroid-induced avascular necrosis of femoral head in rabbit. *Cell Biol Int*. 2017; 41:1379–1390. <https://doi.org/10.1002/cbin.10869> PMID: 28877384
2. Liu H, Jiang T, Huang H, Shen X, Zhu J, Yang J. Estuarine dependency in *Collichthys lucidus* of the Yangtze River Estuary as revealed by the environmental signature of otolith strontium and calcium. *Environ Biol Fish*. 2015; 98:165–172.
3. Sang C, Lin Y, Jiang K, Zhang F, Song W. Molecular cloning and mRNA expression of a hepcidin gene from the spinyhead croaker, *Collichthys lucidus*. *Genet Mol Res*. 2015; 14:16050–16059. <https://doi.org/10.4238/2015.December.7.18> PMID: 26662398
4. Zhuang P, Wang Y, Li S, Deng S, Li C, Ni Y. Fishes of the Yangtze estuary. Shanghai Scientific and Technical Publishers, Shanghai (in Chinese). 2006.
5. Huang LM, Li J, Xie YJ, Zhang YZ. Study of the *Collichthys lucidus* fisheries resources of the Minjiang Estuary and its adjacent waters. *J Oceanogr Taiwan Strait*. 2010; 2:250–256.
6. Couse JF, Yates MM, Walker VR, Korach KS. Characterization of the hypothalamic-pituitary-gonadal axis in estrogen receptor (ER) Null mice reveals hypergonadism and endocrine sex reversal in females lacking ER α but not ER β . *Mol Endocrinol*. 2003; 17:1039–1053. <https://doi.org/10.1210/me.2002-0398> PMID: 12624116
7. Ma Y, Han J, Guo Y, Lam PK, Wu RS, Giesy JP, et al. Disruption of endocrine function in in vitro H295R cell-based and in in vivo assay in zebrafish by 2,4-dichlorophenol. *Aquat Toxicol*. 2012; 106–107:173–181. <https://doi.org/10.1016/j.aquatox.2011.11.006> PMID: 22155427
8. Xu J, Huang W, Zhong C, Luo D, Li S, Zhu Z, et al. Defining Global Gene Expression Changes of the Hypothalamic-Pituitary-Gonadal Axis in Female sGnRH-Antisense Transgenic Common Carp (*Cyprinus carpio*). *PLoS ONE*. 2011; 6(6).
9. Collins PM, O'Neill DF, Barron BR, Moore RK, Sherwood NM. Gonadotropin-releasing hormone content in the brain and pituitary of male and female grass rockfish (*Sebastes rastrelliger*) relation to seasonal changes in reproductive status. *Biol Reprod*. 2001; 65(1):173–179. <https://doi.org/10.1095/biolreprod65.1.173> PMID: 11420237

10. Uzbekova S, Lareyre JJ, Madigou T, Davail B, Jalabert B, Breton B. Expression of prepro-GnRH and GnRH receptor messengers in rainbow trout ovary depends on the stage of ovarian follicular development. *Mol Reprod Dev.* 2002; 62(1):47–56. <https://doi.org/10.1002/mrd.10065> PMID: 11933160
11. Tovo-Neto A, Maira DSR, Habibi HR, Nobrega RH. Thyroid hormone actions on male reproductive system of teleost fish. *Gen Comp Endocrinol.* 2018; 265:230–236. <https://doi.org/10.1016/j.ygcen.2018.04.023> PMID: 29678724
12. Yue Z, Yu M, Zhang X, Wang J, Ru S. The anti-androgenic effect of chronic exposure to semicarbazide on male Japanese flounder (*Paralichthys olivaceus*) and its potential mechanisms. *Comp Biochem Physiol C Toxicol Pharmacol.* 2018; 210:30–34. <https://doi.org/10.1016/j.cbpc.2018.04.005> PMID: 29729480
13. Zhang X, Liu W, Wang J, Tian H, Wang W, Ru S. Quantitative analysis of in-vivo responses of reproductive and thyroid endpoints in male goldfish exposed to monocrotophos pesticide. *Comp Biochem Physiol C Toxicol Pharmacol.* 2018; 211:41–47. <https://doi.org/10.1016/j.cbpc.2018.05.010> PMID: 29803892
14. Hou LP, Chen H, Tian CE, Shi WJ, Liang Y, Wu RR, et al. The progestin norethindrone affects sex differentiation and alters transcriptional profiles of genes along the hypothalamic-pituitary-gonadal and hypothalamic-pituitary-adrenal axes in juvenile zebrafish *Danio rerio*. *Aquat Toxicol.* 2018; 201:31–39. <https://doi.org/10.1016/j.aquatox.2018.05.019> PMID: 29859405
15. Liang YQ, Huang GY, Ying GG, Liu SS, Jiang YX, Liu S, et al. The effects of progesterone on transcriptional expression profiles of genes associated with hypothalamic-pituitary-gonadal and hypothalamic-pituitary-adrenal axes during the early development of zebrafish (*Danio rerio*). *Chemosphere.* 2015; 128:199–206. <https://doi.org/10.1016/j.chemosphere.2015.01.062> PMID: 25706437
16. He Z, Xue L, Jin H. On feeding habits and trophic level of *Collichthys lucidus* in inshore waters of northern East China Sea. *Marine Fisheries.* 2011; 33:265–273.
17. Song N, Ma G, Zhang X, Gao T, Sun D. Genetic structure and historical demography of *Collichthys lucidus* inferred from mtDNA sequence analysis. *Environ Biol Fish.* 2014; 97:69–77.
18. Teaniniuraitemoana V, Huvet A, Levy P, Klopp C, Lhuillier E, Gaertner-Mazouni N, et al. Gonad transcriptome analysis of pearl oyster *Pinctada margaritifera*: identification of potential sex differentiation and sex determining genes. *BMC Genomics.* 2014; 15:491. <https://doi.org/10.1186/1471-2164-15-491> PMID: 24942841
19. Liu LW, Sui YZ, Xu KD, Zhou YD. Development of molecular resources for the spinyhead croaker (*Collichthys lucidus*) using illumina sequencing. *Meta Gene.* 2017; 13:1–4.
20. Chen W, Song W, Chen F, Zhao M, Ma C, Zhang F, et al. Development and characterization of SNP derived from spinyhead croaker (*Collichthys lucidus*) by RNA-seq. *Conserv Genet Resour.* 2017; 9:573–577.
21. Ma C, Cheng Q, Zhang Q, Zhuang P, Zhao Y. Genetic variation of *Coilia ectenes* (Clupeiformes: Engraulidae) revealed by the complete cytochrome b sequences of mitochondrial DNA. *J Exp Ma Biol Ecol.* 2010; 385:14–19.
22. Meng XL, Liu P, Jia FL, Li J, Gao BQ. De novo Transcriptome Analysis of *Portunus trituberculatus* Ovary and Testis by RNA-Seq: Identification of Genes Involved in Gonadal Development. *PLoS ONE.* 2015; 10:e0128659. <https://doi.org/10.1371/journal.pone.0128659> PMID: 26042806
23. Lu J, Luan P, Zhang X, Xue S, Peng L, Mahbooband S, et al. Gonadal transcriptomic analysis of yellow catfish (*Pelteobagrus fulvidraco*): identification of sex-related genes and genetic markers. *Physiol Genomics.* 2014; 46:798–807. <https://doi.org/10.1152/physiolgenomics.00088.2014> PMID: 25185028
24. Gui JF, Zhu ZY. Molecular basis and genetic improvement of economically important traits in aquaculture animals. *Chin Sci Bull.* 2012; 57:1751–1760.
25. Wang ZP, Wang D, Wang CL, Xie WJ, Zhu YF, Chen XW. Transcriptome characterization of HPG axis from Chinese sea perch *Lateolabrax maculatus*. *J Fish Biol.* 2017; 91:1407–1418. <https://doi.org/10.1111/jfb.13459> PMID: 28913865
26. Xu J, Huang W, Zhong C, Luo D, Li S, Zhu Z, et al. Defining global gene expression changes of the hypothalamic-pituitary-gonadal axis in female sGnRH-antisense transgenic common carp (*Cyprinus carpio*). *PLoS ONE.* 2011; 6:e21057. <https://doi.org/10.1371/journal.pone.0021057> PMID: 21695218
27. Takahashi A, Kanda S, Abe T, Oka Y. Evolution of the Hypothalamic-Pituitary-Gonadal Axis Regulation in Vertebrates Revealed by Knockout Medaka. *Endocrinology.* 2016; 157(10):3994–4002. <https://doi.org/10.1210/en.2016-1356> PMID: 27560548
28. Wang Y, Li Y, Chen Q, Liu Z. Diethylstilbestrol impaired oogenesis of yellow catfish juveniles through disrupting hypothalamic-pituitary-gonadal axis and germ cell development. *J Appl Toxicol.* 2017; 38(3):308–317. <https://doi.org/10.1002/jat.3529> PMID: 28960386

29. Liu Y, Hui M, Cui Z, Luo D, Song C, Li Y, et al. Comparative Transcriptome Analysis Reveals Sex-Biased Gene Expression in Juvenile Chinese Mitten Crab *Eriocheir sinensis*. PLoS ONE. 2015; 10: e0133068. <https://doi.org/10.1371/journal.pone.0133068> PMID: 26193085
30. Hermann BP, Heckert LL. Transcriptional regulation of the FSH receptor: new perspectives. Mol Cell Endocrinol. 2007; 260–262:100–108. <https://doi.org/10.1016/j.mce.2006.09.005> PMID: 17084019
31. Yoon S, Chang KT, Cho H, Moon J, Kim JS, Min SH, et al. Characterization of pig sperm hyaluronidase and improvement of the digestibility of cumulus cell mass by recombinant pSPAM1 hyaluronidase in an in vitro fertilization assay. Anim Reprod Sci. 2014; 150:107–114. <https://doi.org/10.1016/j.anireprosci.2014.09.002> PMID: 25261076
32. Syrjanen JL, Pellegrini L, Davies OR. A molecular model for the role of SYCP3 in meiotic chromosome organisation. ELife. 2014; 3.
33. Lin Q, Mei J, Li Z, Zhang X, Zhou L, Gui JF. Distinct and Cooperative Roles of amh and dmrt1 in Self-Renewal and Differentiation of Male Germ Cells in Zebrafish. Genetics. 2017; 207:1007–1022. <https://doi.org/10.1534/genetics.117.300274> PMID: 28893856
34. Fang X, Wu L, Yang L, Song L, Cai J, Luo F, et al. Nuclear progesterin receptor (Pgr) knockouts resulted in subfertility in male tilapia (*Oreochromis niloticus*). J Steroid Biochem Mol Biol. 2018; 182:62–71. <https://doi.org/10.1016/j.jsbmb.2018.04.011> PMID: 29705270
35. Laldinsangi C, Senthilkumaran B. Expression profiling of c-kit and its impact after esiRNA silencing during gonadal development in catfish. Gen Comp Endocrinol. 2018; 266:38–51. <https://doi.org/10.1016/j.ygcen.2018.04.004> PMID: 29625123
36. Tan Q, Zagrodny A, Bernaudo S, Peng C. Regulation of membrane progesterin receptors in the zebrafish ovary by gonadotropin, activin, TGF-beta and BMP-15. Mol Cell Endocrinol. 2009; 312:72–79. <https://doi.org/10.1016/j.mce.2009.03.011> PMID: 19773085
37. Dranow DB, Hu K, Bird AM, Lawry ST, Adams MT, Sanchez A, et al. Bmp15 is an oocyte-produced signal required for maintenance of the adult female sexual phenotype in zebrafish. PLoS genetics. 2016; 12:e1006323. <https://doi.org/10.1371/journal.pgen.1006323> PMID: 27642754
38. Mohindra V, Singh A, Barman A, Tripathi R, Sood N, Lal KK. Development of EST derived SSRs and SNPs as a genomic resource in Indian catfish, *Clarias batrachus*. Mol Biol Rep. 2012; 39:5921–5931. <https://doi.org/10.1007/s11033-011-1404-z> PMID: 22207175
39. Sundaray JK, Rasal KD, Chakrapani V, Swain P, Kumar D, Ninawe AS, et al. Simple sequence repeats (SSRs) markers in fish genomic research and their acceleration via next-generation sequencing and computational approaches. Aquacult Int. 2016; 24:1089–1102.
40. Li YC, Korol AB, Fahima T, Beiles A, Nevo E. Microsatellites: genomic distribution, putative functions and mutational mechanisms: a review. Mol Ecol. 2002; 11:2453–2465. <https://doi.org/10.1046/j.1365-294x.2002.01643.x> PMID: 12453231
41. Yue H, Li C, Du H, Zhang S, Wei Q. Sequencing and De Novo Assembly of the Gonadal Transcriptome of the Endangered Chinese Sturgeon (*Acipenser sinensis*). PLoS ONE. 2015; 10:e0127332. <https://doi.org/10.1371/journal.pone.0127332> PMID: 26030930