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

CURRENT STATUS: UNDER REVIEW

Wei Song
Institute of Hydrobiology, Chinese Academy of Sciences; East China Sea Fisheries Research Institute, Chinese Academy of Fishery Sciences

Yi-Bing Zhang
Institute of Hydrobiology, Chinese Academy of Sciences

Xiao-Juan Zhang
Institute of Hydrobiology, Chinese Academy of Sciences

Jianfang Gui
jfgui@ihb.ac.cn Corresponding Author
ORCID: https://orcid.org/0000-0003-4969-1482

DOI: 10.21203/rs.2.9879/v1

SUBJECT AREAS
Epigenetics & Genomics

KEYWORDS
Collichthys lucidus, de novo transcriptome, GO annotation, reproduction
Abstract

Background

Spinyhead croaker (Collichthys lucidus) is an important kind of fish for commercial fishery because of its delicious meat and abundant resource. However, the resource has been declining with the overfishing and environmental disruption. Unfortunately, the mechanism of sex determination and reproduction in Collichthys lucidus is still confusing, which has largely hindered molecular regulation and reproduction studies.

Results

Based on transcripts sequencing in brain, liver, ovary and testis, a de novo transcriptome assembly of Collichthys lucidus was performed. Totally, 131,168 non-redundant unigenes with an average length of 644 bp and an N50 value of 1033 bp, among which 2,143, 304, 333 and 556 were found specifically expressing in brain, liver, ovary and testis, respectively. Importantly, 1288 genes were differently expressed between ovary and testis, including 442 up- and 846 down-regulated in ovary. The differently expressed genes were involved in function of sexual reproduction, sex differentiation, development of primary male sexual characteristics, female gamete generation, and male sex differentiation. HYAL and SYCP3, two genes which were reported to play dominant roles in reproduction, may regulated the key of fertilization and meiosis in Collichthys lucidus. Meanwhile, 35,476 simple sequence repeats (SSRs) were identified, which would contribute to further genetic and mechanism researches.

Conclusions

The studies of genetic information, molecular regulation and reproduction of Collichthys lucidus would be expand due to the sequencing and de novo assembling on the transcripts of major organs. In addition, the identification of candidate genes involved in reproduction will be useful for protecting the resource of Collichthys lucidus.

Background

*Collichthys lucidus*, which belongs to perciformes, is a kind of economically important fish with small somatotype usually inhabiting in the benthic zone of coastal waters[
1]. It is mainly captured in northwestern Pacific that ranges from Kyushu of Japan to the south China sea and makes a major contribution to demersal fish catches, supporting an important commercial fishery[2]. For its good quality of meat, this species is appreciated by customers. However, because of the overfishing and the damage of environment, the resource of *Collichthys lucidus* has been declining in recent years[3]. Protecting *Collichthys lucidus* resource is quite necessary.

Hypothalamic–pituitary–gonadal axis (HPG) play a dominant role in the reproductive manipulation[4, 5]. As previous article reported, a variety of natural or synthetic compounds, such as monocrotophos pesticide (MCP), semicarbazide (SMC) and thyroid hormones (THs) [6-8], could lead to significant influence on fish reproduction through HPG. With the environment pollution pricking up, zebrafish, Japanese flounder and goldfish were all injured by contaminants[7-9]. However, the regulation of reproduction through this axis in *Collichthys lucidus* is still unknown.

With the rapid development of next-generation high-throughput sequencing technology and ubiquitous application of RNA-seq, gene profiles of many species such as mice, human, and rice[10, 11] were constructed. Moreover, as the paired-read tag sequencing strategy has been widely used, the efficiency of sequencing has been much higher. The transcriptomic information can be used in a really wide field of biological researches, including SSR mining and SNP discovery[12], as well as provides an in-depth understanding of biological processes.

In this study, 4 organizations of *Collichthys lucidus* were sequenced and de novo the transcriptome using Illumina Hiseq 2500 platform. Differentially expressed genes were detected and screened using RSEM and Deseq. The analysis of transcriptome sequences expanded the comprehension of functional genomic studies. Additionally, the differentially enrichment in GO terms contributed to understand the
mechanism of the reproductive development and be helpful for the subsequent research on sex universality of *Collichthys lucidus*.

**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**

*Collichthys lucidus* were obtained from Ningde sea area, Fujian province, and once the capture from the sea, fishes were euthanized by intraperitoneal injection with excessive pentobarbital (150 mg/kg).

Brain, liver, ovary and testis tissues from five healthy fishes were collected randomly and saved in RNAlater (TransGen Biotech, Beijing, China). Moreover, 0.3-0.3 grams of each tissue was taken from each individual and mix into a mix pool for each organ. Animal Total RNA Isolation Kit (Foregene, Chengdu, China) was used to extract the total RNA of each sample following the introductions by manufacture and samples of same organization were pooled.

**cDNA library preparation and sequencing**

RNA quality and quantity were evaluated by Agilent 2100 Bioanalyzer (Agilent, Shanghai, China).

Qualified RNA was processed using Genomic Sample Preparation kit (Illumina, Beijing, China) for library construction and sequenced on Hiseq™ 2500 (Illumina) with a paired-end 150 bp read run.

**De novo assembly**

To remain high-quality clean reads, reads which had adaptors, poly-N or Q-values <20 were removed.

De novo assembly with clean reads was performed using Trinity (version r20140717, K-mer 25 bp) software. Transcriptome assembly was accomplished using a part of Trinity called Butterfly and the longest transcript of each gene was designed as unigenes.

**Functional annotation**
All unigenes were distributed into public databases, containing NCBI nonredundant protein sequences (NR), Gene Ontology (GO), Kyoto Encyclopedia of Genes and Genome (KEGG), evolutionary genealogy of genes: Nonsupervised Orthologous Groups (eggNOG) and Swiss-Prot, for functional annotation to find the relationship between *Collichthys lucidus* and other fishes and detect the functions of unigenes.

**Simple sequence repeat (SSR) identification**

MISA Version 1.0, http://pgrc.ipk.gatersleben.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 and 5, respectively. In addition, if the distance between two SSRs was less than 100 bp, they would be merged into a composite SSR.

**Quantification of gene expression levels**

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 were identified with a standard of fold change and p-value (|log2(fold change)| > 1 & p-value< 0.05)

**Quantitative real-time PCR (qRT-PCR)**

Gene-specific primers were designed by Primer Premier5.0. First strand cDNA was synthesized using the RevertAid First Strand cDNA synthesis kit (ThermoFisher). FastStart Universal SYBR Green Master mix was used to amplify cDNA on QuantStudio 6 Flex Real-Time PCR System (ThermoFisher). The primers were displayed in (Additional file 1). The $2^{-\Delta\Delta Ct}$ method was used to determine the relative gene expression with internal control of B2M.

**Investigation of SSR polymorphism**

A total of 10 primer pairs (Additional file 1) were synthesized and 210 *Collichthys lucidus* from 7 different locations were selected for polymorphism investigation with the SSRs. The 7 locations including Lianyungang (LYG), Zhoushan (ZS), Chongming (CM), Wenzhou (WZ), Xiamen (XM), Ningde
(ND), Dafeng (DF). Total DNA was extracted from dorsal muscles by the CTAB method. The total volume of PCR reactions were 20 μL containing 1 μL template DNA, 1 μL each primer (10 μM), 1 μL 2.5 mM dNTPs, 1 μL EasyTaq DNA polymerase (Beijing Trans Gen Biotech Co., Ltd. China), 2 μL 10×EasyTaq buffer and 13 μL 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 by the SPSS 13.0 statistical software (Chicago, IL, USA). The data were expressed as mean ± standard deviation (SD). Comparison between two difference independent groups were 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, 60,954,060, 57,468,802, 61,486,012, 57,611,124 paired-end reads were generated by samples from brain, liver, ovary and testis of *Collichthys 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 were obtained (8.0 G). De novo assembly of clean reads generated 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 (Table 1). Most transcripts and unigenes were 200 to 299 bp in length (Fig 1A, B). These unigenes were retained for subsequent functional annotation.

**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 five databases (Table 1). The similarity distribution of the top hits showed that 61% of the mapped sequences had similarities higher than 80%, while 37% of the hits had similarities ranging from 40% to 80%. (Fig 2A). 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 2B). The species distribution of NR BLAST matches showed that the top two species was Maylandia zebra (26%) and Oreochromis niloticus (21%) (Fig 2C).

GO and KEGG analysis of Collichthys lucidus transcriptomes

The roles of unigenes were investigated by GO and eggNOG analysis. In GO database, 27,708 unigenes were clustered into 67 terms of three different domains (biological process (BP), cellular component (CC) and molecular function (MF). The number of unigenes in each term are exhibited in Fig 3A. 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). In CC, unigenes were mainly converged on cell (16,229), cell part (16,151) and membrane (11,815). In MF, the most prominent terms were binding (16,123), nucleic acid binding catalytic activity (10,504) and signal transducer activity (1,974). On the other hand, eggNOG analysis annotated 48,042 (36.63%) unigenes, and the result revealed that unigenes mainly contributed to signal transduction mechanisms. The top 3 terms were signal transduction mechanisms, general function prediction only and function unknow (Fig 3B). To further explore the biological functions of unigenes, KEGG database was mapped and 12,883 unigenes were distributed into 32 pathways, 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 4).

Differentially expressed unigenes

The expression pattern of each unigene in brain, liver, ovary and testis was assessed by expected
The number of Fragments Per Kilobase of transcript sequence per Millions base pairs sequenced (FPKM).

The result of examination of unigenes in each organ revealed that brain expressed the most unigenes (74.83%), while liver expressed the least (46.18%) (Fig 5 A). Additionally, a gene was identified as organ-specific when the FPKM was at least 10 in one organ whereas less than 1 in other organs. Under this standard, 2,143, 304, 333, 556 unigenes were found specifically expressing in brain, liver, ovary and testis, respectively. The most genes were specifically expressed in brain, followed by ovary and testis, according with our hypothesis that HPG play a dominant role in *Colichthys lucidus*.

The expression profiles were compared between each two organs. 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 upregulated in testis (Table 2). The differential genes of group ovary vs testis are shown in Fig 5B and the differential genes of other groups are shown in Additional file 2 (Fig S1).

**Potential roles of differentially expressed unigenes**

GO enrichment analysis were established for differentially expressed unigenes between groups and we focused on terms related to reproduction, sex determination and sexual development.

Differentially expressed unigenes between ovary and testis were primarily involved in sexual reproduction (GO: 0019953), development of primary sexual characteristics (GO: 0045137), sex differentiation (GO: 0007548), male gonad development (GO:0008584), female gamete generation (GO: 0007292), male gamete generation (GO: 0048232), development of primary male sexual characteristics (GO: 0046546), male sex differentiation (GO: 0046661) (Additional file 3).

Interestingly, these GO terms were also enriched in other organs, implying their high correlation with
reproduction and sex in *Collichthys lucidus*. For instance, sexual reproduction (GO: 0019953), multicellular organism reproduction (GO: 0032504), reproduction (GO: 0000003) were enriched in differentially expressed unigenes between brain and liver. Developmental process involved in reproduction (GO: 0003006), cellular process involved in reproduction in multicellular organism (BP, GO: 0022412) and sexual reproduction (GO: 0019953) were enriched in differentially expressed unigenes between brain and ovary. Developmental process involved in reproduction (GO:0003006), sexual reproduction (GO: 0019953) and cellular process involved in reproduction in multicellular organism (GO: 0022412) were enriched in differentially expressed unigenes between brain and testis.

Regulation of growth (GO: 0040008), Sexual reproduction (GO: 0019953) and multicellular organism (GO: 0022412) were enriched in differentially expressed unigenes between liver and ovary.

Multicellular organism reproduction (GO: 0032504), sexual reproduction (GO: 0019953) and cellular process involved in reproduction in multicellular organism (GO: 0022412) were enriched in differentially expressed unigenes between liver and testis.

Furthermore, through KEGG pathway enrichment analysis, several pathways were enriched in each group. To investigate pathways connected to reproduction and gonad development, we particularly focused on differentially expressed unigenes between ovary and testis. 165 pathways, including PPAR signaling pathway, ovarian steroidogenesis, oocyte meiosis, steroid hormone biosynthesis, apoptosis – multiple species and proximal tubule bicarbonate reclamation were significantly enriched in differentially expressed unigenes between ovary and testis, which might participate in the regulation of reproduction. The profile of group ovary vs testis is exhibited in Fig 5C and others are shown in Additional file 4 (Fig S2).

**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. 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), ROPN1L (rhophilin 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 3).

Identification of SSRs

Using MISA 1.0 (http://pgrc.ipkgsatersleben.de/misa/misa.html), 35,476 SSRs were predicted and 2,455 compound formations were composed. Of the 35,476 detected SSRs, the most affluent repeat motif was mono- and di-nucleotide repeats, constituting 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). In general, the most abundant repeat motif was poly-T (18.66%, 6617). Among di-nucleotide repeat motifs, (GT/TG) was the dominant type with frequency of (12.24%, 4342) and (CG/GC) was the least type which only occupied (0.09%, 31). These SSRs were valuable for further genetic and mechanism researches (Table 4).

SSR marker polymorphism

In this study, 10 primer pairs were randomly designed to validate the amplification and polymorphism in Collichthys lucidus collected from 7 different regions. Seven of the primer pairs successfully amplified fragments. Among the successful primer pairs, one pair exhibited polymorphisms among the seven colonial Collichthys lucidus (Figure 6 A and B). The polymorphic SSR marker was then used to perform genetic correlation analysis among the Collichthys lucidus groups. The UPGMA clustering produced a dendrogram (Figure 6 C) that separated the Collichthys lucidus into two main groups, from which ND completely separate from the others. Moreover, LYG and ZS formed a sub-cluster, while CM and DF formed a sub-cluster.

Validation of gene expression using qRT-
PCR

To further verify the expression of genes identified by transcriptome sequencing experimentally, 8 genes enriched in reproduction terms were validated by qRT-PCR analysis. In accordant with the RNA-seq, qRT-PCR showed *ROPN1L*, *KLHL10*, *ODF3L2*, *SYCP3* and *SRPK3-like* were specific expressed in testis whereas *BMP15* and *RGS14* were specific expressed in ovary (Fig 7).

Discussion

De novo has been constructed in various of fishes according to published articles. In salmonid fish species including Atlantic salmon, *Brown trout*, *Arctic charr* and European whitefish, 191,977,874, 190,239,319, 180,232,708 and 209,578,198 paired-end reads were obtained and assembling 235,515, 242,899, 200,760 and 209,920 transcripts, respectively[13]. However, little is known about the transcriptome of *Collichthys lucidus*, especially the genes involved in the reproduction and sex differentiation remain unclear. In this 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 samples of brain, liver, ovary and testis of *Collichthys lucidus*, assembled 189,181 transcripts and identified 131,168 unigenes. Moreover, HPG axis is the sex endocrine system of fish and plays an important role in fish reproduction via a cascade of hormones [14]. Transcriptome analysis of HPG axis from *Lateolabrax maculatus* showed 748, 349 and 319 unigenes were uniquely expressed in the brain, ovary and testis tissues, respectively, while 26,623 unigenes were commonly expressed in all HPG tissues [15]. Previous study has defined the overall gene expression of HPG axis in the transgenic fish with blocked gonad development, which displayed 9, 28, and 212 differentially expressed genes in hypothalamus, pituitary, and ovary, respectively [16]. In the present study, the vast majority of the organ-specific expression is found in brains, containing 2,143 unigenes, whereas, only 304, 333 and 556 in liver, ovary and testis of *Collichthys lucidus*, respectively. Moreover, 1288 genes were differently expressed between ovary
and testis. These data may provide a basis for understanding the molecular regulation of gonadal development, sex determination, and reproduction.

In view of that no obvious secondary sexual characteristic has been found in *Collichthys lucidus* and the mechanism of gender development is still not clear, profile of ovary and testis were compared to do help in subsequent researches. We noticed that hyaluronidase was an important component of fertilization in mammals for acrosomal reaction [17], and, HYAL, the gene encoding hyaluronidase, was exactly upregulated in testis. Furthermore, this gene was involved in glycosaminoglycan degradation pathway (ko00531), indicating that hyaluronidase may also contribute to fertilization in *Collichthys lucidus* through giving a hand in dissolving zona pellicuda though there are great differences in fertilization between mammal and fish. On the other hand, SYCP3, the gene encoding synaptonemal complex protein 3, which closely connect to synopsis in meiosis[18] was also upregulated in testis. According to published articles, this gene was examined to reflect the level of meiosis and spermatogenesis[19-21] in zebrafish, tilapia and catfish. Considering the *Collichthys lucidus* resource decline, this gene can be a biomarker in researches about the influence of environmental damage on *Collichthys lucidus* reproduction. Additionally, SYPC3 was involved in homologous recombination pathway (ko03440), meaning that it might a potential target in genetic engineering for reproduction and breeding of *Collichthys lucidus*.

Various genes that master sex determinants have been identified in vertebrates, most of which are known to perform basic functions during sexual development. However, the mechanism of sex determination in teleost fish is largely different from that in mammals and birds. Since their sex chromosomes are less differentiated and often recombine along the length of X and Y chromosomes.

In fishes without Y sex chromosomes, gonad development of female is triggered via Wnt4, and female differentiation is induced by *Fosl2, alpha, Rspo1* and *Ctnnb1*. While in fishes with Y chromosomes, the gene that determines sex is located on the Y chromosome, such as *dmy, gdf6y, sdy, amhy, gsdfy, sox3y*, and *amhr2*, are over-expressed during sex determination[19-21].
In the presence of Foxl2, SdY is transported to the nucleus, and interacts synergistically with Nr5a1 to block the activation of the cyp19a1a promoter in rainbow trout [23]. Therefore, SdY blocked the pathway of female differentiation, thus enabling testicular differentiation. Recently, a novel sex determination mechanism was revealed in teleost fish by comparing RNA-Seq between females and males of the channel catfish [24], which exclusive male sex-specifically expressed BCAR1 gene during the early sex differentiation. Moreover, knockout of BCAR1 significantly converted genetic male (XY) into phenotypic female. At present, little is known about gender determination in perciformes, especially in the genus Collichthys. Vasa RNA is only detected in germ cells from the gonads, and its expression reflects crucial stages of oogenesis in a key stage of oocytogenesis in Collichthys lucidus [25].

The present study displayed that a class of differently expressed genes between ovary and testis were involved in functions of development of primary sexual characteristics, sex differentiation, male gonad development, female gamete generation, male gamete generation, development of primary male sexual characteristics, and male sex differentiation, indicating theses differently expressed genes may contribute to sex-determination of Collichthys lucidus.

Conclusion

The studies of genetic information, molecular regulation and reproduction of Collichthys lucidus would be expand due to the sequencing and de novo assembling on the transcripts of major organs. The identification of candidate genes involved in reproduction will be useful for protecting the resource of Collichthys lucidus. In addition, HYAL and SYCP3 may play dominant roles in the regulation of reproduction and are valuable in further research.

Abbreviations

SSRs: simple sequence repeats
HPG: Hypothalamic-pituitary-gonadal axis
MCP: monocrotophos pesticide
SMC: semicarbazide
THs: thyroid hormones
NR: NCBI nonredundant protein sequences
GO: Gene Ontology
KEGG: Kyoto Encyclopedia of Genes and Genome
Declarations

Ethics approval and consent to participate
All procedures conducted with the fishes were performed following the relevant guidelines and regulations. This study was approved by the Ethics Committee of Chinese Academy of Sciences. No associated permit number was required since commercial fishes sampling was approved.

Consent for publication
All the authors listed have read and approved the manuscript for publication.

Availability of data and material
All data can be obtained by email to the corresponding author.

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

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.

Authors' contributions
Conceptualization and funding: JF G; Data curation: W S, YB Z, and XJ Z; Investigation: W S, YB Z, and XJ Z; Preparation of original draft: W S. All authors have read and approved the manuscript.

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

Reference
1. Li H, Liu D, Li C, Zhou S, Tian D, Xiao D, Zhang H, Gao F, Huang J. Exosomes secreted from mutant-HIF-1α-modified bone-marrow-derived mesenchymal stem cells attenuate early steroid-induced avascular necrosis of femoral head in rabbit. *Cell Biol Int.* 2017; 41(12):1379-1390.
2. Cheng J, Ma GQ, Miao ZQ, Shui BN, Gao TX. Complete mitochondrial genome sequence of the spinyhead croaker *Collichthys lucidus* (Perciformes, Sciaenidae) with phylogenetic considerations. *Mol Biol Rep.* 2012; 39(4):4249-4259.
3. LM H, J L, YJ X, YZ Z. Study of the *Collichthys lucidus* fisheries resources of the Minjiang Estuary and its adjacent waters. *J Oceanogr Taiwan Strait.* 2010; 2:250-256.
4. 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 ERalpha but not ERbeta. *Mol Endocrinol.* 2003; 17(6):1039-1053.
5. Ma Y, Han J, Guo Y, Lam PK, Wu RS, Giesy JP, Zhang X, Zhou B. 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.
6. Tovo-Neto A, da Silva Rodrigues M, Habibi HR, Nobrega RH. Thyroid hormone actions on male reproductive system of teleost fish. *General and comparative endocrinology.* 2018; 265:230-236.
7. 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.
8. 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.
9. Hou LP, Chen H, Tian CE, Shi WJ, Liang Y, Wu RR, Fang XW, Zhang CP, Liang YQ, Xie L. 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 Dario Renio. *Aquat Toxicol.* 2018; 201:31-39.

10. Mortazavi A, Williams BA, McCue K, Schaeffer L, Wold B. Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nature methods.* 2008; 5(7):621-628.

11. Lu T, Lu G, Fan D, Zhu C, Li W, Zhao Q, Feng Q, Zhao Y, Guo Y, Huang X et al. Function annotation of the rice transcriptome at single-nucleotide resolution by RNA-seq. *Genome research.* 2010; 20(9):1238-1249.

12. Iorizzo M, Senalik DA, Grzebelus D, Bowman M, Cavagnaro PF, Matvienko M, Ashrafi H, Van Deynze A, Simon PW. De novo assembly and characterization of the carrot transcriptome reveals novel genes, new markers, and genetic diversity. *BMC genomics.* 2011; 12:389.

13. Carruthers M, Yurchenko AA, Augley JJ, Adams CE, Herzyk P, Elmer KR. De novo transcriptome assembly, annotation and comparison of four ecological and evolutionary model salmonid fish species. *BMC genomics.* 2018; 19(1):32.

14. Gui JF, Zhu ZY. Molecular basis and genetic improvement of economically important traits in aquaculture animals. *Chinese Science Bulletin.* 2012; 57(15):1751-1760.

15. Wang ZP, Wang D, Wang CL, Xie WJ, Zhu YF, Chen XW. Transcriptome characterization of HPG axis from Chinese sea perch Lateolabrax maculatus. *Journal of fish biology.* 2017; 91(5):1407-1418.

16. Xu J, Huang W, Zhong C, Luo D, Li S, Zhu Z, Hu W. 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):e21057.

17. Yoon S, Chang KT, Cho H, Moon J, Kim JS, Min SH, Koo DB, Lee SR, Kim SH, Park KE 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. *Animal reproduction science.* 2014; 150(3-4):107-114.

18. Syrjanen JL, Pellegrini L, Davies OR. A molecular model for the role of SYCP3 in meiotic chromosome organisation. *eLife.* 2014; 3.

19. Lin Q, Mei J, Li Z, Zhang X, Zhou L, Gui J. Distinct and Cooperative Roles of and in Self-Renewal
and Differentiation of Male Germ Cells in Zebrafish. *Genetics*. 2017; 207(3):1007-1022.

20. Fang X, Wu L, Yang L, Song L, Cai J, Luo F, Wei J, Zhou L, Wang D. Nuclear progestin receptor (Pgr) knockouts resulted in subfertility in male tilapia (*Oreochromis niloticus*). *J Steroid Biochem Mol Bio*. 2018; 182:62-71.

21. 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.

22. Li XY, Gui JF. Diverse and variable sex determination mechanisms in vertebrates. *Science China Life sciences*. 2018; 61(12):1503-1514.

23. Li XY, Gui JF. An epigenetic regulatory switch controlling temperature-dependent sex determination in vertebrates. *Science China Life sciences*. 2018; 61(8):996-998.

24. Bertho S, Herpin A, Branthonne A, Jouanno E, Yano A, Nicol B, Muller T, Pannetier M, Pailhoux E, Miwa M et al. The unusual rainbow trout sex determination gene hijacked the canonical vertebrate gonadal differentiation pathway. *Proceedings of the National Academy of Sciences of the United States of America*. 2018; 115(50):12781-12786.

25. Bao L, Tian C, Liu S, Zhang Y, Elaswad A, Yuan Z, Khalil K, Sun F, Yang Y, Zhou T et al. The Y chromosome sequence of the channel catfish suggests novel sex determination mechanisms in teleost fish. *BMC biology*. 2019; 17(1):6.

26. Li Y, Song W, Zhu YF, Zhu TY, Ma LB, Li MY. Evolutionarily conserved vasa identifies embryonic and gonadal germ cells in spinyhead croaker *Collichthys lucidus*. *Journal of fish biology*. 2019.

Tables

| Assembly                   | Number of sequences |
|----------------------------|---------------------|
| Total number of unigenes   | 131,168             |
| Average unigene length (bp)| 644                 |
| N50 (bp)                   | 1033                |
| Annotation                 | Number of matched unigenes (percentages) |
| Transcript BLASTx against NR | 51,590 (39.33%)     |
| Transcript BLASTx against GO | 27,708 (21.12%)     |
| Transcript BLASTx against KEGG | 12,883 (9.82%)     |
| Transcript BLASTx against Swissport | 44,051 (33.58%) |
| All annotated transcripts  | 53,200 (40.56%)     |
| Transcripts identified in all four databases | 8851 (6.75%) |
Table 2 unigenes differently expressed in each two organs

| Case | control | Up-regulated unigenes | Down-regulated unigenes | Total DE unigenes |
|------|---------|------------------------|--------------------------|------------------|
| B    | L       | 992                    | 264                      | 1256             |
| B    | O       | 749                    | 317                      | 1066             |
| B    | T       | 685                    | 388                      | 1073             |
| L    | O       | 211                    | 114                      | 325              |
| L    | T       | 138                    | 315                      | 453              |
| O    | T       | 442                    | 846                      | 1288             |

Table 3 different genes in GO terms related to reproduction

| id          | O    | T    | foldChange(T/O) | log2FoldChange | pval  |
|-------------|------|------|----------------|----------------|-------|
| c35502_g1   | 0    | 1950 | LnF            | Inf            | 0.003 |
| c55252_g1   | 2    | 1765 | 882.5          | 9.79           | 0.008 |
| c55445_g1   | 9245 | 20   | 0.002          | -8.85          | 0.015 |
| c28789_g1   | 15   | 4292 | 286.13         | 8.16           | 0.019 |
| c47852_g1   | 0    | 217  | Inf            | Inf            | 0.023 |
| c65782_g1   | 53   | 20287| 382.77         | 8.58           | 0.024 |
| c66108_g1   | 19878| 48   | 0.003          | -8.18          | 0.026 |
| c54644_g1   | 0    | 157  | Inf            | Inf            | 0.028 |

Table 4 SSRs identified in *Collichthys. Lucidus*

| item                              | number |
|-----------------------------------|--------|
| Total number of SSRs              | 34,476 |
| Number of mono-nucleotide repeats | 14,608 |
| Number of di-nucleotide repeats   | 13,563 |
| Number of tri-nucleotide repeats  | 6,648  |
| Number of tetra-nucleotide repeats| 614    |
| Number of penta-nucleotide repeats| 31     |
| Number of hexa-nucleotide repeats | 12     |

Additional File Legend

Additional file 1: Information of primers used in this study.

Additional file 2: Fig S1. Volcano plot shows differential genes between Brain (B), liver (L), ovary (O) and testis (T) tissues.

Additional file 3: GO terms of differently expressed genes that were related to reproduction.

Additional file 4: Fig S2. KEGG pathway enrichment analysis of differently expressed genes between
Brain (B), liver (L), ovary (O) and testis (T) tissues.

Figures
Figure 1
Length distribution of transcripts. (A) The length distribution of all transcripts showed most transcripts are 200 to 299 bp. (B) The length distribution of all transcripts of unigenes showed most unigenes are 200 to 299 bp.

Figure 2
Transcripts BLAST against NR. (A) The similarity of mapped sequences with genes in NR database. (B) The E-value distribution had a comparable pattern with 51% of the mapped sequences with high homologies (<1e-60), whereas 48% of the homologied sequences ranged between 1e-5 to 1e-60. (C) The species distribution showed that Maylandia zebra and Oreochromis niloticus was the top two species.
Functional analysis of unigenes. (A) Transcripts BLAST against GO. 27,708 unigenes were clustered into 67 terms of three different domains (red columns represent biological process (BP), green columns represent cellular component (CC) and blue columns represent molecular function (MF)) (B) Transcripts BLAST against eggNRG. 48,042 unigenes were annotated, and unigenes mainly contributed to signal transduction mechanisms.
Transcripts BLAST against KEGG. 12,883 unigenes were distributed into 32 pathways. 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.
Figure 5

Analysis of differentially expressed unigenes. (A) Cluster heatmap of differentially expressed unigenes between any two tissues of brain, liver, ovary and testis. The expression profile of genes in different organs. Brain expressed the most unigenes (74.83%), while liver expressed the least (46.18%). (B) The differential genes of group ovary vs T. Between ovary and testis, 442 genes were upregulated in ovary and 846 genes were upregulated in testis. (C) The pathways enrichment profile of group ovary vs T. Nitrogen metabolism, steroid hormone biosynthesis, apoptosis – multiple species and proximal tubule bicarbonate reclamation, were significantly enriched which might participate in the regulation of reproduction.
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 *Collichthys Lucidus* from seven different locations. LYG: lianyungang, ZS: zhoushan, CM: chongming; WZ: wenzhou, XM: xiamen, ND: ningde, DF: dafeng
Genes involved in the reproduction were verified by qRT-PCR. **p < 0.01. compared with the other tissues (t-test).

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- Additional file 3.docx
- Additional file 1 .docx
- Additional file 4 Fig S2.tif
- NC3Rs ARRIVE Guidelines Checklist (fillable).pdf
- Additional file 2 Fig S1.tif