Identification, characterization and functional analysis of gonadal long noncoding RNAs in a protogynous hermaphroditic teleost fish, the ricefield eel (*Monopterus albus*)

Zhi He1†, Lijuan Ye1†, Deying Yang1†, Zhijun Ma1, Faqiang Deng1, Zhide He2, Jiuxiang Hu3, Hongjun Chen1, Li Zheng1, Yong Pu1, Yuanyuan Jiao1, Qiqi Chen1, Kuo Gao1, Jinxin Xiong1, Bolin Lai1, Xiaobin Gu4, Xiaoli Huang1, Shiyong Yang1, Mingwang Zhang1 and Taiming Yan1*

Abstract

**Background:** An increasing number of long noncoding RNAs (lncRNAs) have been found to play important roles in sex differentiation and gonad development by regulating gene expression at the epigenetic, transcriptional and posttranscriptional levels. The ricefield eel, *Monopterus albus*, is a protogynous hermaphroditic fish that undergoes a sequential sex change from female to male. However, the roles of lncRNA in the sex change is unclear.

**Results:** Herein, we performed RNA sequencing to analyse lncRNA expression patterns in five different stages of *M. albus* development to investigate the roles of lncRNAs in the sex change process. A total of 12,746 lncRNAs (1503 known lncRNAs and 11,243 new lncRNAs) and 2901 differentially expressed lncRNAs (DE-lncRNAs) were identified in the gonads. The target genes of the DE-lncRNAs included *foxo1*, *foxm1*, *smad3*, *foxr1*, *camk4*, *ar* and *tgfb3*, which were mainly enriched in signalling pathways related to gonadal development, such as the insulin signalling pathway, MAPK signalling pathway, and calcium signalling pathway. We selected 5 highly expressed DE-lncRNAs (LOC109952131, LOC109954337, LOC109954360 and LOC109958454) for full length amplification and expression pattern verification. They were all expressed at higher levels in ovaries and intersex gonads than in testes, and exhibited specific time-dependent expression in ovarian tissue incubated with follicle-stimulating hormone (FSH) and human chorionic gonadotropin (hCG). The results of quantitative real-time PCR (qRT-PCR) analysis and a dual-luciferase assay showed that *znf207*, as the gene targeted by LOC109958454, was expressed in multiple tissues and gonadal developmental stages of *M. albus*, and its expression was also inhibited by the hormones FSH and hCG.

© The Author(s) 2022. Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/. The Creative Commons Public Domain Dedication waiver (http://creativecommons.org/publicdomain/zero/1.0/) applies to the data made available in this article, unless otherwise stated in a credit line to the data.
Conclusions: These results provide new insights into the role of lncRNAs in gonadal development, especially regarding natural sex changes in fish, which will be useful for enhancing our understanding of sequential hermaphroditism and sex changes in the ricefield eel (M. albus) and other teleosts.

Keywords: RNA-seq, Monopterus albus, Gonadal development, lncRNA

Background
Long noncoding RNAs (lncRNAs) are transcript with a total length of more than 200 nucleotides that are little or no significant protein coding capacity [1, 2]. Compared with mRNAs, lncRNAs are more abundant in the nucleus [3–5]. They can exert their biological functions by acting on protein-coding genes in cis and trans ways [6–9]. Many studies have shown that lncRNAs are involved in several biological processes, including germ cell development [10], meiosis [11], gamete formation [12, 13] and sex differentiation [14].

LncRNAs have been found to play important roles in the reproductive development of organisms. The IncRNAs TCONS_00025195 and TCONS_00025196 may be involved in regulating the expression of SOX9 in human testes [15]. Most of the highly expressed lncRNAs in the testes of adult mice were found to be testis specific and were expressed at the beginning of and after spermatogenesis [16]. A total of 24,601 lncRNAs were identified in ovaries of Columba livia, and 148 lncRNAs were significantly differentially expressed between high and low egg production performance groups, and the target gene FOXX2 of the lncRNA MSTRG.7894.4 is involved in the regulation of an oestrogen receptor [17]. We identified 5317 IncRNAs and 297 differentially expressed IncRNAs (DE-IncRNAs) in Oryzias latipes ovaries and testes, of which IncRNA MSTRG.14827.1 may play a key role in gametogenesis [18]. In total, 28,500 IncRNAs were identified in the ovaries and testes of Pelodiscus sinensis, and the target genes for the 10,495 DE-lncRNAs included several genes involved in gonadal development [19]. In addition, 962 DE-IncRNAs related to sex differentiation were found to be involved in the sex reversal induced by mifepristone (RU486) in XX Oreochromis niloticus gonads [20].

The ricefield eel, Monopterus albus, is increasingly recognized as a new model species for the study of sequential hermaphroditism and sex change [21–23]. M. albus is a hermaphroditic fish with reproductive traits that natural sex reversal from female to male through an intersex stage during its life cycle [22, 24]. To comprehensively reveal the molecular mechanisms of the sex change process and the potential role of IncRNAs in M. albus, RNA sequencing (RNA-seq) of gonadal tissues representing five sexual stages (ovary stage (OV), early intersexual stage (IE), middle intersexual stage (IM), late intersexual stage (IL) and testis stage (TE)) was performed in the present study. The functional IncRNAs, target genes and their correlations associated with both sex determination and gonad development were explored. This study provides novel insights into IncRNA function and the mechanisms of sequential hermaphroditism in fish and other vertebrates.

Results
Identification and characterization of gonadal IncRNAs
To study the potential role of lncRNAs in fish gonads, we used M. albus gonads at five developmental stages (OV, IE, IM, IL, TE) for RNA-seq and cDNA libraries. At least 101,886,556 clean reads were obtained. The Q20 percentage > 97.51%, Q30 percentage > 93.68%, GC content between 47.63–48.71% and percentage of error < 0.01% (Additional file 1: Table S1). These values indicated that the RNA-seq data were reliable.

In this study, a total of 1503 known IncRNAs and 11,243 new IncRNAs (Additional file 2: Fig. S1) were identified in libraries constructed from M. albus gonads. The whole experimental process is shown in Fig. 1A. After each step of screening, the final proportion of different types of IncRNAs (Fig. 1B), including lincRNAs (41.1%), antisense IncRNAs (21.4%), and intronic IncRNAs (37.5%), was determined. The ORF length and average FPKM of IncRNAs were smaller than those of mRNAs (Fig. 1C–D). By comparing the box plots of the quantitative results for different samples (Fig. 1E), the overall expression level in the male stage among different samples was found to be higher than that in the female/intersex stages (P < 0.05).

The DE-IncRNAs (P < 0.05) were counted according to the gonadal development of M. albus (Fig. 2, Additional file 3: Dataset S1). The aggregation of DE-IncRNAs at different gonadal stages indicated that ovarian and intersex expression clustered into one branch, and testis into one branch (Fig. 2A). Venn diagram analysis showed that there was a total of 2901 DE-IncRNAs, and 5 IncRNAs were differentially expressed throughout gonadal development (Fig. 2B). In OV vs. IE, 189 IncRNAs were up regulated and 128 IncRNAs were down regulated (Fig. 2C); in IE vs. IM, 219 IncRNAs were up regulated and 183 IncRNAs were down regulated (Fig. 2D); in IM vs. IL, 181 IncRNAs were up regulated and 299 IncRNAs were down regulated (Fig. 2E); and in IL vs. TE, 398 IncRNAs were up regulated and 1803 IncRNAs were down regulated.
Fig. 1 Global analysis of M. albus gonadal lncRNAs. A Experimental design diagram. B LncRNA type distribution map. C LncRNA and mRNA ORF length density map. D Expression levels of lncRNA and mRNA transcripts depicted as a violin plot. E Comparative analysis of sample expression levels. OV: ovary stage; IE: early intersexual stage; IM: middle intersexual stage; IL: late intersexual stage; TE: testis stage
regulated (Fig. 2F). The results showed that the number of DE-lncRNAs gradually increased from the female to intersex to male stages, especially the number of DE-lncRNAs from IL to TE.

**GO and KEGG analyses of lncRNA target genes**
In this study, we predicted the trans-target genes of lncRNAs, predicted the potential function of lncRNAs, and performed Gene Ontology (GO) cluster analysis (Table 1, Additional file 4: Fig. S2, Additional file 5: Dataset S2,
Additional file 6: Dataset S3) and Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichment analysis (Table 2, Additional file 7: Fig. S3, Additional file 8: Dataset S4, Additional file 6: Dataset S3). The significantly enriched (q-value<0.05) GO terms were involved in regulation of cellular processes, regulation of biological processes and cellular components, and some trans-target genes of the IncRNAs were related to gonadal development, such as *foxo1*, *foxm1*, *smad3*, *foxr1*, *camk4* and *ar*. Genes that were significantly enriched in several KEGG pathways (P<0.05) were identified as being involved in gonadal development, including XLOC_059262 and its target *foxo1* in the insulin signalling pathway, XLOC_091218 and its target *tgfb3* in the MAPK signalling pathway, and XLOC_169543 and its target *camk4* in the calcium signalling pathway. Significantly, XLOC_181771 and its target *smad3* and XLOC_091218

### Table 1 DE-IncRNAs and their targets among the significantly enriched GO terms

| Comparison | DE-IncRNA (number) | Trans-targets | GO terms                                      | Term type  |
|------------|-------------------|----------------|-----------------------------------------------|------------|
| OV vs. IE  | XLOC_059262(45)   | *foxo1*        | Regulation of biological process              | BP         |
|            | XLOC_138801(29)   | *foxm1*        | Regulation of cellular process                | BP         |
| IE vs. IM  | XLOC_181771(35)   | *smad3*        | Regulation of cellular process                | BP         |
|            | XLOC_138801(29)   | *foxm1*        | Regulation of biological process              | BP         |
|            | XLOC_064006(2)    | *foxr1*        | Regulation of biological process              | BP         |
| IM vs. IL  | XLOC_059262(45)   | *foxo1*        | Regulation of biological process              | BP         |
|            | XLOC_181771(35)   | *smad3*        | Regulation of biological process              | BP         |
|            | XLOC_138801(29)   | *foxm1*        | Regulation of cellular process                | BP         |
|            | XLOC_169543(38)   | *camk4*        | Regulation of cellular process                | BP         |
| IL vs. TE  | XLOC_059262(45)   | *foxo1*        | Regulation of cellular process                | BP         |
|            | XLOC_181771(35)   | *smad3*        | Regulation of cellular process                | BP         |
|            | XLOC_138801(29)   | *foxm1*        | Regulation of biological process              | BP         |
|            | XLOC_169543(38)   | *camk4*        | Regulation of cellular process                | BP         |
|            | 109,964,303 (3)   | *ar*           | Regulation of cellular process                | BP         |

Note: IncRNA (number): the listed IncRNA Gene IDs are those with the smallest P-values, and numbers in parentheses are the numbers of IncRNAs; see Schedule Data 4 for details. OV Ovary stage, IE Early intersexual stage, IM Middle intersexual stage, IL Late intersexual stage, TE Testis stage. BP Biological process, CC Cellular component, MF Molecular function

### Table 2 DE-IncRNAs and their targets among the significantly enriched KEGG pathways

| Comparison | IncRNA (number) | Trans-targets | KEGG                          |
|------------|----------------|---------------|-------------------------------|
| OV vs. IE  | XLOC_059262(45)| *foxo1*       | Insulin signaling pathway     |
|            | XLOC_091218(40)| *tgfb3*       | MAPK signaling pathway        |
|            | XLOC_169543(38)| *camk4*       | Calcium signaling pathway     |
| IE vs. IM  | XLOC_059262(45)| *foxo1*       | Endocytosis                   |
|            | XLOC_091218(40)| *tgfb3*       | Endocytosis                   |

Note: IncRNA (number): the listed IncRNA Gene IDs are those with the smallest P-values, and numbers in parentheses are the numbers of IncRNAs; see Schedule Data 4 for details. OV Ovary stage, IE Early intersexual stage, IM Middle intersexual stage, IL Late intersexual stage, TE Testis stage.
and its target \textit{tgfb3} were enriched in endocytosis pathways.

\section*{Validation of DE-IncRNAs}
Through PCR technology, we successfully amplified 5 highly expressed DE-IncRNAs, namely, \textit{LOC109952131} (450 bp), \textit{LOC109953466} (320 bp), \textit{LOC109954337} (239 bp), \textit{LOC109954360} (424 bp) and \textit{LOC109958454} (454 bp). In addition, we showed the location of the IncRNAs on the \textit{M. albus} chromosome and their exon number (Fig. 3A).

To verify the authenticity of the RNA-seq data, we conducted qRT-PCR analysis to determine the expression levels of 5 randomly selected DE-IncRNAs in the gonads at the 5 developmental stages. The IncRNA expression trend obtained by qRT-PCR was consistent with the RNA-seq results, confirming that the RNA-seq data were reliable (Fig. 3 B-F). Five DE-IncRNAs were expressed at significantly lower levels in the male stage than in the female and intersex stages ($P < 0.05$). The significant difference in expression between the male and female/intersex stages further indicated that IncRNAs may regulate the gonadal development of \textit{M. albus}.

\section*{Expression of five DE-IncRNAs in the ovary after incubation with FSH and hCG}
Incubation with FSH increased IncRNA expression in a time-dependent manner, and the overall trend was a decrease followed by an increase (Fig. 4E, G and I). Except for \textit{LOC109952131} (Fig. 4A) and \textit{LOC109953466} (Fig. 4C) incubated at 1 ng/mL and 5 ng/mL, all the samples showed the highest expression level at 1 h of incubation. Incubation with hCG increased IncRNA expression in a time-dependent manner, and the overall trend was an increase followed by a decrease (Fig. 4B, D and H). Except for \textit{LOC109954337} (Fig. 4F) incubated at 10.01U/mL and \textit{LOC109954454} (Fig. 4I) incubated at 10.01U/mL, all the samples showed the highest expression levels at 2 h of incubation.

\section*{Identification of the \textit{LOC109958454} target gene \textit{znf207}}
As shown in Fig. 5A, \textit{LOC109958454} is 11.108 kb away from its colocalized mRNA \textit{znf207}. The results of tissue distribution experiments showed that \textit{znf207} was widely distributed in gonadal tissues (ovary and testis) and nongonadal tissues (eyes, pituitary gland, heart, etc.) of \textit{M. albus} (Fig. 5B). In contrast to \textit{LOC109958454}, the expression of \textit{znf207} continued to increase, and the expression in the male stage was higher than that in the female and intersex stages during the gonadal development process (Fig. 5C). The biological relationship of IncRNAs is often related to their subcellular location. The results showed that \textit{LOC109958454} and \textit{znf207} were specifically expressed in the nucleus (Fig. 5D). The results of the dual-luciferase assay showed that \textit{LOC109958454} could inhibit the expression of \textit{znf207}, and this inhibition disappeared or was weakened when the binding site was mutated (Fig. 5E). The expression of \textit{znf207} at each time point after incubation with FSH and hCG divided by that in the control group (0 h) was mostly less than 1. FSH and hCG inhibited the expression of \textit{znf207} (Fig. 5F-G). In short, \textit{znf207} is the target gene of \textit{LOC109958454}, and its expression level is affected by hormones.

\section*{Discussion}
\subsection*{Function of IncRNAs in gonadal development}
Transcriptome sequencing of Meishan pig and Yorkshire ovarian tissue revealed 3827 IncRNAs and 192 DE-IncRNAs [25]. In Duroc pigs ovaries on days 0, 2, and 4 of follicular development, 140 IncRNAs were found to be differentially expressed in pairwise comparisons [26]. There were 115 DE-IncRNAs between the goat ovarian follicular phase and luteal phase [27]. A total of 1118 DE-IncRNAs were found based on the comparison between the testes of 3 month old and 9 month old sheep [28]. In our research, the number of DE-IncRNAs corresponding to OV vs. IE, IE vs. IM, IM vs. IL and IL vs. TE was 317, 402, 408, and 2201, respectively. Analysis of the results of IncRNA identification in different species and the differential expression at different developmental stages suggested that IncRNAs also play roles in gonadal development in \textit{M. albus}.

IncRNAs have a wide range of functions in animal reproduction and development, including in germ cell growth, meiosis, gametogenesis, sex hormone response, and sex determination [29–31]. A recent study provided the first evidence that the IncRNA \textit{H19} is associated with polycystic ovary syndrome (PCOS) in women and that increased levels of IncRNA \textit{H19} are a risk factor for PCOS [32]. The expression of IncRNAs (such as \textit{XR_001917388.1}) in goat ovaries during the follicular phase is higher than that in the luteal phase and may regulate oogenesis and oocyte maturation [27]. The IncRNA \textit{TCONS_00030774} is located approximately 500bp downstream of the estrogen receptor 1 (\textit{ESR1}), which indicates that it plays a role in porcine ovarian development [25]. The IncRNAs \textit{1700108J01Rik} and \textit{170010I022Rik}, both of which are mouse testis specific IncRNAs, are specifically expressed in the testicular germ cells of premeiosis and round spermatids, which coincides with the reactivation of transcription during spermatogenesis [16]. The IncRNA \textit{NONMMUT074098.2} (\textit{Lnc10}) can cause abnormal spermatogenesis by promoting germ cell apoptosis. After \textit{Lnc10} was knocked down, the morphology of the testis changes, and the average quality decreases [33]. In addition, IncRNAs can regulate
Fig. 3  Verification of 5 DE-lncRNAs. A: The location of lncRNAs on the M. albus chromosome. B-F: Expression of 5 DE-lncRNAs in the gonads.
qRT-PCR: real-time fluorescence quantification; RNA-seq: RNA sequencing. EN: exon number. OV: ovary stage; IE: early intersexual stage; IM: middle intersexual stage; IL: late intersexual stage; TE: testis stage. The results are presented as the means ± SEMs (n = 3). Means marked with different letters are significantly different (P < 0.05).

Fig. 4  In vitro regulatory effect of hCG and FSH on the expression of DE-lncRNAs in the ovaries of M. albus. A, B LOC1099952131 expression patterns in ovaries after incubation with FSH and hCG. C, D LOC1099953466 expression pattern in ovaries after incubation with FSH and hCG. E, F LOC1099954337 expression pattern in ovaries after incubation with FSH and hCG. G, H LOC1099954360 expression pattern in ovaries after incubation with FSH and hCG. I, J LOC1099958454 expression pattern in ovaries after incubation with FSH and hCG. The results are presented as the means ± SEMs (n = 3). Means marked with different letters are significantly different (P < 0.05)
Fig. 4 (See legend on previous page.)
Fig. 5  Expression level of the LOC109958454 cis-target gene znf207. A Positions of LOC109958454 and znf207 on the chromosome. B, C Znf207 expression pattern analysis. D LOC109958454 and znf207 subcellular localization. E LOC109958454 and znf207 targeting verification. F, G In vitro regulatory effect of hCG and FSH on the expression of DE-lncRNAs in the gonads of M. albus. OV: ovary stage; IE: early intersexual stage; IM: middle intersexual stage; IL: late intersexual stage; TE: testis stage; PI: pituitary gland; EY: eye; HE: heart; LI: liver; KI: kidney; IN: intestines; SP: spleen; MU: muscle; Bl: blood. The results are presented as the means ± SEMs (n = 3). Means marked with different letters are significantly different (P < 0.05).
the gonadal development of _Mauremys mutica_ [14], _P. sinensis_ [34] and _Scylla paramamosain_ [35] and affect sex differentiation. In our study, the cis-target genes of the lncRNAs included a large number of genes involved in gonadal development, including _igf1_ [36], _sox9_ [37], _gdf9_ [38], and _smad2_ [24] (Additional file 9: Dataset S5). Therefore, lncRNAs may play roles in the gonadal development process in _M. albus_, and the specific mechanism needs to be further studied.

**DE-lncRNA target genes and their GO and KEGG analyses**

A total of 12,746 lncRNAs were identified from the gonads of _M. albus_ by RNA-seq, and 2901 of the lncRNAs were differentially expressed. In this study, some of the predicted trans-target genes were related to gonadal development, such as _foxo1, foxm1, smad3, foxr1, camk4, ar_ and _tgfb3_ (Table 1 and Table 2). The _XLOC_059262 target gene _foxo1, XLOC_138801 target gene _foxm1_ and _XLOC_064006 target gene _foxr1_ are all fox genes, which play important roles in the sex determination and gonadal development of bony fish [39]. The target gene of _XLOC_181771, _smad3_, is involved in endocytosis and is also a member of the TGF-β family. That the gene _smad3_ may be the main signal transduction molecule that, via activin, stimulates the expression of FSHβ in goldfish [40]. The target gene _camk4_ of _XLOC_169543 in the calcium signalling pathway is expressed at high levels. In mice lacking _camk4_, spermatogenesis and basic nucleoprotein exchange are impaired, leading to male sterility [41]. Similarly, female fertility in _camk4_ deficient muscle groups is significantly reduced due to follicular development and ovulation damage [42]. _ar_ is an important receptor in the supporting cells of the testis. It plays an important role in maintaining the number of spermato- gonia and the integrity of the blood testis barrier, in the completion of meiosis, in the adhesion of sperm cells and in the fertilization of sperm and other spermatogenesis processes [43, 44]. In this study, _ar_ was enriched in only IL vs. TE, indicating that it plays an important role in the development of the testis during the sex reversal process of _M. albus_. Mutation of the _XLOC_003298 target gene _tgfb3_ may be related to male sterility [45]. In addition, _tgfb3_ has a key role related to meiotic arrest and the development of oocytes [46, 47].

The predicted target genes of the DE-lncRNAs were analysed by GO and KEGG cluster analyses. In our research, the insulin signalling pathway, MAPK signalling pathway, and calcium signalling pathway were found to be enriched. The lncRNA GAS5 may regulate the expression of PARP1 by recruiting the transcription factor E2F4 to its promoter, affecting the activity of the MAPK signalling pathway, promoting apoptosis and causing G0/G1 arrest in ovarian cancer cells [48]. The results in zebrafish ( _Danio rerio_ ) show that Corticotropin-releasing hormone (CRH) and its receptor inhibit estrogen production and synthesis, and the inhibition of CRHr is partly mediated through the p38 MAPK signalling pathway [49]. Injection of IRS (an insulin signaling pathway component) dsRNA into adult female oriental fruit fly significantly reduced IRS transcript levels, thereby inhibiting ovarian development, and the average size of the ovaries was reduced by 33% compared to controls [50]. A calcium signaling pathway, initiated by internal stores, is required for acute LH-induced steroidogenesis [51]. And calcium is also an important component of the signal transduction pathway, which is an important regulatory component of meiosis during oogenesis in mouse [52]. In our previous study, insulin, MAPK and calcium signalling pathway were involved in the process of _M. albus_ cell apoptosis [53]. In this study, these pathways were also enriched in the whole process of sexual reversal of _M. albus_, which suggested that it may be involved in the whole process of gonadal development, especially in the process of ovarian development and apoptosis. Receptor-mediated endocytosis uptake of multiple ligands, such as hormones, growth factors, and transport molecules, is required for gonadal development [54]. In teleost fish, vitellogenin (Vtg) is released into the blood circulation through Vtg receptor-mediated endocytosis, and combines with oocytes to form yolk granules, thereby partici- pating in fish ovary growth [55]. In our previous report, endocytosis was associated with oocyte apoptosis during the female-to-male sex transition of _M. albus_ [53]. In this study, endocytosis was enriched during the gonad development from OV to IE and from IL to TE, indicating that endocytosis may be involved in the process of oocyte development and sexual reversal of _M. albus_. The specific mechanism remains to be further explored.

**Effect of FSH and hCG on lncRNAs**

FSH and hCG levels can affect gonadal development [56, 57]. Studies have found that FSH can promote the growth of _M. albus_ oocytes, playing an important role in the transition from primary follicle growth to secondary follicle growth, which is an important step in oocyte maturation [58]. hCG promotes gonad maturation in _Trachinotus blochii_ [59] and _Anguilla japonica_ [60]. Interestingly, the expression levels of lncRNAs can affect the levels of hormones (such as FSH) and are regulated by hormones (such as hCG). For example, after interfering with the expression of the lncRNA _TCONS_00066406, the expression level of FSH is reduced, affecting female reproduction [61, 62]. Moreover, the lncRNA _HAS2-as1_ is an hCG target that promotes the expression of HAS2 and may play a role in regulating cumulus
expansion and migration [63]. In our previous study, FSH and hCG were used to promote the expression of the lncRNA LOC109960696 trans-target gene Smad2 in *M. albus* ovarian tissue in a time- and dose-dependent manner, respectively [24]. We successfully amplified 5 highly expressed DE-lncRNAs (LOC109952131, LOC109953466, LOC109954337, LOC109954360 and LOC109958454). The results showed that the 5 lncRNAs were expressed at higher levels in the ovary and intersex stage than in the testis. Incubation with FSH and hCG increased the expression of the 5 lncRNAs in a time-dependent manner. In summary, it was shown that the expression of lncRNAs is regulated by the levels of hormones, which in turn affects gonadal development in organisms.

**Potential effect of the LOC109958454 cis-target gene znf207 on gonadal development**

Many scholars have also explored the interactions between target genes and lncRNAs. In *C. livia*, RNA-seq analysis showed that FOXX2 was a target gene of the lncRNA MSTRG.7894.4 involved in the regulation of the oestrogen receptor, and qRT-PCR experiments showed that the mRNA expression level of the *Era* gene in the high egg production performance group was significantly higher than that in the low egg production performance group [17, 64]. In *S. paramamosain*, the expression level and complementary sequence of lncRNA-n-cr2 were found to be higher in the ovary than in the testis, and the expression levels of its target genes Sp-jcb2 and Sp-jcb3 in the ovary were significantly lower than those in the testis [35]. In *Fenneropenaeus merguiensis*, the expression of *IncPV13* was inhibited and the expression of vitellogenin (Vg) was significantly decreased in the ovary on day 7 after injection of *IncPV13*-specific double-stranded RNA (ds*IncPV13*). In contrast, the expression of gonad-inhibiting hormone (GIH) was significantly increased in *IncPV13*-knockout shrimp [65]. We found that the cis-target gene znf207 was located 11.108 kb downstream of *LOC109958454*. This gene was widely expressed in various tissues of *M. albus*, and its expression in the testis was higher than that in the intersex stages and ovaries.

Znf207 belongs to the zinc finger protein family, members of which are involved in a variety of biological processes, including chromatin remodelling, transcription activation/inhibition, DNA repair, apoptosis regulation, protein folding and assembly, stress response, and cell proliferation and differentiation [66, 67]. Nuclear lncRNAs are involved in a variety of biological processes, including chromatin organization, transcription and posttranscriptional gene expression, and serve as structural scaffolds for the nuclear domain [68]. Depletion of znf207 leads to an increase in the formation of RNA-DNA hybrids (R-loops), which leads to DNA damage and p53 activation in cells, which in turn leads to cell apoptosis or senescence [69]. Loss of znf207 gene expression is beneficial for the treatment of *tp53* mutant ovarian cancer [70]. In our study, LOC109958454 and znf207 were found to be expressed mainly in the nucleus and to have a targeting relationship; LOC109958454 inhibits the expression of znf207. In addition, FSH and hCG inhibited the expression of znf207. The lncRNA and hormones could regulate the expression of znf207; the specific mechanism remains to be further explored.

**Conclusion**

To the best of our knowledge, our study is the first systematic report of lncRNA sequencing for the five stages of gonadal development during sexual reversal of *M. albus*, and many candidate lncRNAs related to gonadal development were discovered. We showed that 5 lncRNAs had higher levels in the female and intersex stages and that their expression levels were affected by the hormones FSH and hCG. And *LOC109958454* and znf207 were expressed mainly in the nucleus rather than the cytoplasm, verifying their targeting relationship, and found that the expression of znf207 was suppressed by hormones. The above experiments showed that many lncRNAs are involved in the gonadal development of *M. albus* and may function through target genes. This study provides a new theoretical basis for the study of gonadal development in teleost fish and other animals.

**Methods**

**Sample collection and preparation**

Wild *M. albus* (*n* = 300) were purchased from a local market in Chengdu, Sichuan. These fish were kept under the natural temperature and photoperiod in the laboratory. All procedures and investigations used for animal research were subject to approval and performed in accordance with the guidelines of the ethics committee (Approval No.20190031).

Fish were anaesthetized with 0.02% tricaine buffer (80 μg/L) (Sigma, LA, USA) for 10 min after a 24 h fast, and the tissues, including half of the gonads, pituitary gland, eye, heart, kidney, intestines, spleen, muscle and blood, were collected and immediately stored in liquid nitrogen at −80 °C. The other half of the fresh gonads were immediately fixed in Bouin’s solution for 24 h and then stored in 75% ethanol until paraffin embedding.

Sections were serially cut at a thickness of 5 μm using a slicer (Leica, Nussloch, Germany) and stained with haematoxylin/eosin [24]. Through gonadal histological sectioning, gonadal samples from different stages of the *M. albus* sex reversal process were verified. With reference to the previous reports [24, 45, 71], according to the
proportion of male and female germ cell development, the transition from female to male sex was classified into five phases: ovary in mature stage (OV) (Fig. 6A), early intersexual stage gonad (IE) (Fig. 6B), middle intersexual stage gonad (IM) (Fig. 6C), late intersexual stage gonad (IL) (Fig. 6D), and testis (TE) (Fig. 6E).

**Sequencing method and quality control**
We used *M. albus* gonad samples for RNA-seq at the 5 developmental stages, with 3 biological replicates in each stage, and for identification and analysis of lncRNAs. Raw data in FASTQ format were first processed through in-house Perl scripts. In this step, clean data were obtained by removing reads containing adapters, reads containing poly-N sequences and low-quality reads from the raw data. At the same time, the Q20, Q30 and GC contents of the clean data were calculated. All subsequent analyses were based on clean data of high quality.

**LncRNA identification**
Cuffmerge was used to merge the transcripts spliced from each sample, and the transcripts with uncertain chain directions were removed to obtain the complete transcriptome information from this sequencing data [72]. Finally, five basic principles were used to filter the data: (1) transcripts with exon number ≥2; (2) transcripts with transcript length > 200 bp; (3) screening of transcripts with known annotations; (4) calculation of the expression of each transcript by Cuffquant, selecting transcripts with fragments per kilobase of transcript per million mapped reads (FPKM) value ≥0.5; and (5) use of Coding Potential Calculator (CPC) [73] to filter transcripts with coding potential.

**Differential expression of lncRNAs**
After screening, String Tie-eB [74] software was used to quantitatively analyse the transcripts, including lncRNAs, and obtain the FPKM values of the transcripts of each sample. With regard to statistical significance, lncRNAs were analysed as a whole. The DE-lncRNAs were statistically analysed using the edgeR program with default parameters: $p$-value < 0.05 and $|\log_2 \text{(fold change)}| > 1$.

**Target gene prediction, GO and KEGG enrichment analyses**
For position-related target gene analysis, the cis-target genes were predicted based on the positional relationship between lncRNAs and mRNAs, and the screening range was within 100 kb. Target gene prediction was based on the expression correlation between lncRNAs and mRNAs, and the screening condition was Pearson's correlation coefficient > 0.95. It should be noted that the mRNA data came from our laboratory, and RNA-seq was performed using the same batch of samples as those used for lncRNA analysis. GO enrichment analysis of DE-lncRNA target genes was performed by the GOseq R

---

**Fig. 6** Results of identification of the gonad development stages of *M. albus*. **A**: Ovary stage (OV), **B**: early intersexual gonad (IE), **C**: middle intersexual gonad (IM), **D**: late intersexual gonad (II), **E**: testis stage (TE). CAO: cortical alveolar oocyte, PG: primary growth stage, GR: gonadal ridge, SC: spermatocyte, ST: spermatid, BV: blood vessel.
software package [75], and GO entries with q-value < 0.05 were considered to be significantly enriched. KEGG is a high-throughput experimental database generated from a large-scale molecular data set by using a bioinformatics database resource for biological systems. A pathway enriched with P < 0.05 was defined as being significantly enriched by the associated differentially expressed gene. We used KOBAS software [76] to detect the enrichment of KEGG pathways by lncRNA target genes.

Cloning of 5 full length DE-lncRNAs
Using gonadal cDNAs as templates and the primer sequences shown in Additional file 10: Table S2, the initial fragments of the full length DE-lncRNAs were amplified by nested PCR. The final volume of the PCR mixture was 20 μL, which contained 2 μL of gonadal cDNA, 1 μL of each of the upstream and downstream primers (10 μM), 6 μL of water, and 10 μL of 2 × PCR Mix (Vazyme, Nanjing, China). The PCR program included 95 °C for 3 min, 36 cycles (95 °C for 0.5 min, TMC for 0.5 min, 72 °C for 1.5 min), at 72 °C for 10 min and storage at 4 °C until use. The gel recovery product was cloned into the pMD-19 T vector (Vazyme, Nanjing, China) and sequenced using universal primers.

Expression patterns of the DE-lncRNAs determined by qRT-PCR
cDNA was obtained from gonadal and other tissues (pituitary gland, eye, heart, kidney, intestines, spleen, muscle, blood) at the five developmental stages. qRT-PCR analysis (by the double internal reference method) was used to determine the expression levels of lncRNAs in the gonads at the different developmental stages. The genes ef1α and rpl17 (the GenBank accession numbers are kc011266 and kc011267, respectively) were identified as the most stable genes in M. albus at different stages of gonadal development and in different tissues [77] and were therefore used as internal controls. The primer sequences are shown in Additional file 11: Table S3. The CFX system (Bio-Rad, Chicago, USA) was used for qRT-PCR. The PCR mixture contained 1 μL of cDNA template, 0.4 μM upstream and downstream primers, and 10 μL of 2× SYBR Green MasterMix (Takara Bio, Dalian, China). PCR was carried out at 95 °C for 2 min, followed by 40 cycles of 95 °C for 0.5 s and 57-61 °C for 0.5 min and then by signal collection. The melting curve was generated as follows: 95 °C for 5 s, followed by heating from 65 °C-95 °C, increasing the temperature by 5 °C every 5 s, and then by signal collection. The amplification curve showed the expression levels and melting curve analysis confirmed the specificity of qRT-PCR amplification; CFX Manager software was used to analyse gene expression. Calculation of the copy numbers of the IncRNA and internal reference genes in CFX Manager was based on DNA fragments, ranging from 10² to 10⁹ copies [38]. The transcriptional expression level of the DE-lncRNA was calculated as follows:

\[ \frac{C_{\text{lnRNA}}}{\sqrt{C_{\text{ef1α}} \times C_{\text{rpl17}}}} \]

The expression levels of the lncRNA target genes were determined by the same method.

Expression patterns of 5 IncRNAs and znf207 after incubating ovaries with hCG and FSH in vitro
The expression patterns of 5 DE-lncRNAs (LOC109952131, LOC109953466, LOC109954337, LOC109954360 and LOC109958454) and the target gene znf207 of LOC109958454 were analysed after incubating ovaries with human chorionic gonadotropin (hCG) and follicle-stimulating hormone (FSH) in vitro. M199 medium (Sigma, Shanghai, China) was applied on ice, and M. albus ovaries were washed and dissected. Ovarian tissue (50-100 mg) was placed in a 24-well petri dish, and 1 mL of L15 containing penicillin (0.1 μg/mL, Gibco, Massachusetts, USA) and streptomycin (0.1 mg/mL, Gibco, Massachusetts, USA) was added to the basal medium; tissue culture was performed at 28 °C in a humidified incubator. Stock solutions of hCG and FSH (Sigma, Shanghai, China) were prepared in saline solution at 10³ times higher than the final concentration. After pre-incubation in normal saline for 2 h, FSH (0.05, 1.0, or 5.0 ng/ml), hCG (10, 50, or 100 IU/ml) or normal saline (control group) were incubated for 1, 2, 4, or 10 h with 3 replicates per treatment. After incubation, total RNA from ovarian tissue was extracted with TRIzol reagent (Invitrogen, Chicago, USA), cDNA was synthesized by reverse transcription, and qRT-PCR was performed to analyse the expression of lncRNAs. The results of the experimental group were compared to those of the control group as the final mapping data. Please refer to section 2.7 for the method.

Subcellular localization analysis of LOC109958454 and znf207
The ovarian tissue samples of M. albus were cleaned and dissected and then placed in cold Leibovitz’s L-15 medium (Gibco, Massachusetts, USA). Tweezers were used to continuously peel off the oocytes and connective tissue. After mixing the oocytes, the samples were divided into 5 groups, taking care to not break the cells during this process. Resuspend the cell pellet in precooled PBS (Gibco, Massachusetts, USA), and then, prechilled HLB (Foregene, Chengdu, China) was added at least 5 times the volume of the cell pellet, followed by the addition of 100 U of RNase inhibitor (G-Clone, Beijing, China). The mixture
was vortexed for 10 s and centrifuged at 1000×g and 4°C for 5 min. The supernatant (cytoplasm) was carefully aspirated into a new 1.5 mL EP tube, and the pellet (nucleus) was washed with 1 mL of HLB. An RNAprep Pure Micro Kit (Tiangen, Beijing, China) was used to extract cytoplasmic and nuclear RNA, which was then reverse transcribed to cDNA; U6 and 18S were used as internal reference genes to perform qRT-PCR experiments on the lncRNAs. Please refer to section 2.7 for the method.

Plasmid construction and dual-luciferase assay
The full length LOC109958454, znf207 and znf207 mut cDNAs were amplified from M. albus and cloned into the PSE5390 vector (Sangon Biotech, Shanghai, China) to obtain PSE5390-LOC109958454, PSE5390-znf207 and PSE5390-znf207 mut, respectively. Then, 293 T cells (Sangon Biotech, Shanghai, China) were inoculated into a 24 well plate at 30-50% confluence, with 3 replicate wells for each group. Cell transfection was performed using Lipo2000 (Invitrogen, Chicago, USA). Then, 200 μL of diluted 1× PLB was added to each well for cell lysis, and 20 μL of LAR II was added to detect firefly luciferase activity with a microplate reader (Tecan, Shanghai, China). After adding 20 μL of Stop&Glo Reagent, the microplate reader was used to detect Renilla luciferase activity. The Renilla luciferase activity values were divided by the firefly luciferase activity values, and the average values were plotted.

Statistical analysis
All the data are expressed as the mean ± S.E.M. The data were subjected to single-factor analysis of variance, followed by Tukey’s multiple comparison test, with SPSS 20.0 software (SPSS, Inc., Chicago, IL, USA). P < 0.05 was considered to indicate statistical significance.

Abbreviations
IncrRNAs: Long noncoding RNAs; DE-IncRNAs: Differentially expressed IncRNAs; FSH: Follicle stimulating hormone; hCG: Human chorionic gonadotropin; qRT-PCR: Quantitative real-time PCR; OV: Ovary stage; IE: Early intersexual stage; IM: Middle intersexual stage; IL: Late intersexual stage; TE: Testis stage; GO: GENE Ontology; BP: Biological process; CC: Cellular component; MF: Molecular function; KEGG: Kyoto Encyclopedia of Genes and Genomes; EN: Exon number; PI: Pituitary gland; EY: Eye; HE: Heart; LI: Liver; KI: Kidney; IN: Intestines; SP: Spleen; MU: Muscle; BI: Blood; PCOS: Polycystic ovary syndrome; ESR1: The estrogen receptor 1; LH: Luteinizing hormone.

Supplementary Information
The online version contains supplementary material available at https://doi.org/10.1186/s12864-022-08679-2.

Acknowledgements
Not applicable.

Authors’ contributions
ZH: Conceptualization; formal analysis; investigation; writing, original and draft; supervision; and project administration. LYJ: Conceptualization; methodology; investigation; resources; data statistics; and writing, original and draft. DYY: Supervision and validation. ZJM: Writing and review. FQD: Conceptualization and investigation. JXH: Investigation. ZH: Investigation. LZ: Investigation. YP: Resources. YYJ: Resources. QQC: Resources. KG: Resources. JXX: Resources. BLL: Resources. XBG: Validation; writing and editing. XLH: Conceptualization and investigation. SYY: Investigation and resources. MWZ: Investigation. TMY: Project administration; resources; supervision; validation; and writing, review and editing. All authors read and approved the final manuscript.

Funding
This research was supported by the National Natural Science Foundation of China [grant numbers 31972777, 2019, 31402286, 2015].

Availability of data and materials
The datasets generated and analysed during the current study are available in the Sequence Read Archive of National Center for Biotechnology Information database with accession number PRJNA798587 (https://dataview.ncbi.nlm.nih.gov/object/PRJNA798587). The datasets analysed during this study are included in this published article and its supplementary information files. Please contact Zhi He (zhihe@sicau.edu.cn) if someone wants to request the data from this study.

Declarations
Ethics approval and consent to participate
All experimental protocols involved in fishes in this study were conducted in strict accordance with the recommendations in the Guide for the Care and Use of Laboratory Animals of the Sichuan Agricultural University and ARRIVE guidelines (https://arriveguidelines.org). All procedures and investigations were reviewed and approved by the Animal Research and Ethics Committees of Sichuan Agricultural University and performed in accordance with the guidelines of the committee (Approval No.20190031).

Consent for publication
Not applicable.

Competing interests
The authors declare that they have no competing interests.

Author details
1 College of Animal Science and Technology, Sichuan Agricultural University, Chengdu 611130, Sichuan, China. 2 Luzhou Municipal Bureau of Agriculture and Rural Affairs, Luzhou 646600, Sichuan, China. 3 Sichuan Water Conservancy Vocational College, Chengdu 611231, Sichuan, China. 4 College of Veterinary Medicine, Sichuan Agricultural University, Chengdu 611130, Sichuan, China.
References

1. Novikova IV, Hennelly SP, Tung CS, Sanbonmatsu KY. Rise of the RNA machines: exploring the structure of long non-coding RNAs. J Mol Biol. 2013;425(19):3731–46. https://doi.org/10.1016/j.jmb.2013.02.030.

2. Ulitsky I, Bartel DP. lncRNAs: genomics, evolution, and mechanisms. Cell. 2013;154(1):26–46. https://doi.org/10.1016/j.cell.2013.06.020.

3. Deng FQ, Xiong S, Cai YP, He ZD, Wang XY, et al. Transcriptome analysis provides insights into long non-coding RNA functions. Front Genet. 2021;12:571325. https://doi.org/10.3389/fgene.2021.571325.

4. Taher H, Smith CA, Wernig M, Pang S, Tran C, et al. Combinatorial encoding of spatial gene regulatory networks by long noncoding RNAs. Cell. 2020;181(3):621–36. https://doi.org/10.1016/j.cell.2020.03.006.

5. Wang Z, Yang Y, Li S, Li K, Tang Z. Analysis and comparison of long non-coding RNAs expressed in the ovaries of Meishan and Yorkshire pigs. J Biol Chem. 2019;295(50):638–50. https://doi.org/10.1074/jbc.S1190007438.

6. Liu Y, Li MX, Bo W, Li T, Ma LP, Zhai T, et al. Systematic analysis of long non-coding RNAs and mRNAs in the ovaries of duroc pigs during different follicular stages using RNA sequencing. Int J Mol Sci. 2018;19(6):1722. https://doi.org/10.3390/ijms19061722.

7. Liu Y, Qi B, Xie J, Wu KQ, Ling YH, Cao XY, et al. Filtered reproductive long non-coding RNAs by genome-wide analyses of goat ovaries at different estrus periods. BMC Genomics. 2018;19(1):866. https://doi.org/10.1186/s12864-018-5268-7.

8. Yang H, Wang F, Li FZ, Ren CF, Pang J, Wan YJ, et al. Comprehensive analysis of long non-coding RNAs and mRNA expression patterns in sheep testicular maturation. Biol Reprod. 2018;99(3):650–61. https://doi.org/10.1093/biolre/ioy088.

9. Fatica A, Bozzoni I. Long non-coding RNAs: new players in cell differentiation and development. Nat Rev Genet. 2014;15(1):7–21. https://doi.org/10.1038/nrg3606.

10. Taylor DH, Chu TE, Spektor R, Soloway PD. Long non-coding RNA regulation of development and reproduction. Mol Reprod Dev. 2015;82(2):252–66. https://doi.org/10.1002/mrd.22581.

11. Mercer TR, Dinger ME, Mattick JS. Long non-coding RNAs: insights into functions. Nat Rev Genet. 2010;10(3):155–9. https://doi.org/10.1038/nrg27521.

12. Qin L, Huang CC, Yan XM, Wang Y, Li Y, Wei XC. Long non-coding RNA H19 is associated with polycystic ovary syndrome in Chinese women: a preliminary study. Endocr J. 2019;66(7):587–95. https://doi.org/10.1507/endocrj.190004.

13. Li K, Zhong SS, Luo YS, Zou DF, Li MZ, Li YH, et al. Long non-coding RNA binding to Q6-S regulates germ cell apoptosis via p38 MAPK signaling pathway. Cell Death Dis. 2019;10(9):1–9. https://doi.org/10.1038/s41419-019-1941-2.

14. Zhang J, Yu P, Zhou QY, Li XL, Ding G, Su SP, et al. Screening and characterisation of sex differentiation-related long non-coding RNAs in Chinese soft-shell turtle (Pelodiscus sinensis). Sci Rep. 2018;8(1):18363–9. https://doi.org/10.1038/s41598-018-26841-3.

15. Yang XL, Ikhuwuddin M, Li XC, Lin F, Wu QY, Zhang YL, et al. Comparative transcriptome analysis provides insights into differentially expressed genes and long non-coding RNAs between ovary and testis of the mud crab (Scylla paramamosain). Sci Rep. 2018;8(1):18363–9. https://doi.org/10.1038/s41598-018-26841-3.

16. He Z, Wu YS, Xie J, Wang TX, Zhang LH, Zhang WM. Growth differentiation factor 9 (Gdf9) was localized in the female as well as male germ cells in a protogynous hermaphroditic teleost fish, ricerfied eel Monopterus albus.
74. Pertea M, Kim D, Pertea GM, Leek JT, Salzberg SL. Transcript-level expression analysis of RNA-seq experiments with HISAT. StringTie and Ballgown Nat Protoc. 2016;11(9):1650–67. https://doi.org/10.1038/nprot.2016.095.
75. Young MD, Wakefield MJ, Smyth GK, Oshlack A. Gene ontology analysis for RNA-seq: accounting for selection bias. Genome Biol. 2010;11(2):R14. https://doi.org/10.1186/gb-2010-11-2-r14.
76. Mao XZ, Cai T, Olyarchuk JG, Wei LP. Automated genome annotation and pathway identification using the KEGG Orthology (KO) as a controlled vocabulary. Bioinformatics. 2005;21(19):3787–93. https://doi.org/10.1093/bioinformatics/bti430.
77. Hu Q, Guo W, Gao Y, Tang R, Li DP. Reference gene selection for real-time RT-PCR normalization in rice field eel (Monopterus albus) during gonad development. Fish Physiol Biochem. 2014;40(6):1721–30. https://doi.org/10.1007/s10695-014-9962-3.

Publisher’s Note
Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.