

Proteome and Transcriptome Analysis of Gonads Reveals Intersex in *Gigantidas Haimaensis*

Yu Shi

South China Sea Institute Of Oceanology

Gaoyou Yao

South China Sea Institute Of Oceanology

Hua Zhang

South China Sea Institute Of Oceanology

Huixia Jia

South China Sea Institute Of Oceanology

Panpan Xiong

South China Sea Institute Of Oceanology

Maoxian He (✉ hmx2@scsio.ac.cn)

Chinese Academy of Sciences

Research Article

Keywords: Transcriptome, Proteome, Sex change, Intersex, bivalve, *Gigantidas haimaensis*

Posted Date: November 3rd, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-1022235/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License. [Read Full License](#)

Version of Record: A version of this preprint was published at BMC Genomics on March 3rd, 2022. See the published version at <https://doi.org/10.1186/s12864-022-08407-w>.

Abstract

Sex has proven to be one of the most intriguing areas of research across evolution, development, and ecology. Intersex or sex change occurs frequently in molluscs. The deep-sea mussel *Gigantidas haimaensis* often dominates within *Haima* cold seep ecosystems, but details of their reproduction remain unknown. Herein, we conducted a combined proteomic and transcriptomic analysis of *G. haimaensis* gonads to provide a systematic understanding of sexual development in deep-sea bivalves. A total of 2,452 out of 42,238 genes (5.81%) and 288 out of 7,089 proteins (4.06%) were significantly differentially expressed between ovaries and testes with a false discovery rate (FDR) <0.05. Candidate genes involved in sexual development were identified; among 12 differentially expressed genes between sexes, four ovary-biased genes (β -catenin, fem-1, forkhead box L2 and membrane progesterin receptor α) were expressed significantly higher in males than females. Combining histological characteristics implied that these genes may be involved in oogenesis in intersex gonads, consistent with the prevalence of sex change, and gonads may tend to convert from male to female in *G. haimaensis*. The results suggest that this adaptation may be based on local environmental factors, sedentary lifestyles, and patchy distribution, and sex change may facilitate adaptation to a changing environment and expansion of the population. The findings provide a valuable genetic resource to better understand the mechanisms of sex change and survival strategies in deep-sea bivalves.

1. Introduction

Sex has proven to be one of the most intriguing areas of research across evolution, development and ecology [1]. Hermaphroditism, in both sequential and simultaneous forms, occurs in only ~5% of animals but is phylogenetically widespread (70% of phyla) [1, 2]. Sequential hermaphroditism can be expressed as male to female sexual transition (protandry), which is common in molluscs [3, 4].

The mechanism of sex reversal/differentiation involves both genetic and environmental factors [5, 6]. Environmental sex determination (ESD) includes sexual systems that are determined by external factors such as nutritional state, temperature, social structure, or some combination of environmental triggers. The oyster *Ostrea edulis* and other bivalves also change and reverse sex in response to nutritional and temperature cues [7]. Good nutritional conditions are favourable for the conversion of shellfish to females, while in poor nutritional conditions female *Pinctada margaritifera* undergo sex reversal to males [8]. In molluscs, sex change occurs based on local environmental factors such as population density and local mating population size and composition, as well as the age, size and nutritional status of individuals. It is thought that hermaphroditism occurs to facilitate adaptation to certain selective conditions. Sedentary lifestyles, combined with patchy distribution and environmental heterogeneity, appear to promote sequential hermaphroditism to increase reproductive output in molluscs [9].

Although the mechanisms of sex reversal are well studied in vertebrates, in invertebrates, particularly hermaphroditic marine molluscs, data on sex reversal are scarce. Based on high-throughput transcriptome, proteome and draft genome sequencing data, sex determination/differentiation is believed to be controlled by a major gene in the pacific oysters *Crassostrea gigas* [10, 11], *Chlamys farreri* [12–14], *Chlamys nobilis* [15], *Pinctada fucata* [16], *P. margaritifera* [17] and *Patinopecten yessoensis* [18], including double-sex- and mab-3-related transcription factor (DMRT) and SoxE, SOXH (SRY-like) for male sex-determining pathways, and β -catenin and fork head box L2 (foxl2) for female sex-determining pathways.

Much of our current understanding of sexual development comes from a small number of model systems, limiting our ability to make broader conclusions about the evolution of sexual diversity. Deep-sea hydrothermal vents and seeps, characterised by darkness and high concentrations of heavy metals and other toxic substances, can provide sulphide, methane and hydrogen sulphide as chemical energy for use by chemoautotrophic bacteria to support dense populations of invertebrates [19, 20]. Among the deep-sea macro-fauna, *Bathymodiolus* (Bivalvia, *Mytilidae*) mussels often dominate at many cold seep and hydrothermal vent ecosystems worldwide [19, 21]. Previous studies have focused on symbiosis [22], immunity [23], adaptation to abiotic stress [24], ecotoxicology [25], biogeography [26] and genomes [27]. The deep-sea mussel *Gigantidas haimaensis* often dominates at *Haima* cold seep ecosystems on the northwestern slope of the South China Sea [28], but knowledge on reproduction in this species is lacking.

The survival strategies through which *G. haimaensis* adapts to its environment remain poorly understood. To gain insight into the adaptive features of the gonads, we focused genes related to sex. We performed in-depth proteomics and transcriptomics analyses on gonads and analysed the impact of the environment on gonadal development in males and females. The findings expand our understanding of gonadal development in bivalves, and the influence of extreme environments on gonad development.

2. Materials And Methods

2.1 Animals and Collection

G. haimaensis mussels were obtained from the Haima Cold Seep (16.73°N, 110.475°E, depth 1,446 m) using the manned submersible ROV *Haima* during cruise HYDZ6-202005 of the research vessel (R/V) *Haiyang 6* of the Guangzhou Marine Geological Survey (China; September 1st–6th, 2020). Some of the mussels were frozen immediately in liquid nitrogen for 24 h then stored at -80°C, while others were fixed in 100% ethanol. After cruises, mussels were placed on dry ice and transported to the South China Sea Institute of Oceanology, Chinese Academy of Sciences. Gonads were dissected for subsequent RNA extraction and histological procedures. All animal experiments were conducted in accordance with the guidelines and approval of the Animal Research and Ethics Committees of the Chinese Academy of Sciences.

2.2 Protein Digestion

Digestion of protein (250 μ g per sample) from three oysters was performed according to the PASP procedure [29–31]. Protein quality was tested by a Bradford protein assay kit according to the manufacturer's instructions. TMT labelling of peptides was performed according to a procedure described previously [32].

Mobile phase A (2% acetonitrile, adjusted to pH 10.0 using ammonium hydroxide) and B (98% acetonitrile) were used to develop a gradient elution. The lyophilised powder was dissolved in solution A and centrifuged at 12,000 g for 10 min at room temperature. The sample was fractionated using a Waters BEH C18 column (4.6 × 250 mm, 5 µm; Waters) on a Rigol L3000 HPLC system, with a column temperature of 45°C. Eluates were monitored at an absorbance wavelength of 214 nm, fractions were collected at one tube per min, and combined into 10 fractions. All fractions were dried under vacuum, then reconstituted in 0.1% (v/v) formic acid (FA) in water.

2.3 Liquid Chromatography Tandem Mass Spectrometry (LC-MS/MS)

Mobile phase A (100% water, 0.1% FA) and B solution (80% acetonitrile, 0.1% FA) were prepared. Samples (1 µg) were injected into a home-made C18 Nano-Trap column (4.5 cm × 75 µm, 3 µm). A home-made analytical column (25 cm × 150 µm, 1.9 µm) was employed and the column oven was set as 55°C. The separated peptides were analysed by an Orbitrap Exploris 480 instrument coupled with FAIMS (Thermo Fisher) and a Nanospray Flex electrospray ionisation (ESI) device with a spray voltage of 2.1 kV and an ion transport capillary temperature of 320°C. The data-dependent acquisition mode was employed for MS data collection, the FAIMS compensation voltage was set at -45 and -65, and the acquisition parameters were as follows: full scan ranges from m/z 350 to 1500 with a resolution of 60,000 (at m/z 200), automatic gain control (AGC) target value = Auto (the optimal capacity was automatically calculated by the software according to other parameters), and maximum ion injection time = Auto. The scan-round time for MS/MS was set to 1s, and the precursors in the full scan were selected from high to low abundance and fragmented by higher energy collisional dissociation (HCD) with a resolution of 30,000 (at m/z 200), the turbo TMT+precursor Fit function was turned on, and the AGC target value was 1×10^5 . The maximum ion injection time was set to Auto, the normalised collision energy was 36%, the intensity threshold was 5.0×10^3 , and the dynamic exclusion parameter was 45 s. Raw MS data were named "raw".

2.4 Data Analysis

Label-free quantification was carried out in MaxQuant as previously described [33]. The resulting spectra from each run were searched separately against the 720541-X101SC21041130-Z02-unigene.blast.pep.fasta database using the Proteome Discoverer 2.4 (PD 2.4) search engine (Thermo Fisher).

In order to improve the quality of analysis results, PD 2.4 software was used to further filter the retrieval results, and a credibility score >99% identified peptide spectrum matches (PSMs). The identified proteins included at least one unique peptide and all identified PSMs and proteins had a false discovery rate (FDR) <1.0%. The protein quantitation results were statistically analyzed by t-test, and quantitative differences between experimental and control groups with $p < 0.05$ and fold change (FC) ≥ 2.0 and $FC \leq 0.50$ were defined as differentially expressed proteins (DEPs).

The MS proteomics data have been deposited at the Science Data Bank under dataset identifier CSTR 31253.11.sciencedb.01147/ DOI 10.11922/sciencedb.01147.

2.5 Construction of Complementary DNA Library and Illumina Sequencing

Total RNA from three oysters was used as input material for RNA sample preparation. Briefly, mRNA was purified from total RNA using poly-T oligo-attached magnetic beads. Fragmentation was carried out using divalent cations under elevated temperature in First Strand Synthesis Reaction Buffer (5-). First-strand cDNA was synthesised using random hexamer primer and M-MuLV reverse transcriptase, and RNaseH was added to degrade RNA. Second-strand cDNA synthesis was subsequently performed using DNA Polymerase I and dNTPs. Remaining overhangs were converted into blunt ends via exonuclease/polymerase activities. After adenylation of 3' ends of DNA fragments, adaptors with a hairpin loop structure were ligated to prepare for hybridisation. In order to select cDNA fragments mainly 370–420 bp in length, the library fragments were purified with an AMPure XP system (Beckman Coulter, Beverly, USA) [34, 35]. PCR was performed using Phusion High-Fidelity DNA polymerase, Universal PCR primers, and Index (X) Primer. Finally, PCR products were purified using an AMPure XP system and library quality was assessed on a Qubit 2.0 Fluorimeter, a Agilent Bioanalyzer 2100 system, and by quantitative real-time PCR (qRT-PCR).

Clustering of index-coded samples was performed on a cBot Cluster Generation System using a TruSeq PE Cluster Kit v3-cBot-HS (Illumina) according to the manufacturer's instructions. After cluster generation, library preparations were sequenced on an Illumina Novaseq platform and 150 bp paired-end reads were generated.

2.6 Data Filtering and *De Novo* Assembly

Raw data (raw reads) in fastq format were processed via in-house perl scripts. In this step, clean data (clean reads) were obtained by removing reads containing adapter, reads containing N bases, and low-quality reads from raw data. Meanwhile, Q20, Q30 and GC content were calculated for clean data. All downstream analyses were based on clean data of high quality. Transcriptomes were separately assembled *de novo* using Trinity with `min_kmer_cov` set to 2 by default and all other parameters set to default (<http://trinityrnaseq.sourceforge.net/>) [34].

2.7 Gene Functional Annotation

Gene functions were annotated using Nr (NCBI non-redundant protein sequences), Nt (NCBI non-redundant nucleotide sequences), Pfam (Protein family), KOG/COG (Clusters of Orthologous Groups of proteins), Swiss-Prot (a manually annotated and reviewed protein sequence database), KO (KEGG Ortholog database)[36] and GO (Gene Ontology) databases [37].

2.8 Differential Expression Analysis

Differential expression analysis of two conditions/groups (two biological replicates per condition) was performed using the DESeq2 R package (1.20.0). DESeq2 provides statistical routines for determining differential expression in digital gene expression data using a model based on the negative binomial distribution. The resulting p -values were adjusted using the Benjamini and Hochberg's approach for controlling the FDR. Genes with an adjusted p -value <0.05 identified by DESeq2 were assigned as differentially expressed [34].

Prior to differential gene expression analysis, for each sequenced library, read counts were adjusted by then edgeR program through a scaling normalised factor. Differential expression analysis of two conditions was performed using the edgeR R package (3.22.5), and *p*-values were adjusted using the Benjamini & Hochberg method. Corrected *p*-values of 0.05 and an absolute FC value of 2 were set as thresholds for significantly differential expression.

2.9 Histological Procedures

After fixation in alcohol for 24 h, gonadal tissues were dehydrated and embedded in paraffin for histology. Tissues were serially sectioned at 7 µm and stained with hematoxylin and eosin. Classification of the sex stage was determined under light microscopy [14].

2.10 RT-qPCR Validation

A total of 12 genes were selected for RT-qPCR validation. Gonadal tissues of *G. haimaensis* were collected from 5 male and 5 female oysters for qRT-PCR. Tissues were ground in a homogeniser (IKA, Staufen, Germany). Total RNA was isolated from whole tissues with TRIzol Reagent (Invitrogen, Carlsbad, CA, USA), quality was checked by 1.2% (w/v) agarose gel electrophoresis, and quantity was measured using a Quawell Q5000 ultraviolet spectrophotometer (San Jose, CA, USA). All RNAs were treated with DNase I to avoid genomic contamination. A 1 µg sample of isolated RNA was used to synthesise first-strand cDNA using a ReverTra Ace-a First Strand cDNA Synthesis Kit (Toyobo, Tokyo, Japan). Primers designed for each gene are listed in Table 1. qPCR was performed using a Roche LightCycler 480 RT-PCR system with SYBR(R) Premix Ex Taq (Toyobo) according to the manufacturer's protocol. After amplification, fluorescent data were converted to threshold cycle (Ct) values. Concentrations of templates in samples were determined by relating Ct values to standard curves. Target gene transcript levels were normalised against reference gene transcript levels. Reference genes were 60s RP-L15 and β-actin [14].

Table 1
List of primers for quantitative PCR validation

Gene name	Pro/Tran ID	Primer sequence (5' to 3')
<i>beta-catenin</i>	Cluster-1780.1/Cluster-1780.1; orf1	F: AATGCTTGATGGCTGTT R: CTGTATCTTCTTCGGGTG
<i>conodipin</i>	-/Cluster-5499.0	F: AAATAGAAGCGTCAAACG R: GTAGAGGGACCAGCCAAT
<i>Cyp450</i>	Cluster-19168.0;orf1/Cluster-19168.0	F: GGCAAAGTTCTGTCCAAT R: TAAGTCTAAGCAGCGTTC
<i>EGFC</i>	Cluster-18098.1/Cluster-18098.1;	F: CGTGGACAAAGCCATAAA R: CCATATTCTTCGCCATA
<i>fem-1</i>	Cluster-10581.0;orf1/Cluster-10581.0	F: CCTGAACGTTTACTACTCC R: AAATTGCCGAACTACTG
<i>foxl2</i>	-/Cluster-22036.0	F: AATCCAGCAGCACCAACA R: TACGGGAAGGGAACGAG
<i>MMP</i>	Cluster-4389.6602;orf1/Cluster-4389. 6602	F: CGTGCTCTTTGTGAACTT R: GGAGATACTTTGGCTTTT
<i>MPI</i>	Cluster-4389.613;orf1/Cluster-4389.613	F: GCCTATGCGATATTTCTT R: TTGTTTCTGTTTCGGTTGA
<i>MPR</i>	-/Cluster-14850.0	F: TATTTGAGTGCTGGTTGA R: ATCGGTAAGGATGTT
<i>PLAC8</i>	-/Cluster-4389.12722	F: TGGAAACTCGGCTCAACA R: CTGGCAACATAGACAAAGAAAA
<i>SOX2</i>	-/Cluster-16149.0	F: ATGGCACCTCAAACACTACAC R: AACTTGCGAACTCCTCC
<i>vitellogenin</i>	Cluster-20609.0;orf1/Cluster-20609.0	F: AGCCAGAGGAAGTAAGGA R: TATACGAATGTGCCAACA
<i>60s RP-L15</i>	Cluster-4389.7866;orf1/Cluster-4389.7866	F: AGCATCTGACACGGAGCA R: GACACGAGCCAGCAAGAA
<i>β-actin</i>	Cluster-4389.8162/Cluster-4389. 8162	F: GATGAAGCGCAGAGCAAGAG R: TTGTGTCATCTTCTCTGTTTGTCT

3. Results

3.1 Histological Characteristics of *G. haimaensis* Gonads

Gonads were characterised by a majority of spermatogenic cysts filled with spermatids, but some spermatogenic cysts were empty since spermatozoa may have been released. Additionally, some primary oocytes were observed between spermatogenic cysts (Fig. 1b). Ovaries were characterised by mostly mature oocytes, a few oogonia, some empty follicular cavities, and intragonadal somatic cells (Fig. 1c).

3.2 *G. haimaensis* Gonad Transcriptome

As listed in Table 2, Illumina sequencing of ovary and testis transcriptomes generated 128,452,470 raw reads of 100 bp, of which 125,797,626 (97.93%) remained after quality filtering, and 20,966,271 filtered clean reads with Q20 >97.17% were obtained from each library. Raw reads have been submitted to the Science Data Bank under accession numbers CSTR 31253.11.sciencedb.01146/DOI 10.11922/sciencedb.01146. *De novo* Trinity assembly from combined ovary and testis reads produced 78,860 assembled transcripts and 42,238 unigenes. BUSCO revealed a transcriptome completeness of 89.4% of, indicating high-quality *de novo* assembly.

Table 2
Gonad Transcriptome Data of *Gigantidas haimaensis*

sample	raw_reads	clean_reads	clean_bases	error_rate	Q20	Q30	GC_pct
X1	21453392	20994526	6.3G	0.03	97.29	92.07	33.12
X2	20859770	20429242	6.13G	0.03	97.29	92.23	33.39
X3	21341314	20796278	6.24G	0.03	97.42	92.5	34.04
C1	21923435	21424536	6.43G	0.03	97.17	91.9	32.79
C2	20940410	20623937	6.19G	0.03	97.24	92.05	33.19
C3	21934149	21529107	6.46G	0.03	97.17	91.9	32.74

Eventually, 56.79% of 78,860 assembled transcripts were annotated using at least against one of the databases (Table S1), and 17,090 (40.46%) had significant matches against the NR database.

Differential expression analysis revealed that 2,452 out of 42,238 genes (5.81%) were significantly differentially expressed between ovaries and testes with FDR <0.05. Among them, 976 (39.80%) were significantly higher in ovaries (hereafter called ovary-biased genes) compared with 1,476 (60.20%) in testes (hereafter called testis-biased genes; Fig. 2).

All differentially expressed genes (DEGs) were subjected to GO functional analysis. Based on GO analysis, testis-biased genes were linked to 1,399 biological processes, 357 cellular components and 547 molecular functions annotated for GO term assignments, mainly related to membrane, intracellular organelle and transport (Table 3). Ovary-biased genes included 1,090 biological processes, 286 cellular components and 420 molecular functions annotated for GO term assignments, mainly belonging to cellular macromolecule metabolic process, protein metabolic process and metal ion binding (Table 3). Meanwhile, testis-biased genes were associated with 161 KEGG annotation, mainly belonging to cell cycle, purine metabolism, oocyte meiosis, progesterone-mediated oocyte maturation and p53 signalling pathway (Fig. 3a, Table 4). Ovary-biased genes included 215 KEGG annotations, mainly belonging to ribosome, Ras signalling pathway, GnRH signalling pathway, oocyte meiosis and tight junction categories (Fig. 3b, Table 4). Most of these GO term and KEGG pathway enrichments are closely related to sex differentiation or determination.

Table 3
GO terms enriched in ovary- and testis-biased genes of *Gigantidas haimaensis*

ID code	Term	Genes	Pvalue
Ovary-biased			
GO:0044260	cellular macromolecule metabolic process	120	0.049098
GO:0019538	protein metabolic process	70	0.039514
GO:0046872	metal ion binding	66	0.0079916
GO:0043169	cation binding	66	0.0092841
GO:0055114	oxidation-reduction process	36	0.045325
GO:0005509	calcium ion binding	25	9.69E-05
GO:1990234	transferase complex	20	0.01068
GO:0030001	metal ion transport	20	0.021903
GO:0006260	DNA replication	17	0.027564
GO:0065008	regulation of biological quality	17	0.036641
Testis-biased			
GO:0110165	cellular anatomical entity	361	0.0033264
GO:0016020	membrane	200	0.044805
GO:0043226	organelle	183	0.049217
GO:0043229	intracellular organelle	173	0.041224
GO:0051179	localization	143	0.02073
GO:0051234	establishment of localization	136	0.028072
GO:0006810	transport	135	0.028886
GO:0071840	cellular component organization or biogenesis	97	0.025886
GO:0043232	intracellular non-membrane-bounded organelle	94	0.0022039
GO:0043228	non-membrane-bounded organelle	94	0.0041935

Table 4
KEGG pathways enriched in ovary- and testis-biased genes of *Gigantidas haimaensis*

ID code	Pathway	Genes	Pvalue
Ovary-biased			
ko04261	Adrenergic signaling in cardiomyocytes	12	0.000859
ko04260	Cardiac muscle contraction	6	0.003592
ko04912	GnRH signaling pathway	8	0.014003
ko04612	Antigen processing and presentation	4	0.026682
ko00330	Arginine and proline metabolism	5	0.047621
ko04744	Phototransduction	4	0.048889
Testis-biased			
ko04110	Cell cycle	13	2.05E-05
ko04914	Progesterone-mediated oocyte maturation	8	0.000215
ko04115	p53 signaling pathway	7	0.000594
ko04114	Oocyte meiosis	10	0.000635
ko00230	Purine metabolism	11	0.003019
ko04068	FoxO signaling pathway	6	0.014706
ko00910	Nitrogen metabolism	2	0.042186

3.3 *G. haimaensis* Gonad Proteome

To explore sex-related protein expression profiles in *G. haimaensis*, we conducted a large-scale proteomics study using label-free LC-MS/MS data. We studied gonadal tissues from six mature male (M) and mature female (F) *G. haimaensis* samples. We obtained 50,127 unique peptides and 7,089 protein groups of *G. haimaensis* proteins.

A total of 288 DEPs were identified between M and F, with 219 (76%) upregulated in M and 69 (24%) downregulated in M (Fig. 4a). The expression patterns of proteins among M and F groups were quite similar (Fig. 4b).

We analysed GO enrichment between DEPs of M and F groups. The significantly enriched GO categories are shown in Figure 4. Numerous proteins overexpressed in M (compared with F) were enriched in 18 GO terms: 56 proteins in binding, 29 proteins in protein binding, 23 proteins in organic cyclic compound binding, 23 proteins in heterocyclic compound binding, 14 proteins in nucleic acid binding, 13 proteins in metal ion binding, and 10 proteins in carbohydrate derivative binding (Fig. 5a).

Numerous proteins overexpressed in F (compared with M) were enriched in 23 GO terms: 15 proteins in single-organism process, 8 proteins in ion binding, 7 proteins in membrane, 7 proteins in metal ion binding, 7 proteins in single-organism metabolic process, 7 proteins in single-organism cellular process, and 6 proteins in integral component of membrane (Fig. 5b).

Meanwhile, testis-biased genes had 84 KEGG annotations, mainly belonging to oocyte meiosis, cGMP-PKG signalling pathway, apelin signalling pathway, and oxytocin signalling pathway (Fig. 6a). Ovary-biased genes had 70 KEGG annotations, mainly belonging to metabolic pathways, apelin signalling pathway, lysosome, and protein processing in endoplasmic reticulum (Fig. 6b).

Statistical analysis of the subcellular localisation of the different proteins was performed and the results are shown in Figure 7. In total, 47 (37.60%) proteins were nuclear proteins, 24 (19.20%) were cytoplasmic proteins, 15 (12.00%) were plasma membrane proteins, 9 (7.20%) were mitochondrial and endoplasmic reticulum proteins, and 7 (5.60%) were centrosome proteins.

3.4 Comparison of Protein and mRNA Expression Levels

Pearson correlation coefficients for protein abundance measured by label-free assay and mRNA levels measured by high-throughput Illumina HiSeq 2500 sequencing gave a high correlation for F versus M ($r = 0.556$; Fig. 8a, b, Supplementary Table S2). We compared log₂-transformed protein expression FC and mRNA expression FC values for DEPs between F and M ($p < 0.05$; Fig. 7). A total of 108 differed between F and M at both protein and mRNA levels (Fig. 8c); 177 differed between F and M at the protein but not the mRNA level (Fig. 8c); 627 differed between F versus M at the mRNA but not the protein level (Fig. 8c). GO analyses of transcriptomes revealed DEGs for F versus M that were largely involved in the catalytic activity molecular function, consistent with GO analyses of the proteome (Fig. 8d, Supplementary Table S3). The clustering heatmap for KEGG pathway enrichment based on proteome and transcriptome data showed that a large percentage of DEGs with transcriptomic KEGG pathway enrichment were also displayed proteomic KEGG pathway enrichment (Fig. 8e, Supplementary Table S3). These results indicate an overall good correlation between mRNA and protein levels among gonad genes.

3.5 Genes Related to Sexual Development

Two selected genes (CYP450 and Epidermal Growth Factor Receptor (EGFR)) showed significant differential expression at both mRNA and protein levels, five genes (beta-catenin, fem-1, MMP, MPI, vitellogenin) displayed significant differential expression only at the mRNA level, and five genes (conodipin, foxl2, membrane progesterin receptor α (MPR α), PLAC8, SOX2) exhibited significant differential expression only at the protein level (Table 5). Among DEGs identified from transcriptome data, differential expression of 12 sex-related genes was confirmed by quantitative PCR (qPCR), validating the RNA sequencing (RNA-Seq) data (Table 1).

Table 5
Selected sex reversal –related genes detected in transcriptome, proteome and tested by Q-PCR.

Transcriptome						Proteome			
Gene	Gene_ID	Gene annotation	log2FC	pval	Comparison	Protein ID	log2FC	pval	Comparison
beta-catenin	Cluster-1780.1	Armadillo repeat-containing protein 3 [<i>Mizuhopecten yessoensis</i>]	3.2356	9.51E-05	F < M	Cluster-1780.1;orf1	0.937422094	0.071511626	-
conodipin	Cluster-5499.0	Conodipine-M alpha chain [<i>Mizuhopecten yessoensis</i>]	6.7606	1.95E-05	F < M	-	-	-	-
CYP450	Cluster-19494.0	cytochrome P450 family 4, partial [<i>Mytilus galloprovincialis</i>]	-5.1622	0.002184	F > M	Cluster-19168.0;orf1	-1.02459691	0.000583065	F > M
EGFC	Cluster-18602.0	Epidermal growth factor receptor [<i>Mizuhopecten yessoensis</i>]	-7.5487	0.000887	F > M	Cluster-18098.1;	-1.04369702	0.041341124	F > M
fem-1	Cluster-4389.1741	sex determining protein Fem-1 like protein [<i>Pinctada margaritifera</i>]	7.08	8.23E-06	F > M	Cluster-10581.0;orf1	-0.42988855	0.011276669	-
foxl2	Cluster-22036.0	forkhead box protein L2-like [<i>Crassostrea virginica</i>]	-4.5978	0.000749	-	-	-	-	-
MMP	Cluster-4389.6602	PREDICTED: stromelysin-3 [<i>Crassostrea gigas</i>]	-3.0314	0.001984	F > M	Cluster-4389.6602;orf1	-0.59494927	0.421931357	-
MPI	Cluster-4389.8345	metalloproteinase inhibitor 3 [<i>Lingula anatina</i>]	-2.9136	0.000226	F > M	Cluster-20469.0;orf1	-0.61434809	0.310785182	-
MPR	Cluster-14850.0	membrane progesterin receptor alpha-B-like [<i>Mizuhopecten yessoensis</i>]	5.6866	0.000113	F < M	-	-	-	-
PLAC8	Cluster-4389.12722	PREDICTED: placenta-specific gene 8 protein-like [<i>Crassostrea gigas</i>]	4.8244	4.39E-05	F < M	-	-	-	-
sox2	Cluster-16149.0	Sox2 [<i>Pinctada fucata</i>]	-4.7151	2.84E-10	F > M	-	-	-	-
vitellogenin	Cluster-20609.0	vitellogenin [<i>Mimachlamys nobilis</i>]	-9.1347	6.56E-09	F > M	Cluster-20609.0;orf1	-3.14219096	0.14095379	-

The qPCR results showed that differential expression patterns of 11 of the 12 selected genes were generally consistent with the RNA-Seq analysis (Fig. 9, Table 5); expression of four genes (beta-catenin, conodipin, MPR α and PLAC8) was significantly higher in M compared to F, while expression of eight genes (CYP450, EGFR, fem-1, foxl2, MMP, MPI, sox2 and vitellogenin) was significantly higher in F than M. Therefore, logFC and qPCR assays, RNA-Seq data, and label-free data were correlated. The consistency between qPCR, RNA-Seq and label-free data confirmed the reliability of RNA-Seq and label-free data for accurately quantifying gene expression.

4. Discussion

4.1 Determining the Direction of Sex Change based on Histological Characteristics of *G. haimaensis* Gonads

When dissecting gonads of *G. haimaensis*, there was no difference in colour, size or general appearance between males and females; both M and F gonads were white (Fig. 1a). Based on the histological characteristics of gonads, gonads from three females and three male were chosen for transcriptome and proteome analyses.

Spermatogenic cysts were filled with spermatids, some of which were surrounded by follicular cavities and oogonia (Fig. 1a), and 6 of 15 of the examined gonads were testes, all with oogonia. There were no spermatocytes in ovaries, primary oocytes were close to the follicular wall, and they absorbed nutrients from the follicular wall to develop into mature oocytes (Fig. 1b). These characteristics indicate that male *G. haimaensis* may be intersex, and gonads might convert from male to female.

4.2 Candidate Genes Involved in Sexual Development

Although several sex-related genes have been studied in terms of sex determination and/or differentiation, little information about sex reversal/differentiation cascades is available for *G. haimaensis*.

Some female-biased genes were expressed more highly in M than F, including β -catenin, fem-1, mPR α and PLAC8. β -catenin, which transduces the canonical Wnt signalling pathway in mammals by promoting stability in the cytoplasm and nuclear entry, is critical in female ovary differentiation [38]. In molluscs, β -catenin was expressed mostly in mature female gonads, which indicates a conserved role in the maintenance of ovaries [11, 39]. The ankyrin repeat protein Fem-1 is a component of the signal transduction pathway controlling sex determination [40]. The fem-1 gene of *P. margaritifera* was expressed specifically in different reproductive stages (undetermined, sexual inversion, and regression), suggesting that it may be involved in the sperm-oocyte switch [17]. mPR α was primarily localised on the oocyte plasma membrane, where it regulates induction of oocyte maturation by specifically binding progestins [41]. Progestins exert rapid, multifaceted and nongenomic effects on sperm physiology through mPR α in a variety of vertebrate species [42–44], but few on invertebrates have been reported. In *Octopus vulgaris*, progesterone can also induce activation in spermatozoa via mPR α [45]. However, our qPCR results showed that β -catenin, fem-1 and mPR α were expressed higher in M than F, consistent with the transcriptomic data. The qPCR results differed from the proteomic data, which may be because not all transcripts were translated, and mRNA abundance may not correspond to protein expression due to pre-translational, co-translational, or post-translational modification. Combined with the histological characteristics of gonads, these results implied that they may be involved in oogenesis in intersex gonads. A previous study on *Bathymodiolus platifrons* revealed low levels of genetic diversity differences between vent and seep populations [46]. Our results also showed low levels of genetic diversity differences for *G. haimaensis* between vent and seep populations (data not shown, unpublished), which may reflect the survival strategy of sex change that could lead to self-fertilisation to expand the population.

Plac8 is a placental-enriched gene expressed in the spongiotrophoblast layer during mouse development [47]. In mouse, it also plays an important role in spermatocyte differentiation during spermatogenesis [48]. Plac8 has been well studied in vertebrates, yet little is known about its role in invertebrates. In the planarian *Dugesia japonica*, Plac8 plays essential roles in immune responses and development [49]. In *G. haimaensis*, Plac8 was expressed higher in M than F, consistent with the transcriptomic data, implying that Plac8 may be involved in spermatocyte differentiation.

Conodipine-M is a novel phospholipase A2 enzyme isolated from the venom of the marine snail *Conus magus* [50]. Conodipine is secreted by the poison gland in the venom tube and the inner wall of the poison sac. It can specifically act on various ion channels such as potassium, sodium, calcium, and various receptors on the cell membrane, to affect signal transmission in nerve and other cells [50]. Our qPCR results revealed sexual dimorphism for expression of conodipine-M, with significantly higher levels in M than F, suggesting that it may participate in the maintenance of masculinisation. However, the specific role of conodipine-M in testis needs to be clarified in the future.

Some ovary-biased genes reported in other studies and were also identified in our current research. CYP450, EGFR, foxl2, MMP, MPI and vitellogenin were expressed at significantly higher levels in F than M, consistent with the transcriptomic data, and the qPCR results for CYP450 and EGFR were consistent with the proteomic data. In the mammalian female pathway, foxl2 functions by upregulating CYP19A, which encodes the aromatase that converts testosterone into oestrogens [51]. However, CYP19A has been demonstrated to have arisen in the chordate lineage [52], hence it does not exist in bivalves. This finding suggests that despite the possible deep conservation of sex-determining genes among different clades, their regulatory network may have diverged during evolution. The forkhead box L2 gene (foxl2), which encodes a forkhead class transcription factor, is a key factor in the differentiation and maintenance of ovaries in vertebrates. And in molluscs, foxl2 also showed a sexually dimorphic pattern with higher expression levels in females [10, 13, 14, 17, 53], consistent with our qPCR results, which indicates a conserved structure and function of foxl2. Localisation of the Foxl2 protein in spermatogonia and spermatocytes implies that it may also be involved in spermatogenesis in *G. haimaensis*. The EGFR system contributes to key stages of reproduction, such as ovulation, fertilisation, embryo implantation, and the attainment of sexual maturity [54]. Most studies have focused on vertebrates, and few findings have been reported for invertebrates. Our results indicate that EGFR may play a conserved role in ovary development.

Matrix metalloproteinases (MMPs) play an important role in the reproductive process by degrading the extracellular matrix and weakening the follicle wall, leading to follicle rupture [55, 56]. MMP2 and MMP9 stimulate luteinising hormone (LH)-induced steroid production by regulating the release of the EGFR ligand [57]. Tissue inhibitors of metalloproteinases (TIMPs) are proteins secreted by a variety of cells that can selectively inhibit the activity of MMPs. The mechanism of action is to form chelates with metal ions necessary for the active sites of MMPs [58]. Under physiological conditions, TIMPs and MMPs jointly maintain the stability of the biological environment in vivo. However, under pathological conditions, due to their direct action on MMPs, TIMPs are an extremely critical factor for maintaining homeostasis when the activity of MMPs activity is imbalanced. In *G. haimaensis*, MMPs may combine with TIMPs to promote EGFR to stimulate ovary development.

Vitellogenin (VTG), a precursor of yolk protein in oviparous animals, is a molecular carrier that transports nutrients into egg cells. Research on the biological functions of VTG shows that it regulates the osmotic pressure of ovaries [59, 60]. It is also an important immune molecule against pathogenic microorganisms [61–63], and it assists sperm fertilisation [64, 65]. Our transcriptome data and qPCR results showed that vtg was more highly expressed in F than M, indicating an exclusive role in ovaries rather than intersex or testes.

Sox2, a member of the SOX (SRY-related HMG-box) family, is an important transcription factor participating in embryogenesis [66, 67], neurogenesis [68–70], maintenance of stem cells [71–73] and proliferation of primordial germ cells (PGCs) [74]. Also, Sox2 is involved in male germ cell development and stem cell

maintenance in fish [75] and spermatogenesis in Zhikong scallop *Chlamys farreri* [76]. In medaka, Sox2 is specifically expressed in ovary [77], and in *Paralichthys olivaceus* Sox2 is expressed more highly in ovary than testis [78]. Our results showed that *sox2* was expressed significantly higher in F than in M, consistent with the transcriptomic data, and localised in both testis and ovary, suggesting that Sox2 may function in male germ cell development and the maintenance of feminisation.

4.3 Putative Mechanism of Sex Change in *G. haimaensis*

In sex determination, the *foxl2*-leading pathway and RSP0-1/WNT4/ β -catenin signalling pathway act independently and complementary to each other to promote and maintain ovarian development [79–82]. In *G. haimaensis*, *foxl2* was specifically expressed in ovary, while β -catenin was expressed more highly in testis, implying that the *foxl2*-leading pathway may perform a leading role in the ovary determination/maintenance pathway of *G. haimaensis*. Also, *Foxl2* may upregulate CYP450 to increase estrogens, and MMPs may combine with TIMPs to regulate EGFR and stimulate ovary development. The histological characteristics male gonads resembled intersex gonads, and β -catenin may play a role in intersex gonads to initiate ovary development.

Fem-1, mPR α and PLAC8 are downstream genes of the β -catenin signalling pathway regulated by β -catenin that may be involved in oogenesis in testis. We did not identify a testis-determining factor, such as SRY (Sex Determining Region Y) or DMRT, but there may be other male-determining genes. Sedentary lifestyles, combined with patchy distribution and environmental heterogeneity in darkness and in the presence of high concentrations of heavy metals and other toxic substances may stimulate *G. haimaensis* to change sex and thereby increase reproductive output.

In conclusion, our sex-biased proteomics and transcriptomics analysis of testes and ovaries in *G. haimaensis* revealed a strong correlation between mRNA and protein levels of key genes and proteins. Twelve DEGs between sexes were identified, and four ovary-biased genes (β -catenin, fem-1, *foxl2* and mPR α) were expressed significantly higher in M than F. Histological characteristics of gonads implied that they may be involved in oogenesis in intersex gonads, further confirming the occurrence of sex change, and gonads might convert from male to female in *G. haimaensis*. This adaptation may be based on local environmental factors, sedentary lifestyles, and patchy distribution. These findings suggest a deeply conserved function of these genes in sex development, and a diverged regulatory pathway during evolution. This study provides a valuable genetic resource to better understand the mechanisms of sex change and the survival strategies in deep-sea bivalves.

Declarations

Ethics approval and consent to participate

All animal experiments were conducted in accordance with the guidelines and approval of the Animal Research and Ethics Committees of the Chinese Academy of Sciences.

Consent for publication

Not applicable

Availability of data and materials

Data for this manuscript are available at the Science Data Bank (<https://www.scidb.cn/s/2MbaA3>; <https://www.scidb.cn/s/i6N7zy>).

Competing interests

The authors have no conflict of interest to declare.

Funding

This project was supported by the Key Special Project for Introduced Talents Team of Southern Marine Science and Engineering Guangdong Laboratory, Guangzhou, China (grant number GML2019ZD0401), the Major Project of Basic and Applied Basic Research of Guangdong Province (grant number 2019B030302004), and the Science and Technology Planning Project of Guangdong Province, China (grant number 2020B1212060058).

Authors' contributions

Y. S. planned and designed the research. MX. H supervised the experiments. Y. S. performed most of the experiments, analysed data, and wrote the manuscript. GY. Y and H. Z. sailed to the South China sea to collect samples. HX. J and PP. X helped perform RT-qPCR. All authors reviewed, edited and revised the manuscript.

Acknowledgements

Not applicable

References

1. Jarne P, Auld JR: **Animals mix it up too: The distribution of self-fertilization among hermaphroditic animals.** *Evolution* 2006, **60**(9):1816–1824.
2. Vega-Frutis R, Macias-Ordonez R, Guevara R, Fromhage L: **Sex change in plants and animals: a unified perspective.** *J Evol Biol* 2014, **27**(4):667–675.
3. Hoagland KE: **Protandry and Evolution of Environmentally-Mediated Sex Change - Study of Mollusca.** *Malacologia* 1978, **17**(2):365–391.

4. Heller J: **Hermaphroditism in Mollusks.** *Biol J Linn Soc* 1993, **48**(1):19–42.
5. Penman DJ, Piferer F: **Fish Gonadogenesis. Part I: Genetic and Environmental Mechanisms of Sex Determination.** *Rev Fish Sci* 2008, **16**:16–34.
6. Piferer F, Guiguen Y: **Fish Gonadogenesis. Part II: Molecular Biology and Genomics of Sex Differentiation.** *Rev Fish Sci* 2008, **16**:35–55.
7. Coe WR: **Sexual differentiation in mollusks I Pelecypods.** *Q Rev Biol* 1943, **18**(2):154–164.
8. Reddiah K: **Sexuality and Spawning of Manx Pectinids.** *J Mar Biol Assoc Uk* 1962, **42**(3):683–8.
9. Wright WG: **Sex change in the Mollusca.** *Trends Ecol Evol* 1988, **3**(6):137–140.
10. Naimi A, Martinez AS, Specq ML, Diss B, Mathieu M, Sourdain P: **Molecular cloning and gene expression of Cg-Foxl2 during the development and the adult gametogenetic cycle in the oyster *Crassostrea gigas*.** *Comp Biochem Physiol B Biochem Mol Biol* 2009, **154**(1):134–142.
11. Santerre C, Sourdain P, Adeline B, Martinez AS: **Cg-SoxE and Cg-beta-catenin, two new potential actors of the sex-determining pathway in a hermaphrodite lophotrochozoan, the Pacific oyster *Crassostrea gigas*.** *Comp Biochem Physiol A Mol Integr Physiol* 2014, **167**:68–76.
12. Li HL, Zhang ZF, Bi Y, Yang DD, Zhang LT, Liu JG: **Expression Characteristics of beta-Catenin in Scallop *Chlamys farreri* Gonads and Its Role as a Potential Upstream Gene of Dax1 through Canonical Wnt Signalling Pathway Regulating the Spermatogenesis.** *Plos One* 2014, **9**(12).
13. Liu XL, Zhang ZF, Shao MY, Liu JG, Muhammad F: **Sexually dimorphic expression of foxl2 during gametogenesis in scallop *Chlamys farreri*, conserved with vertebrates.** *Dev Genes Evol* 2012, **222**(5):279–286.
14. Shi Y, Liu W, He M: **Proteome and Transcriptome Analysis of Ovary, Intersex Gonads, and Testis Reveals Potential Key Sex Reversal/Differentiation Genes and Mechanism in Scallop *Chlamys nobilis*.** *Mar Biotechnol (NY)* 2018, **20**(2):220–245.
15. Shi Y, Wang Q, He MX: **Molecular identification of dmrt2 and dmrt5 and effect of sex steroids on their expressions in *Chlamys nobilis*.** *Aquaculture* 2014, **426**:21–30.
16. Wang Q, He M: **Molecular characterization and analysis of a putative 5-HT receptor involved in reproduction process of the pearl oyster *Pinctada fucata*.** *Gen Comp Endocrinol* 2014, **204**:71–79.
17. Teaniniuraitemoana V, Huvet A, Levy P, Klopp C, Lhuillier E, Gaertner-Mazouni N, Gueguen Y, Le Moullac G: **Gonad transcriptome analysis of pearl oyster *Pinctada margaritifera*: identification of potential sex differentiation and sex determining genes.** *BMC Genomics* 2014, **15**:491.
18. Li Y, Zhang L, Sun Y, Ma X, Wang J, Li R, Zhang M, Wang S, Hu X, Bao Z: **Transcriptome Sequencing and Comparative Analysis of Ovary and Testis Identifies Potential Key Sex-Related Genes and Pathways in Scallop *Patinopecten yessoensis*.** *Mar Biotechnol (NY)* 2016, **18**(4):453–465.
19. Campbell KA: **The ecology of deep-sea hydrothermal vents.** *Science* 2000, **289**(5480):730–731.
20. Levin LA: **Ecology of cold seep sediments: Interactions of fauna with flow, chemistry and microbes.** *Oceanogr Mar Biol* 2005, **43**:1–46.
21. Sibuet M, Olu K: **Biogeography, biodiversity and fluid dependence of deep-sea cold-seep communities at active and passive margins.** *Deep-Sea Res Pt II* 1998, **45**(1-3):517+.
22. Dubilier N, Bergin C, Lott C: **Symbiotic diversity in marine animals: the art of harnessing chemosynthesis.** *Nat Rev Microbiol* 2008, **6**(10):725–740.
23. Bettencourt R, Pinheiro M, Egas C, Gomes P, Afonso M, Shank T, Santos RS: **High-throughput sequencing and analysis of the gill tissue transcriptome from the deep-sea hydrothermal vent mussel *Bathymodiolus azoricus*.** *BMC Genomics* 2010, **11**:559.
24. Boutet I, Jollivet D, Shillito B, Moraga D, Tanguy A: **Molecular identification of differentially regulated genes in the hydrothermal-vent species *Bathymodiolus thermophilus* and *Paralvinella pandorae* in response to temperature.** *BMC Genomics* 2009, **10**:222.
25. Bougerol M, Boutet I, LeGuen D, Jollivet D, Tanguy A: **Transcriptomic response of the hydrothermal mussel *Bathymodiolus azoricus* in experimental exposure to heavy metals is modulated by the Pgm genotype and symbiont content.** *Mar Genom* 2015, **21**:63–73.
26. Johnson SB, Won YJ, Harvey JBJ, Vrijenhoek RC: **A hybrid zone between *Bathymodiolus* mussel lineages from eastern Pacific hydrothermal vents.** *Bmc Evol Biol* 2013, **13**.
27. Sun J, Zhang Y, Xu T, Zhang Y, Mu HW, Zhang YJ, Lan Y, Fields CJ, Hui JHL, Zhang WP *et al*: **Adaptation to deep-sea chemosynthetic environments as revealed by mussel genomes.** *Nat Ecol Evol* 2017, **1**(5).
28. Xu T, Feng D, Tao J, Qiu JW: **A new species of deep-sea mussel (*Bivalvia: Mytilidae: Gigantidas*) from the South China Sea: Morphology, phylogenetic position, and gill-associated microbes.** *Deep-Sea Res Pt I* 2019, **146**:79–90.
29. Wisniewski JR, Zougman A, Nagaraj N, Mann M: **Universal sample preparation method for proteome analysis.** *Nat Methods* 2009, **6**(5):359–362.
30. Kachuk C, Stephen K, Doucette A: **Comparison of sodium dodecyl sulfate depletion techniques for proteome analysis by mass spectrometry.** *J Chromatogr A* 2015, **1418**:158–166.
31. Gillette MA, Satpathy S, Cao S, Dhanasekaran SM, Vasaikar SV, Petralia F, Li Y, Liang WW, Reva B *et al*: **Proteogenomic Characterization Reveals Therapeutic Vulnerabilities in Lung Adenocarcinoma.** *Cell* 2020, **182**(1):200–225 e235.
32. Zhang H, Liu T, Zhang Z, Payne SH, Zhang B, McDermott JE, Zhou JY, Petyuk VA, Chen L, Ray D *et al*: **Integrated Proteogenomic Characterization of Human High-Grade Serous Ovarian Cancer.** *Cell* 2016, **166**(3):755–765.
33. Lubner CA, Cox J, Lauterbach H, Fancke B, Selbach M, Tschopp J, Akira S, Wiegand M, Hochrein H, O'Keefe M *et al*: **Quantitative proteomics reveals subset-specific viral recognition in dendritic cells.** *Immunity* 2010, **32**(2):279–289.
34. Trapnell C, Pachter L, Salzberg SL: **TopHat: discovering splice junctions with RNA-Seq.** *Bioinformatics* 2009, **25**(9):1105–1111.
35. Wang T, Xiu J, Zhang Y, Wu J, Ma X, Wang Y, Guo G, Shang X: **Transcriptional Responses of *Candida albicans* to Antimicrobial Peptide MAF-1A.** *Front Microbiol* 2017, **8**:894.
36. Mao X, Cai T, Olyarchuk JG, Wei L: **Automated genome annotation and pathway identification using the KEGG Orthology (KO) as a controlled vocabulary.** *Bioinformatics* 2005, **21**(19):3787–3793.

37. Young MD, Wakefield MJ, Smyth GK, Oshlack A: **Gene ontology analysis for RNA-seq: accounting for selection bias.** *Genome Biol* 2010, **11**(2):R14.
38. Moon RT, Kohn AD, De Ferrari GV, Kaykas A: **WNT and beta-catenin signalling: diseases and therapies.** *Nat Rev Genet* 2004, **5**(9):691–701.
39. Tong Y, Zhang Y, Huang J, Xiao S, Zhang Y, Li J, Chen J, Yu Z: **Transcriptomics Analysis of *Crassostrea hongkongensis* for the Discovery of Reproduction-Related Genes.** *Plos One* 2015, **10**(8):e0134280.
40. Doniach T, Hodgkin J: **A Sex-Determining Gene, Fem-1, Required for Both Male and Hermaphrodite Development in *Caenorhabditis-Elegans*.** *Dev Biol* 1984, **106**(1):223–235.
41. Zhu Y, Rice CD, Pang Y, Pace M, Thomas P: **Cloning, expression, and characterization of a membrane progesterin receptor and evidence it is an intermediary in meiotic maturation of fish oocytes.** *Proc Natl Acad Sci U S A* 2003, **100**(5):2231–2236.
42. Harper CV, Barratt CL, Publicover SJ: **Stimulation of human spermatozoa with progesterone gradients to simulate approach to the oocyte. Induction of [Ca(2+)](i) oscillations and cyclical transitions in flagellar beating.** *J Biol Chem* 2004, **279**(44):46315–46325.
43. Lishko PV, Botchkina IL, Kirichok Y: **Progesterone activates the principal Ca2+ channel of human sperm.** *Nature* 2011, **471**(7338):387–391.
44. Sagare-Patil V, Galvankar M, Satiya M, Bhandari B, Gupta SK, Modi D: **Differential concentration and time dependent effects of progesterone on kinase activity, hyperactivation and acrosome reaction in human spermatozoa.** *Int J Androl* 2012, **35**(5):633–644.
45. Tosti E, Di Cosmo A, Cuomo A, Di Cristo C, Gagnaniello G: **Progesterone induces activation in *Octopus vulgaris* spermatozoa.** *Mol Reprod Dev* 2001, **59**(1):97–105.
46. Shen YJ, Kou Q, Chen WT, He SP, Yang M, Li XZ, Gan XN: **Comparative population structure of two dominant species, *Shinkaia crosnieri* (Munidopsidae: *Shinkaia*) and *Bathymodiolus platifrons* (Mytilidae: *Bathymodiolus*), inhabiting both deep-sea vent and cold seep inferred from mitochondrial multi-genes.** *Ecol Evol* 2016, **6**(11):3571–3582.
47. Galaviz-Hernandez C, Stagg C, de Ridder G, Tanaka TS, Ko MSH, Schlessinger D, Nagaraja R: **Plac8 and Plac9, novel placental-enriched genes identified through microarray analysis.** *Gene* 2003, **309**(2):81–89.
48. Ewing B, Green P: **Analysis of expressed sequence tags indicates 35,000 human genes.** *Nat Genet* 2000, **25**(2):232–234.
49. Pang Q, Gao L, Bai Y, Deng H, Han Y, Hu W, Zhang Y, Yuan S, Sun W, Lu Y *et al*: **Identification and characterization of a novel multifunctional placenta specific protein 8 in *Dugesia japonica*.** *Gene* 2017, **613**:1–9.
50. McIntosh JM, Ghomashchi F, Gelb MH, Dooley DJ, Stoehr SJ, Giordani AB, Naisbitt SR, Olivera BM: **Conodipine-M, a novel phospholipase A2 isolated from the venom of the marine snail *Conus magus*.** *J Biol Chem* 1995, **270**(8):3518–3526.
51. Wertheim B, Beukeboom LW, van de Zande L: **Polyploidy in animals: effects of gene expression on sex determination, evolution and ecology.** *Cytogenet Genome Res* 2013, **140**(2-4):256–269.
52. Markov GV, Tavares R, Dauphin-Villemant C, Demeneix BA, Baker ME, Laudet V: **Independent elaboration of steroid hormone signaling pathways in metazoans.** *Proc Natl Acad Sci U S A* 2009, **106**(29):11913–11918.
53. Matsumoto T, Masaoka T, Fujiwara A, Nakamura Y, Satoh N, Awaji M: **Reproduction-related genes in the pearl oyster genome.** *Zool Sci* 2013, **30**(10):826–850.
54. Schneider MR, Wolf E: **The epidermal growth factor receptor and its ligands in female reproduction: Insights from rodent models.** *Cytokine Growth F R* 2008, **19**(2):173–181.
55. Curry TE, Jr., Osteen KG: **The matrix metalloproteinase system: changes, regulation, and impact throughout the ovarian and uterine reproductive cycle.** *Endocr Rev* 2003, **24**(4):428–465.
56. Deady LD, Shen W, Mosure SA, Spradling AC, Sun JJ: **Matrix Metalloproteinase 2 Is Required for Ovulation and Corpus Luteum Formation in *Drosophila*.** *Plos Genet* 2015, **11**(2).
57. Carbajal L, Biswas A, Niswander LM, Prizant H, Hammes SR: **GPCR/EGFR Cross Talk Is Conserved in Gonadal and Adrenal Steroidogenesis but Is Uniquely Regulated by Matrix Metalloproteinases 2 and 9 in the Ovary.** *Mol Endocrinol* 2011, **25**(6):1055–1065.
58. Gomis-Ruth FX, Maskos K, Betz M, Bergner A, Huber R, Suzuki K, Yoshida N, Nagase H, Brew K, Bourenkov GP *et al*: **Mechanism of inhibition of the human matrix metalloproteinase stromelysin-1 by TIMP-1.** *Nature* 1997, **389**(6646):77–81.
59. Finn RN, Kristoffersen BA: **Vertebrate Vitellogenin Gene Duplication in Relation to the "3R Hypothesis": Correlation to the Pelagic Egg and the Oceanic Radiation of Teleosts.** *Plos One* 2007, **2**(1).
60. Williams VN, Reading BJ, Hiramatsu N, Amano H, Glassbrook N, Hara A, Sullivan CV: **Multiple vitellogenins and product yolk proteins in striped bass, *Morone saxatilis*: molecular characterization and processing during oocyte growth and maturation.** *Fish Physiol Biochem* 2014, **40**(2):395–415.
61. Garcia J, Munro ES, Monte MM, Fourrier MC, Whitelaw J, Smail DA, Ellis AE: **Atlantic salmon (*Salmo salar* L.) serum vitellogenin neutralises infectivity of infectious pancreatic necrosis virus (IPNV).** *Fish Shellfish Immunol* 2010, **29**(2):293–297.
62. Wang SH, Wang Y, Ma J, Ding YC, Zhang SC: **Phosvitin Plays a Critical Role in the Immunity of Zebrafish Embryos via Acting as a Pattern Recognition Receptor and an Antimicrobial Effector.** *Journal of Biological Chemistry* 2011, **286**(25):22653–22664.
63. Zhang S, Sun Y, Pang Q, Shi X: **Hemagglutinating and antibacterial activities of vitellogenin.** *Fish Shellfish Immunol* 2005, **19**(1):93–95.
64. Akasaka M, Harada Y, Sawada H: **Vitellogenin C-terminal fragments participate in fertilization as egg-coat binding partners of sperm trypsin-like proteases in the ascidian *Halocynthia roretzi*.** *Biochem Biophys Res Commun* 2010, **392**(4):479–484.
65. Akasaka M, Kato KH, Kitajima K, Sawada H: **Identification of novel isoforms of vitellogenin expressed in ascidian eggs.** *J Exp Zool B Mol Dev Evol* 2013, **320**(2):118–128.

66. Avilion AA, Nicolis SK, Pevny LH, Perez L, Vivian N, Lovell-Badge R: **Multipotent cell lineages in early mouse development depend on SOX2 function.** *Genes Dev* 2003, **17**(1):126–140.
67. Keramari M, Razavi J, Ingman KA, Patsch C, Edenhofer F, Ward CM, Kimber SJ: **Sox2 Is Essential for Formation of Trophectoderm in the Preimplantation Embryo.** *Plos One* 2010, **5**(11).
68. Archer TC, Jin J, Casey ES: **Interaction of Sox1, Sox2, Sox3 and Oct4 during primary neurogenesis.** *Dev Biol* 2011, **350**(2):429–440.
69. Ferri AL, Cavallaro M, Braida D, Di Cristofano A, Canta A, Vezzani A, Ottolenghi S, Pandolfi PP, Sala M, DeBiasi S *et al.*: **Sox2 deficiency causes neurodegeneration and impaired neurogenesis in the adult mouse brain.** *Development* 2004, **131**(15):3805–3819.
70. Cimadamore F, Amador-Arjona A, Chen C, Huang CT, Terskikh AV: **SOX2-LIN28/let-7 pathway regulates proliferation and neurogenesis in neural precursors.** *Proc Natl Acad Sci U S A* 2013, **110**(32):E3017-3026.
71. Ellis P, Fagan BM, Magness ST, Hutton S, Taranova O, Hayashi S, McMahon A, Rao M, Pevny L: **SOX2, a persistent marker for multipotential neural stem cells derived from embryonic stem cells, the embryo or the adult.** *Dev Neurosci* 2004, **26**(2-4):148–165.
72. Masui S, Nakatake Y, Toyooka Y, Shimosato D, Yagi R, Takahashi K, Okochi H, Okuda A, Matoba R, Sharov AA *et al.*: **Pluripotency governed by Sox2 via regulation of Oct3/4 expression in mouse embryonic stem cells.** *Nat Cell Biol* 2007, **9**(6):625–635.
73. Liu K, Lin B, Zhao M, Yang X, Chen M, Gao A, Liu F, Que J, Lan X: **The multiple roles for Sox2 in stem cell maintenance and tumorigenesis.** *Cell Signal* 2013, **25**(5):1264–1271.
74. Campolo F, Gori M, Favaro R, Nicolis S, Pellegrini M, Botti F, Rossi P, Jannini EA, Dolci S: **Essential role of Sox2 for the establishment and maintenance of the germ cell line.** *Stem Cells* 2013, **31**(7):1408–1421.
75. Patra SK, Chakrapani V, Panda RP, Mohapatra C, Jayasankar P, Barman HK: **First evidence of molecular characterization of rohu carp Sox2 gene being expressed in proliferating spermatogonial cells.** *Theriogenology* 2015, **84**(2):268–276.
76. Liang S, Liu D, Li X, Wei M, Yu X, Li Q, Ma H, Zhang Z, Qin Z: **SOX2 participates in spermatogenesis of Zhikong scallop *Chlamys farreri*.** *Sci Rep* 2019, **9**(1):76.
77. Cui J, Shen X, Zhao H, Nagahama Y: **Genome-Wide Analysis of Sox Genes in Medaka (*Oryzias latipes*) and Their Expression Pattern in Embryonic Development.** *Cytogenetic and Genome Research* 2011, **134**(4):283–294.
78. Gao J, Wang Z, Shao K, Fan L, Yang L, Song H, Liu M, Wang Z, Wang X, Zhang Q: **Identification and characterization of a Sox2 homolog in the Japanese flounder *Paralichthys olivaceus*.** *Gene* 2014, **544**(2):165–176.
79. Garcia-Ortiz JE, Pelosi E, Omari S, Nedorezov T, Piao Y, Karmazin J, Uda M, Cao A, Cole SW, Forabosco A *et al.*: **Foxl2 functions in sex determination and histogenesis throughout mouse ovary development.** *BMC Dev Biol* 2009, **9**:36.
80. Kocer A, Pinheiro I, Pannetier M, Renault L, Parma P, Radi O, Kim KA, Camerino G, Pailhoux E: **R-spondin1 and FOXL2 act into two distinct cellular types during goat ovarian differentiation.** *Bmc Developmental Biology* 2008, **8**.
81. Uhlenhaut NH, Jakob S, Anlag K, Eisenberger T, Sekido R, Kress J, Treier AC, Klugmann C, Klasen C, Holter NI *et al.*: **Somatic sex reprogramming of adult ovaries to testes by FOXL2 ablation.** *Cell* 2009, **139**(6):1130–1142.
82. Zhou L, Charkraborty T, Yu X, Wu L, Liu G, Mohapatra S, Wang D, Nagahama Y: **R-spondins are involved in the ovarian differentiation in a teleost, medaka (*Oryzias latipes*).** *BMC Dev Biol* 2012, **12**:36.

Figures

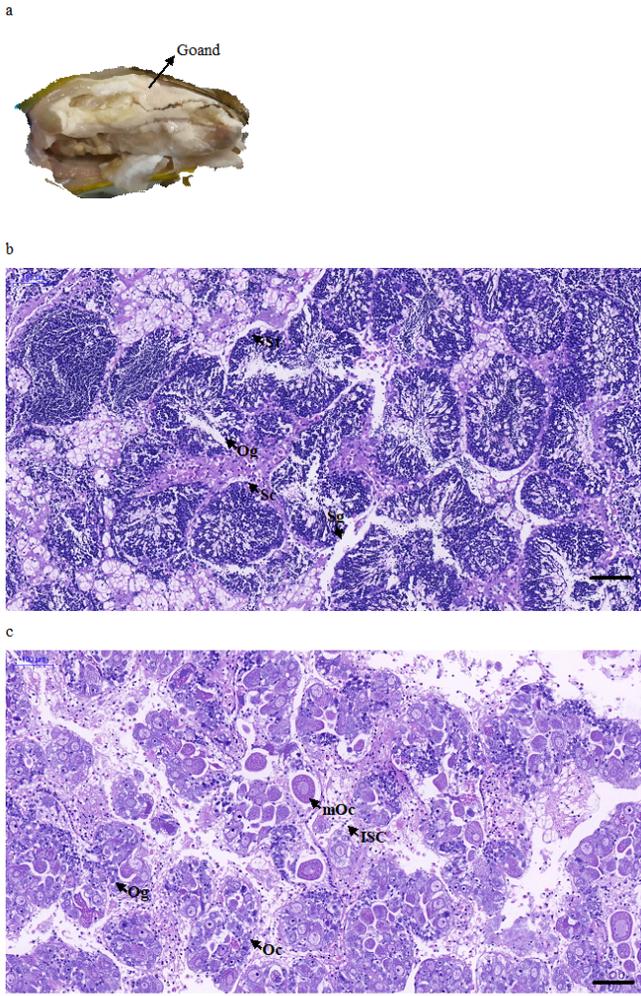


Figure 1
 The appearance and histology of gonads in *G. haimaensis*. a. The appearance of gonads; b. Histology of male gonads; c. Histology of female gonads. ISC, intragonadal somatic cell; Og, oogonium; Oc, oocyte; mOc, mature oocyte; Sg, spermatogonium; Sc, spermatocyte; St, spermatid; Scale bar = 100 μ m.

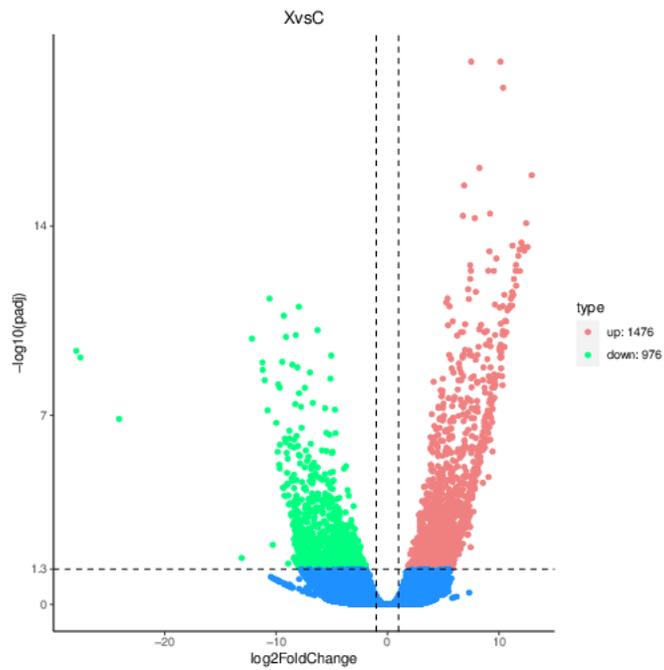


Figure 2

DEGs (differentially expressed genes) identified from six gonad groups. Red indicates upregulated genes and green indicates downregulated genes among the six gonad groups.

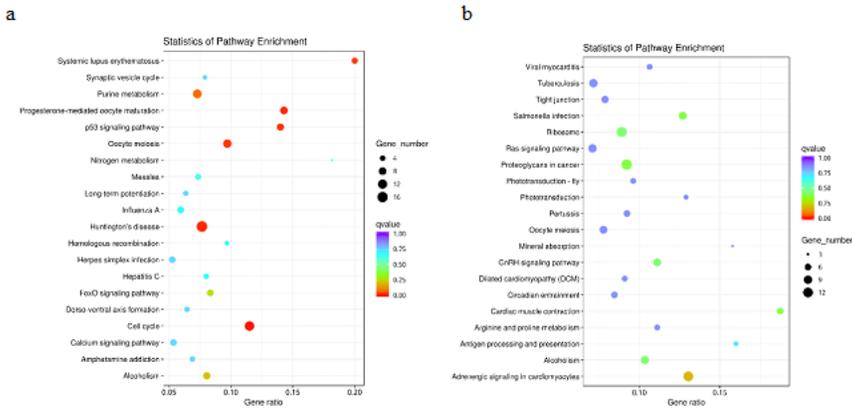


Figure 3
KEGG pathway enrichment scatter diagram based on the transcriptome data. The vertical axis represents the pathway name, and the horizontal axis represents the Rich factor corresponding to the pathway. The magnitude of the q value is represented by the colour of the dots; the smaller the q value, the closer the colour is to red. The number of DEGs (differentially expressed genes) within each pathway is indicated by the dot size.

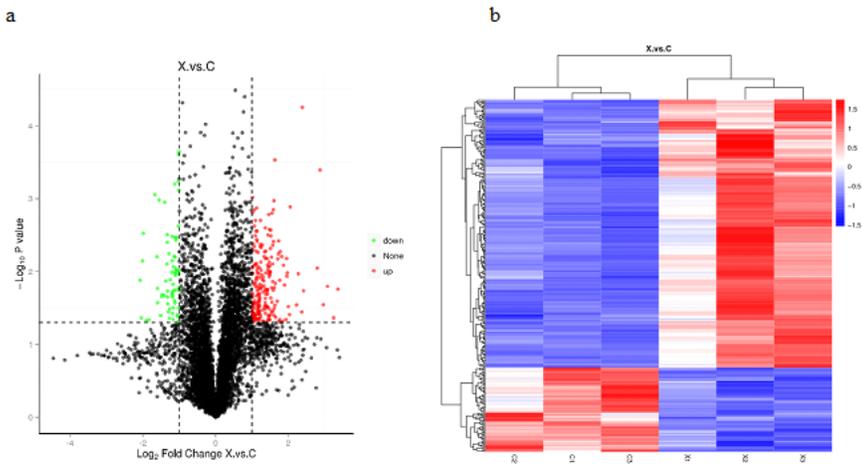
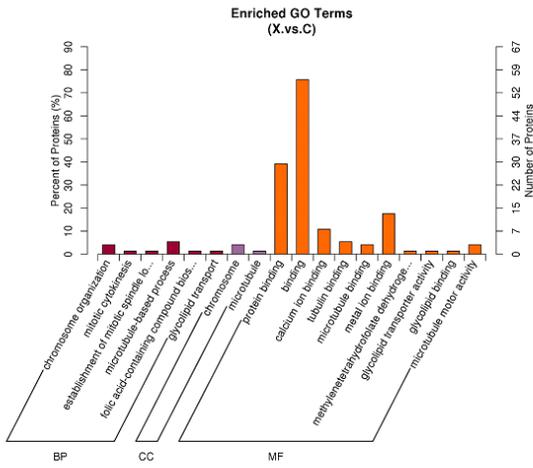


Figure 4
a. Volcano plot of DEPs. Black represents proteins with no significant differences, red represents upregulated proteins, and green represents downregulated proteins. b. Heatmap of DEP clustering. The longitudinal axis indicates the clustering of samples, and the transverse axis shows the clustering of proteins. The shorter the clustering branch, the higher the similarity.

a



b

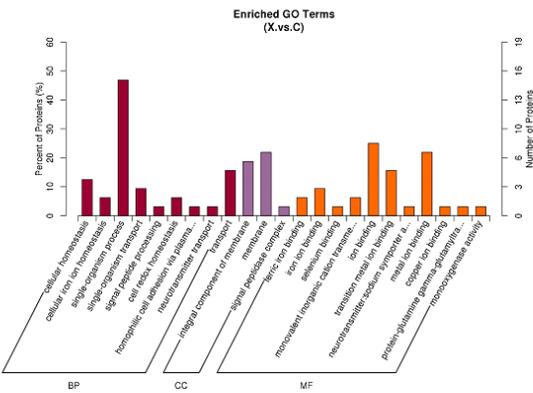
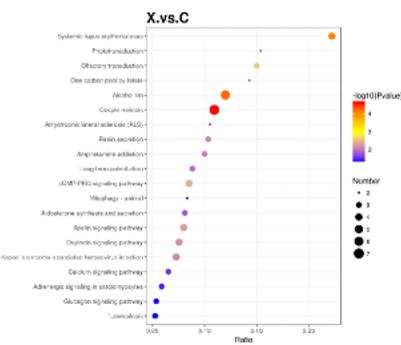


Figure 5

Enriched GO terms for the proteome. a. GO enrichment between DEPs of M and F groups. b. GO terms of proteins overexpressed in F compared with M.

a



b

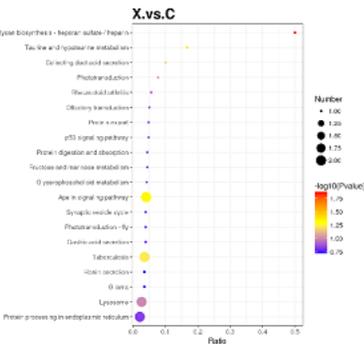


Figure 6

KEGG pathway enrichment scatter diagram for the proteome. The vertical axis represents the pathway name, and the horizontal axis represents the Rich factor corresponding to the pathway. The magnitude of the q value is represented by the colour of the dots; the smaller the q value, the closer the colour is to red. The number of DEGs (differentially expressed genes) within each pathway is indicated by the dot size.

Subcellular localization

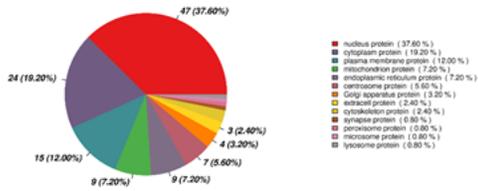


Figure 7

Subcellular localisation analysis of DEPs The subcellular localisation was analysed for DEPs for each comparison.

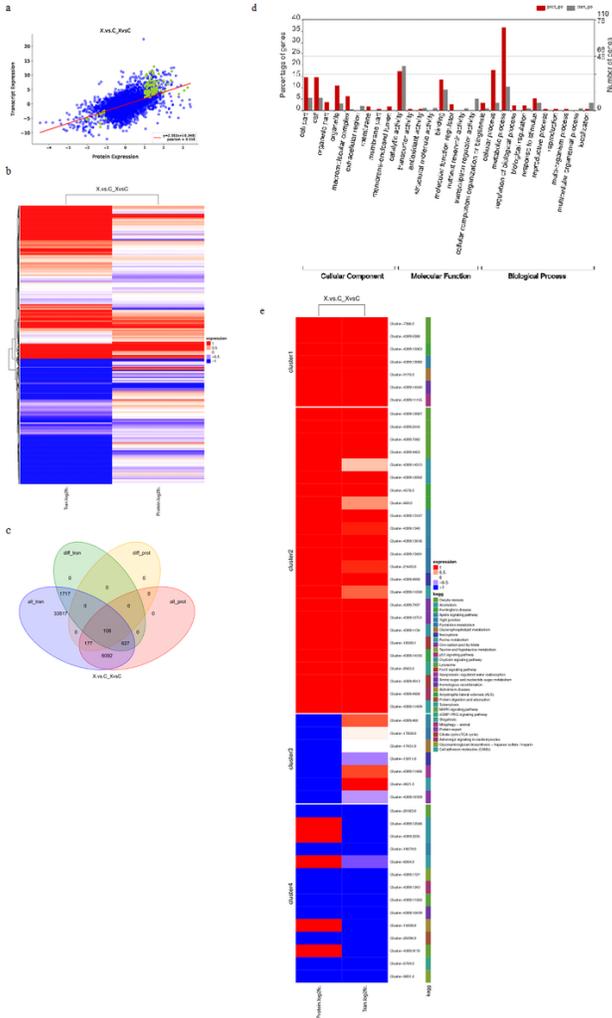


Figure 8

Proteome and transcriptome association analysis. a. Correlation analysis of transcriptome and proteome data. Green represents proteins with significant differences in expression, and blue represents proteins with no significant differences in expression. The horizontal axis represents multiple differences for corresponding proteins identified from the proteome data (log2 value), and the vertical axis represents multiple differences for corresponding genes identified from the transcriptome data (log2 value). b. Heatmap of the correlation analysis of transcriptome and proteome data. Red represents upregulated genes and blue represents downregulated genes. c. Venn diagrams from the comparison of transcriptome and proteome data. all_tran, all genes identified from the transcriptome data; diff_tran, DEGs (differentially expressed genes) identified from the transcriptome data; all_prot, all proteins identified from the proteome data; diff_prot, DEPs identified from the proteome data. d. Correlation analysis of GO functional enrichment. Red columns represent the GO enrichment results for the proteome; grey columns represent the GO enrichment results for the transcriptome. e. Heatmap of KEGG pathway enrichment. Red represents upregulated KEGG pathways; blue represents downregulated KEGG pathways.

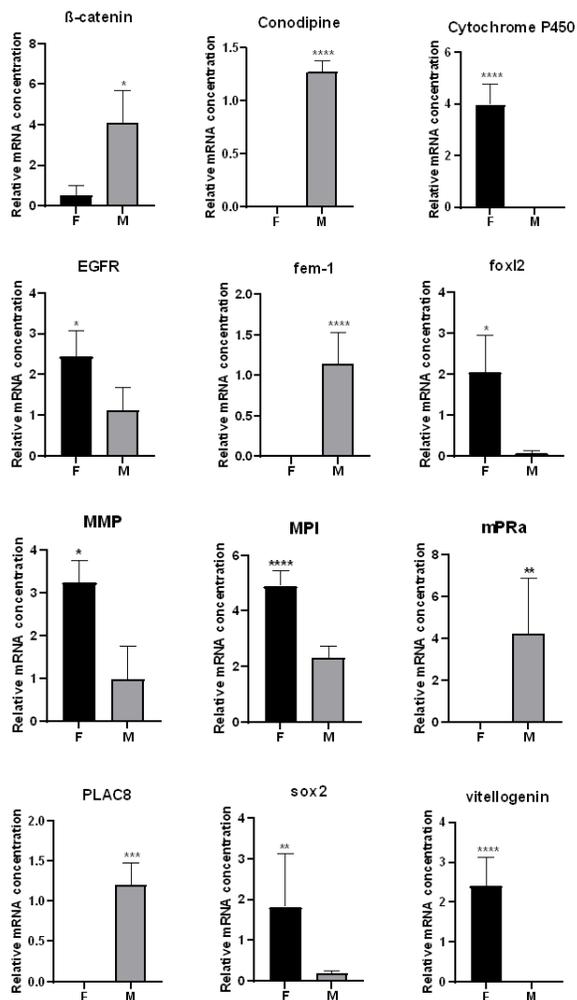


Figure 9

mRNA expression levels for 12 genes in gonads of males and females examined by qPCR to verify the proteome and RNA-Seq data. mRNA levels were quantified by qPCR. 60s RP-L15 and β -actin served as reference genes to normalise expression levels. Results are expressed as fold change (FC). Each bar represents the mean standard deviation (SD) of three samples. Significant differences are indicated by an asterisk (* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$, **** $p < 0.001$). EGFR, epidermal growth factor receptor; MMP, matrix metalloproteinase; MPI, metalloproteinase inhibitor; mPR α , membrane progesterin receptor alpha; PLAC8, placenta-specific gene 8.

Supplementary Files

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

- [TableS1.docx](#)
- [S2.xls](#)
- [S3.xls](#)