Meta-analysis of gonadal transcriptomes in protogynous fishes provides novel perspective into sex change machinery

Ryo Nozu[1,2], Mitsutaka Kadota[3,4], Masaru Nakamura[5], Shigehiro Kuraku[6,7], Hidemasa Bono[1,2#]

[1] Laboratory of Genome Informatics, Graduate School of Integrated Sciences for Life, Hiroshima University, 3-10-23 Kagamiyama, Higashi-Hiroshima, Hiroshima 739-0046, Japan

[2] Laboratory of BioDX, Genome Editing Innovation Center, Hiroshima University, Hiroshima, Japan

[3] Laboratory for Phyloinformatics, RIKEN Center for Biosystems Dynamics Research (BDR), Kobe, Japan

[4] Laboratory for Developmental Genome System, RIKEN Center for Biosystems Dynamics Research (BDR), Kobe, Japan

[5] Okinawa Churashima Research Center, Okinawa Churashima Foundation, Motobu-cho, Okinawa, Japan

[6] Molecular Life History Laboratory, Department of Genomics and Evolutionary Biology, National Institute of Genetics, Mishima, Japan

[7] Department of Genetics, Sokendai (Graduate University for Advanced Studies), Mishima, Japan.

Email:

RN, rnozu22@hiroshima-u.ac.jp; MK, mitsutaka.kadota@riken.jp; MN, m-nakamura@okichura.jp; SK, skuraku@nig.ac.jp

[#]To whom correspondence may be addressed: Hidemasa Bono

bonohu@hiroshima-u.ac.jp

Abstract

Background: Protogynous fishes can change sex from female to male; concomitantly, their ovarian tissue is completely replaced by testicular tissue. The high functional conservation of gonads across organisms suggests the existence of as-yet-undefined common and essential factors in the regulation of such gonadal transformation. To identify these factors, we conducted a meta-analysis of public and newly derived gonadal transcriptome data from seven species.

Results: Classical sex change-related genes, including *cyp19a1a*, *foxl2a*, *hsd17b1*, *amh*, and *gsdf*, were differentially expressed between the ovary and transitional gonads, which corroborated previous reports on protogynous fishes. Moreover, we identified novel factors not previously associated with gonadal sex change in fish. Notably, the expression patterns of acetoacetyl-CoA synthetase (*aacs*) and apolipoprotein Eb (*apoeb*), which are involved in cholesterol synthesis and transport, respectively, suggest that the levels of cholesterol, a precursor of steroid hormones crucial for sex change, are decreased upon sex change onset in the gonads. Thus, cholesterol dynamics may influence gonadal transformation in protogynous fishes.

Conclusion: This finding indicates the potential of transcriptome meta-analysis as a valuable approach to uncover previously overlooked mechanisms underlying sex change. We anticipate that further data-driven research utilizing these novel sex change-related factors will open new horizons in sex change research.

Keywords: Meta-analysis, Protogynous fishes, Gonad, Sex change, Genome, Transcriptome

Background

Sequential hermaphroditism, a phenomenon in which individuals undergo sex change during their lifetime, has been extensively recognized in teleosts over the past several decades [1]. Specifically, protogyny (i.e., the transformation from female to male) is prevalent in 20 families [2]. Phylogenetic analyses have shown that this characteristic has been independently acquired in various evolutionary lineages [3, 4]. On the other hand, the complete replacement of ovarian tissue with testicular tissue is commonly observed during protogynous sex change [5-9], the intricate regulatory mechanisms responsible for inducing these remarkable transformations in gonadal tissue remain incompletely understood. Furthermore, considering the high functional conservation of gonads across organisms, the existence of a regulatory mechanism governing gonadal transformation comprising common and essential factors across protogynous fishes is plausible; however, a comprehensive perspective of this process remains elusive.

Accordingly, elucidation of the molecular mechanisms underlying the gametogenesis switch in sex-changing fish has subsequently become a focus of research. In particular, the expression patterns of multiple sex-related genes during sex change within the family Labridae, which serves as a model for protogynous sex change, have been explored [10-15]. Recently, next-generation sequencing technology has enabled researchers to conduct comprehensive analyses of gene expression, allowing the identification of global gene expression profiles during gonadal transformation in protogynous fish belonging to Labriformes, Perciformes, and Synbranchiformes [16-20]. Thus, these results suggest the involvement of several pivotal genes in the ovary-to-testis transformation in sex-changing fish.

For example, a repertoire of sex-related genes, including cytochrome P450, family 19, subfamily A, polypeptide 1a (*cyp19a1a*), forkhead box L2a (*foxl2a*), folliculogenesis specific bHLH transcription factor (*figla*), anti-Müllerian hormone (*amh*), gonadal somatic cell derived factor (*gsdf*), doublesex and mab-3 related transcription factor 1 (*dmrt1*) were identified as

differentially expressed genes (DEGs) during sex change across protogynous fishes. Notably, the gonadal expression of *cyp19a1a* and *foxl2a*, which are key players in maintaining ovarian status, is reduced upon the onset of sex change in multiple species [14, 21, 22]. *cyp19a1a* encodes aromatase, an enzyme responsible for the conversion of androgens (e.g., testosterone) to estrogens (e.g., estradiol-17ß; E2), whereas *foxl2a* encodes a transcription factor that targets and up-regulates *cyp19a1a* in teleost fishes [23, 24]. Furthermore, the initiation of female-to-male sex change, triggered by a rapid decline in E2 levels [6], is well-coordinated with the expression profiles of these genes. Conversely, *amh*, *gsdf*, and *dmrt1* are crucial for testicular differentiation and development in teleost fishes [25-31]. In several sex-changing species, the upregulation of these genes coincides with the emergence of testicular tissue during transition, suggesting that these genes are involved in testicular formation during gonadal sex change [10-12, 14, 32].

Transcriptome analysis during gonadal transformation has also revealed differential expression of several genes not previously implicated in sex determination or differentiation pathways. In particular, upregulation of apolipoprotein Eb (*apoeb*), ectonucleotide pyrophosphatase/phosphodiesterase 2 (*enpp2*), and l-rhamnose-binding lectin CSL2 (*csl2*) expression, accompanied by downregulation of galectin-3 (*gal3*), was observed during the early phase of gonadal sex change in the protogynous Asian swamp eel *Monopterus albus*. These findings suggest that these factors are involved in gonadal transformation, particularly during the initiation of sex change [16, 19]. However, verifying their direct involvement in the sex change process remains challenging.

Recently, by amalgamating transcriptome data from multiple studies, meta-analysis has attracted attention as a powerful tool capable of providing novel insights overlooked by traditional hypothesis-driven research methods[33-35]. Therefore, the objective of the present study was to identify the shared factors involved in gonadal transformation across species using a meta-analysis of publicly available transcriptome data from multiple experiments and

species. To improve the performance of the meta-analysis, we included data (*de novo* genome assembly data and RNA sequencing (RNA-Seq) data from gonads undergoing sex change) generated in-house from an additional well-studied species, the three-spot wrasse, *Halichoeres trimaculatus*, a protogynous species belonging to the family Labridae [10-13]. Through meta-analysis of RNA-Seq data encompassing seven protogynous species from three orders, we identified 168 DEGs between the transitional gonads and ovaries (Fig. 1). Subsequently, we reevaluated the relationship between these factors and gonadal sex change, providing a novel perspective regarding the mechanisms underlying this process.

Results

De novo genome assembly and transcriptome of gonadal tissue in the three-spot wrasse.

High-throughput genome sequence data were obtained from a locally obtained female three-spot wrasse from Okinawa, Japan. The obtained HiFi data, which comprised approximately 32 Gbp and encompassed approximately 1.9 million DNA sequences, were assembled using the hifiasm program [36]. Subsequently, the acquired Hi-C data were used for scaffolding, resulting in the generation of 227 scaffolds. The longest scaffold measured 44.2 Mbp, with an N50 length (the shortest of the lengths covering 50% of the genome) of 37.9 Mbp, contributing to a total assembly length of approximately 849.9 Mbp. Assessment using BUSCO [37] revealed a completeness > 98%, indicating a high level of coverage and contiguity in the *de novo* genome assembly of the three-spot wrasse (Table 1).

The RNA-Seq reads obtained from transitional gonads and ovaries of additional females were mapped to the genome assembly, and the obtained mapping results were used to assemble the transcriptome for the annotation file (“merged.gtf”). This process yielded 27,573 transcript sequences (Table 2). Subsequently, 24,695 coding sequences (protein sequences) were predicted from the set of transcript sequences using TransDecoder [38].

Data curation of public RNA-Seq data and genomic information. We downloaded Sequence Read Archive (SRA) data from six fish species for which sex changes have been reported [2]: bluehead wrasse *Thalassoma bifasciatum* (order Labriformes), orange-spotted grouper *Epinephelus coioides* (order Perciformes), Hong Kong grouper *E. akaara* (order Perciformes), black seabass *Centropristis striata* (order Perciformes); Asian swamp eel *Monopterus albus* (order Synbranchiformes), zig-zag eel *Mastacembelus armatus* (order Synbranchiformes). Furthermore, using the metadata associated with the SRA data, we curated the RNA-Seq data into comparable sample pairs of intersexual (transitional) gonads and ovaries (a total of 31 paired data, including three-spot wrasse). The data pairs analyzed in this study are listed in Supplementary Table 1.

The genome sequences of five of the seven species utilized in this study are publicly available (Table 2) and served as reference genomes for read mapping. The genome sequences of *T. bifasciatum* (GCA_008086565.1), *E. coioides* (GCA_900536245.1), *M. albus* (GCF_001952655.1), and *M. armatus* (GCF_900324485.2) were retrieved from the National Center for Biotechnology Information (NCBI) Genome Database (accessed in October 2022). We also referenced the genome information of *E. akaara* reported by Ge et al., 2019 [39].

Transcriptome assembly and inference of ortholog relationships. For the five species with public reference genomes, the number of transcript sequences obtained using StringTie [40] with the "--merge" option is shown in Table 2. For black sea bass, (without a reference genome), the number of contiguous sequences assembled using Trinity [41] was 594,726. Among these, 95,941 transcript sequences corresponding to the predicted coding sequences (CDS) were specified as references for expression quantification using Salmon [42]. The number of protein sequences predicted using TransDecoder is shown in Table 2. Orthology inference and consolidation of orthologs within the three-spot wrasse identified 7,289

orthologs that were common to all species used in this study (Table 2), and thus were considered as candidate sex-change factors.

Identification of differentially expressed transcripts. The total intersexual gonad and ovary score (IO-score) was calculated for these 7,289 factors (Supplementary Table 2, IO-score calculation; see Methods). Among these, 99 factors were upregulated (Fig. 2a, Fig. 3, Supplementary Table 3) and 69 factors were downregulated (Fig. 2a, Fig. 4, Supplementary Table 4). Whereas the total IO-score of these factors was occasionally derived from a single species (up: *cep97*, *zmat2*, *tmem237a*, *supt5h*; down: *znf16l*, *lrrc8db*, *sh2b1*, *tmem263*, *cyldb*, *manea*, *nol11*), in most cases, the total IO-scores were based on contributions from two or more species. Notably, the upregulated genes included *amh* and *gsdf*, which have been previously reported to play important roles in gonadal sex change, particularly in testicular formation and development, in some protogynous fish. Conversely, the downregulated genes included *cyp19a1a*, *hsd17b1*, and *foxl2a*, which are crucial for the onset of sex change (Table 3).

Enrichment analysis of DEGs. Sequence homology search of the three-spot wrasse proteins using ggsearch36 [43] against the reference zebrafish proteins (Ensembl Release 108) [44] resulted in the identification of 90 homologs among the 99 upregulated factors (Supplementary Table 3). No corresponding zebrafish orthologs were found for the remaining nine factors that were common to all species analyzed in this study. The results of enrichment analysis using the 90 factors that could be converted to zebrafish orthologs are shown in Fig. 2b. The Gene Ontology (GO) terms “homologous chromosome pairing at meiosis” and “germ cell development,” which are related to reproduction, were significantly enriched among the upregulated factors. The specific genes annotated with these GO terms included *hormad1*, *terb2*, *si:ch211-199g17.9*, *gsdf*, and *sycp3* (Table 3).

We successfully obtained the corresponding zebrafish homologs for all of the 69 downregulated factors. The results of enrichment analysis are shown in Fig. 2c. Among the downregulated factors, GO terms related to the regulation of steroid hormone levels, including "regulation of hormone levels" and "steroid metabolic process," were significantly enriched. Specific genes annotated with these enriched GO terms included *cyp19a1a*, *aacs*, *hsd17b1*, *ece1*, *amacr*, and *sdr42e1* (Table 3).

Pathways mapping of candidate sex-change factors. Visualization of cholesterol and steroid hormone biosynthesis pathways was conducted based on the GO terms obtained from the enrichment analysis. By referencing pathway information for teleost fishes, pathway diagrams were constructed and loaded with 7,289 factors converted to Ensembl zebrafish gene IDs (see Methods). The results showed the mapping of most enzymes involved in the pathway from acetyl-CoA to cholesterol synthesis (Fig. 5). Similarly, when mapping the steroid hormone synthesis pathway, it was confirmed that enzymes involved in the synthesis pathway from cholesterol to estradiol were included among the candidate sex-change factors (supplementary figure 2).

Discussion

In this study, we conducted a meta-analysis of gonadal RNA-Seq data from seven protogynous fish species using a unified analysis pipeline and threshold. The aim was to compare the transcriptomes of ovaries and sexual transition gonads, with the goal of identifying common factors involved in protogynous gonadal sex change. Toward this end, we first assembled the *de novo* genome of the three-spot wrasse. The assembly statistics of the three-spot wrasse (Table 1) were comparable to those of other Labridae species, such as humphead wrasse (*Cheilinus undulatus*), ballan wrasse (*Labrus bergylta*), and New Zealand spotty (*Notolabrus*

celidotus), previously deposited in the NCBI RefSeq database. These results indicated that the *de novo* genome assembly of the three-spot wrasse exhibited high quality, enabling us to perform expression analysis and orthology inference with confidence.

Through this meta-analysis, we successfully identified sex-related factors including *cyp19a1a*, *foxl2a*, *hsd17b1*, *amh*, and *gsdf*, that had been previously reported to play a role in sex change in various protogynous fish species[10, 14, 17, 19]. Furthermore, our enrichment analysis of downregulated factors highlighted a term related to the "regulation of steroid hormone production." This finding connects to those of previous reports, which indicate that a rapid decrease in estrogen levels is crucial for female-to-male sex change [6, 45, 46]. Thus, by pooling data across species, we were able to identify common factors involved in gonadal sex change. Notably, our meta-analysis demonstrates the potential of utilizing transcriptome data from multiple experiments to gain insights regarding the mechanisms underlying sex change in fish.

Our findings demonstrate that *aacs*, annotated with the GO term "regulation of hormone levels," is downregulated during sex change. Notably, no previous reports have suggested the involvement of *aacs* in sex change, indicating that the encoded protein (acetoacetyl-CoA synthetase; Aacs) may represent a novel factor associated with this process. Typically, AACS plays a critical role in cholesterol supply. Studies in mice have shown that *Aacs* knockdown decreases total serum cholesterol levels [47]. Additionally, in rat adrenal glands, inhibition of *AACS* expression through hypermethylation suppresses cholesterol supply and steroid synthesis [48]. Moreover, apolipoprotein Eb (encoded by *apoeb*, listed among the upregulated factors), has been suggested to participate in cholesterol transportation in zebrafish [49]. Apolipoprotein E, encoded by *APOE* which is the human ortholog of *apoeb*, serves as a crucial lipid transport protein involved in the modulation of cholesterol metabolism and is also synthesized within peripheral tissues [50]. In addition, a previous study indicates that APOE gene expression exerts suppressive effects on steroid production in adrenal cells [51].

Specifically, the high expression of APOE in tissues with active steroid production, including the adrenal glands, appears to shift cholesterol metabolism from delivery to the steroidogenic pathway to storage of esterified cholesterol, resulting in a decrease in steroid production levels [52]. These observations strongly support the hypothesis that the downregulation of *aacs* and the upregulation of *apoeb* during sex change contribute to suppression of the cholesterol supply and entire steroidogenesis.

Cholesterol serves as a precursor for steroid hormones, including estrogens, androgens, and glucocorticoids [53-55], which play crucial roles in protogynous sex change. Previous studies have shown that a decline in endogenous estrogens, particularly E2, triggers the transition from ovary to testis in protogynous fish [6, 45, 46]. E2 is synthesized from testosterone via an aromatase enzyme, encoded by *cyp19a1a*. Downregulation of *cyp19a1a* expression at the onset of sex change leads to reduced E2 production. Studies on *M. albus* and *T. bifasciatum* have revealed higher DNA methylation levels of the *cyp19a1a* gene upstream region in the testes than those in the ovaries, suggesting that epigenetic modifications may contribute to the suppression of *cyp19a1a* expression during gonadal sex change [17, 21]. Our present findings could contribute one more “piece of the puzzle”, suggesting that the trigger of gonadal transformation may be attributed not only to the decreased E2 production by the downregulation of aromatase activity but also to the suppression of the cholesterol supply, the precursor of steroid hormones, leading to the inhibition of the entire steroid biosynthesis. Interestingly, a decrease in the expression of lipoprotein lipase (*lpl*) has been observed (Fig 5, supplementary table 4). Lipoprotein lipase is involved in the biosynthetic process of fatty acid which serves as one of the sources of acetyl-CoA, a precursor for cholesterol synthesis. This suggests the possibility of a concurrent suppression of fatty acid supply to the gonads, might be leading to support the hypothesis. Although this hypothesis requires further investigation and an understanding of cholesterol dynamics during sex change, additional exploration is warranted. Notably, a study on medaka (*Oryzias latipes*) subjected to starvation conditions

implied a potential link between masculinization and lower cholesterol levels [56]. Specifically, female-to-male sex reversal was observed in starved medaka larvae, a gonochoristic fish, together with decreased cholesterol levels in the starved samples. However, the precise pathways underlying masculinization and cholesterol dynamics remain unknown. In general, cholesterol is synthesized not only in the liver and adrenal glands but also in the gonads [57-59]. Therefore, gaining a comprehensive understanding of the lipid metabolic mechanisms in fish gonads, including cholesterol synthesis and transportation during sex change, may provide valuable insights regarding the mechanisms underlying gonadal transformation. To clarify this perspective, the remaining challenge lies in providing experimental evidence regarding whether it is possible to control gonadal transformation by actually regulating cholesterol levels within the body.

Conclusions

In this study, our objective was to compare the transcriptomes of sex-transitional gonads with those of ovaries using a unified pipeline to identify common factors involved in gonadal sex change across protogynous fishes. We successfully constructed a high-quality de novo genome assembly applicable for genome-wide analysis in the protogynous three-spot wrasse and acquired newly generated gonadal transcriptome data. Leveraging this information, we conducted a comprehensive meta-analysis by integrating both public and newly derived gonadal transcriptome data from multiple protogynous fishes spanning three orders. Through this analysis, we identified 168 differentially expressed factors between the two gonadal states, suggesting their potential universal role in gonadal sex change across diverse taxa of protogynous fishes. Notably, we identified novel factors that were not previously associated with gonadal sex change. In particular, we identified cholesterol dynamics as an important factor regulating the sex change process. Moreover, this finding highlights the potential of

transcriptomics meta-analysis as a valuable approach to uncover overlooked mechanisms underlying conserved biological processes including sex change. We anticipate that further data-driven research utilizing these novel sex change-related factors will open new horizons in sex change research.

Methods

Animals. The adult female three-spot wrasse specimens collected in the field by local fisherman in the northern region of Okinawa, Japan were purchased and employed for genome and RNA sequencing. These specimens were subsequently cultivated in 1000 L tanks receiving a continuous supply of unregulated seawater, and were subjected to natural photoperiod conditions until sampling.

The maintenance and handling of fish in this study were conducted in accordance with the "Guidelines for Animal Experiments of the Okinawa Churashima Foundation," with the same consideration for animal care and welfare as that for higher vertebrates, including reptiles, birds, and mammals. Animal ethical review and approval were not required for the present study because the guidelines stipulated that approval from the Institutional Animal Care and Use Committee of Okinawa Churashima Foundation is required for "higher" vertebrates, and is waived for "lower" vertebrates including fish.

Genome sequencing and assembly for the three-spot wrasse. Muscle tissue from a mature female individual was used for genome sequencing. The specimen was randomly selected from cultivated tank. After anesthetizing the fish with 0.05% v/v 2-phenoxyethanol (FUJIFILM Wako Pure Chemical), the muscle tissue was dissected, immediately frozen in liquid nitrogen, and stored at −80 °C until genomic DNA extraction. High-molecular-weight DNA was extracted from muscle using a NucleoBond AXG column (Macherey–Nagel), followed by

purification with phenol–chloroform. The concentration of the extracted DNA was measured using Qubit (Thermo Fisher Scientific Inc.), and its size distribution was first analyzed with a TapeStation 4200 (Agilent Technologies) to ensure high integrity, and subsequently analyzed via pulse-field gel electrophoresis on a CHEF DR-II system (BioRad) to ensure a size range of 20–100 kb.

To obtain long-read sequence data, a SMRT sequence library was constructed using the SMRTbell Express Template Prep Kit 2.0 (PacBio) and sequenced on a PacBio Sequel II system. The sequencing output was processed to generate circular consensus sequences to obtain HiFi sequence reads. Adapter sequences were removed from the HiFi reads using HiFiAdapterFilt [60]. The obtained HiFi sequence reads were assembled using the hifiasm (v0.16.1) assembler [36] with default parameters. The obtained contigs were first scaffolded with the P_RNA_scaffolder [61] and further scaffolded using Hi-C data as follows. The Hi-C library was prepared with the in situ method using the muscle tissue of the individual used for DNA extraction according to the iconHi-C protocol [62], employing the restriction enzymes DpnII and HinfI, and sequenced on a HiSeq X sequencing platform (Illumina). The obtained Hi-C read pairs, amounting to 232 million read pairs, were aligned to the HiFi sequence contigs using the Juicer program [63], and the HiFi sequence contigs were scaffolded via the 3d-dna (version 201008) to be consistent with the chromatin contact profiles through manual curation on JuiceBox v1.11.08 [64]. The continuity and completeness of the resulting genome assembly were assessed using the webserver gVolante [65], in which the pipeline BUSCO (v5.4, reference datasets: vertebrata_odb10) was implemented [37].

RNA-Seq of gonadal tissue in the three-spot wrasse. Based on previous studies [10, 45], we triggered a sex change in three-spot wrasse n=5 females (body length; 8.0–9.3 cm, body weight; 12.4–20.9 g) to obtain transitional gonads through treatment with an aromatase inhibitor (AI), which exerts inhibitory effects on the enzyme responsible for estrogen synthesis.

The collected gonads were cut into small pieces, one piece was immersed in TRIzol reagent (Invitrogen) and the rest was fixed in 4% paraformaldehyde/1× phosphate-buffered saline (PBS) for histological observation. Total RNA in each sample was extracted promptly in accordance with the TRIzol instruction manual. The extracted total RNA was dissolved in RNase-free water and stored at −80 °C until RNA-Seq analysis. We selected two treated and two control samples for RNA-Seq analysis on the basis of histological observation (Supplementary fig. 1).

Beads with oligo (dT) were used to isolate poly(A) mRNA after the total RNA was collected from the samples. Fragmentation buffer was added to disrupt mRNA into short fragments. Using these short fragments as templates, a random hexamer primer was used to synthesize the first-strand cDNA. Second-strand cDNA was synthesized using buffer, dNTPs, Rhase H, and DNA polymerase I. Short fragments were purified using the QiaQuick PCR extraction kit (Qiagen) and resolved with EB buffer for end reparation and adding an “A” base. Subsequently, the short fragments were connected to sequencing adapters. Following agarose gel electrophoresis, suitable fragments were selected as templates for PCR amplification. Finally, the library was sequenced using an Illumina HiSeq 2000 as 2 × 90 bp paired-ends. The acquired reads were subjected to the quality control procedures described below and used for downstream expression analysis.

Retrieval and acquisition of RNA-Seq data in protogynous fishes from public resources. To obtain RNA-Seq data from the gonads of protogynous fishes in public databases, we accessed SRA [66] (accessed in October 2022) to search SRA data containing the following keywords: “bony fishes (AND ovary OR gonad).” SRA data were converted into FASTQ files using the Fasterq-dump program of the SRA Toolkit (v3.0.0).

Quality control of RNA-Seq data. To ensure the integrity of the RNA-Seq reads, trimming

and quality control procedures were performed. Specifically, we used Trim Galore (v0.6.7) [67] with the option "--quality 20 --paired" to eliminate low-quality bases from paired-end reads.

Read mapping against the reference genome. For species with available genomic information, RNA-Seq reads were mapped to the reference genome sequences using HISAT2 (v2.2.1) [68] with the parameter "-q -dta -x". The resulting SAM files were sorted and converted into BAM files using SAMtools software (v1.15.1) [69].

Generating General Transfer Format (gtf) files based on genome sequence and RNA-Seq reads for reference gene annotation when quantifying expression values. To ensure consistency in the use of annotation information for expression quantification in species with available reference genomes, we employed the merge function of StringTie (v2.1.7) [40] to generate a gtf file that served as a guide for annotations. The transcriptome assembly for each fish was conducted using StringTie with default settings based on the mapping results (BAM file). The resulting gtf files were then merged using StringTie with the "--merge" option, producing a unified file as the "merged.gtf". This "merged.gtf" file contained a non-redundant set of assembled transcripts for each species. Generated "merged.gtf" files in each species were designated as the reference annotation files for quantifying transcript expression.

Quantification of transcripts expression and detection of differentially expressed transcripts. When reanalyzing high-throughput data from multiple experiments, it is crucial to consider batch effects [70]. In the present study, we therefore calculated the transcript per million (TPM) value for each transcript and utilized the difference in TPM values between the paired data as an indicator for identifying differentially expressed transcripts.

Transcript expression was quantified in species with available genome annotation using

StringTie, specifying the parameters as “-G merged.gtf -e”. The resulting gtf file for each sample provided the TPM values, which were used for downstream analysis.

TPM values were computed for species lacking reference genomes using Salmon (v.1.9.0)[42] with the parameters “quant -I index -l A”. To generate the transcriptome sequences required for Salmon, we performed transcriptome assembly using the Trinity (v2.13.1) assembler [41] with default settings. Open reading frames (ORFs) and CDS were predicted from the transcriptome assembly using TransDecoder (v5.5.0) [38] with the same settings as described below. The resulting gff3 file obtained from the TransDecoder run was used to extract transcript sequences corresponding to the predicted CDS from the transcriptome assembly. The extracted sequence set was then utilized as the reference transcriptome after eliminating duplicate sequences using seqkit (v2.3.0) [71] with the “rmdup -s” command.

To assess changes in transcript expression, the expression ratio was determined using TPM values obtained from intersexual gonad and ovary transcriptomes (referred to as the IO-ratio). The IO-ratio for each transcript was calculated using Equation (1).

$$\text{IO ratio} = \log_2(\text{TPM}_{\text{intersexual}} + 0.01) - \log_2(\text{TPM}_{\text{ovary}} + 0.01), \quad (1)$$

This calculation was performed for all pairs of intersexual gonad and ovary transcriptomes. To convert zeros to logarithms, a value of “0.01” was added to the TPM of each transcript. Transcripts with an IO-ratio > $\log_2(5)$ (indicating a 5-fold expression change) were classified as “upregulated,” whereas transcripts with an IO-ratio < $-\log_2(5)$ (representing a 0.2-fold expression change) were considered “downregulated.” Transcripts outside these thresholds were deemed “unchanged.” To determine differential expression between the intersexual gonad and ovary, an IO-score was calculated for each transcript by subtracting the count of “downregulated” pairs from that of “upregulated” pairs, as shown in Equation (2).

$$\text{IO-score} = (\text{count number}_{\text{upregulated}}) - (\text{count number}_{\text{downregulated}}), \quad (2)$$

Using orthology inference as described below, the total IO-score for each transcript was

calculated by summing the IO-scores of all species. Total IO-score thresholds of 7 and −7 were employed to identify transcripts that were differentially upregulated and downregulated, respectively. The thresholds were arbitrarily set based on 20% above the maximum or below the minimum values (31 or -31). All calculations for IO-ratios and IO-scores were performed using R (v 4.2.1) [72].

Orthology inference among species. GFFread (v0.12.1) [73] was utilized along with the generated “merged.gtf” file to extract transcript sequences from the reference genomes. For species without reference genomes, transcript sequences were obtained via assembly using Trinity, as described previously. TransDecoder was used to identify the coding regions and translate them into amino acid sequences for all species. ORFs were extracted from the transcriptome sequences using TransDecoder Longorfs with the parameters “-m 100”. CDS were then predicted using TransDecoder Predict with default settings, and the protein sequences converted from the predicted CDS were acquired. Subsequently, a reciprocal homology search was conducted between the predicted protein sequences of the three-spot wrasse and those of other species, and the reciprocal best hit (RBH) was identified as the ortholog. The program Diamond (v2.0.15) [74] was used for the RBH search using the parameters “--ultra-sensitive -e 1e-3”. To extract the orthologs shared among all species, we consolidated the RBH results with the three-spot wrasse-predicted protein IDs.

Gene set enrichment analysis. Differentially expressed gene sets were analyzed using the web tool Metascape (v3.5) [75], which enables gene set enrichment analysis. However, Metascape only supports the analysis of zebrafish among teleosts. Therefore, prior to analysis, the predicted protein sequences of the three-spot wrasse were mapped to the corresponding zebrafish Ensembl protein IDs. To obtain orthologous information, we utilized the ggsearch36 program from the FASTA package [43] to globally align the predicted protein sequences of the

three-spot wrasse with the proteome sequences of zebrafish (Ensembl Release 108) [44].

Visualization of cholesterol and steroid hormone biosynthesis pathways and orthologs mapping to the pathways. The hypothetical cholesterol biosynthesis and steroid hormone biosynthesis pathway in teleost fish was illustrated using PathVisio [76]. Modifying information from WikiPathways [77] intended for zebrafish (WP1389, https://www.wikipathways.org/instance/WP1387) and data from the yellow perch, *Perca flavescens* [78], facilitated the depiction of the cholesterol biosynthesis pathway. Additionally, the steroid hormone biosynthesis pathway was constructed by adjusting information sourced from various fish models [22, 79-82]. Moreover, through PathVisio, 7,289 candidate sex-change factors, converted into Ensembl zebrafish gene IDs, were mapped onto the established biosynthesis pathways.

Declarations

Ethics approval

The maintenance and handling of fish in this study were conducted in accordance with the "Guidelines for Animal Experiments of the Okinawa Churashima Foundation," with the same consideration for animal care and welfare as that for higher vertebrates, including reptiles, birds, and mammals. Animal ethical review and approval were not required for the present study because the guidelines stipulated that approval from the Institutional Animal Care and Use Committee of Okinawa Churashima Foundation is required for "higher" vertebrates, and is waived for "lower" vertebrates including fish.

Consent for publication

Not applicable

Availability of data

All sequencing data, including assembled sequences and raw sequence reads, were deposited in the DNA Data Bank of Japan (DDBJ) under the umbrella BioProject accession number PRJDB15835. The genome assembly was deposited in the DDBJ under accession numbers BSYP01000001–BSYP01000227. Additionally, raw sequence reads were deposited in the DDBJ under the following accession numbers: DRR466686 (PacBio HiFi reads), DRR486910 (Illumina HiC reads), and DRR483512-DRR483515 (Illumina RNA-Seq reads).

Competing interests

The authors have no conflicts of competing interest to declare that are relevant to the content of this article.

Funding

This research was supported by the Center of Innovation for Bio-Digital Transformation (BioDX), a program on the open innovation platform for industry-academia co-creation (COI-NEXT), and the Japan Science and Technology Agency (JST, COI-NEXT, JPMJPF2010). This work was partially received financial support by International Research Organization for Advanced Science and Technology, Kumamoto University, Kumamoto, Japan.

Authors' contributions

RN and HB conceived and designed the study. MK and SK performed genome sequence and de novo genome assembly of three-spot wrasse. RN & MN RNA-seq data production of three-spot wrasse. RN performed RNA-seq data analysis and orthology inference. RN, MN, SK and HB wrote a draft manuscript. All authors read and approved the final manuscript.

Acknowledgements

All authors thank Dr. Osamu Nishimura at DNA Sequencing Facility operated by Laboratory for Phyloinformatics in RIKEN BDR for assistance in Hi-C data assessment. We thank the Sesoko Station, Tropical Biosphere Research Center, University of the Ryukyus for allowing us to use their research facility.

Additional information

All supplementary information is accessible in figshare repository (https://doi.org/10.6084/m9.figshare.c.6704160)

Supplementary figure 1. Histological sections of control ovaries and transitional gonads treated AI for 7 days in the three-spot wrasse, *Halichoeres trimaculatus*.

Supplementary figure 2. Hypothetical ovarian steroid hormones biosynthesis pathway in protogynous fishes. The pathway was visualized using PathVisio (https://pathvisio.org), referencing several fish cholesterol synthesis pathways. Gene products are represented in black text, while metabolites are in blue text. Gene product annotations follow Ensembl Zebrafish gene symbols. Red-filled boxes represent gene products with confirmed upregulated expression in this study, while blue-filled boxes denote gene products with downregulated expression. Open boxes represent gene products with no expression variation. Light green-filled boxes represent gene products not included among the 7289 factors but confirmed to be present in at least the three-spot wrasse transcriptome. Dark gray-filled boxes indicate gene products that could not be identified within the three-spot wrasse transcriptome. Circles extending from gene products represent catalytic reactions, while black arrows indicate the conversion of metabolites.

Supplementary table 1

Supplementary table 2

Supplementary table 3

Supplementary table 4

References

1. Atz JW: **Intersexuality in fishes**. In: *Intersexuality in Vertebrates Including Man*. Edited by Armstrong CN, Marshall AJ. London: Academic press; 1964: 145–232.
2. Kuwamura T, Sunobe T, Sakai Y, Kadota T, Sawada K: **Hermaphroditism in fishes: an annotated list of species, phylogeny, and mating system**. *Ichthyol Res* 2020, **67**: 341–360.
3. Avise JC, Mank JE: **Evolutionary perspectives on hermaphroditism in fishes**. *Sex Dev* 2009, **3**: 152–163.
4. Erisman BE, Petersen CW, Hastings PA, Warner RR: **Phylogenetic perspectives on the evolution of functional hermaphroditism in teleost fishes**. *Integr Comp Biol* 2013, **53**: 736–754.
5. Lo Nostro F, Grier H, Andreone L, Guerrero GA: **Involvement of the gonadal germinal epithelium during sex reversal and seasonal testicular cycling in the protogynous swamp eel, *Synbranchus marmoratus* Bloch 1795 (Teleostei, Synbranchidae)**. *J Morphol* 2003, **257**: 107–126.
6. Nakamura M, Hourigan TF, Yamauchi K, Nagahama Y, Grau EG: **Histological and ultrastructural evidence for the role of gonadal steroid hormones in sex change in the protogynous wrasse *Thalassoma duperrey***. *Environ Biol Fishes* 1989, **24**: 117–136.
7. Ohta K, Hirano M, Mine T, Mizutani H, Yamaguchi A, Matsuyama M: **Body color change and serum steroid hormone levels throughout the process of sex change in the adult wrasse, *Pseudolabrus sieboldi***. *Mar Biol* 2008, **153**: 843–852.
8. Muncaster S, Norberg B, Andersson E: **Natural sex change in the temperate protogynous Ballan wrasse *Labrus bergylta***. *J Fish Biol* 2013, **82**: 1858–1870.
9. Bhandari RK, Komuro H, Nakamura S, Higa M, Nakamura M: **Gonadal restructuring and correlative steroid hormone profiles during natural sex change in protogynous honeycomb grouper (*Epinephelus merra*)**. *Zoolog Sci* 2003, **20**: 1399–1404.
10. Horiguchi R, Nozu R, Hirai T, Kobayashi Y, Nagahama Y, Nakamura M: **Characterization of gonadal soma-derived factor expression during sex change in the protogynous wrasse, *Halichoeres trimaculatus***. *Dev Dyn* 2013, **242**: 388–399.
11. Nozu R, Horiguchi R, Kobayashi Y, Nakamura M: **Expression profile of doublesex/male abnormal-3-related transcription factor-1 during gonadal sex change in the protogynous wrasse, *Halichoeres trimaculatus***. *Mol Reprod Dev* 2015, **82**: 859–866.
12. Horiguchi R, Nozu R, Hirai T, Kobayashi Y, Nakamura M: **Expression patterns of sex differentiation-related genes during gonadal sex change in the protogynous wrasse,**

Halichoeres trimaculatus. *Gen Comp Endocrinol* 2018, **257**: 67–73.

13. Kobayashi Y, Horiguchi R, Nozu R, Nakamura M: **Expression and localization of forkhead transcriptional factor 2 (Foxl2) in the gonads of protogynous wrasse, *Halichoeres trimaculatus***. *Biol Sex Differ* 2010, **1**: 3.

14. Thomas JT, Todd EV, Muncaster S, Lokman PM, Damsteegt EL, Liu H, Soyano K, Gléonnec F, Lamm MS, Godwin JR *et al*: **Conservation and diversity in expression of candidate genes regulating socially-induced female-male sex change in wrasses**. *PeerJ* 2019, **7**: e7032.

15. Miyake Y, Sakai Y, Kuniyoshi H: **Molecular cloning and expression profile of sex-specific genes, Figla and Dmrt1, in the protogynous hermaphroditic fish, *Halichoeres poecilopterus***. *Zoolog Sci* 2012, **29**: 690–701.

16. Senthilkumaran B, Chi W, Gao Y, Hu Q, Guo W, Li D: **Genome-wide analysis of brain and gonad transcripts reveals changes of key sex reversal-related genes expression and signaling pathways in three stages of *Monopterus albus***. *PloS one* 2017, **12**: e0173974.

17. Todd EV, Ortega-Recalde O, Liu H, Lamm MS, Rutherford KM, Cross H, Black MA, Kardailsky O, Marshall Graves JA, Hore TA *et al*: **Stress, novel sex genes, and epigenetic reprogramming orchestrate socially controlled sex change**. *Sci Adv* 2019, **5**: eaaw7006.

18. He Z, Deng F, Yang D, He Z, Hu J, Ma Z, Zhang Q, He J, Ye L, Chen H *et al*: **Crosstalk between sex-related genes and apoptosis signaling reveals molecular insights into sex change in a protogynous hermaphroditic teleost fish, ricefield eel *Monopterus albus***. *Aquaculture* 2022, **552**: 737918.

19. Fan M, Yang W, Zhang W, Zhang L: **The ontogenic gonadal transcriptomes provide insights into sex change in the ricefield eel *Monopterus albus***. *BMC Zool* 2022, **7**: 56.

20. Wu X, Yang Y, Zhong C, Guo Y, Wei T, Li S, Lin H, Liu X: **Integration of ATAC-seq and RNA-seq unravels chromatin accessibility during sex reversal in Orange-spotted grouper (*Epinephelus coioides*)**. *Int J Mol Sci* 2020, **21**: 2800.

21. Zhang Y, Zhang S, Liu Z, Zhang L, Zhang W: **Epigenetic modifications during sex change repress gonadotropin stimulation of cyp19a1a in a teleost ricefield eel (*Monopterus albus*)**. *Endocrinology* 2013, **154**: 2881–2890.

22. Li GL, Liu XC, Zhang Y, Lin HR: **Gonadal development, aromatase activity and P450 aromatase gene expression during sex inversion of protogynous red-spotted grouper *Epinephelus akaara* (Temminck and Schlegel) after implantation of the aromatase inhibitor, fadrozole**. *Aquac Res* 2006, **37**: 484–491.

23. Wang DS, Kobayashi T, Zhou LY, Paul-Prasanth B, Ijiri S, Sakai F, Okubo K, Morohashi K, Nagahama Y: **Foxl2 up-regulates aromatase gene transcription in a female-specific manner by binding to the promoter as well as interacting with ad4 binding protein/steroidogenic factor 1**. *Mol Endocrinol* 2007, **21**: 712–725.

24. Zhang C, He Q, Cheng H, Li L, Ruan X, Duan X, Huang F, Yang H, Zhang H, Shi H *et al*: **Transcription factors foxl2 and foxl3 regulate cyp19a1a and cyp11b in orange-spotted grouper (*Epinephelus coioides*)**. *Aquac Rep* 2022, **25**: 101243.

25. Shibata Y, Paul-Prasanth B, Suzuki A, Usami T, Nakamoto M, Matsuda M, Nagahama Y:

Expression of gonadal soma derived factor (GSDF) is spatially and temporally correlated with early testicular differentiation in medaka. *Gene Expr Patterns* 2010, **10**: 283–289.

26. Zhang X, Guan G, Li M, Zhu F, Liu Q, Naruse K, Herpin A, Nagahama Y, Li J, Hong Y: **Autosomal gsdf acts as a male sex initiator in the fish medaka**. *Sci Rep* 2016, **6**: 19738.
27. Kobayashi T, Kajiura-Kobayashi H, Guan G, Nagahama Y: **Sexual dimorphic expression of DMRT1 and Sox9a during gonadal differentiation and hormone-induced sex reversal in the teleost fish Nile tilapia (*Oreochromis niloticus*)**. *Dev Dyn* 2008, **237**: 297–306.
28. Rodríguez-Marí A, Yan Y-L, BreMiller RA, Wilson C, Cañestro C, Postlethwait JH: **Characterization and expression pattern of zebrafish anti-Müllerian hormone (amh) relative to sox9a, sox9b, and cyp19a1a, during gonad development**. *Gene Expr Patterns* 2005, **5**: 655–667.
29. Wang XG, Orban L: **Anti-Müllerian hormone and 11 β-hydroxylase show reciprocal expression to that of aromatase in the transforming gonad of zebrafish males**. *Dev Dyn* 2007, **236**: 1329–1338.
30. Skaar KS, Nobrega RH, Magaraki A, Olsen LC, Schulz RW, Male R: **Proteolytically activated, recombinant anti-mullerian hormone inhibits androgen secretion, proliferation, and differentiation of spermatogonia in adult zebrafish testis organ cultures**. *Endocrinology* 2011, **152**: 3527–3540.
31. Webster KA, Schach U, Ordaz A, Steinfeld JS, Draper BW, Siegfried KR: **Dmrt1 is necessary for male sexual development in zebrafish**. *Dev Biol* 2017, **422**: 33–46.
32. Zhu Y, Wang C, Chen X, Guan G: **Identification of gonadal soma-derived factor involvement in *Monopterus albus* (protogynous rice field eel) sex change**. *Mol Biol Rep* 2016, **43**: 629–637.
33. Bono H: **Meta-analysis of oxidative transcriptomes in insects**. *Antioxidants* 2021, **10**: 345.
34. Rahman MR, Petralia MC, Ciurleo R, Bramanti A, Fagone P, Shahjaman M, Wu L, Sun Y, Turanli B, Arga KY *et al*: **Comprehensive analysis of RNA-Seq gene expression profiling of brain transcriptomes reveals novel genes, regulators, and pathways in autism spectrum disorder**. *Brain Sci* 2020, **10**: 747.
35. Toga K, Bono H: **Meta-analysis of public RNA sequencing data revealed potential key genes associated with reproductive division of labor in social hymenoptera and termites**. *Int J Mol Sci* 2023, **24**: 8353.
36. Cheng H, Concepcion GT, Feng X, Zhang H, Li H: **Haplotype-resolved de novo assembly using phased assembly graphs with hifiasm**. *Nat Methods* 2021, **18**: 170–175.
37. Manni M, Berkeley MR, Seppey M, Simão FA, Zdobnov EM, Kelley J: **BUSCO update: novel and streamlined workflows along with broader and deeper phylogenetic coverage for scoring of eukaryotic, prokaryotic, and viral genomes**. *Mol Biol Evol* 2021, **38**: 4647–4654.
38. **TransDecoder (v5.5.0)** [https://github.com/TransDecoder/TransDecoder]
39. Ge H, Lin K, Shen M, Wu S, Wang Y, Zhang Z, Wang Z, Zhang Y, Huang Z, Zhou C *et al*: ***De novo* assembly of a chromosome-level reference genome of red-spotted grouper**

(*Epinephelus akaara*) using nanopore sequencing and Hi-C. *Mol Ecol Resour* 2019, **19**: 1461–1469.

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

41. Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, Adiconis X, Fan L, Raychowdhury R, Zeng Q *et al*: **Full-length transcriptome assembly from RNA-Seq data without a reference genome**. *Nat Biotechnol* 2011, **29**: 644–652.

42. Patro R, Duggal G, Love MI, Irizarry RA, Kingsford C: **Salmon provides fast and bias-aware quantification of transcript expression**. *Nat Methods* 2017, **14**: 417–419.

43. Pearson WR, Lipman DJ: **Improved tools for biological sequence comparison**. *Proc Natl Acad Sci USA* 1988, **85**: 2444–2448.

44. Cunningham F, Allen JE, Allen J, Alvarez-Jarreta J, Amode M R, Armean Irina M, Austine-Orimoloye O, Azov Andrey G, Barnes I, Bennett R *et al*: **Ensembl 2022**. *Nucleic Acids Res* 2022, **50**: D988–D995.

45. Nozu R, Kojima Y, Nakamura M: **Short term treatment with aromatase inhibitor induces sex change in the protogynous wrasse, *Halichoeres trimaculatus***. *Gen Comp Endocrinol* 2009, **161**: 360–364.

46. Kroon FJ, Liley NR: **The role of steroid hormones in protogynous sex change in the Blackeye goby, *Coryphopterus nicholsii* (Teleostei: Gobiidae)**. *Gen Comp Endocrinol* 2000, **118**: 273–283.

47. Hasegawa S, Noda K, Maeda A, Matsuoka M, Yamasaki M, Fukui T: **Acetoacetyl-CoA synthetase, a ketone body-utilizing enzyme, is controlled by SREBP-2 and affects serum cholesterol levels**. *Mol Genet Metab* 2012, **107**: 553–560.

48. Wu D-M, He Z, Chen T, Liu Y, Ma L-P, Ping J: **DNA hypermethylation of acetoacetyl-CoA synthetase contributes to inhibited cholesterol supply and steroidogenesis in fetal rat adrenals under prenatal nicotine exposure**. *Toxicology* 2016, **340**: 43–52.

49. Gaudet P, Livstone MS, Lewis SE, Thomas PD: **Phylogenetic-based propagation of functional annotations within the Gene Ontology consortium**. *Brief Bioinform* 2011, **12**: 449–462.

50. Blue ML, Williams DL, Zucker S, Khan SA, Blum CB: **Apolipoprotein E synthesis in human kidney, adrenal gland, and liver**. *Proc Natl Acad Sci USA* 1983, **80**: 283–287.

51. Reyland ME, Gwynne JT, Forgez P, Prack MM, Williams DL: **Expression of the human apolipoprotein E gene suppresses steroidogenesis in mouse Y1 adrenal cells**. *Proc Natl Acad Sci USA* 1991, **88**: 2375–2379.

52. Thorngate FE, Strockbine PA, Erickson SK, Williams DL: **Altered adrenal gland cholesterol metabolism in the apoE-deficient mouse**. *J Lipid Res* 2002, **43**: 1920–1926.

53. Payne AH, Hales DB: **Overview of steroidogenic enzymes in the pathway from cholesterol to active steroid hormones**. *Endocr Rev* 2004, **25**: 947–970.

54. Pikuleva IA: **Cholesterol-metabolizing cytochromes P450**. *Drug Metab Dispos* 2006, **34**:

513–520.

55. Miller WL: **Steroidogenic enzymes**. In: *Disorders of the Human Adrenal Cortex*. Edited by Flück CE, Miller WL. Basel: S.Karger AG; 2008: 1–18.

56. Sakae Y, Oikawa A, Sugiura Y, Mita M, Nakamura S, Nishimura T, Suematsu M, Tanaka M: **Starvation causes female-to-male sex reversal through lipid metabolism in the teleost fish, medaka (*Oryzias latipes*)**. *Biol Open* 2020, **9**: bio050054.

57. Gwynne JT, Strauss JF: **The role of lipoproteins in steroidogenesis and cholesterol metabolism in steroidogenic glands**. *Endocr Rev* 1982, **3**: 299–329.

58. Sharpe RL, Drolet M, MacLatchy DL: **Investigation of de novo cholesterol synthetic capacity in the gonads of goldfish (*Carassius auratus*) exposed to the phytosterol beta-sitosterol**. *Reprod Biol Endocrinol* 2006, **4**: 60.

59. Bera A, Chadha NK, Dasgupta S, Chakravarty S, Sawant PB: **Hypoxia-mediated inhibition of cholesterol synthesis leads to disruption of nocturnal sex steroidogenesis in the gonad of koi carp, *Cyprinus carpio***. *Fish Physiol Biochem* 2020, **46**: 2421–2435.

60. Sim SB, Corpuz RL, Simmonds TJ, Geib SM: **HiFiAdapterFilt, a memory efficient read processing pipeline, prevents occurrence of adapter sequence in PacBio HiFi reads and their negative impacts on genome assembly**. *BMC Genomics* 2022, **23**: 157.

61. Zhu B-H, Xiao J, Xue W, Xu G-C, Sun M-Y, Li J-T: **P_RNA_scaffolder: a fast and accurate genome scaffolder using paired-end RNA-sequencing reads**. *BMC Genomics* 2018, **19**: 175.

62. Kadota M, Nishimura O, Miura H, Tanaka K, Hiratani I, Kuraku S: **Multifaceted Hi-C benchmarking: what makes a difference in chromosome-scale genome scaffolding?** *Gigascience* 2020, **9**: giz158.

63. Durand NC, Shamim MS, Machol I, Rao SSP, Huntley MH, Lander ES, Aiden EL: **Juicer provides a one-click system for analyzing loop-resolution Hi-C experiments**. *Cell Syst* 2016, **3**: 95–98.

64. Dudchenko O, Shamim MS, Batra SS, Durand NC, Musial NT, Mostofa R, Pham M, Hilaire BGS, Yao W, Stamenova E *et al*: **The Juicebox assembly tools module facilitates *de novo* assembly of mammalian genomes with chromosome-length scaffolds for under $1000**. *bioRxiv* 2018: 254797.

65. Nishimura O, Hara Y, Kuraku S, Hancock J: **gVolante for standardizing completeness assessment of genome and transcriptome assemblies**. *Bioinformatics* 2017, **33**: 3635–3637.

66. Kodama Y, Shumway M, Leinonen R: **The sequence read archive: explosive growth of sequencing data**. *Nucleic Acids Res* 2011, **40**: D54–D56.

67. **Trim Galore** [https://github.com/FelixKrueger/TrimGalore]

68. Kim D, Paggi JM, Park C, Bennett C, Salzberg SL: **Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype**. *Nat Biotechnol* 2019, **37**: 907–915.

69. Danecek P, Bonfield JK, Liddle J, Marshall J, Ohan V, Pollard MO, Whitwham A, Keane T, McCarthy SA, Davies RM *et al*: **Twelve years of SAMtools and BCFtools**. *Gigascience* 2021, **10**: giab008.

70. Leek JT, Scharpf RB, Bravo HC, Simcha D, Langmead B, Johnson WE, Geman D, Baggerly K,

Irizarry RA: **Tackling the widespread and critical impact of batch effects in high-throughput data**. *Nat Rev Genet* 2010, **11**: 733–739.

71. Zou Q, Shen W, Le S, Li Y, Hu F: **SeqKit: a cross-platform and ultrafast toolkit for FASTA/Q file manipulation**. *PloS one* 2016, **11**: e0163962.

72. **R: A language and environment for statistical computing.** [URL https://www.R-project.org/]

73. Pertea G, Pertea M: **GFF Utilities: GffRead and GffCompare**. *F1000Res* 2020, **9**: 304.

74. Buchfink B, Xie C, Huson DH: **Fast and sensitive protein alignment using DIAMOND**. *Nat Methods* 2014, **12**: 59–60.

75. Zhou Y, Zhou B, Pache L, Chang M, Khodabakhshi AH, Tanaseichuk O, Benner C, Chanda SK: **Metascape provides a biologist-oriented resource for the analysis of systems-level datasets**. *Nat Commun* 2019, **10**: 1523.

76. Murphy RF, Kutmon M, van Iersel MP, Bohler A, Kelder T, Nunes N, Pico AR, Evelo CT: **PathVisio 3: An Extendable Pathway Analysis Toolbox**. *PLoS Comput Biol* 2015, **11**: e1004085.

77. Pico AR, Kelder T, van Iersel MP, Hanspers K, Conklin BR, Evelo C: **WikiPathways: Pathway Editing for the People**. *PLoS Biol* 2008, **6**: e184.

78. Kemski MM, Rappleye CA, Dabrowski K, Bruno RS, Wick M: **Transcriptomic response to soybean meal-based diets as the first formulated feed in juvenile yellow perch (*Perca flavescens*)**. *Sci Rep* 2020, **10**: 3998.

79. Liu Y, Kossack ME, McFaul ME, Christensen LN, Siebert S, Wyatt SR, Kamei CN, Horst S, Arroyo N, Drummond IA *et al*: **Single-cell transcriptome reveals insights into the development and function of the zebrafish ovary**. *Elife* 2022, **11**: e76014.

80. Craft JA, Bain PA, Papanicolaou A, Kumar A: **Identification of Putative Nuclear Receptors and Steroidogenic Enzymes in Murray-Darling Rainbowfish (*Melanotaenia fluviatilis*) Using RNA-Seq and De Novo Transcriptome Assembly**. *PloS one* 2015, **10**: e0142636.

81. Nagahama Y, Chakraborty T, Paul-Prasanth B, Ohta K, Nakamura M: **Sex determination, gonadal sex differentiation, and plasticity in vertebrate species**. *Physiol Rev* 2021, **101**: 1237–1308.

82. Fernandino JI, Hattori RS, Moreno Acosta OD, Strussmann CA, Somoza GM: **Environmental stress-induced testis differentiation: androgen as a by-product of cortisol inactivation**. *Gen Comp Endocrinol* 2013, **192**: 36–44.

Figure legends

Fig. 1. Schematic diagram of the analysis pipeline in this study. The consistent line style of the arrows indicates the utilization of common tools for conducting data processing.

Fig. 2. Scatter plot of the IO-score and gene set enrichment analysis of up- and downregulated genes. a Scatterplot of total IO-scores for the 7,289 factors in descending order. Red circles indicate 99 upregulated factors; blue circles indicate 69 downregulated factors. Gray circles (which have the appearance of lines because they overlap each other) indicate unchanged factors ($6 \leq$ total IO-score ≤ -6). **b,c** Results of enrichment analysis for 99 upregulated (**b**) and 69 downregulated (**c**) factors

Fig. 3. Composition of total IO-scores for upregulated factors (n=99). Numbers in parentheses indicate the number of RNA-Seq data pairs for each species used in this study. Asterisk indicates a gene for which the total IO-score is derived from a single species.

Fig. 4. Composition of total IO-scores for downregulated factors (n=69). Numbers in parentheses indicate the number of RNA-Seq data pairs for each species used in this study. Asterisk indicates a gene for which the total IO-score is derived from a single species.

Fig. 5. Hypothetical ovarian cholesterol biosynthesis pathway in protogynous fishes. The pathway was visualized using PathVisio (https://pathvisio.org), referencing several fish cholesterol synthesis pathways. Gene products are represented in black text, while metabolites are in blue text. Gene product annotations follow Ensembl Zebrafish gene symbols. Red-filled boxes represent gene products with confirmed upregulated expression in this study, while blue-filled boxes denote gene products with downregulated expression. Open boxes represent gene products with no expression variation. Light green-filled boxes represent gene products not included among the 7289 factors but confirmed to be present in at least the three-spot wrasse transcriptome. Dark gray-filled boxes indicate gene products that could not be identified within the three-spot wrasse transcriptome. Circles extending from gene products represent catalytic reactions, while black arrows indicate the conversion of metabolites. The

dashed arrows indicate indirect pathways summarizing multi-step processes. The dashed arrows indicate indirect pathways summarizing multi-step processes.

Table 1. Statistics of genome assembly in Three-spot wrasse (fHalTri1.1) and other Labidae fishes listed in NCBI Genomes.

	fHalTri1.1	***Cheilinus undulatus*** **(humphead wrasse)**	***Notolabrus celidotus*** **(New Zealand spotty)**	***Labrus bergylta*** **(ballan wrasse)**	***Semicossyphus pulcher*** **(California sheephead)**	***Tautogolabrus adspersus*** **(cunner)**
Assembly accession	BSYP01000001–BSYP01000227	GCF_018320785.1	GCF_009762535.1	GCF_900080235.1	GCA_022749685.1	GCA_020745685.1
Total seq length (Mb)	849.9	1173.4	846.7	805.5	794.1	723.6
Max length (Mb)	44.2	59.6	42.4	9.1	38.5	37.1
Num of scaffolds	227	45	467	13,466	178	347
Scaffold N50 (Mb)	37.9	51.5	37.1	0.8	32.1	31.7
Scaffold L50	11	11	11	252	12	11
Num of seqs > 1 Mb	25	24	24	197	35	24
% of seqs > 1 Mb	97.3	99.7	97.6	43.9	97.1	99.9
Num of gaps	117	284	1,133	257	8	283
Complete BUSCO (%)	98.7	99.0	97.1	98.9	98.9	98.1

Table 2. Statistics of orthology relationship among species.

Common name	Scientific name	Order	Reference genome	*Num of transcripts seqs obtained based on the RNA-seq data from gonadal tissue	Num of protein seqs predicted by TransDecoder	Num of RBH (vs Three-spot wrasse)
Three-spot wrasse	*Halichoeres trimaculatus*	Labriformes	Assembly in this study	27,573	24,965	24,965
Bluehead wrasse	*Thalassoma bifasciatum*	Labriformes	GCA_008086565.1	57,699	43,195	13,871
Hong Kong grouper	*Epinephelus akaara*	Perciformes	Ge et al., 2019	26,507	24,752	11,619
Orange-spotted grouper	*Epinephelus coioides*	Perciformes	GCA_900536245.1	30,867	34,191	12,507
Black seabass	*Centropristis striata*	Perciformes	NA	** 95,941	95,941	12,591
Asian swamp eel	*Monopterus albus*	Synbranchiformes	GCF_001952655.1	53,472	50,693	12,765
Zig-zag eel	*Mastacembelus armatus*	Synbranchiformes	GCF_900324485.2	48,724	40,908	13,095

* Based on the annotation file (merged gtf file), the extracted transcripts sequences from the genome by gffread

** Among transcriptome sequences (594,726 seqs) assembled by Trinity, transcripts sequences corresponding to coding seqences predicted via TransDecoder

Table 3. List of genes that contributed to GO term enrichment.

Up/Down	GO ID	Category	GO term	*Genes
Up	GO:0007129	Biological Process	homologous chromosome pairing at meiosis	*hormad1, terb2, si:ch211-199g17.9*
Up	GO:0007369	Biological Process	gastrulation	*zeb1b, tmem237a, col12a1b, pfn1, enpp2*
Up	GO:1902531	Biological Process	regulation of intracellular signal transduction	*cep55l, psd3l, arhgef18b, ccl19a.1, sh3bp1, enpp2, pik3r6b*
Up	GO:0005929	Cellular Component	cilium	*septin6, prom1a, tmem237a, si:ch211-66i15.4, flr*
Up	GO:0007281	Biological Process	germ cell development	*hormad1, gsdf, sycp3*
Down	GO:0010817	Biological Process	regulation of hormone levels	*cyp19a1a, aacs, hsd17b1, ece1*
Down	GO:0006413	Biological Process	translational initiation	*eif4g2b, eif4eb, eif4e1c*
Down	GO:0019752	Biological Process	carboxylic acid metabolic process	*lpl, ldha, aacs, amacr, acaca, dars2*
Down	GO:0008202	Biological Process	steroid metabolic process	*hsd17b1, amacr, sdr42e1*
Down	GO:1904888	Biological Process	cranial skeletal system development	*nog3, cep55l, cyldb, ece1*
Down	GO:1901362	Biological Process	organic cyclic compound biosynthetic process	*tbx5a, hsd17b1, alad, acaca, sdr42e1, adcy6a*

* Gene symbols follow the zebrafish gene nomenclature in Ensembl.

Fig. 1.

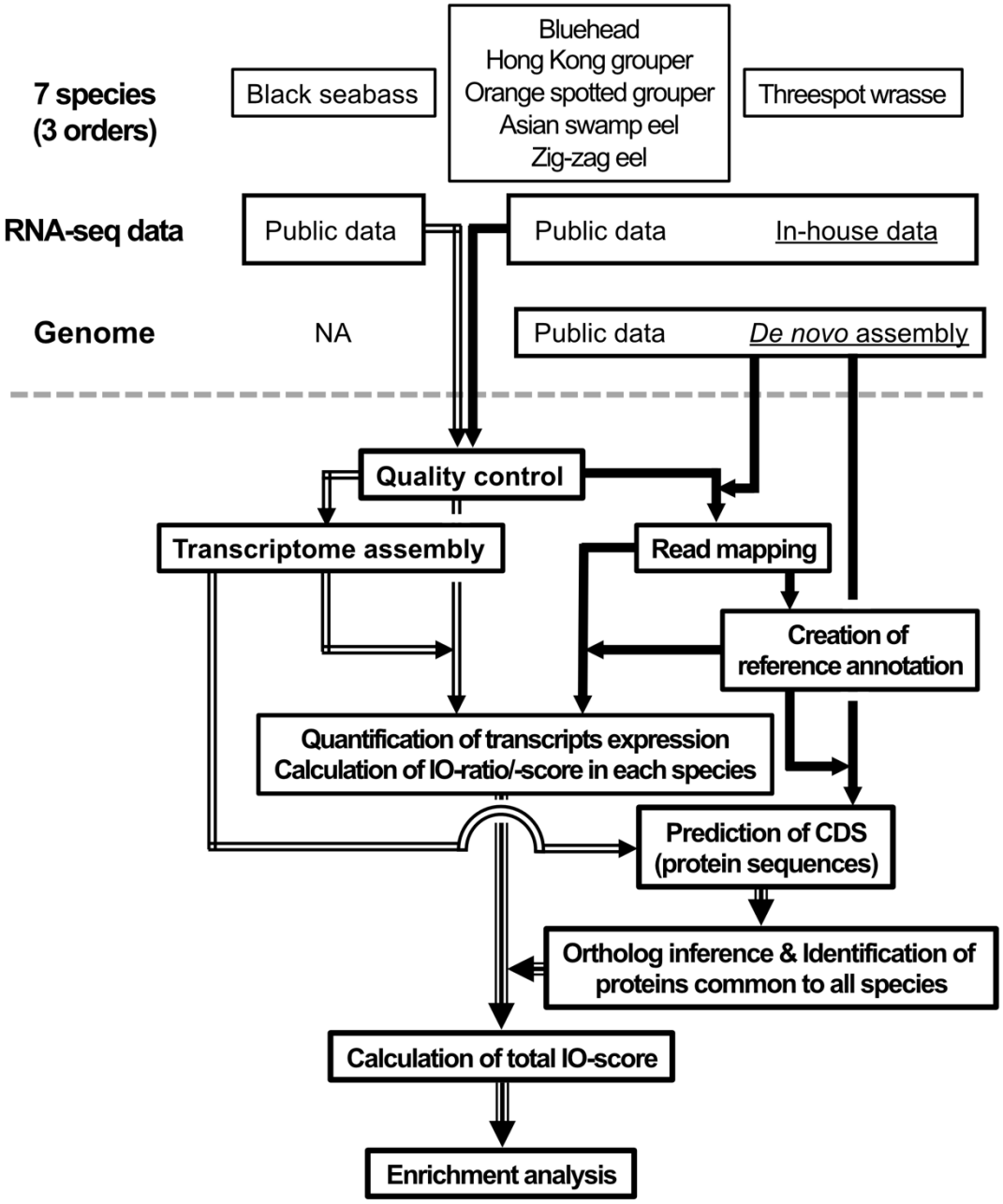

7 species
(3 orders)
Black seabass
Bluehead
Hong Kong grouper
Orange spotted grouper
Asian swamp eel
Zig-zag eel
Threespot wrasse
RNA-seq data
Public data
Public data
In-house data
Genome
NA
Public data
De novo assembly
Quality control
Transcriptome assembly
Read mapping
Creation of
reference annotation
Quantification of transcripts expression
Calculation of IO-ratio/-score in each species
Prediction of CDS
(protein sequences)
Ortholog inference & Identification of
proteins common to all species
Calculation of total IO-score
Enrichment analysis

Fig. 2.

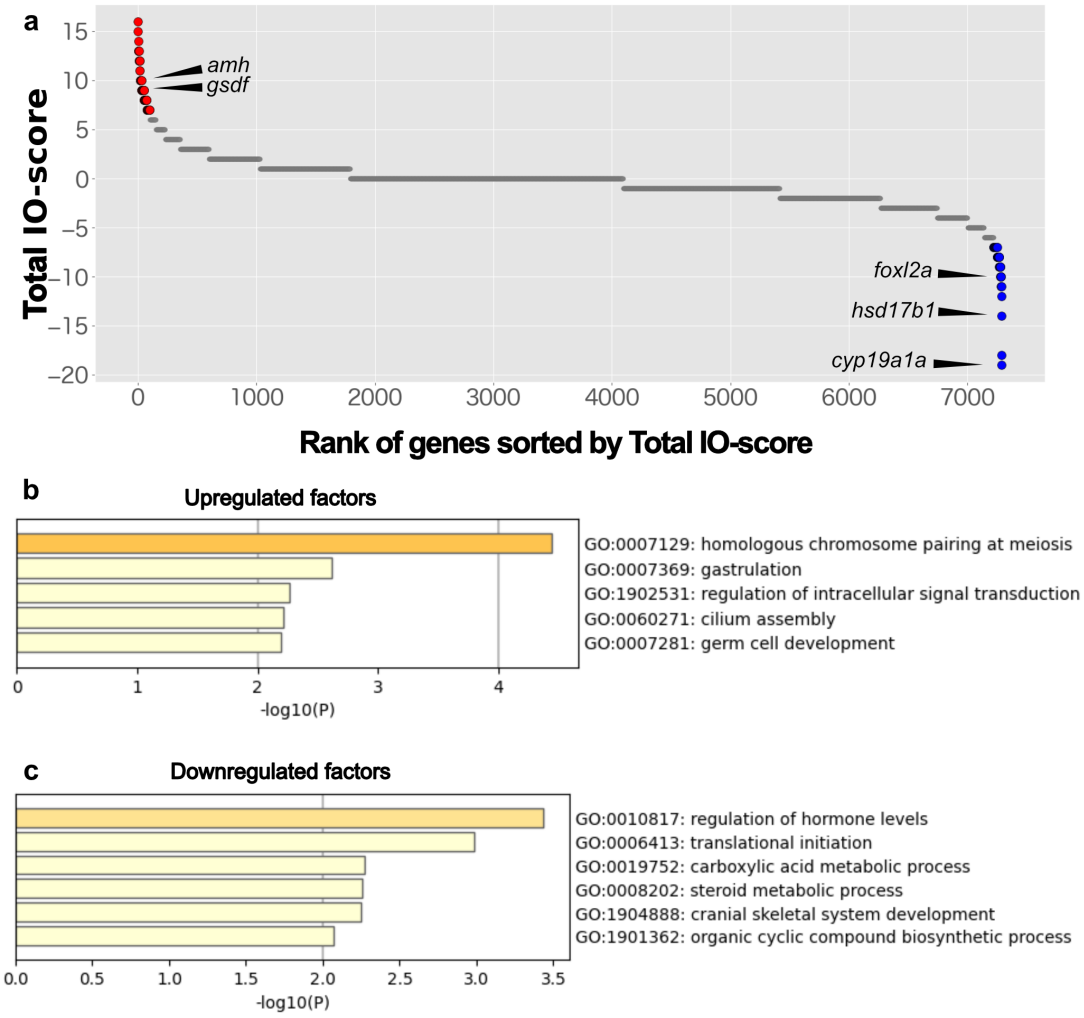

a
Total IO-score
15
10
5
0
−5
−10
−15
−20
amh
gsdf
foxl2a
hsd17b1
cyp19a1a
0
1000
2000
3000
4000
5000
6000
7000
Rank of genes sorted by Total IO-score
b
Upregulated factors
GO:0007129: homologous chromosome pairing at meiosis
GO:0007369: gastrulation
GO:1902531: regulation of intracellular signal transduction
GO:0060271: cilium assembly
GO:0007281: germ cell development
0
1
2
3
4
-log10(P)
c
Downregulated factors
GO:0010817: regulation of hormone levels
GO:0006413: translational initiation
GO:0019752: carboxylic acid metabolic process
GO:0008202: steroid metabolic process
GO:1904888: cranial skeletal system development
GO:1901362: organic cyclic compound biosynthetic process
0.0
0.5
1.0
1.5
2.0
2.5
3.0
3.5
-log10(P)

Fig. 3.

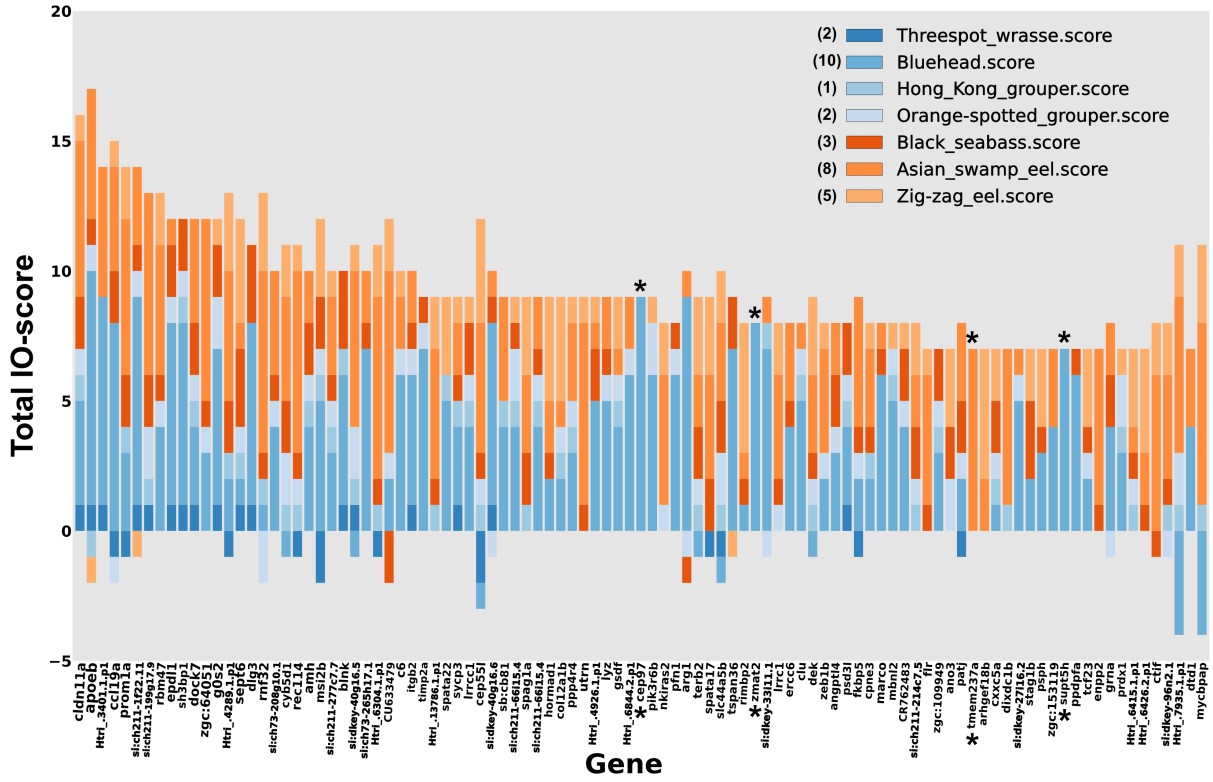

Total IO-score
Gene
(2) Threespot_wrasse.score
(10) Bluehead.score
(1) Hong_Kong_grouper.score
(2) Orange-spotted_grouper.score
(3) Black_seabass.score
(8) Asian_swamp_eel.score
(5) Zig-zag_eel.score

Fig. 4.

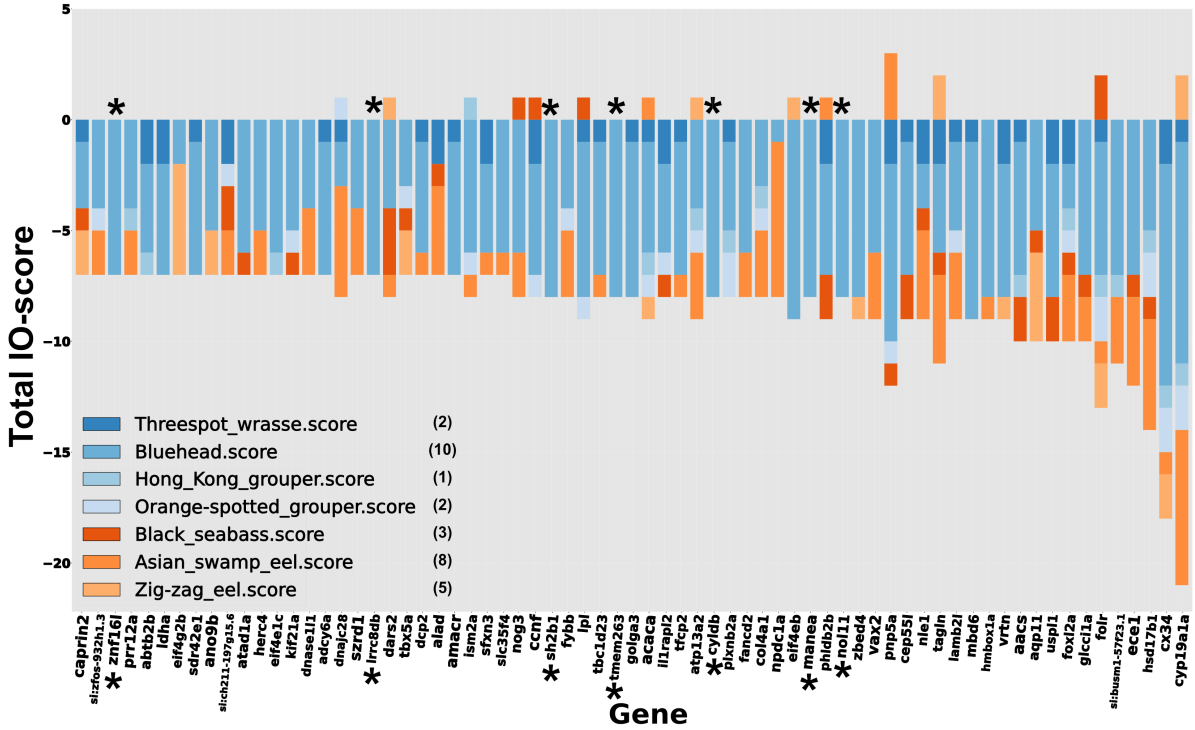
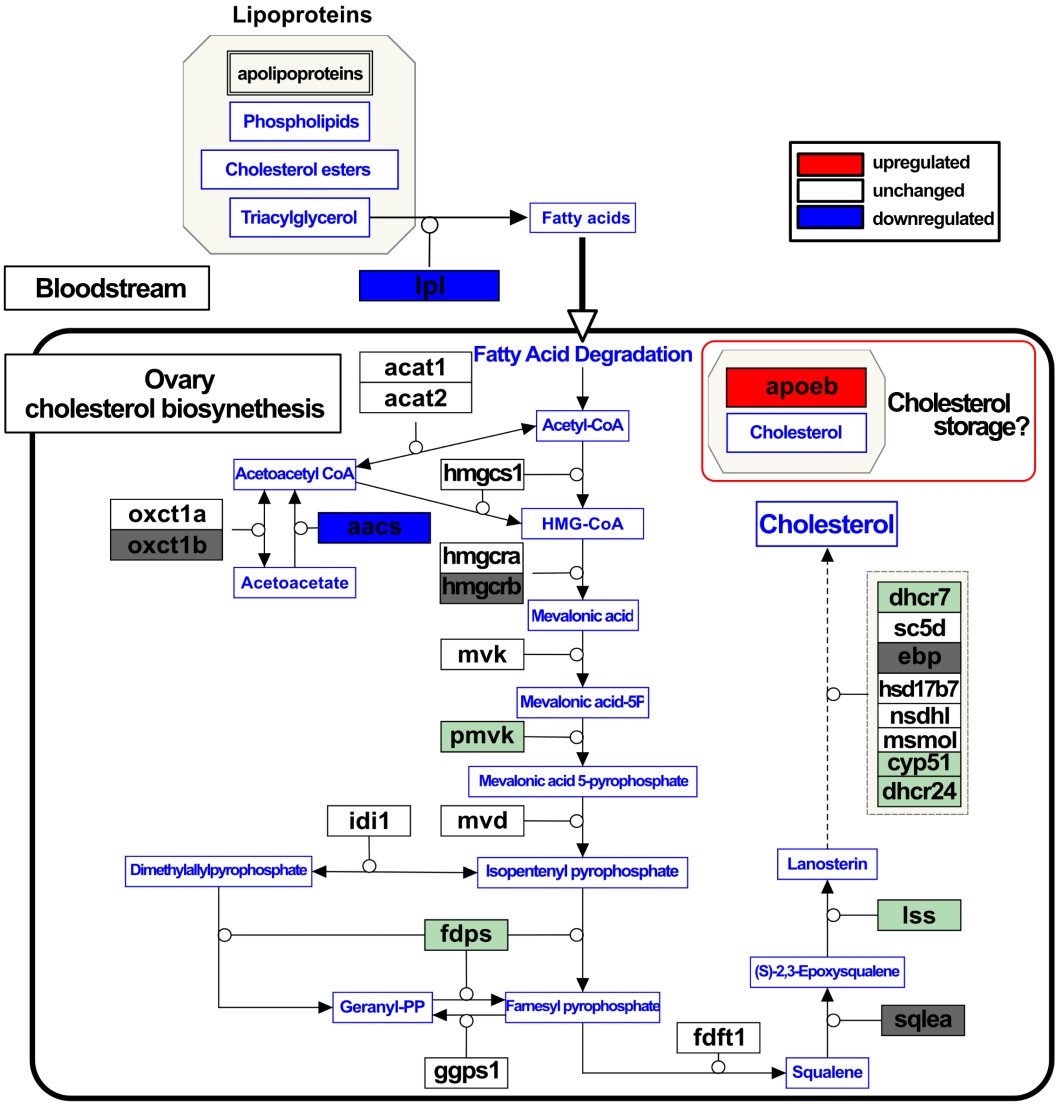

Total IO-score
5
0
−5
−10
−15
−20
Threespot_wrasse.score (2)
Bluehead.score (10)
Hong_Kong_grouper.score (1)
Orange-spotted_grouper.score (2)
Black_seabass.score (3)
Asian_swamp_eel.score (8)
Zig-zag_eel.score (5)
Gene

Fig. 5.


Lipoproteins
apolipoproteins
Phospholipids
Cholesterol esters
Triacylglycerol
Fatty acids
upregulated
unchanged
downregulated
Bloodstream
lpl
Ovary
cholesterol biosynethesis
acat1
acat2
Fatty Acid Degradation
apoeb
Cholesterol
Cholesterol storage?
Acetyl-CoA
Acetoacetyl CoA
hmgcs1
oxct1a
oxct1b
aacs
HMG-CoA
Cholesterol
hmgcra
hmgcrb
Acetoacetate
Mevalonic acid
dhcr7
sc5d
ebp
hsd17b7
nsdhl
msmol
cyp51
dhcr24
mvk
Mevalonic acid-5F
pmvk
Mevalonic acid 5-pyrophosphate
idi1
mvd
Dimethylallylpyrophosphate
Isopentenyl pyrophosphate
Lanosterin
lss
fdps
(S)-2,3-Epoxysqualene
Geranyl-PP
Farnesyl pyrophosphate
sqlea
fdft1
ggps1
Squalene