Gonadal transcriptome of sex change in fishes: quest for causal factors common to phylogenetically distributed protogynous species

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, masaru220101@gmail.com; SK, skuraku@nig.ac.jp

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

bonohu@hiroshima-u.ac.jp

Abstract

Protogyny, capable of changing from female to male during their lifetime, is prevalent in 20 families of teleosts but believed to have evolved within specific evolutionary lineages. Therefore, shared regulatory factors governing this process are expected to be conserved across protogynous fish. However, a comprehensive understanding of this mechanism remains elusive. To identify these factors, we conducted a meta-analysis using gonadal transcriptome data from seven species. We curated data pairs of ovarian tissue and transitional gonad, and employed ratios of expression level as a unified criterion for differential expression, enabling a meta-analysis across species. Our approach revealed that classical sex change-related genes exhibited differential expression levels between the ovary and transitional gonads, consistent with previous reports on protogynous fishes. These results validate our methodology's robustness. Additionally, we identified novel factors not previously linked to gonadal sex change in fish. Notably, changes in the expression levels 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. This implies a potential universal influence of cholesterol dynamics on gonadal transformation in protogyny.

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

Introduction

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]. Phylogenetic analysis has shown that sequential hermaphroditism has been acquired in various evolutionary lineages[2,3] (Fig. 1a). Especially, protogyny (i.e., the transformation from female to male) is prevalent in 20 families but is thought to have emerged once in evolutionally the Percomorpha (Fig. 1a)[4]. Therefore, there are common features in the intricate regulatory mechanisms during protogynous sex change. Specifically, the replacement of ovarian tissue with testicular tissue is commonly observed during protogynous sex change[5-9]. Additionally, 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 and commonality 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]. In other words, it is expected that by reanalyzing studies across multiple species and studies, universal factors controlling gonadal transformation, which have not been detected in individual studies, could be identified among

protogynous fishes. 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. Furthermore, in this study, we focused on another well-studied species, the three-spot wrasse, *Halichoeres trimaculatus*, a protogynous species belonging to the family Labridae and produced *de novo* genome assembly and gonad transcriptome data undergoing sex change. This new data consolidated our dataset encompassing seven protogynous species from three orders allowing exhaustive genome-wide investigation. Through this dataset and uniform pipeline, we identified factors with differential expression levels between the transitional gonads and ovaries (Fig. 1b). 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

125 the transcriptome for the annotation file ("merged.gtf"). This process yielded 27,573 transcript
126 sequences (Table 2). Subsequently, 24,695 coding sequences (protein sequences) were predicted
127 from the set of transcript sequences using TransDecoder[38].
128
129 **Data curation of public RNA-Seq data and genomic information.** We downloaded Sequence
130 Read Archive (SRA) data from six fish species for which sex changes have been reported[4]:
131 bluehead wrasse *Thalassoma bifasciatum* (order Labriformes), orange-spotted grouper
132 *Epinephelus coioides* (order Perciformes), Hong Kong grouper *E. akaara* (order Perciformes),
133 black seabass *Centropristis striata* (order Perciformes); Asian swamp eel *Monopterus albus*
134 (order Synbranchiformes), zig-zag eel *Mastacembelus armatus* (order Synbranchiformes).
135 Furthermore, using the metadata associated with the SRA data, we curated the RNA-Seq data
136 into comparable sample pairs of intersexual (transitional) gonads and ovaries (a total of 31
137 paired data, including three-spot wrasse). The data pairs analyzed in this study are listed in
138 Supplementary Table 1.
139 The genome sequences of five of the seven species utilized in this study are publicly
140 available (Table 2) and served as reference genomes for read mapping. The genome sequences
141 of *T. bifasciatum* (GCA_008086565.1), *E. coioides* (GCA_900536245.1), *M. albus*
142 (GCF_001952655.1), and *M. armatus* (GCF_900324485.2) were retrieved from the National
143 Center for Biotechnology Information (NCBI) Genome Database (accessed in October 2022).
144 We also referenced the genome information of *E. akaara* reported by Ge et al., 2019[39].
145
146 **Transcriptome assembly and inference of ortholog relationships.** For the five species with
147 public reference genomes, the number of transcript sequences obtained using StringTie [40] with
148 the "--merge" option is shown in Table 2. For black sea bass, (without a reference genome), the
149 number of contiguous sequences assembled using Trinity [41] was 594,726. Among these, 95,941
150 transcript sequences corresponding to the predicted coding sequences (CDS) were specified as

151 references for expression quantification using Salmon[42]. The number of protein sequences
152 predicted using TransDecoder is shown in Table 2. Orthology inference and consolidation of
153 orthologs within the three-spot wrasse identified 7,289 orthologs that were common to all
154 species used in this study (Table 2), and thus were considered as candidate sex-change factors.
155
156 **Identification of differentially expressed factors.** The total intersexual gonad and ovary score
157 (IO-score) was calculated for these 7,289 factors (Fig. 2a, Supplementary Table 2, IO-score
158 calculation; see Methods). Among these, 99 factors were upregulated (Fig. 2b, Supplementary
159 Table 3) and 69 factors were downregulated (Fig. 2c, Supplementary Table 4). Whereas the total
160 IO-score of these factors was occasionally derived from a single species (up: *cep97*, *zmat2*,
161 *tmem237a*, *supt5h*; down: *znf16l*, *lrrc8db*, *sh2b1*, *tmem263*, *cyldb*, *manea*, *nol11*), in most cases,
162 the total IO-scores were based on contributions from two or more species. Notably, the
163 upregulated genes included *amh* and *gsdf*, which have been previously reported to play
164 important roles in gonadal sex change, particularly in testicular formation and development, in
165 some protogynous fish. Conversely, the downregulated genes included *cyp19a1a*, *hsd17b1*, and
166 *foxl2a*, which are crucial for the onset of sex change (Table 3).
167
168 **Enrichment analysis of differentially expressed factors.** Sequence homology search of the
169 three-spot wrasse proteins using ggsearch36 [43] against the reference zebrafish proteins (Ensembl
170 Release 108) [44]resulted in the identification of 90 homologs among the 99 upregulated factors
171 (Supplementary Table 3). No corresponding zebrafish orthologs were found for the remaining
172 nine factors that were common to all species analyzed in this study. The results of enrichment
173 analysis using the 90 factors that could be converted to zebrafish orthologs are shown in Fig. 3a.
174 The Gene Ontology (GO) terms “homologous chromosome pairing at meiosis” and “germ cell
175 development,” which are related to reproduction, were significantly enriched among the
176 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. 3b. 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. 4). 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*

203 *undulatus*), ballan wrasse (*Labrus bergylta*), and New Zealand spotty (*Notolabrus celidotus*),
204 previously deposited in the NCBI RefSeq database. These results indicated that the *de novo*
205 genome assembly of the three-spot wrasse exhibited high quality, enabling us to perform
206 expression analysis and orthology inference with confidence.

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

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

229 APOE in tissues with active steroid production, including the adrenal glands, appears to shift
230 cholesterol metabolism from delivery to the steroidogenic pathway to storage of esterified
231 cholesterol, resulting in a decrease in steroid production levels[52]. These observations strongly
232 support the hypothesis that the downregulation of *aacs* and the upregulation of *apoeb* during sex
233 change contribute to suppression of the cholesterol supply and entire steroidogenesis.
234 Cholesterol serves as a precursor for steroid hormones, including estrogens, androgens, and
235 glucocorticoids[53-55], which play crucial roles in protogynous sex change. Previous studies have
236 shown that a decline in endogenous estrogens, particularly E2, triggers the transition from ovary
237 to testis in protogynous fish[6,45,46]. E2 is synthesized from testosterone via an aromatase enzyme,
238 encoded by *cyp19a1a*. Downregulation of *cyp19a1a* expression at the onset of sex change leads
239 to reduced E2 production. Studies on *M. albus* and *T. bifasciatum* have revealed higher DNA
240 methylation levels of the *cyp19a1a* gene upstream region in the testes than those in the ovaries,
241 suggesting that epigenetic modifications may contribute to the suppression of *cyp19a1a*
242 expression during gonadal sex change[17,21]. Our present findings could contribute one more
243 "piece of the puzzle", suggesting that the trigger of gonadal transformation may be attributed
244 not only to the decreased E2 production by the downregulation of aromatase activity but also to
245 the suppression of the cholesterol supply, the precursor of steroid hormones, leading to the
246 inhibition of the entire steroid biosynthesis. Interestingly, a decrease in the expression of
247 lipoprotein lipase (*lpl*) has been observed (Fig 4, supplementary table 4). Lipoprotein lipase is
248 involved in the biosynthetic process of fatty acid which serves as one of the sources of acetyl-
249 CoA, a precursor for cholesterol synthesis. This suggests the possibility of a concurrent
250 suppression of fatty acid supply to the gonads, might be leading to support the hypothesis.
251 Although this hypothesis requires further investigation and an understanding of cholesterol
252 dynamics during sex change, additional exploration is warranted. Notably, a study on medaka
253 (*Oryzias latipes*) subjected to starvation conditions implied a potential link between
254 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

307 4200 (Agilent Technologies) to ensure high integrity, and subsequently analyzed via pulse-field
308 gel electrophoresis on a CHEF DR-II system (BioRad) to ensure a size range of 20–100 kb.
309 To obtain long-read sequence data, a SMRT sequence library was constructed using the
310 SMRTbell Express Template Prep Kit 2.0 (PacBio) and sequenced on a PacBio Sequel II
311 system. The sequencing output was processed to generate circular consensus sequences to
312 obtain HiFi sequence reads. Adapter sequences were removed from the HiFi reads using
313 HiFiAdapterFilt[60]. The obtained HiFi sequence reads were assembled using the hifiasm
314 (v0.16.1) assembler [36] with default parameters. The obtained contigs were first scaffolded with
315 the P_RNA_scaffolder [61] and further scaffolded using Hi-C data as follows. The Hi-C library
316 was prepared with the in situ method using the muscle tissue of the individual used for DNA
317 extraction according to the iconHi-C protocol[62], employing the restriction enzymes DpnII and
318 HinfI, and sequenced on a HiSeq X sequencing platform (Illumina). The obtained Hi-C read
319 pairs, amounting to 232 million read pairs, were aligned to the HiFi sequence contigs using the
320 Juicer program[63], and the HiFi sequence contigs were scaffolded via the 3d-dna (version
321 201008) to be consistent with the chromatin contact profiles through manual curation on
322 JuiceBox v1.11.08[64]. The continuity and completeness of the resulting genome assembly were
323 assessed using the webserver gVolante[65], in which the pipeline BUSCO (v5.4, reference
324 datasets: vertebrata_odb10) was implemented[37].
325
326 **RNA-Seq of gonadal tissue in the three-spot wrasse.** Based on previous studies[10,45], we
327 triggered a sex change in three-spot wrasse n=5 females (body length; 8.0–9.3 cm, body weight;
328 12.4–20.9 g) to obtain transitional gonads through treatment with an aromatase inhibitor (AI),
329 which exerts inhibitory effects on the enzyme responsible for estrogen synthesis. The collected
330 gonads were cut into small pieces, one piece was immersed in TRIzol reagent (Invitrogen) and
331 the rest was fixed in 4% paraformaldehyde/1× phosphate-buffered saline (PBS) for histological
332 observation. Total RNA in each sample was extracted promptly in accordance with the TRIzol

333 instruction manual. The extracted total RNA was dissolved in RNase-free water and stored at
334 −80 °C until RNA-Seq analysis. We selected two treated and two control samples for RNA-Seq
335 analysis on the basis of histological observation (Supplementary fig. 1).
336 Beads with oligo (dT) were used to isolate poly(A) mRNA after the total RNA was collected
337 from the samples. Fragmentation buffer was added to disrupt mRNA into short fragments.
338 Using these short fragments as templates, a random hexamer primer was used to synthesize the
339 first-strand cDNA. Second-strand cDNA was synthesized using buffer, dNTPs, Rhase H, and
340 DNA polymerase I. Short fragments were purified using the QiaQuick PCR extraction kit
341 (Qiagen) and resolved with EB buffer for end reparation and adding an "A" base. Subsequently,
342 the short fragments were connected to sequencing adapters. Following agarose gel
343 electrophoresis, suitable fragments were selected as templates for PCR amplification. Finally,
344 the library was sequenced using an Illumina HiSeq 2000 as 2 × 90 bp paired-ends. The acquired
345 reads were subjected to the quality control procedures described below and used for
346 downstream expression analysis.
347
348 **Retrieval and acquisition of RNA-Seq data in protogynous fishes from public resources.** To
349 obtain RNA-Seq data from the gonads of protogynous fishes in public databases, we accessed
350 SRA [66] (accessed in October 2022) to search SRA data containing the following keywords:
351 "bony fishes (AND ovary OR gonad)." SRA data were converted into FASTQ files using the
352 Fasterq-dump program of the SRA Toolkit (v3.0.0).
353
354 **Quality control of RNA-Seq data.** To ensure the integrity of the RNA-Seq reads, trimming and
355 quality control procedures were performed. Specifically, we used Trim Galore (v0.6.7) [67]with
356 the option "--quality 20 --paired" to eliminate low-quality bases from paired-end reads.
357
358 **Read mapping against the reference genome.** For species with available genomic

359 information, RNA-Seq reads were mapped to the reference genome sequences using HISAT2
360 (v2.2.1) [68]with the parameter "-q -dta -x". The resulting SAM files were sorted and converted
361 into BAM files using SAMtools software (v1.15.1)[69].
362
363 **Generating General Transfer Format (gtf) files based on genome sequence and RNA-Seq**
364 **reads for reference gene annotation when quantifying expression values.** To ensure
365 consistency in the use of annotation information for expression quantification in species with
366 available reference genomes, we employed the merge function of StringTie (v2.1.7) [40]to
367 generate a gtf file that served as a guide for annotations. The transcriptome assembly for each
368 fish was conducted using StringTie with default settings based on the mapping results (BAM
369 file). The resulting gtf files were then merged using StringTie with the "--merge" option,
370 producing a unified file as the "merged.gtf". This "merged.gtf" file contained a non-redundant
371 set of assembled transcripts for each species. Generated "merged.gtf" files in each species were
372 designated as the reference annotation files for quantifying transcript expression.
373
374 **Quantification of transcripts expression and detection of differentially expressed**
375 **transcripts.** When reanalyzing high-throughput data from multiple experiments, it is crucial to
376 consider batch effects[70]. In the present study, we therefore calculated the transcript per million
377 (TPM) value for each transcript and utilized the difference in TPM values between the paired
378 data as an indicator for identifying differentially expressed transcripts.
379 Transcript expression was quantified in species with available genome annotation using
380 StringTie, specifying the parameters as "-G merged.gtf -e". The resulting gtf file for each
381 sample provided the TPM values, which were used for downstream analysis.
382 TPM values were computed for species lacking reference genomes using Salmon (v.1.9.0)[42]
383 with the parameters "quant -I index -l A". To generate the transcriptome sequences required for
384 Salmon, we performed transcriptome assembly using the Trinity (v2.13.1) assembler [41] with

385 default settings. Open reading frames (ORFs) and CDS were predicted from the transcriptome
386 assembly using TransDecoder (v5.5.0) [38]with the same settings as described below. The
387 resulting gff3 file obtained from the TransDecoder run was used to extract transcript sequences
388 corresponding to the predicted CDS from the transcriptome assembly. The extracted sequence
389 set was then utilized as the reference transcriptome after eliminating duplicate sequences using
390 seqkit (v2.3.0) [71]with the "rmdup -s" command.
391 To assess changes in transcript expression, the expression ratio was determined using TPM
392 values obtained from intersexual gonad and ovary transcriptomes (referred to as the IO-ratio).
393 The IO-ratio for each transcript was calculated using Equation (1).
394 $$\text{IO ratio} = \log_2(\text{TPM}_{\text{intersexual}} + 0.01) - \log_2(\text{TPM}_{\text{ovary}} + 0.01), \quad (1)$$
395 This calculation was performed for all pairs of intersexual gonad and ovary transcriptomes. To
396 convert zeros to logarithms, a value of "0.01" was added to the TPM of each transcript.
397 Transcripts with an IO-ratio > $\log_2(5)$ (indicating a 5-fold expression change) were classified as
398 "upregulated," whereas transcripts with an IO-ratio < $-\log_2(5)$ (representing a 0.2-fold
399 expression change) were considered "downregulated." Transcripts outside these thresholds were
400 deemed "unchanged." To determine differential expression between the intersexual gonad and
401 ovary, an IO-score was calculated for each transcript by subtracting the count of
402 "downregulated" pairs from that of "upregulated" pairs, as shown in Equation (2).
403 $$\text{IO-score} = (\text{count number}_{\text{upregulated}}) - (\text{count number}_{\text{downregulated}}), \quad (2)$$
404 Using orthology inference as described below, the total IO-score for each transcript was
405 calculated by summing the IO-scores of all species. Total IO-score thresholds of 7 and −7 were
406 employed to identify transcripts that were differentially upregulated and downregulated,
407 respectively. The thresholds were arbitrarily set based on 20% above the maximum or below the
408 minimum values (31 or -31). All calculations for IO-ratios and IO-scores were performed using
409 R (v 4.2.1)[72].
410

411 **Orthology inference among species.** GFFread (v0.12.1) [73]was utilized along with the
412 generated "merged.gtf" file to extract transcript sequences from the reference genomes. For
413 species without reference genomes, transcript sequences were obtained via assembly using
414 Trinity, as described previously. TransDecoder was used to identify the coding regions and
415 translate them into amino acid sequences for all species. ORFs were extracted from the
416 transcriptome sequences using TransDecoder Longorfs with the parameters "-m 100". CDS
417 were then predicted using TransDecoder Predict with default settings, and the protein sequences
418 converted from the predicted CDS were acquired. Subsequently, a reciprocal homology search
419 was conducted between the predicted protein sequences of the three-spot wrasse and those of
420 other species, and the reciprocal best hit (RBH) was identified as the ortholog. The program
421 Diamond (v2.0.15) [74]was used for the RBH search using the parameters "--ultra-sensitive -e 1e-
422 3". To extract the orthologs shared among all species, we consolidated the RBH results with the
423 three-spot wrasse-predicted protein IDs.
424
425 **Gene set enrichment analysis.** Differentially expressed gene sets were analyzed using the web
426 tool Metascape (v3.5)[75], which enables gene set enrichment analysis. However, Metascape only
427 supports the analysis of zebrafish among teleosts. Therefore, prior to analysis, the predicted
428 protein sequences of the three-spot wrasse were mapped to the corresponding zebrafish Ensembl
429 protein IDs. To obtain orthologous information, we utilized the ggsearch36 program from the
430 FASTA package [43] to globally align the predicted protein sequences of the three-spot wrasse
431 with the proteome sequences of zebrafish (Ensembl Release 108)[44].
432
433 **Visualization of cholesterol and steroid hormone biosynthesis pathways and orthologs**
434 **mapping to the pathways.** The hypothetical cholesterol biosynthesis and steroid hormone
435 biosynthesis pathway in teleost fish was illustrated using PathVisio[76]. Modifying information
436 from WikiPathways [77] intended for zebrafish (WP1387,

437 https://www.wikipathways.org/instance/WP1387) and data from the yellow perch, *Perca*
438 *flavescens*[78], facilitated the depiction of the cholesterol biosynthesis pathway. Additionally, the
439 steroid hormone biosynthesis pathway was constructed by adjusting information sourced from
440 various fish models[22,79-82]. Moreover, through PathVisio, 7,289 candidate sex-change factors,
441 converted into Ensembl zebrafish gene IDs, were mapped onto the established biosynthesis
442 pathways.
443
444
445 **References**

446 1 Atz, J. W. in *Intersexuality in Vertebrates Including Man* (eds C. N. Armstrong & A. J.
447 Marshall) 145–232 (Academic press, 1964).
448 2 Avise, J. C. & Mank, J. E. Evolutionary perspectives on hermaphroditism in fishes. *Sex. Dev.* **3**,
449 152–163, doi:10.1159/000223079 (2009).
450 3 Erisman, B. E., Petersen, C. W., Hastings, P. A. & Warner, R. R. Phylogenetic perspectives on the
451 evolution of functional hermaphroditism in teleost fishes. *Integr. Comp. Biol.* **53**, 736–754,
452 doi:10.1093/icb/ict077 (2013).
453 4 Kuwamura, T., Sunobe, T., Sakai, Y., Kadota, T. & Sawada, K. Hermaphroditism in fishes: an
454 annotated list of species, phylogeny, and mating system. *Ichthyological research* **67**, 341–360,
455 doi:10.1007/s10228-020-00754-6 (2020).
456 5 Lo Nostro, F., Grier, H., Andreone, L. & Guerrero, G. A. Involvement of the gonadal germinal
457 epithelium during sex reversal and seasonal testicular cycling in the protogynous swamp eel,
458 *Synbranchus marmoratus* Bloch 1795 (Teleostei, Synbranchidae). *Journal of morphology* **257**,
459 107–126, doi:10.1002/jmor.10105 (2003).
460 6 Nakamura, M., Hourigan, T. F., Yamauchi, K., Nagahama, Y. & Grau, E. G. Histological and
461 ultrastructural evidence for the role of gonadal steroid hormones in sex change in the protogynous
462 wrasse *Thalassoma duperrey*. *Environmental biology of fishes* **24**, 117–136,
463 doi:10.1007/bf00001282 (1989).
464 7 Ohta, K. *et al.* Body color change and serum steroid hormone levels throughout the process of sex
465 change in the adult wrasse, *Pseudolabrus sieboldi*. *Marine Biology* **153**, 843–852,
466 doi:10.1007/s00227-007-0856-0 (2008).
467 8 Muncaster, S., Norberg, B. & Andersson, E. Natural sex change in the temperate protogynous
468 Ballan wrasse *Labrus bergylta*. *J Fish Biol* **82**, 1858–1870, doi:10.1111/jfb.12113 (2013).
469 9 Bhandari, R. K., Komuro, H., Nakamura, S., Higa, M. & Nakamura, M. Gonadal restructuring and
470 correlative steroid hormone profiles during natural sex change in protogynous honeycomb grouper
471 (*Epinephelus merra*). *Zoolog Sci* **20**, 1399–1404 (2003).

10 Horiguchi, R. *et al.* Characterization of gonadal soma-derived factor expression during sex change in the protogynous wrasse, *Halichoeres trimaculatus*. *Dev Dyn* **242**, 388–399, doi:10.1002/dvdy.23929 (2013).

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* **82**, 859–866, doi:10.1002/mrd.22527 (2015).

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* **257**, 67–73, doi:10.1016/j.ygcen.2017.06.017 (2018).

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*. *Biology of sex differences* **1**, 3, doi:10.1186/2042-6410-1-3 (2010).

14 Thomas, J. T. *et al.* Conservation and diversity in expression of candidate genes regulating socially-induced female-male sex change in wrasses. *PeerJ* **7**, doi:10.7717/peerj.7032 (2019).

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* **29**, 690-701, doi:10.2108/zsj.29.690 (2012).

16 Senthilkumaran, B. *et al.* 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* **12**, doi:10.1371/journal.pone.0173974 (2017).

17 Todd, E. V. *et al.* Stress, novel sex genes, and epigenetic reprogramming orchestrate socially controlled sex change. *Science Advances* **5**, doi:10.1126/sciadv.aaw7006 (2019).

18 He, Z. *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* **552**, doi:10.1016/j.aquaculture.2022.737918 (2022).

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 zoology* **7**, doi:10.1186/s40850-022-00155-4 (2022).

20 Wu, X. *et al.* Integration of ATAC-seq and RNA-seq unravels chromatin accessibility during sex reversal in Orange-spotted grouper (*Epinephelus coioides*). *International Journal of Molecular Sciences* **21**, doi:10.3390/ijms21082800 (2020).

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* **154**, 2881–2890, doi:10.1210/en.2012-2220 (2013).

22 Li, G. L., Liu, X. C., Zhang, Y. & Lin, H. R. 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. *Aquaculture research* **37**, 484–491 (2006).

23 Wang, D. S. *et al.* 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 **21**, 712–725, doi:10.1210/me.2006-0248 (2007).

24 Zhang, C. *et al.* Transcription factors foxl2 and foxl3 regulate cyp19a1a and cyp11b in orange-spotted grouper (*Epinephelus coioides*). *Aquaculture reports* **25**, doi:10.1016/j.aqrep.2022.101243 (2022).

25 Shibata, Y. *et al.* Expression of gonadal soma derived factor (GSDF) is spatially and temporally correlated with early testicular differentiation in medaka. *Gene Expr Patterns* **10**, 283–289, doi:10.1016/j.gep.2010.06.005 (2010).

26 Zhang, X. *et al.* Autosomal gsdf acts as a male sex initiator in the fish medaka. *Sci Rep* **6**, doi:10.1038/srep19738 (2016).

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* **237**, 297–306, doi:10.1002/dvdy.21409 (2008).

28 Rodríguez-Marí, A. *et al.* Characterization and expression pattern of zebrafish anti-Müllerian hormone (amh) relative to sox9a, sox9b, and cyp19a1a, during gonad development. *Gene expression patterns* **5**, 655–667, doi:10.1016/j.modgep.2005.02.008 (2005).

29 Wang, X. G. & 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* **236**, 1329–1338, doi:10.1002/dvdy.21129 (2007).

30 Skaar, K. S. *et al.* Proteolytically activated, recombinant anti-mullerian hormone inhibits androgen secretion, proliferation, and differentiation of spermatogonia in adult zebrafish testis organ cultures. *Endocrinology* **152**, 3527–3540, doi:10.1210/en.2010-1469 (2011).

31 Webster, K. A. *et al.* Dmrt1 is necessary for male sexual development in zebrafish. *Developmental biology* **422**, 33–46, doi:10.1016/j.ydbio.2016.12.008 (2017).

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. *Molecular biology reports* **43**, 629–637, doi:10.1007/s11033-016-3997-8 (2016).

33 Bono, H. Meta-analysis of oxidative transcriptomes in insects. *Antioxidants* **10**, doi:10.3390/antiox10030345 (2021).

34 Rahman, M. R. *et al.* Comprehensive analysis of RNA-Seq gene expression profiling of brain transcriptomes reveals novel genes, regulators, and pathways in autism spectrum disorder. *Brain sciences* **10**, doi:10.3390/brainsci10100747 (2020).

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. *International Journal of Molecular Sciences* **24**, doi:10.3390/ijms24098353 (2023).

36 Cheng, H., Concepcion, G. T., Feng, X., Zhang, H. & Li, H. Haplotype-resolved de novo assembly using phased assembly graphs with hifiasm. *Nature methods* **18**, 170–175, doi:10.1038/s41592-020-01056-5 (2021).

37 Manni, M. *et al.* BUSCO update: novel and streamlined workflows along with broader and deeper phylogenetic coverage for scoring of eukaryotic, prokaryotic, and viral genomes. *Molecular Biology and Evolution* **38**, 4647–4654, doi:10.1093/molbev/msab199 (2021).

38 Haas, B. *TransDecoder (v5.5.0)*, <https://github.com/TransDecoder/TransDecoder> (

39 Ge, H. *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* **19**, 1461–1469, doi:10.1111/1755-0998.13064 (2019).

40 Pertea, M. *et al.* StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nature biotechnology* **33**, 290–295, doi:10.1038/nbt.3122 (2015).

41 Grabherr, M. G. *et al.* Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nature biotechnology* **29**, 644–652, doi:10.1038/nbt.1883 (2011).

42 Patro, R., Duggal, G., Love, M. I., Irizarry, R. A. & Kingsford, C. Salmon provides fast and bias-aware quantification of transcript expression. *Nature methods* **14**, 417–419, doi:10.1038/nmeth.4197 (2017).

43 Pearson, W. R. & Lipman, D. J. Improved tools for biological sequence comparison. *Proceedings of the National Academy of Sciences* **85**, 2444–2448, doi:10.1073/pnas.85.8.2444 (1988).

44 Cunningham, F. *et al.* Ensembl 2022. *Nucleic Acids Res* **50**, D988–D995, doi:10.1093/nar/gkab1049 (2022).

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* **161**, 360–364, doi:10.1016/j.ygcen.2009.01.024 (2009).

46 Kroon, F. J. & Liley, N. R. The role of steroid hormones in protogynous sex change in the Blackeye goby, *Coryphopterus nicholsii* (Teleostei: Gobiidae). *Gen Comp Endocrinol* **118**, 273–283, doi:10.1006/gcen.2000.7459 (2000).

47 Hasegawa, S. *et al.* Acetoacetyl-CoA synthetase, a ketone body-utilizing enzyme, is controlled by SREBP-2 and affects serum cholesterol levels. *Molecular Genetics and Metabolism* **107**, 553–560, doi:10.1016/j.ymgme.2012.08.017 (2012).

48 Wu, D.-M. *et al.* DNA hypermethylation of acetoacetyl-CoA synthetase contributes to inhibited cholesterol supply and steroidogenesis in fetal rat adrenals under prenatal nicotine exposure. *Toxicology* **340**, 43–52, doi:10.1016/j.tox.2016.01.002 (2016).

49 Gaudet, P., Livstone, M. S., Lewis, S. E. & Thomas, P. D. Phylogenetic-based propagation of functional annotations within the Gene Ontology consortium. *Briefings in Bioinformatics* **12**, 449–462, doi:10.1093/bib/bbr042 (2011).

50 Blue, M. L., Williams, D. L., Zucker, S., Khan, S. A. & Blum, C. B. Apolipoprotein E synthesis in human kidney, adrenal gland, and liver. *Proc Natl Acad Sci USA* **80**, 283–287, doi:10.1073/pnas.80.1.283 (1983).

51 Reyland, M. E., Gwynne, J. T., Forgez, P., Prack, M. M. & Williams, D. L. Expression of the human apolipoprotein E gene suppresses steroidogenesis in mouse Y1 adrenal cells. *Proc Natl Acad Sci USA* **88**, 2375–2379, doi:10.1073/pnas.88.6.2375 (1991).

52 Thorngate, F. E., Strockbine, P. A., Erickson, S. K. & Williams, D. L. Altered adrenal gland cholesterol metabolism in the apoE-deficient mouse. *J Lipid Res* **43**, 1920–1926, doi:10.1194/jlr.M200205-JLR200 (2002).

53 Payne, A. H. & Hales, D. B. Overview of steroidogenic enzymes in the pathway from cholesterol

592 to active steroid hormones. *Endocr Rev* **25**, 947–970, doi:10.1210/er.2003-0030 (2004).
593 54 Pikuleva, I. A. Cholesterol-metabolizing cytochromes P450. *Drug metabolism and disposition :*
594 *the biological fate of chemicals* **34**, 513–520, doi:10.1124/dmd.105.008789 (2006).
595 55 Miller, W. L. in *Disorders of the Human Adrenal Cortex* (eds C.E. Flück & Walter L. Miller)
596 1–18 (S.Karger AG, 2008).
597 56 Sakae, Y. *et al.* Starvation causes female-to-male sex reversal through lipid metabolism in the
598 teleost fish, medaka (*Olyzias latipes*). *Biol Open* **9**, doi:10.1242/bio.050054 (2020).
599 57 Gwynne, J. T. & Strauss, J. F. The role of lipoproteins in steroidogenesis and cholesterol
600 metabolism in steroidogenic glands. *Endocr Rev* **3**, 299–329, doi:10.1210/edrv-3-3-299 (1982).
601 58 Sharpe, R. L., Drolet, M. & MacLatchy, D. L. Investigation of de novo cholesterol synthetic
602 capacity in the gonads of goldfish (*Carassius auratus*) exposed to the phytosterol beta-sitosterol.
603 *Reproductive biology and endocrinology* **4**, doi:10.1186/1477-7827-4-60 (2006).
604 59 Bera, A., Chadha, N. K., Dasgupta, S., Chakravarty, S. & Sawant, P. B. Hypoxia-mediated
605 inhibition of cholesterol synthesis leads to disruption of nocturnal sex steroidogenesis in the gonad
606 of koi carp, *Cyprinus carpio*. *Fish Physiol Biochem* **46**, 2421–2435, doi:10.1007/s10695-020-
607 00887-5 (2020).
608 60 Sim, S. B., Corpuz, R. L., Simmonds, T. J. & Geib, S. M. HiFiAdapterFilt, a memory efficient
609 read processing pipeline, prevents occurrence of adapter sequence in PacBio HiFi reads and their
610 negative impacts on genome assembly. *BMC genomics* **23**, doi:10.1186/s12864-022-08375-1
611 (2022).
612 61 Zhu, B.-H. *et al.* P_RNA_scaffolder: a fast and accurate genome scaffolder using paired-end RNA-
613 sequencing reads. *BMC genomics* **19**, doi:10.1186/s12864-018-4567-3 (2018).
614 62 Kadota, M. *et al.* Multifaceted Hi-C benchmarking: what makes a difference in chromosome-scale
615 genome scaffolding? *GigaScience* **9**, doi:10.1093/gigascience/giz158 (2020).
616 63 Durand, N. C. *et al.* Juicer provides a one-click system for analyzing loop-resolution Hi-C
617 experiments. *Cell Systems* **3**, 95–98, doi:10.1016/j.cels.2016.07.002 (2016).
618 64 Dudchenko, O. *et al.* The Juicebox assembly tools module facilitates *de novo* assembly of
619 mammalian genomes with chromosome-length scaffolds for under $1000. *bioRxiv*,
620 doi:10.1101/254797 (2018).
621 65 Nishimura, O., Hara, Y., Kuraku, S. & Hancock, J. gVolante for standardizing completeness
622 assessment of genome and transcriptome assemblies. *Bioinformatics* **33**, 3635–3637,
623 doi:10.1093/bioinformatics/btx445 (2017).
624 66 Kodama, Y., Shumway, M. & Leinonen, R. The sequence read archive: explosive growth of
625 sequencing data. *Nucleic Acids Res* **40**, D54–D56, doi:10.1093/nar/gkr854 (2011).
626 67 *Trim Galore*, <https://github.com/FelixKrueger/TrimGalore> (
627 68 Kim, D., Paggi, J. M., Park, C., Bennett, C. & Salzberg, S. L. Graph-based genome alignment and
628 genotyping with HISAT2 and HISAT-genotype. *Nature biotechnology* **37**, 907–915,
629 doi:10.1038/s41587-019-0201-4 (2019).
630 69 Danecek, P. *et al.* Twelve years of SAMtools and BCFtools. *GigaScience* **10**,
631 doi:10.1093/gigascience/giab008 (2021).

70 Leek, J. T. *et al.* Tackling the widespread and critical impact of batch effects in high-throughput data. *Nature reviews. Genetics* **11**, 733–739, doi:10.1038/nrg2825 (2010).

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* **11**, doi:10.1371/journal.pone.0163962 (2016).

72 R Core Team. *R: A language and environment for statistical computing.*, <URL https://www.R-project.org/> (2022).

73 Pertea, G. & Pertea, M. GFF Utilities: GffRead and GffCompare. *F1000Research* **9**, doi:10.12688/f1000research.23297.1 (2020).

74 Buchfink, B., Xie, C. & Huson, D. H. Fast and sensitive protein alignment using DIAMOND. *Nature methods* **12**, 59–60, doi:10.1038/nmeth.3176 (2014).

75 Zhou, Y. *et al.* Metascape provides a biologist-oriented resource for the analysis of systems-level datasets. *Nature communications* **10**, doi:10.1038/s41467-019-09234-6 (2019).

76 Murphy, R. F. *et al.* PathVisio 3: An Extendable Pathway Analysis Toolbox. *PLoS Comput Biol* **11**, doi:10.1371/journal.pcbi.1004085 (2015).

77 Pico, A. R. *et al.* WikiPathways: Pathway Editing for the People. *PLoS Biol* **6**, doi:10.1371/journal.pbio.0060184 (2008).

78 Kemski, M. M., Rappleye, C. A., Dabrowski, K., Bruno, R. S. & Wick, M. Transcriptomic response to soybean meal-based diets as the first formulated feed in juvenile yellow perch (*Perca flavescens*). *Sci Rep* **10**, doi:10.1038/s41598-020-59691-z (2020).

79 Liu, Y. *et al.* Single-cell transcriptome reveals insights into the development and function of the zebrafish ovary. *Elife* **11**, doi:10.7554/eLife.76014 (2022).

80 Craft, J. A., Bain, P. A., 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* **10**, doi:10.1371/journal.pone.0142636 (2015).

81 Nagahama, Y., Chakraborty, T., Paul-Prasanth, B., Ohta, K. & Nakamura, M. Sex determination, gonadal sex differentiation, and plasticity in vertebrate species. *Physiol Rev* **101**, 1237–1308, doi:10.1152/physrev.00044.2019 (2021).

82 Fernandino, J. I., Hattori, R. S., Moreno Acosta, O. D., Strussmann, C. A. & Somoza, G. M. Environmental stress-induced testis differentiation: androgen as a by-product of cortisol inactivation. *Gen Comp Endocrinol* **192**, 36-44, doi:10.1016/j.ygcen.2013.05.024 (2013).

83 Rabosky, D. L. *et al.* An inverse latitudinal gradient in speciation rate for marine fishes. *Nature* **559**, 392-395, doi:10.1038/s41586-018-0273-1 (2018).

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.

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.

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).

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

Competing interests

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

721 **Figure legends**

722 **Fig. 1. Summarized phylogenetic tree of Actinopterygii and schematic diagram of the**
723 **analysis pipeline in this study. a** The order-level phylogenetic tree is shown. The phylogenetic
724 tree was derived from Daniel et al., (2018)[83] and modified. Diamonds indicate orders with
725 sequential hermaphroditism. Numbers in parentheses indicate the number of summarized
726 orders. Orders containing the fish species used in this study are shown in bold. **b** The consistent
727 line style of the arrows indicates the utilization of common tools for conducting data processing.
728 The letters in parentheses next to the species names represent the following order names; L:
729 Labriformes, P: Perciformes, S: Synbranchiformes.

730

731 **Fig. 2. Scatter plot of the total Intersexual gonad and Ovary (IO)-score and Composition**
732 **of total IO-scores for up- and downregulated factors. a** Scatterplot of total IO-scores for the
733 7,289 factors in descending order. Red circles indicate 99 upregulated factors; blue circles
734 indicate 69 downregulated factors. Gray circles (which have the appearance of lines because
735 they overlap each other) indicate unchanged factors ($6 \leq$ total IO-score ≤ -6). Composition of
736 total IO-scores for up- (**b**, n=99) and downregulated factors (**c**, n=69). Numbers in parentheses
737 next to species name indicate the number of RNA-Seq data pairs for each species used in this
738 study. Asterisk indicates a factor for which the total IO-score is derived from a single species.

739

740 **Fig. 3. Results of enrichment analysis for differentially expressed factors using Metascape.**
741 **a** 99 upregulated factors. **b** 69 downregulated factors.

742

743 **Fig. 4. Hypothetical ovarian cholesterol biosynthesis pathway in protogynous fishes.** The
744 pathway was visualized using PathVisio (https://pathvisio.org), referencing several fish
745 cholesterol synthesis pathways. Gene products are represented in black text, while metabolites
746 are in blue text. Gene product annotations follow Ensembl Zebrafish gene symbols. Red-filled

747 boxes represent gene products with confirmed upregulated expression in this study, while blue-
748 filled boxes denote gene products with downregulated expression. Open boxes represent gene
749 products with no expression variation. Light green-filled boxes represent gene products not
750 included among the 7289 factors but confirmed to be present in at least the three-spot wrasse
751 transcriptome. Dark gray-filled boxes indicate gene products that could not be identified within
752 the three-spot wrasse transcriptome. Circles extending from gene products represent catalytic
753 reactions, while black arrows indicate the conversion of metabolites. The dashed arrows
754 indicate indirect pathways summarizing multi-step processes.
755
756

757 **Table 1.** Statistics of genome assembly in Three-spot wrasse (fHalTri1.1) and other Labidae

758 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

759 * BUSCO version 5.4, reference datasets: vertebrata_odb10

760

761

762

763

764

765

766

767

768

769

770

771

772

773 **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

774 * Based on the annotation file (merged gtf file), the extracted transcripts sequences from the
775 genome by gffread
776 ** Among transcriptome sequences (594,726 seqs) assembled by Trinity, transcripts sequences
777 corresponding to coding seqences predicted via TransDecoder
778
779
780
781
782
783
784
785
786
787
788
789
790

791 **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*

792 * Gene symbols follow the zebrafish gene nomenclature in Ensembl.

793

Fig. 1.

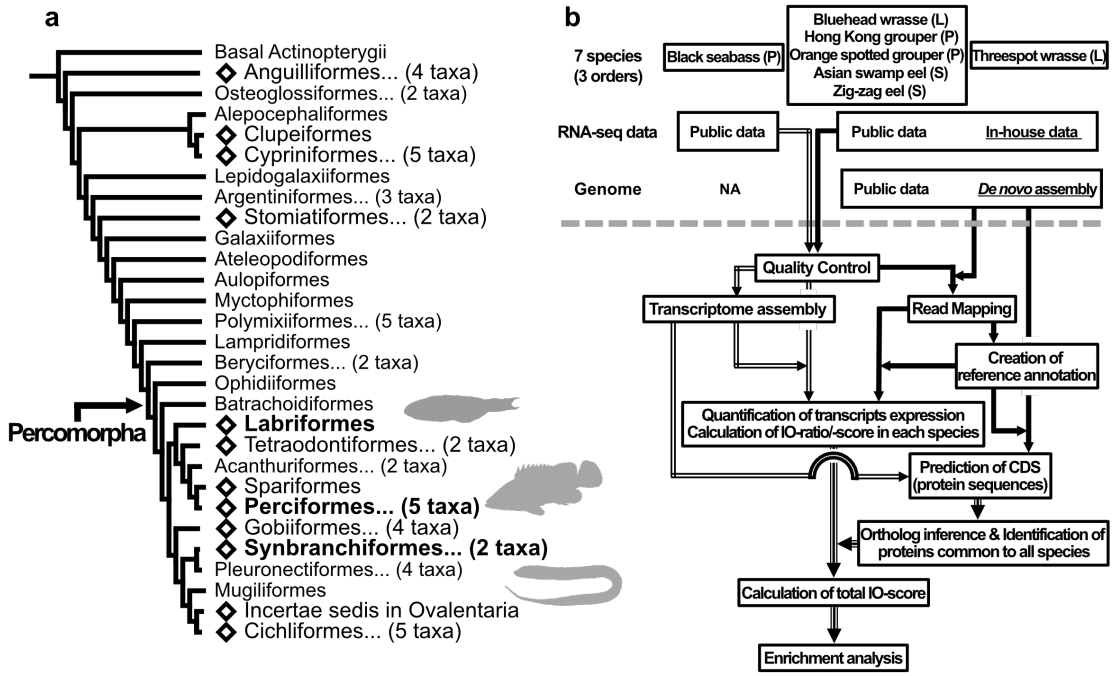

a
Basal Actinopterygii
Anguilliformes... (4 taxa)
Osteoglossiformes... (2 taxa)
Alepocephaliformes
Clupeiformes
Cypriniformes... (5 taxa)
Lepidogalaxiiformes
Argentiniformes... (3 taxa)
Stomiatiformes... (2 taxa)
Galaxiiformes
Ateleopodiformes
Aulopiformes
Myctophiformes
Polymixiiformes... (5 taxa)
Lampridiformes
Beryciformes... (2 taxa)
Ophidiiformes
Batrachoidiformes
Percomorpha
Labriformes
Tetraodontiformes... (2 taxa)
Acanthuriformes... (2 taxa)
Spariformes
Perciformes... (5 taxa)
Gobiiformes... (4 taxa)
Synbranchiformes... (2 taxa)
Pleuronectiformes... (4 taxa)
Mugiliformes
Incertae sedis in Ovalentaria
Cichliformes... (5 taxa)
b
7 species
(3 orders)
Black seabass (P)
Bluehead wrasse (L)
Hong Kong grouper (P)
Orange spotted grouper (P)
Asian swamp eel (S)
Zig-zag eel (S)
Threespot wrasse (L)
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.

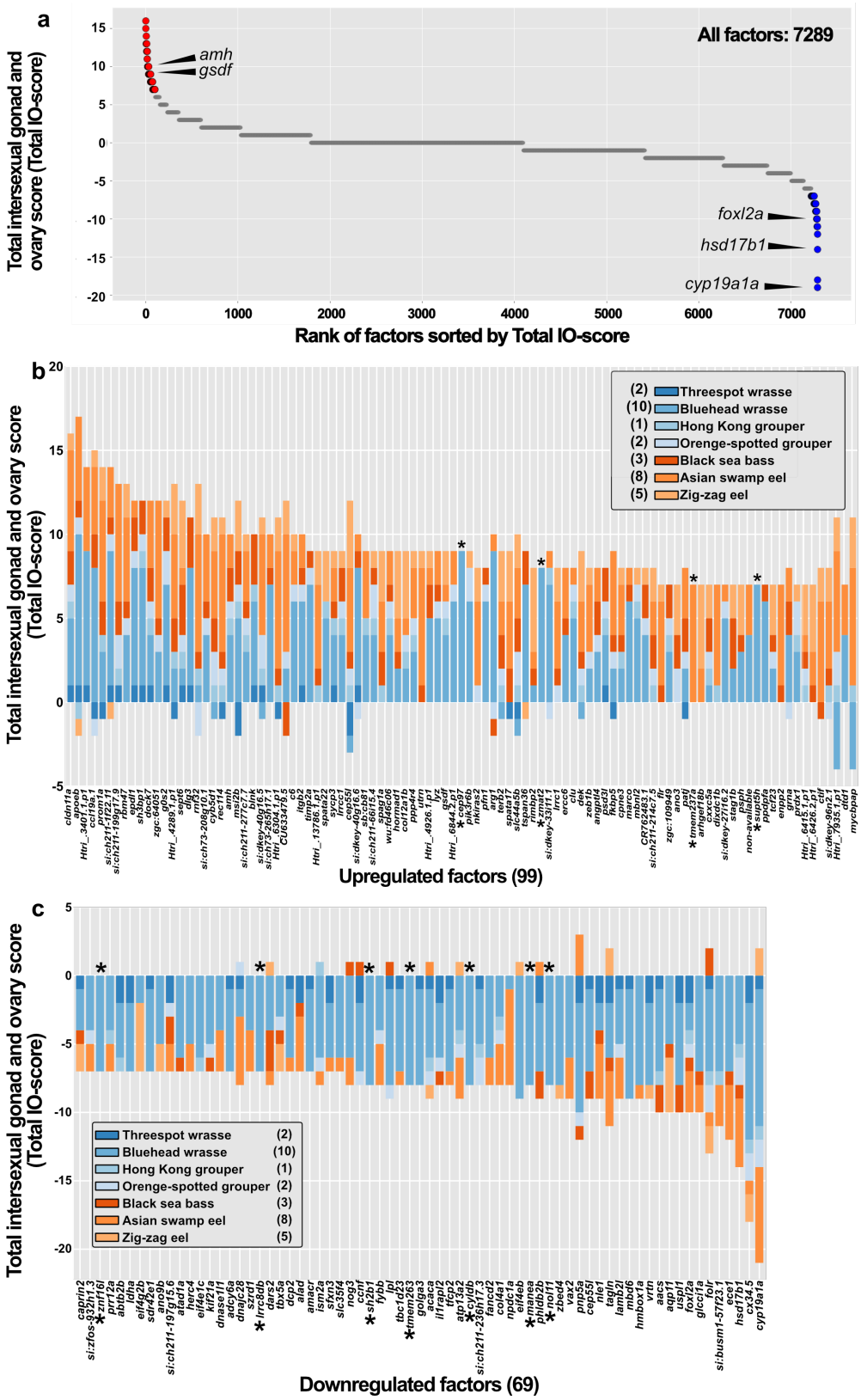

a
All factors: 7289
amh
gsdf
foxl2a
hsd17b1
cyp19a1a
Total intersexual gonad and ovary score (Total IO-score)
Rank of factors sorted by Total IO-score
b
(2) Threespot wrasse
(10) Bluehead wrasse
(1) Hong Kong grouper
(2) Orenge-spotted grouper
(3) Black sea bass
(8) Asian swamp eel
(5) Zig-zag eel
Total intersexual gonad and ovary score (Total IO-score)
Upregulated factors (99)
c
Threespot wrasse (2)
Bluehead wrasse (10)
Hong Kong grouper (1)
Orenge-spotted grouper (2)
Black sea bass (3)
Asian swamp eel (8)
Zig-zag eel (5)
Total intersexual gonad and ovary score (Total IO-score)
Downregulated factors (69)

Fig. 3.

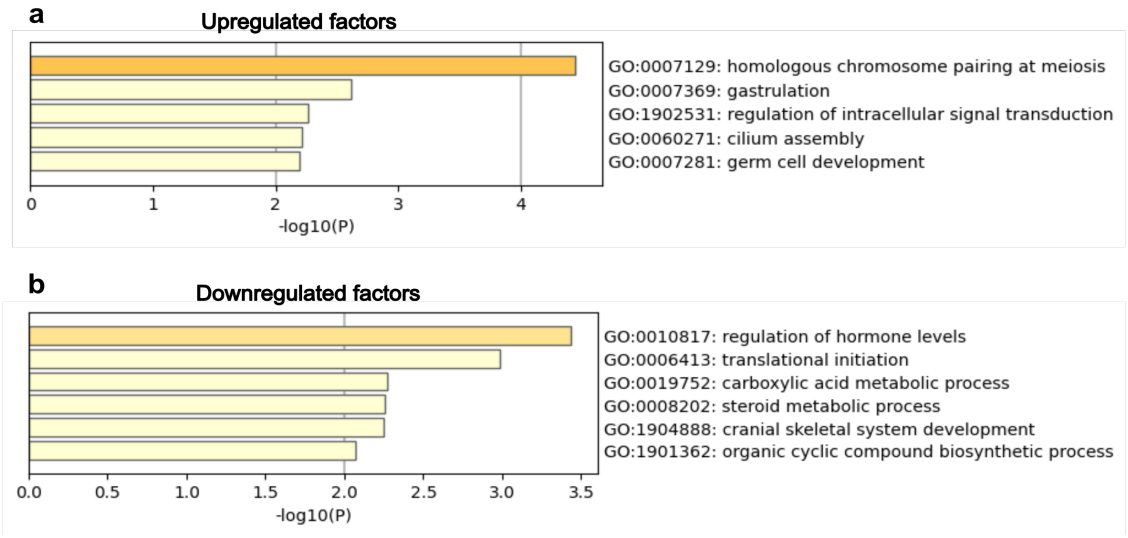
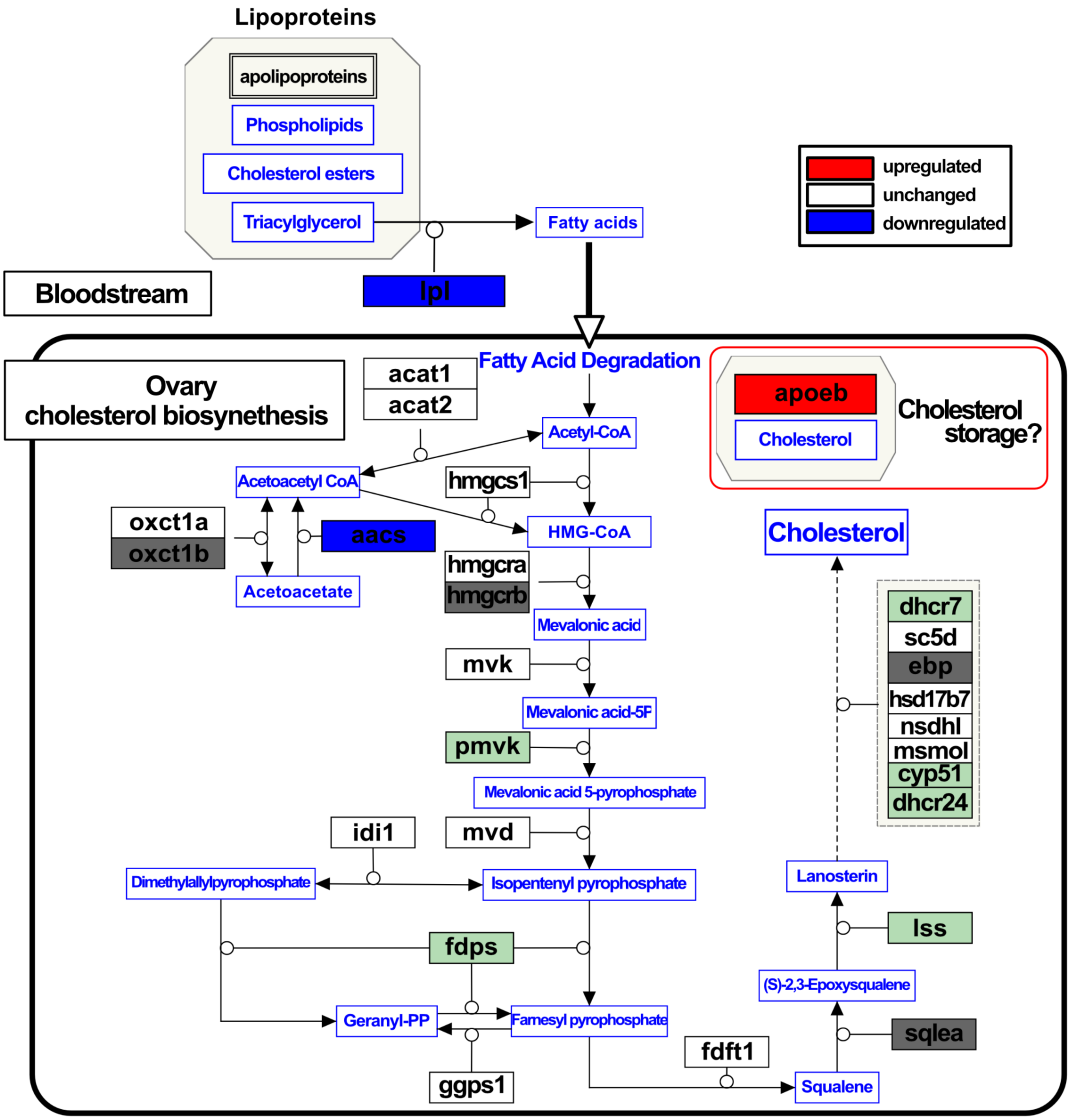

a
Upregulated factors
GO:0007129: homologous chromosome pairing at meiosis
GO:0007369: gastrulation
GO:1902531: regulation of intracellular signal transduction
GO:0060271: cilium assembly
GO:0007281: germ cell development
0
1
2
3
4
-log10(P)
b
Downregulated factors
GO:0010817: regulation of hormone levels
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. 4.


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