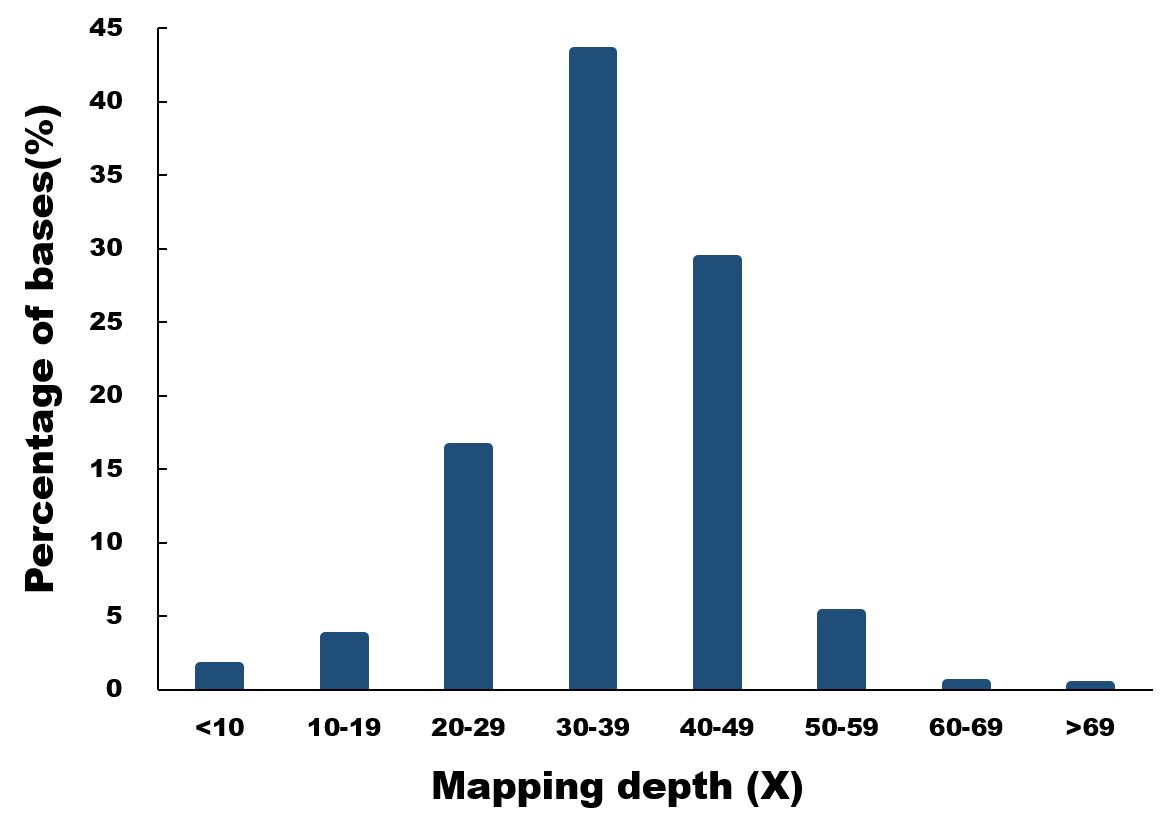
Supplementary Materials for

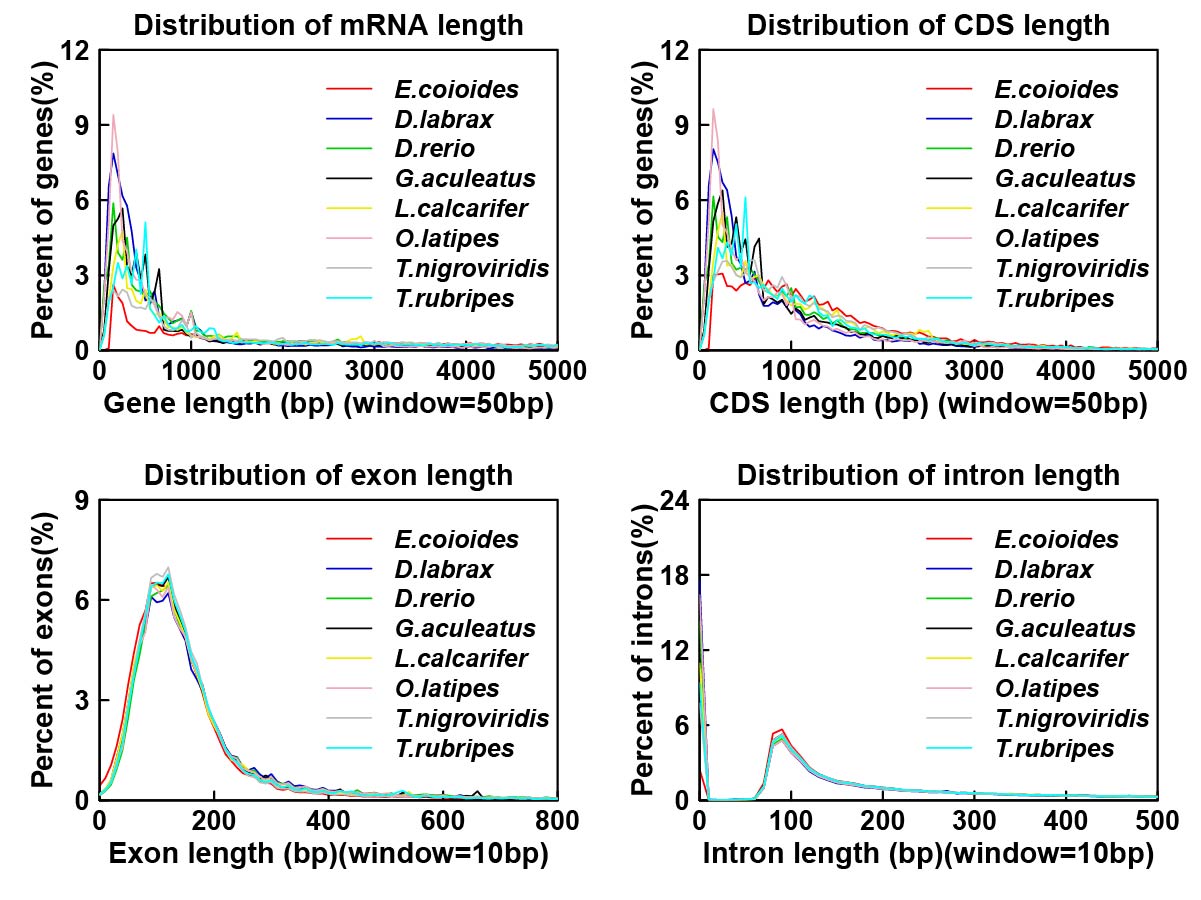
**The genome and transcriptome of protogynous orange-spotted grouper (*Epinephelus coioides*) provide insights into mechanisms of sex differentiation and sex reversal**

**Supplementary Figures**

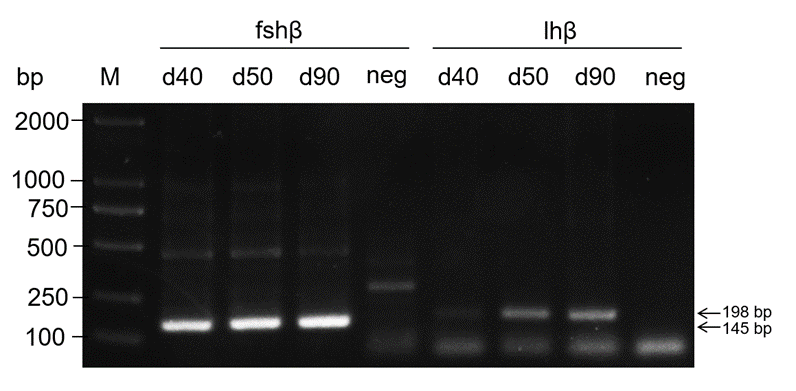
**Supplementary Figure 1| Estimation of the *E. coioides* genome size based on K-mer analysis.** Genome size G is estimated as G = K\_num/ Peak\_depth, where the K\_num is the total number of K-mers, and the Peak\_depth is the expected value of K-mer depth (in this graph it is 20). The X-axis is the depth of K-mers derived from the sequenced reads, and the Y-axis is the frequency of the K-mer depth.



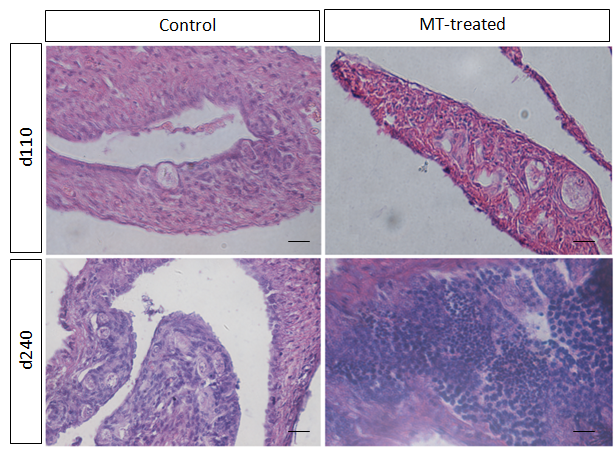
**Supplementary Figure 2| Depth of single-base distribution based on short-read alignment.** To validate genome assembly completeness, PacBio reads were aligned against the assembly using the blasr software.



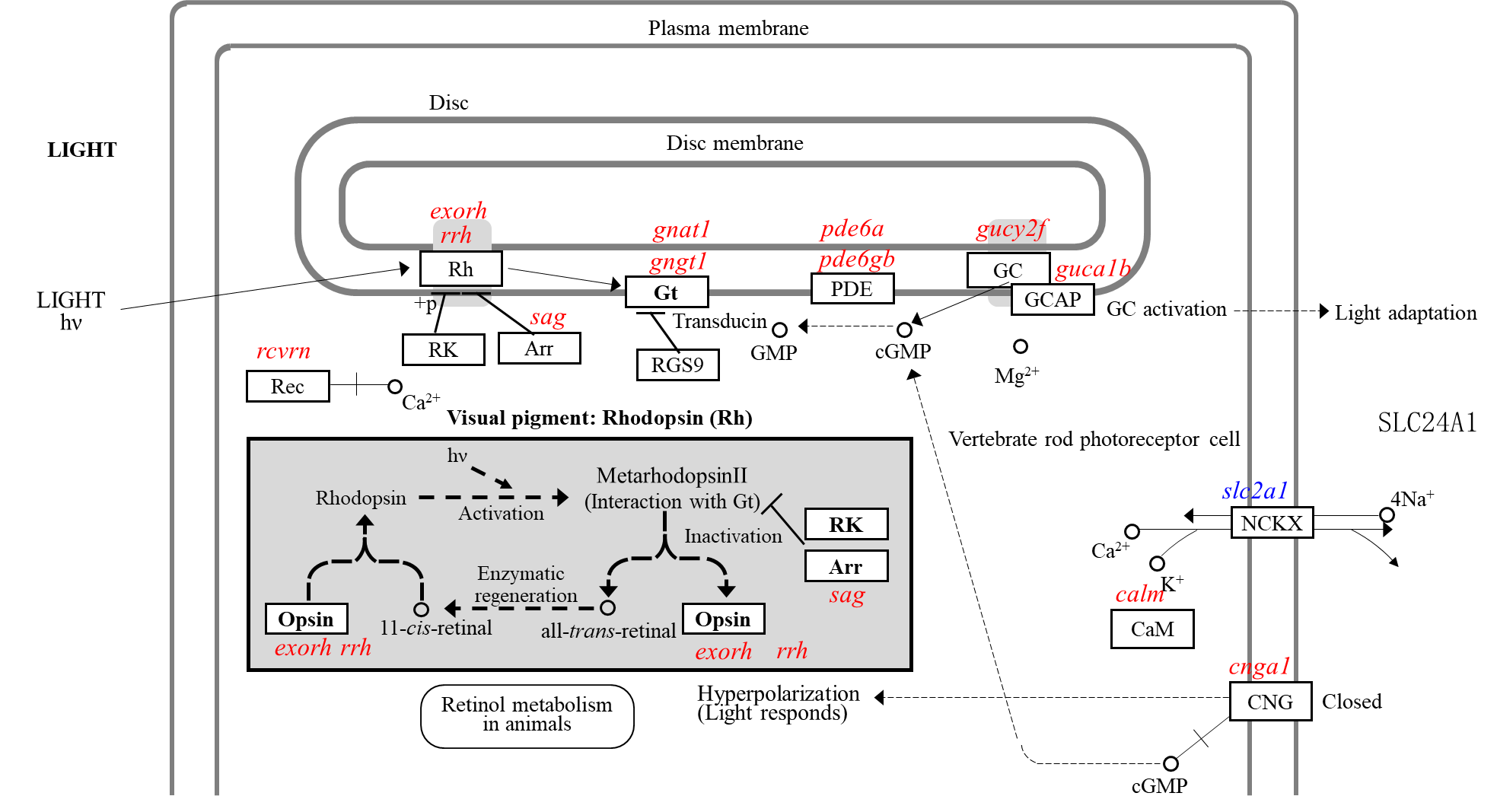
**Supplementary Figure 3| CDS length, exon length, intron length and gene length distributions in the assembled genomes of *E. coioides* and seven other teleost species.**



**Supplementary Figure 4|** **RT-PCR analysis of *fshβ* and *lhβ* expression at 40, 50 and 90 dpf. M, marker; Neg, negative control.** Target bands are indicated by arrows.



**Supplementary Figure 5| Gonadal histology of experiment fish.** Gonadal tissues from control and MT-feeding group were collected at 110 dpf and 240 dpf. Scar bar, 25 um.



**Supplementary Figure 6| Activation of phototransduction pathways in dominant female fish.** Gene symbols in red and blue indicate up- and downregulated genes, respectively. This figure was adapted from the KEGG phototransduction pathway.

**Supplementary Tables**

**Supplementary Table 1| Statistics of PacBio and Illumina sequencing data for the *E. coioides* genome.**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Sequencing** | **Insert Size** | **Total Data (Gb)** | **Reads Length** | **Sequence coverage (×)** |
| Illumina (paired-end) | 170 bp | 26.58 | 100x2 | 25.31 |
| 500 bp | 28.35 | 100x2 | 27.00 |
| 800 bp | 11.02 | 100x2 | 10.50 |
| PacBio | 20 kb | 54.36 | -- | 51.77 |
| **Total** | -- | 120.31 | -- | 114.58 |

Note: Assuming the genome size is 1.05 Gb.

**Supplementary Table 2| Statistics of the 17-kmer analysis.**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Kmer** | **Kmer num** | **Kmer depth** | **genome size** | **used base** | **used read** | **X** |
| 17 | 22,097,528,231 | 20 | 1,050,876,411 | 32,920,770,450 | 658,415,409 | 29.4118 |

**Supplementary Table 3: Statistics of de novo assembly of *E. coioides* genome.**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Parameter** | **Scaffold** | | **Contig** | |
| **Size (bp)** | **Number** | **Size (bp)** | **Number** |
| N90 | 457,668 | 468 | 454,834 | 471 |
| N80 | 911,490 | 314 | 907,401 | 316 |
| N70 | 1,407,599 | 225 | 1,373,451 | 226 |
| N60 | 1,975,747 | 163 | 1,949,530 | 164 |
| N50 | 2,567,054 | 117 | 2,494,093 | 118 |
| Longest | 11,453,851 |  | 11,453,851 |  |
| Total Size | 1,023,562,069 |  | 1,023,556,695 |  |
| Total Number(>500bp) |  | 1,450 |  | 1,459 |
| Total Number(>2kb) |  | 1,450 |  | 1,457 |

**Supplementary Table 4| Assessment of sequence coverage of the genome assembly by homology search with *de novo* assembled transcriptome data.**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Transcription Data** | **Number** | **align-hit in one scaffold** | | **with >90% sequence in one scaffold** | |
| **Number** | **Percentage (%)** | **Number** | **Percentage (%)** |
| >200 bp | 236,140 | 234,671 | 99.38% | 224,557 | 95.09% |
| >500 bp | 128,195 | 127,791 | 99.68% | 121,395 | 94.70% |
| >1000 bp | 88,145 | 87,998 | 99.83% | 83,240 | 94.44% |

**Supplementary Table 5| Assessment of genome assemblies and gene sets in BUSCO notation (C: complete [D: duplicated], F: fragmented, M: missing, n: gene number)**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Species | **Genome size/Gene Number** | **Complete Single-Copy BUSCOs** | **Complete Duplicated BUSCOs** | **Fragmented BUSCOs** | **Missing BUSCOs** | **Total searched BUSCOs** |
| ***E.coioides*** | genome | 97.00% | 2.80% | 1.50% | 1.50% | 4,584 |
|  | 26934 genes | 92.80% | 2.90% | 4.60% | 2.60% | 4,584 |
| ***D.rerio*** | genome | 91.20% | 3.40% | 4.50% | 4.30% | 4,584 |
|  | 26459 genes | 96.30% | 6.30% | 1.90% | 1.80% | 4,584 |
| ***O.niloticus*** | genome | 95.90% | 2.00% | 2.50% | 1.60% | 4,584 |
|  | 26763 genes | 98.90% | 24.10% | 0.60% | 0.50% | 4,584 |
| ***G.aculeatus*** | genome | 96.50% | 2.00% | 1.90% | 1.60% | 4,584 |
|  | 27576 genes | 97.50% | 23.00% | 1.90% | 0.60% | 4,584 |

**Supplementary Table 6| Statistics ofeach pseudo-chromosome.**

|  |  |  |  |
| --- | --- | --- | --- |
| **Chromosome** | **Length (Mb)** | **Scaffold number** | **% of genome** |
| 1 | 38.61 | 24 | 3.77 |
| 2 | 38.17 | 25 | 3.73 |
| 3 | 22.34 | 17 | 2.18 |
| 4 | 46.80 | 28 | 4.57 |
| 5 | 37.69 | 25 | 3.68 |
| 6 | 37.79 | 21 | 3.69 |
| 7 | 38.74 | 24 | 3.78 |
| 8 | 53.14 | 24 | 5.19 |
| 9 | 48.73 | 33 | 4.76 |
| 10 | 39.56 | 30 | 3.87 |
| 11 | 39.22 | 19 | 3.83 |
| 12 | 35.65 | 18 | 3.48 |
| 13 | 44.38 | 66 | 4.34 |
| 14 | 42.44 | 34 | 4.15 |
| 15 | 37.78 | 27 | 3.69 |
| 16 | 35.24 | 26 | 3.44 |
| 17 | 37.20 | 25 | 3.63 |
| 18 | 31.19 | 45 | 3.05 |
| 19 | 32.92 | 15 | 3.22 |
| 20 | 21.85 | 16 | 2.13 |
| 21 | 43.52 | 38 | 4.25 |
| 22 | 39.17 | 33 | 3.83 |
| 23 | 41.88 | 19 | 4.09 |
| 24 | 43.11 | 33 | 4.21 |

**Supplementary Table 7| Statistics of repeats in the *E. coioides* genome.**

|  |  |  |
| --- | --- | --- |
| **Type** | **Repeat Size** | **% of genome** |
| ***Trf*** | 30,119,396 | 2.94 |
| ***Repeatmasker*** | 103,041,340 | 10.07 |
| ***Proteinmask*** | 42,674,494 | 4.17 |
| ***De novo*** | 392,192,794 | 38.32 |
| **Total** | 422,160,512 | 41.24 |

**Supplementary Table 8| TE content in the assembled genome.**

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Type** | **RepBase TEs** | | **TE Proteins** | | ***De novo*** | | **Combined TEs** | |
|  | **Length（bp）** | **%in Genome** | **Length（bp）** | **% in Genome** | **Length（bp）** | **% in Genome** | **Length（bp）** | **% in Genome** |
|
| DNA | 34,979,426 | 3.43 | 8,478,084 | 0.83 | 156,992,557 | 15.38 | 174,827,228 | 17.13 |
| LINE | 23,005,923 | 2.25 | 29,009,006 | 2.84 | 135,699,697 | 13.3 | 152,466,521 | 14.94 |
| LTR | 6,600,312 | 0.65 | 5,097,646 | 0.5 | 28,608,537 | 2.8 | 33,592,106 | 3.29 |
| SINE | 6,374,798 | 0.62 | 0 | 0 | 43,224,347 | 4.24 | 46,234,893 | 4.53 |
| Other | 11,518 | 0 | 0 | 0 | 0 | 0 | 11,518 | 0 |
| Unknown | 950,014 | 0.09 | 0 | 0 | 85,076,375 | 8.34 | 85,841,116 | 8.41 |
| Total | 68,940,074 | 6.76 | 42,565,081 | 4.17 | 363,118,277 | 35.58 | 382,503,106 | 37.48 |

**Supplementary Table 9| *De novo* annotated repeats in the assembled genome.**

|  |  |  |
| --- | --- | --- |
|  | | |
| **Type** | Length (bp) | % in genome |
| **DNA** | 166,052,858 | 16.22 |
| **LINE** | 152,461,353 | 14.90 |
| **SINE** | 11,200,708 | 1.09 |
| **LTR** | 58,768,563 | 5.74 |
| **Other** | 0 | 0.00 |
| **Satellite** | 2,764,416 | 0.27 |
| **Simple repeat** | 11,924,508 | 1.17 |
| **Unknown** | 38,191,680 | 3.73 |
| **Total** | 392,192,794 | 38.32 |

**Supplementary Table 10| Statistics of functional annotation.**

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Values** | **Total** | **Nr** | **Swissprot** | **KEGG** | **TrEMBL** | **Interpro** | **GO** | **Overall** |
| **Number** | 26,931 | 25,213 | 22,251 | 22,160 | 25,173 | 23,193 | 15,193 | 25,345 |
| **Percentage** | 100% | 93.62% | 82.62% | 82.28% | 93.47% | 86.12% | 62.20% | 94.11% |

**Supplementary Table 11| Summary of genes and gene families in the 10 representative species.**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Species** | **Total genes** | **Unclustered genes** | **Families** | **Unique families** | **Average genes per family** |
| ***D.rerio*** | 26,405 | 980 | 14,466 | 388 | 1.76 |
| ***E.coioides*** | 24,425 | 1,690 | 14,247 | 532 | 1.6 |
| ***G.aculeatus*** | 20,772 | 882 | 13,576 | 161 | 1.47 |
| ***O.latipes*** | 19,671 | 854 | 13,055 | 228 | 1.44 |
| ***T.rubripes*** | 18,507 | 199 | 12,642 | 62 | 1.45 |
| ***T.nigroviridis*** | 19,583 | 632 | 13,038 | 270 | 1.45 |
| ***D.labrax*** | 25,050 | 1,321 | 14,999 | 281 | 1.58 |
| ***L.calcarifer*** | 21,370 | 403 | 13,981 | 207 | 1.5 |
| ***O.niloticus*** | 21,437 | 138 | 13,357 | 44 | 1.59 |
| ***H.sapiens*** | 22,256 | 1,474 | 15,116 | 1,604 | 1.37 |

Unclustered genes refer to species-specific genes; unique families refer to species-specific gene families.

**Supplementary Table 12| The expanded genes in *E.coioides* and functional annotations**

(See separate excel file)**.**

**Supplementary Table 13| The contracted genes in *E.coioides* and function annotations**

(see separate excel file)**.**

**Supplementary Table 14| The 64 sex differentiation related genes**

(see separate excel file)**.**

**Supplementary Table S15| Summary of sequenced and mapping statistics.**

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Sample** | **Sample’s name in database** | | | **Q20 (%)** | **Coverage rate (%)** | **Mapping reads (%)** | **Sequence depth** |
| primary-male | | A12sex | 94.50 | | 98.26 | 93.88 | 28.91 |
| primary-male | | B1sex | 95.06 | | 98.17 | 94.74 | 27.85 |
| normal-female | | F2 | 97.33 | | 97.93 | 95.74 | 30.78 |
| normal-female | | F3 | 97.18 | | 97.94 | 95.52 | 30.42 |
| normal-female | | F5 | 97.41 | | 97.94 | 95.69 | 30.49 |
| normal-female | | F6 | 96.29 | | 98.02 | 95.73 | 31.2 |
| normal-female | | F7 | 97.23 | | 97.99 | 95.92 | 30.77 |
| normal-female | | M11 | 98.19 | | 97.91 | 95.84 | 31.41 |
| normal-female | | M3 | 96.99 | | 97.99 | 95.57 | 33.23 |
| normal-female | | M7 | 97.08 | | 97.8 | 95.87 | 26.55 |
| normal-female | | M8 | 97.11 | | 98.00 | 96.02 | 31.29 |
| normal-female | | M9 | 98.37 | | 97.74 | 95.64 | 26.64 |
| normal-female | | SCA-female | 94.40 | | 97.77 | 94.68 | 22.62 |
| normal-male | | SXA-male | 96.92 | | 97.63 | 94.71 | 20.51 |

**Supplementary Table S16| Annotation statistics for population SNPs**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **SNP** | **Syn CDS** | **Nonsyn CDS** | **5’UTR** | **3’UTR** | **mRNA** |
| 17,238,457 | 219,857 | 142,225 | 34,229 | 157,191 | 6,862,753 |

**Supplementary Table S17| Annotation statistics for population indels**

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Indel** | **Insertion** | **Deletion** | **Frame-shift** | **Non-frame-shift** | **5’UTR** | **3’UTR** | **mRNA** |
| 5,262,780 | 2,123,742 | 3,139,038 | 30,385 | 8,056 | 8,715 | 50,799 | 1,968,902 |

**Supplementary Table S18| summary of significant variations between two primary-male and twelve normal individuals.**

|  |  |  |
| --- | --- | --- |
| **Category** | **Variation number** | **Gene number** |
| Non-syn SNPs | 16 | 13 |
| Large effect SNP | 1 | 1 |
| frameshift indels | 11 | 8 |

**Supplementary Table 19| Gonadal gene expression at different sex developmental stages**

(see separate excel file)**.**

**Supplementary Table 20| Expression of sex differentiation related genes**

(see separate excel file)**.**

**Supplementary Table 21| Group 1: Early sex differentiation related genes**

(see separate excel file)**.**

**Supplementary Table 22| Group 2: Early sex differentiation and male stage related genes**

(see separate excel file)**.**

**Supplementary Table 23| Group 3: Female stage related genes**

(see separate excel file)**.**

**Supplementary Table 24| Group 4: Male stage related genes**

(see separate excel file)**.**

**Supplementary Table 25| The expression of aromatase during the sex differentiation and ovarian development**

(see separate excel file)**.**

**Supplementary Table 26| MT-feeding transcriptomes**

(see separate excel file)**.**

**Supplementary Table 27| Gene expression in the brain and pituitary during sex reversal**

(see separate excel file)**.**

**Supplementary Table 28| Socially controlled sex change transcriptomes**

(see separate excel file)**.**

**Supplementary Table 29| KEGG enrichment of differentially expressed genes between FFMR and FFAS**

(see separate excel file)**.**

**Supplementary Table 30| KEGG enrichment of differentially expressed genes between FFAS and DFF**

(see separate excel file)**.**

**Supplementary Table 31| KEGG enrichment of differentially expressed genes between FFAS and SFF**

(see separate excel file)**.**

**Supplementary Table 32| KEGG enrichment of differentially expressed genes between FFMR and DFF**

(see separate excel file)**.**

**Supplementary Table 33| KEGG enrichment of differentially expressed genes between SFF and DFF**

(see separate excel file)**.**

**Supplementary Table 34| Statistic of sequencing, mapping and fish age for RNA-seq samples.**

|  |  |  |  |
| --- | --- | --- | --- |
| **Samples of sex differentiation, sex reversal and 17α-methyltestosterone (MT) treatment** | | | |
| **Sample** | **Total reads(M)** | **Age** | **Reads length** |
| ASD\_50d\_c\_G (gonad, control, 50days) | 49.17 | 50days | 2\*90 |
| ASD\_90d\_c\_G (gonad, control, 90days) | 48.37 | 90days | 2\*90 |
| ASD\_90d\_f\_G (gonad, MT-feeding , 90days) | 50.12 | 90days | 2\*90 |
| ASD\_110d\_c\_G (gonad, control, 110days) | 50.3 | 110days | 2\*90 |
| ASD\_110d\_f\_G (gonad, MT-feeding , 110days) | 50.12 | 110days | 2\*90 |
| ASD\_180d\_c\_G (gonad, control, 180days) | 48.55 | 180days | 2\*90 |
| ASD\_180d\_f\_G (gonad, MT-feeding , 180days) | 46.91 | 180days | 2\*90 |
| ASD\_240d\_c\_G (gonad, control, 240days) | 48.38 | 240days | 2\*90 |
| ASD\_240d\_f\_G (gonad, MT-feeding , 240days) | 48.4 | 240days | 2\*90 |
| NSR\_G2\_B1 (brain and pituitary) | 48.52 | 5years | 2\*90 |
| NSR\_G2\_B2 (brain and pituitary) | 45.52 | 5years | 2\*90 |
| NSR\_G2\_G1 (gonad) | 48.42 | 5years | 2\*90 |
| NSR\_G2\_G2 (gonad) | 51.71 | 5years | 2\*90 |
| NSR\_O3\_B1 (brain and pituitary) | 48.94 | 5years | 2\*90 |
| NSR\_O3\_B2 (brain and pituitary) | 49.94 | 5years | 2\*90 |
| NSR\_O3\_G1 (gonad) | 50.35 | 5years | 2\*90 |
| NSR\_O3\_G2 (gonad) | 50.02 | 5years | 2\*90 |
| NSR\_O4\_B1 (brain and pituitary) | 48.3 | 5years | 2\*90 |
| NSR\_O4\_B2 (brain and pituitary) | 49.03 | 5years | 2\*90 |
| NSR\_O4\_G1(gonad) | 41.13 | 5years | 2\*90 |
| NSR\_O4\_G2 (gonad) | 50.43 | 5years | 2\*90 |
| NSR\_EBi\_B1 (brain and pituitary) | 49.04 | 5years | 2\*90 |
| NSR\_EBi\_G1 (gonad) | 49.4 | 5years | 2\*90 |
| NSR\_LBi\_B1 (brain and pituitary) | 48.78 | 5years | 2\*90 |
| NSR\_LBi\_G1 (gonad) | 47.33 | 5years | 2\*90 |
| NSR\_T\_B1 (brain and pituitary) | 49.08 | 5years | 2\*90 |
| NSR\_T\_B2 (brain and pituitary) | 49.31 | 5years | 2\*90 |
| NSR\_T\_G1 (gonad) | 48.5 | 5years | 2\*90 |
| NSR\_T\_G2 (gonad) | 50.1 | 5years | 2\*90 |
| **Samples of socially controlled sex change experiment** | | | |
| **Sample** | **Total reads(M)** | **Age** | **Reads length** |
| B1 (brain, FFMR) | 42.34 | Adult | 2\*100 |
| B2 (brain, FFMR) | 39.57 | Adult | 2\*100 |
| B3 (brain, FFMR) | 33.92 | Adult | 2\*100 |
| F1 (brain, FFAS) | 37.18 | Adult | 2\*100 |
| F2 (brain, FFAS) | 34.88 | Adult | 2\*100 |
| N1 (brain, FFAS) | 38.16 | Adult | 2\*100 |
| N3 (brain, FFAS) | 44.48 | Adult | 2\*100 |
| J1 (brain, DFF) | 29.68 | Adult | 2\*100 |
| M1 (brain, DFF) | 26.945 | Adult | 2\*100 |
| J4B (brain, DFF) | 33.17 | Adult | 2\*100 |
| J3 (brain, SFF) | 30.91 | Adult | 2\*100 |
| M2 (brain, SFF) | 30.825 | Adult | 2\*100 |
| J5B (brain, SFF) | 38.49 | Adult | 2\*100 |

**ASD**, MT treatment experiment; **NSR\_G2**, the female fish whose gonad was filled with primary-growth stage oocytes; **NSR\_O3**, the female fish whose gonad were characterized by a majority of cortical-alveolus stage oocytes; **NSR\_O4**, the female fish whose gonad was dominated by the vitellogenic stage oocytes; **NSR\_EBi**, the fish whose gonad was at early bisexual stage; **NSR\_LBi**, the fish whose gonad was at late bisexual stage; **NSR\_T**, the male fish whose gonad was the functional testis with spermatozoa released into sperm sinuses; FFMR, female fish before male removal; FFAS, female fish in aggressive interaction; DFF, dominant female fish; SFF, subdominant female fish.

**Supplementary URLs**

Ensembl, http://www.ensembl.org/index.html;

NR, NCBI Non-redundant protein sequence, <https://blast.ncbi.nlm.nih.gov/>;

KEGG, <http://www.genome.jp/kegg/>;

Repbase, <http://www.girinst.org/repbase/index.html>;

RepeatMasker, <http://repeatmasker.org/>;

Fossil calibration times, http://www.fossilrecord.net/dateaclade/index.html.