## Supplementary Tables and Figures

**Supplementary Table S1.** Sequencing data generated for *D. mawsoni* genome assembly and annotation.

| | Platform | Total reads No. | Total read base |
|---|---:|---:|---:|
| Genome | Sequel | 7,541,877 | 75,510,543,787 |
| | Novaseq | 874,915,446 | 132,112,232,346 |
| Transcriptome | Sequel | 993,300 | 72,377,817,589 |

**Supplementary Table S2.** Statistics for genome assembly of *D. mawsoni*.

| | Falcon-unzip | 3d-DNA |
|---|---:|---:|
| Number of contigs (scaffolds) | 2,440 | 1,062 |
| Total size of contigs (scaffolds) | 924,749,474 | 926,329,754 |
| Longest contig (scaffold) | 24,490,688 | 46,819,977 |
| N50 contig (scaffold) length | 3,228,001 | 36,983,948 |
| L50 contig (scaffold) count | 65 | 12 |

**Supplementary Table S3.** Lengths of 24 chromosome scale scaffolds in *D. mawsoni* genome.

| **Scaffolds** | **Length (bp)** |
|---|---:|
| HiC_scaffold_5 | 46,819,977 |
| HiC_scaffold_25 | 42,898,023 |
| HiC_scaffold_10 | 42,290,140 |
| HiC_scaffold_38 | 41,797,624 |
| HiC_scaffold_22 | 41,781,106 |
| HiC_scaffold_35 | 41,196,903 |
| HiC_scaffold_26 | 40,017,477 |
| HiC_scaffold_31 | 39,955,282 |
| HiC_scaffold_1 | 39,888,709 |
| HiC_scaffold_6 | 39,811,704 |
| HiC_scaffold_21 | 38,643,262 |
| HiC_scaffold_30 | 36,984,448 |
| HiC_scaffold_4 | 36,385,023 |
| HiC_scaffold_8 | 35,318,215 |
| HiC_scaffold_41 | 35,000,612 |
| HiC_scaffold_11 | 33,398,360 |
| HiC_scaffold_3 | 32,861,413 |
| HiC_scaffold_34 | 30,779,419 |
| HiC_scaffold_32 | 29,174,432 |
| HiC_scaffold_28 | 29,051,562 |
| HiC_scaffold_13 | 26,820,266 |
| HiC_scaffold_33 | 26,316,149 |
| HiC_scaffold_24 | 24,955,770 |
| HiC_scaffold_23 | 13,864,730 |
| **Sum** | **846,010,606** |

**Supplementary Table S4.** Benchmarking Universal Single-Copy Orthologs (BUSCO) evaluated for completeness of *D. mawsoni* genome assembly.

| **actinopterygii_odb9** | **No** | **%** |
|---|---:|---:|
| Complete BUSCOs (C) | 4,197 | 91.5 |
| Complete and single-copy BUSCOs (S) | 3,962 | 86.4 |
| Complete and duplicated BUSCOs (D) | 235 | 5.1 |
| Fragmented BUSCOs (F) | 194 | 4.3 |
| Missing BUSCOs (M) | 193 | 4.2 |
| Total BUSCO groups searched | 4,584 | |

**Supplementary Table S5.** Statistics of annotated transposable elements of *D. mawsoni.*

| Class | | Count | Length occupied (bp) | Percentage of sequences |
|---|---|---:|---:|---:|
| DNA | | 62,227 | 17,248,270 | 1.86 |
| | Academ-1 | 2,496 | 641,403 | 0.07 |
| | CMC-Chapaev | 177 | 42,051 | 0.00 |
| | CMC-EnSpm | 37,730 | 15,509,198 | 1.67 |
| | Crypton-A | 5,427 | 892,138 | 0.10 |
| | Crypton-C | 56 | 5,995 | 0.00 |
| | Crypton-H | 396 | 741,879 | 0.08 |
| | Crypton-V | 2,079 | 706,938 | 0.08 |
| | Dada | 779 | 375,943 | 0.04 |
| | Ginger-1 | 4,223 | 752,390 | 0.08 |
| | IS3EU | 8,841 | 2,103,489 | 0.23 |
| | Kolobok-T2 | 63,460 | 14,974,973 | 1.62 |
| | MULE-MuDR | 932 | 297,762 | 0.03 |
| | Merlin | 2,218 | 785,005 | 0.08 |
| | Novosib | 6,734 | 692,647 | 0.07 |
| | P | 9,864 | 2,212,850 | 0.24 |
| | PIF | 163 | 72,029 | 0.01 |
| | PIF-Harbinger | 20,079 | 6,994,636 | 0.76 |
| | PIF-ISL2EU | 2,632 | 1,314,889 | 0.14 |
| | PiggyBac | 5,272 | 1,659,800 | 0.18 |
| | Sola-1 | 2,537 | 234,328 | 0.03 |
| | Sola-2 | 302 | 116,271 | 0.01 |
| | TcMar-Fot1 | 625 | 229,359 | 0.02 |
| | TcMar-ISRm11 | 1,361 | 433,465 | 0.05 |
| | TcMar-Tc1 | 13,921 | 3,162,543 | 0.34 |
| | TcMar-Tc2 | 633 | 152,091 | 0.02 |
| | Zator | 1,009 | 216,560 | 0.02 |
| | Zisupton | 8,959 | 3,053,901 | 0.33 |
| | hAT | 4,660 | 989,755 | 0.11 |
| | hAT-Ac | 114,317 | 31,354,354 | 3.38 |
| | hAT-Blackjack | 5,550 | 2,062,972 | 0.22 |
| | hAT-Charlie | 52,808 | 15,859,065 | 1.71 |
| | hAT-Pegasus | 3,123 | 581,016 | 0.06 |
| | hAT-Tag1 | 2,040 | 325,427 | 0.04 |
| | hAT-Tip100 | 47,528 | 11,133,775 | 1.20 |
| | hAT-hAT19 | 122 | 73,829 | 0.01 |
| | hAT-hAT5 | 4,189 | 1,323,182 | 0.14 |
| | hAT-hAT6 | 421 | 98,292 | 0.01 |

|  |  |  |  |  |
|---|---|---|---|---|
|  | hAT-hATm | 685 | 180,033 | 0.02 |
|  | hAT-hobo | 2,730 | 716,015 | 0.08 |
| LINE |  | 353 | 251,342 | 0.03 |
|  | CR1 | 375 | 86,049 | 0.01 |
|  | I | 8,065 | 3,323,406 | 0.36 |
|  | I-Jockey | 1,584 | 307,115 | 0.03 |
|  | L1 | 6,680 | 3,081,651 | 0.33 |
|  | L1-Tx1 | 2,884 | 1,249,498 | 0.13 |
|  | L2 | 74,102 | 25,380,324 | 2.74 |
|  | Penelope | 609 | 759,171 | 0.08 |
|  | Proto2 | 353 | 164,506 | 0.02 |
|  | R2-Hero | 706 | 312,401 | 0.03 |
|  | R2-NeSL | 136 | 121,874 | 0.01 |
|  | RTE-BovB | 4,019 | 3,595,177 | 0.39 |
|  | RTE-X | 1,399 | 565,362 | 0.06 |
|  | Rex-Babar | 50,677 | 10,388,128 | 1.12 |
| LTR |  | 108 | 89,952 | 0.01 |
|  | Caulimovirus | 136 | 50,747 | 0.01 |
|  | Copia | 2,788 | 1,354,829 | 0.15 |
|  | DIRS | 12,989 | 18,948,235 | 2.05 |
|  | ERV1 | 5,658 | 5,844,206 | 0.63 |
|  | ERVK | 1,710 | 184,858 | 0.02 |
|  | Gypsy | 21,970 | 9,951,831 | 1.07 |
|  | Ngaro | 1,264 | 659,928 | 0.07 |
|  | Pao | 2,942 | 1,873,823 | 0.20 |
| RC | Helitron | 7,371 | 4,182,969 | 0.45 |
| SINE |  | 7,211 | 1,182,027 | 0.13 |
|  | 5S | 564 | 121,010 | 0.01 |
|  | Alu | 61 | 56,296 | 0.01 |
|  | ID | 437 | 104,956 | 0.01 |
|  | MIR | 19,474 | 2,654,456 | 0.29 |
|  | tRNA | 696 | 116,016 | 0.01 |
|  | tRNA-Core | 1,036 | 127,774 | 0.01 |
|  | tRNA-L1 | 1,317 | 252,943 | 0.03 |
| Unknown |  | 452,844 | 98,352,995 | 10.62 |
| **Total transposable elements** |  | **1,195,823** | **336,016,373** | **36.27** |

**Supplementary Table S6.** Genome assemblies used in this study.

| **Common name** | **Species** | **Assembly** |
|---|---|---|
| Antarctic toothfish | *Dissostichus mawsoni* | This study |
| Antarctic icefish | *Chaenocephalus aceratus* | Kim et al., 2019* |
| Antarctic dragonfish | *Parachaenichthys charcoti* | Ahn et al., 2017** |
| Antarctic bullhead notothen | *Notothenia coriiceps* | GCA_000735185 |
| Stickleback | *Gasterosteus aculeatus* | BROAD S1 |
| Spiny chromis | *Acanthochromis polyacanthus* | ASM210954v1 |
| Midas cichlid | *Amphilophus citrinellus* | Midas_v5 |
| Orange clownfish | *Amphiprion percula* | Nemo_v1 |
| Climbing perch | *Anabas testudineus* | fAnaTes1.1 |
| Pachon cavefish | *Astyanax mexicanus* | Astyanax_mexicanus-1.0.2 |
| Atlantic herring | *Clupea harengus* | Ch_v2.0.2 |
| Zebrafish | *Danio rerio* | GRCz11 |
| Cod | *Gadus morhua* | gadMor1 |
| Coelacanth | *Latimeria chalumnae* | LatCha1 |
| Nile tilapia | *Oreochromis niloticus* | O_niloticus_UMD_NMBU |
| Japanese medaka HNI | *Oryzias latipes* | ASM223471v1 |
| Amazon molly | *Poecilia formosa* | Poecilia_formosa-5.1.2 |
| Yellowtail amberjack | *Seriola lalandi dorsalis* | Sedor1 |
| Fugu | *Takifugu rubripes* | fTakRub1.2 |
| Platyfish | *Xiphophorus maculatus* | X_maculatus-5.0-male |

* Kim et al., 2019, Nature Ecology & Evolution, 3, 469–478.

** Ahn et al., 2017, GigaScience, 6, gix060.

**Supplementary Table S7.** Statistics for genome annotation

| Annotation database | Annotated number | Percentage (%) |
|---|---|---|
| No. Genes | 32,914 | - |
| nr Annotation | 30,048 | 91.3 |
| GO Annotation | 24,920 | 75.7 |
| KEGG Annotation | 14,474 | 44.0 |
| KOG Annotation | 19,205 | 58.3 |
| Pfam Annotation | 20,937 | 63.6 |
| Swissprot Annotation | 24,263 | 73.7 |
| TrEMBL Annotation | 29,559 | 89.8 |
| Interpro Annotation | 29,994 | 91.1 |
|  | **Count** | **Length Sum (bp)** |
| Exon | 319,003 | 55,190,635 |
| CDS | 316,045 | 51,218,172 |

**Supplementary Table S8.** Genome families of *D. mawsoni* and other fish species**.**

| Species | No. genes | No. genes in orthogroups | No. f unassigned genes | Percentage of genes in orthogroups | Percentage of unassigned genes | No. orthogroups containing species | Percentage of orthogroups containing species | No. species-specific orthogroups | No. genes in species-specific orthogroups | Percentage of genes in species-specific orthogroups |
|---|---|---|---|---|---|---|---|---|---|---|
| *D. mawsoni* | 32,914 | 29,718 | 3,196 | 90.3 | 9.7 | 16,896 | 67.3 | 434 | 1,431 | 4.3 |
| *N. coriiceps* | 32,661 | 29,195 | 3,466 | 89.4 | 10.6 | 18,378 | 73.2 | 101 | 238 | 0.7 |
| *P. charcoti* | 32,713 | 31,541 | 1,172 | 96.4 | 3.6 | 18,544 | 73.8 | 74 | 357 | 1.1 |
| *C. aceratus* | 30,773 | 28,579 | 2,194 | 92.9 | 7.1 | 17,133 | 68.2 | 187 | 611 | 2.0 |
| *A. polyacanthus* | 24,027 | 23,624 | 403 | 98.3 | 1.7 | 15,781 | 62.8 | 65 | 156 | 0.6 |
| *A. citrinellus* | 23,696 | 23,353 | 343 | 98.6 | 1.4 | 15,327 | 61.0 | 18 | 44 | 0.2 |
| *A. percula* | 23,926 | 23,385 | 541 | 97.7 | 2.3 | 15,467 | 61.6 | 22 | 102 | 0.4 |
| *A. testudineus* | 24,643 | 24,274 | 369 | 98.5 | 1.5 | 15,407 | 61.3 | 41 | 154 | 0.6 |
| *A. mexicanus* | 26,698 | 25,399 | 1,299 | 95.1 | 4.9 | 15,579 | 62.0 | 135 | 643 | 2.4 |
| *C. harengus* | 24,095 | 23,356 | 739 | 96.9 | 3.1 | 14,453 | 57.5 | 113 | 489 | 2.0 |
| *D. rerio* | 30,313 | 29,438 | 875 | 97.1 | 2.9 | 15,227 | 60.6 | 211 | 1,663 | 5.5 |
| *G. morhua* | 20,095 | 19,504 | 591 | 97.1 | 2.9 | 13,918 | 55.4 | 17 | 70 | 0.3 |
| *G. aculeatus* | 20,787 | 20,223 | 564 | 97.3 | 2.7 | 14,161 | 56.4 | 16 | 46 | 0.2 |
| *L. chalumnae* | 19,569 | 18,915 | 654 | 96.7 | 3.3 | 13,319 | 53.0 | 130 | 980 | 5.0 |
| *O. niloticus* | 28,189 | 27,449 | 740 | 97.4 | 2.6 | 15,359 | 61.2 | 98 | 629 | 2.2 |
| *O. latipes* | 22,127 | 21,519 | 608 | 97.3 | 2.7 | 14,544 | 57.9 | 33 | 224 | 1.0 |
| *P. formosa* | 23,615 | 23,447 | 168 | 99.3 | 0.7 | 15,191 | 60.5 | 26 | 82 | 0.3 |
| *S. lalandi dorsalis* | 24,755 | 24,224 | 531 | 97.9 | 2.1 | 15,811 | 63.0 | 66 | 141 | 0.6 |
| *T. rubripes* | 21,411 | 20,804 | 607 | 97.2 | 2.8 | 14,158 | 56.4 | 50 | 204 | 1.0 |
| *X. maculatus* | 23,774 | 23,285 | 489 | 97.9 | 2.1 | 15,304 | 60.9 | 36 | 137 | 0.6 |

**Supplementary Table S9.** Gene Ontology of expanded gene families in *D. mawsoni* genome among 20 fish species.

| **Category** | **GO ID** | **GO Name** | **No. genes** | ***P*-Value** |
|---|---|---|---|---|
| Biological Process | GO:0048468 | cell development | 116 | 1.80E-15 |
| | GO:0030154 | cell differentiation | 153 | 3.30E-13 |
| | GO:0000904 | cell morphogenesis involved in differentiation | 49 | 9.70E-12 |
| | GO:0010951 | negative regulation of endopeptidase activity | 35 | 1.50E-11 |
| | GO:0006427 | histidyl-tRNA aminoacylation | 9 | 1.80E-11 |
| | GO:0052548 | regulation of endopeptidase activity | 36 | 3.60E-11 |
| | GO:0022008 | neurogenesis | 82 | 5.40E-11 |
| | GO:0010466 | negative regulation of peptidase activity | 35 | 1.80E-10 |
| | GO:0048667 | cell morphogenesis involved in neuron differentiation | 42 | 5.90E-10 |
| | GO:0030030 | cell projection organization | 68 | 7.80E-10 |
| | GO:0052547 | regulation of peptidase activity | 36 | 9.70E-10 |
| | GO:0120036 | plasma membrane bounded cell projection organization | 67 | 1.00E-09 |
| | GO:0048699 | generation of neurons | 73 | 1.30E-09 |
| | GO:0035803 | egg coat formation | 13 | 2.60E-09 |
| | GO:0030182 | neuron differentiation | 69 | 2.80E-09 |
| | GO:0048812 | neuron projection morphogenesis | 42 | 3.00E-09 |
| | GO:0048858 | cell projection morphogenesis | 42 | 3.40E-09 |
| | GO:0120039 | plasma membrane bounded cell projection morphogenesis | 42 | 3.40E-09 |
| | GO:0051346 | negative regulation of hydrolase activity | 37 | 4.70E-09 |
| | GO:0007409 | axonogenesis | 38 | 4.90E-09 |
| | GO:0032990 | cell part morphogenesis | 42 | 5.60E-09 |
| | GO:0007339 | binding of sperm to zona pellucida | 13 | 7.90E-09 |
| | GO:0031175 | neuron projection development | 47 | 9.00E-09 |
| | GO:0048666 | neuron development | 58 | 9.10E-09 |
| | GO:0061564 | axon development | 39 | 9.60E-09 |
| | GO:0009988 | cell-cell recognition | 13 | 1.30E-08 |
| | GO:0000902 | cell morphogenesis | 51 | 1.40E-08 |
| | GO:0051336 | regulation of hydrolase activity | 65 | 2.10E-08 |
| | GO:0032989 | cellular component morphogenesis | 47 | 2.60E-08 |
| | GO:0007399 | nervous system development | 106 | 3.50E-08 |
| | GO:0007292 | female gamete generation | 15 | 4.90E-08 |
| | GO:0016043 | cellular component organization | 215 | 8.90E-08 |
| | GO:0030162 | regulation of proteolysis | 36 | 1.20E-07 |
| | GO:0006418 | tRNA aminoacylation for protein | 17 | 1.40E-07 |

| | | | | |
|---|---|---|---|---|
| | | translation | | |
| | GO:0048731 | system development | 209 | 2.50E-07 |
| | GO:0043038 | amino acid activation | 17 | 3.00E-07 |
| | GO:0043039 | tRNA aminoacylation | 17 | 3.00E-07 |
| | GO:0048477 | oogenesis | 13 | 3.60E-07 |
| | GO:0007411 | axon guidance | 25 | 4.60E-07 |
| | GO:0097485 | neuron projection guidance | 25 | 5.40E-07 |
| | GO:0015914 | phospholipid transport | 17 | 7.80E-07 |
| | GO:0031290 | retinal ganglion cell axon guidance | 12 | 1.00E-06 |
| | GO:0090156 | cellular sphingolipid homeostasis | 5 | 1.10E-06 |
| | GO:0097742 | de novo centriole assembly | 5 | 1.10E-06 |
| | GO:0098535 | de novo centriole assembly involved in multi-ciliated epithelial cell differentiation | 5 | 1.10E-06 |
| | GO:0051639 | actin filament network formation | 5 | 1.10E-06 |
| | GO:0051248 | negative regulation of protein metabolic process | 40 | 1.30E-06 |
| | GO:0006820 | anion transport | 44 | 2.70E-06 |
| | GO:0006378 | mRNA polyadenylation | 9 | 2.70E-06 |
| | GO:0015748 | organophosphate ester transport | 17 | 2.80E-06 |
| | GO:0007338 | single fertilization | 13 | 5.50E-06 |
| | GO:0043086 | negative regulation of catalytic activity | 37 | 5.70E-06 |
| | GO:1903251 | multi-ciliated epithelial cell differentiation | 5 | 6.20E-06 |
| | GO:0006379 | mRNA cleavage | 8 | 7.00E-06 |
| | GO:0051960 | regulation of nervous system development | 27 | 8.60E-06 |
| | GO:0015698 | inorganic anion transport | 22 | 9.80E-06 |
| | GO:0002062 | chondrocyte differentiation | 9 | 9.90E-06 |
| Molecular Function | GO:0004821 | histidine-tRNA ligase activity | 9 | 1.80E-11 |
| | GO:0004866 | endopeptidase inhibitor activity | 35 | 2.30E-11 |
| | GO:0061135 | endopeptidase regulator activity | 35 | 4.10E-11 |
| | GO:0030414 | peptidase inhibitor activity | 35 | 1.50E-10 |
| | GO:0061134 | peptidase regulator activity | 35 | 8.70E-10 |
| | GO:0140326 | ATPase-coupled intramembrane lipid transporter activity | 12 | 5.50E-08 |
| | GO:0004857 | enzyme inhibitor activity | 35 | 2.10E-07 |
| | GO:0016875 | ligase activity, forming carbon-oxygen bonds | 17 | 1.20E-06 |
| | GO:0004812 | aminoacyl-tRNA ligase activity | 17 | 1.20E-06 |
| | GO:0004396 | hexokinase activity | 6 | 1.70E-06 |
| | GO:0004127 | cytidylate kinase activity | 5 | 6.20E-06 |
| | GO:0051427 | hormone receptor binding | 10 | 7.60E-06 |

|  |  |  |  |  |
|---|---|---|---|---|
|  | GO:0035257 | nuclear hormone receptor binding | 9 | 9.90E-06 |
| Cellular Component | GO:0002178 | palmitoyltransferase complex | 7 | 3.40E-08 |
|  | GO:0035339 | SPOTS complex | 5 | 1.10E-06 |
|  | GO:0031211 | endoplasmic reticulum palmitoyltransferase complex | 5 | 6.20E-06 |
|  | GO:0017059 | serine C-palmitoyltransferase complex | 5 | 6.20E-06 |
|  | GO:1905961 | protein-cysteine S-palmitoyltransferase complex | 5 | 6.20E-06 |

**Supplementary Table S10.** Gene Ontology of contracted gene families in *D. mawsoni* genome among 20 fish species.

| Category | GO ID | GO Name | No. genes | *P*-Value |
|---|---|---|---|---|
| Biological Process | GO:0032482 | Rab protein signal transduction | 5 | 3.40E-06 |
| | GO:1990778 | protein localization to cell periphery | 6 | 1.10E-07 |
| | GO:0006904 | vesicle docking involved in exocytosis | 4 | 2.00E-06 |
| | GO:0017157 | regulation of exocytosis | 5 | 1.10E-06 |
| | GO:0030011 | maintenance of cell polarity | 3 | 3.90E-06 |
| | GO:0048731 | system development | 25 | 9.50E-06 |
| | GO:1903361 | protein localization to basolateral plasma membrane | 3 | 2.80E-07 |
| | GO:0007399 | nervous system development | 16 | 2.60E-06 |
| | GO:0035090 | maintenance of apical/basal cell polarity | 3 | 2.40E-06 |
| | GO:0045199 | maintenance of epithelial cell apical/basal polarity | 3 | 2.40E-06 |
| Molecular Function | GO:0004652 | polynucleotide adenylyltransferase activity | 3 | 8.30E-06 |
| | GO:1990817 | RNA adenylyltransferase activity | 3 | 7.00E-07 |
| Cellular Component | GO:0098793 | presynapse | 7 | 2.90E-06 |
| | GO:0032593 | insulin-responsive compartment | 3 | 7.10E-08 |

**Supplementary Table S11.** Gene Ontology of specific gene families in *D. mawsoni* genome among four Antarctic fish.

| **Category** | **GO ID** | **GO Name** | **No. genes** | ***P*-Value** |
|---|---|---|---|---|
| Biological Process | GO:1903432 | regulation of TORC1 signaling | 7 | 3.50E-05 |
| | GO:0032008 | positive regulation of TOR signaling | 8 | 4.30E-05 |
| | GO:0038202 | TORC1 signaling | 7 | 5.20E-05 |
| | GO:0090114 | COPII-coated vesicle budding | 7 | 5.20E-05 |
| | GO:1904263 | positive regulation of TORC1 signaling | 7 | 5.40E-06 |
| | GO:0048194 | Golgi vesicle budding | 7 | 1.90E-04 |
| | GO:0006259 | DNA metabolic process | 107 | 5.20E-20 |
| | GO:0070278 | extracellular matrix constituent secretion | 7 | 3.00E-06 |
| | GO:0007029 | endoplasmic reticulum organization | 7 | 1.90E-04 |
| | GO:0071897 | DNA biosynthetic process | 53 | 1.60E-29 |
| | GO:0015074 | DNA integration | 24 | 1.20E-04 |
| | GO:0036372 | opsin transport | 7 | 7.70E-07 |
| | GO:0006278 | RNA-dependent DNA biosynthetic process | 53 | 1.10E-34 |
| | GO:0098529 | neuromuscular junction development, skeletal muscle fiber | 3 | 1.50E-04 |
| Molecular Function | GO:0038024 | cargo receptor activity | 19 | 2.80E-07 |
| | GO:0140097 | catalytic activity, acting on DNA | 62 | 3.50E-22 |
| | GO:0030246 | carbohydrate binding | 43 | 5.00E-09 |
| | GO:0005044 | scavenger receptor activity | 19 | 1.30E-07 |
| | GO:0008061 | chitin binding | 8 | 1.30E-04 |
| | GO:0034061 | DNA polymerase activity | 51 | 3.30E-31 |
| | GO:0003964 | RNA-directed DNA polymerase activity | 51 | 2.30E-35 |
| | GO:0016779 | nucleotidyltransferase activity | 58 | 2.90E-20 |
| Cellular Component | GO:0031080 | nuclear pore outer ring | 7 | 5.40E-06 |
| | GO:0030127 | COPII vesicle coat | 8 | 1.00E-04 |
| | GO:0070971 | endoplasmic reticulum exit site | 7 | 1.60E-06 |
| | GO:0012507 | ER to Golgi transport vesicle membrane | 9 | 4.20E-05 |
| | GO:0030134 | COPII-coated ER to Golgi transport vesicle | 9 | 1.70E-04 |

**Supplementary Table S12.** Characterization of transcripts based on splice junctions.

| **Category** | **No. Isoforms** | **No. Genes** |
|---|---:|---:|
| Full Spice Match (FSM) | 1,189 | 1,007 |
| Incomplete Splice Match (ISM) | 8,166 | 3,679 |
| Novel In Catalog (NIC) | 1,288 | 902 |
| Novel Not in Catalog (NNC) | 9,165 | 3,966 |
| Genic Genomic | 1,981 | 1,757 |
| Antisense | 1,911 | 257 |
| Fusion | 280 | 160 |
| Intergenic | 7,500 | 7,500 |

**Supplementary Table S13.** Summary of transcripts based on structural categories and subcategories.

| **Categories/Subcategories** | **anti-sense** | **FSM** | **fusion** | **genic** | **ISM** | **intergenic** | **NIC** | **NNC** | **Total** |
|---|---|---|---|---|---|---|---|---|---|
| **Annotated Transcripts** | | | | | | | | | |
| 3prime_fragment | | | | | 200 | | | | **200** |
| 5prime_fragment | | | | | 4,009 | | | | **4,009** |
| internal_fragment | | | | | 409 | | | | **409** |
| intron_retention | | | | | 257 | | | | **257** |
| mono-exon | | 39 | | | 3,291 | | | | **3,330** |
| multi-exon | | 1,150 | | | | | | | **1,150** |
| | | **1,189** | | | **8,166** | | | | **9,355** |
| **Novel Transcripts** | | | | | | | | | |
| at_least_one_novel_splicesite | | | | | | | | 7,579 | **7,579** |
| combination_of_known_junctions | | | | | | | 1 | | **1** |
| combination_of_known_splicesites | | | | | | | 87 | | **87** |
| intron_retention | | | 108 | | | | 538 | 1586 | **2,232** |
| mono-exon | 1,185 | | | 1918 | | 7032 | 479 | | **10,614** |
| mono-exon_by_intron_retention | | | | | | | 183 | | **183** |
| multi-exon | 726 | | 172 | 63 | | 468 | | | **1,429** |
| | **1,911** | | **280** | **1,981** | | **7,500** | **1,288** | **9,165** | **22,125** |
| **Total** | **1,911** | **1,189** | **280** | **1,981** | **8,166** | **7,500** | **1,288** | **9,165** | **31,480** |

**Supplementary Table S14.** Genes with more than 10 splicing variants.

| Cluster ID | No. Isoform | Description | Scaffold | Gene start | Gene End | Transcript length | No. exons |
|---|---|---|---|---|---|---|---|
| Cluster00001 | 94 | protein FAM133-like | HiC_scaffold_35 | 26365291 | 26385663 | 2731 | 26 |
| Cluster00002 | 80 | collagen alpha-2(I) chain isoform X1 | HiC_scaffold_3 | 13801711 | 13810315 | 3489 | 27 |
| Cluster00003 | 74 | collagen alpha-1(I) chain-like | HiC_scaffold_41 | 34480271 | 34501552 | 5600 | 50 |
| Cluster00004 | 68 | collagen alpha-1(I) chain-like | HiC_scaffold_28 | 21631468 | 21649356 | 5597 | 51 |
| Cluster00005 | 35 | keratin, type I cytoskeletal 13-like | HiC_scaffold_41 | 6557802 | 6661239 | 3348 | 21 |
| Cluster00006 | 33 | intermediate filament protein ON3-like isoform X1 | HiC_scaffold_10 | 7466753 | 7478205 | 3205 | 17 |
| Cluster00007 | 31 | nucleolar protein 56 | HiC_scaffold_25 | 16890203 | 16896544 | 2200 | 12 |
| Cluster00008 | 29 | nucleolin isoform X1 | HiC_scaffold_38 | 19807945 | 19816808 | 3292 | 18 |
| Cluster00009 | 27 | collagen alpha-1(XVII) chain-like isoform X1 | HiC_scaffold_8 | 15681495 | 15693908 | 3839 | 36 |
| Cluster00010 | 25 | desmoplakin-like isoform X2 | HiC_scaffold_38 | 29900296 | 29918571 | 5578 | 25 |
| Cluster00011 | 24 | leukocyte elastase inhibitor-like isoform X1 | HiC_scaffold_33 | 9263919 | 9289857 | 4143 | 32 |
| Cluster00012 | 24 | polyadenylate-binding protein 1 | HiC_scaffold_3 | 17579339 | 17585539 | 2813 | 14 |
| Cluster00013 | 22 | major histocompatibility complex class I-related gene protein-like | HiC_scaffold_3 | 2729636 | 3134308 | 3374 | 21 |
| Cluster00014 | 22 | gelsolin-like | HiC_scaffold_38 | 37524652 | 37537161 | 2532 | 16 |
| Cluster00015 | 22 | keratin, type I cytoskeletal 13-like | HiC_scaffold_31 | 2677155 | 2691323 | 1498 | 8 |
| Cluster00016 | 21 | myosin-9 isoform X1 | HiC_scaffold_41 | 28581753 | 28630836 | 5892 | 36 |
| Cluster00017 | 21 | actin, cytoplasmic 1 | HiC_scaffold_41 | 26070013 | 26074393 | 1855 | 7 |
| Cluster00018 | 20 | arachidonate 15-lipoxygenase B-like | HiC_scaffold_30 | 25625705 | 25657011 | 5087 | 27 |
| Cluster00019 | 19 | cadherin-like protein 26 | HiC_scaffold_10 | 24885335 | 24895171 | 3425 | 17 |
| Cluster00020 | 18 | von Willebrand factor A domain-containing protein 5A-like | HiC_scaffold_35 | 26670764 | 26683684 | 2959 | 18 |
| Cluster00021 | 17 | elongation factor 2-like | HiC_scaffold_38 | 13441256 | 13448740 | 3253 | 13 |
| Cluster00022 | 17 | calpastatin isoform X3 | HiC_scaffold_25 | 17073302 | 17087875 | 2895 | 34 |
| Cluster00023 | 16 | heat shock protein HSP 90-beta | HiC_scaffold_32 | 16767846 | 16787170 | 2595 | 13 |
| Cluster00024 | 16 | heat shock cognate 71 kDa protein | HiC_scaffold_11 | 31918180 | 31946563 | 3271 | 19 |
| Cluster00025 | 16 | serine/arginine repetitive matrix protein 1 isoform X1 | HiC_scaffold_4 | 12547785 | 12558681 | 3248 | 16 |
| Cluster00026 | 16 | gelsolin-like isoform X1 | HiC_scaffold_21 | 2434770 | 2511589 | 2916 | 25 |
| Cluster00027 | 15 | keratin, type I cytoskeletal 13-like | HiC_scaffold_58 | 3393820 | 3413886 | 2818 | 15 |
| Cluster00028 | 15 | 14-3-3 protein beta/alpha | HiC_scaffold_4 | 17676840 | 17683696 | 2101 | 6 |
| Cluster00029 | 15 | Tubulin beta chain | HiC_scaffold_4 | 8929775 | 8940692 | 2661 | 14 |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| Cluster00030 | 15 | eukaryotic translation initiation factor 4B isoform X2 | HiC_scaffold_25 | 17730695 | 17742324 | 2912 | 15 |
| Cluster00031 | 14 | poly(U)-binding-splicing factor PUF60 isoform X2 | HiC_scaffold_33 | 17506101 | 17514878 | 2114 | 12 |
| Cluster00032 | 14 | B-cadherin-like | HiC_scaffold_8 | 32179861 | 32193991 | 2508 | 17 |
| Cluster00033 | 14 | vigilin isoform X2 | HiC_scaffold_38 | 39094368 | 39124447 | 3780 | 28 |
| Cluster00034 | 14 | plasminogen activator inhibitor 1 RNA-binding protein-like isoform X1 | HiC_scaffold_38 | 15108054 | 15114110 | 2636 | 9 |
| Cluster00035 | 14 | integrin alpha-6-like isoform X1 | HiC_scaffold_21 | 23521889 | 23548589 | 3736 | 25 |
| Cluster00036 | 13 | T-complex protein 1 subunit beta | HiC_scaffold_13 | 24962254 | 24972478 | 1654 | 15 |
| Cluster00037 | 13 | T-complex protein 1 subunit alpha | HiC_scaffold_32 | 14758680 | 14764325 | 2345 | 12 |
| Cluster00038 | 13 | 14-3-3 protein beta/alpha-B-like | HiC_scaffold_3 | 17568819 | 17577116 | 2304 | 6 |
| Cluster00039 | 13 | SPARC | HiC_scaffold_22 | 37884453 | 37917685 | 1031 | 11 |
| Cluster00040 | 13 | casein kinase I isoform X4 | HiC_scaffold_22 | 5822369 | 5846780 | 1802 | 12 |
| Cluster00041 | 13 | T-complex protein 1 subunit zeta | HiC_scaffold_11 | 13756192 | 13763360 | 1839 | 14 |
| Cluster00042 | 13 | periostin-like isoform X1 | HiC_scaffold_35 | 13775685 | 13788301 | 3207 | 21 |
| Cluster00043 | 13 | nucleolar RNA helicase 2 | HiC_scaffold_5 | 17425608 | 17434827 | 3184 | 16 |
| Cluster00044 | 13 | probable ATP-dependent RNA helicase DDX5 isoform X1 | HiC_scaffold_41 | 30357857 | 30376133 | 2603 | 16 |
| Cluster00045 | 13 | 26S proteasome non-ATPase regulatory subunit 1 | HiC_scaffold_38 | 18843247 | 18879448 | 3660 | 27 |
| Cluster00046 | 13 | little elongation complex subunit 2 | HiC_scaffold_26 | 22168996 | 22181557 | 4049 | 24 |
| Cluster00047 | 13 | collagen alpha-2(VI) chain isoform X1 | HiC_scaffold_21 | 27611019 | 27626632 | 3223 | 32 |
| Cluster00048 | 12 | actin-related protein 2-like | HiC_scaffold_24 | 14963435 | 15009575 | 2856 | 18 |
| Cluster00049 | 12 | eukaryotic peptide chain release factor subunit 1 | HiC_scaffold_22 | 20296399 | 20304537 | 2143 | 11 |
| Cluster00050 | 12 | signal peptide peptidase-like 2A isoform X2 | HiC_scaffold_5 | 14080868 | 14099489 | 2101 | 15 |
| Cluster00051 | 12 | probable global transcription activator SNF2L2 | HiC_scaffold_31 | 29394350 | 29460810 | 4719 | 34 |
| Cluster00052 | 12 | ATP synthase subunit alpha, mitochondrial-like | HiC_scaffold_31 | 24443187 | 24449589 | 2035 | 11 |
| Cluster00053 | 12 | clustered mitochondria protein homolog | HiC_scaffold_31 | 14133807 | 14146588 | 4553 | 29 |
| Cluster00054 | 12 | coatomer subunit beta' | HiC_scaffold_21 | 23637773 | 23646440 | 3093 | 21 |
| Cluster00055 | 12 | deoxycytidylate deaminase | HiC_scaffold_10 | 21871921 | 21884100 | 1596 | 8 |
| Cluster00056 | 11 | elongation factor 1-gamma | HiC_scaffold_22 | 22132085 | 22139863 | 1524 | 10 |
| Cluster00057 | 11 | Eukaryotic translation initiation factor 3 subunit C | HiC_scaffold_28 | 14233086 | 14244750 | 3116 | 21 |
| Cluster00058 | 11 | eukaryotic translation initiation | HiC_scaffold_30 | 21370913 | 21378972 | 2743 | 19 |

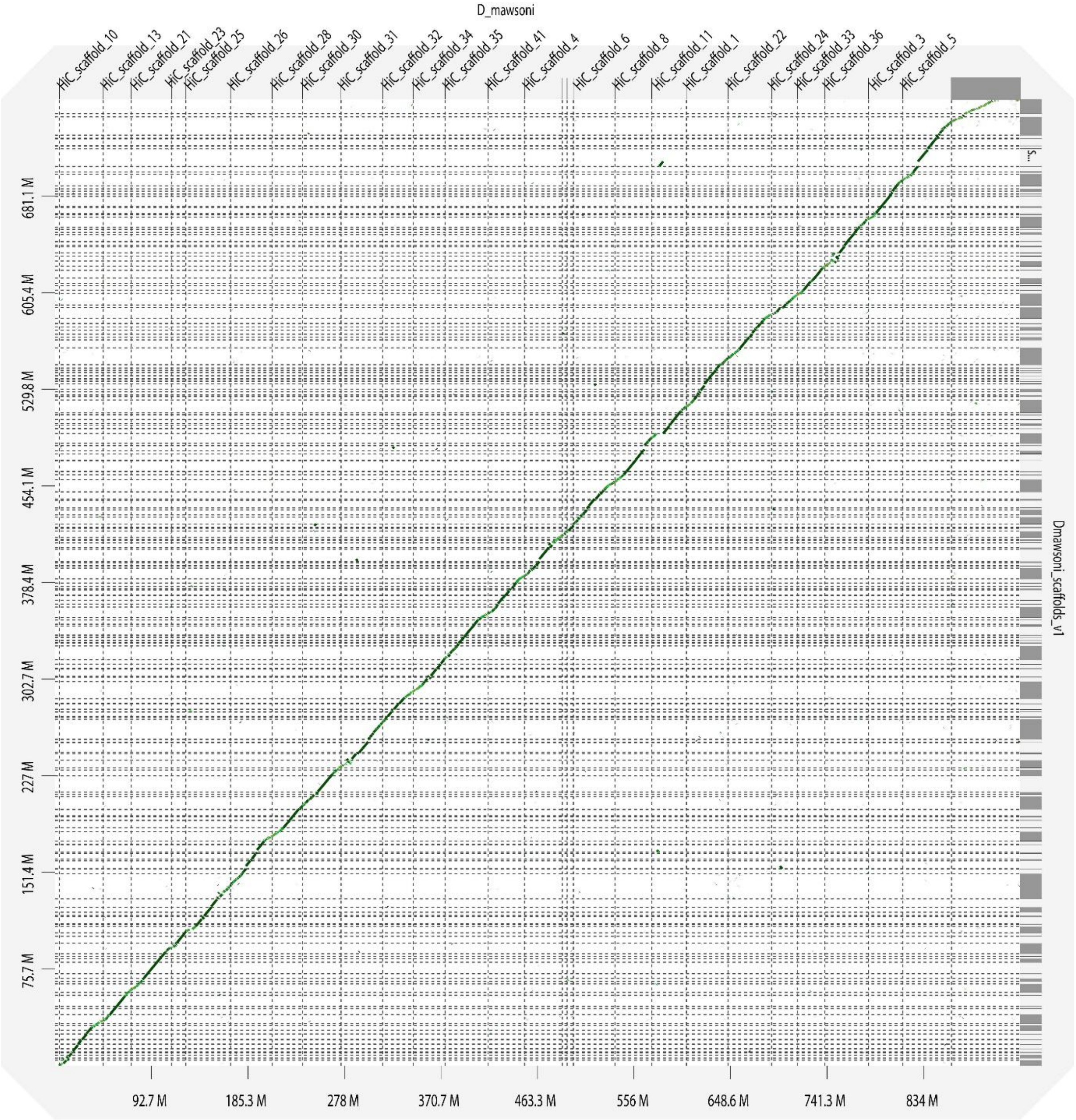
|  |  |  |  |  |  |  |  |
|---|---|---|---|---|---|---|---|
|  |  | factor 3 subunit B |  |  |  |  |  |
| Cluster00059 | 11 | sciellin isoform X1 | HiC_scaffold_6 | 24416009 | 24428016 | 1669 | 16 |
| Cluster00060 | 11 | collagen alpha-2(V) chain-like | HiC_scaffold_6 | 7838977 | 7850871 | 2400 | 24 |
| Cluster00061 | 11 | eukaryotic translation initiation factor 5 | HiC_scaffold_4 | 23203238 | 23213907 | 1687 | 10 |
| Cluster00062 | 11 | trichohyalin-like isoform X1 | HiC_scaffold_1 | 16522098 | 16529634 | 2877 | 15 |
| Cluster00063 | 11 | plakophilin-3 isoform X2 | HiC_scaffold_26 | 23672015 | 23689174 | 3769 | 14 |
| Cluster00064 | 11 | deoxyribonuclease-1-like | HiC_scaffold_10 | 21273262 | 21356630 | 2227 | 14 |
| Cluster00065 | 10 | ATP synthase subunit beta, mitochondrial | HiC_scaffold_23 | 9541936 | 9545604 | 1943 | 10 |
| Cluster00066 | 10 | nucleolar protein 58 | HiC_scaffold_23 | 8965292 | 8970790 | 2095 | 15 |
| Cluster00067 | 10 | T-complex protein 1 subunit eta | HiC_scaffold_24 | 6520313 | 6525351 | 2224 | 12 |
| Cluster00068 | 10 | protein max isoform X4 | HiC_scaffold_32 | 18428157 | 18440808 | 656 | 4 |
| Cluster00069 | 10 | H-2 class I histocompatibility antigen, Q10 alpha chain-like isoform X2 | HiC_scaffold_3 | 28771169 | 28780450 | 2418 | 9 |
| Cluster00070 | 10 | alcohol dehydrogenase class-3 | HiC_scaffold_22 | 29815705 | 29848625 | 2472 | 21 |
| Cluster00071 | 10 | heat shock 70 kDa protein 4-like | HiC_scaffold_22 | 18421659 | 18430960 | 2968 | 19 |
| Cluster00072 | 10 | histidine--tRNA ligase, cytoplasmic isoform X2 | HiC_scaffold_22 | 5176609 | 5190002 | 2000 | 12 |
| Cluster00073 | 10 | heterogeneous nuclear ribonucleoprotein A/B-like | HiC_scaffold_11 | 2874581 | 2886698 | 1937 | 9 |
| Cluster00074 | 10 | N-acetylgalactosaminyltransferase 7 isoform X1 | HiC_scaffold_30 | 1393678 | 1422330 | 3272 | 18 |
| Cluster00075 | 10 | 26S proteasome non-ATPase regulatory subunit 2 | HiC_scaffold_35 | 34459933 | 34477617 | 2898 | 21 |
| Cluster00076 | 10 | Actin-related protein 3 | HiC_scaffold_6 | 3244610 | 3277669 | 2716 | 15 |
| Cluster00077 | 10 | ras GTPase-activating-like protein IQGAP1 | HiC_scaffold_5 | 34372247 | 34419569 | 4674 | 37 |
| Cluster00078 | 10 | volume-regulated anion channel subunit LRRC8D-like | HiC_scaffold_41 | 28488992 | 28512796 | 3628 | 15 |
| Cluster00079 | 10 | calpain-9 | HiC_scaffold_30 | 30296192 | 30299552 | 537 | 4 |
| Cluster00080 | 10 | oxysterol-binding protein-related protein 7-like isoform X2 | HiC_scaffold_41 | 4378065 | 4404346 | 3508 | 22 |
| Cluster00081 | 10 | alpha-actinin-1 isoform X3 | HiC_scaffold_4 | 27863264 | 27909704 | 3708 | 26 |
| Cluster00082 | 10 | kinectin isoform X1 | HiC_scaffold_4 | 16399032 | 16417531 | 4583 | 33 |
| Cluster00083 | 10 | putative aminopeptidase W07G4.4 | HiC_scaffold_31 | 25246354 | 25252805 | 1633 | 16 |
| Cluster00084 | 10 | coronin-1C isoform X2 | HiC_scaffold_31 | 13463564 | 13471002 | 990 | 9 |
| Cluster00085 | 10 | E3 ubiquitin-protein ligase UBR5 isoform X3 | HiC_scaffold_1 | 31203150 | 31233999 | 8845 | 59 |
| Cluster00086 | 10 | desmocollin-2-like | HiC_scaffold_21 | 19071982 | 19078911 | 3294 | 17 |

|  |  |  |  |  |  |  |  |
|---|---|---|---|---|---|---|---|
| Cluster00087 | 10 | nucleolar GTP-binding protein 1 | HiC_scaffold_21 | 6882345 | 6896824 | 2519 | 17 |
| Cluster00088 | 10 | WD repeat-containing protein 1 | HiC_scaffold_25 | 17170593 | 17179958 | 1843 | 14 |
| Cluster00089 | 10 | mucin-5AC-like | HiC_scaffold_25 | 6481679 | 6585019 | 12786 | 87 |
| Cluster00090 | 10 | cellular nucleic acid-binding protein isoform X1 | HiC_scaffold_10 | 24689606 | 24693643 | 1564 | 5 |
| Cluster00091 | 10 | 14-3-3 protein beta/alpha-1 | HiC_scaffold_10 | 23663526 | 23667072 | 1749 | 7 |
| Cluster00092 | 10 | deoxynucleoside triphosphate triphosphohydrolase SAMHD1 | HiC_scaffold_10 | 21733242 | 21745122 | 2541 | 15 |

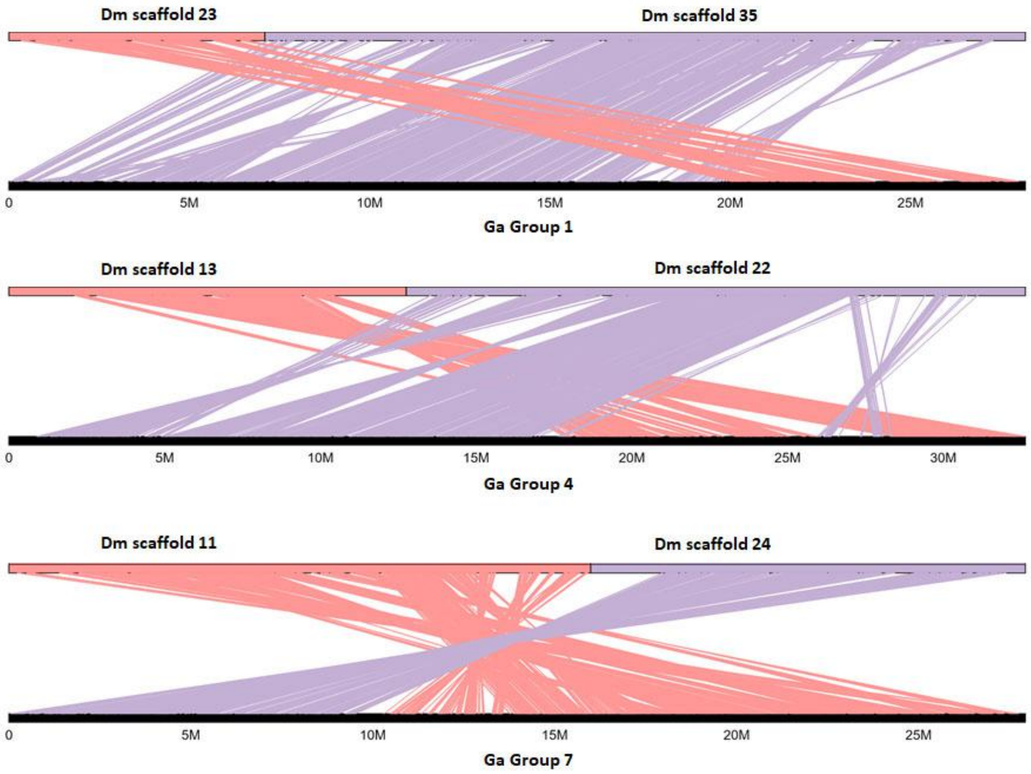
**Supplementary Table S15.** Gene Ontology of genes with more than 10 splicing variants.

| **GO Type** | **GO ID** | **GO Name** | ***P*-Value** |
|---|---|---|---|
| Biological Process | GO:0007010 | cytoskeleton organization | 2.0E-05 |
| | GO:0009894 | regulation of catabolic process | 2.4E-05 |
| | GO:0043933 | protein-containing complex subunit organization | 3.7E-05 |
| | GO:0022613 | ribonucleoprotein complex biogenesis | 4.4E-05 |
| | GO:0030036 | actin cytoskeleton organization | 7.1E-05 |
| | GO:0034622 | cellular protein-containing complex assembly | 1.2E-04 |
| | GO:0097435 | supramolecular fiber organization | 1.6E-04 |
| | GO:0032989 | cellular component morphogenesis | 2.4E-04 |
| | GO:0006412 | translation | 2.9E-04 |
| | GO:0043043 | peptide biosynthetic process | 3.3E-04 |
| | GO:0022607 | cellular component assembly | 4.4E-04 |
| | GO:0006518 | peptide metabolic process | 4.8E-04 |
| | GO:0043604 | amide biosynthetic process | 5.3E-04 |
| | GO:1901566 | organonitrogen compound biosynthetic process | 6.4E-04 |
| | GO:0048731 | system development | 7.2E-04 |
| | GO:0065003 | protein-containing complex assembly | 7.2E-04 |
| | GO:0061061 | muscle structure development | 9.2E-04 |
| | GO:0043603 | cellular amide metabolic process | 1.3E-03 |
| | GO:0006996 | organelle organization | 1.8E-03 |
| | GO:0009888 | tissue development | 2.0E-03 |
| | GO:0050790 | regulation of catalytic activity | 6.1E-03 |
| | GO:0044265 | cellular macromolecule catabolic process | 6.5E-03 |
| | GO:0048513 | animal organ development | 6.7E-03 |
| | GO:0009057 | macromolecule catabolic process | 8.8E-03 |
| | GO:0048468 | cell development | 9.1E-03 |
| | GO:0030154 | cell differentiation | 1.2E-02 |
| | GO:0006508 | proteolysis | 2.0E-02 |
| | GO:0009790 | embryo development | 2.1E-02 |
| | GO:0060429 | epithelium development | 2.3E-02 |
| | GO:0019538 | protein metabolic process | 3.3E-02 |
| | GO:1901564 | organonitrogen compound metabolic process | 3.5E-02 |
| | GO:0072359 | circulatory system development | 3.8E-02 |
| | GO:0051246 | regulation of protein metabolic process | 4.2E-02 |
| Molecular Function | GO:0003723 | RNA binding | 1.2E-11 |
| | GO:0008135 | translation factor activity, RNA binding | 4.3E-06 |
| | GO:0035639 | purine ribonucleoside triphosphate binding | 4.1E-04 |
| | GO:0032555 | purine ribonucleotide binding | 4.8E-04 |
| | GO:0097367 | carbohydrate derivative binding | 8.1E-04 |
| | GO:0000166 | nucleotide binding | 1.0E-03 |
| | GO:1901265 | nucleoside phosphate binding | 1.0E-03 |

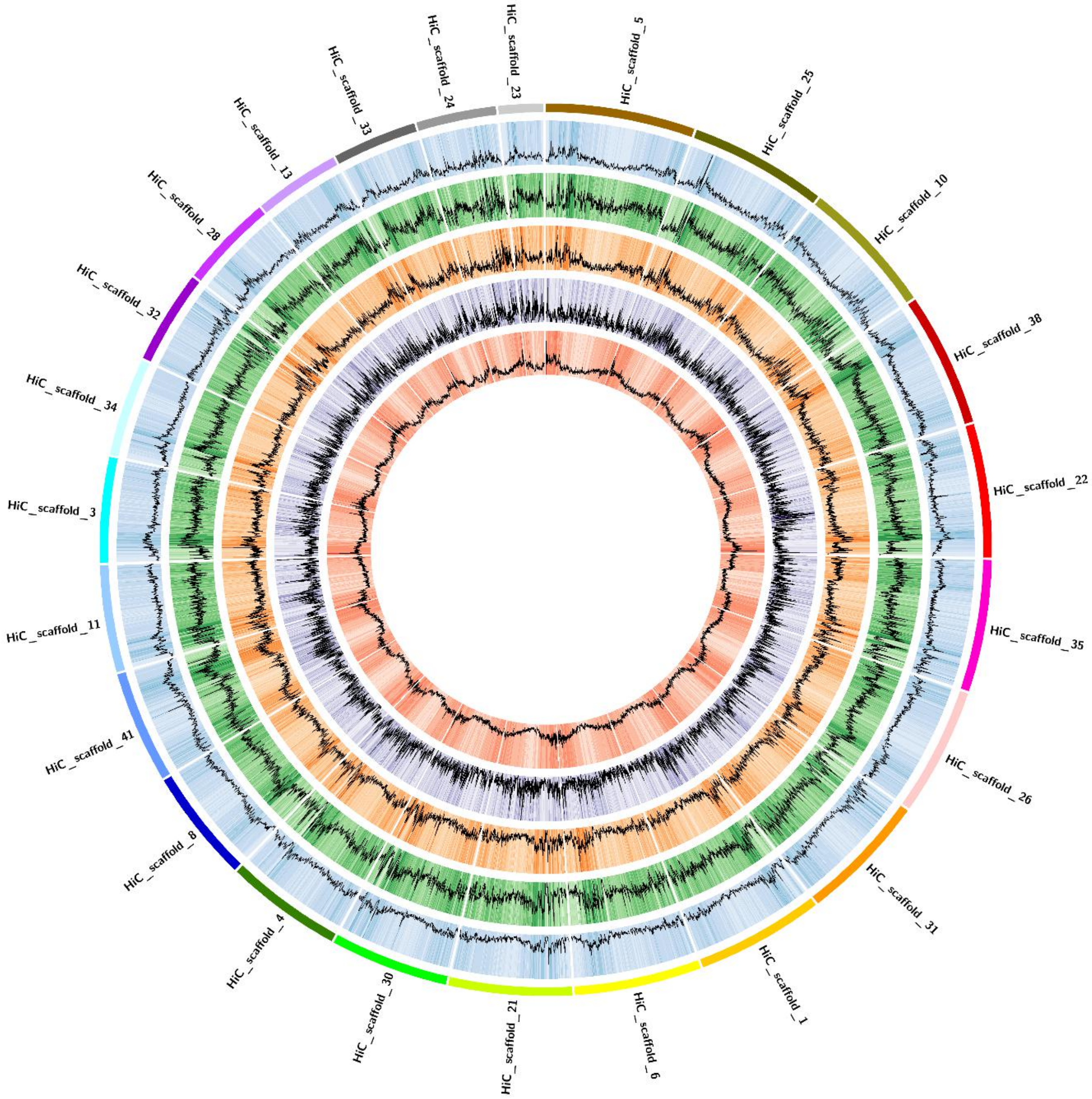
| | | | |
|---|---|---|---|
| | GO:0043168 | anion binding | 1.1E-03 |
| | GO:0005524 | ATP binding | 1.4E-03 |
| | GO:0032559 | adenyl ribonucleotide binding | 1.6E-03 |
| | GO:0030554 | adenyl nucleotide binding | 1.6E-03 |
| | GO:0036094 | small molecule binding | 2.3E-03 |
| | GO:0003676 | nucleic acid binding | 4.6E-03 |
| | GO:0005509 | calcium ion binding | 4.7E-03 |
| | GO:0016887 | ATPase activity | 4.8E-03 |
| | GO:0008092 | cytoskeletal protein binding | 6.0E-03 |
| | GO:0004175 | endopeptidase activity | 9.4E-03 |
| | GO:0017111 | nucleoside-triphosphatase activity | 9.5E-03 |
| | GO:0003779 | actin binding | 9.8E-03 |
| | GO:0016462 | pyrophosphatase activity | 1.1E-02 |
| | GO:0016818 | hydrolase activity, acting on acid anhydrides, in phosphorus-containing anhydrides | 1.1E-02 |
| | GO:0016817 | hydrolase activity, acting on acid anhydrides | 1.2E-02 |
| | GO:0016787 | hydrolase activity | 2.7E-02 |
| | GO:0070011 | peptidase activity, acting on L-amino acid peptides | 3.0E-02 |
| Cellular component | GO:0099081 | supramolecular polymer | 1.2E-08 |
| | GO:0099512 | supramolecular fiber | 1.2E-08 |
| | GO:0099513 | polymeric cytoskeletal fiber | 2.9E-06 |
| | GO:0005882 | intermediate filament | 9.8E-06 |
| | GO:0045111 | intermediate filament cytoskeleton | 1.1E-05 |
| | GO:0005856 | cytoskeleton | 1.3E-05 |
| | GO:0043232 | intracellular non-membrane-bounded organelle | 1.4E-05 |
| | GO:0043228 | non-membrane-bounded organelle | 1.5E-05 |
| | GO:0031981 | nuclear lumen | 7.2E-04 |
| | GO:0070013 | intracellular organelle lumen | 1.7E-03 |
| | GO:0043233 | organelle lumen | 1.7E-03 |
| | GO:0005634 | nucleus | 2.0E-03 |
| | GO:0043231 | intracellular membrane-bounded organelle | 4.1E-03 |
| | GO:0043227 | membrane-bounded organelle | 4.5E-03 |



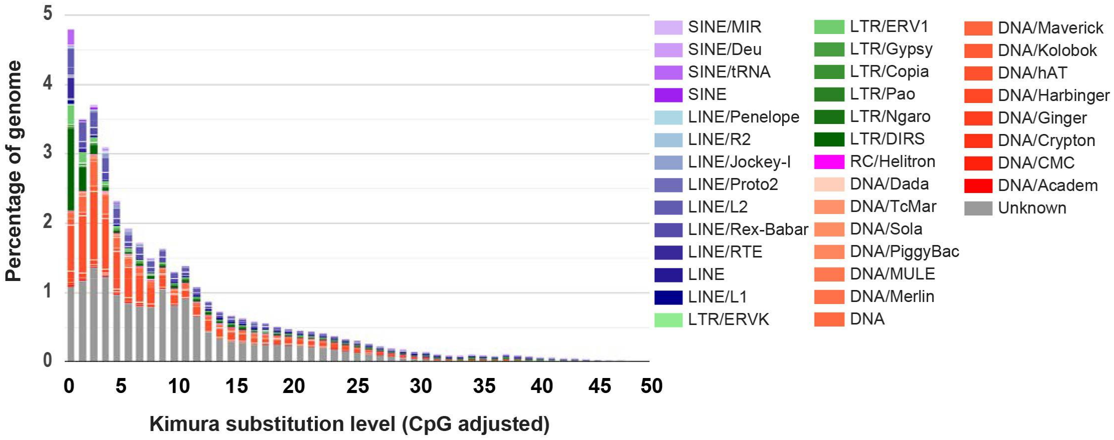
**Supplementary Figure S1.** Whole-genome alignment between *D. mawsoni* chromosome assembly and short-read assembly. X-axis is newly assembled in this study and Y-axis is Dmawsoni_scaffolds_v1 from GigaScience. Analysis at genomic scale using D-GENIES (http://dgenies.toulouse.inra.fr/).



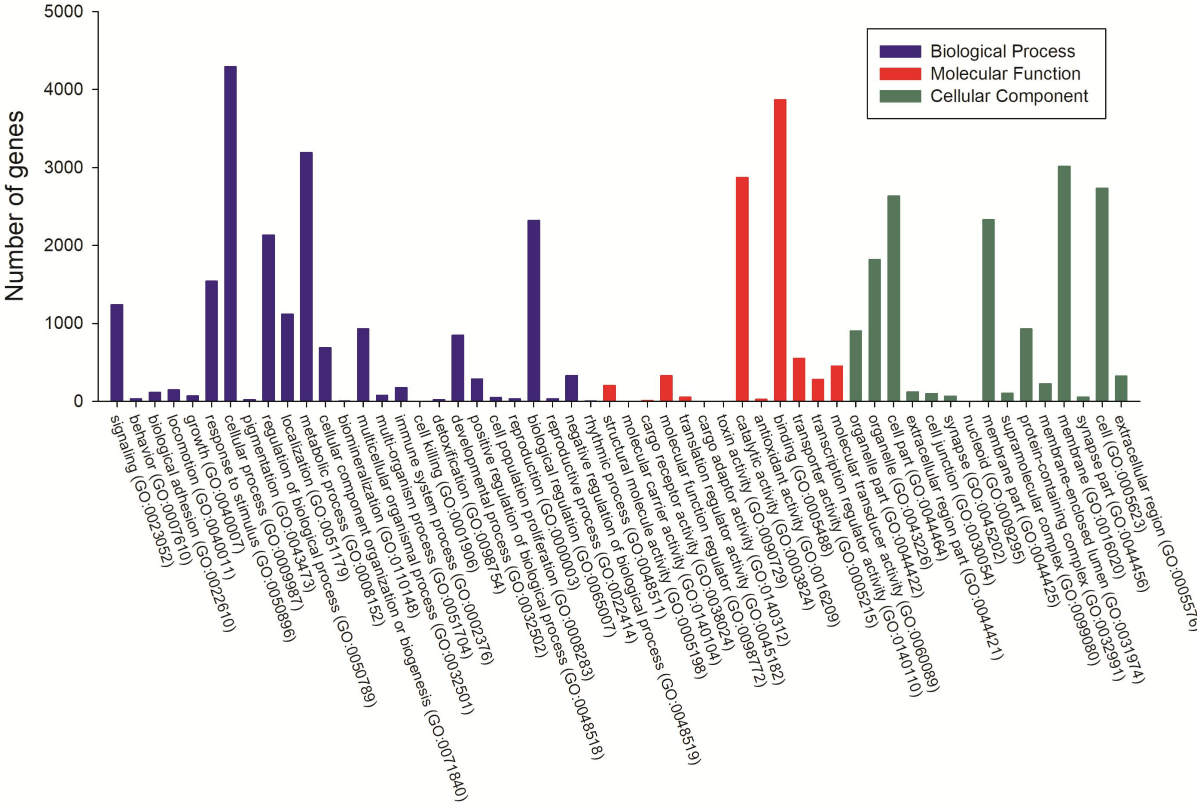
**Supplementary Figure S2.** *Gasterosteus aculeatus* chromosomes Group 1, Group 4, and Group 7 showing mapped each scaffolds of *D. mawsoni*. Scaffold23 and scaffold35 of *D. mawsoni* aligned to Group 1 of *G. aculeatus* chromosome, scaffold13 and scaffold22 to Group 4, and scaffold11 and scaffold24 to Group 7, respectively. Syntenic gene pairs are connected by lines.



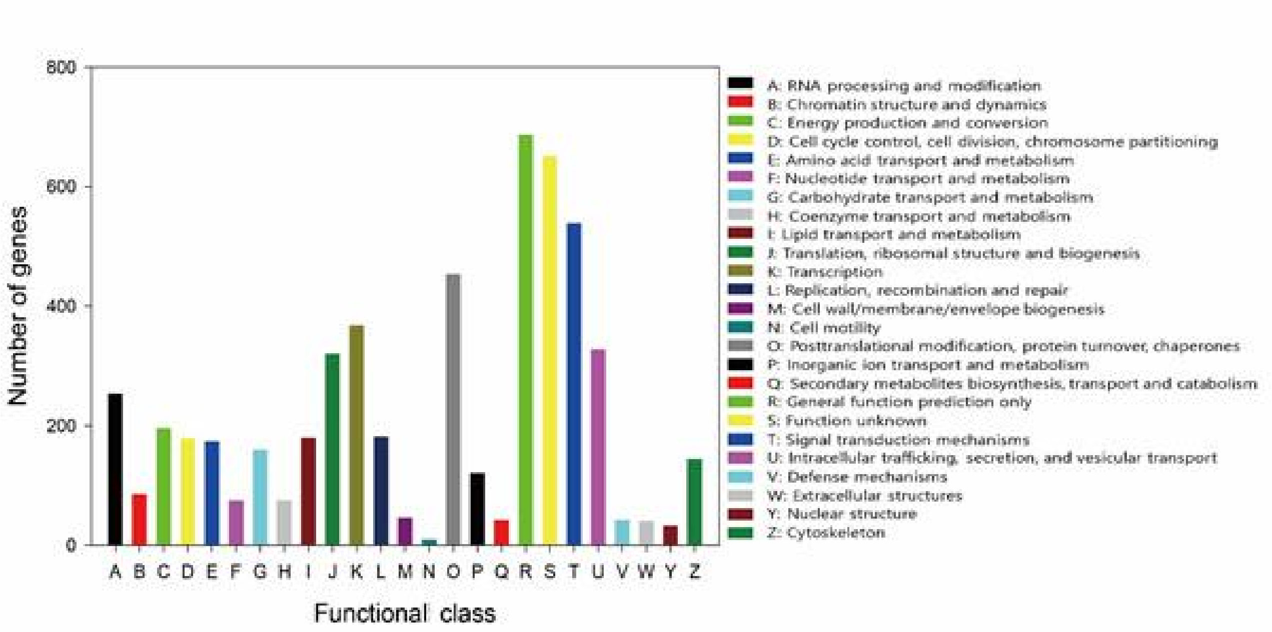
**Supplementary Figure S3.** Overview of *D. mawsoni* genome and mutations. Features are arranged as SNP (single nucleotide polymorphism), insertion, deletion, gene density, and repeat density from outside to inside in 100 Kb intervals across 24 chromosomes. Density of each feature is plotted as a heatmap with different colors and a line plot.



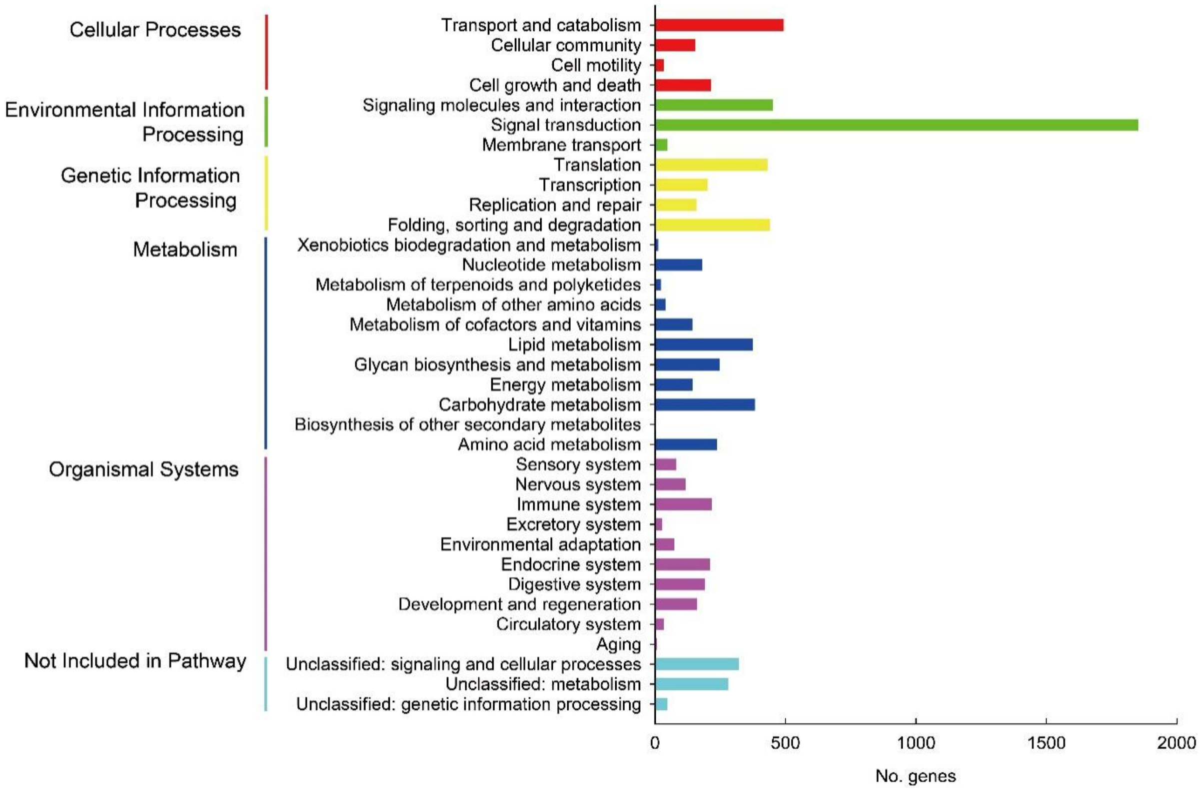
**Supplementary Figure S4.** Kimura distance-based copy divergence analysis of transposable elements in *D. mawsoni* genome.



**Supplementary Figure S5.** Gene Ontology (GO) annotation of predicted genes in *D. mawsoni* genome. Horizontal axis indicates second-level GO-annotation class and vertical axis indicates number of genes in each class.



**Supplementary Figure S6.** Eukaryotic Orthologous Groups (KOG) classification of predicted genes. Results are summarized into 24 functional classes. Horizontal axis indicates each class and vertical axis indicates number of genes in each class.



**Supplementary Figure S7.** Kyoto Encyclopedia of Genes and Genomes (KEGG) classifications.