

Supplementary Figures

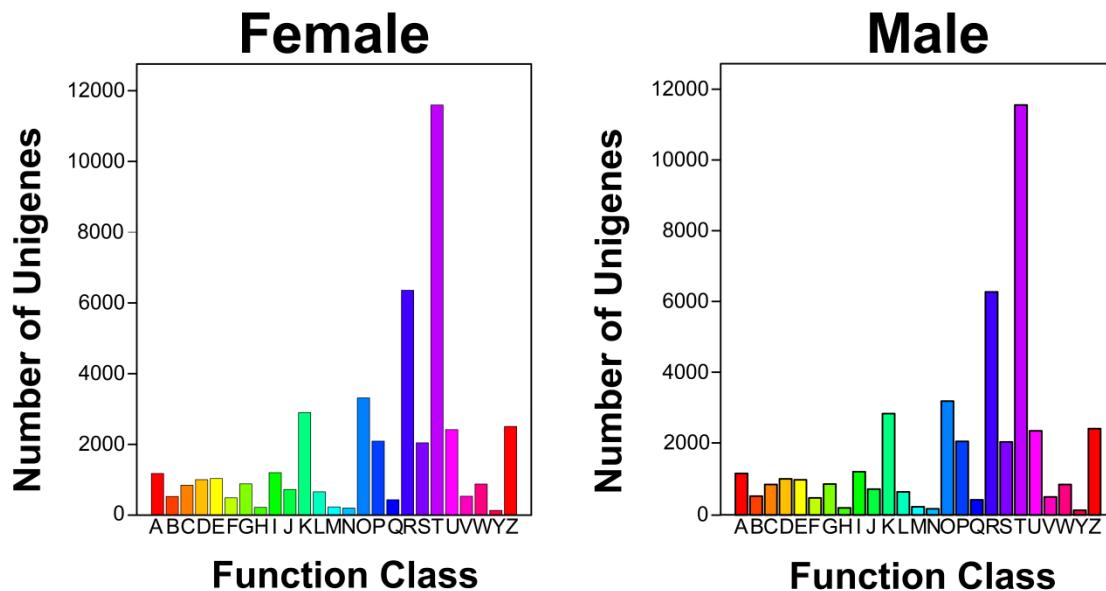


Figure. S1. KOG function classification of predicted genes in *S. argus*. The different color code (A–Z) represented different category as listed at right of the histogram. A: RNA processing and modification. B: Chromatin structure and dynamics. C: Energy production and conversion. D: Cell cycle control, cell division, chromosome partitioning. E: Amino acid transport and metabolism. F: Nucleotide transport and metabolism. G: Carbohydrate transport and metabolism. H: Coenzyme transport and metabolism. I: Lipid transport and metabolism. J: Translation, ribosomal structure and biogenesis. K: Transcription. L: Replication, recombination and repair. M: Cell wall/membrane/envelope biogenesis. N: Cell motility. O: Posttranslational modification, protein turnover, chaperones. P: Inorganic ion transport and metabolism. Q: Secondary metabolites biosynthesis, transport and catabolism. R: General function prediction only. S: Function unknown. T: Signal transduction mechanisms. U: Intracellular trafficking, secretion, and vesicular transport. V: Defense mechanisms. W: Extracellular structures. Y: Nuclear structure. Z: Cytoskeleton.

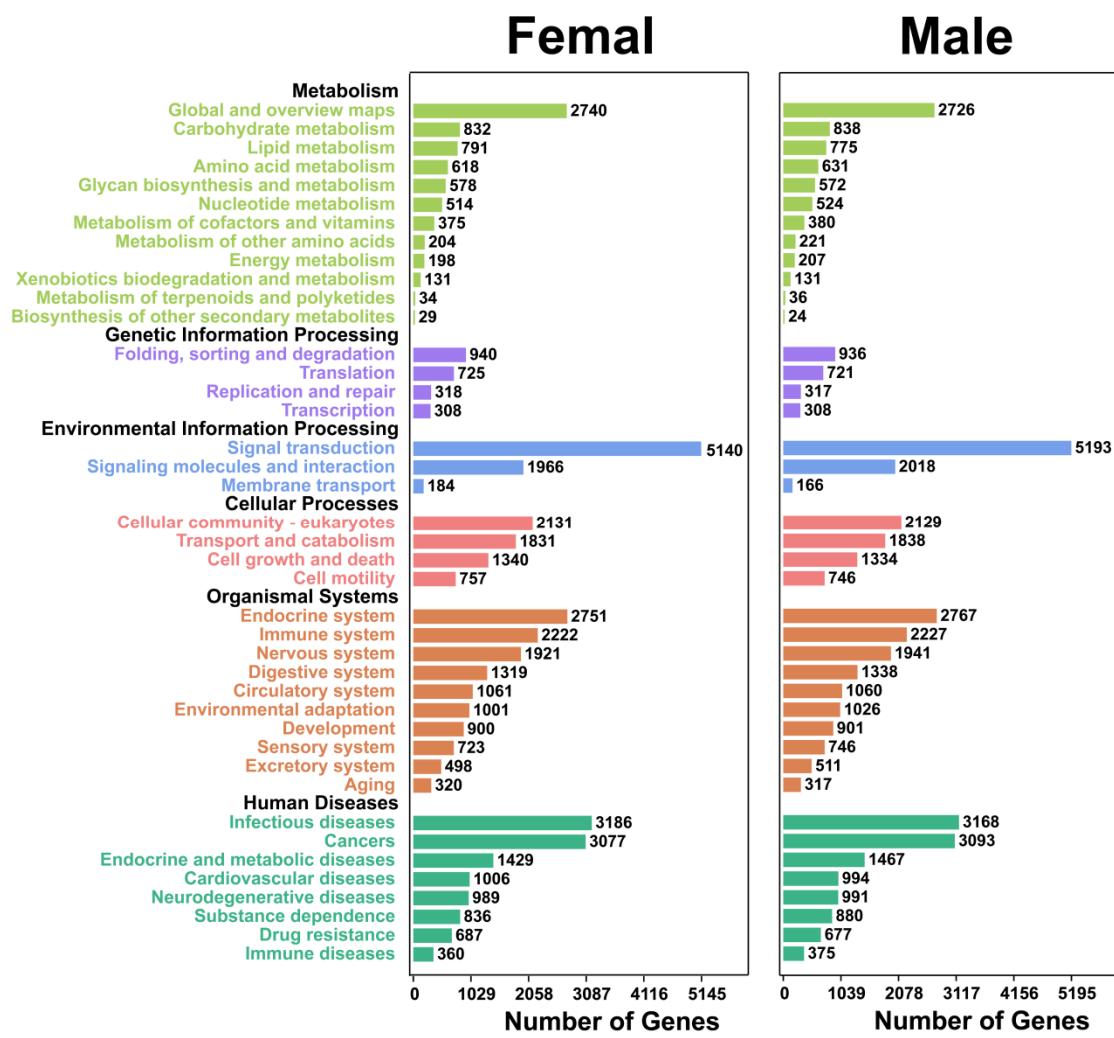


Figure. S2. KEGG pathway annotation of predicted genes in *S. argus*. The different color code represented different category.

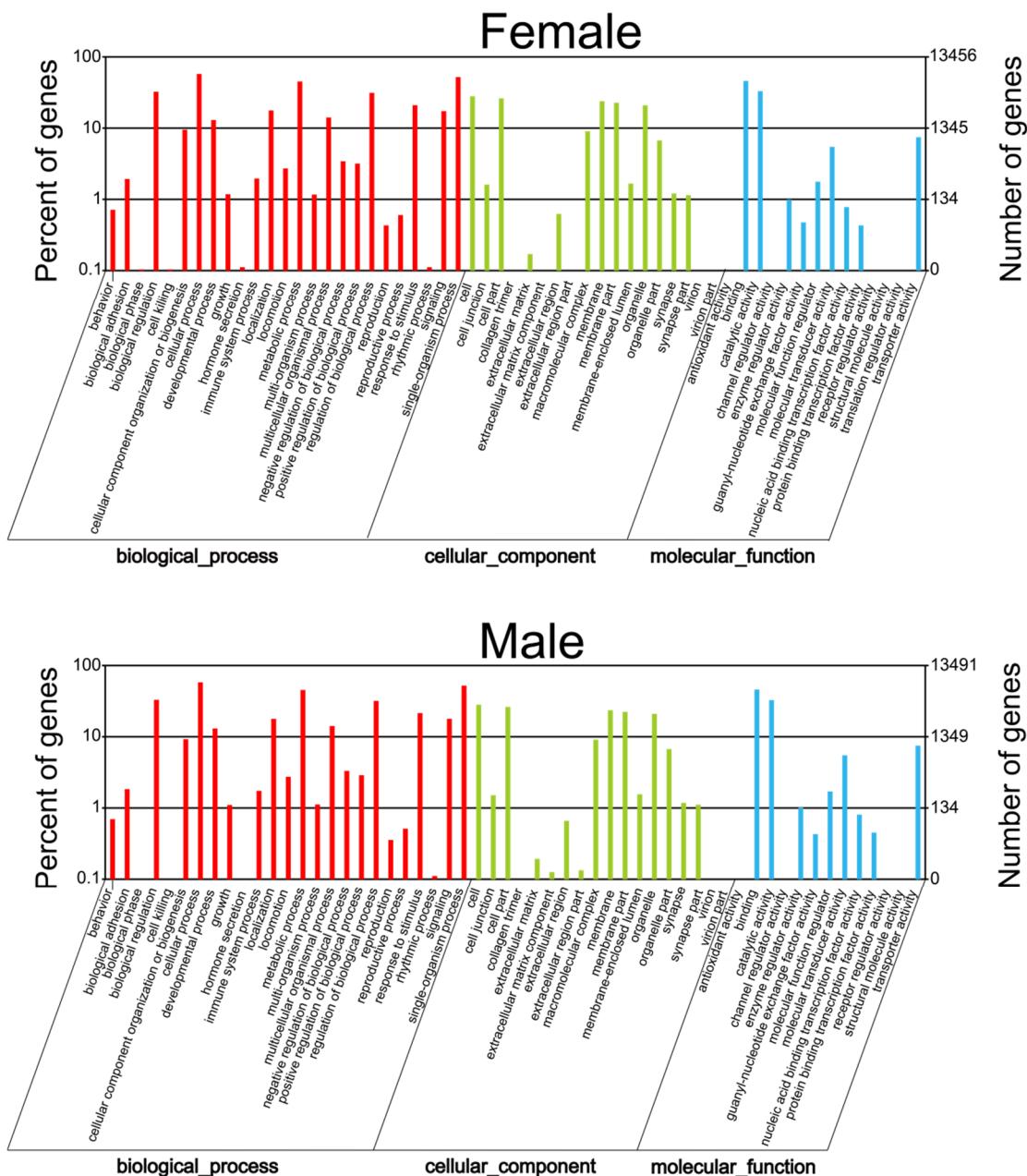


Figure. S3. Gene Ontology classification of predicted genes in *S. argus*. Genes were assigned to three categories: biological process, cellular component and molecular function.

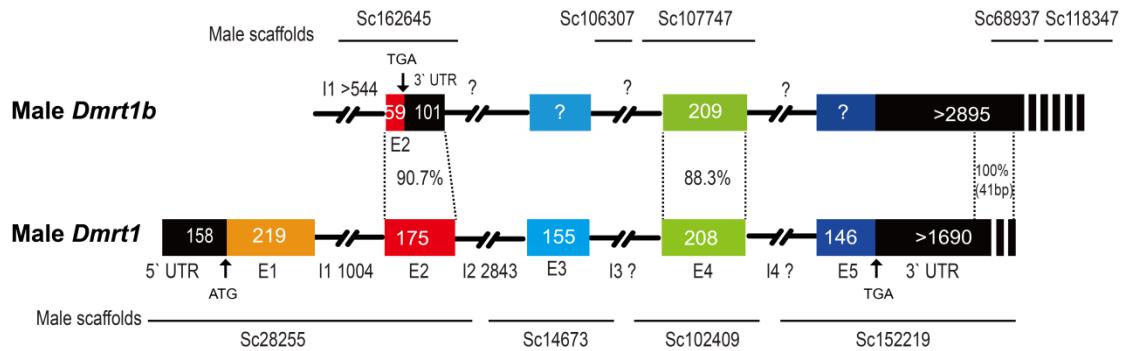


Figure. S4. The *Dmrt1b* gene from male genome. Numbers indicate base pairs (loci) of exon and intron sequences. Percentages indicate the similarity of *Dmrt1* and *Dmrt1b*. Arrows indicate the start and stop codons. Different colored rectangles represent different exons.

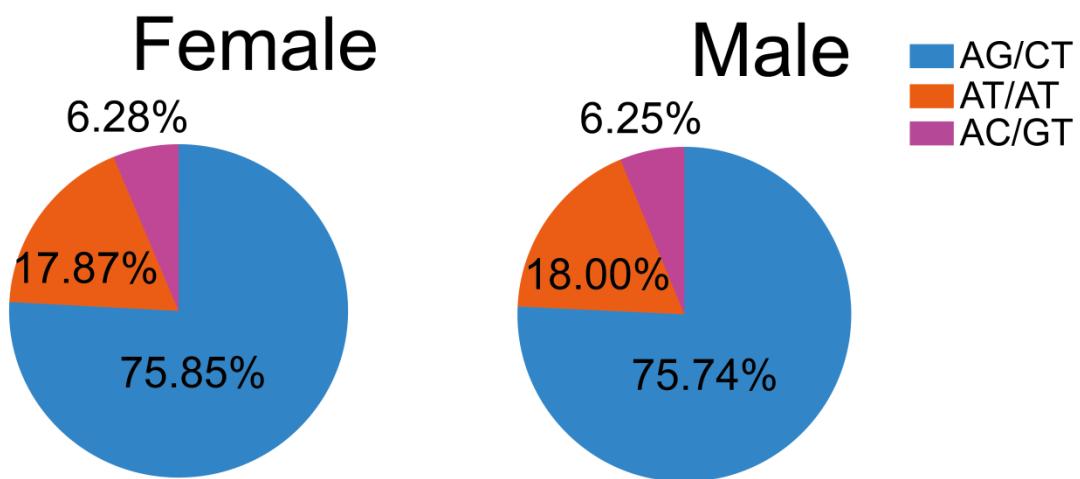


Figure. S5. Percentage of different motifs in di-nucleotide repeats of female and male *S. argus* genome.

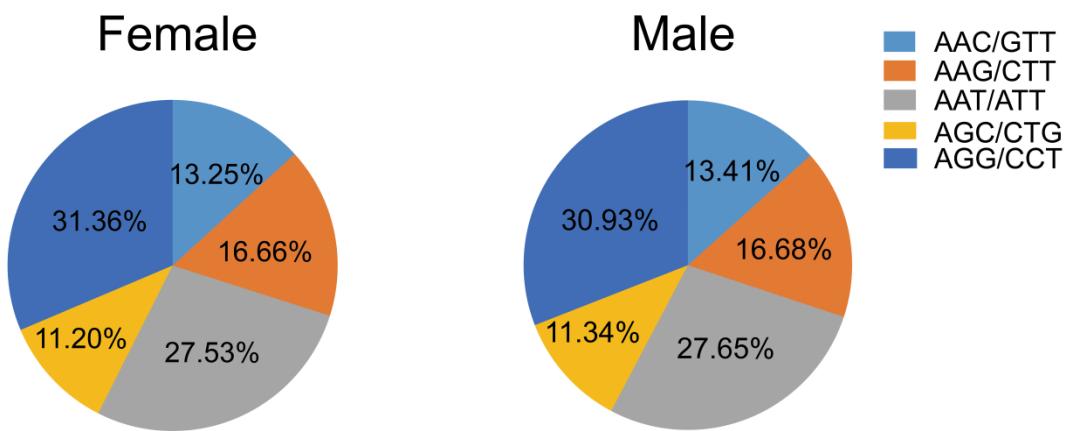


Figure. S6. Percentage of different motifs in trinucleotide repeats of female and male *S. argus* genome.

Supplementary Tables

Table S1. Statistics of top 4 similar species blasted against the NCBI nonredundant nucleotide database.

Species	Percentage (female/male)
<i>Dicentrarchus labrax</i>	0.91 / 0.91
<i>Scatophagus argus</i>	0.49/0.19
<i>Haplochromis burtoni</i>	0.36 / 0.37
<i>Oreochromis niloticus</i>	0.26 / 0.30

Note: Percentage of the similarity more than 90% reads in 5000 reads of female and male *S. argus*.

Table S2. Statistics of KOG function classification of predicted genes in *S. argus*

KOG code	KOG code description	Gene number (female/male)
A	RNA processing and modification	1178/1193
B	Chromatin structure and dynamics	528/543
C	Energy production and conversion	846/877
D	Cell cycle control, cell division, chromosome partitioning	1003/1037
E	Amino acid transport and metabolism	1035/1011
F	Nucleotide transport and metabolism	489/493
G	Carbohydrate transport and metabolism	884/889
H	Coenzyme transport and metabolism	219/208
I	Lipid transport and metabolism	1201/1242
J	Translation, ribosomal structure and biogenesis	725/744
K	Transcription	2901/2904
L	Replication, recombination and repair	659/665
M	Cell wall/membrane/envelope biogenesis	232/239
N	Cell motility	201/181
O	Posttranslational modification, protein turnover, chaperones	3315/3262
P	Inorganic ion transport and metabolism	2092/2110
Q	Secondary metabolites biosynthesis, transport and catabolism	436/438
R	General function prediction only	6357/6386
S	Function unknown	2045/2094
T	Signal transduction mechanisms	11595/11733
U	Intracellular trafficking, secretion, and vesicular transport	2418/2412
V	Defense mechanisms	535/521
W	Extracellular structures	880/877
X	Mobilome: prophages, transposons	0/0
Y	Nuclear structure	134/140
Z	Cytoskeleton	2507/2474

Table S3. Assessment of *S. argus* genome assembly and completeness using BUSCO
(Benchmarking Universal Single-Copy Orthologs).

	Genome BUSCO (female / male)	Percentage (female / male)
Complete BUSCOs	3,055/3,108	66.6%/67.8%
Complete Single-Copy BUSCOs	2,990/3,042	65.2%/66.4%
Complete Duplicated BUSCOs	65/66	1.4%/1.4%
Fragmented BUSCOs	881/838	19.2%/18.3%
Missing BUSCOs	648/638	14.2%/13.9%

Supplementary Sequences

>Male scaffold28255 (Exon1 and 2 of *Dmrt1*)

>Male scaffold14673 (Exon3 of *Dmrt1*)

nnnaagagggccaagacc
gcatatgccagaggatccaggacaattacatccaacgtccatggagcatgtggagggcataaagtgc

>Male scaffold102409 (Exon4 of *Dmrt1*)

>Male scaffold152219 (Exon5 of *Dmrt1*)

gaggtttattttcaagtgcataaacatacttaccagagcacgacaacaatttaatgtttatgtgtttataatttttaaccataataatgg
ttataccaaagcctttatttgacatacttattttatgtcaatgtgaaatttgataaatattaaaaaaaacaattttaaaattgtcatgtcaca
gttatcctaccaaataaaagcatttacacacagtaaactgtaaaattacacaggtgcctgtgttagtgaagaggcttgcactgaaaa
gaggaaaaaaaacttagtggcctgctactatcacacctccgcatactgtccaataattagtgaaagtgcataactgttgcgtgtactata
gtagccagcttgacaactttatgaagtgcctccaaatggactttgactggcaaaaacacagaattggcagttcatgtttaaaattcagtgttc
ttcaacactcttcagtgacacaggatggcagcgagttagctgactttggfggctgttctctgacaacttaagatgtacgaagatc
tttgatccagttcaagacaaagttaaaacaagcaagatctaagaatcttagacagtagtggtggctgaaaattgtgtctggccatcatgccagc
attggacacttagggaaaacgcgtctttcattgtctactttgaaaagttaaaataaaccggattctgtgtctcattctgcagCCTTCT
CCCCCAGTCACATCTCCATTGTTACGACTCCGCCTGCAGGACCCGCTTAACCTGTGAG
GTCAATGTTGAATGTGAGGCCAGCGATGAGATGCCAAACTCACTGTCAGCTCCATCAT
CGATGGTGACGCTGCCAAAT**TGA**AAAAAAGCTAAGGATAAAATTATAACACAAAGAGCAG

>Predicated amino acid sequence of *Dmrt1*

MNKDTQRKQVPESTGTLSPSKHQKPSRMPKCSRNCRNHGYISPLKGHKHFCNWRDCQCPCPKLIAERQKVMAAQVALRRQQDQEELGICSPVILPGPEVMVKDEAGADCMMLSAEGRSPTPTSTSNSFLAATGCSMSSSLSTGARAHTEEPSDLLLENSYYNFHRPLCYPTYYGNLYNYQQYQQMPHDDGLSSHNMSSQYYMHSYYPAATYLTQSLVSTPCVPGVFNLEDNNNSNKNNNNNNNDNNCSEAMAASFSPSHISIVHDSACRTPLNSEVNVECEASDEMPNFTVSSIIDGDAAK.

>Female scaffold83349 (Exon1 and 2 of *Dmrt1b*)

atataatatactgtcaataaagctgattcttctttactccactttattttgtctaactttgtcactttgcagaatgggattaataatgccaaatca
atcaacatcaaattattttatgttttagttaatggtaatgtttattatattatcataatataatgtatccatcccacagggtgcac
gtgaaagtataatgcagaaaattgtatacagtgatgtattttctcgactaagtagttaaatatttgctgctatgttattttgcagtt
ggccaaattaaaaatcgaaagttgaatttgaaactgttagagctgacttgagtcgtgaactatacagcAGGGACAAATC
CGACATAGTATCTTAAGTGACTGCTATTCCAGTCGCC**ATG**AACAAGGACACGCAGAG
CAAGCAGGTGCCGGAGTCCACCGGAACTCTCTCCCCATCCACGGGCCAGAACACCG
CAGGATCTTCAAGTGCTCCCTCTGTAGGAACCACGGCTGTATAATTCTGTGAAGGGC

ACAAACGCTTCTGCAAGTGGAGGGACTGCCAGTGTCCCAAATGTAAACTGACAGCTG
ACTTGACTGACATTCAAGgtaacggtttaccggtagcacggtagctgagactgaaaacagtgcattaaacaa
aaaaaaagggtttagtgcataaaaccgtggtagaaacagggaaatacaacatggaaagggtcacataactatttcaattatgttttatatttt
aatacagcgacaagttactgttagttcttaattctatttgttatcaacacgtcaaagcttagtccacagctagcgtacagagcttacta
actcttacacgcgttaatgagtggccactgtgttactacactgttactaatagaatttagtaatacagtaggttagtctcgtattaaag
agcaattactttccatttcatgtcattacttgtaacattgtacccgttagctgaagtaacattttagtggggaaacgcattatcattga
tcgtagctttgacagtagtaacaccgggaggtagcagggactaccctgtggccaagggtcaagctgcagatacgcacggctctgg
cggtcagaaaatcataagcatccaaaatgctccatttacctgagaatttaacgcattactgtgttg
tgcgtgtgcgcctcgtaattgtgtcttcaagcatattcgacatgtccgtctgtatgcacagcgttggacatattaaacttactat
tcatatagctttagtattacaagatagattgttatttgagaagggttatcgtaatgtattctaaacaaattcaatttgcatttgcatttgc
tggtagagacagaataacatgcataatggcatcgtccgtccattttatgtctgttgcacagacaaactg
ttgttattattgttataattatcatgttatgttatattctatacttctgttgcatttgcacagataatgtctgtttccctccccacca
gAGGGCAAGAAAAAGGAAGCAAGCTCGGCAGGAAAAGCTCGGTATTGCCAGGGCC
TGAGGTGTAGGTGAAGGATGAAGCCGGAGCAGACTGCATGCTCTGCTGAGGGACG
ATCCCCGACACCTACCAGTACCTCCAATTCTTCTCTGCTGCCACAGgttaggtattgcagtcatttta
gttccacagtttagtgcacttccatgtcaaaaccttcatgtgttactactaatgttagcaacagctgttgcactccactggtagtgc
aaaaacccataataaaactagtagtggccatgcacatgttagtgcacatggaaacatggaaacgttcaatagccaaacgcacattcagatgg
aaaacatgcataatggaaacatggaaagaaaacttcagacccgtcgactgcattgcacatgcataattacaaaacaaacaaaacaaaaccaaca
acacttcatttgcaccaactagcatattttatgttctgttatttagtgcatttgcataatggaa

>Predicated amino acid sequence of *Dmrt1b*

MNKDTQSKQVPESTGTLSPSTGQKHRRIFKCSLCRNGCIIPVKGHKRFCWRDCQCPKC
KLTADLTDIQEKKKEASSAGKARYFARA.

> Female scaffold49157 (Contains homologous sequences to the 3' UTR of *Dmrt1*)

GAAGGCCAGGCTATAAAGTGGAGTGAATAGATAGAACTTATGTTGTGATAAAAACACCT
GACAGAACTCAGGTGTGCCTCAGCTGCCAACAGGAGGAGTCAGCCACACCCCCCTGCC
ACCCACAGGCTCTGCAAAAGGAACAGAAAATGGAAAGGTAAACAGAGAGAGCACAA
AAAGAAAAGCACACAGAAAACATACTACACCAAACCATAACAACAATAAGCCTG
TCTCCATGGCTTCACTTTCAGACATTATTACTGTCTATGGTCAGCACCTGCTCTCT
TTGAAATATTGTCATTAAAATGCTGCTGAAACAACGGTTAGGAGCAACTGTCTGTCAA
CAAAACTGTCAACCAACAGATGCTGTTACTGATATTGATATTTCAGAGTACAATAGTA
TAGAGTGTAAAACCTGTATAATCATCGTACTTTTTCTTACAATTTGATAGGG
AAAACATTAAATAATTGTAACCATTGTACTGAGCAGGATAATTCATCAGACCATTGT
TTCTAAATGAGCTCAAACATGTCTTGTCTTGACAATAATTCCATGTCAGTCATCA
GGTGACACATACAAATATCTGCAATAAGCTATTCTTGACAATTATTGATGGTGGAT
GACAAGACACCTTGACTTGCAACATTGAAAGTACTCTGAATTAAATACTGTGATTT
ACCTCAAACAAACGATCGGATCAATGGAAAATCAGATGATTGCGAAGCTGCTGAAACT
GCACACGTTCTTCTTTCTGGTATAGATGAAAGTTAATTAAAGCTCAGAGCAGGGAA
CCTGCTGACAGAGTGTCTCCTGCAGGACAATGTGCAATTGAGGGTTAGGGGG
ACAAAAAAGCCAACATGAATCATTTCCCTTGTGTGTTGGTGTGATGAGCAACG
ACTGCTTATTATGGCGCTGATCCGGCAGCACCGTTATACCACAGAATTGATGTCTG
TTGTATTTAACCCCTCTGCCAGGAGTTGCTGTGCACACGGTTACTCTTACAC
ACTTTGGTACTGCCTCGTAGATACAGCAGAGGATGAGGGGGTAATCTGTAGCAGTGA
TGTGTAGAGCAGGAATATTGAGTTATTGATTAAGTCATCTAGAAAATGAGTCAG
CATCTAATTGAGTCGTTGGTCAGTCATTAAAGGCAATTCCAAGACAAAAAAGTAAAA
CATTTATAGTTTCAGCATCTTGGATTGCTGTTCTGTCCCTGAGTTGAGTAA
ATTGAATATCTTGAGTTCTGGACTGTTGGATCTGCCATTTCATGTTGTTG
ATGATTACAGACCCAGTAACATTAATCAAACCCCAAATCAAACAAAGTTGTTTTA
CTTCATGAACTAATGATAGTAATCCTGTTGCCACTATGAGTTGTTGTTAAGTG
TTAATTAAAAAACACAGATGAGATTGCAATTCTCTGCATTGTGAAGCTGACT
AGTACATGGCTGTATTGTATATTGCTGTGTTATGGAGATTCAACTGCATGCT
TTGAAAGTGTATTGTGTTGTTAGAATAAGAGAGCAACAAAGTAATTGGAGT
CTTGGAGTTATGTCGCTACAAGTTAGGAAAATAACGCCCTAAGTGATGAACTG
TAAATGTGATTGTGTTGAAACAACAATATAACAAAGAAAATTACTGTATT
TTTCTGCTGCCTGCCAACATTGATTCTGCCACTCTGTTGTAATCAATAACG
CGACCACCTGTTTATGAAACGCACCTACTCTTAATTCCACCAAGTGGTTTA
GAGAGAAGAATGTGGAGTTACCTATAAGACTTGTACATTAAATCACCCCTGAGACC
AGATTGTACAGCAGTATCGATTGATCCTGATATATCACAGAAATTACTGTCACCTCTGT
TTTCAATTAAAGATAACATCAGTCAGATAATGACTTAATTGGATTTCAGAAGATTGAG
TTTATTGCCCGTGCTTACGAGTTAGTTAAGGAAGGCCATTATCATGGTTGTCAG
GGGTGAGCTGCTGTGCTGCCCTGCTGAGCTGCTGCCCTGTGACAGCGCA
GGGGAGAAAGTGTACAGGATTGGTTCTGAGCTGCCACCTCCATGAATCAT
GCCCTCCCATAATAGATAAAATCTGCTCGATCCCACCGTGACACCTCCTGGTGGAGTG
ATGGCCGCTGACACAGACGAGAAGTGTGGAAGAGGACGCCGTAAGATGTTGGTGTAT
ACTTGTTGATACTTGGATGTTGAGTTCATCTGAGTGTAGATGAAATATGTTGATGA
GGTCATCATAACAGGATCATAACAGGATTGGAGCTTTATGCCCTGAGGTAGGGCT

GAACATGTTCAATAAGACAAAAATTGGTCTGAAATGTGTACCACTTGAAGTTAGGCA
ACATGATCTGCAACTTATTGCATTCTCCTTTATCCTCTTACCTTTGGTGAA
AATCTGTAACAGCATACTTAATGCTGAGCCTTAGGTTAAAGTTAGTATATACCTGTT
AAATGTTGAAGATTAGTTAGCAAACCAGTACAAACTATATGCTCATTCAATAAAGTACG
TGACAAGGCCTGTTAATGTTGACAGACATCTCTTGGCTAAATACAACGTATTGAATG
ATTCAAGATATTAAATAATATTATTCTGTTTCCAAGGTTGATTATTGTGTGTATTGTTCC
AAATATTATATGAAGAAAGATCTGCTTTCTTTATGACGTGATCTGTTGTGGGTTTC
AATACCTCAGCGTTACTGTGATATATTCAAGATTTTGAATGAGAGATCAGAAATATCA
GTTTATCGTGATAGATCTACTGGCGGATTTGGCATGTTT

>Male scaffold162645 (Exon 2 of *Dmrt1b*)

>Male scaffold106307 (Contains homologous sequences to the intron 3 of *Dmrt1*)

>Male scaffold107747 (Contains homologous sequences to the exon 4 of *Dmrt1*)

ctcgtggcatgaaatgcatttagaaacaggctggagaggttttatcaagtttatcatcctctatctgatttcgagaacacccctt
agttcatgtcagtgaagaaccaaccctaactgttaaggaaaataataattacattacacacaattaatcagctaattttactccat
gtccaggatagcttaattttactggacaaccacaacaaggtaatagaggatgtatgagaaggcttatgccttgactctgttggatgtac
ctaccacgcccagaaaacctctgaaggagggtctgaggagatattccagatcagatctgccccgtctttgaaggagcagtggcttact
tgagccccatccagcctgcagagggactcatttatgcagcctgtatctgtgaagtcatcttcggcgtctacccaaagctcctaccacagact
tccagctcagcttgcattttactgttgatctgcaccaatctgcctgtcaacatctcagaacaatgtctattttgtcattttggatactgaaca
tacagacacgacttcagccccaccaataacactgagaaagggttatgaaactgttaatattggcttcacattgcacnnnnnnnnnnnnnnnnnn
nn
tggatcccttcgtggcagaacctgccttttgcataaagacagttttggctgttatgttgcattttgtcataactgttaatattggcttcacatt
actctcaagacccagttcaatctcaatgtatttcacagcctaatacagagactgttagcttaataacagttataaagctgtcgagaaat
tgcgttagacagactcgaaaaagtttcagaggactgaaagcagcgtttgcagetcatttttagaaaaggccacaacccatgaattcaaa
caactactaattctctgttatttttacctctatttcatccctcagatttagtttgcgttagctatgttgcataactgtcgccatgttggagaact
tacagtcaaacatgggtgaaactgagtcataatgttcatattgtatttgcataatataagaaaaggacccatgtttaaaaaaaaaagtt
atatgataaaaaaaaacaatctgtacaatgcgtactcgatgttgcacatcgacatACGACAATTCTTATATCAG
CCTCCTGTCGAGCCGCAACATGTCCCTCGGTACTGCATGCATTCTACTACCCAGCAG
ACACCTACTCGGCTCAGAGACTGGATTCCACCCCTGTGTGCCAAGTGTCTAACCT
GGAGGACAACAAACAGCAACAAACAATGACAACAAATGACAACAACTGCTTGAGG
CCATGGCTGCATGTNNnn
nnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnntacacacaaaacactcacacacacttccagggccataaagtgttactc
ggcatttacccacagtgttagttactgtcggtcaaggagaat

>Male scaffold68937 (Contains homologous sequences to the 3' UTR of *Dmrt1*)
TTTGATAGGGAAAACATTAAATAATTGTAACCATTGACTGAGCAGGATAATTCA
GACCATTGTTCTAAATGAGCTCAAACATGTCTTGTACAATAATTCCATG
TCAGTCATCAGGTGACACATAACAAATATATCTGCAATAAGCTATTCTGGACAATT
TATTGGTGGATGACAAGACACCTTGACTGCAACATTGGAAAGTACTCTGAATTAAA
TACTGTGATTACCTCAAACAAACGATCGGATCAATGGAAAATCAGATGATTGCGAAGC
TGCTGAAACTGCACACGTTCTTCTTCTGGTATAGATGAAAGTTAATTAAAGCTCA
GAGCAGGGAACCTGTCTGAGTGTCTCCTCCTGCAGGACAATGTCAATTGAGGGTTA
GGGGGGACAAAAAGCCAACATGAATCATATTCCCTTGTGTTGTTGGTGTGATGA
GCAACGACTGCTTATTATGGCGCTGATCCGGCAGCACCGTTATACCACAGAATTG
ATGTCGTGTATATTAAACCCCTCTGGCCAGGAGTTGCTGTGCACACGGTTACTC
TTTACACACTTGGTACTGCCTCGTAGATAACAGCAGAGGATGAGGGGGTGAATCTGTAG
CAGTGATGTGTAGAGCAGGAATTTGAGTTATTATCTGATTAAGTCATCTAGAAAATG
AGTCAGCATCTAATTGAGTCGTTGGTCAGTCATTAAAGCATTCCAAGACAAAAAA
GTAACACATTATAGTTCAGCATCTTGGATTGCTGTTTCTGTGCCTGAGTTG
TAGTAAATTGAATATCTTGAGTTCTGGACTGTTGGATCTGCCATTTCATGTTGT
TTTGTGATGATTACAGACCCAGTAACATTAATCAAACCCAAATCAAACAAAGTTG
TTTTACTTCATGAACATGATAGTAATCCTGTTGCCACTTATGAGTTGTTGTT
AAGTGTAAATTAAAAACACACAGATGAGATTGCCAATTCTCTGCATTGTGAAGC
TGACTAGTACATGGCTGTATTGTATATTGCTGTGTTATGGAGATTCAAACTGC
ATGCTTGAA

>Male scaffold118347 (Contains homologous sequences to the 3` UTR of *Dmrt1*)
GACTTAATTGGATTTACAGAAGATTGAGTTTATTGCCCGTGCTTACGAGTTAGTT
AAGGAAGGCCATTTATCATGGTTGTACAGGGTGAGCTGCTGCTGCCTGCTCCT
GTAGCTGCTGCCCTGTGTACAGCGAGGGGAGAAAGTGTACAGGATTGGTTTC
CTTGAGCTGCCACCTCCATGAATCATGCCCTCCATAATAGATAAAATCTGCTCGGA
TCCCACCGTGACACCTCCTGGTGGAGTGATGCCGCTGACACAGACGAGAAGTGTG
GAAGAGGACGCCGTAAGATGTTGGTGTACTTGTGTACTTGATGAGGATGTT
GAGTGATGATGAAATATGTGTGATGAGGTGATCAGGATCATACAGGATT
GGAAGCTCTTATGCCTGTAGGTAGGGCTGAACATGTTCAATAAGACAAAAATTGGTCT
GAAATGTGTACCACTTGAAGTTAGGCAACATGATCTGCAACTTATTGCATTCTCCT
TTTATTCCCTTTACCTTTGGTAAAATCTGTAACAGCATCCTAATGCTGAGCCTT
AGGTAAAGTTAGTATACCTGTTAAATGTTGAAGATTAGTTAGCAAACCCAGTACA
AACTATATGCTATTCAATAAAAGTACGTGACAAGGCCTGTTAATGTTGACAGACATCTT
CTTGCGCTAAATACATCTGATTGAATGATTGAGATATTAAAGAAATATTCTGTTCCA
AGGTTGATTATTGTGTATTGTTCCAATATTATGAAGAAAGATATGCTTTCTT
TTTATGACGTGATCTGTTGGGTTCAATACCTCCCGGTTACTGTGATATATTGAGATT
TTTCGAATAAGAGATCAGAAATATGAGTTATCGTGTAGATCTACTGGCGGATTG
ATGTTGCAATAATCATCTGATTCTACAGTTGTCAGTTCTCCATCCATCCATCT
TCCGCTCACCCGGGACCGGGCGTGGGGCAGCAGCTGAGCAGGGATGCCAGAC
TTCCCTCTCCCCAGACACTCTCCAGCTCTCCGGCAGACCCCCAAGCGTCCCCAG
GCCAGCCGAGTGACATAGTCTCAATATAAAAGTAAAATTCTCACAGGTAAGAGTGGAG
CTGATTGAACTATTTATAATGTTCAAGAGAAAACATAATTAAATTATTGAGTCTAAAAG
TATTTTCACAACATAAGTGGAGAGTCAAGCTACATTATTCAATTAGAAAACAAC
GTTTCAGGAACCTTTAGAAATGAGACAGTAAAACAGAATAAGGGATAGCTGCTTAG
ACTCAAGAGAAGTGTGACTTAATCCCAGAAAATAGTTCCCTGTTAAAACAAGCTG
AAATTGGAAATATCCAACACTACAGTACAAGTGCCTGTAATTGGTCACATAAAATGTGG
TAATCTGAAACTAATCTGTATCGACAATTAAAGCAACAGGGGCACATATTGAGAAC
CCCTAAACAAAATCGAACACTTCAAAATCATACAGCAGTTTATCAGCATCTTATATG
CTGAGATGCTGATTGTACAATTATCCAATTAAATCATTGGCCACTTCCGTG
ACTTCCCTAACAAATAAGATAATGTCAATCATGTGTTCATAGGGTTGGTCAAGTC
AAGTCAAGTCAAGTCAACTTATTGTACA