Science Advances
AAAS

Supplementary Materials for

The genetic architecture of phenotypic diversity in the Betta fish (*Betta splendens*)

Wanchang Zhang *et al.*

Corresponding author: Rasmus Nielsen, rasmus_nielsen@berkeley.edu; Yijiang Hong, yjhong2008@163.com

Sci. Adv. **8**, eabm4955 (2022)
DOI: 10.1126/sciadv.abm4955

This PDF file includes:

Detailed Table of Contents

Text S1. Genome assembling and annotation

Before assembly, the genome size of Siamese fighting fish was estimated to be 487.89 Mb using Jellyfish (v2.0) (*79*) with a heterozygosity of 0.27% determined by *k*-mer (*k*=17) frequency analysis using 43.51 Gb Illumina sequencing data. The genome was assembled using 53.27Gb (~110×) PacBio Sequel SMRTbell data, augmented with 10X genomics, BioNano optimal map and Hi-C data for scaffolding and anchoring to chromosomes.

Genome assembling

The Falcon (v0.3.0) (https://github.com/PacificBiosciences/FALCON/) (*80*) was used to preliminary assemble the PacBio reads using parameters "--max_diff 100 --max_cov 100 --min_cov 2 --bestn 10 --min_len 6000". The Falcon assembly was polished with Quiver5 (*81*), and error-corrected with 53.05 Gb Illumina HiSeq X Ten data from the same individual using Pilon (v1.18) (*82*) with default parameters. Then, purge_haplotigs (*83*) was applied to remove the heterozygous contigs under parameters "-l 25 -m 70 -h 125 -a 75". Pre-scaffolding of the contigs were performed using fragScaff (version2.1) (*84*) with parameters "-m 3000 -q 30 -E 30000 -o 60000 -C 5 -j 1.5" to associate contigs into scaffolds based on 114.98 Gb data of linked reads generated with 10X Genomics technology. Contigs scaffolding with optical mapping was performed with the BioNano Saphyr system (BioNano Inc.). To construct the optimal map, DNA extraction and labeling were conducted with standard protocols from BioNano Genomics. Single molecules with length exceeding 100 Kb were used for assembly with RefAligner and Assembler from BioNano Solve (version 3.1) (*85*) software package. The assembled optical map was aligned to the sequence assembly with RefAligner using default parameters. Assembly into the chromosome level was performed with Lachesis (v-201701) (*86*) software with the 52.58 G Hi-C data.

Genome annotation

We first annotated the tandem repeats with Tandem Repeats Finder (TRF, v4.09) (*87*) by ab initio prediction. For other repeats, RepeatModeler (v2.0.1) (*88*) , LTR_FINDER (v1.07) (*89*) and RepeatScout (v1.0.5) (*90*) were used for de novo repeat detection. And RepeatMasker (4.1.0) (*91*) against the Repbase TE library (*92*) and RepeatProtein Mask (4.1.0) (*93*) against the TE protein database (*91*) were used for homology-based repeat detection. To annotate non-coding RNA including tRNA, rRNA, miRNA and snRNA, tRNAscan-SE (v1.4) (*94*) (http://lowelab.ucsc.edu/tRNAscan-SE/, default options) was used to find the tRNA sequences according to their structural features. We use the rRNA sequences of related species as the references and find the highly conserved rRNA sequences in our genome assembly through Blast (v2.2.26). The sequences of miRNAs and snRNAs were predicted

with the Infernal ("INFERence of RNA ALignment") (http://infernal.janelia.org/) software (v14.1) (*95*) using the covariance model of the Rfam family.

To annotate the protein-coding genes, we combined de novo predictions, homolog-based predictions and RNA-Seq based predictions approaches. Five ab initio gene prediction programs were used to predict genes, including Augustus (v3.0.2) (*96*), Genescan (v1.0) (*97*), Geneid (v1.4) (*98*), GlimmerHMM (v3.0.2) (*99*), and SNAP (v2006-07-28) (*100*). For homolog-based predictions, protein sequences of related species were downloaded from Ensembl or NCBI. Homologous sequences were aligned against the repeat-marked *Betta splendens* genome using TBLASTN (v2.2.26, e-value ≤ 1e-05). Genewise (v2.4.1, "-tfor-genesf") (*101*) was employed to predict gene models based on the alignment sequences. For RNA-Seq based predictions, Trinity (v2.1.1, "--normalize_reads --full_cleanup --min_glue 2 --min_kmer_cov 2 --KMER_SIZE 25 --no_distributed_trinity_exec") (*102*) and SMRTLink (v6.0, https://www.pacb.com/support/software-downloads/, default parameters) were used to assemble the RNA-Seq data, and then PASA software (http://pasapipeline.github.io/) (*103*) was used to improve the gene structures. A weighted and non-redundant gene set was generated by EVidenceModeler (v1.1.1) (*104*), which merged all gene models predicted by the above three approaches. To annotate the gene functions, we extracted the longest transcript of each gene. The translated protein sequence of the transcript was used to 'blastp' search the Swissprot database (http://www.uniprot.org) (*105*) with the cutoff "E-value <10-5", then the gene annotations and gene ontology terms from the closest hit were extracted for gene prediction. The conserved protein domains and functional motifs of the transcripts were identified with InterproScan (*106*) with default settings. The gene pathway annotation was obtained by mapping to the KEGG database (http://www.genome.jp/kegg/) (*107*).

To evaluate the quality of our assembly, we mapped the Illumina short reads to the assembled genome using BWA (*64*), and 97.51% of reads were successfully mapped. 99.87% of the assembly was covered, indicating the near-completion of the genome. Filtered variants called from Illumina data yielded 198,849 heterozygous SNPs (0.0936%) and 532 homozygous SNPs (0.0003%), supporting high base level accuracy of the genome assembly.

Text S2. Background on the Betta fish and the samples

The common name Betta fish, refers to the wild and domesticated fish in the genus *Betta* distributed in Southeast Asia, especially in Thailand, Cambodia, Vietnam, Malaysia and Indonesia. The Siamese fighting fish, *Betta splendens*, primarily originated from Central Thailand and the lower Mekong, is well-known for its aggressive behavior and various domesticated forms (*2*). This fish was traditionally used in gambling matches and has been domesticated for at least 300 years (*2*). Today, multiple varieties have been raised differing in aggressiveness, fin shape, color morphology and body size. The inheritance of several traits such as sex-determination, long-fin vs short-fin, and coloration of Turquoise-green, Royal-blue and Steel-blue colors, were explored since the 1930s and 1940s (*3-5, 7, 8*). However, there have been relatively few molecular studies on the genetics of *B. splendens* domestication traits, until a recent study of the double tail, elephant ear, albino and fin spot and sex determination phenotypes (*10-12*). In addition to the interest in its phenotypic diversity, the Siamese fighting fish is also considered as a popular model for toxicological and behavioral research.

Although initially domesticated for its aggressive behavior, numerous other breeds have been selected for ornamental purposes. In total, we have sampled five major caudal fin morphological types: The Fighter and four types of non-Fighters, namely Crowntail, Veiltail, Halfmoon and Halfmoon Plakat (HMPK). The Crowntail, Veiltail and Halfmoon all have distinctive long fins, and therefore considered as long-fin morphotypes. The Halfmoon has a "D" shaped caudal fin which forms a 180-degree angle, and HMPK are short-tailed Halfmoon. The Fighters have short tail, and their tail spread is usually less than 180 degrees. Within HMPK, breeds were further classified based on body color (such as Red, Yellow etc.), body size (e.g., Giant), and pectoral fin morphology (e.g., Dumbo breed with enlarged pectoral fins) (fig. S4). We collected 13 breeds with HMPK tail in total.

We note here that although we are using the term "breed" in this section and throughout the paper, these are not breeds in the traditional sense of inbred lines, since these morphotypes are often interbred. One example is the Royal-blue phenotype, which produces Royal-blue, Turquoise and Steel-blue offspring when selfed (*5*), showing that a segregating locus underlies all those phenotypes. Another example is the Dumbo phenotype, which exists in both Betta fish with HMPK and Halfmoon tail types. Being aware that what we are calling "breeds" are akin to morphotypes, we have sought to assign individuals to breeds not only based on phenotypic similarity, but also considering the potential for a shared genetic basis for the phenotype based on past breeding experiments.

We also note that the 'wild' individuals we sampled here are actually brought from the fish farms, not collected from their natural habitats. These phenotypically 'wild' individuals

have been raised in fish farms for many generations. They are frequently used as donors of traits for improving domesticated *B. splendens*, and at the same time, they themselves are frequently improved by hybridizing with domesticated *B. splendens*.

Text S3. Population genetics of the Betta fish

The phylogeny of the *Betta* species

To infer the phylogeny of the *Betta* species, we constructed a maximum-likelihood (ML) tree using concatenated sequences at genome-wide SNP sites (fig. S5). We note that, due to recombination and independent assortment, different regions of a genome will have different trees. The ML tree can be considered an estimate of the average tree across the genome. While it is important to avoid confusion of the interpretation of the branch lengths in such a tree, it is useful for elucidating individuals, and groups, that in average are more closely related to each other genetically than other such groups. In the estimated tree, all individuals from *Betta splendens* species form a clade relative to other wild species, providing genetic evidence for a single origin of all domesticated Siamese fighting fish from the wild *B. splendens*, despite their extraordinarily high phenotypic diversity. This result is also compatible with the PCA in which all *B. splendens* individuals form a cluster (Fig. 1C).

The wild species on the most basal position of the tree, the *B. smaragdina* and *B. smaragdina guitar* individuals are interspersed with each other, showing that these two groups are not genomically differentiated (fig. S5). All the *B. stiktos* individuals form a monophyletic clade which is positioned as a sub-clade in the *B. smaragdina* (fig. S5). These observations suggest that all the three nominal species should be considered as one species group even if they are phenotypically distinct. *B. imbellis* and *B. siamorientalis* are the wild species that are most closely related to the *B. splendens*. Individuals of the two species form a cluster, and *B. siamorientalis* individuals form a monophyletic clade which is a subclade within the *B. imbellis* cluster, suggesting that these two species should also be considered as one species group. All the *B. mahachaiensis* individuals form a monophyletic clade (fig. S5). The relationships of different species are also recapitulated by the first principal component (PC) which divided the samples into four clusters corresponding to the *B. splendens* group, the *B. imbellis* and *B. siamorientalis* group, the *B. mahachaiensis* group, the group consisting of *B. smaragdina*, *B. smaragdina guitar* and *B. stiktos* (Figs. 1C and S7). We note that all our domesticated samples are purchased from commercial vendors in China and local farmers in Thailand. It is, therefore, possible that some of the evidence for admixture between nominal species could be a consequence of captive breeding.

Population structure of the *B. splendens*

Within the population of *B. splendens*, individuals of the same breed tend to cluster together in the genome-wide tree, showing that individuals of the same phenotype are also genomically similar. Individuals of the Steel-blue, Turquoise-green and Royal-blue are

interspersed with each other, which is consistent with the fact that the Royal-blue individuals are offspring of the hybridization between Turquoise-green and Steel-blue breeds (Fig. 1B). Genetically, the Red phenotype is dominant over the Yellow phenotype, and they form independent clusters on the ML tree (Fig. 1B). The Orange phenotype is considered an intermediate phenotype between Red and Yellow, but all Orange individuals cluster with Yellow individuals. This observation is also confirmed by admixture and PCA showing Orange and Yellow are more closely related and genetically differentiated from the Red breed (Figs. 1C and S8). The Halfmoon individuals are divided into two clusters, which is also consistent with the admixture analysis in which two subgroups show different component profiles (fig. S6). The Dumbo-Halfmoon breed is featured for both its distinctively enlarged pectoral fin (the Dumbo phenotype) and its semicircular tail that is similar to the Halfmoon breed, but its genetic origin remains elusive. In the genome-wide average phylogeny (Fig. 1B), the Dumbo-Halfmoon clusters with Dumbo-HMPK rather than Halfmoon, which is also supported by the admixture result as the two Dumbo breeds share the similar genetic components that is distinctive from the Halfmoon breed (fig. S6). These results suggest that the Dumbo-Halfmoon was bred by introducing the long-fin phenotype into the genomic background of Dumbo.

The population structure of the domesticated *B. splendens* is largely driven by several breeds that carry substantial breed-specific genetic drift, including the Black, Orange and Yellow, the clade consisting of Steel-blue, Royal-blue and Turquoise-green, and Opaque (Figs. 1B-D). These breeds have long breed-specific branch lengths on the genome-wide tree (Fig. 1B) and show distinctive components in the admixture analysis (Fig. 1D). The most noticeable is the group consisting of Steel-blue, Royal-blue and Turquoise-green, as they form an independent cluster in the PCA (Fig. 1C). The breed-specific genetic drift is likely due to the strong bottleneck effect imposed by local breeders to maintain pure lines.

Admixture and gene flow among the Betta species

Interspecies hybridization can be viable in Betta fish and it is a common practice to capture wild Betta species to hybridize with domesticated fish to produce new phenotypes. Hybridization among different breeds of domesticated *B. splendens* is also a common practice for producing new variates. Moreover, the domesticated *B. splendens* can easily survive in nature when they escape or are released by humans (*108*). Release or escape of domesticated *B. splendens* in Thailand may possibly facilitate hybridization in the wild between wild and domesticated breeds of *B. splendens*.

To explore the gene flow among different Betta species, we performed D-statistics analyses in the form of D(P1, P2; P3, P4), which tests whether the four populations (individuals) are compatible with the unrooted phylogeny ((P1, P2), (P3, P4)). When setting

P4 as the outgroup, then D(P1, P2; P3, Outgroup) < 0 provides evidence of gene flow between P2 and P3 (or external gene-flow into P1). We randomly picked three individuals from each of the species or breeds, and iteratively tested all non-redundant combinations of D(P1, P2; P3, Outgroup) where P1 and P2 are *B. splendens* individuals, P3 are individuals of *B. mahachaiensis*, *B. imbellis* or *B. siamorientalis*, and Outgroups are *B. smaragdina* individuals (fig. S11). For each of the P3 species, the Z score distribution of D(P1, P2; P3, Outgroup) greatly deviates from the null expectation, which shows pervasive rejection of D(P1, P2; P3, Outgroup) ~ 0, suggesting ubiquitous gene flow between the domesticated Betta fish and all three wild Betta species. When testing D(P1, P2; P3 = *B. mahachaiensis*, Outgroup) with *B. splendens* as P1 and P2 (fig. S12A), we found mostly negative D values when P2 is represented by either Royal-blue, Steel-blue, Turquoise-green, Dragon and Copper individuals, and mostly positive D values when using Veiltail, Crowntail, Fighter or Black individuals as P2 (fig. S12A). While these results are not necessarily driven by gene flow specifically from or into *B. mahachaiensis*, they do provide strong evidence that different breeds have different levels of gene flow with *B. mahachaiensis* related individuals. Moreover, individuals of the same breed also show different patterns of gene-flow: among the three wild *B. splendens* individuals, two of which (BSP1 and BSP2) show largely positive scores in the D test (P, wild *B. splendens*; *B. mahachaiensis*, Outgroup) (fig. S12A), but one of them (BSP9) shows largely negative scores. When switching P3 with *B. imbellis* or *B. siamorientalis*, we observe similar patterns which collectively suggest extensive heterogeneity in the gene flow between different breeds and wild species, or even between different individuals of the same breed and wild species (fig. S12). This level of heterogeneity is only compatible with a scenario in which gene flow between domesticated and wild species is common and recent.

To capture the major patterns of gene-flow within the Betta species, we performed TreeMix analyses by grouping individuals of the same species or breeds as one population, allowing migration edges to vary from 0 to 7 (Figs. S9-S10). As discussed above, there is extensive and frequent gene-flow between Betta species, especially between the domesticated and wild Betta fish. In this case, TreeMix may actually be unable to resolve the true scenario by falling into local optima (*109*). We also note that, when too many populations are admixed, and each population has undergone multiple waves of admixture, the gene flow signals captured can actually be represented by different TreeMix models. In this case, the migration edges on the TreeMix graph cannot simply be interpreted as gene-flow between the two populations. With this information in mind, we attempt to highlight the most prominent and consistent patterns in the TreeMix graphs.

Most notably, when allowing migrations, the wild *B. splendens* group is best described as a mixture between domesticated breeds and wild populations that is basal to the clade

consisting of *B. imbellis* group and *B. splendens* group. This model is also consistent with the PCA (Fig. 1C) where the wild *B. splendens* individuals fall within the cline consisting of *B. splendens* group individuals on the y-axis, but shift towards other wild species on the x-axis. In the model with 4 and 5 migration edges, Mosaic and Giant breeds can be described as having received gene flow from the *B. imbellis* related branch. This hypothesis is also supported by D-statistics testing D(P, Giant/Mosaic; Imbellis/Siamorientalis, Outgroup) in which the D statistics tend to be negative (Figs. S12B-C).

Another prominent observation from the D(P1, P2; Mahachaiensis, Outgroup) tests (Fig. S12A) is that when setting P2 as Royal-blue, Turquoise-green or Steel-blue, the D values tend to be negative, suggesting gene flow between *B. mahachaiensis* and these breeds. In TreeMix, we observe consistent results when assuming 4 migration edges indicating gene-flow from *B. mahachaiensis* into the clade consisting of Steel-blue, Royal-blue and Turquoise-green breeds (fig. S10).

Supplementary Figures

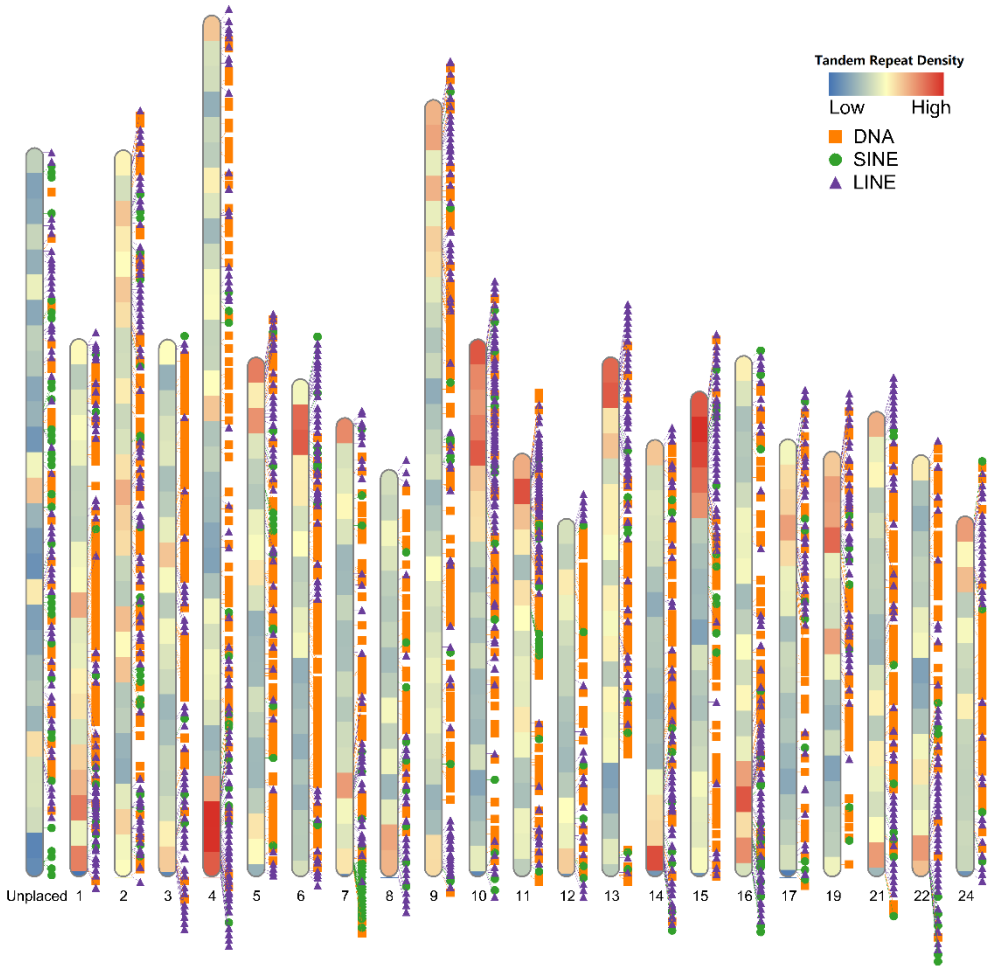

Figure S1. Genomic landscape of repeat sequences in the Siamese fighting fish genome.

The color of the chromosome fragments corresponds to the repeat sequence density (shades of red represent higher density and shades of blue, lower density), and the DNA transposons, SINEs and LINEs are denoted as solid rectangles, circles and triangles, respectively. All the unplaced scaffolds are stacked together on the left “Unplaced” block. All the elements are anchored into the genome using the “RIdeogram” package (*110*) in R.

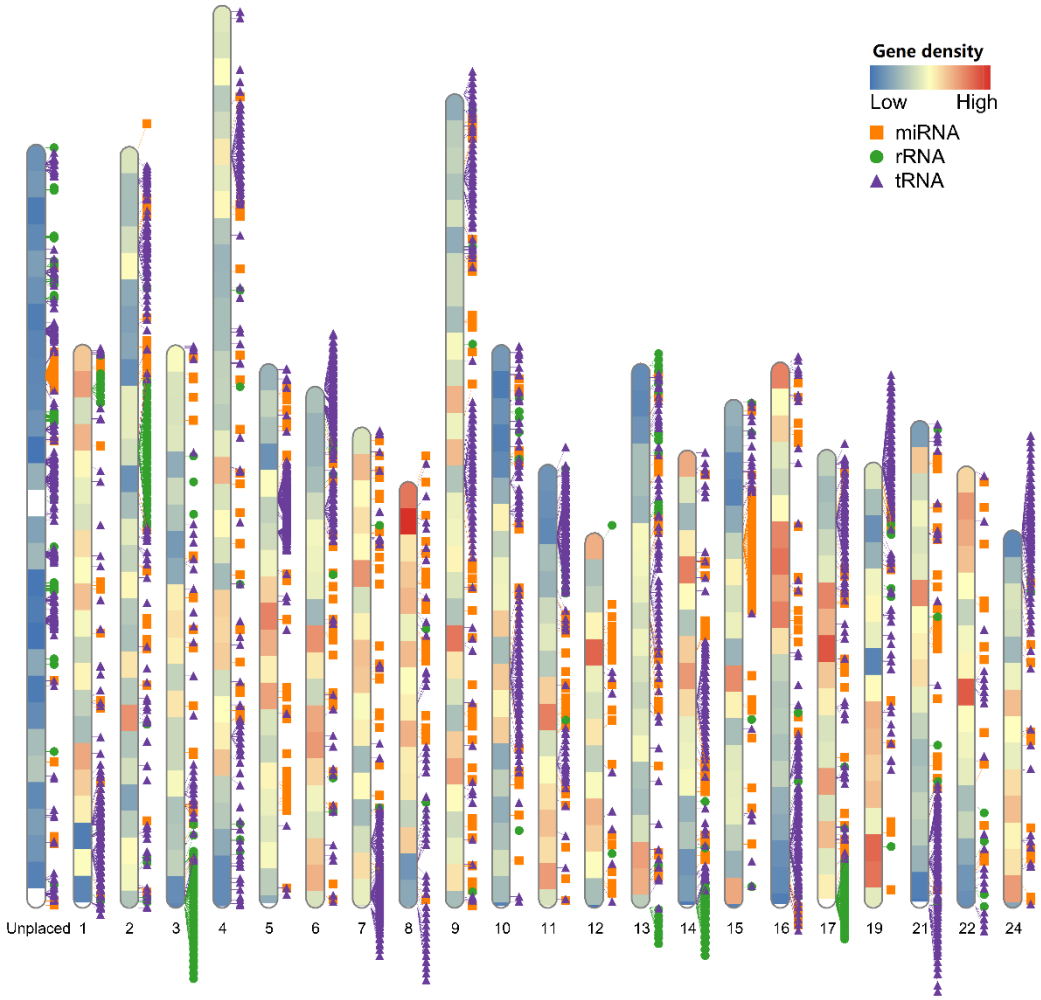

Figure S2. Genomic landscape of annotated protein-coding genes and non-coding RNA elements in the Siamese fighting fish genome.

The color of the chromosome fragments corresponds to the gene density (shades of red represent higher density and shades of blue as lower density), and the miRNA, tRNA and rRNA are denoted as solid rectangles, circles and triangles, respectively. All the unplaced scaffolds are stacked together on the left "Unplaced" block. Note the gene hotspot close to the beginning of chromosome 8. All the elements are anchored into the genome using "RIdeogram" package (*110*) in R.

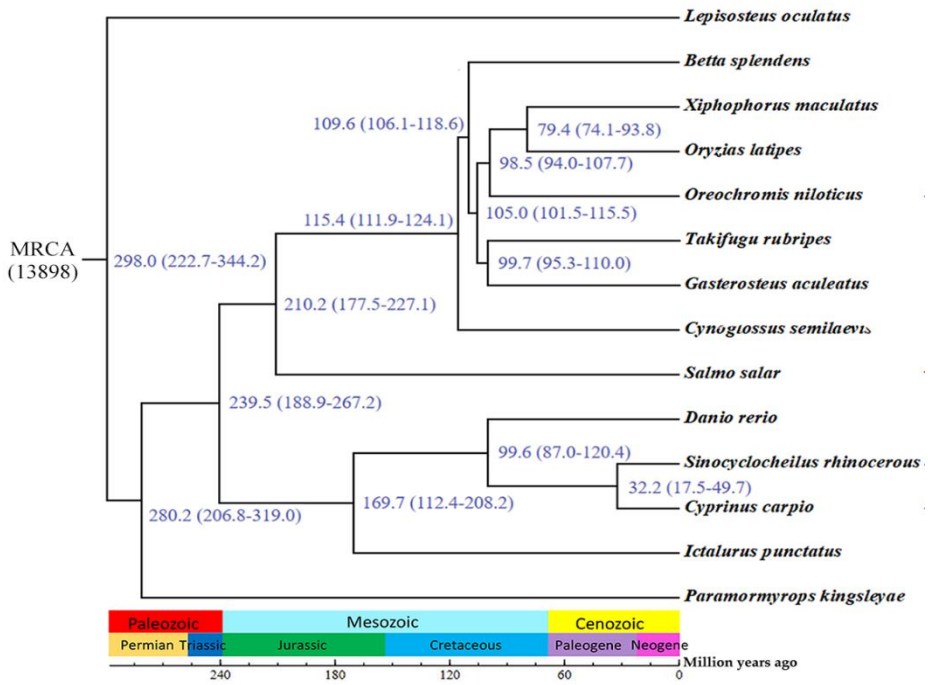

Figure S3. Time-calibrated phylogeny of the Siamese fighting fish and other 13 teleosts.

The phylogeny is constructed with genome wide single-copy orthologous genes from 14 teleost species including the Siamese fighting fish (*Betta splendens*). The numbers close to the nodes of the tree are the estimated divergence times (in units of millions of years ago).

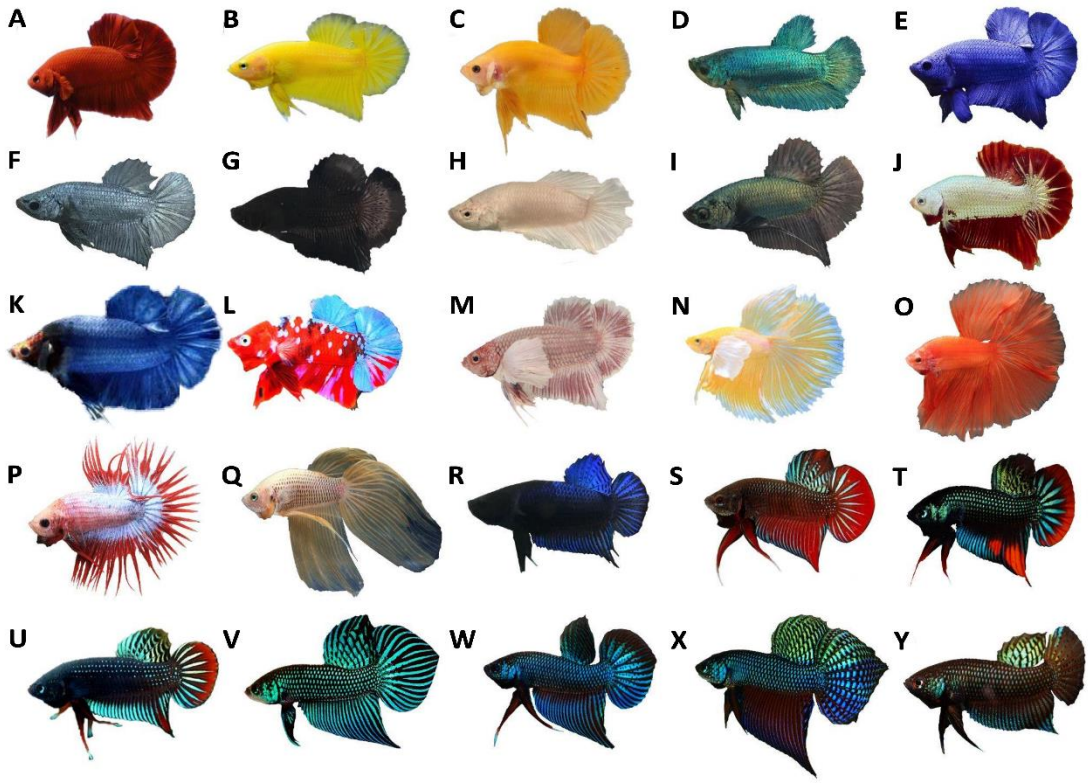

Figure S4. Illustration of samples used in this study.

Images (**A**-**R**) are different domesticated breeds of the Siamese fighting fish and (**S**-**Y**) are different wild Betta species. (**A**) Solid Red Halfmoon Plakat (HMPK); (**B**) Solid Yellow HMPK; (**C**) Solid Orange HMPK; (**D**) Turquoise-green HMPK; (**E**) Royal-blue HMPK; (**F**) Steel-blue HMPK; (**G**) Solid Black HMPK; (**H**) Solid Opaque HMPK; (**I**) Copper HMPK; (**J**) Dragon HMPK; (**K**) Giant HMPK; (**L**) Mosaic HMPK; (**M**) Dumbo HMPK; (**N**) Dumbo Halfmoon; (**O**) Halfmoon; (**P**) Crowntail; (**Q**) Veiltail; (**R**) Fighter; (**S**) wild *Betta splendens*; (**T**) *Betta imbellis*; (**U**) *Betta siamorientalis*; (**V**) *Betta mahachaiensis*; (**W**) *Betta smaragdina*; (**X**) *Betta smaragdina guitar*; (**Y**) *Betta stiktos.*

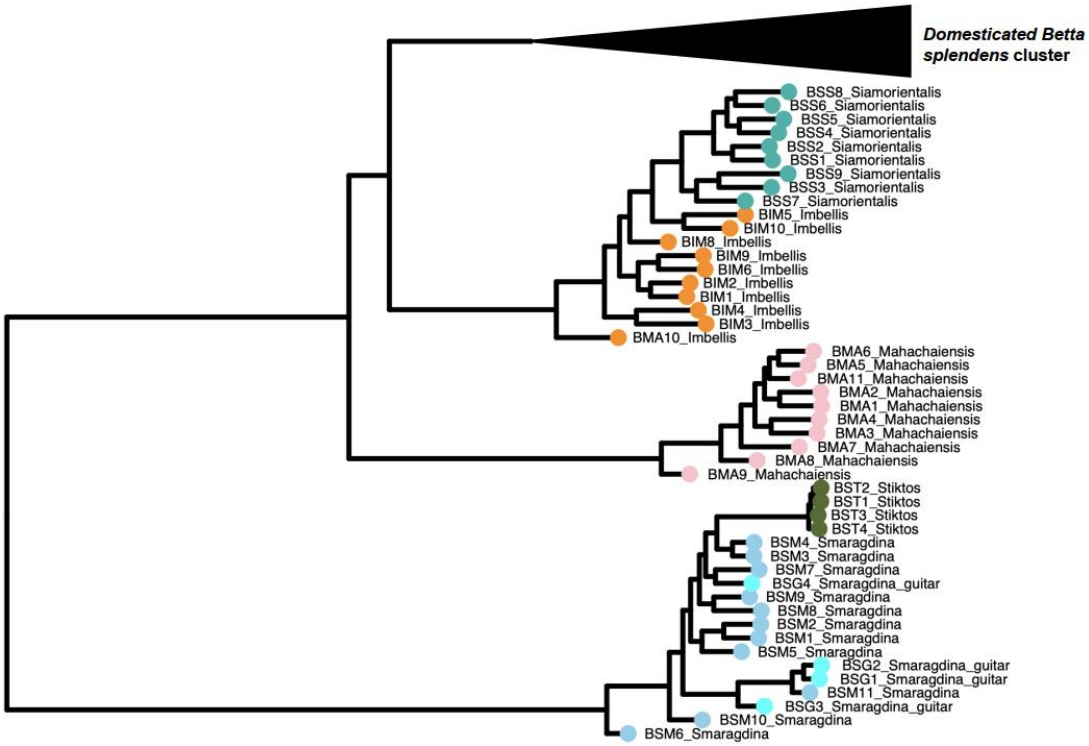

Figure S5. Maximum-likelihood phylogeny of the Betta fish constructed with genome-wide markers.

This is related to Fig. 1B. The black triangle contains both wild and domesticated *B. splendens*.

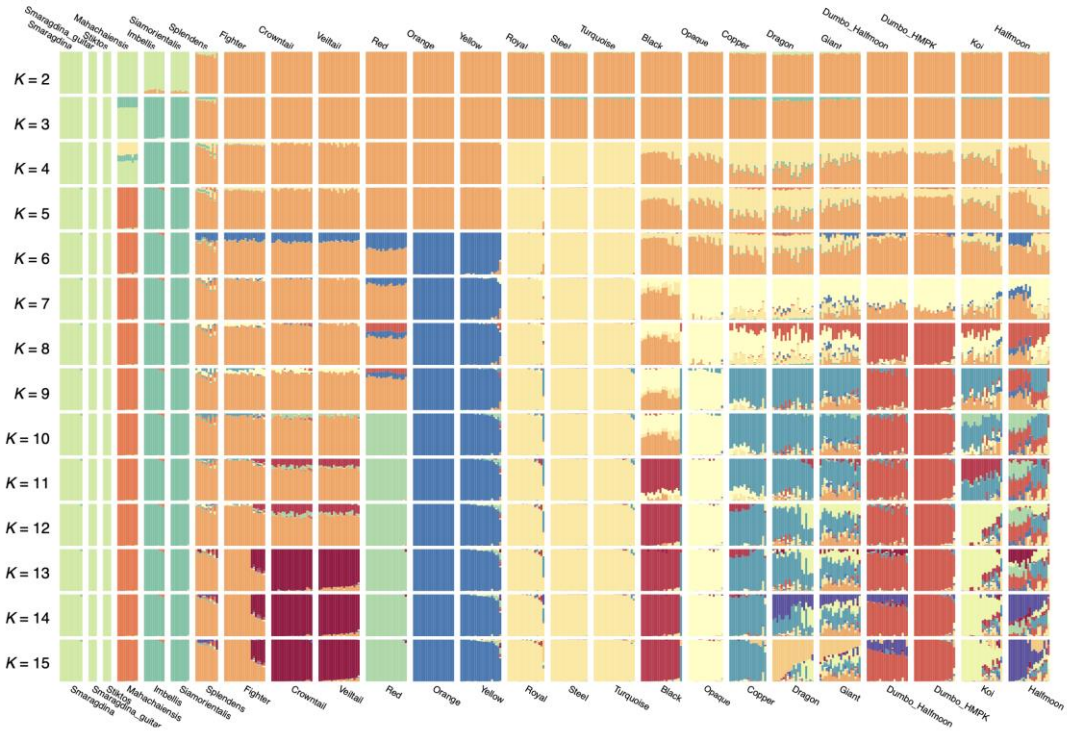

Figure S6. Admixture analysis of the Betta fish with K varying from 2 to 15.

Individuals of the same domesticated breeds that are phenotypically defined by color and fin morphology (listed on the x-axis) generally cluster together.

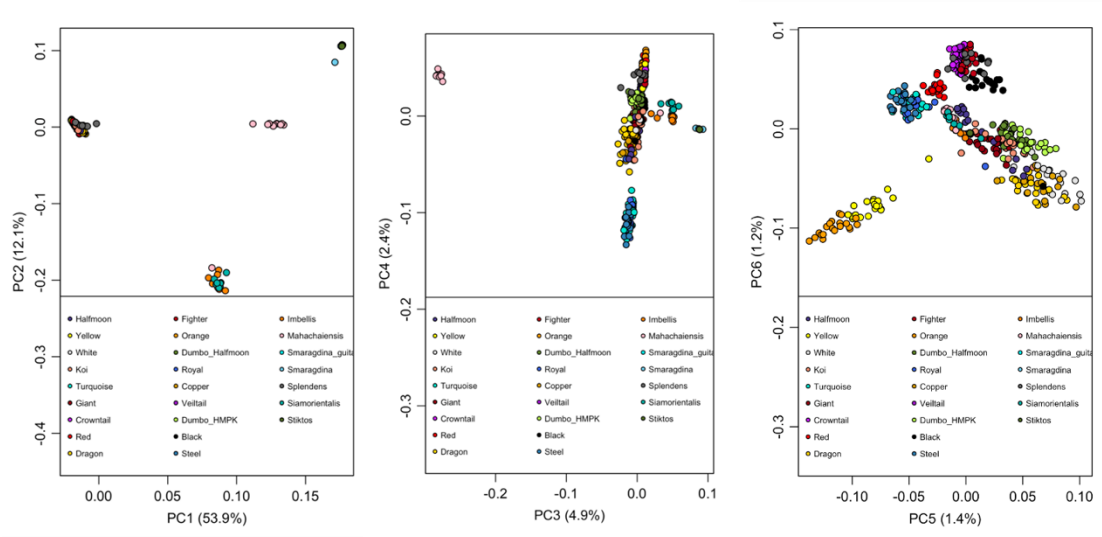

Figure S7. PCA of all Betta fish samples in this study.

Principal components 1 through 6 are shown.

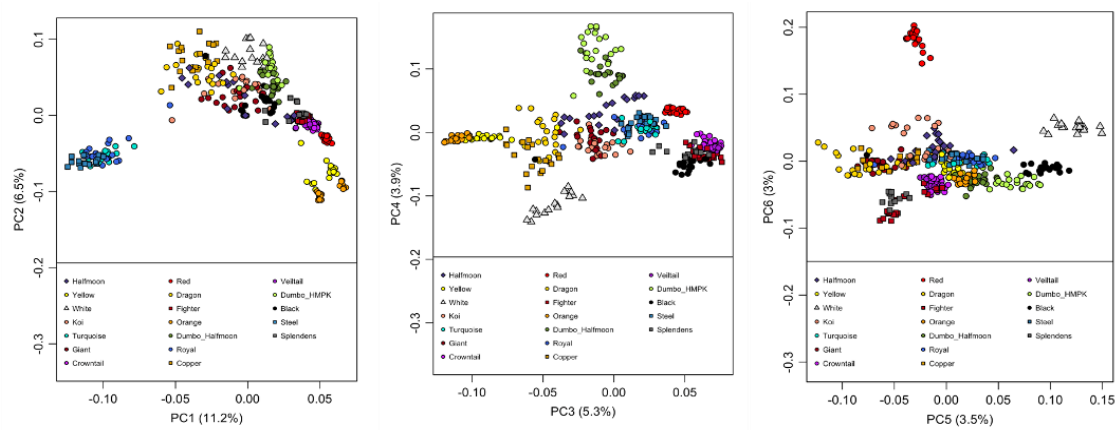

Figure S8. PCA of domesticated *Betta splendens* samples.

Principal components 1 through 6 are shown.

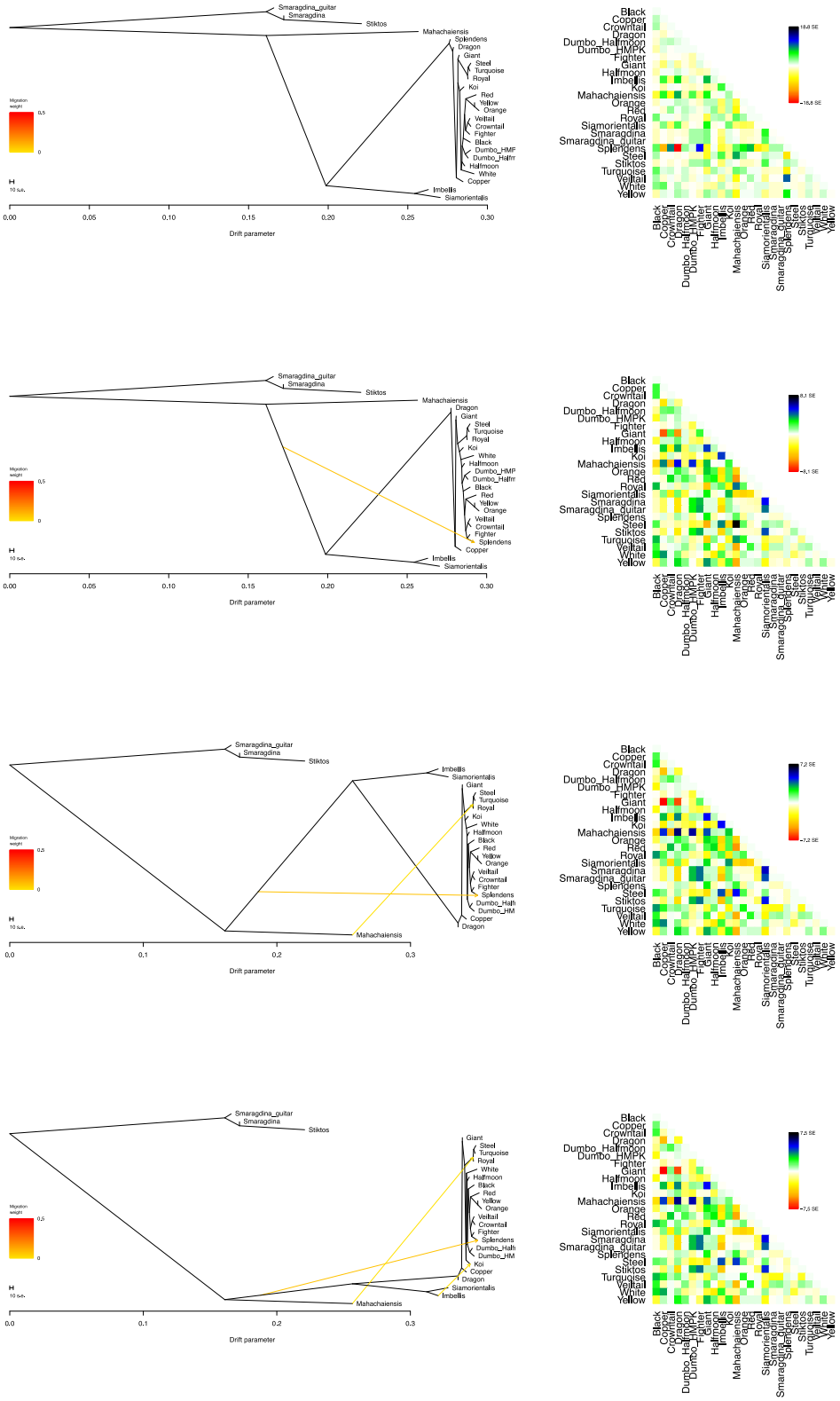

Figure S9. TreeMix analysis of Betta species populations with 0 to 3 migration edges.

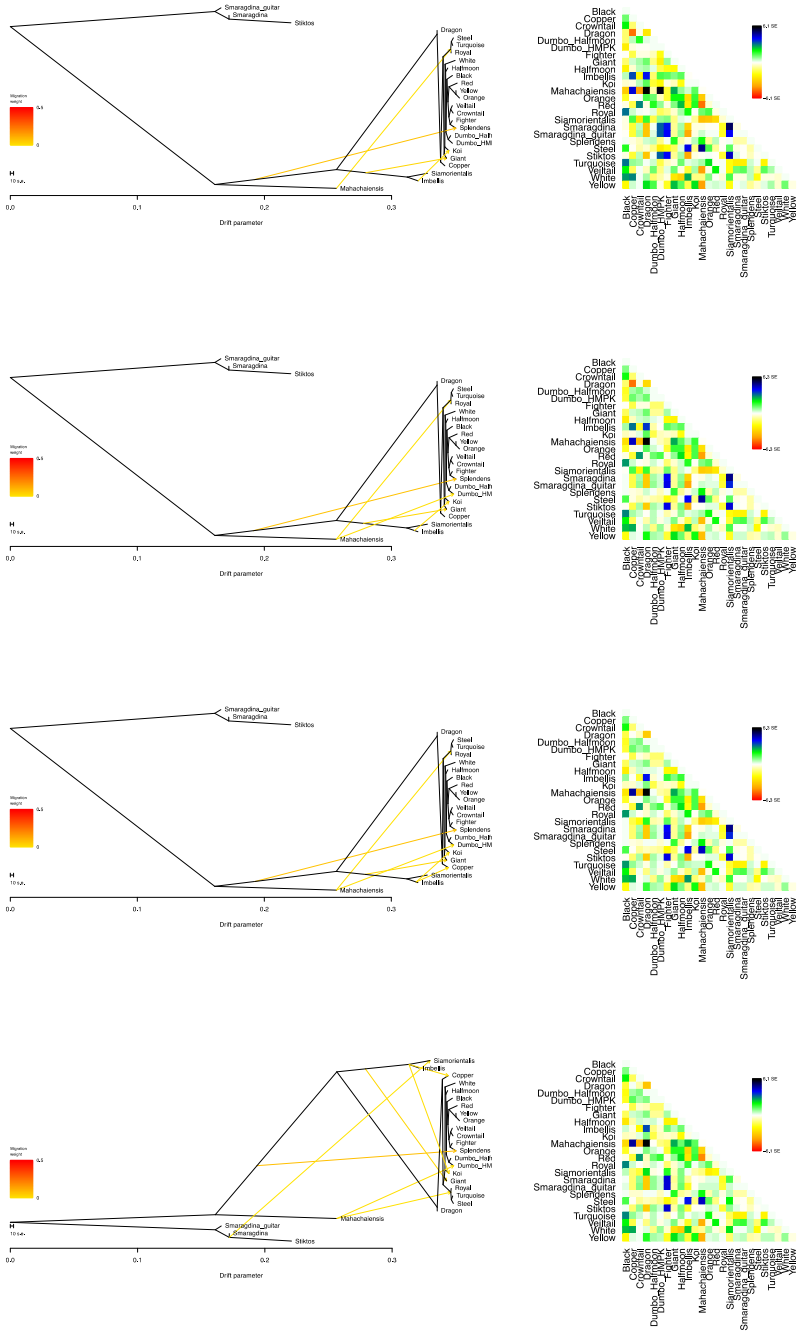

Figure S10. TreeMix analysis of Betta species populations with 4 to 7 migration edges.

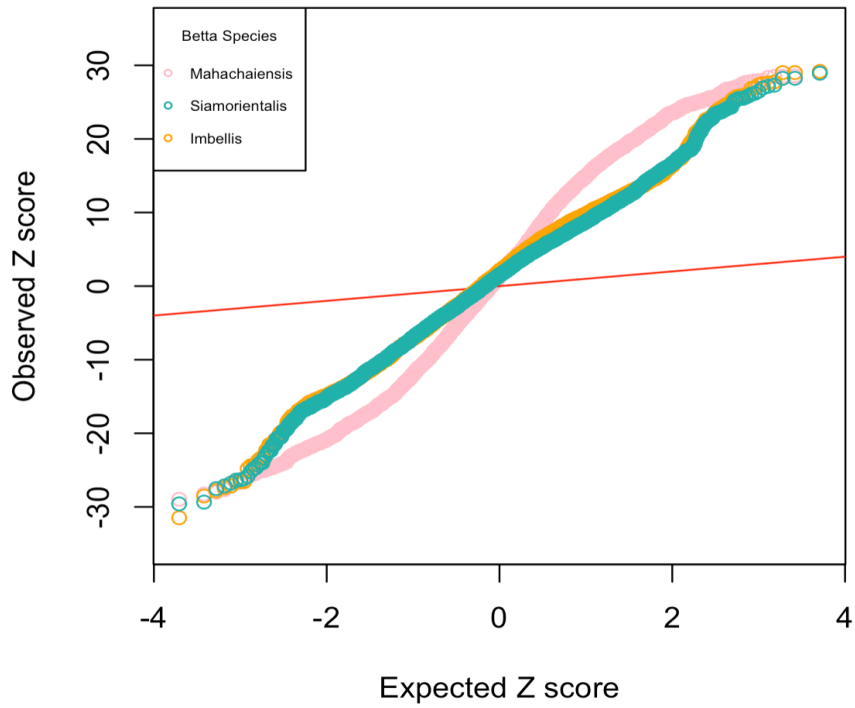

Figure S11. Testing all combinations of D(P1, P2; P3, Outgroup).

P1 and P2 are two individuals from the *Betta splendens* complex and Z are individuals from the three wild Betta species. Outgroups are *B. smaragdina* individuals.

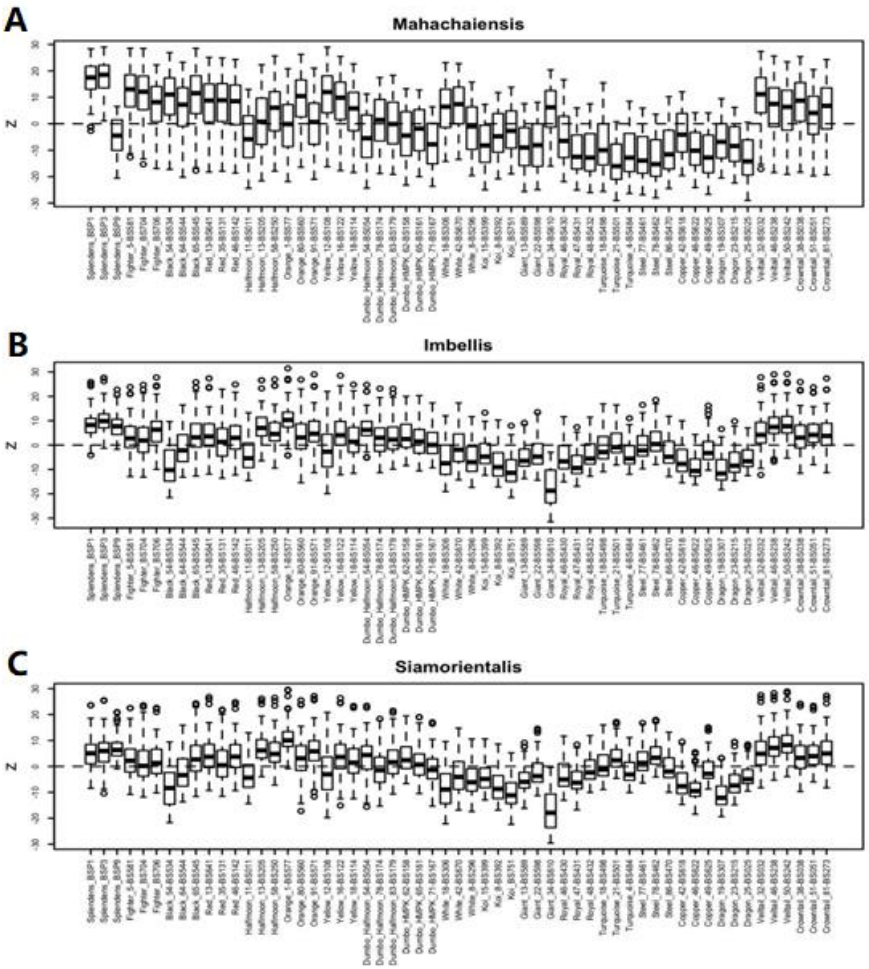

Figure S12. D-stats in the form of D(P1, P2; Mahachaiensis/Imbellis/Siamorientalis, Outgroup).

In each panel (**A**, **B** and **C**), along the x-axis, each of the boxplot shows the distribution of Z scores of all tests in the form of D (P, individual on X; Mahachaiensis/Imbellis/Siamorientalis, Outgroup) where P are all individuals from other domesticated breeds.

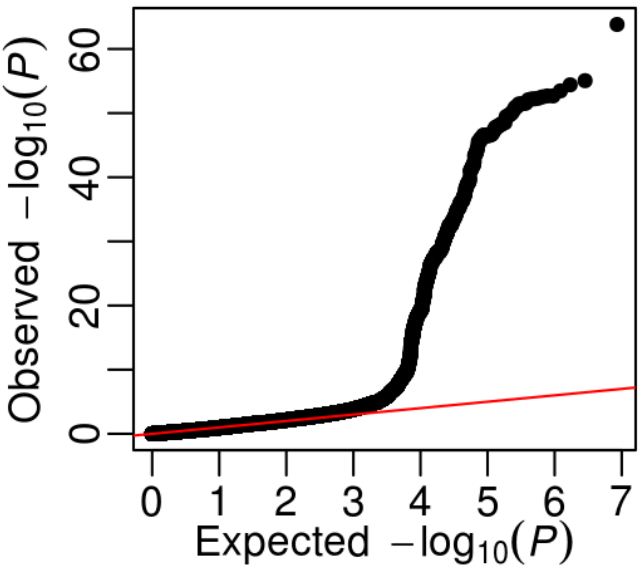

Figure S13. QQ plot of GWAS of sex in the Siamese fighting fish.

The inflation factor is 0.785.

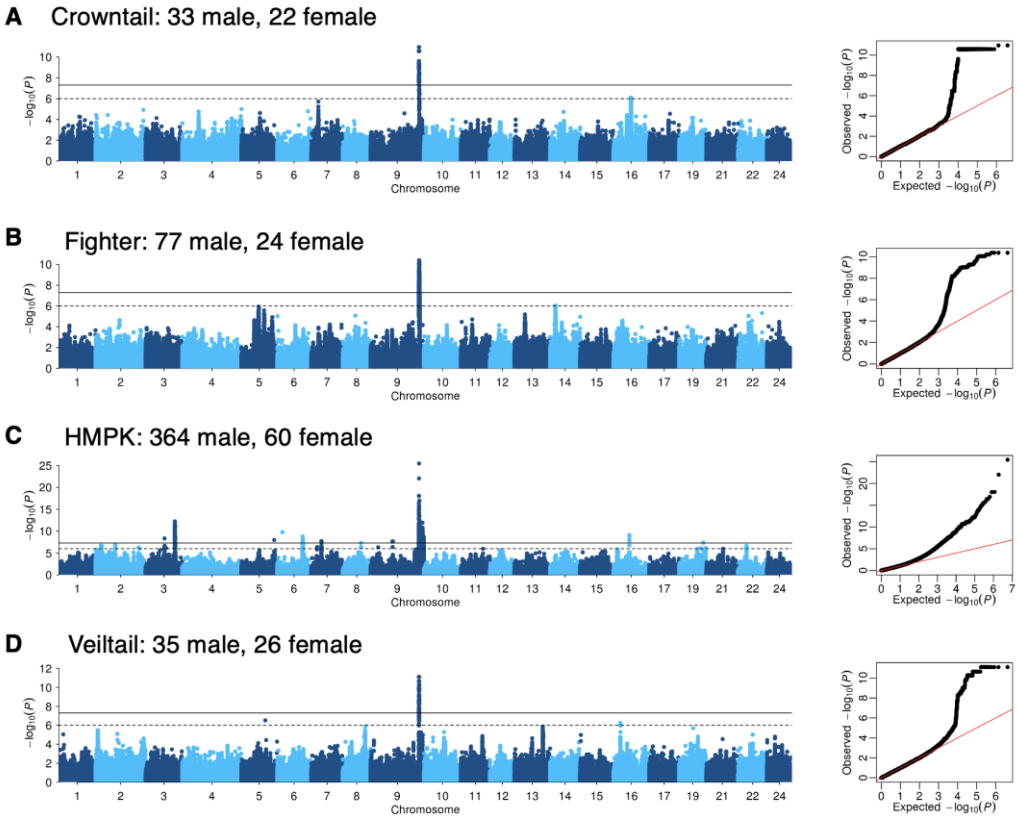

Figure S14. Manhattan and QQ plots of GWAS of Sex in different domesticated breeds.

(**A**) Crowntail, (**B**) Fighter, (**C**) HMPK and (**D**) Veiltail. Number of male and female are shown on the Manhattan plots.

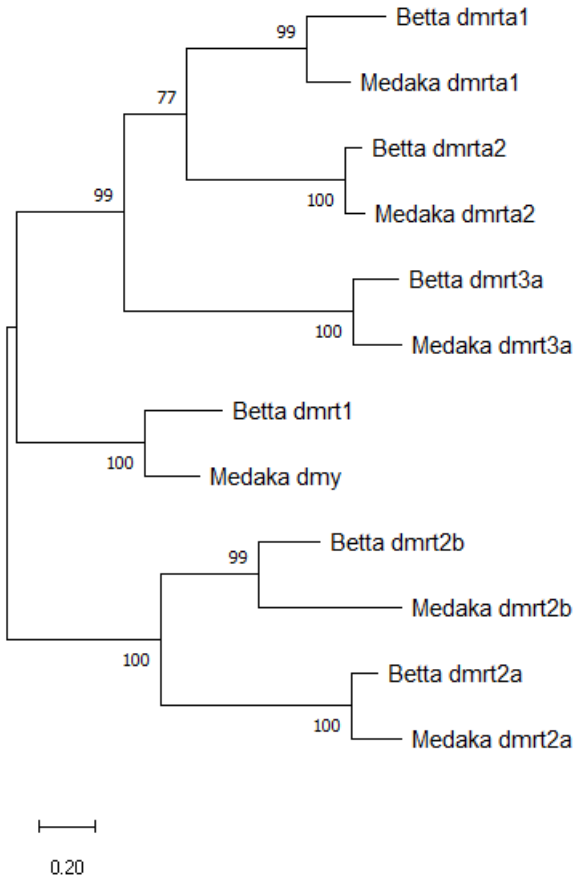

Figure S15. Maximum-likelihood phylogeny of *dmrt* homologs in Medaka and *Betta splendens*.

The bootstrap values are based on 1,000 replicates.

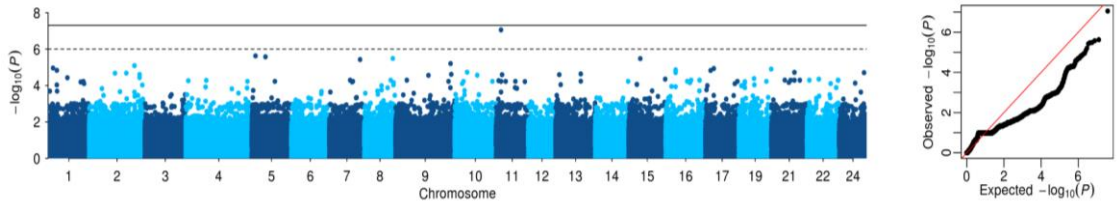

Figure S16. Manhattan and QQ plot for GWAS of Sex in wild individuals.

The analyses were based on 29 wild individuals of *B. imbellis*, *B. siamorientalis*, *B. mahachaiensis*, *B. smaragdina*, *B. smaragdina* guitar and *B. stiktos*, which were phenotyped for sex.

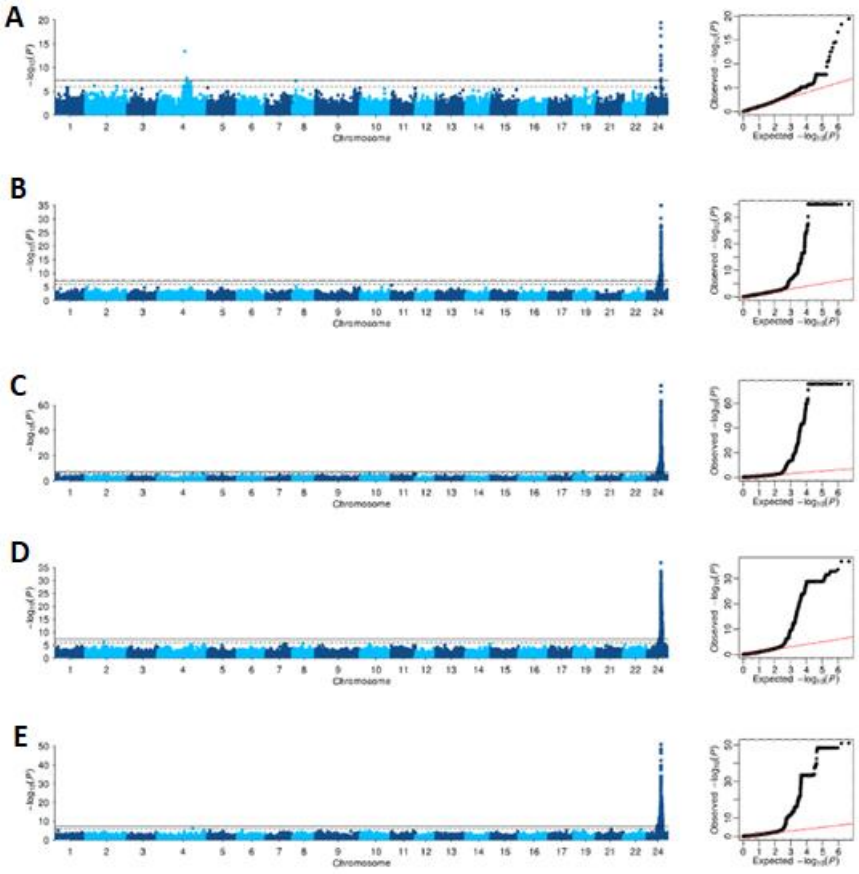

Figure S17. Manhattan and QQ plots of GWAS for the Turquoise-green, Royal-blue and Steel-blue breeds.

(**A**) Quantitative GWAS with coding Turquoise-green as 1, Royal-blue as 2 and Steel-blue as 3. (**B**) Case-control GWAS of Royal-blue vs. Steel-blue. (**C**) Case-control of Royal-blue vs. (Turquoise-green+Steel-blue). (**D**) Case-control GWAS of Turquoise-green vs. (Steel-blue+Royal-blue). (**E**) Turquoise-green vs. Steel-blue.

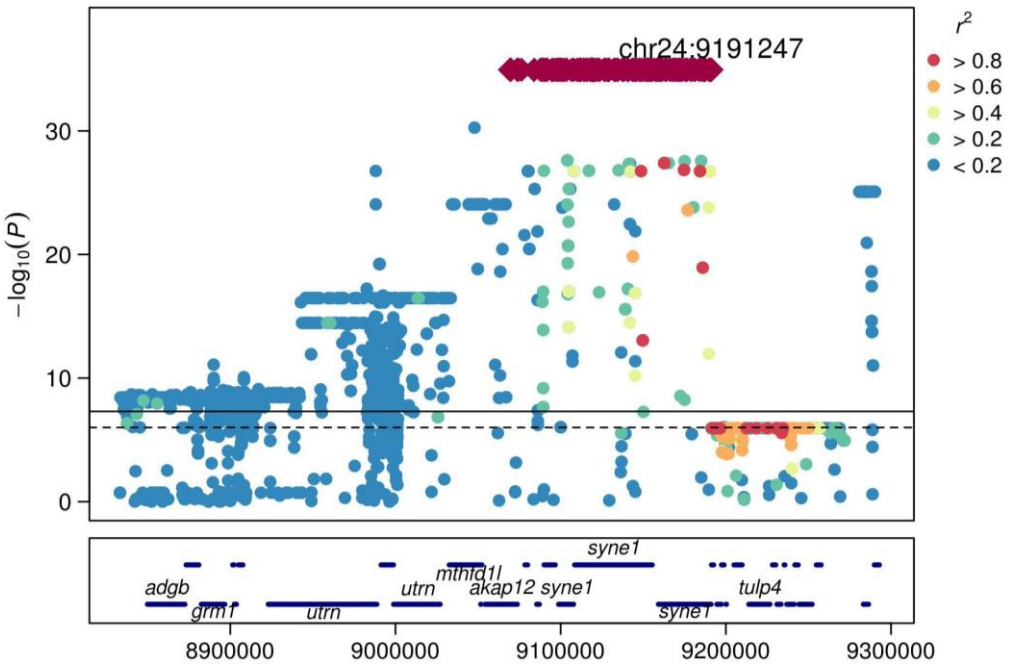

Figure S18. LocusZoom plot for locus on chromosome 24 associated with the Royal-blue, Turquoise-green and Steel-blue variation.

The solid and dashed lines in the Manhattan plots represent genome-wide (5×10^{-8}) and suggestive (1×10^{-6}) significance thresholds, respectively. Colors of dots represent linkage disequilibrium with the lead variant, indicated by a diamond symbol. Bottom box shows the location of annotated genes in this region.

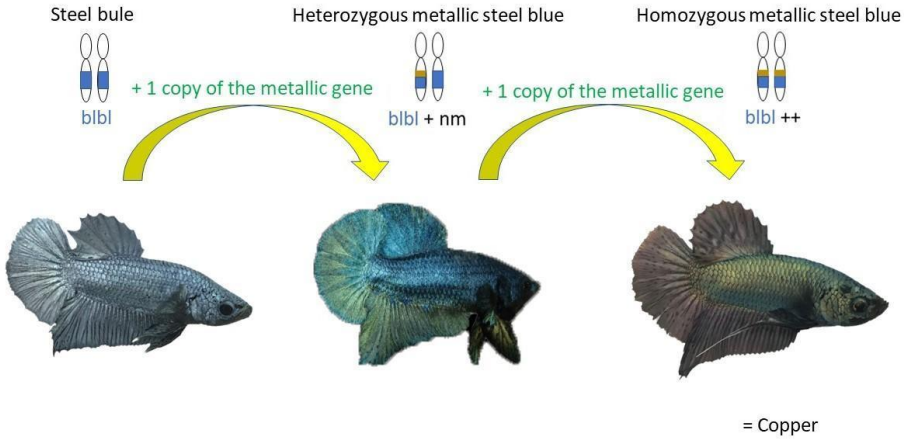

Figure S19. Breeding of Copper from Steel-blue.

The Copper phenotype is derived from two copies of presumed “metallic” genes on a Steel-blue genetic background (*24*).

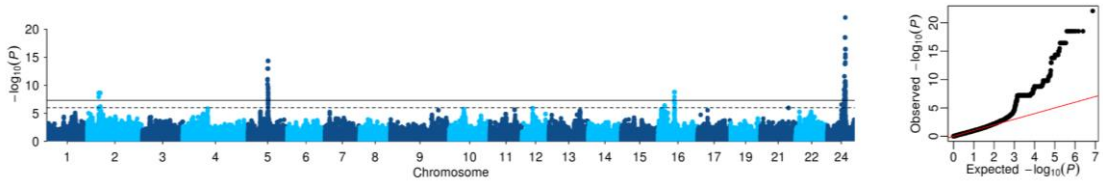

Figure S20. Manhattan and QQ plot of GWAS for Steel-blue vs. Copper.

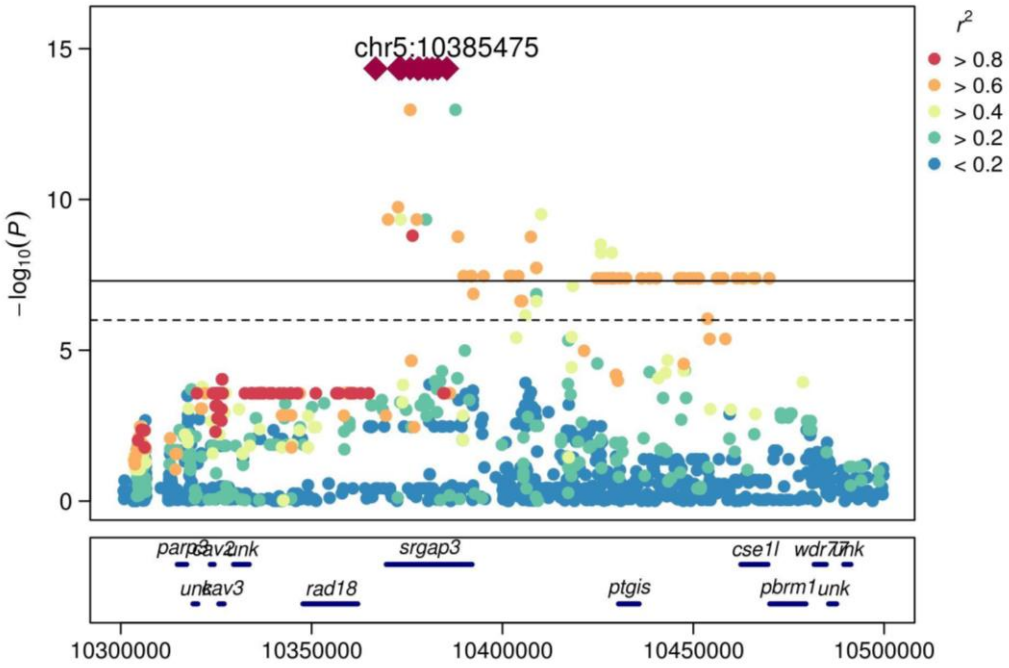

Figure S21. LocusZoom plot for the locus on chromosome 5 in case-control GWAS of Steel-blue vs. Copper.

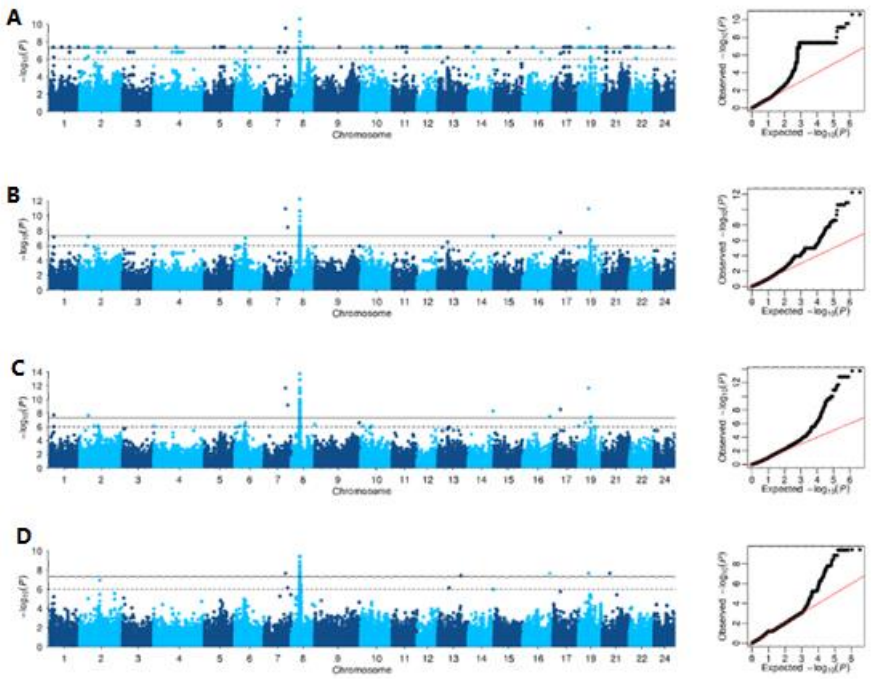

Figure S22. Manhattan and QQ plots of GWAS for the Red, Orange and Yellow breeds.

(**A**) GWAS coding Red as 1, Orange as 2 and Yellow as 3. (**B**) Case-control GWAS of Yellow vs. (Red+Orange). (**C**) Case-control of Orange vs. (Red+Yellow). (**D**) Case-control GWAS of Orange vs. Yellow.

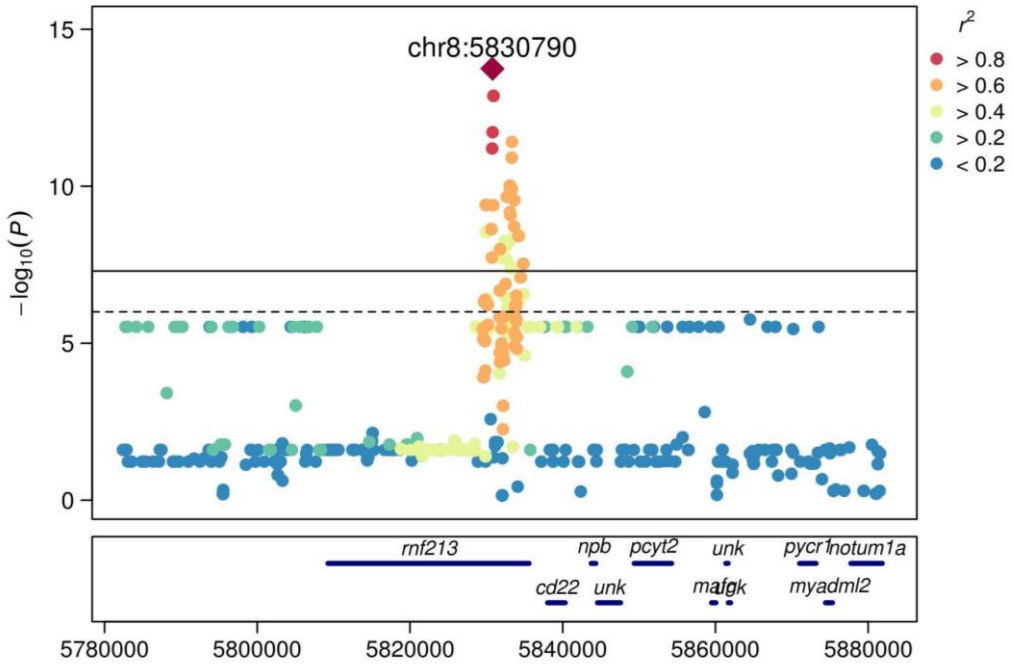

Figure S23. LocusZoom plot for locus on chromosome 8 by case-control GWAS of Orange vs. Red and Yellow.

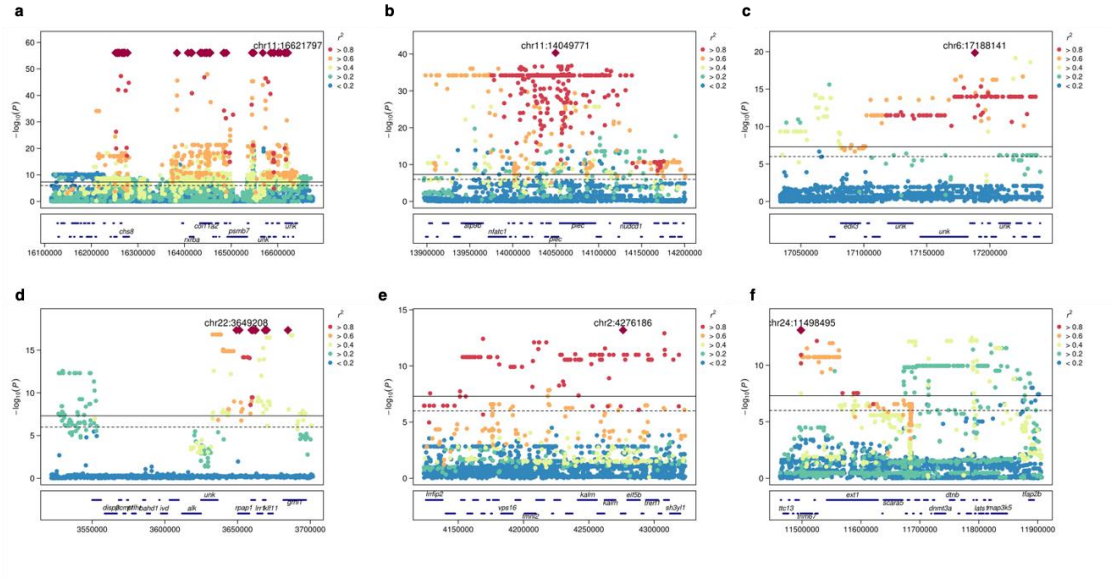

Figure S24. LocusZoom plots for loci that showed significant associations with the Mosaic color phenotype.

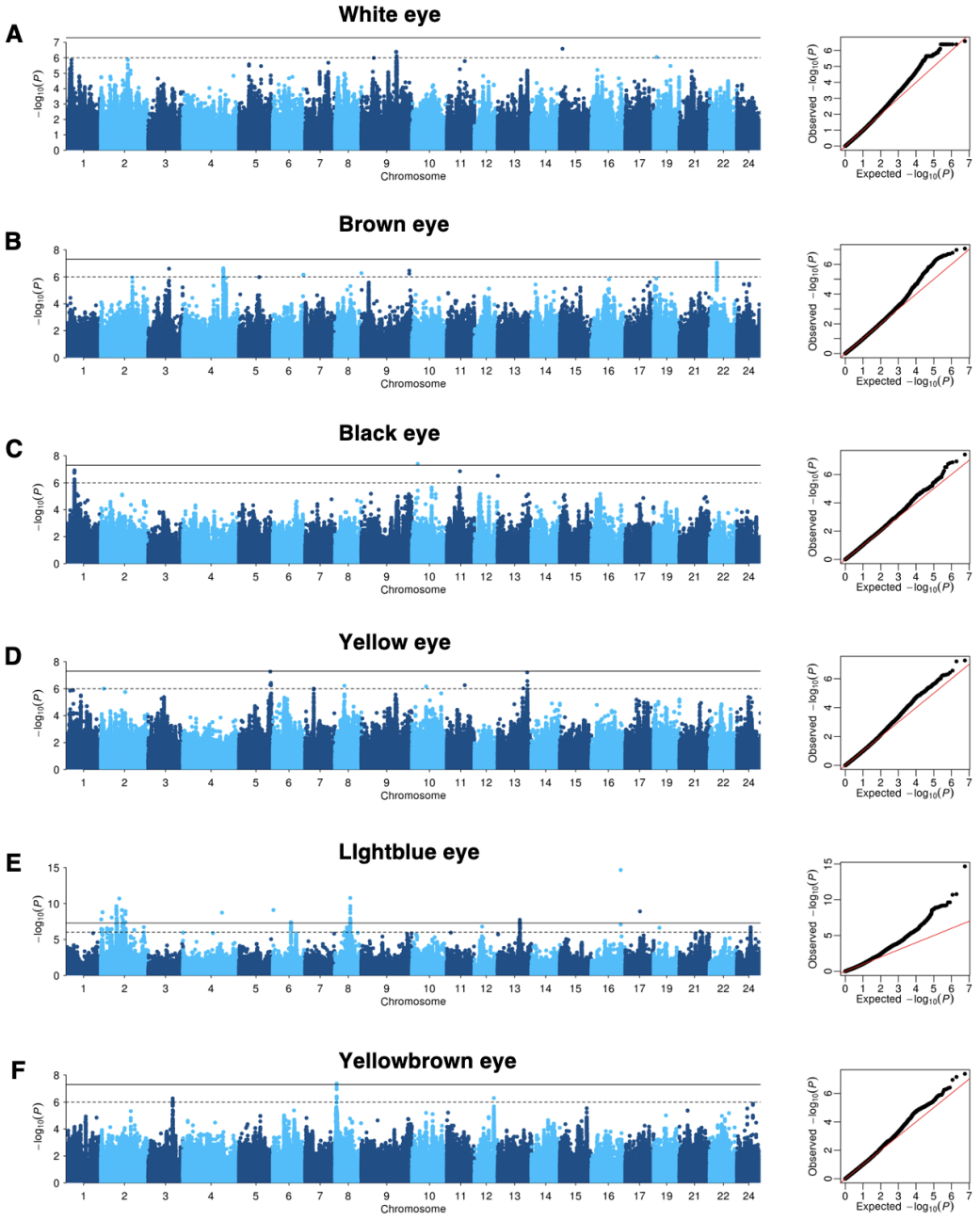

Figure S25. Manhattan and QQ plots of GWAS of eye colors.

The GWAS was performed by setting individuals of each color category as cases iteratively and all individuals with other colors as controls. (**A**) white eye color (n = 196); (**B**) brown eye color (n = 104); (**C**) black eye color (n = 102); (**D**) yellow eye color (n = 112); (**E**) light blue eye color (n = 15) and (**F**) yellow brown eye color (n = 77).

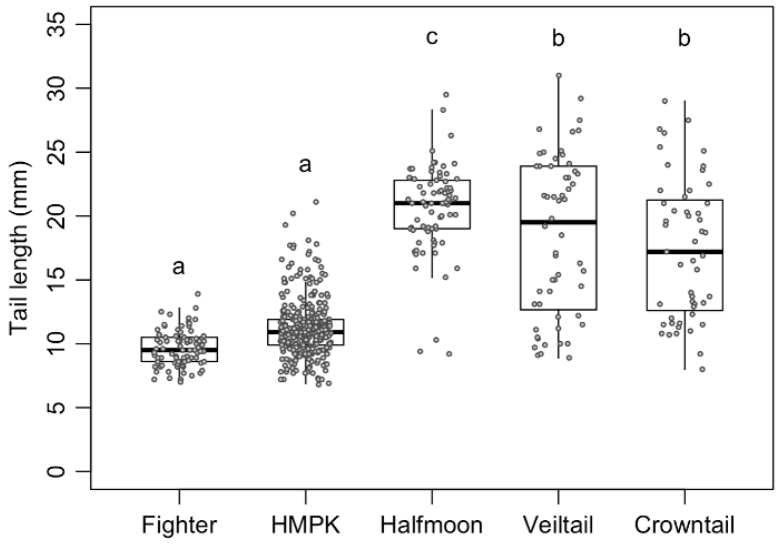

Figure S26. Caudal fin length in five breeds of the Siamese fighting fish.

Caudal fins of the long-finned breeds (Crowntail, Veiltail and Halfmoon) are significantly longer than those of Fighter and HMPK breeds. Different letters indicate significant differences (P.adj < 0.003 for all comparisons, one-way ANOVA, Tukey's HSD test).

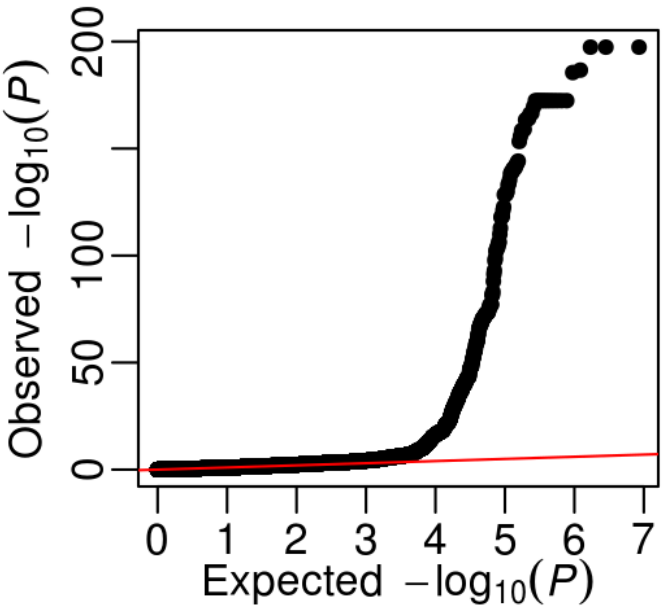

Figure S27. QQ plot for the case-control GWAS of short and long fin.

This is based on the results of **Fig. 3A**. The inflation factor is 0.853.

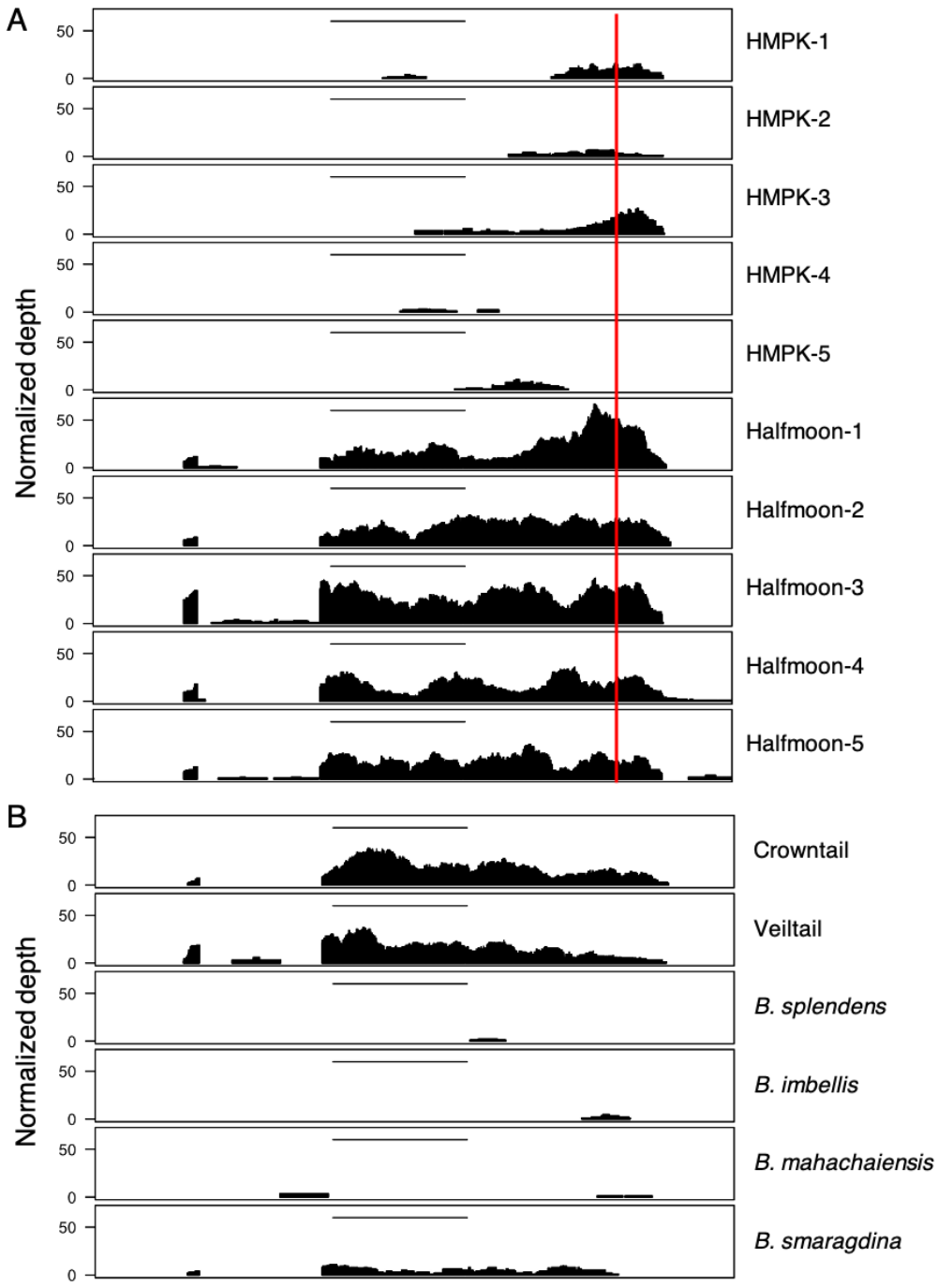

Figure S28. RNA-Seq reads mapped to the *kcnj15* gene region.

(**A**) comparison in short-fin (HMPK) and long-fin (Halfmoon) fishes, the red vertical line indicates the position of the GWAS lead SNP. (**B**) normalized depth of RNA-Seq reads mapped to the *kcnj15* gene region in Crowntail, Veiltail and wild Betta splendens complex. All the depths are normalized by a coefficient that accounts for the number of reads sequenced for each library. Horizontal black line shows the coding sequence of *kcnj15* in the draft genome annotation.

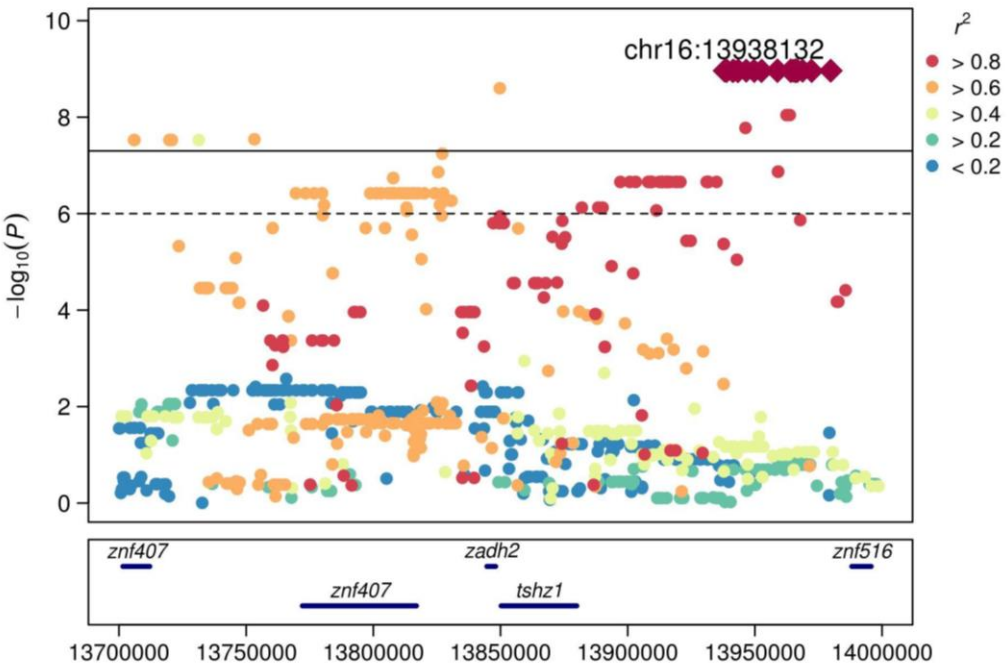

Figure S29. LocusZoom plot of the locus in GWAS between Halfmoon and Veiltail.

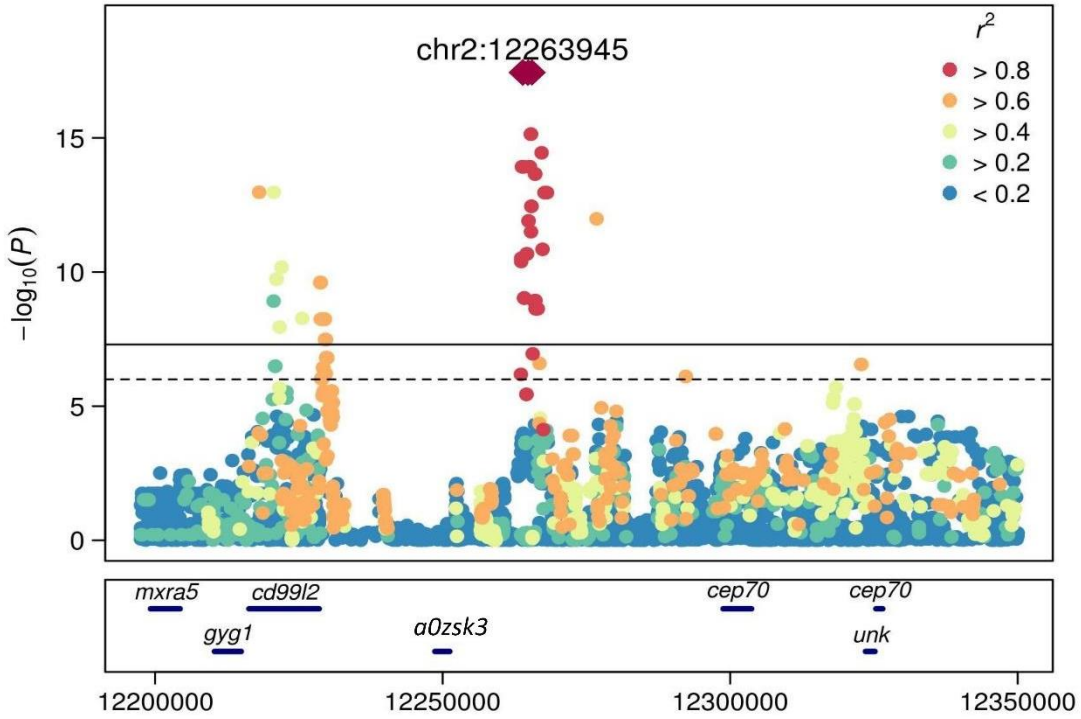

Figure S30. LocusZoom plot of the locus in GWAS between Crowntail and Halfmoon.

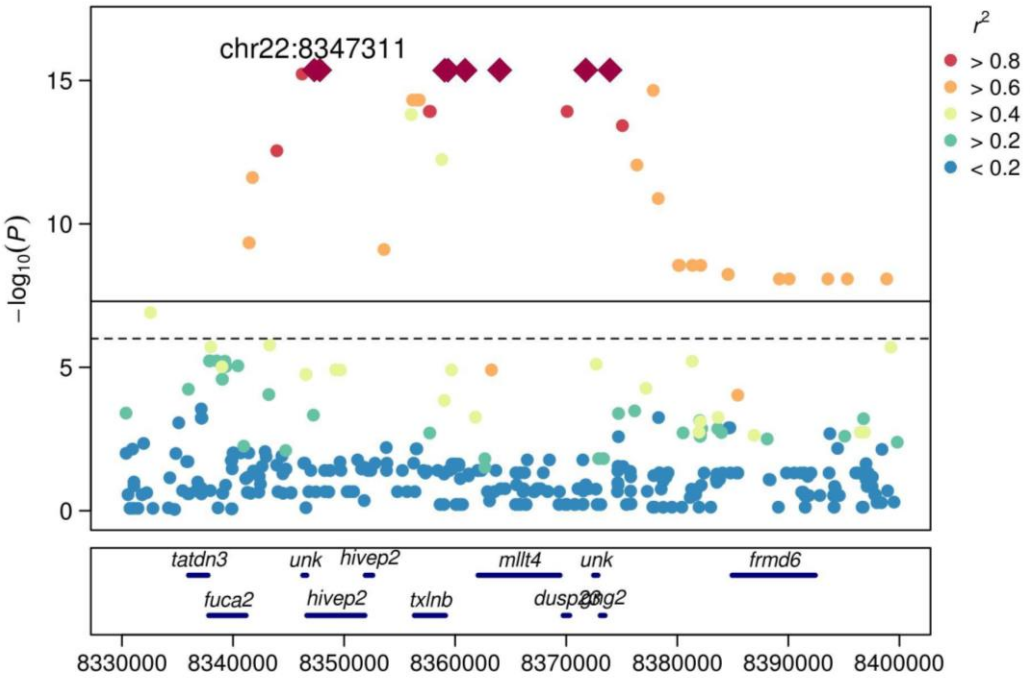

Figure S31. LocusZoom plot of the locus in GWAS between Crowntail and Veiltail.

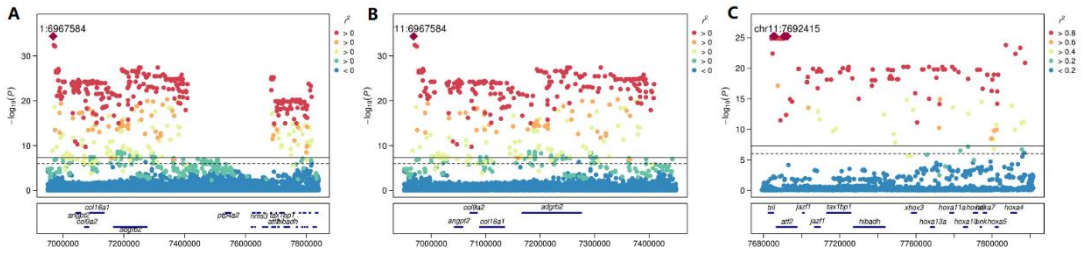

Figure S32. LocusZoom plots of the Dumbo loci on chromosome 11.

(**A**) 800 kb region containing associated SNPs. (**B**) Zoom into the first half of plot A. (**C**) Zoom into the second part of plot A, containing the *hoxa* gene cluster.

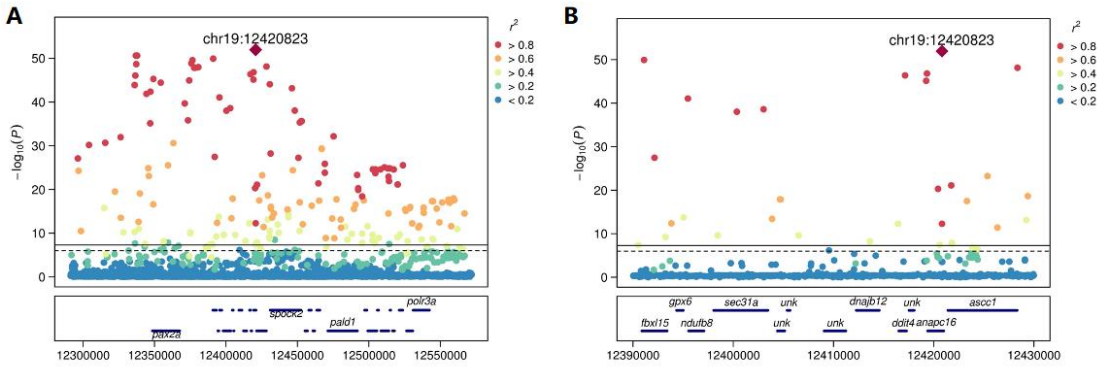

Figure S33. Regional plots of the Dumbo locus on chromosome 19.

(**A**) 200-kb region containing SNPs strongly associated with the Dumbo phenotype. (**B**) Zoom into the region containing the lead SNPs.

Gene_name	FPKM_Pectoral_fins_Dumbo_HMPK	FPKM_Pectoral_fins_HMPK	log2FoldChange	Padj
Kcnh8_1	0.991664409	0	2.243785332	1
Kcnh8_2	0.422032668	0	0.980482564	1
Evx1	58.49828164	60.96740196	-0.064559954	0.887202391
Col16a1	949.5730747	500.4649942	0.923114638	0.214906958
Hoxa3a	18.07948281	10.59606287	0.7615032	0.681277353
Hoxa4	24.24635614	31.25977298	-0.371613547	0.801821016
Hoxa5	0.594998645	0.604161546	-0.098310127	1
Hoxa7	5.902276205	3.480277658	0.736768436	0.698367986
Hoxa9	29.63767345	32.10366475	-0.118055324	0.931610201
Hoxa10	48.29335684	66.45856483	-0.450692049	0.561503533
Hoxa11a	44.97582757	31.60816499	0.524278058	0.495570478
Hoxa13a	960.6635627	766.89889	0.325523846	0.345766168
Fbxl15	42.2225873	20.95027251	1.024741706	0.156198473

Figure S34. Expression profile of genes in the locus associated with the Dumbo phenotype in the pectoral fins of Dumbo and non-Dumbo phenotypes.

The numbers are shaded in blue-to-red color gradients corresponding to large-to-small numbers to facilitate visualization.

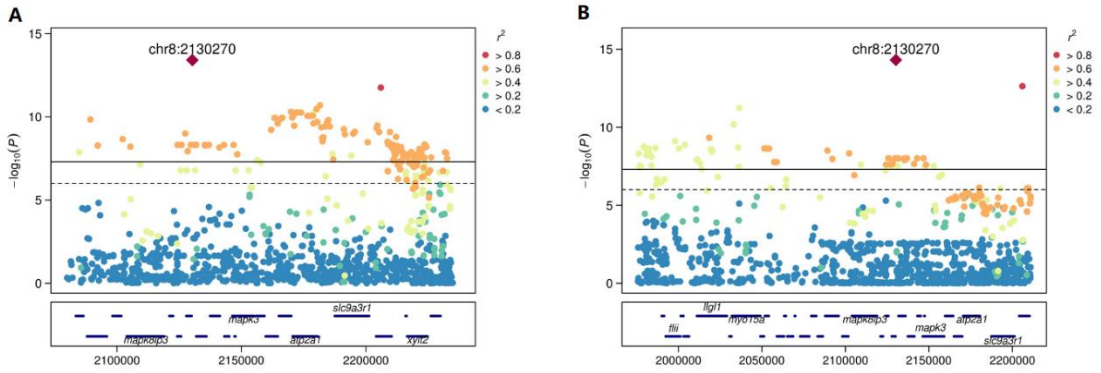

Figure S35. LocusZoom plot for the chromosome 8 peak in the (A) GWAS of standard-length and (B) Giant case-control GWAS.

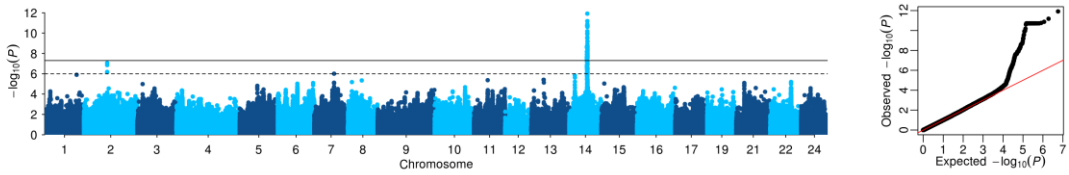

Figure S36. GWAS of standard length in non-Giant population.

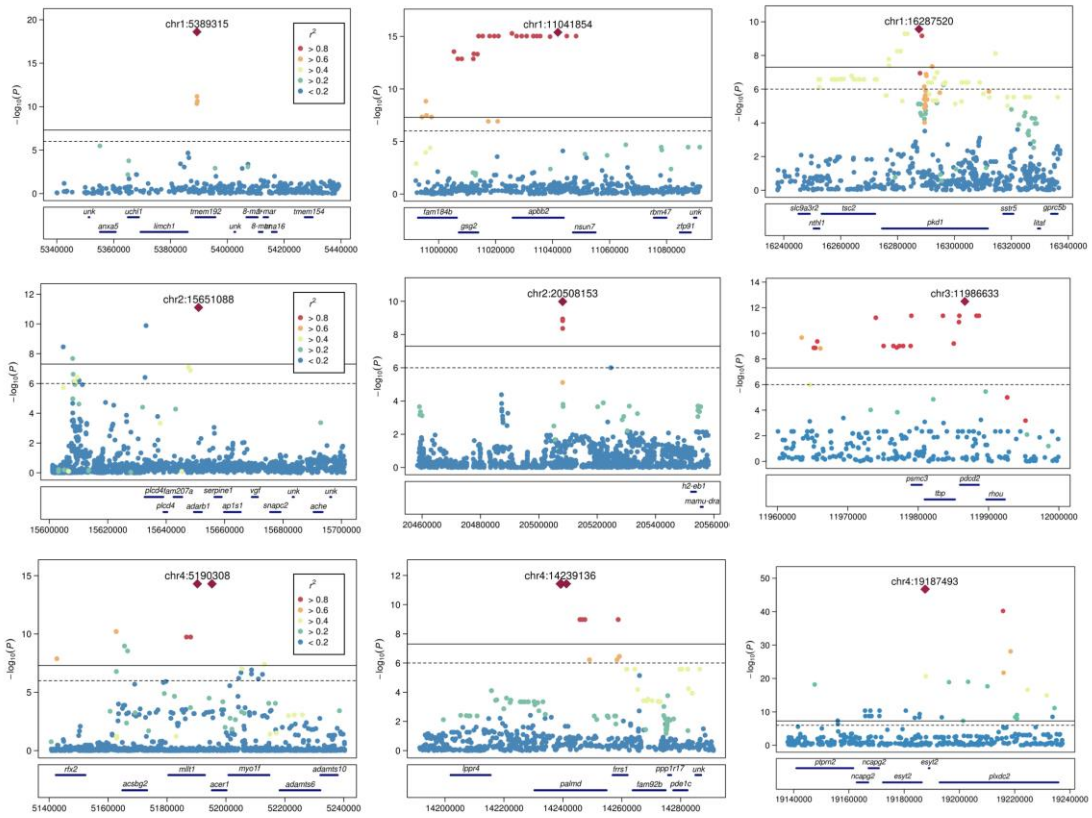

Figure S37. LocusZoom plot for peaks on chromosome 1 to 4 significantly associated with the Fighter breed.

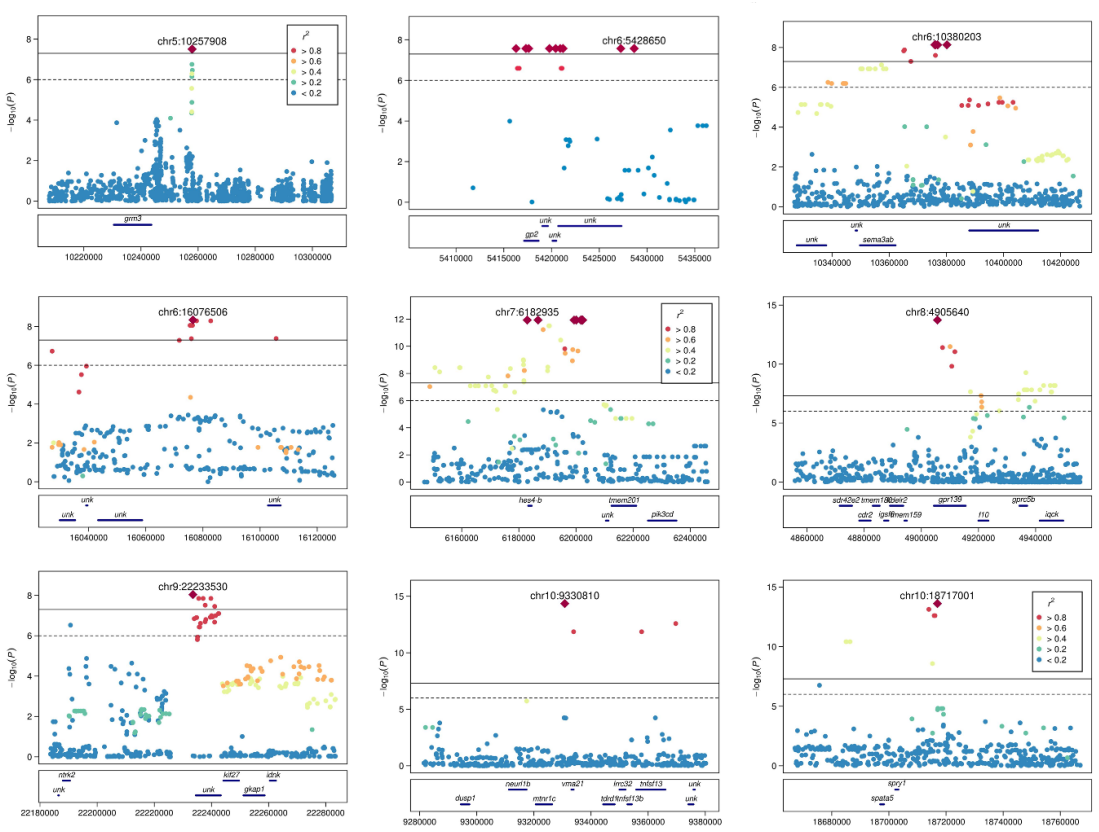

Figure S38. LocusZoom plot for peaks on chromosome 5 to 10 significantly associated with the Fighter breed.

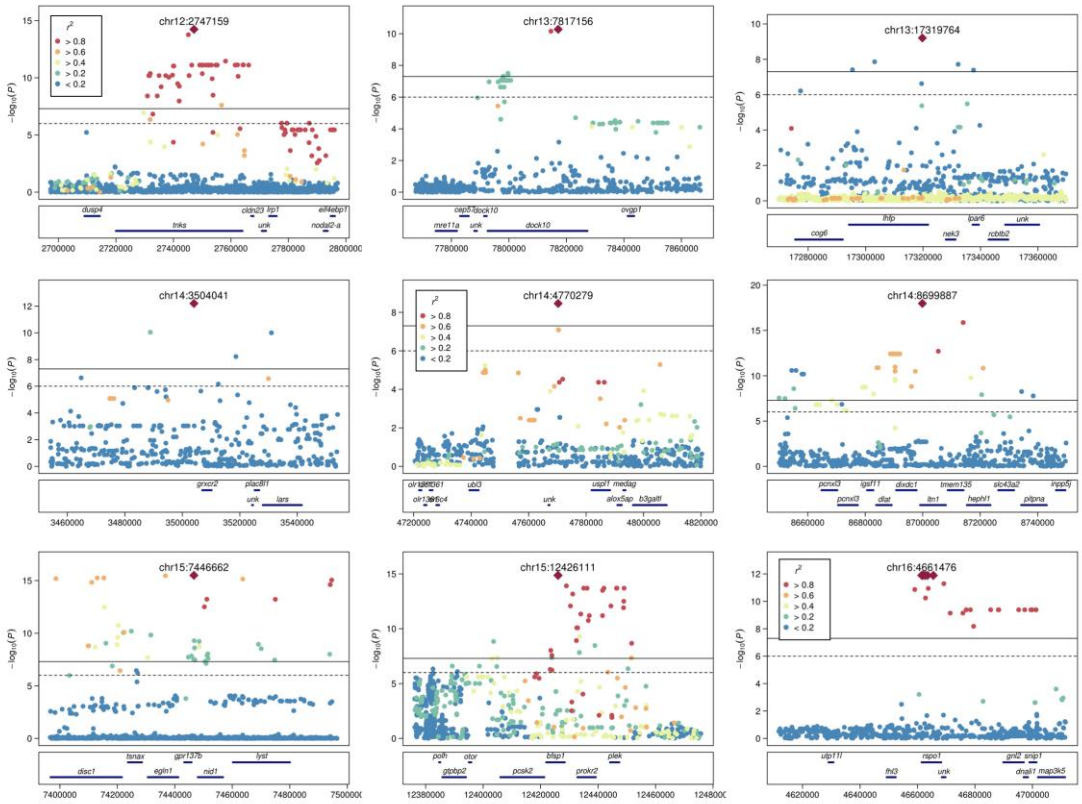

Figure S39. LocusZoom plot for peaks on chromosome 12 to 16 significantly associated with the Fighter breed.

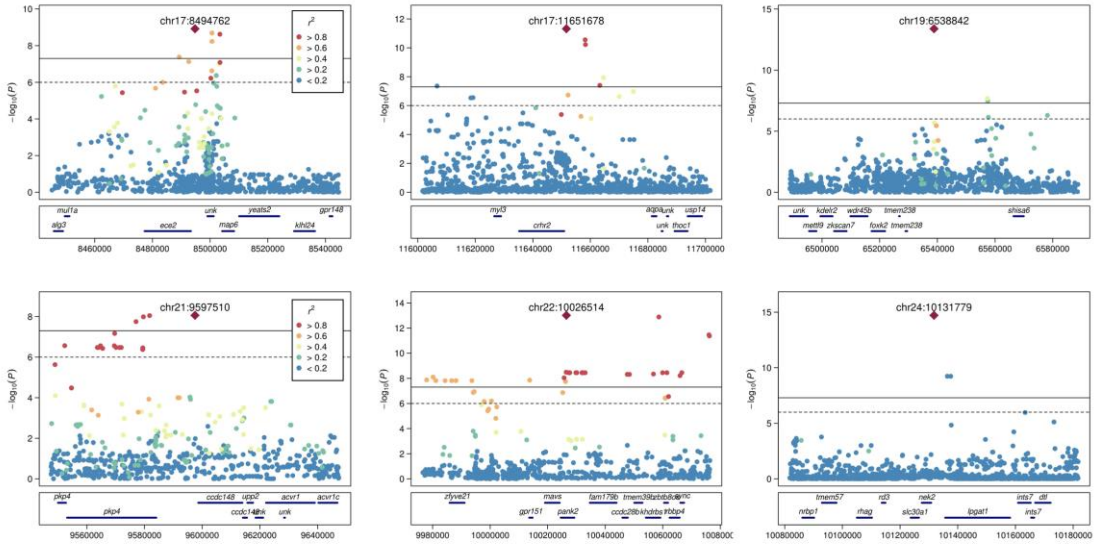

Figure S40. LocusZoom plot for peaks on chromosome 17 to 24 significantly associated with the Fighter breed.

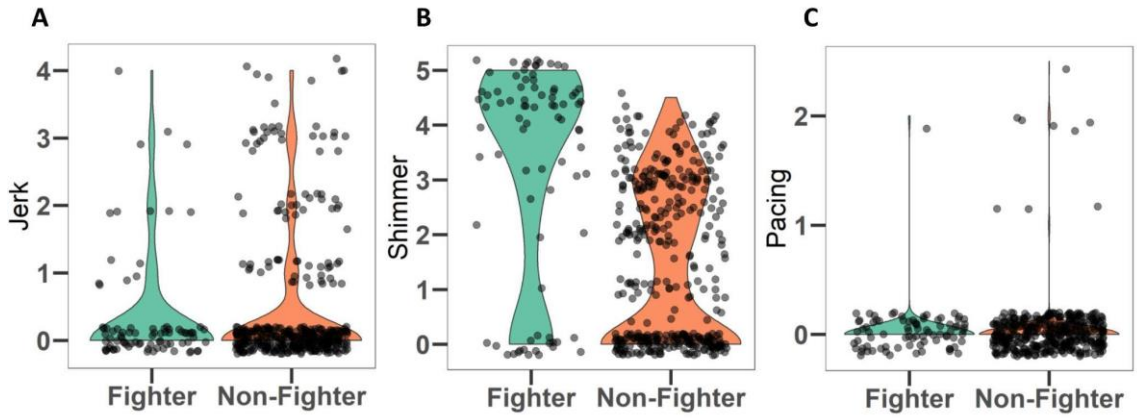

Figure S41. Boxplot for aggression related phenotypes composing the aggression index, showing their distributions in Fighter and non-Fighter breeds.

Other indices of aggression are shown in **Fig. 5**.

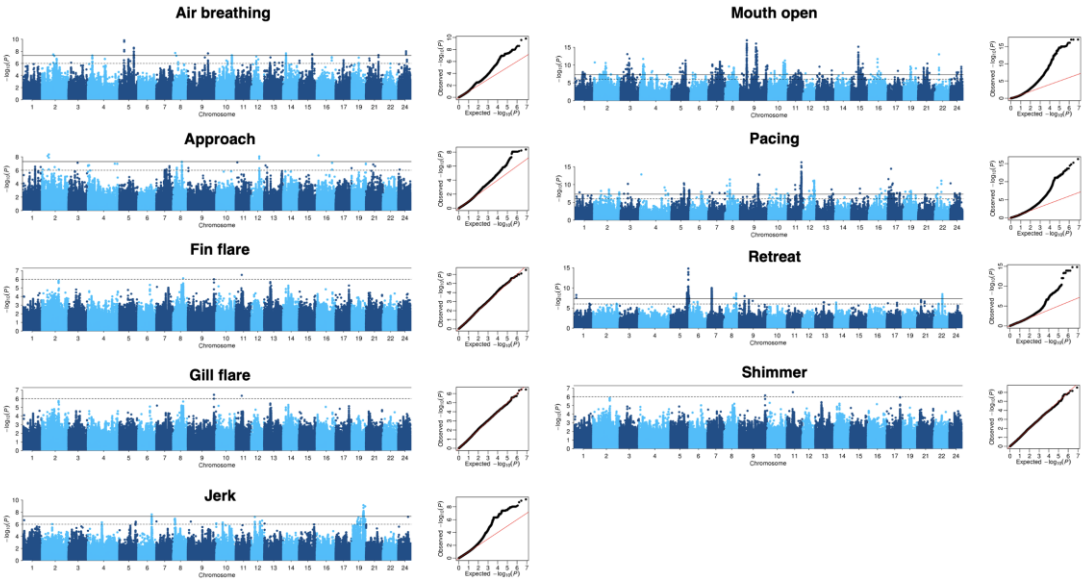

Figure S42. Manhattan and QQ plots for the Aggression index and aggression behaviors in simulative fighting.

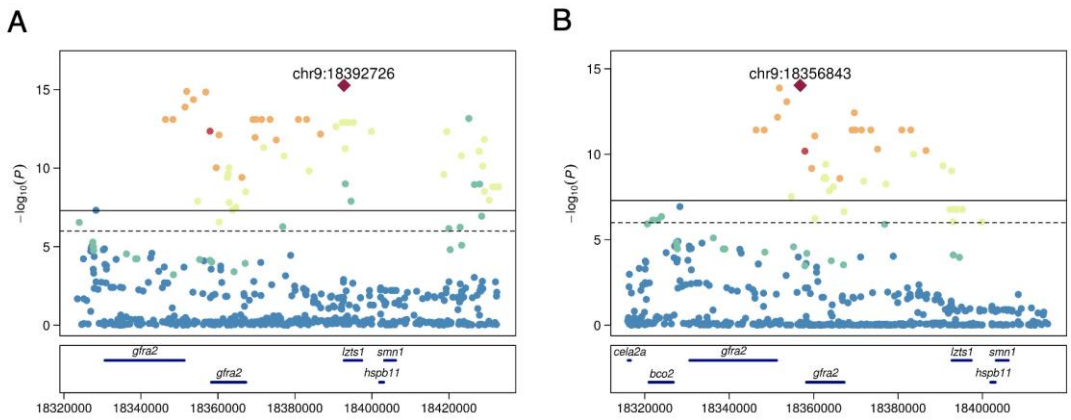

Figure S43. LocusZoom plots for the shared GWAS signal in Charge (A) and Mouth open (B).

Supplementary Tables

Table S1. Sequencing platforms and data output

Platform	System	Insertion (bp)	Read length (bp)	Data (Gb)	Depth
PacBio	SMRTbell	20 kb	NA	53.27	109.18×
Illumina	HiSeq X Ten	350	150	53.55	109.76×
10X genomics	HiSeq X Ten	350	150	114.98	235.68×
BioNano	NA	NA	NA	155.66	319.05×
Hi-C	HiSeq X Ten	350	150	52.58	107.77×

Table S2. Repeat annotation of the Siamese fighting fish genome

Type	Repeat Size (bp)	% of genome
Trf	34,495,406	7.64
Repeatmasker	111,386,726	24.68
Proteinmask	21,751,452	4.82
Total*	119,389,200	26.45

* There are overlaps in annotated sequence by different tools.

Table S3. Transposable element annotation of the Siamese fighting fish genome

Category	Denovo+Repbase Length(bp)	% in genome	TE proteins Length(bp)	% in genome	Combined TEs Length(bp)	% in genome
DNA	26,631,246	5.9	1,739,683	0.39	27,825,348	6.17
LINE	36,099,475	8	14,184,228	3.14	39,930,260	8.85
SINE	3,488,471	0.77	0	0	3,488,471	0.77
LTR	34,288,688	7.6	5,868,612	1.3	35,407,679	7.85
Simple repeat	19,853,701	4.4	0	0	19,853,701	4.4
Unknown	2,152,019	0.48	0	0	2,152,019	0.48
Total	111,386,726	24.68	21,751,452	4.82	113,804,129	25.22

Table S4. Gene annotation of the Siamese fighting fish genome

	Gene set	Number	Average transcript length (bp)	Average CDS length (bp)	Average exons per gene	Average exon length (bp)	Average intron length (bp)
De novo	Augustus	24,391	6,165.89	1,448.60	8.23	176.11	652.86
	GlimmerHMM	116,146	3,153.40	606.86	3.27	185.51	1,121.14
	SNAP	45,506	4,805.18	914.46	5.73	159.54	822.22
	Geneid	25,185	11,273.55	1,357.08	8.36	162.41	1,348.09
	Genscan	25,854	12,073.09	1,744.03	9.65	180.82	1,194.78
	Cca	22,661	5,271.45	1,300.62	6.83	190.44	681.14
	Cse	20,221	7,410.23	1,633.96	8.82	185.36	739.10
	Dre	20,403	6,869.51	1,560.61	8.30	187.99	727.10
	Gac	23,876	5,723.22	1,329.39	7.41	179.29	684.96
	Ipu	22,203	6,522.39	1,465.08	7.79	188.15	745.17
	Loc	20,327	6,875.17	1,474.38	8.02	183.88	769.55
Homolog	Ola	21,838	6,915.09	1,641.71	8.51	192.91	702.16
	Oni	21,398	7,214.29	1,656.79	8.81	188.14	711.93
	Pki	21,771	6,590.64	1,478.89	7.92	186.79	738.98
	Srh	21,101	6,682.92	1,513.33	8.03	188.55	735.77
	Ssa	24,121	5,977.32	1,427.03	7.40	192.74	710.57
	Tru	20,568	7,037.84	1,582.11	8.73	181.13	705.36
	Xma	21,285	7,431.05	1,689.68	8.91	189.67	725.98
RNAseq	PASA	183,242	5,808.30	1,149.68	7.11	161.63	762.08
	Cufflinks	70,695	13,208.26	3,448.98	10.88	317.01	987.81
EVM		28,224	7,098.17	1,403.07	8.15	172.09	796.16
Pasa-update		28,064	7,146.26	1,414.99	8.23	171.92	792.65
Final set		25,104	7,695.22	1,500.39	8.90	168.67	784.59

Table S5. Non-coding RNA annotation of the Siamese fighting fish genome

	Type	Copy(w*)	Average length(bp)	Total length(bp)	% of genome
miRNA		1,818	88.74	161,327	0.035748
tRNA		6,214	75.02	466,163	0.10
	rRNA	1,439	129.18	185,888	0.041190
	18S	21	537.81	11,294	0.002503
rRNA	28S	110	237.87	26,166	0.005798
	5.8S	7	157	1,099	0.000244
	5S	1,301	113.24	147,329	0.032646
	snRNA	498	127.66	63,573	0.014087
	CD-box	87	92.36	8,035	0.001780
snRNA	HACA-box	76	156.71	11,910	0.002639
	splicing	319	126.80	40,449	0.008963

Table S6. Statistics of functional annotation of structural genes in the Siamese fighting fish genome

	Number	Percent (%)
Total	25,104	-
Swissprot	21,132	84.20
Nr	13,459	53.60
KEGG	19,127	76.20
InterPro	21,506	85.70
GO	16,079	64.00
Pfam	19,521	77.80
Annotated	22,788	90.80
Unannotated	2,316	9.20

Table S7. Statistics of three Siamese fighting fish genome assemblies

	ASM365015v1[1] (male: XY)	fBetSpl5.3[2] (male: XY)	Bspl.v1.2018[3] (female: XX)	BettaSRF1.0[4] (female: XX)
Assembly Statistics				
Sequencing platforms (Depth)	Illumina (114×) Hi-C (75×)	PacBio (48×) Illumina (83×) 10X Genomics (183×) BioNano	ONT (50×) Illumina (114×) Hi-C (75×)	PacBio (109×) Illumina (109×) 10X Genomics (235×) BioNano (319×) Hi-C (107×)
Total sequence length (Mb)	456.232	441.388	440.784	451.294
Total ungapped length (Mb)	412.702	441.359	440.71	448.889
Total Gap length (Mb)	43.53 (9.54%)	0.029 (0.0065%)	-	2.405 (0.53%)
Number of scaffolds	12,333	70	109	318
Scaffold N50 (Mb)	19.754	20.129	20.67	19.63
Scaffold L50	10	9	9	9
Number of contigs	35,779	398	857	526
Contig N50 (bp)	38,053	2,497,747	2,171,468	4,077,672
Contig L50	3,417	48	62	-
GC content	40.04%	45.10%	45.10%	45.08%
Assembled pseudochromosomes	21	21	21	21
Genome annotation				
Gene number	23,981	-	21,535	25,104
Average transcript length (bp)	6,421.71	-	-	7,695.22
Average CDS length (bp)	1,626.98	-	-	1,500.39
Average exon per gene	8.9	-	-	8.9
Average exon length (bp)	182.65	-	-	168.67
Average intron length (bp)	606.37	-	-	784.59
BUSCO genome assessment				
Complete	95.4% (4,375/4,584)	-	96.9% (3,528/3,640)	97.5% (2,522/2,586)
Single-copy complete	92.3% (4,232/4,584)	-	96.1% (3,499/3,640)	96.1 (2,486/2,586)
Duplicated complete	3.1% (142/4,584)	-	0.8% (29/3,640)	1.4% (36/2,586)
Fragmented	2.8% (128/4,584)	-	0.3% (11/3,640)	1.4% (36/2,586)
Missing	1.8% (82/4,584)	-	2.8% (101/3,640)	1.1% (28/2,586)

Reference 1, 2 and 3 are assemblies generated by Fan et al (*111*), Rhie et al (*112*) and Prost et al (*113*). Reference 4 is the assembly built by this study. BUSCO for ASM365015v1 used Actinopterygii_odb9 as the reference lineage. BUSCO in this study used vertebrata_odb9 as the reference lineage. CEGMA assessment assembled 239 (238 complete and 1 partial) out of 248 CEGs (core eukaryotic genes) in the Siamese fighting fish lineage.

Table S8. Description of all samples used in this study

Breed	Abbreviation	Brief description	Num.	Illustration
Wild Siamese fighting fish		Wild Betta splendens	11	Fig. S4S
Betta imbellis		Wild Betta imbellis	10	Fig. S4T
Betta siamorientalis		Wild Betta siamorientalis	9	Fig. S4U
Betta mahachaiensis		Wild Betta mahachaiensis	10	Fig. S4V
Betta smaragdina		Wild Betta smaragdina	11	Fig. S4W
Betta smaragdina guitar		Wild Betta smaragdina guitar	4	Fig. S4X
Betta stiktos		Wild Betta stiktos	4	Fig. S4Y
Fighter Siamese fighting fish	Fighter	Short fin, selected for more aggressive performances over centuries	101	Fig. S4R
Veiltail Siamese fighting fish	Veiltail	Long fin, Veil-shaped caudal fin	61	Fig. S4Q
Crowntail Siamese fighting fish	Crowntail	Long fin, Crown-like appearance of the caudal fin, reduction in webbing/tissue between all fin rays	55	Fig. S4P
Halfmoon Siamese fighting fish	Halfmoon	Long fin, symmetrical D-shaped caudal fin	85	Fig. S4O
Dumbo Halfmoon Siamese fighting fish	Dumbo_Halfmoon	Long fin, symmetrical D-shaped caudal fin and enlarged pectoral fins	38	Fig. S4N
Halfmoon Plakat Siamese fighting fish	HMPK	Short fin, symmetrical D-shaped caudal fin	424	Fig. S4A-M
Solid Red Siamese fighting fish	Red	Solid red on the fish appearance	26	Fig. S4A
Solid Yellow Siamese fighting fish	Yellow	Solid yellow on the fish appearance	29	Fig. S4B
Solid Orange Siamese fighting fish	Orange	Solid orange on the fish appearance	28	Fig. S4C
Solid Turquoise green Siamese fighting fish	Turquoise-green	Solid Turquoise-green on the fish appearance	26	Fig. S4D
Solid Royal Blue Siamese fighting fish	Royal-blue	Solid Royal-blue on the fish appearance	28	Fig. S4E
Solid Steel Blue Siamese fighting fish	Steel-blue	Solid Steel-blue on the fish appearance	35	Fig. S4F
Solid Black Siamese fighting fish	Black	Solid Black on the fish appearance	20	Fig. S4G
Solid Opaque Siamese fighting fish	Opaque	Solid Opaque on the fish appearance	17	Fig. S4H
Copper Siamese fighting fish	Copper	Solid Copper (homozygous metallic steel blue) on the fish appearance	43	Fig. S4I
Dragon Siamese fighting fish	Dragon	Solid Dragon (black, red, orange) on the fish appearance	28	Fig. S4J
Giant Siamese fighting fish	Giant	Variation with a larger body size than normal ones; SL: male > 4.5 cm, female > 4 cm	102	Fig. S4K
Mosaic color Siamese fighting fish	Mosaic	Different color exhibits a mosaic pattern on the fish appearance like koi	55	Fig. S4L
Dumbo HMPK Siamese fighting fish	Dumbo_HMPK	Enlarged and extended paired pectoral fins	64	Fig. S4M

Table S9. Genotype distribution of the top SNP in Sex GWAS in all Betta fishes

Genotype	Domesticated		Wild						
	Male	Female	BSP	BIM	BSS	BMA	BSM	BST	BSG
	590	137	11 (3♂, 6♀, 2 NA)	10 (4♂, 6♀)	9 (4♂, 5♀)	10 (5♂, 5♀)	11 (6♂, 5♀)	4 (2♂, 2♀)	4 (2♂, 2♀)
GG	8	0	0	9	9	11	11	4	4
GA	530	7	3	0	0	0	0	0	0
AA	52	130	8	0	0	0	0	0	0

BSP: B. splendens, BIM: B. imbellis, BSS: B. siamorientalis, BMA: B. mahachaiensis, BSM: B. smaragdina, BST: B. stiktos, BSG: B. smaragdina guitar.

Table S10. Genotype distribution of the top SNP (chr24:9,191,247) in GWAS of Turquoise-green, Royal-blue and Steel-blue

Group	Royal-blue	Steel-blue	Turquoise-blue	Other
AA	1	35	0	291
GA	27	0	0	184
GG	0	0	26	163

Table S11. Genotype at the peak SNP on chromosome 5 of the case-control GWAS of Copper vs. Steel in different fish groups

Group	Copper	Steel	Other
CC	40	0	324
TC	3	0	210
TT	0	35	115

Table S12. Genotype distribution of the top SNP in the case-control GWAS of Orange vs. Red and Yellow.

Group	Orange	Red	yellow	Other
TA/T-	6	0	0	79
TA/TA	20	0	0	55
T-/T-	2	26	29	510

Table S13. Genotype frequency of the lead SNP (chr14:9,596,738) in GWAS of long fin in wild individuals of *B. splendens* complex

	BSP	BIM	BSI	BMA	BSM	BSG	BST
Total	11	10	9	10	11	4	4
GG	11	10	9	10	4	0	4
GA	0	0	0	0	5	3	0
AA	0	0	0	0	2	1	0

BSP: *Betta splendens*, BIM: *Betta imbellis*, BSI: *Betta siamorientalis*, BMA: *Betta mahachaiensis*, BSM: *Betta smaragdina*, BSG: *Betta smaragdina guitar*, BST: *Betta stiktos*.

Table S14. Annotated genes in the locus of GWAS of body size on chromosome 8

Gene	Annotation
tmem11	transmembrane protein 11--play a role in mitochondrial morphogenesis
dhrs7b	Dehydrogenase/Reductase 7B—function in steroid hormone regulation
cramp1l	Cramped chromatin regulator homolog 1 like– body height
hn1l	Hematological and Neurological Expressed 1-like
mapk8ip3	Mitogen-activated protein kinase 8 interacting protein—function as a regulator of vesicle transport, through interactions with the JNK-signaling components and motor proteins
nme3	Nucleoside Diphosphate kinase--Major role in the synthesis of nucleoside triphosphates other than ATP
mrps34	Mitochondrial Ribosomal Protein S34--plays a role in maintaining the stability of the small ribosomal subunit for mitoribosome formation.
spsb3	Sp1a/Ryanodine Receptor Domain and SOCS Box Containing 3
atp6v0c	encodes a component of vacuolar ATPase (V-ATPase), a multisubunit enzyme that mediates acidification of eukaryotic intracellular organelles.
ypel3	Yippee Like 3, involved in proliferation and apoptosis in myeloid precursor cells.
gdpd3	Glycerophosphodiester Phosphodiesterase Domain Containing 3
mapk3	a group of mitogen-activated protein kinases (MAPK) that mediate intracellular signaling, transduce signals from growth factors and phorbol esters.
ehd1	The protein encoded by this gene is thought to play a role in the endocytosis of IGF1 receptors.
rabep2	Plays a role in membrane trafficking and in homotypic early endosome fusion
atp2a1	encodes one of the SERCA Ca(2+)-ATPases, which are intracellular pumps located in the sarcoplasmic or endoplasmic reticula of muscle cells.
slc9a3r1	encodes a sodium/hydrogen exchanger regulatory cofactor.

Table S15. GWAS statistics for the lead SNPs of the 36 peaks associated with the Fighter breed

Chr	Position	MAF	Beta	Beta_SE	P_wald	Annotation
1	5389315	0.021	1.26E-01	1.36E-02	2.47E-19	*tmem192*
1	11041854	0.024	1.03E-01	1.23E-02	4.08E-16	*apbb2*
1	16287520	0.05	6.21E-02	9.70E-03	2.74E-10	*pkd1*
2	11481489	0.05	5.23E-02	8.63E-03	2.26E-09	intergenic
2	15651088	0.021	9.57E-02	1.38E-02	7.89E-12	*adarb1*; another closest gene, *surpine1*
2	20508153	0.022	6.85E-02	1.05E-02	1.05E-10	intergenic
3	11986633	0.028	9.36E-02	1.26E-02	3.20E-13	*pdcd2*
4	5190308	0.026	1.02E-01	1.28E-02	5.07E-15	*mllt1/acer1*
4	14241211	0.028	9.54E-02	1.35E-02	3.75E-12	*palmd*
4	19187735	0.028	1.95E-01	1.25E-02	1.83E-47	intergenic; closest gene, two copies of *esyt2*
5	10257908	0.021	5.84E-02	1.04E-02	3.12E-08	intergenic; closet gene, *grm3*
6	5416281	0.022	7.59E-02	1.35E-02	2.68E-08	intergenic
6	10376891	0.043	5.90E-02	1.01E-02	7.30E-09	intergenic
6	16076506	0.032	6.63E-02	1.12E-02	4.62E-09	intergenic
7	6196086	0.03	7.17E-02	1.10E-02	1.52E-10	intergenic
8	4905640	0.026	9.59E-02	1.23E-02	1.86E-14	*gpr139*
9	22233530	0.337	2.82E-02	4.86E-03	9.23E-09	intergenic
10	9330810	0.023	1.05E-01	1.31E-02	4.43E-15	intergenic
10	18717001	0.026	9.45E-02	1.21E-02	2.44E-14	intergenic
11	7182252	0.051	9.27E-02	1.47E-02	5.55E-10	*adgrb2*
12	2747159	0.092	1.02E-01	1.28E-02	5.80E-15	*tnks*
13	7817156	0.026	7.99E-02	1.20E-02	5.33E-11	*dock10*
13	17319764	0.07	6.88E-02	1.10E-02	6.25E-10	*lhfp*
14	3504041	0.03	8.97E-02	1.22E-02	6.30E-13	intergenic
14	4770279	0.067	5.28E-02	8.83E-03	3.47E-09	intergenic
14	8699887	0.033	9.80E-02	1.08E-02	1.05E-18	*ltn1*
15	7446662	0.026	1.03E-01	1.23E-02	3.20E-16	intergenic
15	12426111	0.159	7.92E-02	9.69E-03	1.36E-15	*bfsp1*
16	4661706	0.021	8.58E-02	1.19E-02	1.28E-12	*rspo1*
17	8494762	0.037	5.94E-02	9.64E-03	1.20E-09	intergenic; closest gene, *ece2*
17	11651678	0.025	9.05E-02	1.29E-02	4.79E-12	intergenic; closet gene, *crhr2*
19	6538842	0.021	1.15E-01	1.50E-02	4.11E-14	intergenic
21	9597510	0.209	3.72E-02	6.39E-03	8.74E-09	intergenic
22	10026514	0.052	6.69E-02	8.80E-03	9.69E-14	*pank2*
24	738503	0.025	8.48E-02	1.32E-02	2.43E-10	intergenic
24	10131779	0.028	9.21E-02	1.13E-02	1.92E-15	intergenic

Table S16. Statistics of lead SNPs in GWAS of aggression behavior

Phenotype	Chr	Position	MAF	Beta	Beta_SE	P_wald
Aggression Index	8	9323926	0.12	1.04E+01	1.83E+00	2.36E-08
Aggression Index	9	28220286	0.464	7.80E+00	1.38E+00	2.69E-08
Air Breathing	2	12210826	0.068	-1.62E+00	2.89E-01	3.36E-08
Air Breathing	4	4671604	0.05	-1.72E+00	3.10E-01	4.95E-08
Air Breathing	5	5240300	0.035	2.00E+00	3.05E-01	1.50E-10
Air Breathing	5	15699411	0.069	1.71E+00	2.82E-01	2.56E-09
Air Breathing	9	21760819	0.102	1.46E+00	2.57E-01	2.21E-08
Air Breathing	10	16969635	0.117	1.27E+00	2.28E-01	4.41E-08
Air Breathing	14	3057582	0.111	-9.76E-01	1.71E-01	2.16E-08
Air Breathing	15	14278551	0.028	1.92E+00	3.40E-01	3.13E-08
Air Breathing	21	13514679	0.053	-1.52E+00	2.72E-01	4.11E-08
Air Breathing	24	8388991	0.054	1.57E+00	2.69E-01	9.71E-09
Approach	2	7366880	0.112	-6.22E-01	1.04E-01	4.31E-09
Approach	12	8720757	0.029	7.92E-01	1.35E-01	8.97E-09
Approach	16	1891638	0.026	1.01E+00	1.70E-01	6.09E-09
Charge	2	888879	0.021	6.86E-01	1.15E-01	5.37E-09
Charge	2	1266786	0.025	4.86E-01	8.16E-02	5.22E-09
Charge	2	6021620	0.039	4.64E-01	8.20E-02	2.77E-08
Charge	2	10463284	0.122	3.43E-01	5.24E-02	1.68E-10
Charge	2	15761083	0.035	6.53E-01	1.05E-01	1.26E-09
Charge	3	3165876	0.031	7.99E-01	1.05E-01	1.47E-13
Charge	4	3286888	0.043	-7.66E-01	1.04E-01	8.51E-13
Charge	7	12483332	0.023	8.68E-01	1.16E-01	4.66E-13
Charge	7	15324331	0.029	-7.13E-01	1.28E-01	4.39E-08
Charge	8	1991002	0.043	4.84E-01	8.37E-02	1.37E-08
Charge	8	2165505	0.123	2.94E-01	5.01E-02	8.61E-09
Charge	9	18392726	0.043	7.50E-01	8.92E-02	5.36E-16
Charge	9	19653104	0.029	5.83E-01	1.05E-01	4.22E-08
Charge	10	8851073	0.106	3.74E-01	6.14E-02	2.22E-09
Charge	11	6029533	0.149	2.75E-01	4.48E-02	1.95E-09
Charge	15	5903530	0.049	5.27E-01	8.65E-02	2.41E-09
Charge	15	7760308	0.027	6.29E-01	1.11E-01	2.39E-08
Charge	21	5765521	0.046	5.78E-01	9.17E-02	6.60E-10
Charge	22	2276262	0.03	6.74E-01	1.10E-01	2.08E-09
Charge	24	9212039	0.068	5.05E-01	7.74E-02	1.82E-10
Jerk	6	14544405	0.039	7.68E-01	1.35E-01	2.38E-08
Jerk	19	12771854	0.023	9.97E-01	1.58E-01	7.05E-10
Jerk	19	14663789	0.029	8.88E-01	1.43E-01	1.12E-09
Mouth_open	1	8946456	0.256	-2.93E-01	5.00E-02	8.94E-09
Mouth_open	1	11152960	0.031	4.88E-01	8.63E-02	2.68E-08

Mouth_open	2	688638	0.024	9.10E-01	1.32E-01	1.82E-11
Mouth_open	2	6040705	0.032	6.08E-01	1.09E-01	4.09E-08
Mouth_open	2	10095538	0.069	4.73E-01	8.02E-02	7.28E-09
Mouth_open	2	12503755	0.032	-8.30E-01	1.21E-01	2.57E-11
Mouth_open	2	18173080	0.476	-2.93E-01	5.24E-02	3.77E-08
Mouth_open	3	3165876	0.031	6.43E-01	1.10E-01	8.90E-09
Mouth_open	3	7287377	0.026	9.15E-01	1.19E-01	8.43E-14
Mouth_open	3	9599591	0.04	6.86E-01	9.42E-02	1.42E-12
Mouth_open	3	11972749	0.05	-6.01E-01	9.63E-02	9.44E-10
Mouth_open	3	13292692	0.04	-6.04E-01	1.07E-01	3.04E-08
Mouth_open	4	3351554	0.031	-9.22E-01	1.27E-01	1.82E-12
Mouth_open	4	13155139	0.077	-4.94E-01	8.29E-02	4.88E-09
Mouth_open	5	12164849	0.033	6.68E-01	9.84E-02	3.65E-11
Mouth_open	5	13338703	0.046	5.36E-01	9.28E-02	1.44E-08
Mouth_open	5	14351227	0.056	4.31E-01	7.48E-02	1.54E-08
Mouth_open	5	15266993	0.055	6.79E-01	9.58E-02	5.07E-12
Mouth_open	7	2870858	0.028	-5.65E-01	9.77E-02	1.38E-08
Mouth_open	7	4224900	0.06	6.48E-01	9.71E-02	7.30E-11
Mouth_open	7	8016229	0.031	7.72E-01	1.23E-01	8.54E-10
Mouth_open	7	10177774	0.039	6.71E-01	1.09E-01	1.42E-09
Mouth_open	7	12427490	0.08	-4.88E-01	7.00E-02	1.11E-11
Mouth_open	7	14671220	0.094	-4.31E-01	6.62E-02	2.03E-10
Mouth_open	8	7463244	0.034	6.24E-01	1.08E-01	1.47E-08
Mouth_open	8	8826437	0.06	5.12E-01	8.79E-02	1.07E-08
Mouth_open	8	9323926	0.12	4.44E-01	6.99E-02	5.18E-10
Mouth_open	9	7387782	0.039	8.47E-01	9.49E-02	1.04E-17
Mouth_open	9	12522097	0.087	4.19E-01	7.52E-02	4.24E-08
Mouth_open	9	14367483	0.031	6.75E-01	1.14E-01	5.60E-09
Mouth_open	9	15251036	0.043	6.40E-01	1.03E-01	1.20E-09
Mouth_open	9	17359196	0.031	1.02E+00	1.18E-01	9.11E-17
Mouth_open	9	18356843	0.056	7.06E-01	8.81E-02	9.26E-15
Mouth_open	9	28162819	0.07	6.79E-01	1.05E-01	3.11E-10
Mouth_open	10	2882719	0.108	-4.11E-01	7.16E-02	1.81E-08
Mouth_open	10	6708619	0.034	6.23E-01	1.04E-01	4.48E-09
Mouth_open	10	16384692	0.074	-6.29E-01	9.12E-02	1.81E-11
Mouth_open	10	18005353	0.028	7.96E-01	1.12E-01	5.10E-12
Mouth_open	10	20199582	0.027	-5.76E-01	1.02E-01	3.10E-08
Mouth_open	11	8305920	0.021	-7.59E-01	1.33E-01	2.29E-08
Mouth_open	11	11501465	0.038	6.05E-01	1.00E-01	3.00E-09
Mouth_open	11	13504051	0.029	-6.69E-01	1.16E-01	1.55E-08
Mouth_open	12	2430401	0.09	4.83E-01	8.48E-02	2.13E-08
Mouth_open	13	16707525	0.028	7.41E-01	1.23E-01	3.31E-09
Mouth_open	14	8133765	0.029	6.80E-01	1.18E-01	1.64E-08

Mouth_open	15	7760308	0.027	9.33E-01	1.11E-01	6.94E-16
Mouth_open	15	9982570	0.077	4.21E-01	7.03E-02	4.34E-09
Mouth_open	15	10418179	0.087	4.27E-01	6.80E-02	7.73E-10
Mouth_open	15	12617566	0.029	-7.05E-01	1.04E-01	4.17E-11
Mouth_open	16	2086114	0.094	-3.99E-01	7.15E-02	4.25E-08
Mouth_open	16	9389326	0.048	7.02E-01	9.70E-02	1.96E-12
Mouth_open	16	10862572	0.077	4.24E-01	7.12E-02	5.17E-09
Mouth_open	16	12805719	0.021	7.70E-01	1.34E-01	1.71E-08
Mouth_open	17	10937179	0.041	-6.81E-01	1.14E-01	4.23E-09
Mouth_open	19	13164163	0.057	4.90E-01	7.82E-02	8.44E-10
Mouth_open	21	5941809	0.05	-7.06E-01	1.11E-01	5.41E-10
Mouth_open	21	6139552	0.062	-4.21E-01	7.37E-02	1.96E-08
Mouth_open	21	9979386	0.031	6.99E-01	1.03E-01	3.58E-11
Mouth_open	21	12315581	0.049	-6.29E-01	1.09E-01	1.66E-08
Mouth_open	22	4291697	0.025	9.93E-01	1.29E-01	9.48E-14
Mouth_open	24	6750410	0.03	-6.39E-01	1.10E-01	1.18E-08
Mouth_open	24	8393907	0.036	-5.79E-01	9.94E-02	1.12E-08
Mouth_open	24	11837916	0.036	6.87E-01	1.07E-01	3.03E-10
Pacing	2	7405263	0.025	3.30E-01	5.57E-02	6.23E-09
Pacing	3	7960607	0.033	3.06E-01	4.57E-02	6.32E-11
Pacing	4	30370545	0.021	3.32E-01	5.94E-02	3.86E-08
Pacing	5	13528298	0.024	3.18E-01	4.71E-02	4.39E-11
Pacing	5	19345107	0.02	2.92E-01	4.84E-02	3.41E-09
Pacing	6	11724880	0.074	1.75E-01	3.00E-02	1.03E-08
Pacing	7	3917434	0.026	2.98E-01	5.22E-02	2.08E-08
Pacing	8	2347817	0.033	2.66E-01	4.56E-02	1.02E-08
Pacing	8	5021010	0.034	3.30E-01	4.62E-02	3.63E-12
Pacing	9	19566680	0.049	2.52E-01	3.76E-02	6.58E-11
Pacing	10	8277069	0.083	1.65E-01	2.93E-02	2.82E-08
Pacing	10	8703633	0.02	3.53E-01	5.79E-02	2.20E-09
Pacing	11	7833331	0.021	3.80E-01	5.77E-02	1.30E-10
Pacing	11	13356668	0.038	2.78E-01	4.27E-02	2.00E-10
Pacing	11	14659061	0.034	3.62E-01	4.16E-02	5.28E-17
Pacing	12	6413797	0.045	2.26E-01	3.82E-02	6.27E-09
Pacing	12	11813544	0.058	2.17E-01	3.09E-02	7.57E-12
Pacing	13	16507512	0.041	2.60E-01	4.26E-02	2.39E-09
Pacing	15	9510311	0.029	2.73E-01	4.42E-02	1.38E-09
Pacing	16	12661328	0.02	3.40E-01	5.53E-02	1.72E-09
Pacing	16	13852715	0.033	2.49E-01	4.40E-02	2.56E-08
Pacing	17	676561	0.023	4.21E-01	5.87E-02	2.86E-12
Pacing	17	5313515	0.031	-7.05E-01	1.05E-01	4.83E-11
Pacing	17	9863738	0.021	3.29E-01	5.79E-02	2.45E-08
Pacing	17	12960185	0.061	2.09E-01	3.67E-02	2.18E-08

Pacing	19	6558844	0.028	2.82E-01	4.83E-02	1.05E-08
Pacing	22	2583595	0.031	3.12E-01	4.97E-02	7.64E-10
Pacing	24	10163599	0.02	3.37E-01	5.98E-02	3.15E-08
Retreat	5	19319551	0.02	6.77E-01	8.20E-02	1.56E-15
Retreat	7	4530008	0.059	3.91E-01	5.91E-02	1.05E-10
Retreat	8	12929786	0.033	5.25E-01	8.64E-02	2.53E-09
Retreat	9	5961575	0.039	4.87E-01	8.37E-02	1.13E-08
Retreat	22	8949086	0.036	4.84E-01	8.04E-02	3.51E-09

REFERENCES AND NOTES

1. W. J. Rainboth, *Fishes of the Cambodian Mekong* (Food and Agriculture Organization, 1996).

2. H. M. Smith, L. P. Schultz, *The Fresh-water Fishes of Siam, or Thailand* (Smithsonian Institution, U.S. National Museum Bulletin, Government Printing Office, 1945).

3. H.-W. Lissmann, Die umwelt des kampffisches (Betta splendens Regan). *Z. Vgl. Physiol.* **18**, 65–111 (1932).

4. H. Goodrich, R. N. Mercer, Genetics and colors of the Siamese fighting, *Betta splendens*. *Science* **79**, 318–319 (1934).

5. K. Umrath, Über die Vererbung der Farben und des Geschlechts beim Schleierkampffisch, *Betta splendens*. *Z.Ver-erbungslehre* **77**, 450–454 (1939).

6. K. Eberhardt, Die Vererbung der Farben bei Betta Splendens Regan. *Z.Ver-erbungslehre* **79**, 548–560 (1941).

7. K. Eberhardt, Geschlechtsbestimmung und -Differenzierung bei Betta Splendens Regan I. *Z.Ver-erbungslehre* **81**, 363–373 (1943).

8. K. Eberhardt, Ein Fall von geschlechtskontrollierter Vererbung bei Betta splendens Regan. *Z. Indukt. Abstammungs. Vererbungsl.* **81**, 72–83 (1943).

9. G. Svärdson, T. Wickbom, The chromosomes of two species of Anabantidae (Teleostei), with a new case of sex reversal. *Hereditas* **28**, 212–216 (1942).

10. L. Wang, F. Sun, Z. Y. Wan, B. Ye, Y. Wen, H. Liu, Z. Yang, H. Pang, Z. Meng, B. Fan, Y. Alfiko, Y. Shen, B. Bai, M. S. Q. Lee, F. Piferrer, M. Schartl, A. Meyer, G. H. Yue, Genomic basis of striking fin shapes and colours in the fighting fish. *Mol. Biol. Evol.* **38**, 3383–3396 (2021).

11. Y. M. Kwon, N. Vranken, C. Hoge, M. R. Lichak, A. L. Norovich, K. X. Francis, J. Camacho-Garcia, I. Bista, J. Wood, S. McCarthy, W. Chow, H. H. Tan, K. Howe, S. Bandara, J. von Lintig, L. Ruber, R.

Durbin, H. Svardal, A. Bendesky, Genomic consequences of domestication of the Siamese fighting fish. *Sci. Adv.* **8**, eabm4950 (2022).

12. L. Wang, F. Sun, Z. Y. Wan, Z. Yang, Y. X. Tay, M. Lee, B. Ye, Y. Wen, Z. Meng, B. Fan, Y. Alfiko, Y. Shen, F. Piferrer, A. Meyer, M. Schartl, G. H. Yue, Transposon-induced epigenetic silencing in the X chromosome as a novel form of dmrt1 expression regulation during sex determination in the fighting fish. *BMC Biol.* **20**, 5 (2022).

13. F.-S. Grazyna, D. Fopp-Bayat, M. Jankun, S. Krejszeff, A. Mamcarz, Note on the karyotype and NOR location of Siamese fighting fish *Betta splendens* (Perciformes, Osphronemidae). *Caryologia* **61**, 349–353 (2008).

14. G. Parra, K. Bradnam, I. Korf, CEGMA: A pipeline to accurately annotate core genes in eukaryotic genomes. *Bioinformatics* **23**, 1061–1067 (2007).

15. F. A. Simão, R. M. Waterhouse, P. Ioannidis, E. V. Kriventseva, E. M. Zdobnov, BUSCO: Assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* **31**, 3210–3212 (2015).

16. X. Zhou, M. Stephens, Genome-wide efficient mixed-model analysis for association studies. *Nat. Genet.* **44**, 821–824 (2012).

17. R. J. Vanzo, H. Twede, K. S. Ho, A. Prasad, M. M. Martin, S. T. South, E. R. Wassman, Clinical significance of copy number variants involving KANK1 in patients with neurodevelopmental disorders. *Eur. J. Med. Genet.* **62**, 15–20 (2019).

18. P. Martínez, A. M. Viñas, L. Sánchez, N. Díaz, L. Ribas, F. Piferrer, Genetic architecture of sex determination in fish: Applications to sex ratio control in aquaculture. *Front. Genet.* **5**, 340 (2014).

19. C. A. Smith, K. N. Roeszler, T. Ohnesorg, D. M. Cummins, P. G. Farlie, T. J. Doran, A. H. Sinclair, The avian Z-linked gene DMRT1 is required for male sex determination in the chicken. *Nature* **461**, 267–271 (2009).

20. C. Ge, J. Ye, C. Weber, W. Sun, H. Zhang, Y. Zhou, C. Cai, G. Qian, B. Capel, The histone demethylase KDM6B regulates temperature-dependent sex determination in a turtle species. *Science* **360**, 645–648 (2018).

21. I. Nanda, M. Kondo, U. Hornung, S. Asakawa, C. Winkler, A. Shimizu, Z. Shan, T. Haaf, N. Shimizu, A. Shima, M. Schmid, M. Schartl, A duplicated copy of *DMRT1* in the sex-determining region of the Y chromosome of the medaka, *Oryzias latipes*. *Proc. Natl. Acad. Sci. U.S.A.* **99**, 11778–11783 (2002).

22. N. Ospina-Alvarez, F. Piferrer, Temperature-dependent sex determination in fish revisited: Prevalence, a single sex ratio response pattern, and possible effects of climate change. *PLOS ONE* **3**, e2837 (2008).

23. X. Zhang, N. Yang, F. Jiang, H. Huang, Inheritance of body colors of Siamese fighting fishes of different strains. *Chinese J. Trop. Agric.* **34**, 109–113 (2014).

24. D. Lee, I. M.-J. Xu, D. K.-C. Chiu, R. K.-H. Lai, A. P.-W. Tse, L. L. Li, C.-T. Law, F. H.-C. Tsang, L. L. Wei, C. Y.-K. Chan, C.-M. Wong, I. O.-L. Ng, C. C.-L. Wong, Folate cycle enzyme MTHFD1L confers metabolic advantages in hepatocellular carcinoma. *J. Clin. Invest.* **127**, 1856–1872 (2017).

25. J. T. Bagnara, J. Matsumoto, W. Ferris, S. K. Frost, W. A. Turner Jr., T. T. Tchen, J. D. Taylor, Common origin of pigment cells. *Science* **203**, 410–415 (1979).

26. J. H. M. v. Esch, Understanding metallic genetics (2008); http://www.bettaterritory.nl/BT-AABcoppergenetics.htm.

27. C. Bacon, V. Endris, G. A. Rappold, The cellular function of srGAP3 and its role in neuronal morphogenesis. *Mech. Dev.* **130**, 391–395 (2013).

28. T. Hama, Chromatophores and iridocytes, in *Medaka (Killifish): Biology and Strains*, T. Yamamoto, Ed. (Keigaku Inc, 1975), pp. 138–153.

29. W. Liu, D. Morito, S. Takashima, Y. Mineharu, H. Kobayashi, T. Hitomi, H. Hashikata, N. Matsuura, S. Yamazaki, A. Toyoda, K.-I. Kikuta, Y. Takagi, K. H. Harada, A. Fujiyama, R. Herzig, B. Krischek, L. Zou, J. E. Kim, M. Kitakaze, S. Miyamoto, K. Nagata, N. Hashimoto, A. Koizumi, Identification of

RNF213 as a susceptibility gene for moyamoya disease and its possible role in vascular development. *PLOS ONE* **6**, e22542 (2011).

30. C. Pietsch, P. Hirsch, *Biology and Ecology of Carp* (CRC Press, Taylor & Francis Group, 2015).

31. M. Tsutsumi, S. Imai, Y. Kyono-Hamaguchi, S. Hamaguchi, A. Koga, H. Hori, Color reversion of the albino medaka fish associated with spontaneous somatic excision of the Tol-1 transposable element from the tyrosinase gene. *Pigment Cell Res.* **19**, 243–247 (2006).

32. E. P. Ahi, L. A. Lecaudey, A. Ziegelbecker, O. Steiner, R. Glabonjat, W. Goessler, V. Hois, C. Wagner, A. Lass, K. M. Sefc, Comparative transcriptomics reveals candidate carotenoid color genes in an East African cichlid fish. *BMC Genomics* **21**, 54 (2020).

33. C. Li, H. Chen, Y. Zhao, S. Chen, H. Xiao, Comparative transcriptomics reveals the molecular genetic basis of pigmentation loss in *Sinocyclocheilus* cavefishes. *Ecol. Evol.* **10**, 14256–14271 (2020).

34. G. A. Lucas, "A study of variation in the Siamese fighting fish, *Betta splendens*, with emphasis on color mutants and the problem of sex determination," thesis, Iowa State University (1968).

35. L. Zhu, L. Li, Y. Qi, Z. Yu, Y. Xu, Cryo-EM structure of SMG1–SMG8–SMG9 complex. *Cell Res.* **29**, 1027–1034 (2019).

36. S. Perathoner, J. M. Daane, U. Henrion, G. Seebohm, C. W. Higdon, S. L. Johnson, C. Nüsslein-Volhard, M. P. Harris, Bioelectric signaling regulates size in zebrafish fins. *PLOS Genet.* **10**, e1004080 (2014).

37. S. Stewart, H. K. Le Bleu, G. A. Yette, A. L. Henner, A. E. Robbins, J. A. Braunstein, K. Stankunas, longfin causes cis-ectopic expression of the kcnh2a ether-a-go-go K^+ channel to autonomously prolong fin outgrowth. *Development* **148** dev199384 (2019).

38. M. R. Silic, Q. Wu, B. H. Kim, G. Golling, K. H. Chen, R. Freitas, A. A. Chubykin, S. K. Mittal, G. Zhang, Potassium channel-associated bioelectricity of the dermomyotome determines fin patterning in zebrafish. *Genetics* **215**, 1067–1084 (2020).

39. J. S. Lanni, D. Peal, L. Ekstrom, H. Chen, C. Stanclift, M. E. Bowen, A. Mercado, G. Gamba, K. T. Kahle, M. P. Harris, Integrated K^+ channel and K^+Cl- cotransporter functions are required for the coordination of size and proportion during development. *Dev. Biol.* **456**, 164–178 (2019).

40. J. H. Laity, B. M. Lee, P. E. Wright, Zinc finger proteins: New insights into structural and functional diversity. *Curr. Opin. Struct. Biol.* **11**, 39–46 (2001).

41. L. Angus, S. Moleirinho, L. Herron, A. Sinha, X. Zhang, M. Niestrata, K. Dholakia, M. B. Prystowsky, K. F. Harvey, P. A. Reynolds, Willin/FRMD6 expression activates the Hippo signaling pathway kinases in mammals and antagonizes oncogenic YAP. *Oncogene* **31**, 238–250 (2012).

42. O. Larouche, M. L. Zelditch, R. Cloutier, Fin modules: An evolutionary perspective on appendage disparity in basal vertebrates. *BMC Biol.* **15**, 32 (2017).

43. T. Nakamura, A. R. Gehrke, J. Lemberg, J. Szymaszek, N. H. Shubin, Digits and fin rays share common developmental histories. *Nature* **537**, 225–228 (2016).

44. Y. Cui, S. He, C. Xing, K. Lu, J. Wang, G. Xing, A. Meng, S. Jia, F. He, L. Zhang, SCF^{FBXL15} regulates BMP signalling by directing the degradation of HECT-type ubiquitin ligase Smurf1. *EMBO J.* **30**, 2675–2689 (2011).

45. K. E. Kemper, P. M. Visscher, M. E. Goddard, Genetic architecture of body size in mammals. *Genome Biol.* **13**, 244 (2012).

46. N. Niepoth, A. Bendesky, How natural genetic variation shapes behavior. *Annu. Rev. Genomics Human Genet.* **21**, 437–463 (2020).

47. A. Ramos, D. Gonçalves, Artificial selection for male winners in the Siamese fighting fish *Betta splendens* correlates with high female aggression. *Front. Zool.* **16**, 34 (2019).

48. K. Kikuma, X. Li, D. Kim, D. Sutter, D. K. Dickman, Extended synaptotagmin localizes to presynaptic ER and promotes neurotransmission and synaptic growth in *Drosophila*. *Genetics* **207**, 993–1006 (2017).

49. Y. Li, P. Hollingworth, P. Moore, C. Foy, N. Archer, J. Powell, P. Nowotny, P. Holmans, M. O'Donovan, K. Tacey, L. Doil, R. van Luchene, V. Garcia, C. Rowland, K. Lau, J. Cantanese, J. Sninsky, J. Hardy, L. Thal, J. C. Morris, A. Goate, S. Lovestone, M. Owen, J. Williams, A. Grupe, Genetic association of the APP binding protein 2 gene (APBB2) with late onset Alzheimer disease. *Human Mutat.* **25**, 270–277 (2005).

50. J.-H. Seo, S.-K. Song, P. H. Lee, A novel PANK2 mutation in a patient with atypical pantothenate-kinase-associated neurodegeneration presenting with adult-onset parkinsonism. *J. Clin. Neurol.* **5**, 192–194 (2009).

51. T. Kishimoto, J. Radulovic, M. Radulovic, C. R. Lin, C. Schrick, F. Hooshmand, O. Hermanson, M. G. Rosenfeld, J. Spiess, Deletion of crhr2 reveals an anxiolytic role for corticotropin-releasing hormone receptor-2. *Nat. Genet.* **24**, 415–419 (2000).

52. I. Y. Buchsbaum, P. Kielkowski, G. Giorgio, A. C. O'Neill, R. Di Giaimo, C. Kyrousi, S. Khattak, S. A. Sieber, S. P. Robertson, S. Cappello, ECE2 regulates neurogenesis and neuronal migration during human cortical development. *EMBO Rep.* **21**, e48204 (2020).

53. D. Wang, H. M. Stoveken, S. Zucca, M. Dao, C. Orlandi, C. Song, I. Masuho, C. Johnston, K. J. Opperman, A. C. Giles, M. S. Gill, E. A. Lundquist, B. Grill, K. A. Martemyanov, Genetic behavioral screen identifies an orphan anti-opioid system. *Science* **365**, 1267–1273 (2019).

54. R. Kandaswamy, A. McQuillin, S. I. Sharp, A. Fiorentino, A. Anjorin, R. A. Blizard, D. Curtis, H. M. D. Gurling, Genetic association, mutation screening, and functional analysis of a Kozak sequence variant in the metabotropic glutamate receptor 3 gene in bipolar disorder. *JAMA Psychiatry.* **70**, 591–598 (2013).

55. M. F. Egan, R. E. Straub, T. E. Goldberg, I. Yakub, J. H. Callicott, A. R. Hariri, V. S. Mattay, A. Bertolino, T. M. Hyde, C. Shannon-Weickert, M. Akil, Jeremy Crook, R. K. Vakkalanka, R. Balkissoon, R. A. Gibbs, J. E. Kleinman, D. R. Weinberger, Variation in GRM3 affects cognition, prefrontal glutamate, and risk for schizophrenia. *Proc. Natl. Acad. Sci. U.S.A.* **101**, 12604–12609 (2004).

56. M. R. Dyer, J. E. Walker, Sequences of members of the human gene family for the c subunit of mitochondrial ATP synthase. *Biochem. J.* **293**, 51–64 (1993).

57. J. Egawa, S. Hoya, Y. Watanabe, A. Nunokawa, M. Shibuya, M. Ikeda, E. Inoue, S. Okuda, K. Kondo, T. Saito, N. Kaneko, T. Muratake, H. Igeta, N. Iwata, T. Someya, Rare *UNC13B* variations and risk of schizophrenia: Whole-exome sequencing in a multiplex family and follow-up resequencing and a case-control study. *Am. J. Med. Genet. B Neuropsychiatr. Genet.* **171**, 797–805 (2016).

58. J. Wang, J.-D. Qiao, X.-R. Liu, D.-T. Liu, Y.-H. Chen, Y. Wu, Y. Sun, J. Yu, R.-N. Ren, Z. Mei, Y.-X. Liu, Y.-W. Shi, M. Jiang, S.-M. Lin, N. He, B. Li, W.-J. Bian, B.-M. Li, Y.-H. Yi, T. Su, H.-K. Liu, W.-Y. Gu, W.-P. Liao, *UNC13B* variants associated with partial epilepsy with favourable outcome. *Brain* **144**, 3050–3060 (2021).

59. J.-M. Belton, R. P. McCord, J. H. Gibcus, N. Naumova, Y. Zhan, J. Dekker, Hi-C: A comprehensive technique to capture the conformation of genomes. *Methods* **58**, 268–276 (2012).

60. L. Li, C. J. Stoeckert, D. S. Roos, OrthoMCL: Identification of ortholog groups for eukaryotic genomes. *Genome Res.* **13**, 2178–2189 (2003).

61. R. C. Edgar, MUSCLE: Multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **32**, 1792–1797 (2004).

62. A. Stamatakis, RAxML-VI-HPC: Maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* **22**, 2688–2690 (2006).

63. Z. Yang, PAML 4: Phylogenetic analysis by maximum likelihood. *Mol. Biol. Evol.* **24**, 1586–1591 (2007).

64. H. Li, R. Durbin, Fast and accurate long-read alignment with Burrows–Wheeler transform. *Bioinformatics* **26**, 589–595 (2010).

65. Broad Institute, Picard toolkit (2019).

66. R. Poplin, V. Ruano-Rubio, M. A. DePristo, T. J. Fennell, M. O. Carneiro, G. A. Van der Auwera, D. E. Kling, L. D. Gauthier, A. Levy-Moonshine, D. Roazen, K. Shakir, J. Thibault, S. Chandran, C. Whelan, M. Lek, S. Gabriel, M. J. Daly, B. Neale, D. G. MacArthur, E. Banks, Scaling accurate genetic variant discovery to tens of thousands of samples. bioRxiv 201178 [**Preprint**]. 24 July 2018. https://doi.org/10.1101/201178.

67. T. S. Korneliussen, A. Albrechtsen, R. Nielsen, ANGSD: Analysis of next generation sequencing data. *BMC Bioinformatics* **15**, 356 (2014).

68. H. Li, B. Handsaker, A. Wysoker, T. Fennell, J. Ruan, N. Homer, G. Marth, G. Abecasis, R. Durbin; 1000 Genome Project Data Processing Subgroup, The sequence alignment/map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).

69. B. L. Browning, S. R. Browning, Genotype imputation with millions of reference samples. *Am. J. Hum. Genet.* **98**, 116–126 (2016).

70. J. Meisner, A. Albrechtsen, Inferring population structure and admixture proportions in low-depth NGS data. *Genetics* **210**, 719–731 (2018).

71. R Core Team, *R: A Language and Environment for Statistical Computing* (R Foundation for Statistical Computing, 2013).

72. L.-T. Nguyen, H. A. Schmidt, A. Von Haeseler, B. Q. Minh, IQ-TREE: A fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol. Biol. Evol.* **32**, 268–274 (2015).

73. J. Pickrell, J. Pritchard, Inference of population splits and mixtures from genome-wide allele frequency data. *PLOS Genet.* **8**, e1002967 (2012).

74. C. Palmer, I. Pe'er, Statistical correction of the winner's curse explains replication variability in quantitative trait genome-wide association studies. *PLOS Genet.* **13**, e1006916 (2017).

75. A. Mortazavi, B. A. Williams, K. McCue, L. Schaeffer, B. Wold, Mapping and quantifying mammalian transcriptomes by RNA-seq. *Nat. Methods* **5**, 621–628 (2008).

76. Y. Liao, G. K. Smyth, W. Shi, featureCounts: An efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* **30**, 923–930 (2014).

77. M. D. Robinson, D. J. McCarthy, G. K. Smyth, edgeR: A Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**, 139–140 (2010).

78. M. I. Love, W. Huber, S. Anders, Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).

79. G. Marçais, C. Kingsford, A fast, lock-free approach for efficient parallel counting of occurrences of k-mers. *Bioinformatics* **27**, 764–770 (2011).

80. C. S. Chin, P. Peluso, F. J. Sedlazeck, M. Nattestad, G. T. Concepcion, A. Clum, C. Dunn, R. O'Malley, R. Figueroa-Balderas, A. Morales-Cruz, G. R. Cramer, M. Delledonne, C. Luo, J. R. Ecker, D. Cantu, D. R. Rank, M. C. Schatz, Phased diploid genome assembly with single-molecule real-time sequencing. *Nat. Methods* **13**, 1050–1054 (2016).

81. C. S. Chin, D. H. Alexander, P. Marks, A. A. Klammer, J. Drake, C. Heiner, A. Clum, A. Copeland, J. Huddleston, E. E. Eichler, S. W. Turner, J. Korlach, Nonhybrid, finished microbial genome assemblies from long-read SMRT sequencing data. *Nat. Methods* **10**, 563–569 (2013).

82. B. J. Walker, T. Abeel, T. Shea, M. Priest, A. Abouelliel, S. Sakthikumar, C. A. Cuomo, Q. Zeng, J. Wortman, S. K. Young, A. M. Earl, Pilon: An integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PLOS ONE* **9**, e112963 (2014).

83. M. J. Roach, S. A. Schmidt, A. R. Borneman, Purge Haplotigs: Allelic contig reassignment for third-gen diploid genome assemblies. *BMC Bioinformatics* **19**, 460 (2018).

84. A. Adey, J. O. Kitzman, J. N. Burton, R. Daza, A. Kumar, L. Christiansen, M. Ronaghi, S. Amini, K. L. Gunderson, F. J. Steemers, J. Shendure, In vitro, long-range sequence information for de novo genome assembly via transposase contiguity. *Genome Res.* **24**, 2041–2049 (2014).

85. E. T. Lam, A. Hastie, C. Lin, D. Ehrlich, S. K. Das, M. D. Austin, P. Deshpande, H. Cao, N. Nagarajan, M. Xiao, P. Y. Kwok, Genome mapping on nanochannel arrays for structural variation analysis and sequence assembly. *Nat. Biotechnol.* **30**, 771–776 (2012).

86. J. N. Burton, A. Adey, R. P. Patwardhan, R. Qiu, J. O. Kitzman, J. Shendure, Chromosome-scale scaffolding of de novo genome assemblies based on chromatin interactions. *Nat. Biotechnol.* **31**, 1119–1125 (2013).

87. G. Benson, Tandem repeats finder: A program to analyze DNA sequences. *Nucleic Acids Res.* **27**, 573–580 (1999).

88. A. F. Smit, R. Hubley, *RepeatModeler Open* (2008).

89. Z. Xu, H. Wang, LTR_FINDER: An efficient tool for the prediction of full-length LTR retrotransposons. *Nucleic Acids Res.* **35**, W265–W268 (2007).

90. A. L. Price, N. C. Jones, P. A. Pevzner, De novo identification of repeat families in large genomes. *Bioinformatics* **21**, i351–i358 (2005).

91. A. Smit, R. Hubley, P. Green, *RepeatMasker Open v4.0* (2013).

92. W. Bao, K. K. Kojima, O. Kohany, Repbase update, a database of repetitive elements in eukaryotic genomes. *Mobile DNA* **6**, 11 (2015).

93. A. Smit, R. Hubley, P. Green, *RepeatMasker Open v3.0* (2004).

94. P. P. Chan, T. M. Lowe, tRNAscan-SE: Searching for tRNA genes in genomic sequences. *Methods Mol. Biol.* **1962**, 1–14 (2019).

95. E. P. Nawrocki, D. L. Kolbe, S. R. Eddy, Infernal 1.0: Inference of RNA alignments. *Bioinformatics* **25**, 1335–1337 (2009).

96. M. Stanke, O. Schöffmann, B. Morgenstern, S. Waack, Gene prediction in eukaryotes with a generalized hidden Markov model that uses hints from external sources. *BMC Bioinformatics* **7**, 62 (2006).

97. A. A. Salamov, V. V. Solovyev, Ab initio gene finding in *Drosophila* genomic DNA. *Genome Res.* **10**, 516–522 (2000).

98. G. Parra, E. Blanco, R. Guigo, GeneID in *Drosophila*. *Genome Res.* **10**, 511–515 (2000).

99. W. H. Majoros, M. Pertea, S. L. Salzberg, TigrScan and GlimmerHMM: Two open source ab initio eukaryotic gene-finders. *Bioinformatics* **20**, 2878–2879 (2004).

100. I. Korf, Gene finding in novel genomes. *BMC Bioinformatics* **5**, 59 (2004).

101. E. Birney, M. Clamp, R. Durbin, GeneWise and genomewise. *Genome Res.* **14**, 988–995 (2004).

102. M. G. Grabherr, B. J. Haas, M. Yassour, J. Z. Levin, D. A. Thompson, I. Amit, X. Adiconis, L. Fan, R. Raychowdhury, Q. Zeng, Z. Chen, E. Mauceli, N. Hacohen, A. Gnirke, N. Rhind, F. di Palma, B. W. Birren, C. Nusbaum, K. Lindblad-Toh, N. Friedman, A. Regev, Full-length transcriptome assembly from RNA-seq data without a reference genome. *Nat. Biotechnol.* **29**, 644–652 (2011).

103. B. J. Haas, A. L. Delcher, S. M. Mount, J. R. Wortman, R. K. Smith Jr., L. I. Hannick, R. Maiti, C. M. Ronning, D. B. Rusch, C. D. Town, S. L. Salzberg, O. White, Improving the *Arabidopsis* genome annotation using maximal transcript alignment assemblies. *Nucleic Acids Res.* **31**, 5654–5666 (2003).

104. B. J. Haas, S. L. Salzberg, W. Zhu, M. Pertea, J. E. Allen, J. Orvis, O. White, C. R. Buell, J. R. Wortman, Automated eukaryotic gene structure annotation using EVidenceModeler and the program to assemble spliced alignments. *Genome Biol.* **9**, R7 (2008).

105. R. Apweiler, A. Bairoch, C. H. Wu, W. C. Barker, B. Boeckmann, S. Ferro, E. Gasteiger, H. Huang, R. Lopez, M. Magrane, M. J. Martin, D. A. Natale, C. O'Donovan, N. Redaschi, L.-S. L. Yeh, UniProt: The universal protein knowledgebase. *Nucleic Acids Res.* **32**, D115–D119 (2004).

106. P. Jones, D. Binns, H.-Y. Chang, M. Fraser, W. Li, C. McAnulla, H. McWilliam, J. Maslen, A. Mitchell, G. Nuka, S. Pesseat, A. F. Quinn, A. Sangrador-Vegas, M. Scheremetjew, S.-Y. Yong, R. Lopez, S. Hunter, InterProScan 5: Genome-scale protein function classification. *Bioinformatics* **30**, 1236–1240 (2014).

107. M. Kanehisa, Y. Sato, M. Kawashima, M. Furumichi, M. Tanabe, KEGG as a reference resource for gene and protein annotation. *Nucleic Acids Res.* **44**, D457–D462 (2016).

108. C. C. F. Pleeging, C. P. H. Moons, Potential Welfare issues of the Siamese fighting fish (*Betta splendens*) at the retailer and hobbyist aquarium. *Vlaams Diergeneeskundig Tijdschrift* **86**, 213–223 (2017).

109. M. Lipson, P.-R. Loh, A. Levin, D. Reich, N. Patterson, B. Berger, Efficient moment-based inference of admixture parameters and sources of gene flow. *Mol. Biol. Evol.* **30**, 1788–1802 (2013).

110. Z. Hao, D. Lv, Y. Ge, J. Shi, D. Weijers, G. Yu, J. Chen, RIdeogram: Drawing SVG graphics to visualize and map genome-wide data on the idiograms. *PeerJ Comput. Sci.* **6**, e251 (2020).

111. G. Fan, J. Chan, K. Ma, B. Yang, H. Zhang, X. Yang, C. Shi, H. Chun-Hin Law, Z. Ren, Q. Xu, Q. Liu, J. Wang, W. Chen, L. Shao, D. Gonçalves, A. Ramos, S. D. Cardoso, M. Guo, J. Cai, X. Xu, J. Wang, H. Yang, X. Liu, Y. Wang, Chromosome-level reference genome of the Siamese fighting fish *Betta splendens*, a model species for the study of aggression. *Gigascience* **7**, giy087 (2018).

112. A. Rhie, S. A. McCarthy, O. Fedrigo, J. Damas, G. Formenti, S. Koren, M. Uliano-Silva, W. Chow, A. Fungtammasan, J. Kim, C. Lee, B. J. Ko, M. Chaisson, G. L. Gedman, L. J. Cantin, F. Thibaud-Nissen, L. Haggerty, I. Bista, M. Smith, B. Haase, J. Mountcastle, S. Winkler, S. Paez, J. Howard, S. C. Vernes, T. M. Lama, F. Grutzner, W. C. Warren, C. N. Balakrishnan, D. Burt, J. M. George, M. T. Biegler, D. Iorns, A. Digby, D. Eason, B. Robertson, T. Edwards, M. Wilkinson, G. Turner, A. Meyer, A. F. Kautt, P. Franchini, H. W. Detrich III, H. Svardal, M. Wagner, G. J. P. Naylor, M. Pippel, M. Malinsky, M. Mooney, M. Simbirsky, B. T. Hannigan, T. Pesout, M. Houck, A. Misuraca, S. B. Kingan, R. Hall, Z. Kronenberg, I. Sović, C. Dunn, Z. Ning, A. Hastie, J. Lee, S. Selvaraj, R. E. Green, N. H. Putnam, I. Gut, J. Ghurye, E. Garrison, Y. Sims, J. Collins, S. Pelan, J. Torrance, A. Tracey, J. Wood, R. E. Dagnew, D. Guan, S. E. London, D. F. Clayton, C. V. Mello, S. R. Friedrich, P. V. Lovell, E. Osipova, F. O. Al-Ajli, S. Secomandi, H. Kim, C. Theofanopoulou, M. Hiller, Y. Zhou, R. S. Harris, K. D. Makova, P. Medvedev, J. Hoffman, P. Masterson, K. Clark, F. Martin, K. Howe, P. Flicek, B. P. Walenz, W. Kwak, H. Clawson, M. Diekhans, L. Nassar, B. Paten, R. H. S. Kraus, A. J. Crawford, M. T. P. Gilbert, G. Zhang, B. Venkatesh, R. W. Murphy, K.-P. Koepfli, B. Shapiro, W. E. Johnson, F. D. Palma, T. Marques-Bonet, E. C. Teeling, T. Warnow, J. M. Graves, O. A. Ryder, D. Haussler, S. J.

O'Brien, J. Korlach, H. A. Lewin, K. Howe, E. W. Myers, R. Durbin, A. M. Phillippy, E. D. Jarvis, Towards complete and error-free genome assemblies of all vertebrate species. *Nature* **592**, 737–746 (2021).

113. S. Prost, M. Petersen, M. Grethlein, S. J. Hahn, N. Kuschik-Maczollek, M. E. Olesiuk, J.-O. Reschke, T. E. Schmey, C. Zimmer, D. K. Gupta, T. Schell, R. Coimbra, J. De Raad, F. Lammers, S. Winter, A. Janke, Improving the chromosome-level genome assembly of the Siamese fighting fish (*Betta splendens*) in a university master's course. *G3 (Bethesda)* **10**, 2179–2183 (2020).