Supplementary Information for

Repeated translocation of a supergene underlying rapid sex chromosome turnover in *Takifugu* pufferfish.

Ahammad Kabir, Risa Ieda, Sho Hosoya, Daigaku Fujikawa, Kazufumi Atsumi, Shota Tajima, Aoi Nozawa, Takashi Koyama, Shotaro Hirase, Osamu Nakamura, Mitsutaka Kadota, Osamu Nishimura, Shigehiro Kuraku, Yasukazu Nakamura, Hisato Kobayashi, Atsushi Toyoda, Satoshi Tasumi, Kiyoshi Kikuchi

Corresponding author: Kiyoshi Kikuchi and Sho Hosoya
Email: akikuchi@mail.ecc.u-tokyo.ac.jp, ahosoya@mail.ecc.u-tokyo.ac.jp

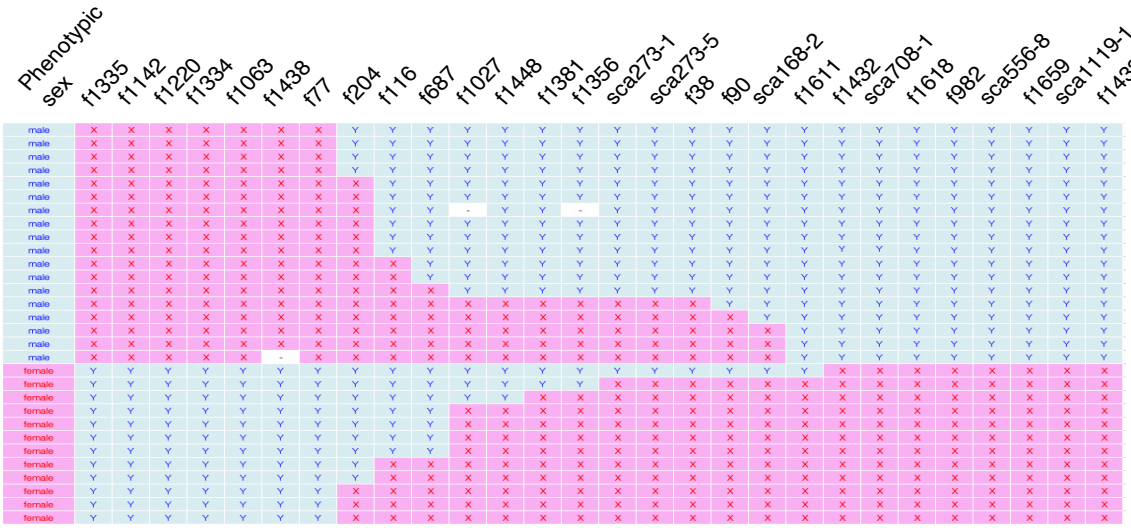
This PDF file includes:

Other supplementary Dataset for this manuscript include the following:

DNA sequences around the exon9 of *Amhr2* (Fasta1_amhr2exon9.text)
DNA sequences for Y-specific genes and their autosomal paralogs (Fasta2_genes_on_SDR_and_their_paralogs.txt)
Genotype data for genetic mapping (mapping genotype.xlsx)

Supplementary Figures

(A)



(B)

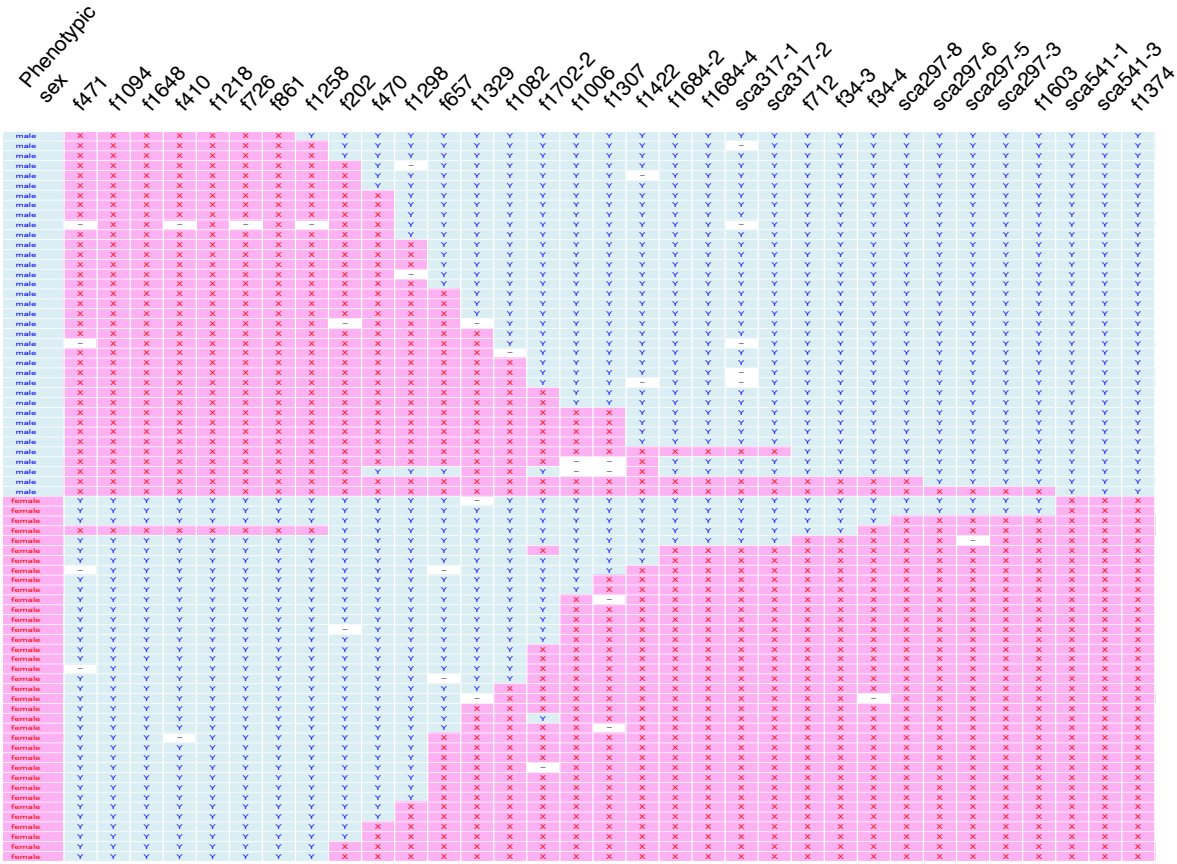


Fig. S1. The search for individuals with recombination near the sex-determining locus in *T. snyderi* (A) and *T. vermicularis* (B). (A) To fine map the sex-determining locus in *T. snyderi*, we first used

two markers, f1335 and f1433, that were previously mapped on Chr18 in *T. rubripes* (fugu). We genotyped the two markers in 765 fish from three families (342, 158, and 265 fish in Families A, B, and C, respectively), and identified 30 recombinants. We then genotyped 26 microsatellite markers flanked by the two markers on Chr18 in those individuals. The first row of the table contains the marker names. The rows below show the data of recombinant individuals identified by the screening. "X" and "Y" indicate female- and male-associated alleles, respectively, inherited from the father. Empty blocks indicate that genotypes are not assigned. The results suggest tight linkage of the sex-determining locus with markers near the distal end of Chr18, as well as linkage between the marker loci. Thus, the synteny of this region is conserved between *T. snyderi* and *T. rubripes* (fugu), except for the sex-determining locus. (B) For *T. vermicularis*, 35 markers, most of which have been previously mapped on Chr10 in *T. rubripes*, were used to screen 226 fish in Family D and E. We found 73 fish with recombination between the most distant markers, f471 and f1374. This analysis suggests both tight linkage of the sex-determining locus with the distal end of Chr10 and conserved synteny of this region between *T. vermicularis* and *T. rubripes*, except for the sex-determining locus.

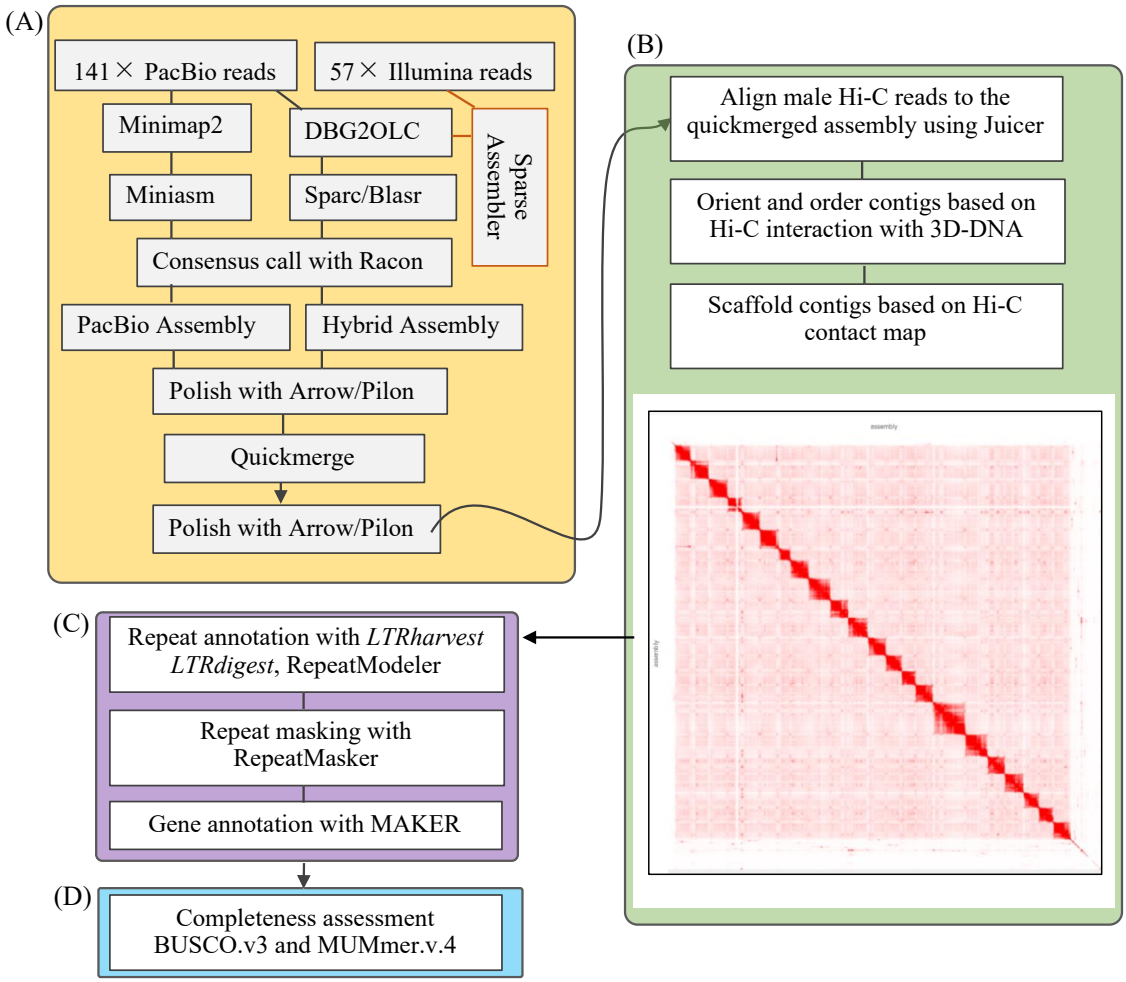


Fig. S2. Schematic overview of the genome assembly process in *T. niphobles* with the YY genotype. (A) Initial genome contigs were generated by two strategies: (1) PacBio long-read-only assembly using Minimap2 and Miniasm, and (2) hybrid assembly using DBG2OLC and SparseAssembler by the combination of PacBio long reads and Illumina short reads. The two assemblies were merged using Quickmerge. (B) The Hi-C reads were aligned to the merged assembly using Juicer. Hi-C scaffolding was performed with 3D-DNA using the Juicer output. (C) Gene annotation was conducted using MAKER with evidence from RNA-seq transcriptomes data, predicted protein sequences from FUGU5, and ab initio gene predictions from SNAP and AUGUSTUS, followed by repeat annotation using *LTRharvest*, *LTRdigest*, and RepeatModeler. Repeat-rich regions were identified by RepeatMasker. (D) Genome completeness was assessed by BUSCO (v. 3.0.2).

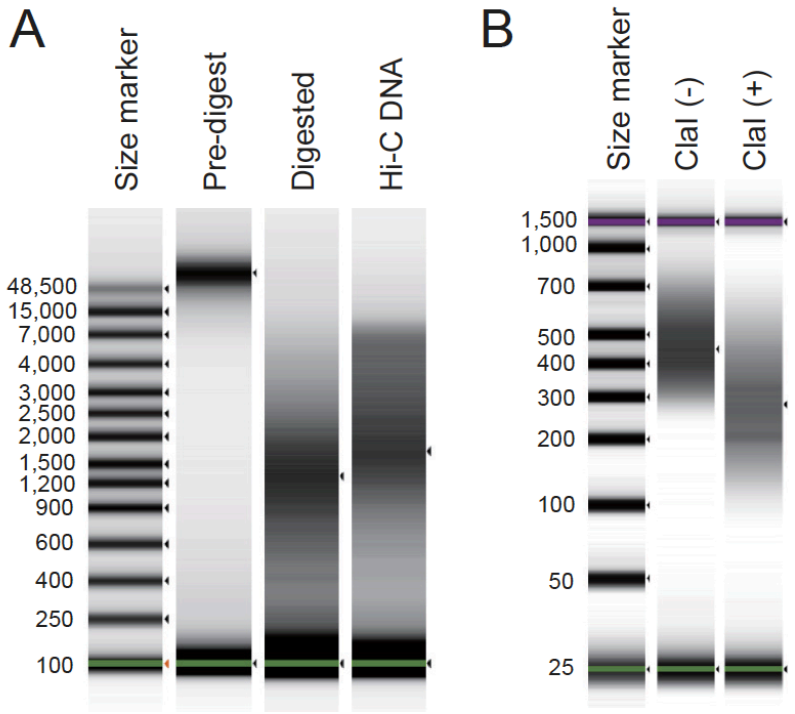


Fig. S3A. Quality control of Hi-C DNA and the Hi-C library.
(A) Size-shift analysis of pre-digested, digested, and ligated DNA (Hi-C DNA) with an Agilent TapeStation using the Genomic DNA ScreenTape assay. (B) Size-shift analysis of the Hi-C library, with or without ClaI restriction enzyme digestion, with an Agilent TapeStation using the High-Sensitivity D1000 ScreenTape assay.

	Pre-HiC	i12000_r2 (with misjoin)	i12000_r0 (without misjoin)
Number of scaffolds	728	1,628	424
Number of scaffolds > 1 K nt	728	1,628	415
Number of scaffolds > 10 K nt	713	1,508	358
Number of scaffolds > 100 K nt	242	215	86
Number of scaffolds > 1 M nt	68	24	23
Number of scaffolds > 10 M nt	8	20	22
Largest scaffold length	15,050,114	24,532,149	29,787,861
N50 scaffold length	6,083,713	13,322,009	16,184,500
Sum of sequence length > 1 M (%)	81.8	77.3	92.8
Sum of sequence length > 10 M (%)	24.9	72	91.2
Number of complete genes	4,340	4,301	4,326
Number of complete + partial genes	4,484	4,467	4,478

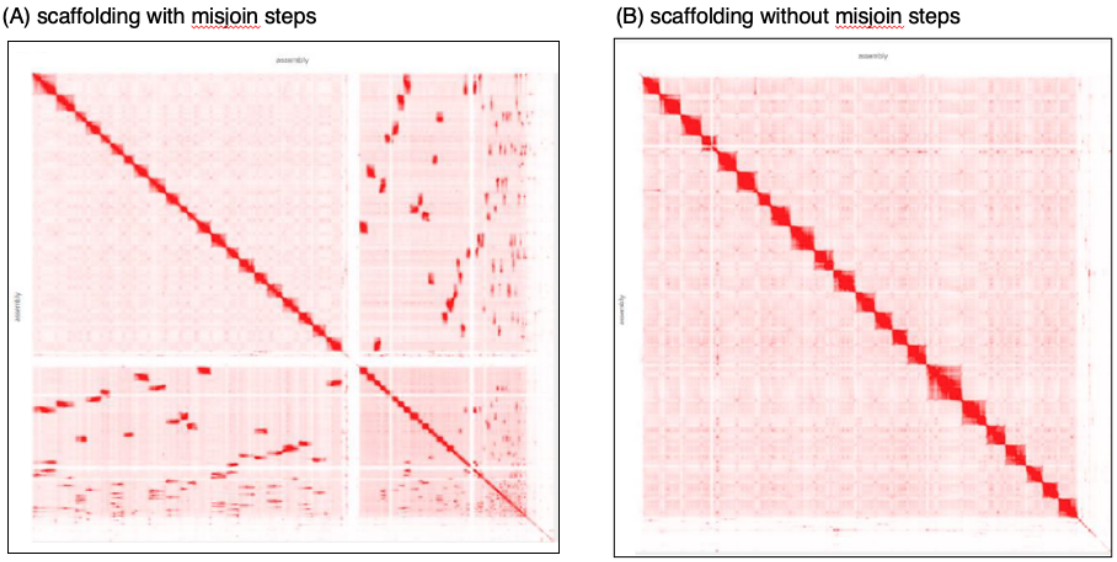


Fig. S3B. Hi-C scaffolding increased N50 lengths and gene space completeness scores.
The table shows a summary of the assemblies before scaffolding, after scaffolding with misjoin steps, and after scaffolding without misjoin steps. Disabling misjoin correction resulted in acceptable chromosome-scale genome sequences, which was validated by the formation of larger and fewer blocks in the Hi-C contact maps (scaffolding with (A) and without (B) misjoin steps), and in an increase in gene space completeness scores (table). In the final assembly, scaffolds of 10 Mb or longer comprise > 90% of the whole genome length, and the number of scaffolds matches the number of chromosomes (n = 22) in *T. niphobles* (1).

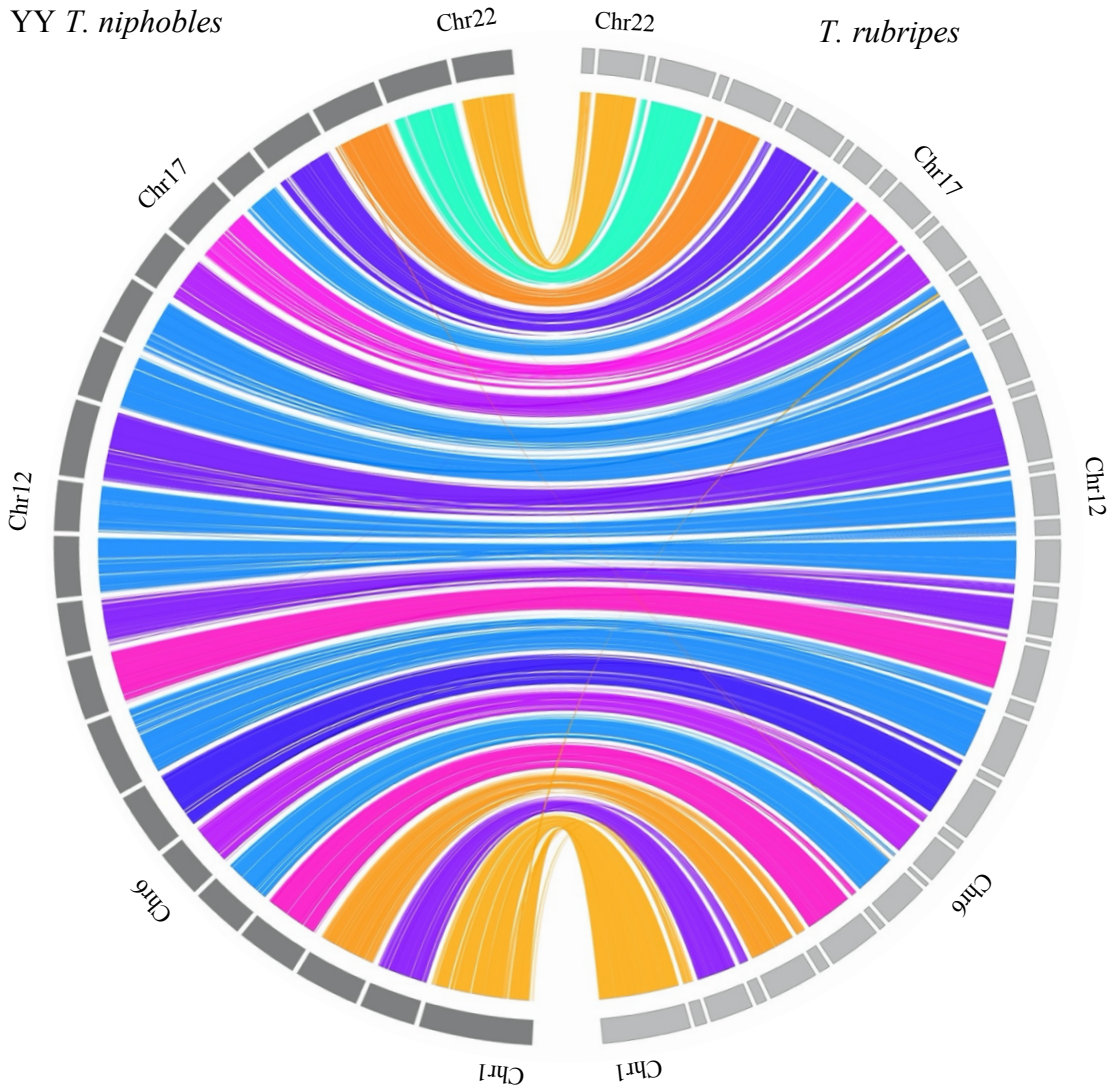


Fig. S4. Conserved synteny between the assembled genome of *T. niphobles* with the YY genotype and a reference genome sequence of fugu, *T. rubripes* (FUGU5).
Based on conserved synteny, chromosome identities for the assembled genome of *T. niphobles* were assigned according to those of *T. rubripes*.

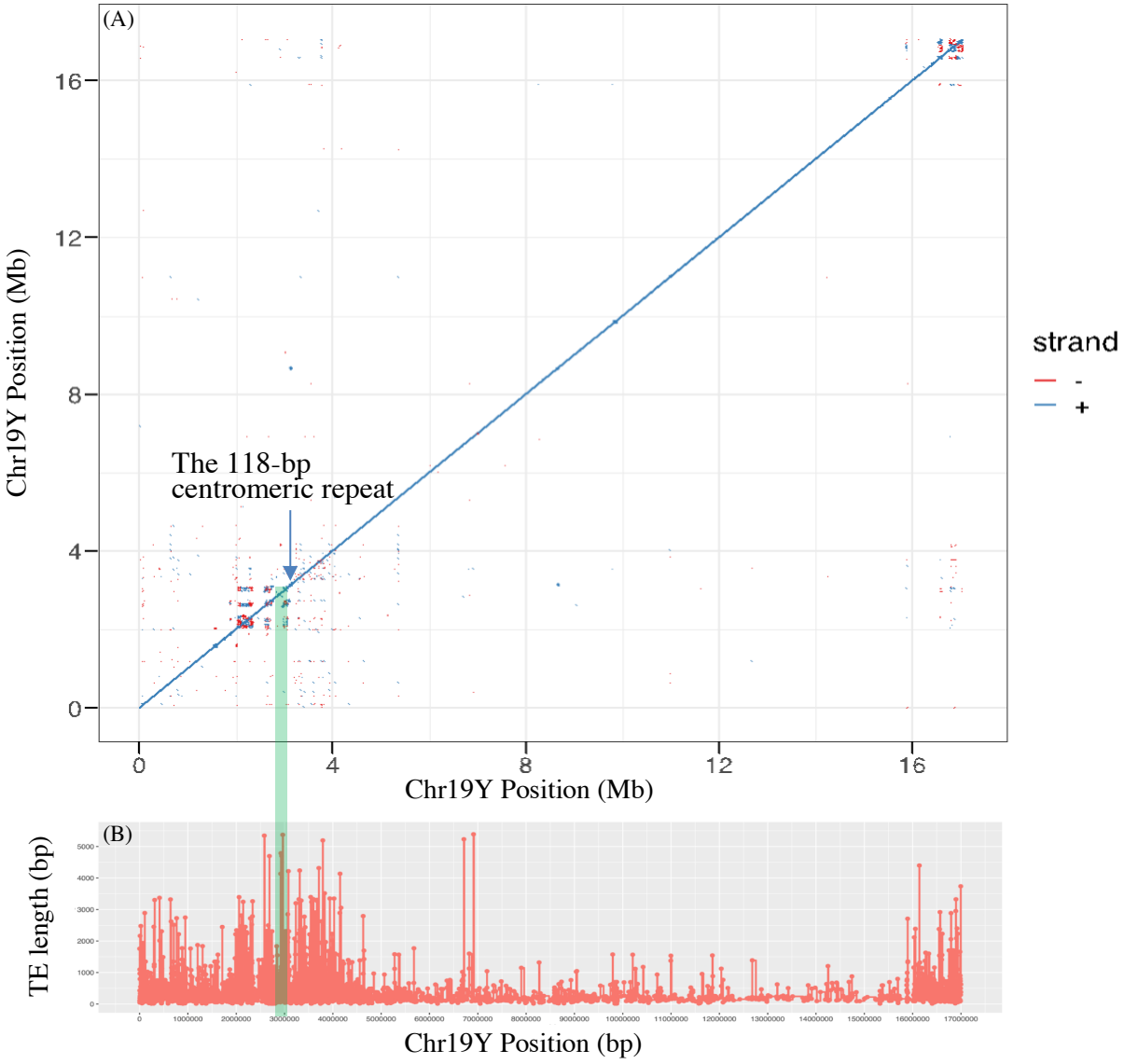


Fig. S5. (A) Dot plot of the self-comparison of Chr19Y.
The presence of direct (blue) and inverted (red) repeats are visualized as the accumulation of dots off the central diagonal line. The blue arrow indicates the accumulation of a centromeric repeat. The green shadow represents the male-specific region (245 kb). Genomic position is scaled in Mb. (B) The TE length distribution across Chr19Y. Y-axis is scaled in bp.

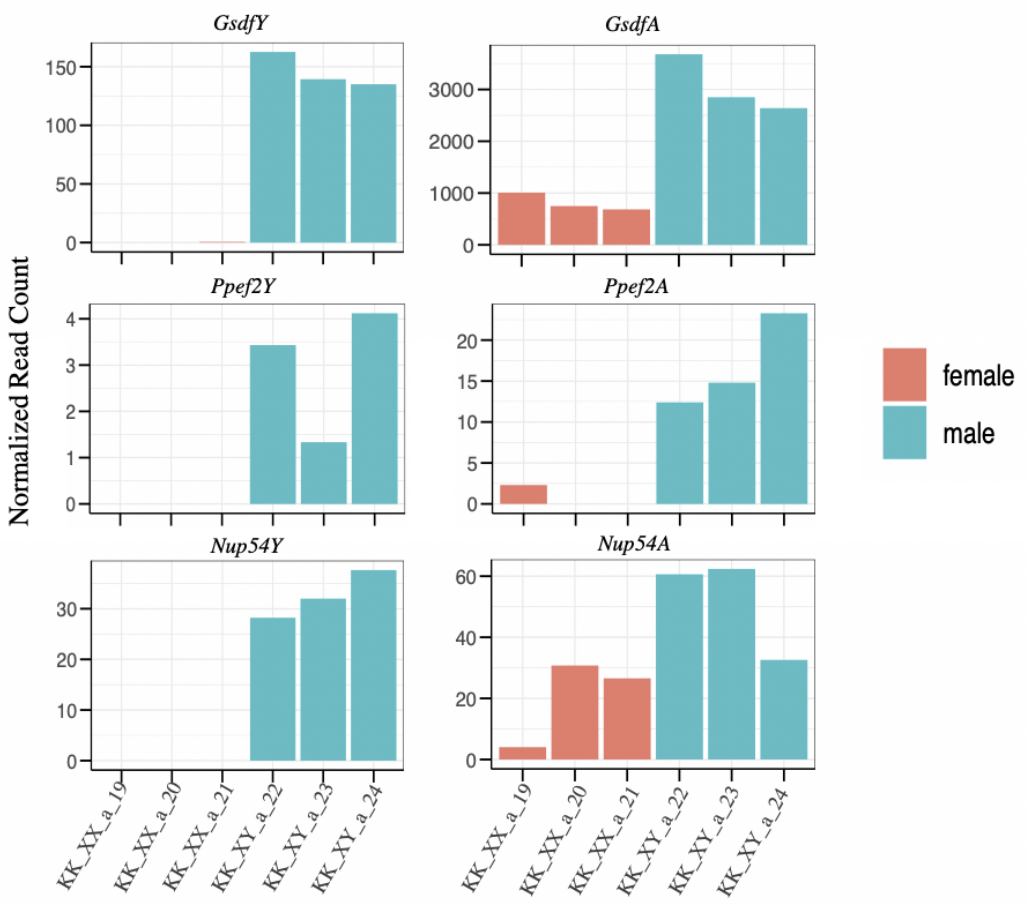


Fig. S6. Expression of the genes in the male-specific region and their autosomal paralogs in the developing gonads.

We first identified diagnostic single-nucleotide sites on the coding region of *Gsdf*, *Ppef2*, and *Nup54* that could distinguish the male-specific and autosomal paralogs. We then aligned RNA-seq reads from developing gonads (90 dpf) on the reference genome of *T. niphobles* and counted the number of reads mapped on the diagnostic sites. The paralog-specific read counts across libraries were normalized for the library size (total number of paired mapped reads) and expressed per 23 M reads. The candidate sex-determining gene *GsdfY* was expressed in the developing gonads in males only. In addition, autosomal *GsdfA* was more highly expressed in males than in females (three libraries each of females and males).

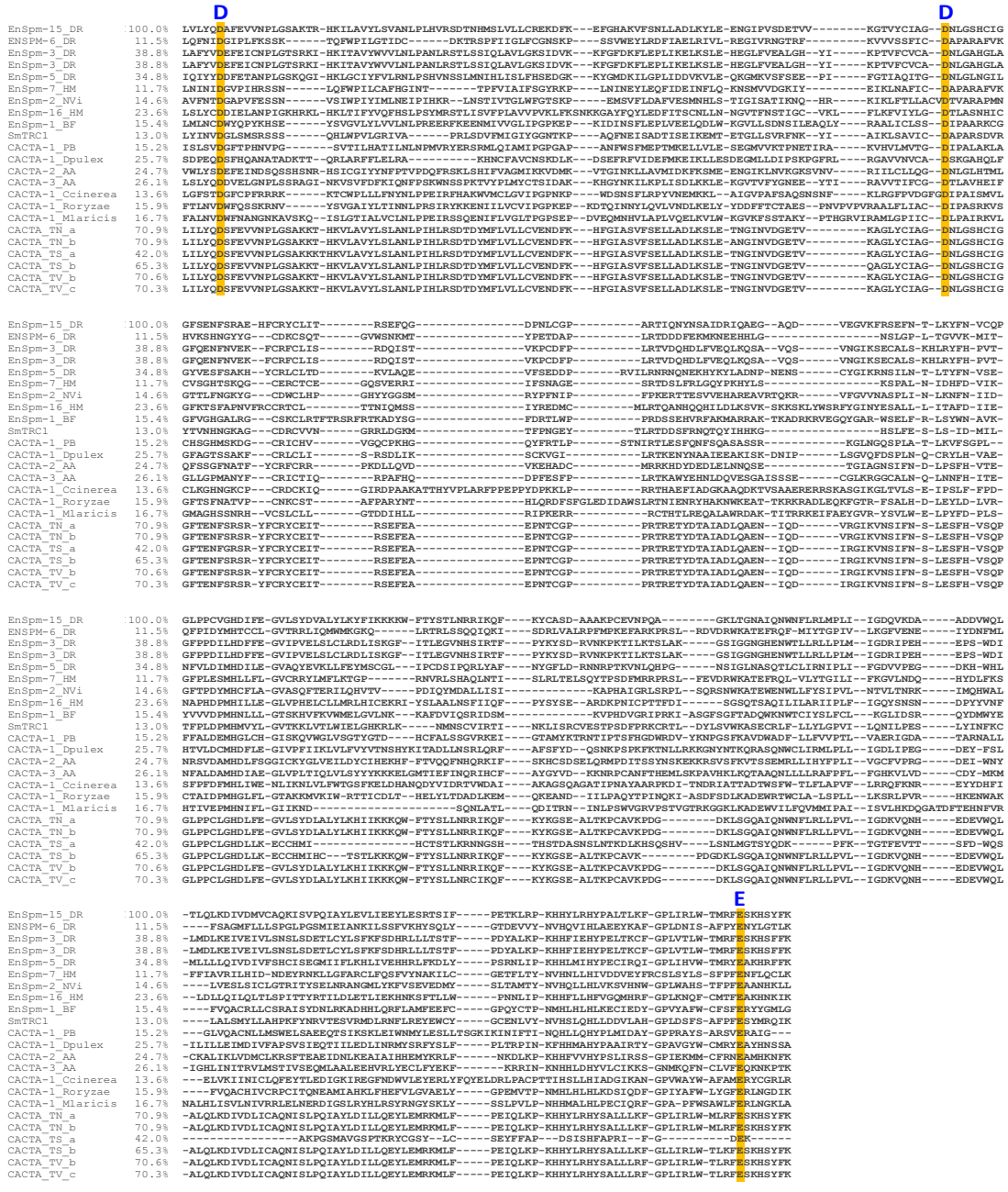


Fig. S7. The catalytic center of the transposase contained in the CACTA transposon.
The DDD/E triad is highlighted in orange in the protein sequence alignment and marked with blue letters according to (2, 3). CACTA_TN_a and CACTA_TN_b in *Takifugu niphobles* are 70% similar to the EnSpm-15_DR in zebrafish in terms of the protein-coding sequences residing within the DDD/E triad. Sequences were aligned by MUSCLE (https://www.ebi.ac.uk/Tools/msa/muscle/) and visualized in Mview (https://www.ebi.ac.uk/Tools/msa/mview/). AA, BF, DR, HM, NVi, PB, TN, TS, and TV represent *Aedes aegypti*, *Branchiostoma floridae*, *Danio rerio*, *Hydra vulgaris*, *Nasonia vitripennis*, *Phycomyces blakesleeanus*, *Takifugu niphobles*, *Takifugu snyderi*, and *Takifugu vermicularis*, respectively. The sequences except for those in *Takifugu* are found in sd01.txt at https://www.pnas.org/doi/10.1073/pnas.1104208108#sec-1.

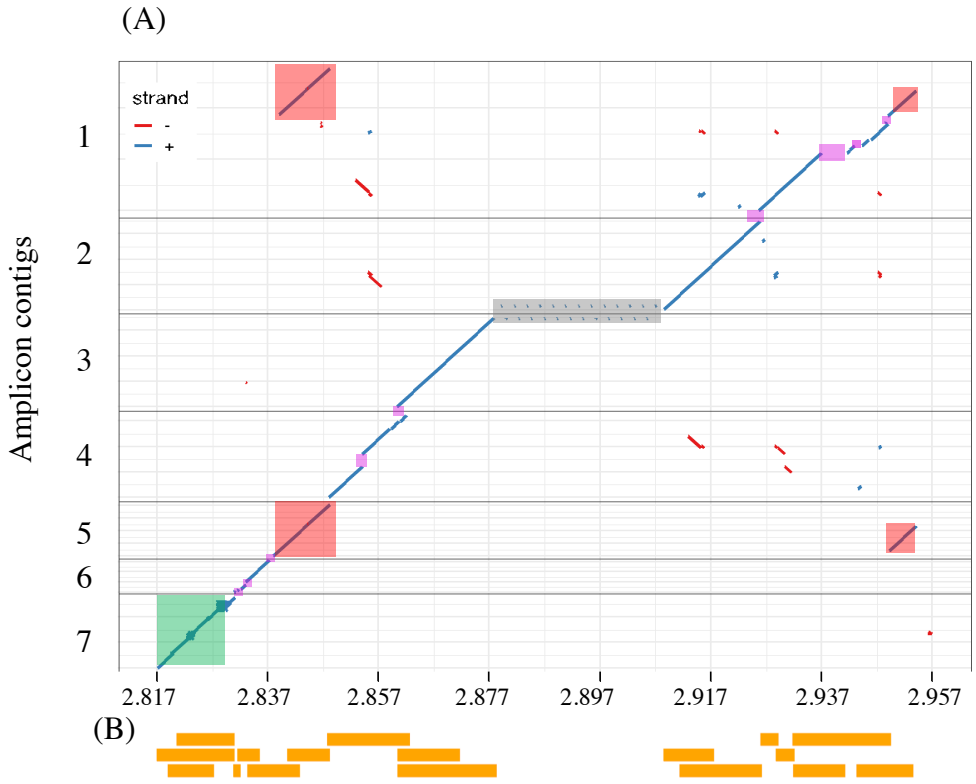
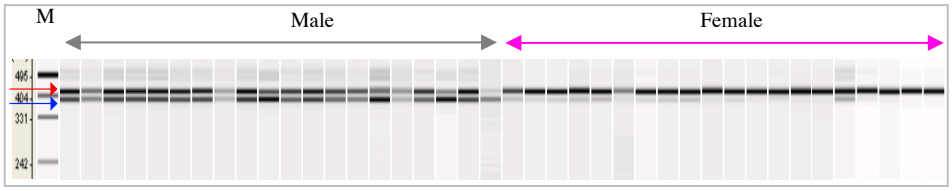
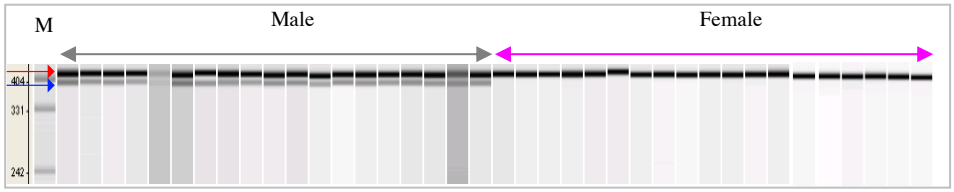


Fig. S8. (A) Comparison between amplicon sequences and their targeted regions spanning the boundary between the pseudoautosomal region and the male-specific region in *T. niphobles*. Amplicon contigs were assembled using the sequences of tiled PCR products determined by Oxford Nanopore sequencing technology. No amplicons were generated in the grey shadow region due to the presence of ~30-kb repetitive sequence at the 3' end of *GsdfY* gene. Disconcordance between amplicon contigs and corresponding sequences on the *T. niphobles* Chr19Y assembly are shown in magenta shadows. Pink shadows indicate duplicated sequences in the male-specific region predicted from the Chr19Y assembly. The green shadow represents the pseudoautosomal region. (B) The predicted lengths of the PCR amplicons based on the Chr19Y assembly.

T. niphobles



T. snyderi



T. vermicularis

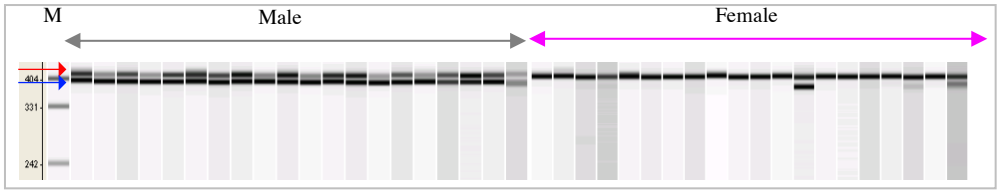


Fig. S9. The presence of the male-specific region in wild populations of *T. niphobles*, *T. snyderi*, and *T. vermicularis*.
Primer pairs that can produce paralog-specific amplicons (single-ended arrows) were used. Genotyping was performed in wild-caught individuals of each species. Amplified PCR products were analyzed on a MultiNA instrument. The lower band is associated with the male phenotype in each species. “M” denotes a set of molecular weight markers.

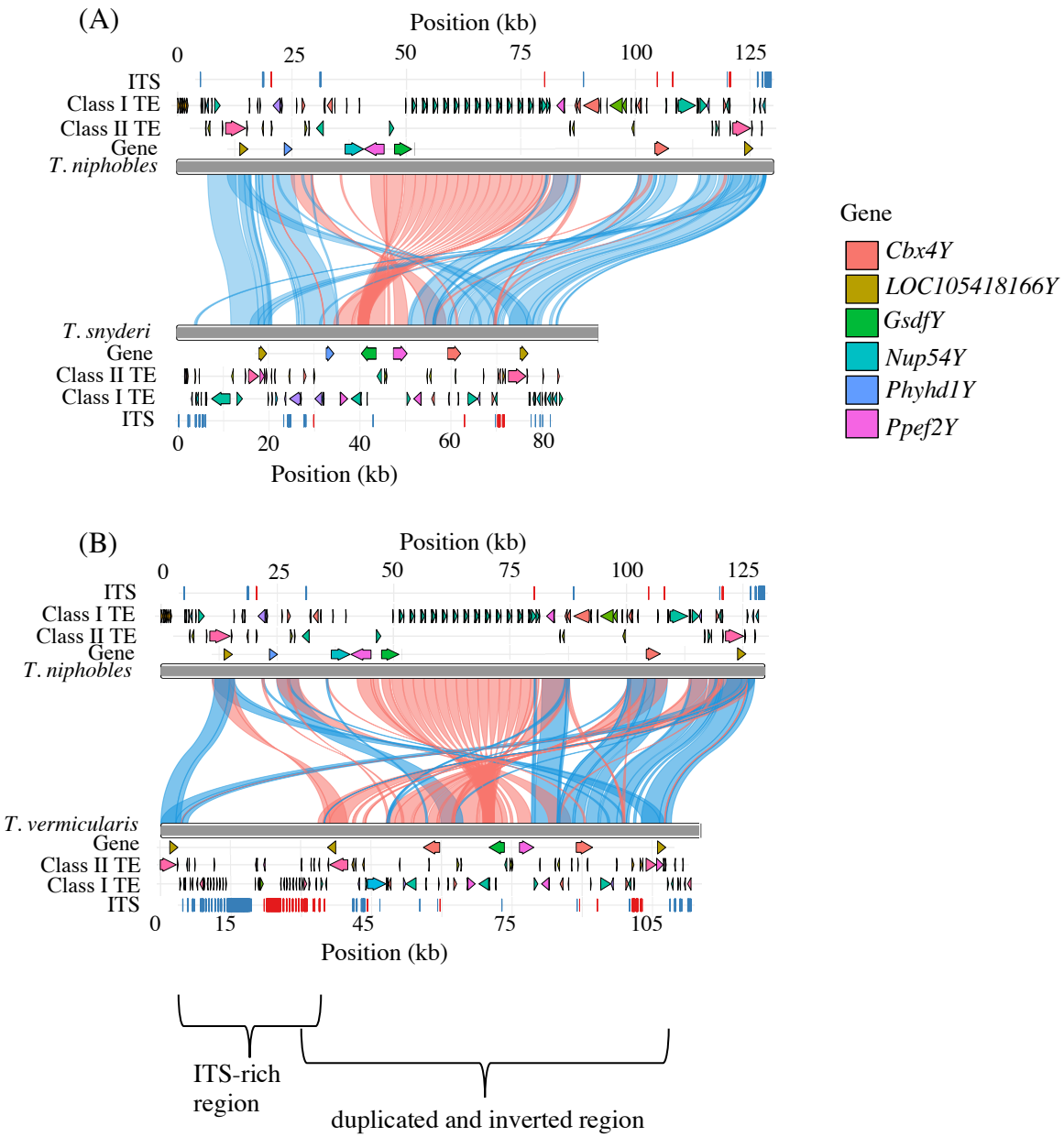


Fig. S10. Schematic representations of the repeat annotation in the male-specific region in *T. niphobles*, *T. snyderi*, and *T. vermicularis*.

Arrows depict Class I transposable elements (TEs), Class II TEs, and full-length genes. Rectangles represent interstitial telomeric sequences (ITSs) $(TTAGGG)_n$ Syntenic and inverted segments are connected by blue and red ribbons, respectively. (A) Comparison of the male-specific region between *T. niphobles* and *T. snyderi*. (B) Comparison of the male-specific region between *T. niphobles* and *T. vermicularis*.

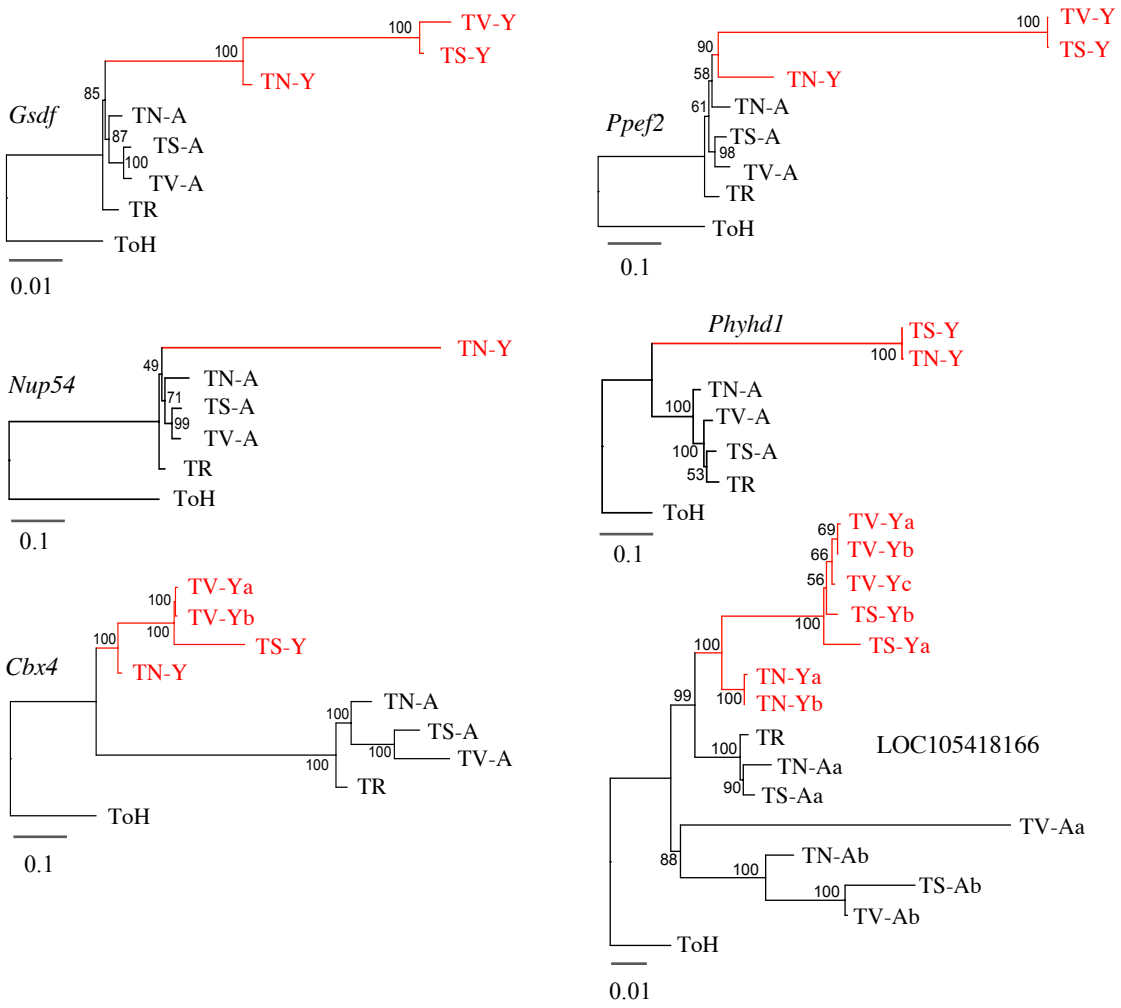


Fig. S11. Maximum-likelihood clustering of the male-specific and autosomal paralogs in *Takifugu niphobles*, *T. snyderi*, and *T. vermicularis* incorporating InDels information.
To reduce the possible effects of long-branch attraction, gaps were treated as fifth states in the multiple sequence alignment. A maximum-likelihood tree was built for each of the genes using RAxML (v. 8.0) with the MULTICAT model, GTR substitution model, and -V option. TR, TN, TS, and TV represent *T. rubripes*, *T. niphobles*, *T. snyderi*, and *T. vermicularis*, respectively. *Torquigener hypselogeneion* (ToH) sequences were used as the outgroup. The reliability of the inferred tree was tested by 1,000 fast bootstrap replicates. Red and black colors represent the male-specific (-Y) and autosomal (-A) sequences, respectively.

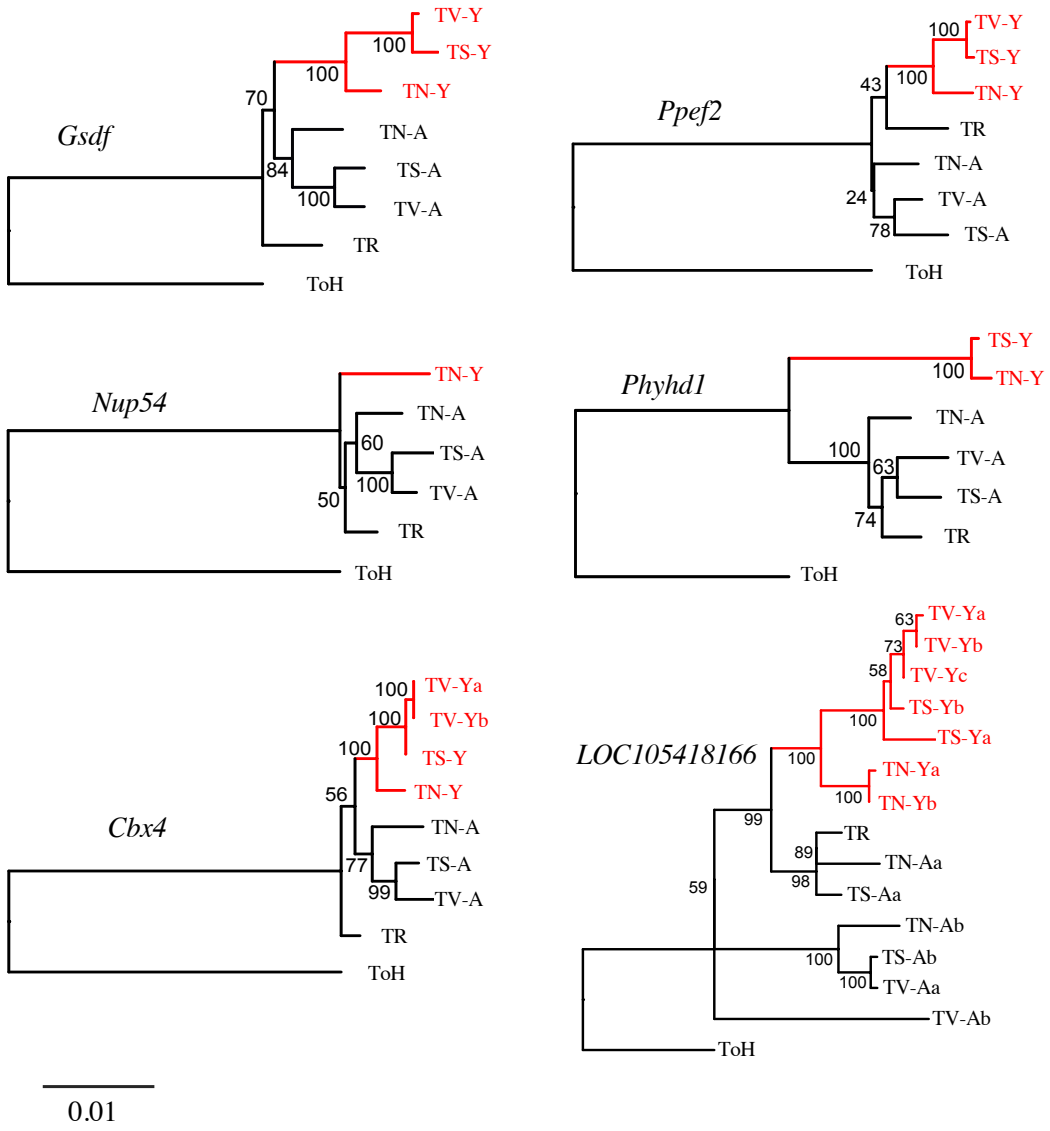


Fig. S12. Maximum-likelihood clustering of the male-specific and autosomal paralogs in *Takifugu niphobles*, *T. snyderi*, and *T. vermicularis*.

In contrast to Fig. S11, gap information was not incorporated in phylogenies. A maximum-likelihood tree was built for each of the genes using RAxML (v. 8.0) with the *GTRGAMMA* model and *--JC69* option. TR, TN, TS, and TV represent *T. rubripes*, *T. niphobles*, *T. snyderi*, and *T. vermicularis*, respectively. *Torquigener hypselogeneion* (ToH) sequences were used as the outgroup. The reliability of the inferred tree was tested by 1,000 fast bootstrap replicates. Red and black colors represent the male-specific (-Y) and autosomal (-A) sequences, respectively.

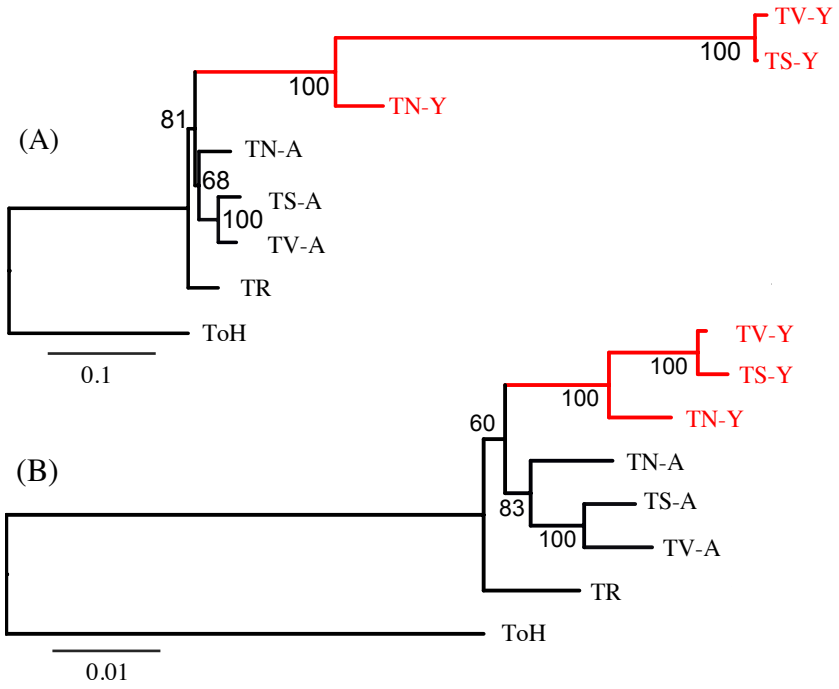


Fig. S13. Maximum-likelihood clustering for the concatenated sequence of the male-specific and autosomal *Gsdf* and *Ppef2* genes in *Takifugu niphobles*, *T. snyderi*, and *T. vermicularis*.
Since segmental duplication encompassing *Gsdf*, *Ppef2*, and *Nup54* on Chr6 likely contributed to the formation of the male-specific sequence in an ancestor of the three species, we concatenated *Gsdf* and *Ppef2* sequences and constructed phylogenetic trees. Note that *Nup54* was excluded since the male-specific paralog of this gene is absent in *T. snyderi* and *T. vermicularis*. (A) The tree incorporating InDels information. We treated gaps as fifth states in the multiple sequence alignment to incorporate InDels information. The tree was built using RAxML (v. 8.0) with the *MULTICAT* model and *GTR* substitution model. (B) The tree without incorporation of the InDels information. The tree was built using RAxML (v. 8.0) with the *GTRGAMMA* model. TR, TN, TS, and TV represent *T. rubripes*, *T. niphobles*, *T. snyderi*, and *T. vermicularis*, respectively. *Torquigener hypselogeneion* (ToH) sequences were used as the outgroup. The reliability of the inferred tree was tested by 1,000 fast bootstrap replicates. Red and black colors represent the male-specific (-Y) and autosomal (-A) sequences, respectively.

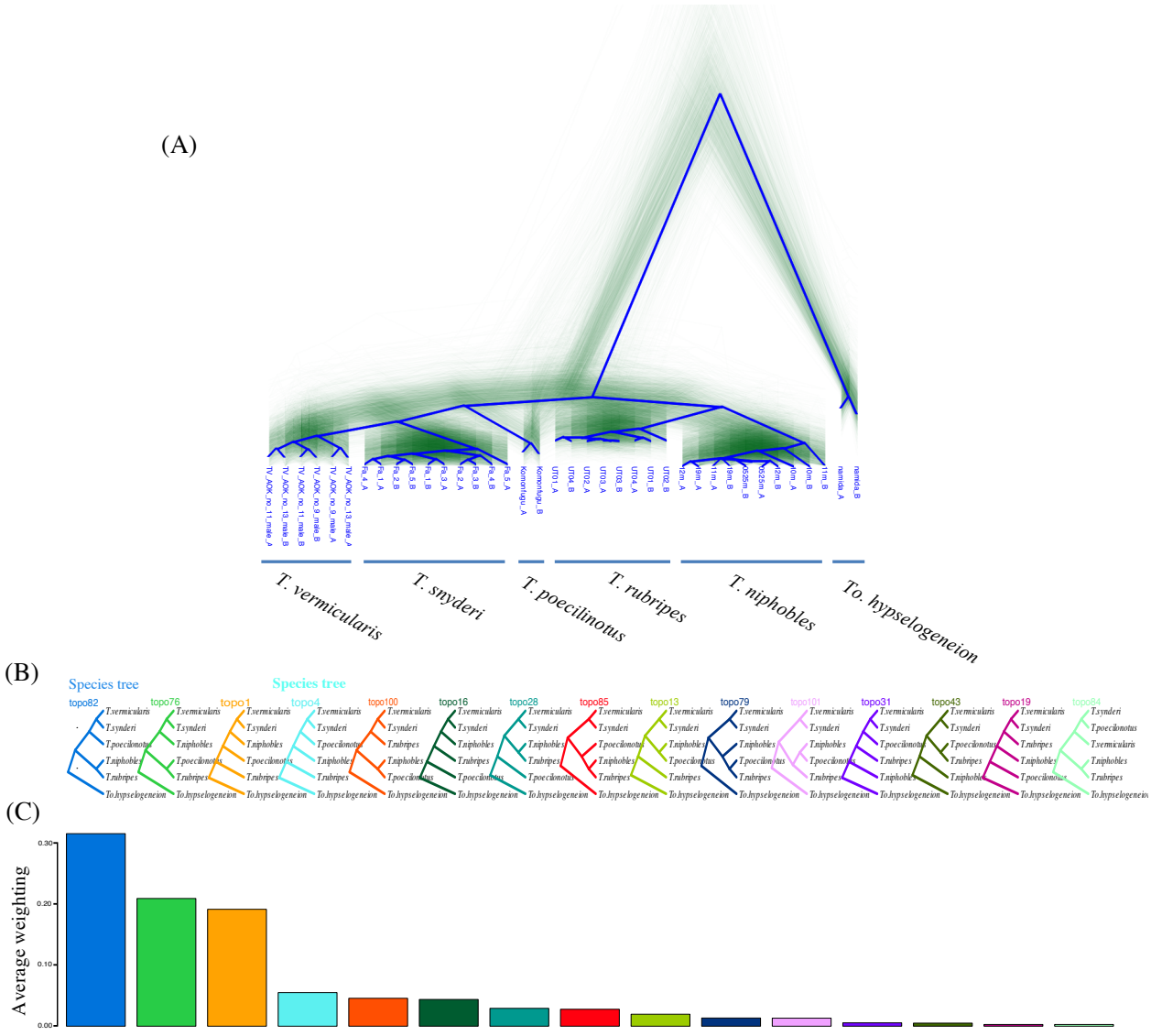


Fig. S14. Topology weighting identifies widespread phylogenetic discordance.
We investigated variations in species relationships across the genomes of *T. rubripes*, *T. niphobles*, *T. snyderi*, *T. vermicularis*, and *T. poecilonotus* in windows of 2,000 SNPs using TWISST, which assessed the weighting of alternative topological relationships among the species. *Torquigener hypselogeneion* was used as the outgroup. (A) Rooted consensus tree. Two thousand subsampled species trees are plotted with the consensus tree. (B) The 15 most abundant potential rooted topologies represent phylogenetic relationships among five in-group taxa. (C) The topology weightings for each of the 15 most abundant topologies in panel B, averaged across 2,000 SNP windows across the genome. The topology weighting indicates that large-scale phylogenetic discordance has shaped the relationships among these species. The 15 most abundant topologies represent 93% of the 105 potential topologies that describe the relationships between the five species. Of the 15 most abundant topologies, the most common are topo82, topo76, topo1, and topo4. The topo82 and topo4 topologies are consistent with the species tree estimated by RAxML (Fig. 1A), in which *T. poecilonotus*, *T. snyderi*, and *T. vermicularis* are monophyletic sister taxa. By contrast, topo76 and topo1 are alternative topologies in which *T. niphobles*, *T. snyderi*, and *T. vermicularis* are monophyletic sister taxa.

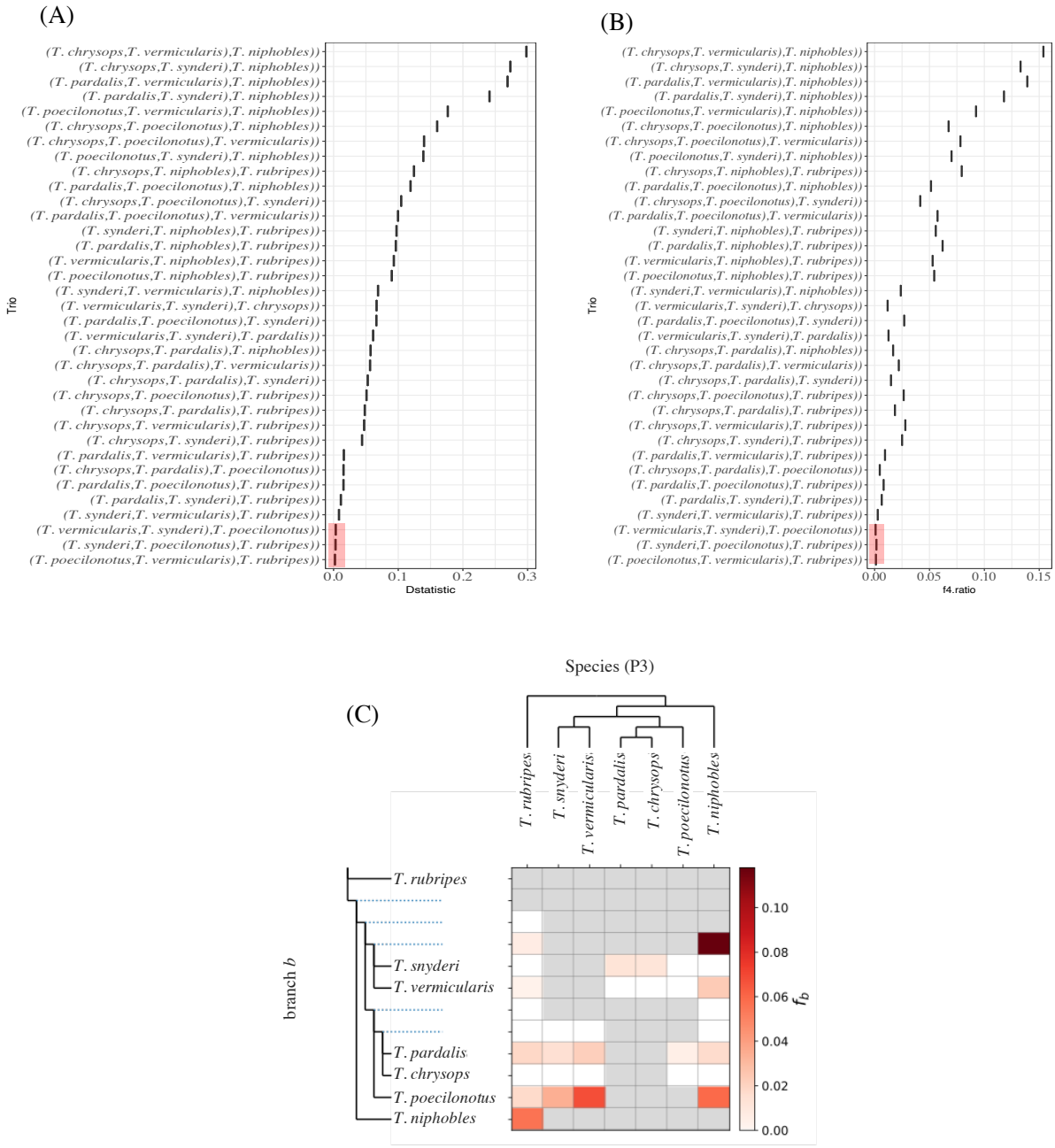


Fig. S15. Past introgression implied by *D* statistics and f_4-ratio statistics.
D statistics and f_4 ratio statics for all taxon trio combinations guided by the species tree (Fig. 1A) of *T. rubripes*, *T. niphobles*, *T. poecilinotus*, *T. snyderi*, *T. pardalis*, *T. chrysops*, and *T. vermicularis*. *Torquigener hypselogeneion* was used as the outgroup for all trio combinations. (A) *D* statistics. (B) f_4-ratio statistics. Red shades in (A) and (B) show the trios with non-significant statistics. The *P*-values for each quartet were corrected for multiple testing using the Bonferroni method. (C) The heatmap represents the scores of *f*-branch statistics showing the excess allele sharing between the species pairs. The f_4-ratio values were mapped to internal branches for the given species tree using the *f*-branch method. Both the *D* statistics and f_4-ratio statistics support past introgression when they significantly differ from zero. The most extreme *D* and f_4 values are

observed in quartets in which *T. chrysops* or *T. pardalis* is in position 1, and *T. niphobles* in position 3, and *T. vermicularis* ($D = 0.29, 0.26$; $f_4 = 0.154, 0.139$) or *T. snyderi* ($D = 0.27, 0.24$; $f_4 = 0.133, 0.118$) in position 2, suggesting a substantial amount of gene sharing between *T. niphoble* and *T. vermicularis*, and *T. niphoble* and *T. snyderi*. Moreover, the result of the *f*-branch test suggested that a strong gene flow occurred in the past between *T. niphobles* and the last common ancestor of *T. snyderi* and *T. vermicularis*.

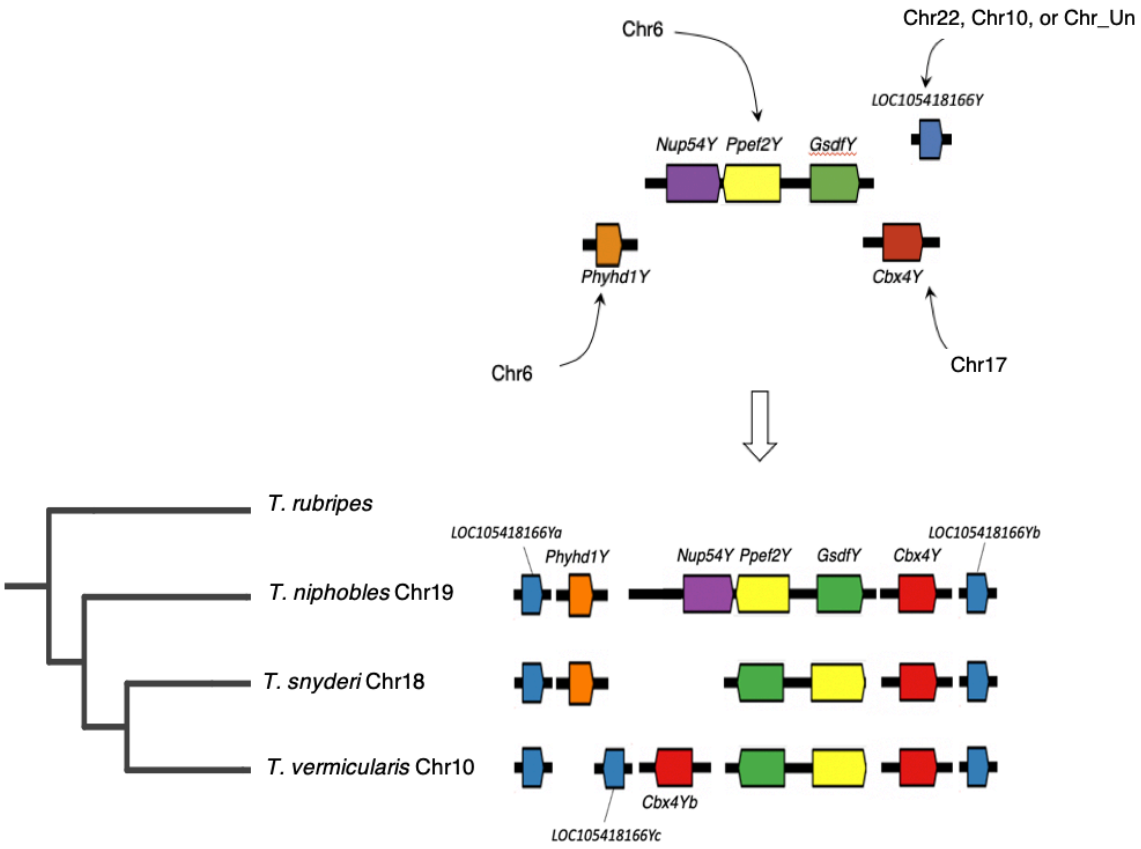


Fig. S16. A hypothetical model for the evolutionary development of the core male-specific region. It is likely that the core male-specific region evolved in the common ancestor of the three species through the combination of at least two processes. One includes segmental duplication of the region encompassing *Nup54*, *Ppef2*, and *Gsdf* and the translocation of this region to the future sex chromosome. The other process is the gathering of other unlinked genes (*Phyhd1*, *Cbx4*, and *LOC105418166*) through independent duplications and translocations. *Nup54Y* was likely lost before the divergence of *T. snyderi* and *T. vermicularis*, while the gene content was retained in the lineage leading to *T. niphobles*. Then, in the *T. vermicularis* lineage, while *Phyhd1Y* was lost, the segmental duplication of part of the male-specific region resulted in the emergence of *LOC105418166Yc* and *Cbx4Yb*.

References for Supplementary Figures.

1. K. Miyaki, O. Tabeta, H. Kayano, Karyotypes in six species of pufferfishes genus Takifugu (Tetraodontidae, Tetraodontiformes). *Fish. Sci.* **61,** 594-598 (1995).

2. Y. W. Yuan, S. R. Wessler, The catalytic domain of all eukaryotic cut-and-paste transposase superfamilies. *Proc. Natl. Acad. Sci. U. S. A.* **108**, 7884–7889 (2011).

3. V. V. Kapitonov, J. Jurka, A universal classification of eukaryotic transposable elements implemented in Repbase. *Nat. Rev. Genet.* **9,** 411-412 (2008).

Supplementary text

Supplementary Materials and Methods

1. Linkage mapping

1-1. Genotyping of SNP markers in *T. snyderi*

Reads were trimmed using Trimmomatic (v. 0.36) (1) as described in (2), and mapped onto a reference genome sequence of *T. rubripes* (fugu), FUGU5 (3) (GenBank: GCA_000180615.2) using BWA-MEM (4) with the default parameters. SNP calling was performed using GATK HaplotypeCaller (v. 3.8) (5) with the default setting. VCFtools (v. 0.1.14) (6) was used to extract SNPs meeting the following criteria: minor allele frequency between 0.1 and 0.4, genotyped for ≥ 90% of individuals, allele count of two, and minimum depth of 10. The SNP loci heterozygous in both parents were removed. The sequence data were registered in the DDBJ SRA (Accession No. DRA012888) The genotype information is provided in SI (mapping genotype.xlsx).

1-2. Linkage map construction and mapping of the sex-determining locus in *T. snyderi*

The male and female linkage maps were constructed based on the paternally and maternally inherited alleles at marker loci, respectively, using the R/qtl package (v. 1.41) (7). Linkage groups were inferred with the *formLinkageGroups* function (max.rf = 0.45, min.lod = 6.0), and the markers were ordered by the *orderMarkers* function. Linkage groups with fewer than five loci were excluded from the subsequent analysis. The sex-determining locus was analyzed by the interval mapping implemented in R/qtl as described in (8). The genome-wide significance level was determined with a permutation test with 10,000 permutations (9). The chromosome and linkage group identities of *T. snyderi* were assigned based on synteny with those of *T. rubripes* established in (3).

2. Phylogenetic framework

2-1. Genome sequencing

In order to generate a species phylogenetic tree, whole-genome resequencing data were obtained from 12 *Takifugu* species and one outgroup (one individual per species): *Takifugu rubripes*, *T. snyderi*, *T. vermicularis*, *T. niphobles*, *T. pardalis*, *T. poecilonotus*, *T. chrysops*, *T. stictonotus*, *T. obscurus*, *T. ocellatus*, *T. xanthopterus*, *T. porphyreus*, and *Torquigener hypselogeneion* (outgroup). Genomic DNA was extracted from the caudal fin using the Gentra Puregene tissue kit (Qiagen). The TruSeq DNA PCR-free Kit (Illumina) was used to construct the libraries (one library per sample) (150-, 100-, or 90-bp paired-end reads). Library preparation and whole-genome sequencing were performed with HiSeq 2000, HiSeq 2500, or HiSeq X Ten platforms at the NODAI genome research center, National Institute of Genetics, or BGI Corporation. Summary statistics of sequencing data are described in Table S4. The data have been registered in the DDBJ SRA (Accession No. DRA012890 and

DRA012891). Whole-genome resequencing data of *T. rubripes* were obtained from (10) (Accession No. DRA007464 in DDBJ SRA).

2-2. Genotyping and construction of species tree

Reads were trimmed with Trimmomatic (v. 0.36) (1) using the following parameters: ILLUMINACLIP TruSeq3-PE-2.fa:2:30:10, LEADING:10, TRAILING:10, SLIDING WINDOW:30:20, AVGQUAL:20, and MINLEN:80, MINLEN:90, or MINLEN:140. Then, the read pairs that survived at both paired ends were mapped onto the FUGU5 assembly using BWA-MEM (4) with the default parameters. We marked the PCR duplicated reads and filtered out secondary reads using SAMtools (v. 1.9) (11). SNP calling from mapped reads was done using *Freebayes* (v. 1.3.1) (12) with the following settings: min-mapping-quality = 30, use-best-n-alleles = 4, min-alternate-count = 2, min-alternate-fraction = 0.2, min-coverage = 4, and ploidy = 2. Variant filtering was done by *vcffilter* from *vcflib* (https://github.com/vcflib/vcflib) using the following parameters: QUAL / AO > 20 & SAF > 0 & SAR > 0 & RPR > 1 & RPL > 1. InDels were excluded from all individuals. Variants residing in the gaps, repeats, and non-oriented scaffolds in FUGU5 were subsequently eliminated. We defined "divergent sites" as the loci where genotypes were homozygous for the derived alleles. These sites (12,983,719 in total) were concatenated and subjected to a phylogenetic analysis using RAxML (v. 8.0) (13) with the *ASC_GTRCAT* model and *–V* option. The reliability of the inferred tree was tested by 1,000 fast bootstrap replicates.

3. *k*-mer analysis of *T. niphobles*, *T. snyderi*, and *T. vermicularis*

3-1. Samples and resequencing

For the *k*-mer analysis, we sequenced six females and five males of *T. niphobles* from Lake Hamana (Shizuoka Prefecture, Japan), three pools (eight individuals per pool) each of male and female *T. snyderi* from Suruga Bay, and four females and four males of *T. vermicularis* from Ariake Sea (Table S5). Phenotypic sex was visually determined under a microscope. Genomic DNA extraction and library preparation were carried out as described above. Sequencing was performed using HiSeq 2000, HiSeq 2500, or HiSeq X Ten by BGI Corporation or Macrogen Corporation. A summary of the sequencing statistics is presented in Table S5. (Accession No. DRA012877, DRA012889, DRA012878 in DDBJ SRA).

3-2. *k*-mer counting

We counted 35-mer occurrences in each individual (*T. niphobles* and *T. vermicularis*) or each sample pool (*T. snyderi*) using Jellyfish (v. 2.2.6) (14), followed by quality trimming with Trimmomatic (v. 0.36) (1) as described above except for the use of the option MINLEN:140. The "dump" subcommand

was executed to output all 35-mers for each of the samples/pools. The male-specific 35-mers were extracted by comparing females and males. Reads containing the male-specific 35-mers and their read pairs were extracted from the original FASTQ files and assembled into contigs by MetaPlatanus (v. 1.2.2) (15) with a default setting. Contigs assembled with depths of coverage ranging from 9 to 200 (*T. niphobles*), 18 to 500 (*T. snyderi*), or 9 to 300 (*T. vermicularis*) were retained.

3-3. Annotation of the male-specific sequences obtained by the *k*-mer approach

To characterize the male-specific sequence in *T. niphobles*, *T. snyderi*, and *T. vermicularis*, we first took advantage of a well-annotated fugu genome (FUGU5). We masked repetitive sequences in the contigs using a FUGU5 repeat annotation database, and conducted a BLASTn search (16, 17) of the contigs against FUGU5 was conducted with a word_size of 50. We retained the genes meeting one or more of the following criteria: genes on which more than two contigs hit, genes on which two contigs covering at least 40% of the sequence hit, or genes on which a single contig covering at least 50% of the sequence hit. Contigs hitting these genes were then reassembled using the *CAP* option (Contig Assembly Program) of BioEdit (v. 7.2.5) (18).

4. Genome assembly in a *T. niphobles* YY male

4-1. Hybrid genome assembly and Hi-C scaffolding of the *T. niphobles* genome

Illumina short reads and PacBio long reads of the YY male were deposited in DDBJ SRA under the accession number DRA012992. Illumina reads were trimmed with Trimmomatic (v. 0.36) (1) using the following parameters: ILLUMINACLIP TruSeq3-PE-2.fa:2:30:10, LEADING:10, TRAILING:10, SLIDING WINDOW:30:20, AVGQUAL:20, and MINLEN:240, and the resulting reads were assembled with SparseAssembler (v. 20160205) (19) into contigs with k-mer size = 51 and skip length = 15. The contiguity of the short-read assembly was improved by anchoring the contigs to the PacBio long reads using DBG2OLC (v. 20180222) (20) with *k*-mer size = 17, minimum overlap = 30, and AdaptiveTh = 0.01. After removing chimeric reads with the option of *ChimeraTh* = 1, we obtained an initial hybrid assembly. Because raw PacBio long reads are prone to contain errors, we corrected potential sequencing errors in the initial assembly by realigning the long reads and the Illumina contigs using Sparc (v. 20160205) (21). We also aligned the long reads to the assembly with pbalign (v. 0.4.1) (https://github.com/PacificBiosciences/pbalign/) using the BLASR (v. 5.3.1) (22) algorithm and polished it with Arrow (v. 2.3.3) with the –minCoverage 12 option (23). For error correction at the base level, Illumina reads were aligned to the assembly using BWA-MEM (4), and the assembly was polished three times with Pilon (v. 1.22) (24). For long-read-only assembly, the Minimap2/Miniasm pipeline (25, 26) was used. In brief, a Pairwise mApping Format (PAF) file was generated using Minimap2 with the -x ava-pb option. The PAF file was converted into an assembly graph as a

Graphical Fragment Assembly (GFA) format file using Miniasm. Then, a *de novo* assembly sequence (unitig) was extracted from the GFA file. The long reads were mapped to the unitig sequences using Minimap2, and three rounds of consensus corrections were performed using Racon (v. 1.3.0) (27). The resultant assembly was polished three times with Pilon as described above. Finally, the hybrid assembly and long-read-only assembly were merged using Quickmerge (v. 0.2) (28), employing the latter as the query assembly. The merged assembly was then merged to the long-read-only assembly to produce a more contiguous assembly. Furthermore, Hi-C data from an XY male individual were used to scaffold this merged genome. An overview of the genome assembly is shown in Fig. S2.

4-2. Tissue collection, Hi-C sequencing, and Hi-C scaffolding of the *T. niphobles* genome

The liver tissue was dissected from an XY male *T. niphobles* and frozen in liquid nitrogen. Sample fixation, chromatin isolation, Hi-C library preparation, and post-sequencing quality control were performed following the iconHi-C protocol using the DpnII restriction enzyme (29) (Fig. S3A). Quality control of the post-ligated DNA (Hi-C DNA) and the Hi-C library was performed based on the DNA size shift as described in (29) (Fig. S3A). Sequencing of the Hi-C library was performed on an Illumina HiSeq X Ten platform with 151-bp paired-end reads, yielding approximately 92 million read-pairs. Post-sequencing quality control of the Hi-C library was performed by HiC-Pro (v. 2.11.1) (30) using one million subsampled read-pairs from the large-scale sequencing data (Table S10). After removing low-quality and adapter sequences using Trim Galore (v. 0.6.0) (https://www.bioinformatics.babraham.ac.uk/projects/trim_galore/), the trimmed reads were mapped to the Quickmerged assembly using Juicer (v. 20190226) (31). Using the Juicer output, Hi-C scaffolding was performed using 3D-DNA (v. 20180929) (32) without the misjoin-correction step (i = 12,000 and r = 0) (Fig. S3B). The Hi-C dataset was deposited in DDBJ SRA under the accession number DRA012992. The resultant scaffolded assembly was used for the following analysis.

4-3. Gene annotation of the *T. niphobles* genome

Gene annotation was conducted using MAKER (v. 3.01.02) (33), which utilizes both evidence-based methods and ab initio predictions. For the evidence-based annotation, the RNA-seq data of XY *T. niphobles* and the annotated protein sequences from FUGU5 were used with default parameters. RNA-seq data of XY *T. niphobles* were generated on an Illumina HiSeq 2000 platform at Novogene (Tokyo, Japan). Sample collection, RNA extraction, library preparation, and sequencing strategies are described in Section 5. RNA-seq in *T. niphobles*. The ab initio annotation was performed using SNAP (34), trained on the gene model from the evidence-based annotation, and AUGUSTUS (35), trained on the Actinopterygii gene model (36) from BUSCO (v. 3.0.2). The curated libraries of repeats for *T.*

niphobles were also used for the annotation (see details of the repeat annotation in Section 6. Repeat annotation of the *T. niphobles* genome).

4-4. Repeat annotation of the *T. niphobles* genome

To identify repeat families in the assembly, a de novo repeat library was constructed using RepeatModeler (v. 1.0.11) (37) with RECON and RepeatScout (38, 39). Full-length, long terminal repeat retrotransposons were identified using *LTRharves*t and *LTRdigest* (40, 41). MITE-Hunter was used to identify the miniature inverted-repeat transposable elements (42). Tandem Repeats Finder was used to identify the simple, low-complexity, and satellite repeats (43). The annotated repeat sequences from zebrafish in Repbase (44) were also used. With these repeat libraries, we identified the repeat region in the de novo genome assembly using RepeatMasker (v. 4.0.7) (http://www.repeatmasker.org).

4-5. Assembly quality assessment

The quality of the assemblies was assessed by gVolante (v. 1.2.1) (45) with BUSCO (v. 3.0.2) (36) and the ortholog gene set of Actinopterygii. In addition, the quality of the Hi-C–based proximity-guided assembly was evaluated by mapping the Illumina short reads (sample IDs: 19m, 12m, 10m, 11m, and 0525m in Table S5) to the assembled genome using BWA-MEM with default parameters. A comparison of this genome assembly with that of *T. rubripes* (FUGU5) was also conducted (see details in the below.).

4-6. Comparison of genome assemblies of *T. niphobles* and *T. rubripes*

We examined the level of conserved synteny between *T. niphobles* and *T. rubripes* (FUGU5) using the nucmer algorithm implemented in MUMmer (v. 4.0) (46) with option -l 100 -c 100. Synteny blocks longer than 5 kb with 95% sequence identity were visualized by Circos Plot using circos (v. 0.69-7) (47). Based on the conserved synteny between the two species and the genomic positions of the sex-linked markers reported previously (8), we identified the pseudochromosome corresponding to the sex chromosome of *T. niphobles* and denoted it Chr19Y. To find the centromere on Chr19Y, the following centromeric repeat reported in *T. rubripes* (3) was BLAST searched against the genome assembly of *T. niphobles*:

"ACGAGAAAACGTCAAAAACGTCATAATGTGAGCGCAGCATGAGTTTTCAGGTGATCATG
TTGAATTTACCTCTGTTTTGAGAAACTTGTATATCCTGACCAAAAGTGATGGTTTCCCC."

5. RNA-seq in *T. niphobles*

5-1. RNA extraction and sequencing

To obtain transcriptomic data from the differentiating gonads of *T. niphobles*, we produced one family from a pair of wild-caught parents collected from Lake Hamana. Conditions for rearing and feeding fish were set as above. The genotypic sex of siblings was inferred by sex-linked microsatellite markers (f1413, fi-4, f1595, and f645) (8). Fish samples at 90 dpf were dissected, and gonads were preserved in RNA*later* solution (Qiagen). The observation of sectioned gonads stained with hematoxylin and eosin suggested that the gonads were in an early stage of morphological differentiation of the ovary and testis. Total RNA from the gonads was extracted with TRIzol (Invitrogen). TURBO (Thermo Fisher Scientific) was used to remove the genomic DNA from total RNA. Because a sufficient amount of total RNA was not obtained from the gonads, the same amount of total RNA from three individuals with the same genotypic sex was pooled for each library preparation. The average body lengths of the XX and XY fish were 30.8 mm (SD = 1.5 mm, n = 9) and 28.5 mm (SD = 1.4 mm, n = 9), respectively. Library preparation and paired-end sequencing (~4 Gb/pool) were performed on an Illumina HiSeq 2000 platform at Novogene. Summary of the RNA-seq data is presented in Table S14. Raw reads were trimmed by removing adapters and low-quality reads as described above except for the use of the option MINLEN:140. Then all sequences were aligned to the reference *T. niphobles* genome sequence using HISAT2 (v. 2.0.4) with the options of -k 1, --no-mixed, and --no-discordant (48). The data were registered in the DDBJ SRA (Accession No. DRA012876).

5-2. Expression of male-specific genes in the developing gonads

The expression of the male-specific paralogs in the developing gonads was examined from the RNA-seq data obtained above. For this purpose, the diagnostic nucleotide sites that can distinguish transcripts from the male-specific sequences and their autosomal paralogs were identified. This was done manually on IGV (v. 2.11.1) (49). Then, the number of reads with and without the diagnostic sites were counted from the HISAT2 output file (BAM file) using the AESReadCounter function implemented in GATK4 (5). The paralog-specific read counts were normalized for the total number of paired mapped reads in each library and expressed per 23 M reads. Because the diagnostic sites on the coding region of the paralog pair were only found in *Gsdf*, *Ppef2*, and *Nup54*, the paralog-specific expressions were not assessed for *Phyhd1Y*, *Cbx4Y*, *LOC105418166Y*, *Cyc1*, *Hipk1a*, *Hipk1b*, *Hipk1c*, *Hipk1d* or *Ffra2*.

6. Characterization of the male-specific region

6-1. Samples and additional resequencing of *T. niphobles*

For a comparison of the relative depth of read coverage between males and females, we sequenced an additional two females and four males of *T. niphobles* from Lake Hamana. Phenotypic sexing and genomic DNA extraction were carried out as described above. Library preparation and sequencing on

HiSeq 2000 platform were performed by BGI Corporation (Kobe, Japan). Information about the sequencing statistics is presented in Table S5. These reads were trimmed by removing adapters and low-quality reads as described above, with either the MINLEN:80 (m8, m, m30, and m31) or MINLEN:140 (35fe and 11fe) option. The data were registered in the DDBJ SRA (Accession No. DRA012877).

6-2. Comparison of the relative depth of read coverage between males and females

To determine the male-specific region in the genome assembly of the *T. niphobles* YY male, we mapped the resequencing data of 10 males and 8 females (Table S5) onto the *T. niphobles* reference sequence using BWA-MEM (4) with default settings, and compared the relative depth of coverage between sexes. To count the depth, we used Mosdepth (v. 0.3.0) (50) with a 1-kb window size and excluded regions with depth greater than 70x. Mean coverage values were calculated separately for females and males, after normalization of each sample with the genome coverage. The male-to-female coverage ratio was calculated as (average male coverage+1) / (average female coverage+1).

6-3. Characterization of the male-specific region

To identify duplicate and satellite sequences within the male-specific region, the sequences of the male-specific region were aligned against themselves using BLASTn with the "word_size 50" option. To test if the male-specific region comprised segments that were duplicated and translocated from other regions of the genome, we first masked repetitive sequences from the two male-specific regions using the repeat database of *T. niphobles* YY genome assembly. We then performed a BLAST search of the male-specific region against the reference genome of *T. niphobles*, and then aligned the male-specific region to the genome assembly of *T. niphobles* using the nucmer algorithm with option -l 50 -c 65 implemented in MUMmer (v. 4.0) (46). We restricted our analyses to the genic regions and visualized synteny blocks longer than 300-bp with 90% sequence identity.

6-4. Verification of the assembled male-specific sequence of *T. niphobles* by short- and long-range tiling PCR

To verify the overall accuracy of the assembly for the male-specific region and its adjacent pseudoautosomal regions, we used tiling-path PCR combining long- and short-range PCRs. We first designed primers for short-range PCRs around the genomic position Chr19Y: 2.828 Mb to 2.955 Mb (Table S15). No PCR primers were constructed for the region from 2.879 Mb to 2.909 Mb due to the presence of large repeated sequences at the 3' ends of *GsdfY* gene. The region from 2.956 Mb to 3.118 Mb was also highly repetitive; therefore, it was not possible to design primers for this region either. PCRs were carried out in a total reaction volume of 15 μl containing 0.3 μl (10 μM) of each primer,

1.5 µl of Takara 10X PCR buffer (Mg2+ plus), 1.2 µl of 2.5 mM dNTP mixture (Takara Bio), 0.15 µl of Taq DNA Polymerase, 10.55 µl DW and 1 µl (~30 ng) of genomic DNA. PCR reactions were operated on a BIO-RAD T100 thermal cycler (Bio-Rad), under the following thermal conditions: initial denaturation at 94°C for 3 min, followed by 35 cycles at 94°C for 30 s and 64.5°C for 3.3 min, and a final one-cycle elongation step at 64.5°C for 5 min. The primers for long-range PCR were selected from those that were successful in the short-range PCR amplification. We also identified additional primer sequences from the adjacent regions of the male-specific sequence. In total, 12 and 5 primer pairs targeting the ~110-kb male-specific and adjacent regions, respectively, were selected (Table S15). PCRs were carried out with KOD One PCR Master Mix (Toyobo). In brief, the amplification was performed in 10 µl of KOD One 2× PCR Master Mix, 8 µl of DW, 0.5 µl (10 µM) of each primer, and 1 µl (~30 ng) of genomic DNA in a total reaction volume of 20 µl. PCR reactions were carried out on a BIO-RAD T100 thermal cycler (Bio-Rad, CA, USA) under the following thermal conditions: initial denaturation at 94°C for 2 min, 5 cycles of 94°C for 10 s and 74°C for 5 min, 5 cycles of 94°C for 10 s and 72°C for 5 min, 5 cycles of 94°C for 10 s and 70°C for 5 min, 25 cycles of 94°C for 10 s and 68°C for 5 min, and final extension at 68°C for 5 min. In total, 17 long-range PCR amplicons were generated. To determine the sequences of long-range PCR amplicons, we used Oxford Nanopore sequencing technology. The amplicons were purified with AmpureXP beads and pooled in a single tube. Sequencing was carried out using a single GridION cell with the Ligation Sequencing Kit 1D (SQK-LSK109) at Genebay. Basecalling was performed using Guppy (v. 3.0.3) (http://nanoporetech.com) and adapter trimming was conducted with Porechop (v. 0.2.3) (https://github.com/rrwick/Porechop.git), followed by a quality trimming step using fastp (-q 7) (51). In total, 1.09-Gb (yielded reads = 192,777; N50 = 8 kb) was sequenced, resulting in approximately ~9000× coverage relative to the length of the targeted region. Reads with a mean qscore greater than 7 and a read length greater than 1 kb were kept for the following steps. Canu (v. 1.7.1) (52) was used for raw read correction with default parameters except for corOutCov = 2,000 and corMinCoverage = 0. Corrected reads were aligned to Chr19Y using BWA-MEM (4). The chimeric reads were filtered by Alvis (53). Soft-clipped reads were flagged using the biostar84452.jar script (https://github.com/lindenb/jvarkit.git). After these filtering steps, reads with a mapping quality of 60 were retained. The mapped BAM file was converted to FASTQ file and fed into Canu (v. 1.7.1) (52) to assemble the reads into contigs with default parameters except for corOutCoverage = 500 and corMinCoverage = 0. Contigs were aligned on Chr19Y for dot plot visualization by ggplot2 (54).

6-5. Confirmation of the male-specific region in wild populations by diagnostic primers

To obtain more data supporting the male-specific presence of this region in a wild population (n = 20 for each sex of each species), we designed the primers “TS_chr14_8812k_0.4k_F” and

"TS_chr14_8812k_0.5k_R," which produce distinct amplicons from the male-specific region (352 bp) and its paralogous region on Chr14 (408 bp) (Table S15). PCRs were carried out in a total reaction volume of 15 µl containing 0.3 µl (10 µM) of each primer, 1.5 µl of 10X brilliant buffer (Mg2+ plus), 1.2 µl of 2.5 mM dNTP mixture, 0.15 µl of Hot-Start Gene Taq (Nippon Gene), 10.55 µl DW, and 1 µl (25 to 30 ng/µl) of genomic DNA. The following thermal conditions were used: initial denaturation at 95°C for 2 min, followed by 30 cycles at 94°C for 30 s and 64°C for 30 s and a final one-cycle elongation step at 65°C for 30 s. The two amplicons were identified using the DNA-500 Kit on a MultiNA instrument (Shimadzu, Kyoto, Japan). The designed PCR primers were referred to as the diagnostic primers.

7. Genome assembly in a *T. snyderi* XY male

7-1. Long- and linked-read sequencing

The whole-genome sequence of an XY individual from Lake Hamana was obtained using PromethION (Nanopore Technologies) and MGI single-tube long-fragment reads (stLFRs, MGISEQ-2000RS) (Table S16 and S17). Genomic DNA was isolated as described above. To enhance recovery of DNA fragments longer than 10-kb, the Short Read Eliminator XS kit (Circulomics, Baltimore, MD, USA) was used. Library preparation and sequencing for both technologies were performed by GeneBay. As for PromethION sequencing, the raw signal intensity data were used for base calling using Guppy (v. 3.0.3). In total, ~9.09-Gb data of 890,724 (N50 = 25 kb) reads were generated. Low-quality reads (with a mean quality score less than 7) and the adapters were removed using fastp (51) and Porechop (v. 0.2.3) (https://github.com/rrwick/Porechop), respectively. The retained reads were corrected in Canu (v. 1.7.1) (52) with the following settings: minimum read length = 5,000; minimum overlap length = 1,200; and corMinCoverage = 0. This resulted in a total of 7.05-Gb corrected reads. As for stLFR sequencing, the MGIEasy stLFR Library Prep Kit was used to prepare co-barcoding DNA libraries (55). The libraries were sequenced using the MGISEQ-2000RS. In total, ~38-Gb (without barcodes 32.17 Gb) of paired-end sequences was generated. The stLFR reads were filtered using SOAPfilter (v. 2.2) (56) with a Phred score ≤10 to remove adapter sequences and low-quality reads containing more than 60% bases. The data were registered in the DDBJ SRA (Accession No. DRA014181).

7-2. Hybrid genome assembly

We generated an initial genome assembly of *T. snyderi* using Canu (v. 1.7.1) (52) with Nanopore long reads. For consensus correction, Nanopore long reads were mapped to the unitig sequences by Minimap2 (26), and three rounds of corrections were performed using Racon (27) followed by one round of correction using medaka (v. 1.2.1) (https://nanoporetech.github.io/medaka/). To improve the

Nanopore assembly with stLFR data, we first transformed the stLFR data to pseudo- 10X Genomics reads using the stlfr2supernova pipeline (https://github.com/BGIQingdao/stlfr2supernova_pipeline). Then the Nanopore assembly was scaffolded with the transformed stLFR reads using ARKS (v. 1.0.6) with LINKS (v. 1.8.7) (57, 58), using default parameters. The resultant assembly was polished twice using POLCA (59) with the transformed stLFR data.

The gene functional annotation of the assembly and its completeness assessment were conducted as described for *T. niphobles*. For the identification of the male-specific region, the relative depth of coverage between males and females was analyzed using 24 individuals for each sex (three pools each of females and males, with each pool containing eight individuals) as described for *T. niphobles*. The characterization of the male-specific region was conducted as described for *T. niphobles*.

8. Genome assembly in a *T. vermicularis* XY male

We sequenced the genome of an XY individual from Ariake Sea using PacBio long reads (Table S18) and Illumina paired-end short reads (Sample ID: S9, Table S5). The genomic DNA was extracted as described for *T. niphobles*. DNA fragments shorter than 10-kb were removed using the Short Read Eliminator XS kit. Library preparation and sequencing for PacBio Sequel II were performed at Macrogen Corporation. In total, ~128.8-Gb data composed of 11,257,208 subreads (N50 = 15 kb) were generated from a single PacBio SMRT cell. The PacBio subreads were corrected by Canu (v. 1.7.1) with the following settings: minimum read length = 7,000; minimum overlap length = 1,200; and coreOutcoverage = 200. This resulted in ~67.17-Gb of corrected reads. The data were registered in the DDBJ SRA (Accession No. DRA014182). For Illumina paired-end short reads, library preparation and sequencing were performed using the NovaSeq 6000 platform of GeneBay. We also sequenced additional eight female and one male *T. vermicularis* from Ariake Sea using the NovaSeq 6000 platform following the aforementioned protocol. Information about the resequencing statistics is presented in Table S5. The data were registered in the DDBJ SRA (Accession No. DRA012878). The short reads were trimmed by removing adapters and low-quality reads as described above, except that the option MINLEN:140 was used. We generated an initial genome assembly of *T. vermicularis* from the error-corrected PacBio reads using Minimap2 and Miniasm as described for *T. niphobles*. Sequence annotation of the assembly and assessment of its completeness were conducted as described for *T. niphobles*. For identification of the male-specific region, the relative depth of coverage between males and females was analyzed using 7 males and 12 females as described for *T. niphobles*. Characterization of the male-specific region was conducted as described for *T. niphobles*.

9. Phylogenetic analysis of male-specific genes and their autosomal paralogs

To determine the phylogenetic relationships between the male-specific genes and their autosomal paralog(s), we further identified their orthologs in *T. rubripes* (FUGU5) by combining BLAST results and their syntenic relationships. We used *Torquigener hypselogeneion* as the outgroup. Since the genome assembly for this species was not available, we constructed a de novo draft genome by assembling the Illumina short reads, obtained for the species phylogenic analysis (Table S4), using Platanus Assembler (v. 1.2.4) with default parameters (15). The orthologous sequences of each targeted gene from *Torquigener hypselogeneion* were identified by reciprocal BLAST search approaches using the male-specific genes, *Torquigener hypselogeneion* genome, and FUGU5 annotation database. When multiple contigs of *Torquigener hypselogeneion* showed high similarity between a target gene from FUGU5, these contigs were further assembled using a reference-guided contig assembly strategy using ragtag (60). Finally, multiple sequence alignment files were generated for each gene from all species using webPRANK (61). Phylogenetic relationships were inferred using RAxML (v. 8.0) (13) with the *GTRGAMMA* model and *--JC69* option. The reliability of the inferred tree was tested by 1,000 fast bootstrap replicates. We also created alternative trees that incorporated InDels into the phylogeny construction. The gaps in the multiple sequence alignment were treated as fifth states and subjected to RAxML (v. 8.0) (13) with the *MULTICAT* model and *GTR* substitution model with the -V option. Furthermore, the concatenated sequences of the *Gsdf* and *Ppef2* genes from the male-specific and autosomal paralogs were used to construct the phylogeny with and without incorporation of the InDels information. Phylogenetic relationships without InDels information were inferred using RAxML (v. 8.0) (13) with the *GTRGAMMA* model, and those that did incorporate InDels information were inferred as described above. Source data are provided in SI (Fasta2_genes_on_SDR_and_their_paralogs.txt).

10. Admixture analysis

10-1. Tree topology weighting by topology weighting by
iterative sampling of sub-trees (TWISST):

We quantified the genealogical relationships throughout the genomes of five *Takifugu* species using TWISST (62). To this end, we used 16,343,174 SNP sites from five Takifugu species (*T. rubripes* (n=4), *T. niphobles* (n=5), *T. snyderi* (n=5), *T. vermicularis* (n=3), *T. poecilonotus* (n=1)) and one outgroup species (*Torquigener hypselogeneion* (n=1)). Whole-genome resequencing data for *T. rubripes* were previously obtained in our laboratory (DDBJ SRA; DRA007464) (10). Quality control of the raw resequencing reads, mapping of the reads onto FUGU5, and SNP calling were performed as described above ("Phylogenetic framework" section in the supplementary Information). Variant filtering was achieved using *vcffilter* from *vcflib* (63) with the following parameters "QUAL / AO > 20 & SAF > 0 & SAR > 0 & RPR > 1 & RPL > 1." Variants residing in the gaps, repeats, and non-

oriented scaffolds in FUGU5 were subsequently eliminated. The genotype dataset was phased using the program Beagle ver. 4.0 (64) with default parameters. We performed a sliding window–based estimation of the local phylogenetic relationships using the phased genotype dataset. The raxml_sliding_windows.py script from the genomics_general package (https://github.com/simonhmartin/genomics_general/tree/master/phylo) was used to reconstruct the phylogenetic trees in windows of 2,000 polymorphic sites. The TWISST program was used to calculate the weighting of each local window.

10-2. *D* statistics and f_4-ratio statistics

To aid the interpretation of the TWISST analysis, we investigated admixture and introgression across the genomes of *Takifugu* species using Patterson's *D* statistics (65, 66), f_4-ratio statistics (67), and *f*-branch statistics (68). We added resequencing data of two *Takifugu* individuals (*T. chrysops* (n=1), *T. pardalis* (n=1)) to the data used in the TWISST analysis and obtained 17,543,194 SNP sites. The past gene flow in seven species was inferred from *D* statistics calculated utilizing the four-taxon test (((P1, P2), P3), outgroup). Ancestral alleles were designated as "A" and derived alleles as "B," and the genome-wide *D* statistics were calculated using all trio combinations among the seven species with Dsuite v.0.2 r20 (69), guided by the species tree that was constructed by RAxML (Fig. 1A). *Torquigener hypselogeneion* was used as the outgroup for all trio combinations. Gene flow among species was examined, and it was determined if *D* statistics were greater than those expected under the model without gene flow. Test results were interpreted as follows: in four-taxon tests, a significant positive *D* statistic indicated gene flow between P2 and P3. In the four-taxon pattern (((P1, P2), P3), outgroup), the positions of P1 and P2 for a species are arbitrary; thus, Dsuit always assigns them so that P2 and P3 share more derived alleles, and the values of the ABBA–BABA statistics are then limited to between 0 and 1. Block jackknife resampling was used for the blocks of 5,000 informative sites to evaluate significant deviations from zero in Patterson's *D* statistics. Furthermore, we used f_4-ratio statistics to infer the mixing proportions of an admixture event. The results presented in this study are based on the "tree" output of the Dsuite function Dtrios, with each trio arranged according to the species tree based on the maximum-likelihood topology estimated by RAxML (Fig. 1A).

SI References

1. A. M. Bolger, M. Lohse, B. Usadel, Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120 (2014).

2. S. Hosoya, *et al.*, Random PCR-based genotyping by sequencing technology GRAS-Di (genotyping by random amplicon sequencing, direct) reveals genetic structure of mangrove

fishes. *Mol. Ecol. Resour.* **19**, 1153–1163 (2019).

3. W. Kai, *et al.*, Integration of the genetic map and genome assembly of fugu facilitates insights into distinct features of genome evolution in teleosts and mammals. *Genome Biol. Evol.* **3**, 424–442 (2011).

4. H. Li, R. Durbin, Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).

5. A. McKenna, *et al.*, The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* **20**, 1297–1303 (2010).

6. P. Danecek, *et al.*, The variant call format and VCFtools. *Bioinformatics* **27**, 2156–2158 (2011).

7. K. W. Broman, H. Wu, S. Sen, G. A. Churchill, R/qtl: QTL mapping in experimental crosses. *Bioinformatics* **19**, 889–890 (2003).

8. R. Ieda, *et al.*, Identification of the sex-determining locus in grass puffer (Takifugu niphobles) provides evidence for sex-chromosome turnover in a subset of Takifugu species. *PLoS ONE* **131**, e0190635 (2018).

9. R. W. Doerge, G. A. Churchill, Permutation tests for multiple loci affecting a quantitative character. *Genetics* **142**, 285–294 (1996).

10. M. Sato, *et al.*, A highly flexible and repeatable genotyping method for aquaculture studies based on target amplicon sequencing using next-generation sequencing technology. *Sci. Rep.* **9**, 6904 (2019).

11. H. Li, *et al.*, The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).

12. E. Garrison, G. Marth, Haplotype-based variant detection from short-read sequencing. *arXiv* [Preprint] (2012). https://arxiv.org/abs/1207.3907 (Accessed 20 July 2012).

13. A. Stamatakis, RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30**, 1312–1313 (2014).

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

15. R. Kajitani, *et al.*, Efficient de novo assembly of highly heterozygous genomes from whole-genome shotgun short reads. *Genome Res.* **24**, 1384–1395 (2014).

16. C. Camacho, *et al.*, BLAST+: architecture and applications. *BMC Bioinformatics* **10**, 421 (2009).

17. Z. Zhang, S. Schwartz, L. Wagner, W. Miller, A greedy algorithm for aligning DNA sequences. *Journal of Computational Biology* **7** (2000).

18. T. A. Hall, BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symposium Series* **41**, 95–98 (1999).

19. C. Ye, Z. S. Ma, C. H. Cannon, M. Pop, D. W. Yu, Exploiting sparseness in de novo genome assembly. *BMC Bioinformatics* **13 Suppl 6**, S1 (2012).

20. C. Ye, C. M. Hill, S. Wu, J. Ruan, Z. S. Ma, DBG2OLC: Efficient Assembly of Large Genomes Using Long Erroneous Reads of the Third Generation Sequencing Technologies. *Sci. Rep.* **6**, 31900 (2016).

21. C. Ye, Z. S. Ma, Sparc: a sparsity-based consensus algorithm for long erroneous sequencing reads. *PeerJ* **4**, e2016 (2016).

22. M. J. Chaisson, G. Tesler, Mapping single molecule sequencing reads using basic local alignment with successive refinement (BLASR): application and theory. *BMC Bioinformatics* **13**, 238 (2012).

23. C.-S. Chin, *et al.*, Nonhybrid, finished microbial genome assemblies from long-read SMRT sequencing data. *Nat. Methods* **10**, 563–569 (2013).

24. B. J. Walker, *et al.*, Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PLoS One* **9**, e112963 (2014).

25. H. Li, Minimap and miniasm: fast mapping and de novo assembly for noisy long sequences. *Bioinformatics* **32**, 2103–2110 (2016).

26. H. Li, Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* **34**, 3094–3100 (2018).

27. R. Vaser, I. Sović, N. Nagarajan, M. Šikić, Fast and accurate de novo genome assembly from long uncorrected reads. *Genome Res.* **27**, 737–746 (2017).

28. M. Chakraborty, J. G. Baldwin-Brown, A. D. Long, J. J. Emerson, Contiguous and accurate de novo assembly of metazoan genomes with modest long read coverage. *Nucleic Acids Res.* **44**, e147 (2016).

29. M. Kadota, *et al.*, Multifaceted Hi-C benchmarking: what makes a difference in chromosome-scale genome scaffolding? *Gigascience* **9, giz158** (2020).

30. N. Servant, *et al.*, HiC-Pro: an optimized and flexible pipeline for Hi-C data processing. *Genome Biol.* **16**, 259 (2015).

31. N. C. Durand, *et al.*, Juicer Provides a One-Click System for Analyzing Loop-Resolution Hi-C Experiments. *Cell Syst* **3**, 95–98 (2016).

32. O. Dudchenko, *et al.*, De novo assembly of the Aedes aegypti genome using Hi-C yields chromosome-length scaffolds. *Science* **356**, 92–95 (2017).

33. B. L. Cantarel, *et al.*, MAKER: an easy-to-use annotation pipeline designed for emerging model organism genomes. *Genome Res.* **18**, 188–196 (2008).

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

35. M. Stanke, A. Tzvetkova, B. Morgenstern, AUGUSTUS at EGASP: using EST, protein and genomic alignments for improved gene prediction in the human genome. *Genome Biol.* **7 Suppl 1**, S11.1-8 (2006).

36. 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).

37. A. Smit, R. Hubley, *RepeatModeler open-1.0* (2008–2015). Seattle, USA: Institute for Systems Biology. Available from: httpwww.repeatmasker.org, Last Accessed May 1, 2018.

38. Z. Bao, S. R. Eddy, Automated de novo identification of repeat sequence families in sequenced genomes. *Genome Res.* **12**, 1269–1276 (2002).

39. A. L. Price, N. C. Jones, P. A. Pevzner, De novo identification of repeat families in large genomes. *Bioinformatics* **21 Suppl 1**, i351-8 (2005).

40. D. Ellinghaus, S. Kurtz, U. Willhoeft, LTRharvest, an efficient and flexible software for de novo detection of LTR retrotransposons. *BMC Bioinformatics* **9**, 18 (2008).

41. S. Steinbiss, U. Willhoeft, G. Gremme, S. Kurtz, Fine-grained annotation and classification of de novo predicted LTR retrotransposons. *Nucleic Acids Res.* **37**, 7002–7013 (2009).

42. Y. Han, S. R. Wessler, MITE-Hunter: a program for discovering miniature inverted-repeat transposable elements from genomic sequences. *Nucleic Acids Res.* **38**, e199 (2010).

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

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

45. O. Nishimura, Y. Hara, S. Kuraku, gVolante for standardizing completeness assessment of genome and transcriptome assemblies. *Bioinformatics* **33**, 3635–3637 (2017).

46. G. Marçais, *et al.*, MUMmer4: A fast and versatile genome alignment system. *PLoS Comput. Biol.* **14**, e1005944 (2018).

47. M. Krzywinski, *et al.*, Circos: an information aesthetic for comparative genomics. *Genome Res.* **19**, 1639–1645 (2009).

48. D. Kim, B. Langmead, S. L. Salzberg, HISAT: a fast spliced aligner with low memory requirements. *Nat. Methods* **12**, 357–360 (2015).

49. H. Thorvaldsdóttir, J. T. Robinson, J. P. Mesirov, Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Brief. Bioinform.* **14**, 178–192 (2013).

50. B. S. Pedersen, A. R. Quinlan, Mosdepth: quick coverage calculation for genomes and exomes. *Bioinformatics* **34**, 867–868 (2018).

51. S. Chen, Y. Zhou, Y. Chen, J. Gu, Fastp: An ultra-fast all-in-one FASTQ preprocessor in *Bioinformatics* 34, i884-i890 (2018).

52. S. Koren, *et al.*, Canu: Scalable and accurate long-read assembly via adaptive κ-mer weighting and repeat separation. *Genome Res*. **27** 722-736 (2017).

53. S. Martin, R. M. Leggett, Alvis: a tool for contig and read ALignment VISualisation and chimera detection. *BMC Bioinformatics* **22**, 124 (2021).

54. H. Wickham, Ggplot2. *Wiley Interdiscip. Rev. Comput. Stat.* **3**, 180–185 (2011).

55. O. Wang, *et al.*, Efficient and unique cobarcoding of second-generation sequencing reads from long DNA molecules enabling cost-effective and accurate sequencing, haplotyping, and de novo assembly. *Genome Res.* **29**, 798–808 (2019).

56. R. Luo, *et al.*, SOAPdenovo2: An empirically improved memory-efficient short-read de novo assembler. *GigaScience* **1, 18** (2012).

57. L. Coombe, *et al.*, ARKS: chromosome-scale scaffolding of human genome drafts with linked read kmers. *BMC Bioinformatics* **19**, 234 (2018).

58. R. L. Warren, *et al.*, LINKS: Scalable, alignment-free scaffolding of draft genomes with long reads. *Gigascience* **4**, 35 (2015).

59. A. V. Zimin, S. L. Salzberg, The genome polishing tool POLCA makes fast and accurate corrections in genome assemblies. *PLoS Comput. Biol.* **16**, e1007981 (2020).

60. M. Alonge, *et al.*, RaGOO: fast and accurate reference-guided scaffolding of draft genomes. *Genome Biol.* **20**, 224 (2019).

61. A. Löytynoja, N. Goldman, webPRANK: a phylogeny-aware multiple sequence aligner with interactive alignment browser. *BMC Bioinformatics* **11**, 579 (2010).

62. S. H. Martin, S. M. Van Belleghem, Exploring evolutionary relationships across the genome using topology weighting. *Genetics* **206**, 429–438 (2017).

63. E. Garrison, Z. N. Kronenberg, E. T. Dawson, B. S. Pedersen, P. Prins, Vcflib and tools for processing the VCF variant call format. *bioRxiv* [Priprint] (2021) https:/doi.org/10.1101/2021.05.21.445151 (Last Accessed by July 23 2021).

64. S. R. Browning, B. L. Browning, Rapid and accurate haplotype phasing and missing-data inference for whole-genome association studies by use of localized haplotype clustering. *Am. J. Hum. Genet.* **81**, 1084–1097 (2007).

65. R. E. Green, *et al.*, A draft sequence of the Neandertal genome. *Science* **328**, 710–722 (2010).

66. E. Y. Durand, N. Patterson, D. Reich, M. Slatkin, Testing for ancient admixture between closely related populations. *Mol. Biol. Evol.* **28**, 2239–2252 (2011).

67. N. Patterson, *et al.*, Ancient admixture in human history. *Genetics* **192**, 1065–1093 (2012).

68. M. Malinsky, *et al.*, Whole-genome sequences of Malawi cichlids reveal multiple radiations interconnected by gene flow. *Nat. Ecol. Evol.* **2**, 1940–1955 (2018).

69. M. Malinsky, M. Matschiner, H. Svardal, Dsuite - Fast D-statistics and related admixture evidence from VCF files. *Mol. Ecol. Resour.* **21**, 584–595 (2021).

Species	Sampled location	The proportion of the G/C genotype at SNP7271 in *Amhr2*		Total number of fish examined	P-value for associtaion between the genotype and phenotypic sex	Method	Previously repotated or this study.
		Phenotypic male	Phenotypic female				
*Takifugu rubripes**	Ensyu Nada, Ise Bay, Japan	7/7	0/8	15	0.0001554	Direct sequencing	Kamiya et al. (2012).
Takifugu pardalis	Lake Hamana and Atusmi Peninsula, Japan	20/20	0/8	28	1.21315E-07	Direct sequencing	This study and Kamiya et al. (2012).
Takifugu poecilonotus	Lake Hamana, Japan	7/7	0/6	13	0.000311491	Direct sequencing	Kamiya et al. (2012).
Takifugu chrysops	Izu Peninsula, Japan	4/4	0/4	8	0.00467773	Direct sequencing	This study
Takifugu stictonotus	Shimonoseki, Fukui, and Ofunato, Japan	16/16	0/26	42	9.13E-11	Direct sequencing	This study
*Takifugu obscurus***	Aquaculture fish in Yamaguchi, Japan	1/1	0/2	3	0.333333	Direct sequencing	This study
Takifugu ocellatus	Aquarium fish	10/10	0/2	12	0.000532006	Direct sequencing	This study
*Takifugu xanthopterus****	Shimonoseki, Japan	9/10	1/25	35	3.62E-07	Direct sequencing	This study
Takifugu porphyreus	Shimonoseki, Japan	8/8	0/8	16	6.33425E-05	Direct sequencing	This study
*Takifugu niphobles*****	Lake Hamana, Japan	0/8 0/100	0/7 0/100	15 200	1 1	Direct sequencing HRM	Ieda et al. (2018).
*Takifugu snyderi******	Sagara, Atsumi and Hitachi, Japan	1/16	0/12	29	0.377822	Direct sequencing	This study
	Sagara, Japan	0/50	0/50	100	1	HRM	
Takifugu vermicularis	Ariake Bay (Shimabara), Japan	0/4 0/50	0/4 0/50	8 100	1 1	Direct sequencing HRM	This study

Table S1. SNP genotype in *Amhr2* and phenotypic sex of *Takifugu* species

*A significant association between the genotype and phenotype has been reported for other individuals in Kamiya et al. (2012).
**A significant association between the genotype and phenotype in three genetically independent aquaculture populations has been reported in the following paper. A rapid and reliable method for identifying genetic sex in obscure pufferfish (*Takifugu obscurus*). Aquaculture (2020) 519:734749. doi: 10.1016/j.aquaculture.2019.734749
***One males and one female showed the mismach with the genotype at the SNP site in *Amhr2*.
****A significant association between the genotype and phenotype has been reported for other individuals in Ieda et al. (2018).
*****One out of 15 male individuals has the G/C genotype at the SNP site in *Amhr2*.

Table S2A. The families used for linkage mapping in *T. snyderi*

Family name	AKA	Total number of fish	Phenotypic male	Phenotypic female	Type of genetic markers
Family A	shousai1202	83	40	43	Microsatellites on Chr18
		98	53	45	Genome-wide SNP markers
Family B	shousai1201	25	13	12	Microsatellites on Chr18
Fanily C	shousai1203	62	27	35	Microsatellites on Chr18

Three wild-caught males were mated with a wild-caught female.

Table S2B. The families used for linkage mapping in *T. vermicularis*

Family name	AKA	Total number of fish	Phenotypic male	Phenotypic female	Type of genetic markers
Family D	nashi12	62	31	31	Genome-wide microsatellites
Family E	nashi15	164	81	83	Microsatellites on Chr10

Two wild-caught parents were mated independently.

Table S3A. Numbers of fish showing concordance and discordance between phenotypic and genotypic sex among those used for genetic mapping in *T. snyderi*

Family name	Phenotypic male		Phenotypic female		P-value for associtaion between the markers and sex
	Expected male genotype (XY)*	Expected female genotype (XX)*	Expected male genotype (XY)*	Expected female genotype (XX)*	
Family A	40	0	0	43	1.86E-23
Family B	13	0	0	12	1.92E-20
Family C	27	0	1	34	1.00E-16

* The expected male and female genotypes were determined based on the paternally inherited allele of markers (f1618 and f1659) near the distal end of LG18 (Chr18).

Table S3B. Numbers of fish showing concordance and discordance between phenotypic and genotypic sex among those used for genetic mapping in *T. vermicularis*

Family name	Phenotypic male		Phenotypic female		P-value for associtaion between the markers and sex
	Expected male genotype (XY)*	Expected female genotype (XX)*	Expected male genotype (XY)*	Expected female genotype (XX)*	
Family D	31	0	0	31	4.29E-15
Family E	80	1	5	78	1.80E-39

* The expected male and female genotypes were determined based on the paternally inherited allele of markers sca541-1 near the distal end of LG10 (Chr10).

Table S4. Sample information and genome resequencing statistics of *Takifugu* species for the phylogenetic analysis

Species	AKA	Platform	Paired-end (bp)	Reads Number	Base (Mb)	Coverage (x)
Takifugu rubripes	UT01	HiSeq 2000	100x2	45,150,676	4,515	11.3
Takifugu pardalis	higanfugu	HiSeq 2500	90x2	142,361,704	12,812	32.0
Takifugu poecilonotus	komonfugu	HiSeq 2000	100x2	49,668,804	4,967	12.4
Takifugu chrysops	Aka_1	HiSeq X Ten	150x2	122,637,398	18,395	46.0
Takifugu stictonotus	Gomafugu	HiSeq 2500	90x2	139,026,904	12,512	31.3
Takifugu obscurus	mefugu	HiSeq 2000	100x2	46,304,194	4,630	11.6
Takifugu ocellatus	meganefugu	HiSeq 2000	100x2	55,867,410	5,587	14.0
Takifugu xanthopterus	shimafugu	HiSeq 2000	100x2	41,690,572	4,169	10.4
Takifugu porphyreus	mafugu	HiSeq 2500	90x2	144,594,610	13,013	32.5
Takifugu niphobles	kusafugu11_male32	HiSeq 2500	90x2	149,777,588	13,478	33.7
Takifugu snyderi	TS_ATM_no_13	HiSeq 2500	150x2	28,346,608	4,251	10.6
Takifugu vermicularis	nashifugu	HiSeq 2000	100x2	48,787,416	4,879	12.2
Torquigener hypselogeneion	namidafugu	HiSeq 2000	100x2	46,268,560	4,627	11.6

Table S5A. Sample information and genome resequencing statistics of *Takifugu niphobles*

Sample ID	AKA	Sex	Platform	Paired-end (bp)	Reads Number	Base (Mb)	Coverage (x)	Sampling location	The analysis used for
12m	TN_2010_no_12_male	M	Hiseq 2000	150x2	28,501,980	4,275	10.7	Lake Hamana	k-mer, coverage, k-mer based assembly
19m	TN_2010_no_19_male	M	Hiseq 2000	150x2	28,497,398	4,274	10.7	Lake Hamana	k-mer, coverage, k-mer based assembly
10m	TN_2010_no_10_male	M	Hiseq 2000	150x2	28,508,256	4,276	10.7	Lake Hamana	k-mer, coverage, k-mer based assembly
11m	TN_2010_no_11_male	M	Hiseq 2000	150x2	28,512,916	4,276	10.7	Lake Hamana	k-mer, coverage, k-mer based assembly
m8	kusafugu male 1	M	Hiseq 2000	90x2	140,463,248	12,641	31.6	Lake Hamana	coverage
m1	kusafugu09_male1	M	Hiseq 2000	90x2	148,625,678	13,374	33.4	Lake Hamana	coverage
m30	kusafugu11_male30	M	Hiseq 2000	90x2	149,786,688	13,479	33.7	Lake Hamana	coverage
m31	kusafugu11_male31	M	Hiseq 2000	90x2	149,062,258	13,410	33.5	Lake Hamana	coverage
m32	kusafugu11_male32	M	Hiseq 2000	90x2	149,777,588	13,478	33.7	Lake Hamana	coverage
0525m	TN_20150525_male	M	Hiseq 2000	150x2	28,330,108	4,249	10.6	Lake Hamana	k-mer, coverage, k-mer based assembly
Bfe	TN_2015B_female	F	Hiseq 2000	150x2	28,404,224	4,260	10.7	Lake Hamana	k-mer, coverage
Afe	TN_2015A_female	F	Hiseq 2000	150x2	28,375,024	4,256	10.6	Lake Hamana	k-mer, coverage
65fe	TN_2010_no_65_female	F	Hiseq 2000	150x2	28,396,346	4,259	10.6	Lake Hamana	k-mer, coverage
44fe	TN_2010_no_44_female	F	Hiseq 2000	150x2	28,493,116	4,273	10.7	Lake Hamana	k-mer, coverage
35fe	TN_yyno_35_xxfemale	F	Hiseq 2000	150x2	28,282,010	4,242	10.6	Lake Hamana	coverage
Misakife	TN_2012Misaki_xxfemale	F	Hiseq 2000	150x2	28,197,746	4,229	10.6	Lake Hamana	k-mer, coverage
41fe	TN_2010_no_41_female	F	Hiseq 2000	150x2	28,353,496	4,253	10.6	Lake Hamana	coverage
1103fe	TN_1103_xxfemale	F	Hiseq 2000	150x2	28,356,500	4,253	10.6	Lake Hamana	k-mer, coverage

Table S5B. Sample information and genome resequencing statistics of *Takifugu snyderi*

Sample ID	The pooled number of individuals	Sex	Platform	Paired-end (bp)	Reads Number	Base (Mb)	Coverage (x)	Sampling location	The analysis used for
fP1	8	F	HiSeq X Ten	150x2	455,991,960	68,855	21.5	Sagara	k-mer, coverage
fP2	8	F	HiSeq X Ten	150x2	471,561,858	71,206	22.3	Sagara	k-mer, coverage
fP3	8	F	HiSeq X Ten	150x2	467,422,088	70,581	22.1	Sagara	k-mer, coverage
mP1	8	M	HiSeq X Ten	150x2	477,727,398	72,137	22.5	Sagara	k-mer, coverage
mP2	8	M	HiSeq X Ten	150x2	455,008,940	68,706	21.5	Sagara	k-mer, coverage
mP3	8	M	HiSeq X Ten	150x2	472,824,138	71,396	22.3	Sagara	k-mer, coverage

Table S5C. Sample information and genome resequencing statistics of *Takifugu vermicularis*

Sample ID	AKA	Sex	Platform	Paired-end (bp)	Reads Number	Base (Mb)	Coverage (x)	Sampling location	The analysis used for
89	NA	F	HiSeq 2500	150x2	28,151,646	4,223	10.6	Ariake Sea	k-mer, coverage
90	NA	F	HiSeq 2500	150x2	28,427,218	4,264	10.7	Ariake Sea	k-mer, coverage
92	NA	F	HiSeq 2500	150x2	28,334,946	4,250	10.6	Ariake Sea	k-mer, coverage
93	NA	F	HiSeq 2500	150x2	28,375,900	4,256	10.6	Ariake Sea	k-mer, coverage
94	NA	M	HiSeq 2500	150x2	28,326,008	4,249	10.6	Ariake Sea	k-mer, coverage
95	NA	M	HiSeq 2500	150x2	28,333,434	4,250	10.6	Ariake Sea	k-mer, coverage
96	NA	M	HiSeq 2500	150x2	28,416,252	4,262	10.7	Ariake Sea	k-mer, coverage
97	NA	M	HiSeq 2500	150x2	28,430,436	4,265	10.7	Ariake Sea	k-mer, coverage
1	NA	M	HiSeq 2500	100x2	48,787,416	4,878	12.2	Ariake Sea	coverage
S9	NA	M	NovaSeq 6000	150x2	60,399,446	9,059	22.6	Ariake Sea	coverage, assembly construction with long and short reads
S11	NA	M	NovaSeq 6000	150x2	64,961,632	9,744	24.4	Ariake Sea	coverage
S1	NA	F	NovaSeq 6000	150x2	67,223,726	10,083	25.2	Ariake Sea	coverage
S3	NA	F	NovaSeq 6000	150x2	62,019,898	9,302	23.3	Ariake Sea	coverage
S4	NA	F	NovaSeq 6000	150x2	67,825,690	10,173	25.4	Ariake Sea	coverage
S7	NA	F	NovaSeq 6000	150x2	60,123,744	9,018	22.5	Ariake Sea	coverage
S2	NA	F	NovaSeq 6000	150x2	72,475,682	10,871	27.2	Ariake Sea	coverage
S5	NA	F	NovaSeq 6000	150x2	70,944,158	10,641	26.6	Ariake Sea	coverage
S6	NA	F	NovaSeq 6000	150x2	64,852,998	9,727	24.3	Ariake Sea	coverage
S10	NA	F	NovaSeq 6000	150x2	60,425,356	9,063	22.7	Ariake Sea	coverage

Table S6A. Assembly statistics of the male-specific contigs of *T. niphobles, T. snyderi,* and *T. vermicularis* obtained from 35-mer analysis

	T. niphobles	*T. snyderi*	*T. vermicularis*
Total length	218,693 bp	490,306 bp	712,486 bp
Total number of contigs	551	1,499	2,267
Longest contig	3,867 bp	2,354 bp	1,637 bp
N50 contigs size	361 bp	302 bp	291 bp
Shortest contig	203 bp	128 bp	131 bp
Average contig size	397 bp	327 bp	314 bp

Table S6B. Assembly statistics of the repeat masking male-specific contigs of *T. niphobles*, *T. snyderi*, and *T. vermicularis* obtained from 35-mer analysis

	T. niphobles	*T. snyderi*	*T. vermicularis*
Total length	189,350 bp	412,993 bp	636,715 bp
Total number of contigs	504	1,321	2,143
Longest contig	3,395 bp	2,054 bp	1,637 bp
N50 contigs size	352 bp	292 bp	283 bp
Shortest contig	103 bp	102 bp	101 bp
Average contig size	376 bp	313 bp	297 bp

Table S7. Protein-coding genes in the repeat masking male-specific contigs of *T. niphobles*, *T. snyderi*, and *T. vermicularis* identified by the 35-mer analysis

T. niphobles				
Transcript	Chromosome	Start	End	Gene name
ENSTRUT00000050560.1	Chr6	3,843,340	3,848,045	*Gsdf*
ENSTRUT00000036284.2	Chr6	3,847,801	3,854,105	*Ppef2*
ENSTRUT00000036506.2	Chr6	3,855,437	3,860,495	*Nup54*
ENSTRUT00000047035.2	Chr6	4,322,729	4,325,103	*Phyhd1*
ENSTRUT00000026952.2	Chr12	2,412,069	2,413,877	*Uqcrb*
ENSTRUT00000004332.2	HE591965 (linked to Chr17)	5,569	11,125	*Cbx4*
	Chr22	10,421,960	10,423,745	*LOC105418166*
ENSTRUT00000000264.2	HE592879	5,023	6,581	
ENSTRUT00000049960.1	HE592881	4,321	5,864	
ENSTRUT00000050972.1	HE593486	4,284	5,854	
ENSTRUT00000033706.2	HE593884	6,668	7,747	
ENSTRUT00000008519.2	HE594729	398	1,566	
ENSTRUT00000006976.2	HE595883	381	2,179	

T. snyderi				
Transcript	Chromosome	Start	End	Gene name
ENSTRUT00000050560.1	Chr6	3,843,340	3,848,045	*Gsdf*
ENSTRUT00000036284.2	Chr6	3,847,801	3,854,105	*Ppef2*
ENSTRUT00000047035.2	Chr6	4,322,729	4,325,103	*Phyhd1*
ENSTRUT00000049362.1	Chr13	249,922	251,189	
ENSTRUT00000004696.2	Chr19	5,543,117	5,547,028	
ENSTRUT00000028499.2	Chr20	590,965	593,234	
ENSTRUT00000053310.1	Chr21	3,903,120	3,907,840	
	Chr22	10,421,960	10,423,745	*LOC105418166*
ENSTRUT00000051340.1	HE591759	67,310	68,298	
ENSTRUT00000053632.1	HE591843	283,793	285,052	
ENSTRUT00000006207.2	HE591877	1,747	10,877	
ENSTRUT00000004332.2	HE591965 (linked to Chr17)	5,569	11,125	*Cbx4*
ENSTRUT00000005695.2	HE592014	11,953	13,339	
ENSTRUT00000054941.1	HE592045	34,942	37,166	
ENSTRUT00000002195.2	HE592247	840	5,029	
ENSTRUT00000003378.2	HE592566	13,102	16,210	
ENSTRUT00000057277.1	HE592566	4,546	6,019	
ENSTRUT00000005528.2	HE592608	521	8,296	
ENSTRUT00000005165.2	HE592831	6,766	8,194	
ENSTRUT00000000264.2	HE592879	5,023	6,581	

ENSTRUT00000001494.2	HE592881	9,873	10,567
ENSTRUT00000049960.1	HE592881	4,321	5,864
ENSTRUT00000049117.1	HE592908	8,974	10,242
ENSTRUT00000049458.1	HE592972	44	4,148
ENSTRUT00000052718.1	HE593008	11,482	12,178
ENSTRUT00000052794.1	HE593106	21	1,160
ENSTRUT00000034896.2	HE593292	1,595	8,108
ENSTRUT00000049469.1	HE593327	8,328	10,161
ENSTRUT00000007069.2	HE593424	4,501	5,859
ENSTRUT00000051117.1	HE593431	1,054	2,443
ENSTRUT00000050972.1	HE593486	4,284	5,854
ENSTRUT00000055318.1	HE593539	4,414	5,937
ENSTRUT00000005027.2	HE593607	3,122	7,140
ENSTRUT00000055210.1	HE593740	1,027	2,672
ENSTRUT00000033706.2	HE593884	6,668	7,747
ENSTRUT00000051955.1	HE594621	26	2,549
ENSTRUT00000012543.2	HE594752	1,424	4,265
ENSTRUT00000035679.2	HE594850	4,400	6,263
ENSTRUT00000055929.1	HE595010	5,440	6,442
ENSTRUT00000050001.1	HE595161	25	2,037
ENSTRUT00000006228.2	HE596788	947	2,433
ENSTRUT00000054060.1	HE597519	370	2,307
ENSTRUT00000054619.1	HE598190	1,342	2,148

T. vermicularis				
Transcript	chromosome	start	end	gene name
ENSTRUT00000003077.2	Chr4	5,033,623	5,038,911	*Cyp3A27*
ENSTRUT00000050560.1	Chr6	3,843,340	3,848,045	*Gsdf*
ENSTRUT00000036284.2	Chr6	3,847,801	3,854,105	*Ppef2*
ENSTRUT00000000482.2	Chr7	5,879,689	5,880,456	*Hepcidin-like*
ENSTRUT00000057327.1	Chr7	6,637,502	6,643,374	
	Chr22	10,421,960	10,423,745	*LOC105418166*
ENSTRUT00000004332.2	HE591965 (linked to Chr17)	5,569	11,125	*Cbx4*
ENSTRUT00000055318.1	HE593539	4,414	5,937	
ENSTRUT00000033706.2	HE593884	6,668	7,747	
ENSTRUT00000006319.2	HE597518	1,831	2,746	

Table S8. Summary of PacBio long read (Sequel) (141.5× coverage) for a male of *T. niphobles*

Number of SMRT cell	7
Number of Subread	5,445,262
Total Bases of Subread (bp)	56,602,699,569
Mean read length of Subread (bp)	10,395
Median read length of Subread (bp)	7,699
Subread N50 (bp)	17,204

Table S9. Summary of Illumina paired-end short-read (57.6× coverage) for a male of *T. niphobles*

Read length	Reads	Base (bp)	Platform	Coverage
245x2	92567888	22679132560	HiSeq 2500	56.7

Table S10. Post-sequencing quality control of the Hi-C library

Read-pair category	Liver-DpnII library
Genome mapping	
Unique mapped	62.9%
Unmapped	1.6%
Multiple mapped	21.2%
Singleton mapped	14.3%
Structure of Hi-C library	
Valid interaction	60.6%
Dangling-end	0.9%
Religation	1.3%
Self circle	0.1%
Single-end	0.0%
Filtered	0.0%
Dumped	0.0%

Table S11. Pseudochromosomes in *T. niphobles* assembly

Pseudochromosome	length (bp)
chromosome_1	29,787,861
chromosome_2	14,645,703
chromosome_3	17,155,000
chromosome_4	16,156,946
chromosome_5	13,743,954
chromosome_6	12,714,000
chromosome_7	16,596,000
chromosome_8	19,988,554
chromosome_9	15,883,851
chromosome_10	13,966,000
chromosome_11	16,244,946
chromosome_12	13,266,446
chromosome_13	20,217,573
chromosome_14	16,184,500
chromosome_15	15,720,327
chromosome_16	13,444,393
chromosome_17	16,811,000
chromosome_18	10,770,500
chromosome_19	17,051,500
chromosome_20	17,331,000
chromosome_21	18,322,500
chromosome_22	15,772,500

Table S12. Summary of the complete, fragmented, missing, and duplicated orthologs inferred from Benchmarking Universal Single-Copy Orthologs (BUSCO) search against the 4,584 highly conserved orthologs for Actinopterygii

BUSCO statistic	YY *T. niphobles*	XY *T. snyderi*	XY *T. vermicularis*
Complete BUSCOs	4,326 (94.4%)	4,440 (96.86%)	4,426 (96.6%)
Fragmented BUSCOs	152 (3.3%)	77 (1.7%)	64 (1.4%)
Missing BUSCOs	106 (2.3%)	66 (1.44%)	94 (2.1%)
Duplicated BUSCOs	143 (3.1%)	110 (2.41%)	213 (4.7%)

Table S13A. Repeat annotation of the 130-kb male-specific region (2.828 Mb to 2.957 Mb on Chr19Y)

Total length: 129,662 bp			
GC level: 45.68%			
Total length of the repeat rich regions: 55,121 bp (42.51%)			
Elements	number of elements	length occupied	percentage of sequence (%)
SINEs:	0	0	0
ALUs	0	0	0
MIRs	0	0	0
LINEs:	10	5,272	4.07
LINE1	0	0	0
LINE2	0	0	0
L3/CR1	0	0	0
LTR elements:	27	22,173	0
ERVL	0	0	0
ERVL-MaLRs	0	0	0
ERV_classI	6	828	6.39
ERV_classII	0	0	0
DNA elements:	13	4,853	3.74
hAT-Charlie	2	438	0.34
TcMar-Tigger	0	0	0
Unclassified:	28	20,642	0
Total interspersed repeats:		52,940	0
Small RNA:	0	0	0
Satellites:	0	0	0
Simple repeats:	27	2,706	2.09
Low complexity:	2	91	0.07

Table S13B. Repeat annotation of the 115-kb male-specific region (2.970 Mb to 3.085 Mb on Chr19Y)

Total length: 115,000 bp			
GC level: 44.85%			
Total length of the repeat rich regions: 71,663 bp (62.32%)			
Elements	number of elements	length occupied	percentage of sequence (%)
SINEs:	1	209	0.18
ALUs	0	0	0.00
MIRs	0	0	0.00
LINEs:	24	14,898	12.95
LINE1	0	0	0.00
LINE2	13	6,972	6.06
L3/CR1	0	0	0.00
LTR elements:	7	4,967	4.32
ERVL	0	0	0.00
ERVL-MaLRs	0	0	0.00
ERV_classI	3	3,226	2.81
ERV_classII	0	0	0.00
DNA elements:	53	23,145	20.13
hAT-Charlie	8	2,471	2.15
TcMar-Tigger	0	0	0.00
Unclassified:	47	26,181	22.77
Total interspersed repeats:		69,400	60.35
Small RNA:	0	0	0.00
Satellites:	0	0	0.00
Simple repeats:	20	2,130	1.85
Low complexity:	2	133	0.12

Table S13C. Repeat annotation for the genome assembly of *T. niphobles*

Total length of the assembly: 396,489,440 bp			
GC level: 45.65%			
Total length of the repeat rich regions: 63,245,816 bp (15.95%)			
Elements	number of elements	length occupied (bp)	Percent of sequence (%)
SINEs:	4,588	794,469	0.2
ALUs	0	0	0
MIRs	103	10,054	0
LINEs:	27,519	15,149,565	3.82
LINE1	112	107,478	0.03
LINE2	14,394	7,302,237	1.84
L3/CR1	0	0	0
LTR elements:	8,406	5,020,062	1.27
ERVL	0	0	0
ERVL-MaLRs	0	0	0
ERV_classI	1,156	1,016,554	0.26
ERV_classII	0	0	0
DNA elements:	31,531	13,122,993	3.31
hAT-Charlie	5,361	2,541,555	0.64
TcMar-Tigger	1,169	427,845	0.11
Unclassified:	35,632	17,042,877	4.3
Total interspersed repeats:		51,129,966	12.9
Small RNA:	40	6,366	0
Satellites:	330	184,413	0.05
Simple repeats:	173,261	10,993,224	2.77
Low complexity:	13,391	960,562	0.24

Table S14. Summary of *T. niphobles* RNA sequence data

Sample ID	Sex	Raw reads	Clean reads	Raw base (Gb)	Clean base (Gb)	Mapping percentage (%)	Platform	Read length
KK_XX_a_19	female	18,833,080	18,586,998	5.65	5.58	78.4	HiSeq 2000	150x2
KK_XX_a_20	female	13,329,149	13,135,808	4	3.94	78.94	HiSeq 2000	150x2
KK_XX_a_21	female	13,088,469	12,839,537	3.93	3.85	81.06	HiSeq 2000	150x2
KK_XY_a_22	male	15,587,826	15,262,247	4.68	4.58	82.69	HiSeq 2000	150x2
KK_XY_a_23	male	20,181,900	19,801,853	6.05	5.94	81.81	HiSeq 2000	150x2
KK_XY_a_24	male	14,065,137	13,726,104	4.22	4.12	77.86	HiSeq 2000	150x2

Table S15A. Primers for the tiling PCR targeting the male-specific and adjacent regions in *T. niphobles* (long-range)

Long-range					
Forward primer	Sequence	Reverse primer	Sequence	Amplicon length (bp)	location
Sd-20f	GGTTCCAATCCAGACTCAGCA	sd-s-5r	TCGACTGTGTTGGGCCTTTG	7,924	Y-specific
sd_s_4f	CCGACTGAGTCCGTGTTGGA	sd-s-9r	GCTCATCTGCCTGGCAATATTTGT	17,703	Y-specific
dup-st-3f	CTGGCTACTTTAGCTGTCAGTGTTCC	dup-st-3r	CGGTGTATTCTTCAGCGGTACTGT	9,344	Y-specific
sd-s-10f	AAATCAGCACAGCCGCACAC	sd-s-10r	TCTTCATCACGTCGGGGTCA	3,336	Y-specific
sd-s-11f	ACTTCGGACCTGGCCATCTG	sd-s-11r	GCTCCCTGGTCTCCCGTTCT	3,200	Y-specific
sd-s-12f	TGCCAGCCTCCATCTTCTCC	sd-s-16r	TTCTCCTGTTGGCCCTCCTC	14,776	Y-specific
sd-s-15f	CCCAGAGGTGGTTCCCGTAA	sd-s-17r	GCCTTCGCAAGGAGGTCAGA	9,051	Y-specific
commonR	ACATGCAGCTCTTCCTCCTTAC	sd-9r	TGGACAGCCTTTTCCAGAGAGA	17,824	Y-specific
sd-4f	CCTCCAACCTGATTATCGTCACC	sd-9r	TGGACAGCCTTTTCCAGAGAGA	11,215	Y-specific
sd-9f	CCGTATGCGGAAGATCCAGA	sd-16r	AGCCGAGGTAAGGAGGGACA	14,848	Y-specific
sd-17f	GTTTGGCTGGTCGGCATAAA	sd-20r	TTTGCGCCACTATTCAGCAC	7,668	Y-specific
sd-20f	GGTTCCAATCCAGACTCAGCA	sd-23r	GCACTTTTGCGCATTTGGTT	9,434	Y-specific
sd-23f	GGGGAGCCGACACTTTCAG	sd-24r	CCCCATTTGTCACGTTCCAG	2,176	pseudoautosomal region
sd-25f	GGGCTCACAGGACCCCCTAT	break-1r	GAGGAAGCTTCACCAGATGCTAATGT	2,127	Pseudoautosomal region
105f	CCATCCACCACAGCTCCATC	out-sd-1r	GAGGCCGGTATTCCCCTGAA	10,425	Pseudoautosomal region
105f	CCATCCACCACAGCTCCATC	appl1-22r	GACAGTCCGCAGGTAACTACCTTCTTT	13,969	Pseudoautosomal region
Appl-280f	GAGTTCTTGTCACGCCGTCCTTTCC	Appl1-2100r	GCCTCAGACTCGCCCTTCTTCTTCC	8,308	Pseudoautosomal region

Standard PCR					
sd-s-1f	GGAGGGGGAAAGGAGTGAGG	sd-s-1r	TTGGGCTTCTTGGACGAGGA	3,259	Y-specific
sd-s-2f	CGACAGCGGGCAGCAATAAT	sd-s-2r	GCTTCGTGACGGAGGCTGAT	3,245	Y-specific
sd-s-3f	CTGGGCTACCGGAGGAGGTT	sd-s-3r	GGCCCGCCAAACTTGAAAAA	3,212	Y-specific
sd-s-4f	CCGACTGAGTCCGTGTTGGA	sd-s-4r	TCTGGAGAAAAGTGGGTCGAAGA	3,208	Y-specific

sd-s-5f	TCCTCTTGGGCATTGGGAAA	sd-s-5r	TCGACTGTGTTGGGCCTTTG	3,238	Y-specific
sd-s-6f	TGTGCTTTCCTGGACCACTCC	sd-s-6r	GGAGCGGGAAGAGAGGGTTC	3,263	Y-specific
sd-s-7f	GCCAGTCCCACGGGTGTAAT	sd-s-7r	CCACAGATGCCTCGGATGGA	3,224	Y-specific
sd-s-8f	GCCCACTGCACTCAGCAAAA	sd-s-8r	CCAGCTTGAAGCCATGATGC	3,209	Y-specific
sd-s-9f	AGGCAGCTCCAGAGCCTTCA	sd-s-9r	GCTCATCTGCCTGGCAATATTTGT	3,244	Y-specific
sd-s-10f	AAATCAGCACAGCCGCACAC	sd-s-10r	TCTTCATCACGTCGGGGTCA	3,337	Y-specific
sd-s-11f	ACTTCGGACCTGGCCATCTG	sd-s-11r	GCTCCCTGGTCTCCCGTTCT	3,201	Y-specific
sd-s-12f	TGCCAGCCTCCATCTTCTCC	sd-s-12r	GCGCCTGGCTGTCTTTCAGT	3,277	Y-specific
sd-s-13f	AAGGATGCTCCTGGGCCTTC	sd-s-13r	GCAAGGGTAACGTGGGGACA	3,223	Y-specific
sd-s-14f	GGGGTGTTGGCGGACATTTA	sd-s-14r	TGCTCCTGGTCCCAGAAAGG	3,278	Y-specific
sd-s-15f	CCCAGAGGTGGTTCCCGTAA	sd-s-15r	GACCCGCAACATCATCAACG	3,225	Y-specific
sd-s-16f	GCGAGCTTGTCGGGACCTAT	sd-s-16r	TTCTCCTGTTGGCCCTCCTC	3,205	Y-specific
sd-s-17f	GTCGCCAGCGTCCTGTCCTA	sd-s-17r	GCCTTCGCAAGGAGGTCAGA	3,206	Y-specific
sd-1f	GGCAATCAGGTGAGGTCACAG	sd-1r	CAAATCAGACGGGAGGCAAA	2,229	Y-specific
sd-2f	CAGTACCGGTATGTGTTTTTATTACCG	sd-2r	TCACCTGTGGCCGCTTTTTA	2,201	Y-specific
sd-3f	TGGAAGCGCTACAACCCAGA	sd-3r	AGCCAGTCTGCAGGGAGTCA	2,224	Y-specific
sd-4f	CCTCCAACCTGATTATCGTCACC	sd-4r	GCGGAAGATGGTTTCCAGGT	2,208	Y-specific
sd-5f	CGTCCACAAGTCGGATCTCA	sd-5r	TCCAGAAGAATCAGGCAACCA	2,206	Y-specific
sd-6f	TAGTTGGGGTTCCGCTCAGA	sd-6r	CATTTTTGTGCCCGTAAGATCC	2,224	Y-specific
sd-7f	AGCTGGGTGGGCTGAGCTT	sd-7r	CTGGTCAAGGGCGACAGAAA	2,205	Y-specific
sd-8f	TGCAAGGAGAACGGAGAGATG	sd-8r	TGGTCAAATTTTGTGCCCTGA	2,216	Y-specific
sd-9f	CCGTATGCGGAAGATCCAGA	sd-9r	TGGACAGCCTTTTCCAGAGAGA	2,237	Y-specific
sd-10f	CAGTACAGCAGCCACGCAAA	sd-10r	GGCGCATGCTGAAACATACA	2,221	Y-specific
sd-11f	GTCGTCGTTCACCGTGGTTC	sd-11r	TTTTCCGACACAATATGAGGGTTT	2,200	Y-specific
sd-12f	CGCAGTTTGGCATGAAATCC	sd-12r	CACGAAGCAACACTTTCAGCA	2,201	Y-specific
sd-13f	TGTCTCCTCCTGGCTCCAAA	sd-13r	TGAGACAAGCAGAGCTGAACAGG	2,204	Y-specific
sd-14f	GGCAGCTGTCCACTTAAACACC	sd-14r	TCCACCCAGAGTTCAGGTGAG	2,226	Y-specific
sd-15f	GGATGAACCACAGGCAGCTC	sd-15r	CTGCGGCGTTTGTTTCTCTC	2,205	Y-specific
sd-16f	TGGATGACACCTGGGTCAAAA	sd-16r	AGCCGAGGTAAGGAGGGACA	2,252	Y-specific

sd-17f	GTTTGGCTGGTCGGCATAAA	sd-17r	AAGGGGTTGGAGGATGATGC	2,204	Y-specific
sd-18f	TCTCCACACAGCGACCTCCT	sd-18r	GGTCTTCATTCCCCAAATGTGA	2,232	Y-specific
sd-19f	ACAAAATGGGGAGGGGAACC	sd-19r	TGGCCATTTTTGTTTGACGAA	2,205	Y-specific
sd-20f	GGTTCCAATCCAGACTCAGCA	sd-20r	TTTGCGCCACTATTCAGCAC	2,278	Y-specific
sd-21f	GCAAAAGCTTGCCAAACTCC	sd-21r	TTGTGTGTCCGCCATTTTCA	2,200	Y-specific
sd-22f	CGTGCACATTTTGCATTTCC	sd-22r	CGAGTCTGCCCTTCCGTTTT	2,223	Y-specific
sd-23f	CTGGAATGGATCGTCGGAAA	sd-23r	TGCACCAGATTTTGCATGGT	2,009	Pseudoautosomal region
sd-24f	GGGGAGCCGACACTTTCAG	sd-24r	GCACTTTTGCGCATTTGGTT	2,220	Pseudoautosomal region
sd-25f	CCCCATTTGTCACGTTCCAG	sd-25r	CTGGTGCGAGCCACTGAAG	2,206	Pseudoautosomal region
sd-26f	GGGCTCACAGGACCCCCTAT	sd-26r	TCACGGCTGCCCTGTCAC	2,219	Pseudoautosomal region

Table S15B. Primers for the PCR to confirm absence of female sequence in the male-specific region in *T. niphobl*es, *T.syderi* and *T. vermicularis*

Forward primer	Sequence	Reverse primer	Sequence	Amplicon length (bp)	location
TS_chr14_8812k_0.4k_F	AAAATGGAGTTAACAGCAGAGGTT	TS_chr14_8812k_0.5k_R	CGTCTCCTGAAGGATTTCCA	352 for TN, 380 for TS, 382 for TV	Y-specific
				408 for TN, 408 for TS, 407 for TV	Chr14

Table S16. Summary of Nanopore long read (PromethION) (22.73× coverage) for a male *T. snyderi*

Number of cell	1
Number of Subread	890,724
Total Bases of Subread (bp)	9,093,477,165
Mean read length of Subread (bp)	19,723
Median read length of Subread (bp)	16,866
Subread N50 (bp)	25,290

Table S17. Summary of stLFR paired-end short-read (80.43× coverage) for a male *T. snyderi*

Read length	Paired-end	Reads	Base (Gb)	Platform	Coverage
245x2	100×2	160,874,075	32	DNBSEQ-G400 (MGISEQ-2000RS)	80.43

Table S18. Summary of PacBio long read (Sequel II) (322.09× coverage) for a male *T. vermicularis*

Number of SMRT cell	1
Number of Subread	11,257,208
Total Bases of Subread (bp)	128,836,539,683
Mean read length of Subread (bp)	11,444
Median read length of Subread (bp)	9,777
Subread N50 (bp)	14,928

Table S19. Primer sequences for microsatellite markers and their genomic position used for fine mapping in Family A, B, and C of *T. snyderi*

Marker	Foward primer	Reverse primer	Position in FUGUv5	
			chromosome	bp
f1335	GAGCAACAAGCTGCATCAAA	AGCCTGCTGGGAAATACTTG	18	161,825
f1142	CTTCTCAGAACGGGTGTGGT	ATCAGCTGCCATGTGATGTC	18	199,012
f1220	GGTTGGCAGCCAGACATAAT	GGATGAAATCTCCCTTTTGC	18_un	842,772
f1334	GCAGCTCTGTGTGTGATCGT	AGGAGGATAAAAGGGCATGG	18	864,742
f1063	TGACCTCCAGCCTGACAAAT	TCAGAGCTGTGGAACCTTGA	18	1,451,492
f1438	GTTCCTGCGCCTGTATTGTT	GATCTATCCAAAGGCCGTCA	18_un	1,505,004
f77	TTGGCTAAGATGGTTTTCAC	AGATGCTGCTGCTGGTTTTACTGG	18	2,384,479
f1601	AATATCTGCTCCCTCCTCTGC	AGCAGCTGTTTCAGGCAACT	18_un	1,742,409
f204	GGTACGCTGTTCCACGAG	CACCACTACCATCAACCCCATCTT	18	4,836,362
f116	AGCGGGGTGAGAACAAATTA	ATGAGGGAGAGGAGACCACA	18	6,323,528
f687	TTCCTGCTTCTGGTTTTGCT	CGCCAATTACGCGATTATCT	18_un	115,569
f348	TTGAATGCAACCACCTTTCA	AGGTGCAGGAAACTGACAGC	18	6,885,207
f1027	GCTTTAATTGCGCGTGTGTA	TGCTTTGTGCCGCAATATAC	18	7,373,495
f1448	TGAGAATGTCCACTGTGTCCTC	GACCTCCTGGACCAGTTCAA	18_un	1,525,229
f1381	GGGGAGGGCCGTTAAATTAG	CTGCCTGAATCTCGGTGACT	18_un	1,155,540
f1356	CCATGGGAAAGTGGAACC	GAAAGGAGCTGCCAGACTGA	18_un	1,104,420
sca273-1	GCCATATGAGTTAGCCCCCTA	TCCTGTCCAGCTTAACCACA	18_un	490,760
sca273-5	GAGGGATGAAAGCAGACAGG	CACAGGAACTGACGGTTCAA	18_un	537,289
f38	AAGGGACTCTGGTTCACAGC	AACTTGGGATGTTGGGAAAA	18	7,519,234
f90	CCTACGACGAGGTGAAGGAG	CAGGTCCTGGTAGGACTGGA	18	7,852,540
sca168-2	ATGGGCGCCAAACATAATAG	ATGGTGCGTTCAAGAGCAG	18	7,976,009
f1432	AACCAAAGAAGGACGCGTTA	GTCTGGGAAGGTTCCTGTGA	18_un	1,419,398
sca708-1	AGCAGGTTGGGTGTGAAAAG	CGTTGAGTTGAATCGCTGAG	18_un	1,697,768
f1611	TTCCTGCCTCCAACATTTTC	TGCTGAGGATCATCACAAGA	18_un	1,790,732
f1618	CCGGGGTTCCTGTAGCTAAT	CAGCTACGGGCTTTGTTCTC	18_un	1,817,835
f982	CCAACAGCCAAATCCACTTT	GCTGCTCTGTCTGACAATGG	18_un	1,000,724
sca556-2	GCGAAAAGTGTGTCACAAATG	TGGTCCTGGTCTTCACACTG	18_un	1,306,897
f1385	AGCCGTTCCAGCTGTTACTC	CTGTGACGCAGGATTCCTCT	18_un	1,246,518
f1659	CATGCTGGGCTACTGTGGTA	CACAGACGATGAGCAGGAGA	18_un	1,335,649
sca1119-1	GGCATCACTGTTTTGACTGC	GGAAGCTGTTCCAGGTGTGT	18_un	1,336,102
f1433	AAGGAGGTCGTCCTCTGACA	TCACCTGTCACAACAGGTACG	18_un	1,325,055
f1393	TTCACTGGGTGGAACAATGA	TCGGAGTCTTTTGGATCAGG	19_un	1,637,578

Table S20. Primer sequences for microsatellite markers and their genomic position used for fine mapping in Family E of *T. vermicularis*

Marker	Foward primer	Reverse primer	Position in FUGUv5	
			chromosome	bp
f471	GGGTGCATTTCAGTGCTTCT	TCTCTTGCCTCCTTTCTCTCTC	10	2,726,410
f1094	TGAGAAATGCCTTGACAGCA	TAATGTGGGGGAGAGAGCTG	10	279,958
f1648	ACTCATCCACGCTTCCTGAT	AATCGAACCCAGGACCTTCT	10_un	2,750,871
f410	GATGAAATCCCTCACGGAGA	TGCGCAACATACGATGTCA	10	1,336,695
f1218	AATCCAAATCCCGATTTTCC	CCAGTATGTTTTGGCGGAGT	10_un	1,218,659
f726	TTCCTACTTCCACGCCTGTC	CCACTATGCGACTGAGGTGA	10_un	948,203
f861	AGTGGTGGAGGGTCTGTCTG	GCTTCATTAGATTGCGTGCAT	10_un	1,492,219
f1258	GCCAAGACAAACGTTCCAGT	TTCAATGGGGACACTCACAA	10	3,996,882
f202	CCCAGCAGTGCTCTGAAATA	GACGCAGGGAATCTCACAAT	10	4,777,201
f470	TGGCCTCAACTCTGTTTCAA	GTTACCACCGAAACCCTCTG	10	6,081,561
f1298	CAGGAAGCCAAAGTGCTGTT	GTGCAGGAGGAACCATTGAC	10	5,754,685
f657	GCTGTGTTTGCTCTCCATCA	GCCTGTGTGGGTTTCACCTA	10	6,673,873
f1329	ACGACATCTGGCTGGTTGAT	CATCATCCCACATCCTACCC	10	6,874,354
f1082	GTTAGGTGGCCTTGGAAACA	CCTCCATCAACCACATCTCC	10	7,148,077
f1702-2*	AGCATCCAACAGGCAGAAAT	AGCCATCTTGAATCCTGTGC	10_un	2,875,022
f1006	ACACGCACTCACTCACTTGC	AGCTTTCAGAATGCAGCACA	10	8,623,283
f1307	AGTTCGAGCTCCGCAAAGTC	TAGCCCACGTCTAACCAGTG	10	8,694,162
f1422	GCTAAAACGGACAGGCAGAG	TTGAGCTCATCCTCCCACTC	10	8,754,095
f1684-2*	GACAGAGAGGGCGATTTGAC	CCACCAGTGCAGAAACATTG	10_un	2,868,401
f1684-4*	GCGCTCAGATCAGAACACAG	TGGATCGTTGATGAAGCAGA	10_un	2,852,214
sca317-1*	ATTTGCGGAGGACAGATTCA	AGCTGCCTCATTGGTTGTTT	10_un	1,751,153
sca317-2*	TGGATGAGGTTTCCTTGACC	ATGTTTCCATGACCCAGAGC	10_un	1,753,339
f724	CCACCTGGAGAAAAGTGCTC	AAATTGATGCCATCGGACTC	10_un	1,568,157
f34-3*	TTTCCCCACTAGGAGAGCTG	GCCTGGCAGAGAGAGAGATG	10_un	2,816,704
f34-4*	AATAATGAGGCGGAGGGATT	CCCCCAACATGTTCTGATGT	10_un	2,787,913
sca297-8*	ACAGATGGAACCAGGGTCAG	GATTTGCCGAAGGTCAAAAA	10	8,813,985
sca297-6*	TGGAAGCTGCGACAATAAAG	GGAACGGTCCAAGAACCATA	10	8,852,028
sca297-5*	CCCAAGATCCTGAACAGGTG	GAACCTCCGATACGTCCAGA	10	8,854,961
sca297-3*	TTCAGATTGACGACACCTGTTT	CGGTATGGAAACGGAATGTT	10	9,088,101
f1603	GACCCCACAGGTACAGCATT	ACCTTCGCCACAACAAACTC	10_un	2,673,931
sca541-1*	AGATCAACGGGGCTGAAGT	GGATGCAGTGTGTCCTGTTT	10_un	2,334,935
sca541-3*	CACAAAAGCAGGTGCGAGTA	AAATGCTTTTGCTCTGAGGTG	10_un	2,400,464
f1374	AGGAGAGCTGTGAGGCTACG	CGGGGAAACCAGAGGATTAC	10_un	2,335,369

* Markers newly developed in this study