Recent studies have revealed a surprising diversity of sex chromosomes in vertebrates. However, the detailed mechanism of their turnover is still elusive. To understand this process, it is necessary to compare closely related species in terms of sex-determining genes and the chromosomes harboring them. Here, we explored the genus *Takifugu*, in which one strong candidate sex-determining gene, *Amhr2*, has been identified. To trace the processes involved in transitions in the sex-determination system in this genus, we studied 12 species and found that while the *Amhr2* locus likely determines sex in the majority of *Takifugu* species, three species have acquired sex-determining loci at different chromosomal locations. Nevertheless, the generation of genome assemblies for the three species revealed that they share a portion of the male-specific supergene that contains a candidate sex-determining gene, *GsdfY*, along with genes that potentially play a role in male fitness. The shared supergene spans ~100 kb and is flanked by two duplicated regions characterized by CACTA transposable elements. These results suggest that the shared supergene has taken over the role of sex-determining loci from *Amhr2* in lineages leading to the three species, and repeated translocations of the supergene underlie the turnover of sex chromosomes in these lineages. These findings highlight the underestimated role of a mobile supergene in the turnover of sex chromosomes in vertebrates.

*Significance*

Although turnover of sex chromosomes is very common in many vertebrate lineages, the transition process is still elusive. We studied the sex-determining region (SDR) of 12 congeneric fish species of *Takifugu* fish that compose an important model for the study of genomics and sex determination. We found that while nine species retained their ancestral SDR, three species had acquired derived SDRs. Although the derived SDRs resided in three different chromosomes, they harbored a shared supergene flanked by two putative transposable elements. The results highlight the understimated role of a mobile supergene in turnover of sex chromosomes in vertebrates.
one of the congeneric species, *T. niphobles*, in which the sex-determining locus mapped to a location distinct from that of *Amhr2* (15, 16). Moreover, the strong suppression of recombination around the sex-determining locus in *T. niphobles* suggested that this species potentially has sex chromosomes with an early stage of sequence differentiation. Since the ancestor of the Takifugu genus underwent rapid speciation 2–5 million years ago, the genus will provide a unique opportunity for understanding the process of rapid sex chromosome turnover and the initiation of sex chromosome differentiation, as well as the genomic changes associated with each of these.

To gain insight into the transition process between the sex-determination systems, we expanded our previous work to include 12 Takifugu species and found that sex chromosome replacements have occurred in three species. We then generated a genome assembly for each of the three species and unexpectedly found that despite the fact that the chromosomal location of the sex-determining locus is distinct among the three species, these loci share a very similar male-specific region. We speculated that repeated translocation of a preexisting sex-determining locus may be more prevalent in vertebrates than previously thought and may have caused rapid turnover of sex chromosomes.

![Fig. 1](https://doi.org/10.1073/pnas.2121469119)
Results

The Sex-Determining SNP in Amhr2 Is Absent in 3 of 12 species.

Previous studies suggested that the sex-determining role of SNP7271 in Amhr2 is not conserved in one of four Takifugu species, namely T. niphobles (16) (shown in green in Fig. 1A). To gain a more comprehensive view of the transition of sex-determining loci in this genus, we sequenced the genomic segment harboring the SNP7271 site from females and males of eight Takifugu species that had not been examined previously (Fig. 1B and SI Appendix; Table S1). We found that while the association between SNP7271 polymorphisms and gonadal sex was conserved among six species, this was not the case in the other two; the SNP locus was homozygous for the C allele (the female allele) in T. snyderi and T. vermicularis, regardless of their sex, except in one T. snyderi male individual (Fig. 1B and SI Appendix, Table S1, shown in red in Fig. 1A). We performed additional genotyping of the SNP site and found that the G allele (the male allele) was absent in wild populations of the two species examined (n = 50 for each sex of each species) (SI Appendix, Table S1).

Linkage Mapping Identifies Sex-Determining Loci. The genotyping results implied that the sex-determining loci have turned over in lineages leading to T. snyderi and T. vermicularis. Alternatively, the role of master sex determiner in these species may have been taken over by environmental factors. To test these hypotheses, we produced 98 full-sibling progenies (Family A) in T. snyderi and 62 full-sibling progenies (Family D) in T. vermicularis from wild-caught parents (SI Appendix, Table S2), and we performed genome-wide linkage mapping of the sex-determining locus. We found a strong linkage between gonadal sex and paternally inherited alleles at loci near the distal end of Chr18 in T. snyderi (logarithm of the odds [LOD] = 29.3; P = 1.8 × 10−23) and Chr10 in T. vermicularis (LOD = 18.7; P = 4.3 × 10−15) (Fig. 1C). Indeed, the markers f1659 on Chr18 and sca541-1 on Chr10 exhibited perfect segregation with phenotypic sex among the 83 progeny of T. snyderi and the 62 progeny of T. vermicularis (SI Appendix, Table S3). We further confirmed this result in additional families of T. snyderi (P < 1.0 × 10−16) and T. vermicularis (P = 1.8 × 10−20) by marker association (Fisher exact test) (SI Appendix, Table S3). The linkage analysis also revealed that the marker loci and the synteny of these markers are largely conserved between the species (SI Appendix, Fig. S1).

A Phylogenetic Framework Suggests Three Independent Turnovers.

Summing up the results we have reported here thus far and those of previous studies (15, 16), it is now evident that while the role of the Amhr2 locus in sex determination is largely conserved among Takifugu species, at least three turnover events have occurred in this genus, resulting in four distinct sex-determining loci (Fig. 1D). To infer the ancestral and derived states of the four sex-determining loci, we generated a phylogenetic tree using whole-genome resequencing data of 12 Takifugu species and one outgroup species (Fig. 1A and SI Appendix, Table S4). According to the most parsimonious interpretation of the phylogeny, the results corroborate the previous hypothesis that T. niphobles has obtained a derived sex-determining locus on Chr19, while having lost the ancestral male-determining allele (the G allele) at the Amhr2 locus on the same chromosome (16) (green in Fig. 1A). Moreover, the phylogenetic tree indicates that sex chromosomes are likely to have turned over at least twice in the lineage leading to the sister species, T. snyderi and T. vermicularis (red in Fig. 1A).

Having detected the three turnover events related to sex chromosomes within Takifugu species, we turned to the detailed characterization of the three sex-determining loci that have likely arisen recently (shown in green and red in Fig. 1A).

k-mer Analysis Identifies Male-Specific Sequences in T. niphobles.

Although we previously showed that the sex-determining locus in T. niphobles maps to Chr19 with an XX/XY system (16) (shown in green in Fig. 1A), no male-specific sequences have been found. To identify male-specific sequences in T. niphobles, we first searched for the male-specific k-mers in the resequencing data of six females and five males (SI Appendix, Table S5) and obtained the 35-mer sequences unique to males. Using these sequences, we extracted the pair-end reads from the same resequencing data and assembled them into 551 contigs with a total of 218-kb sequences (SI Appendix, Table S6). We then masked repetitive sequences in the contigs and used the Basic Local Alignment Search Tool (BLAST) to search against a fugu reference genome sequence, FUGU5 (23). This revealed that similar sequences exist on multiple chromosomes in T. rubripes, including Chr6, Chr12, and Chr22 (SI Appendix, Table S7).

Construction of a Chromosomal-Scale Genome Assembly of a T. niphobles YY Male.

As the k-mer approach suggested that the male-specific region of T. niphobles is likely absent in the fugu reference genome, we attempted to generate a reference genome sequence for T. niphobles, including its Y chromosome and the male-specific region. Complete sequence assemblies of Y chromosomes have not been generated for many organisms (but see some exceptions in refs. 24–28) because of technical difficulties related to the discrimination of X and Y in the male genome and/or the assembly of regions containing high levels of repetitive sequences. To alleviate these problems, we sequenced the genome of a male T. nobiles carrying two Y chromosomes (instead of X and Y) and generated a draft genome sequence using an hybrid assembly strategy with PaCBio long reads and Illumina short reads (SI Appendix, Fig. S2 and Tables S8 and S9). The assembly consists of 728 contigs with an N50 of 6.08 Mb and a total size of ~396 Mb. Further scaffolding with Hi-C data placed 288 contigs into 22 superscaffolds (pseudo-chromosomes) that had chromosomal-level length (longer than 10 Mb) (SI Appendix, Fig. S3B). The size of these pseudo-chromosomes covered 91.2% of the total genome and their number was consistent with the karyotype of this species (29) (SI Appendix, Fig. S3B and Tables S10 and S11). The total number of scaffolds in the Hi-C–based, proximity-guided assembly was 424. The scaffolding increased N50 from 6.1 Mb to 16.2 Mb and that of the longest scaffold from 15 Mb to 29.8 Mb (Table 1 and SI Appendix, Fig. S3B). A Benchmarking Universal Single-Copy Orthologs (BUSCO) search against the 4,584 single-copy orthologs for Actinopterygii suggested that the assembly was missing only 2.3% of the core genes (SI Appendix, Table S12). Furthermore, ~95% of the short reads from multiple male individuals were uniquely mapped to the assembled genome with a mapping quality above 20. These metrics indicated the high quality of the chromosomal scale assembly. The assembly showed highly conserved synteny with the fugu reference genome (SI Appendix, Fig. S4) and, therefore, chromosome identities for the T. niphobles were assigned based on those of T. rubripes. Based on the conserved synteny shown here and the genomic positions of the sex-linked...
markers reported previously (16), we identified the pseudoautosomal chromosome corresponding to the *T. niphobles* sex chromosome (Chr19), which comprises 23 contigs totaling 17 Mb in length (*SI Appendix*, Fig. S5 and Table S11). Because this pseudoautosomal chromosome was assembled from the YY male, we denoted it as Chr19Y.

The Male-Specific Region Is Flanked by Pseudoautosomal Regions in *T. niphobles*. To determine the male-specific region in this assembly, we aligned resequencing reads from 8 females and 10 males against the assembly and compared the relative depth of coverage between males and females. We detected a clear difference in the coverage in the region (~245 kb) on Chr19Y (Fig. 2A), suggesting that the region may be specific to males. The male-specific region is composed of two subregions with different relative coverage: the left-half region (130 kb) showed a relatively higher male to female ratio than the right-half region (115 kb) (Fig. 2A). The absence of differences in coverage depth between sexes along the flanking regions indicates that these regions are shared between Y and X chromosomes.

Accumulation of Large Repetitive Sequences Is Heterogeneous in the Male-Specific Region of *T. niphobles*. Accumulations of large repetitive sequences have been observed in the nonrecombining regions of sex chromosomes across diverse taxa (30, 31). To determine if this is the case in *T. niphobles*, we aligned Chr19Y to itself and found that the repeat accumulation was still modest in about half of the male-specific region (130 kb) (Fig. 2B; shaded in green), while the other 115 kb contained many large, repetitive sequences (Fig. 2B; shaded in pink). Furthermore, the self-alignment plot implied that the 115-kb region is likely derived from the duplication of a region residing on the pseudoautosomal region near the male-specific 130-kb region (Fig. 2B; shaded in orange).

Repetitive Sequences Are Accumulated in the Male-Specific Region of *T. niphobles*. The repeat sequence analysis revealed accumulation of repeat sequences within the 245-kb male-specific regions; repeat sequences occupied 42.5% and 62.2% of the 130-kb and 115-kb regions, respectively (Fig. 2C and *SI Appendix*, Table S13). While these values are greater than that of the whole Chr19Y (14.9%), the pseudoautosomal regions near the male-specific region also harbor repeat-rich sequences (*SI Appendix*, Fig. S5). For example, repeat sequences occupied 46.3% of the ~1-Mb pseudoautosomal region (ranging from the position 2 Mb to the proximal boundary to the male-specific region).

### Table 1. Assembly statistics for the genomes of *Tkifugu niphobles*, *T. snyderi*, and *T. vermicularis*

| Genotype          | T. niphobles | T. snyderi | T. vermicularis |
|-------------------|--------------|------------|-----------------|
| Sequencing and scaffolding strategies | PacBio, Illumina, Hi-C | Nanopore, stLF, Illumina | PacBio, Illumina |
| No. of scaffolds/contigs | 424         | 1,097      | 1,756           |
| Total length, Mb | 396         | 399        | 429             |
| Average length, kb | 935         | 363        | 244             |
| Longest scaffold/contig, Mb | 29.8 Mb      | 22.3 Mb    | 8,669 kb        |
| Shortest scaffold/contig, bp | 1,000        | 162        | 2,979           |
| Scaffold/contig N50, Mb | 16          | 6          | 1.6             |

stLF, single-tube long-fragment read.

A Candidate Sex-Determining Gene in *T. niphobles* Resides in the Male-Specific Region Flanked by Putative Transposable Elements. Since the *k*-mer analysis suggested that part of the male-specific region may have been derived from an autosomal region, we tested this possibility by performing BLAST searches of the male-specific sequences (130 kb + 115 kb) against the reference genome sequences of *T. niphobles* and *T. rubripes* (*FUGU5*).

We found that the majority of the 130-kb male-specific region was composed of small segments that showed high sequence similarity to portions of other chromosomes, suggesting that the transposed sequences were the main source of the male-specific region (Fig. 2D). Furthermore, we annotated seven protein-coding genes that are paralogous to the genes on three different autosomes (Fig. 2C; *GsdfY, Ppef2Y, Nup54Y, Phyh1Y, Cbx4Y*, and two LOC105418166Y (uncharacterized protein-coding genes that appear to be conserved in teleost fish). Of these, *GsdfY* is worth noting since the ortholog of this gene has been reported as the master sex-determining gene in a teleost species, *Oryzias luzonensis* (18). Furthermore, loss of function of this gene was found to result in female-to-male sex reversal in two teleost species, Nile tilapia and medaka (32, 33). Our RNA-seq analysis confirmed the paralog-specific expression from this locus in differentiating male gonads, along with the expressions of *Ppef2Y* and *Nup54Y* (*SI Appendix*, Fig. S6 and Table S14). As for *Phyh1Y, Cbx4Y*, and *LOC105418166Y*, their expression of either the male-specific or autosomal paralogs was observed, but paralog-specific expression could not be ensured, due to the lack of diagnostic nucleotide sites that can distinguish the transcripts of these paralogs.

Intriguingly, the sequence annotation also revealed that the male-specific region is flanked at both ends by two CACTA transposable elements containing the catalytic center of the transposase (Fig. 2C). CACTA elements, also described as EnSpm elements, belong to the class II transposons that utilize a cut-and-paste mechanism for their transposition (34–36). A full-length CACTA element contains two open reading frames, one encoding a transposase and the other a protein of unknown function. Although we were unable to show solid evidence that the two CACTA elements of *T. niphobles* are structurally intact, we found that the catalytic center of the transposase called the DDD/E domain is conserved (*SI Appendix*, Fig. S7).

The search involving the other half of the male-specific region (115 kb) identified six protein-coding genes (*Cyc1Y, Hipk1Yb, Hipk1Yc, Hipk1Yd, Hipk1Ye*, and *Pfbr2Y*), though their roles in sex determination have not been reported (Fig. 2C). Each gene had more than one untruncated or truncated paralog on Chr19Y at the outside of the male-specific region (Fig. 2D), corroborating the hypothesis that the 115-kb region originated from duplication events on the pseudoautosomal regions (Fig. 2B). Among
these paralogs, the expression of either the male-specific or pseudautosomal paralogs of Cyc1Y was observed in the developing gonads. However, the paralog-specific expression was not assessed, due to the lack of diagnostic nucleotide sites.

**Verification of the Male-Specific Region of T. niphobles.** To verify the overall accuracy of the assembled sequences of the 130-kb male-specific region and its adjacent pseudautosomal regions, we designed 12- and 5-primer pairs targeting each of the two regions, respectively ([SI Appendix, Table S15]), and successfully obtained all amplicons. Sequence alignment revealed that the amplicon contigs were largely concordant with the target region of the Chr19Y assembly ([SI Appendix, Fig. S8]). Discordant regions comprised only 1.2% of the amplicon contigs and 4.5% of the target region of the Chr19Y assembly. The mismatch may have been due to misassembly, misamplification of amplicons, or true polymorphisms between individuals used for the tiling polymerase chain reaction (PCR) and the Chr19Y construction. We also confirmed the male specificity of the region using a wild population (n = 20 for each sex; P = 3.0 × 10⁻¹⁰) with a diagnostic primer pair that could yield amplicons of different size from the male-specific region (352 bp) and its paralogous region (408 bp) ([SI Appendix, Fig. S9]). Although one potential sex-reversed fish was observed, similar incomplete penetration of the sex-determining locus has been reported in experimental families of this species (16). Thus, the reliability of the assembly results of the 130-kb male-specific region and one of its adjacent region in Chr19Y was confirmed. Note that we were not able to apply this approach to the 115-kb region, due to its highly repetitive nature.

PNAS 2022 Vol. 119 No. 23 e2121469119

https://doi.org/10.1073/pnas.2121469119
k-mer Analysis Identifies Male-Specific Sequences in T. snyderi and T. vermicularis. We used similar strategies to characterize the sex-determining locus of T. snyderi on Chr18 and that of T. vermicularis on Chr10 (shown in red in Fig. 1A). First, we searched for the male-specific 35-mers in the resequencing data of 24 individuals for each sex of T. snyderi (three pools of both females and males, each pool containing eight individuals) and four individuals for each sex of T. vermicularis (SI Appendix, Table S5), and assembled them into contigs (SI Appendix, Table S6). BLAST searches of the contigs against the T. niphobles assembly revealed that many of the male-specific contigs in both species showed high similarity to part of the male-specific sequence in T. niphobles. For example, the male-specific contigs in both species contain a coding sequence of GsdfY (SI Appendix, Table S7). This result was unexpected, as the chromosomal locations of the sex-determining loci in the three species are distinct from each other, raising the possibility that the male-specific region is shared among the three species, at least in part.

Construction of the Genome Assemblies of T. snyderi and T. vermicularis. We then generated a genome assembly for T. snyderi and T. vermicularis, including their Y chromosomes. Because individuals with the YY genotype were not available, the genome of an XY male was sequenced for each species. We used Oxford Nanopore long reads and MGI single-tube long-read fragment reads for T. snyderi, and PacBio long reads and Illumina short reads for T. vermicularis (SI Appendix, Tables S16–S18). The male assembly for T. snyderi consisted of 1,097 scaffolds with an N50 of 6.0 Mb and a total size of ~399 Mb, whereas the male assembly for T. vermicularis comprised 1,756 contigs with an N50 of 1.57 Mb and a total size of ~428 Mb (Table 1). A BUSCO search against the 4,584 single-copy orthologs for Actinopterygii suggested that only 1.2% and 2.5% of the orthologs for Actinopterygii appeared to be conserved but located on different chromosomes in T. niphobles, 61 kb in T. snyderi, and 105 kb in T. vermicularis. The gene contents in these regions were found to be similar but distinct. For example, while GsdfY, Ppef2Y, Cbx4Y, LOC105418166Y, and LOC105418166B were shared among all three species, Nup54Y was observed only in T. niphobles. Moreover, T. vermicularis appeared to have lost Phyd1Y but acquired an additional copy of Cbx4Y and LOC105418166Y (Fig. 3C and SI Appendix, Fig. S10).

To understand the history of the male-specific region that appeared to be conserved but located on different chromosomes in the three species, we constructed phylogenetic trees for each gene in the male-specific region and its autosomal paralog(s) across the three species (Fig. 3D). This phylogenetic analysis identified three main branching patterns (Fig. 3D and SI Appendix, Figs. S11 and S12). The pattern of Gsdf indicated that the male-specific region (red in Fig. 3D) and the autosomal paralogs (blue in Fig. 3D) in the three species diverged after their split with the lineage leading to T. rubripes, but before the diversification of the three species. The second pattern, including Ppef2 and Nup54, suggests that the topology at the crown age of the genus is ambiguous. The third pattern, including Cbx4, Phyd1, and LOC105418166, suggested that these paralog pairs diverged before the split of the three species with T. rubripes. These results indicate that the male-specific region may have been formed due to capture of GsdfY by the preexisting segment containing other paralogs. However, the presence of a single homolog for these genes (excluding LOC105418166) in T. rubripes (Fig. 3C and D) raises questions about this theory, although lineage-specific gene loss in T. rubripes is one possible explanation. Alternatively, an elevated rate of evolution due to the reduced efficacy of purifying selection in the nonrecombining region may underlie the gene topology in which gene trees are discordant with the species phylogeny (37). For example, the rapid evolution of the male-specific paralogs (e.g., Phyd1Y) may result in “long branch attraction” in which these paralogs are clustered with an outgroup homolog (38). Furthermore, positive selection in the male-specific paralogs after the gene duplication but before speciation would give rise to the third pattern (39).

Despite the heterogeneous topologies observed in the gene trees, all paralog pairs in the three species were clustered not according to species genealogy but rather to paralog groups (red in Fig. 3D and SI Appendix, Figs. S11–S13), suggesting that the male-specific genes were likely transposed or translocated as a unit into the current genomic positions after the initial formation of the supergene. When focusing on the male-specific paralogs for each species, duplicates of the gene flanking the male-specific region, LOC105418166Ys, were clustered according to species group (Fig. 3C). Duplicates of Cbx4Y in T. vermicularis were also clustered according to species group. Thus, it is likely that the lineage-specific duplication of genes occurred in the gene casette of the male-specific region.

Male-Specific Regions Are Largely Conserved among the Three Species. Manual annotation of the male-specific regions in T. niphobles (130 kb), T. snyderi (79 kb), and T. vermicularis (105 kb) revealed similar characteristics of these regions; they harbor GsdfY and some other shared genes, and are flanked by two CACTA transposons (Fig. 3C). The sizes of the segments are also similar: 115 kb in T. niphobles, 61 kb in T. snyderi, and 105 kb in T. vermicularis. The gene contents in these regions were found to be similar but distinct. For example, while GsdfY, Ppef2Y, Cbx4Y, LOC105418166Y, and LOC105418166B were shared among all three species, Nup54Y was observed only in T. niphobles. Moreover, T. vermicularis appeared to have lost Phyd1Y but acquired an additional copy of Cbx4Y and LOC105418166Y (Fig. 3C and SI Appendix, Fig. S10).

The Assembled Genomes for T. snyderi and T. vermicularis Contain Their Male-Specific Regions. To determine the male-specific region in the assembly for T. snyderi, we aligned resequencing reads from females and males and compared the relative depth of coverage between them. This analysis revealed a clear difference in coverage between sexes on scaffold373 (79 kb) (Fig. 3A). A BLAST search revealed conserved synteny between this region and the 130-kb male-specific region harboring GsdfY in T. niphobles (Fig. 3C and SI Appendix, Fig. S10A). We used the same approach for T. vermicularis, and found coverage differences between sexes on contig utg0005841 (97 kb) (Fig. 3B). As in the case of T. snyderi, sequence similarity was observed between this contig and the 130-kb male-specific region in T. niphobles. In addition, this contig contained large inverted duplications and a repeat-rich region where paired-end reads could not be properly mapped (SI Appendix, Fig. S10B).

The association of the male-specific region with phenotypic sex in a wild population of both species (n = 19 and 20 for each sex, respectively, of both T. snyderi and T. vermicularis) was supported using a diagnostic primer pair that could produce amplicons of different sizes from the male-specific region and its paralogous region (P = 5.7 × 10−11 and P = 3.3 × 10−9, respectively) (SI Appendix, Fig. S9).
autosomal paralogs) excludes a simple introgression as an explanation of the discordance. To examine the degree to which ancestral polymorphisms have affected the divergence of Takifugu, we used topology weighting by iterative sampling of sub-trees analysis and the f4 test (SI Appendix, Figs. S14 and S15). Although the distinction between the two scenarios (i.e., incomplete lineage sorting and ancient introgression of the supergene) was unclear, these analyses suggested that a substantial amount of shared polymorphism has shaped the speciation of Takifugu, with the evidence of gene flow between T. niphobles and the last common ancestor of T. niphobles and T. vermicularis (SI Appendix, Fig. S15).

Discussion

Based on a combination of extensive genetic and genomic data, this study documents evidence that the evolution of the male-specific supergene underlies the replacement of the sex-determining gene in an ancestor of a subset of Takifugu fishes and that subsequent translocation of the supergene drove a rapid sex chromosome turnover without a simultaneous change in the sex-determining gene. Presumably, the accumulation of genes in the male-specific region conferred a selective advantage of this region over the ancestral sex-determining gene, Amhr2. However, the insertion of the supergene could have caused sex chromosome differentiation that resulted in a selective disadvantage, namely the accumulation of repeats (toxic Y hypothesis) (42), which may, in turn, have facilitated translocation of the core part of the supergene at relatively short intervals.

Formation of the Male-Specific Supergene. Supergenes are sets of tightly linked genes that contribute to a complex phenotype (43–48). Sex-determination regions can be regarded as supergenes because suppressed recombination, if present, causes cosegregation of genes linked to the sex-determining gene (49).

Although direct evidence is lacking in Takifugu fish, the genes linked to Gsdf/Y in the core male-specific region may play...
a role in testis differentiation, because either they or their paralogs are expressed in the male gonad during gonadal differentiation (SI Appendix, Fig. S6). Moreover, studies in other animals have suggested that some genes in the male-specific region of the three *Takifugu* fishes are involved in gonadal differentiation and/or sexual behavior. For example, *Nup54* regulates nuclear-cytoplasmic transport, whose possible role in gonadal differentiation has been proposed based on its conserved expression in teleost fish (50). In *Drosophila*, this gene is essential in transposon silencing in the ovary (51), and there is evidence suggesting that it plays a role in sexual conflict by influencing female behavior (52). Moreover, this gene is contained in a supergene under sexual antagonistic selection in rainbow trout (53). *Ppef2* is predominantly expressed in photoreceptors and the pinacle gland, and it participates in phototransduction in mammals. This gene is also contained in the rainbow trout supergene under sexual antagonistic selection (53). *Cbx4* codes for a member of the Polycomb-group proteins that control cell identity and development (54). Although its role in sexual dimorphism has not been reported, *Cbx2* codes for another member of the group and is required to stabilize the testis pathway in mice (55).

**The Evolutionary Development of the Core Male-Specific Region.** Although supergenes contribute to many fascinating phenotypes (11–14, 44–48), their evolutionary development is often only incompletely understood. In particular, the analysis of sex-determining regions has been difficult due to the extensive accumulation of repetitive sequences (see some exceptions in refs. 24–28). Our study provided a unique opportunity to infer the process involved in assembling the male-specific supergene and its subsequent modifications during repeated translocation.

It is likely that the core male-specific region evolved in the common ancestor of the three species after diverging from the lineages leading to *T. rubripes* and *T. obscurus* through the combination of at least two processes. One includes segmental duplication encompassing *Nup54, Ppef2*, and *Gasdf* on Chr6 and the translocation of this region to the future sex chromosome (Fig. 3C and SI Appendix, Figs. S13 and S16), given that the genomic arrangement of the three genes adjacent to each other is a conserved feature among teleost fish (50). The other process is the gathering of other unlinked genes (Phyhd1, *Cbx4*, and *LOC105418166* through independent duplications and translocations (Fig. 3C and SI Appendix, Fig. S16). It is possible that the latter process partly predates the former, as the phylogenetic analysis in this study implied that the emergence of *Phyhd1Y* and *LOC105418166Y* occurred prior to that of *GasdfY*. The original segment, if this is the case, was probably not male specific until it captured *GasdfY*. Based on the core structure formation, it is parsimonious to consider that *Nup54Y* was lost before the divergence of *T. snyderi* and *T. vermicularis*, while the gene content was retained in the lineage leading to *T. niphobles* (Fig. 3C and SI Appendix, Fig. S16). 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 *LOC105418166Y* and *Cbx4Yb* (Fig. 3C and SI Appendix, Figs. S10 and S16). The variations in gene content in the shared male-specific region suggest that adaptive replacements of genes may have taken place in these species and could be ongoing.

**The Jumping Sex-Determining Locus.** In vertebrates, transposition or translocation of a preexisting sex-determining locus has only been reported in salmonids (56). Comparative analyses of the sex-determining loci in the three salmonid species (57) suggested that an ∼4.1-kb sequence containing the sex-determining gene *sdY* is conserved along with transposable elements. Although an RNA-mediated mechanism has been proposed for the transposition of the *sdY* locus in salmonids, different mechanisms likely played a role in the movement of the male-specific region in *Takifugu*, as the size of the supergene spans ∼100 kb. A notable feature of the gene arrangement of the conserved male-specific region is that the region is flanked by two duplicated regions harboring CACTA transposable element and *LOC105418166Y* (Fig. 3C). Therefore, the flanking transposons are candidate drivers for relocating the male-specific region. However, we were unable to find the terminal inverted repeats that are expected to reside at the ends of transposons if a cut-and-paste process is involved in transposition. Although it is possible that the terminal inverted repeats have degenerated, the same orientation of the two transposons is incompatible with the standard model of cut-and-paste transposition (34). Based on the orientation of the two flanking regions, we prefer the hypothesis that nonallelic homologous recombination played a role in the translocation. This process is thought to contribute to duplication and translocation events in animals and plants (58–60) and suggests that recombination between two directly oriented homologous sequences (typically, repeats or transposable elements) can lead to deletion and insertion of the genomic region flanked by these sequences (61). As a result, duplicated and translocated fragments are frequently bordered by transposable elements (62). Of note, this model is only speculative in *Takifugu*, and the mechanism underlying the translocation of supergene is still undetermined.

**Roles of Translocations in Accumulation and Purging of Deleterious Sequences.** It is hypothesized that sex chromosomes start to differentiate when recombination around the sex-determining region stops. While the accumulation of deleterious mutations on the differentiating sex chromosomes will eventually reduce fitness (6, 42, 63, 64), the “jumping” ability of the sex-determining locus is a way to purge the deleterious mutations (56). Consistent with these hypotheses, our analysis of k-mers and read coverage suggested that half of the male-specific region (115 kb), which contains many large repetitive sequences in *T. niphobles*, appeared to be purged from *T. snyderi* and *T. vermicularis* after the translocations. The 115-kb region may have arisen due to the reduced recombination triggered by the insertion of the 130-kb region containing *GasdfY* in an ancestor of the three species. However, it is possible that the 115-kb region has recently emerged in *T. niphobles* in a species-specific manner. The distinction of the two scenarios was challenging due to the presence of multiple duplicated loci corresponding to genes in the *T. niphobles* 115-kb region in the genomes of *Takifugu* fishes, which made it difficult to distinguish orthologs and paralogs. To fully evaluate the consequence of repeated translocation of the sex-determining supergene in the three species, more contiguous sequences of the genomes for *T. snyderi* and *T. vermicularis* are needed.

**Remarks** This study uncovered a process involved in the replacement of both sex chromosomes and sex-determining genes in closely related species and highlighted the importance of repeated translocation of the sex-determining region. Moreover, the construction of contigs for the jumping sex-determining supergene revealed that it has a unique structure flanked by two putative transposable elements. Outside of vertebrates, the translocation or transposition of a preexisting sex-determining gene was recently characterized in insects (65) and plants (66, 67). The repeated
translocation of a sex-determining locus may be more prevalent in vertebrates than previously thought and may have caused rapid turnover of sex chromosomes.

**Materials and Methods**

**Conservation of the Sex-Determining SNP in Amhr2.** The genotype at SNP7271 on Amhr2 is associated with phenotypic sex in fugu (15, 16). We previously examined whether the association is conserved in *T. pardalis, T. poecilhon tus,* and *T. niphobles,* and found that while it was identified in *T. pardalis* and *T. poecilhon tus,* the locus was fixed with the female allele (C allele) in *T. niphobles* (15, 16). In this study, we extended the analysis to eight Takifugu species, namely *T. chrysops, T. stictodon, T. obscurus, T. ocellatus, T. xanthopterus,* *T. porphyreus,* *T. snyderi,* and *T. vermicularis,* and genotyped a total of 342 wild-caught and 15 aquacultured fish by a direct-sequencing approach. The genomic region containing exon2 of Amhr2 and its adjacent regions was PCR amplified and directly sequenced, as previously reported (16). The sequence data are provided in SI Appendix (Fasta1_amhr2 exon9.fas). Additional genotyping of SNP7271 of *T. snyderi* (*n* = 100) and *T. vermicularis* (*n* = 100) was performed by high-resolution melting analysis using a Lightscanner (Idaho Technology) following the method described in ref. 68. Information on the sampling locations, sex, and association analysis (Fisher exact test) results of the fish of each species is provided in SI Appendix, Table S1.

**Linkage Mapping.**

**Experimental families of *T. snyderi.*** Three families (families A, B, and C) of *T. snyderi* were produced between three wild males and a single wild female caught at Suruga Bay, Shizuoka Prefecture, Japan (SI Appendix, Table S2). Procedures for artificial fertilization and rearing conditions were defined as previously reported for *T. niphobles* (69). Fish at 122 and 174 d postfertilization (dpf) were combined and used for the following analyses. Phenotypic sex was visually determined under a microscope.

**Genotyping of SNP markers in *T. snyderi.*** Genotyping by random ampiclon sequencing-direct technology (70, 71) was used to obtain the genome-wide genotype data from 98 fish in Family A. Genomic DNA was extracted from the caudal fin using the Gentra Puregene tissue kit (Qiagen). Library construction and sequencing on an Illumina HiSeq 2500 platform (paired end 100 bp) were conducted by Eurofin Genomics as described in ref. 71. Genotyping procedures are described in the SI Appendix, Methods.

**Linkage map construction and mapping of the sex-determining locus in *T. snyderi.*** The sex-determining locus was analyzed by the interval mapping observed at the distal end of Chr18 is family specific. To test this possibility, we used 83, 25, and 62 fish from families A, B, and C, respectively, and geno
typed markers f1618 and f1659, which were previously mapped near the distal end of Chr18 in the fugu genome. The association between gonadal sex and genotype was assessed by Fisher exact test (SI Appendix, Table S3). To increase the resolution, we genotyped two marker loci (f1335 and f1433) on Chr18 for 765 fish from the three families (342, 158, and 265 fish in families A, B, and C, respectively). After identifying 30 recombinants, 26 microsatellite markers flanked by the two markers on Chr18 were genotyped for these individuals (SI Appendix, Fig. S1). Genotyping procedures are described in ref. 23. Information on the markers is provided in SI Appendix, Table S19.

**Experimental families of *T. vermicularis.*** For *T. vermicularis,* two pairs of wild-caught parents from Ariake Sea (Nagasaki Prefecture, Japan) were mated independently to produce two families, families D and E (SI Appendix, Table S2). Fish production and phenotyping were performed as for *T. snyderi.*

**Linkage map construction and mapping of the sex-determining locus in *T. vermicularis.*** We first conducted a genome-wide linkage analysis of the sex-determining locus using 62 fish at 127 dpf in family D with 138 genome-wide microsatellite markers (SI Appendix, Table S2). The linkage between phenotype and genotype observed at the distal end of Chr10 in family D was confirmed using 164 fish in family E sampled at 131 to 161 dpf by genotyping 35 markers anchored on Chr10 in *T. rubripes.* The genotype information is provided in SI Appendix (mapping genotype.xlsx).

**The phylogenetic framework of Takifugu.** To generate a species phylogenetic tree, we first obtained whole-genome resequencing data from 12 Takifugu species and one outgroup (one individual per species): *T. rubripes, T. snyderi, T. vermicularis, T. niphobles, T. pardalis, T. poecilhon tus, T. chrysops, T. stictodon, T. obscurus, T. ocellatus, T. xanthopterus, T. porphyreus,* and *Torquigener hypselogeneon* (outgroup). We then used RaxML (version 0.8) to construct the phylogenetic tree (see details in the SI Appendix, Methods).

**k-mer Analysis to identify male-specific sequences.** A detailed description of k-mer analyses of *T. niphobles,* *T. snyderi,* and *T. vermicularis* is provided in the SI Appendix, Methods.

**Genome Assembly in a *T. niphobles* YY Male.**

**Long- and short-read sequencing.** We sequenced the genome of a YY individual on a PacBio Sequel platform. The production of the YY fish was reported previously (16). In brief, we first identified an XY sex-reversed female using sex-linked microsatellite markers (f2003 and f2006) to cross with an XY male. The offspring consisted of individuals carrying XX, XY, or YY sex chromosome sets. Then the YY fish was selected based on the sex-linked markers f2003 and f2006. Genomic DNA of the YY fish was isolated from the caudal fin by the conventional phenol/chloroform/isoamyl-alcohol method. Size selection, library preparation, and sequencing were conducted at the National Institute of Genetics. Approximately 56.6-Gb sequences with total reads of 5,445,262 (N50 = 17 kb) were derived from eight PacBio SMRT cells (SI Appendix, Table S8). Genomic DNA from the YY fish was also used for whole-genome Illumina resequencing. A TruSeq DNA PCR-Free library was prepared and sequenced (2 × 245 bp) on the HiSeq 2500 platform at the National Institute of Genetics. The data have been registered in the DNA Data Bank of Japan Sequence Read Archive (accession no. DRA012992).

**Hybrid genome assembly, Hi-C scaffolding, and gene annotation.** Detailed procedures are provided in the SI Appendix, Methods. In brief, Illumina reads were first trimmed and assembled into contigs. We then obtained an initial hybrid assembly by anchoring the contigs to the PacBio long reads. We corrected potential sequencing errors in the initial assembly by realigning the long reads and the Illumina contigs and polishing it. For long-read-only assembly, a Pairwise mApping Format file was generated using Minimap2 (73) and converted into an assembly graph using Miniasm (74). Then, a de novo assembly sequence (unitig) was extracted from the assembly graph file. We polished the two assemblies and merged them. 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 SI Appendix, Fig. S2. Gene annotation was conducted using MAKER (version 3.01.02) (75), which utilizes both evidence-based methods (RNA-seq data) and ab initio predictions. RNA-seq data were obtained from the differentiating gonads of *T. niphobles* at 90 dpf.

**Characterization of the male-specific region.** 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 (SI Appendix, Methods and Table S5). The data were registered in the DNA Data Bank of Japan Sequence Read Archive (accession no. DRA012877). We then mapped the resequencing data of 10 males and 8 females (SI Appendix, Table S5) onto the genome assembly of the *T. niphobles* YY male using the Burrows-Wheeler Aligner Maximal Exact Match (BWA-MEM) algorithm (76) (SI Appendix, Methods). 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 aligned the male-specific regions to the genome assembly using the numper algorithm with option -l 50 -c 65 implemented in MUMmer (version 4.0) (77). We restricted our analyses to the genomic regions and visualized synten blocks longer than 300 bp with 90% sequence identity. We verified the contiguity of the assembled sequence by tiling PCR. Moreover, the presence of the male-specific region in the wild population of *T. niphobles* was confirmed by a diagnostic primer. A detailed description of these strategies is presented in the SI Appendix, Methods.

**Genome assembly in a *T. snyderi* XY male and a *T. vermicularis* XY male.** The whole-genome sequence of a *T. snyderi* XY individual from Lake Hamana was obtained using PromethION long reads (Nanopore Technologies) and MGI single-tube long-read fragments (MGISEQ-2000RS) (SI Appendix, Tables S16.
and S17), whereas that of T. vermiculifer XY individual from Ariake Sea was obtained using PacBio long reads (SI Appendix, Table S18) and Illumina paired-end short reads (Sample identifier: 59; SI Appendix, Table S5). Detailed procedures for DNA extraction, sample preparation, sequencing, genome assembly, genome annotation, and identification and characterization of the male-specific region are provided in the SI Appendix, Methods.

**Phylogenetic analysis of male-specific genes and their autosomal paralogs.** To determine the phylogenetic relationships of the male-specific genes and their autosomal paralogs, we used RAxML (version 0.8) (SI Appendix, Methods).

**Admixture analysis.** To investigate the traces of incomplete lineage sorting and/or past introgression in Takifugu species, we used resequencing data from four T. rubripes individuals, five T. nigripinnis individuals, five T. synderi individuals, three T. vermiculifer individuals, and a single individual for T. chrysops, T. poecilonotus, and T. pardalis. A single individual of Torquigener hypsolege-neon was used as outgroup for all analyses. We quantified the genealogical relationships throughout the genome of Takifugu using topology weighting by iterative sampling of sub-trees (78). We also investigated admixture/introgression across the genomes using Patterson D statistics (79, 80), fₐ-ratio statistics (81), and Fbranch statistics (82) (SI Appendix, Methods).

**Data Availability.** Next-generation sequencing (NGS) reads data have been deposited in DNA Data Bank of Japan Sequence Read Archive (https://www.ddbj.nig.ac.jp/di/index.html) (DBA012890, DBA012891, DBA012877, DBA012899, DRA012676, DRA012992, DRA012876, DRA012885, DRA014181, and DRA014182). All study data, other than the NGS data, are included in the article and/or supporting information.

**ACKNOWLEDGMENTS.** This work was in part supported by Grants-in-Aid for Scientific Research (grants 22H00377, 2210002, 24380102, 15H04542, 18H02277, 16K14966, 18050815, and 17H06425). This work was also supported by the Cooperative Research Grant of the Genome Research for Bioresource, NODAI Genome Research Center, Tokyo University of Agriculture, and JST-Mirai Program (grants JPMMJ18CH and JPMMJ121C). The authors are grateful to Shuya Kato for helping with the data analysis and anonymous reviewers for their valuable comments and suggestions that improved this article.

Author affiliations: *Fisheries Laboratory, University of Tokyo, Hamamatsu 431-0214, Japan; †School of Marine Biosciences, Kitasato University, Sagamihara 252-0373, Japan; Laboratory for Phyloinformatics, RIKEN Biosystems Dynamics Research, Kobe 650-0047, Japan; ‡Genome Informatics Laboratory, National Institute of Genetics, Mishima 411-8540, Japan; §Nodal Genome Research Center, Tokyo University of Agriculture, Setagaya 156-8502, Japan; and † Comparative Genomics Laboratory, Center for Information Biology, National Institute of Genetics, Mishima 411-8540, Japan

Author contributions: S. Hosoya and K.K. designed research; A.K., R.I., S. Hosoya, D.F., K.A., T. Tajima, A.N., M.K., O. Nishimura, S.K., A.T., S. Tasumi, and K.K. performed research; T.K. and O. Nakamura contributed new reagents/analytic tools; A.K., R.I., S. Hosoya, D.F., K.A., T. Tajima, T.K., S. Hirase, M.K., O. Nishimura, S.K., Y.N., A.T., S. Tasumi, and K.K. analyzed data; and A.K., R.I., S. Hosoya, S. Hirase, M.K., O. Nishimura, S.K., H.K., A.T., S. Tasumi, and K.K. wrote the paper.
54. N. M. Luis et al., Regulation of human epidermal stem cell proliferation and senescence requires polycomb-dependent and -independent functions of Cbx4. Cell Stem Cell 9, 233-246 (2011).

55. S. A. Garcia-Moreno et al., CB2 is required to stabilize the testis pathway by repressing Wnt signaling. PLoS Genet. 15, e1007895 (2019).

56. S. Bertho, A. Herpin, M. Schartl, Y. Guiguen, Lessons from an unusual vertebrate sex-determining gene. Philos. Trans. R. Soc. Lond. B Biol. Sci. 376, 20200092 (2021).

57. J. J. Faber-Hammond, R. B. Phillips, K. H. Brown, Comparative analysis of the shared sex-determination region (SDR) among salmonid fishes. Genome Biol. Evol. 7, 1972-1987 (2015).

58. A. Kuzminov, Homologous recombination—Experimental systems, analysis, and significance. Ecol. Plus 4(2)(2011).

59. S. Yang et al., Repetitive element-mediated recombination as a mechanism for new gene originations in Drosophila. PLoS Genet. 4, e3 (2008).

60. M. M. Parks, C. E. Lawrence, B. J. Raphael, Detecting non-allelic homologous recombination from high-throughput sequencing data. Genome Biol. 16, 72 (2015).

61. M. R. Woodhouse, B. Pedersen, M. Freeling, Transposed genes in Aridobopsis are often associated with flanking repeats. PLoS Genet. 6, e1000949 (2010).

62. T. Wicker, J. P. Buchmann, B. Keller, Patching gaps in plant genomes results in gene movement and erosion of colinearity. Genome Res. 20, 1229-1237 (2010).

63. O. Blaser, C. Grossen, S. Neuenschwander, S. Neuenschwander, Sex-chromosome turnovers induced by deleterious mutation load. Evolution 67, 635-645 (2013).

64. B. Charlesworth, D. Charlesworth, The degeneration of Y chromosomes. Philos. Trans. R. Soc. Lond. B Biol. Sci. 355, 1563-1572 (2000).

65. A. Sharma et al., Male sex in houseflies is determined by Mdmd, a paralog of the generic splice factor gene CWC22. Science 356, 642-645 (2017).

66. J. A. Tennesen et al., Repeated translocation of a gene cassette drives sex-chromosome turnover in strawberries. PLoS Biol. 16, e2006062 (2018).

67. W. Yang et al., A general model to explain repeated turnovers of sex determination in the Salicaceae. Mol. Biol. Evol. 38, 968-980 (2021).

68. T. Matsumura et al., An efficient molecular technique for sexing tiger pufferfish (fugu) and the occurrence of sex reversal in a hatchery population. Fish. Sci. 80, 933-942 (2014).

69. S. Hosoya et al., The genetic architecture of growth rate in juvenile Takifugu species. Evolution 67, 590-598 (2013).

70. H. Enoki, “The construction of pseudomolecules of a commercial strawberry by DeNovoMAGIC and new genotyping technology, GRAS-Di” in Proceedings of the Plant and Animal Genome Conference XXVII (San Diego, CA, 2019), pp. 37002

71. 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).

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

73. H. Li. Minimap2: Pairwise alignment for nucleotide sequences. Bioinformatics 34, 3094-3100 (2018).

74. H. Li, Minimap and miniasm: Fast mapping and de novo assembly for noisy long sequences. Bioinformatics 32, 2103-2110 (2016).

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

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

77. G. Marçais et al., MuMmer4: A fast and scalable genome alignment system. PLoS Comput. Biol. 14, e1005944 (2018).

78. S. H. Martin, S. M. Van Belleghem, Exploring evolutionary relationships across the genome using topology weighting. Genetica 206, 429-438 (2017).

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

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

81. N. Patterson et al., Ancient admixture in human history. Genetics 192, 1065-1093 (2012).

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