# Characterization of the Sex Associated Region in the Sablefish Genome

by

Tasmin Carlow-Ujiie

A Thesis Submitted in Partial Fulfillment of the
Requirements for the Degree of

BACHELOR OF SCIENCE (HONS.)

in the Department of Biology


University of Victoria

Characterization of the Sex Associated Region in the Sablefish Genome

by

Tasmin Carlow-Ujiie

Supervisory Committee

Dr. John S. Taylor, Supervisor
Department of Biology

Dr. Barbara Hawkins, Honours Advisor
Department of Biology

Dr. Terri Lacourse, Honours Advisor
Department of Biology

# Abstract

The locus that determines sex in sablefish (*Anoplopoma fimbria*) and a putative master sex-determining gene, *gsdf*, were mapped to chromosome 14 in 2013. In 2021, the role of *gsdf* was confirmed and the allele responsible for sex was characterized to be a 936 transposable element insertion upstream of *gsdf* that upregulates expression in developing testis and determines maleness via a male heterogametic (XX/XY) system. According to classical theory of sex chromosome evolution, a sex-determining mutation creating a master sex-determining (MSD) gene is followed by recombination suppression spreading from that locus and eventual chromosome degradation and differentiation which ultimately produces heteromorphic sex chromosomes. In sablefish, it was not clear whether X and Y versions of chromosome 14 differed beyond the *gsdf* insertion. In this study, we used genome wide association studies (GWAS) and linkage disequilibrium (LD) analyses to define boundaries of the sex-associated region (SAR). We found that loci with genotype frequencies that differed between males and females were distributed across a large region of chromosome 14 (38,000 bp). Genes within this block include *LOCI129102360*, *nup54*, *LOCI129102793*, and *gsdf*. A weaker male-specific block of LD beyond the SAR included *aff1*. Alleles of *nup54*, *LOCI129102793,* and *aff1* on the *gsdfY*-bearing (Y) chromosome differed nonsynonymously from X-linked counterparts. These genes therefore appear to be 'founder loci' of an emerging sex-determining region. Moving forward, characterization of the SAR at this stage of sex chromosome evolution will explicate the rate and progression of the sex-determination system in sablefish.

# Table of Contents

## List of Tables



## List of Figures



## List of Abbreviations

| | |
|---|---|
| MSD | Master Sex-Determining |
| GWAS | Genome-Wide Association Study |
| LD | Linkage Disequilibrium |
| SAR | Sex-Associated Region |
| SDR | Sex-Determining Region |
| SNP | Single Nucleotide Polymorphism |

# Acknowledgements

I would like to thank Dr. John S. Taylor for his generous support and guidance on this project, without whom this work would not have been possible. I would also like to thank Zhaoze Liu for his mentorship with data handling and software usage.

I thank Golden Eagle Sablefish for samples, which included fish raised at the farm in Kyuquot Sound and wild-caught samples that made up the reference panel. This research was funded by an NSERC Alliance/MITACS/UVic Faculty of Science Grant, and a DFO/BCSRIF Grant to Drs. John Taylor and Briony Campbell. Raw sequence data were generated by the Center for Aquaculture Technology (CAT) and variant-calling was carried out by Zhaoze Liu using a sablefish reference genome that was the product of a collaboration between the Taylor and Koop labs at UVic.

Finally, I would like to thank my parents for their endless encouragement and advice.

# Introduction

## Classic Theory of Sex Chromosome Evolution

Mechanisms that drive sex determination vary widely across eukaryotes. This diversity contradicts the expectation for conservation of traits with strong influence over fitness. In the case of sex-determining systems, incomplete or incompatible sex determining mechanisms ensure rapid truncation of genetic lineages. A sex-determining system must produce reproductively viable offspring for the system to persist. Given this, the rate and diversity of sex-determining system turnover is unexpected. Methods other than genetic sex determination exist; for example, nonavian reptiles (*e.g.* tortoises and alligators) modulate the sex of broods depending on the incubation temperature of their nest (Ciofi & Swingland, 1997). Other environmental factors that modulate sex include photoperiod in crustaceans (Bulnheim, 1978), pheromones in angiosperms (Gregg, 1982), and water availability in mosses (Benassi *et al.*, 2011). Several species integrate multiple environmental stimuli to determine sex. Nevertheless, genetic sex determination is most common across eukaryotes and is the focus of this project.

Genetic sex-determining mechanisms are broadly classified as male heterogametic (XX/XY) or female heterogametic (ZZ/ZW) systems. Clades in which males are heterozygous and females are homozygous include mammals (Waters *et al.*, 2007), most bony fishes (Mank *et al.*, 2006), most true frogs (Jeffries *et al.*, 2018), and fruit flies (Aida, 1930). Clades where females are heterozygous for the sex-determining locus include birds (Brant, 1952), several species of bony fishes (Mank *et al.*, 2006), some true frogs (Jeffries *et al.*, 2018), and silkworm moths (Strunnikov, 1975).

Male and female heterogametic systems are further subdivided into homomorphic and heteromorphic sex chromosomes. Homomorphic sex chromosomes are karyotypically indistinct, whereas heteromorphic sex chromosomes are different sizes (*e.g.* X and Y chromosomes in humans). While distinction between homomorphic and heteromorphic chromosomes is not strict, decreased size observed in the heterogametic sex chromosome (Y or W) is maintained by recombination suppression. Across taxa, individuals with heteromorphic sex chromosomes tend to die earlier than those with homomorphic sexes due to imperfect gene dosage compensation of the heterogametic sex (Xirocostas *et al.*, 2020). In humans, lacking an X chromosome is incompatible with life because the Y chromosome cannot compensate for functionally essential genes on the X chromosome.

The locus directly responsible for sex determination is most often an allele of one gene, called the master sex-determining (MSD) gene. A rare exception to this is Tasmanian Atlantic salmon, which determine sex depending on genotypes of three separate loci (Eisbrenner *et al.*, 2014). More frequently, sex is determined by one MSD, such as a single nucleotide polymorphism (SNP) in the pufferfish gene *amhr2*, that alone determines an individual's sex (Kamiya *et al.*, 2012). In humans, the Y-linked *SRY* gene (Sex-Determining Region Y protein) determines whether an individual grows into a male or female (Sinclair *et al.*, 1990). That this gene alone determines sex is evidenced by rare cases of human XX males with a fragment of the Y chromosome including *SRY* translocated to one of the X chromosomes (Vergnaud *et al.*, 1986). Not every gene can mutate into a MSD; there are families of genes that tend to determine sex. For example, homologs of the anti-Müllerian hormone gene, *amh,* modulate sex in pufferfish (Kamiya *et al.*, 2012), Nile tilapia (Li *et al.*, 2015), and monotreme mammals (Shearwin-Whyatt *et al.*, 2025). Another common MSD is *dmrt1*, which determines sex in avian reptiles (Smith *et al.*, 2009), fruit flies (Erdman & Burtis, 1993), and the medaka species *Oryzias latipes* (Matsuda *et al.*, 2002). The ancestral homolog of *SRY* in mammals, *SOX3*, determines sex in another species of medaka, *Oryzias dancena,* (Takehana *et al.*, 2014) and the Japanese frog *Glandirana rugosa* (Miura *et al.*, 2024). Genes that encode transcription and growth factors generally lead genetic cascades that direct development of sexual organs.

Despite the reliance of sex determination on one gene, the sex-associated region (SAR) often spans beyond the MSD. Recombination suppression spreading from a MSD is characteristic but is not restricted to sex-determining regions. Inversions and centromeric regions (Koury, 2023) also display recombination suppression in regions unrelated to sex. In genetic sex determining systems, it is often the case that recombination suppression starts from the MSD and grows outward, covering a region beyond the MSD. Before alleles in the region are fixed, this is called the sex-associated region (defined above). Over generations, the SAR becomes fixed into a sex-determining region (SDR). In a SDR, the allele that determines sex still holds the only influence over sex. However, due to suppressed recombination, a marker located physically far away from the MSD on the sex chromosome predicts maleness just as accurately as the MSD itself. Without recombination, genes on the heterogametic sex chromosome are free to degrade and differentiate. Heteromorphic sex chromosomes, where the Y is smaller than the X, occur in humans, rainbow trout (Thorgaard, 1977), and gingko trees (Lee, 1954). There exist exceptions

to this; for example, recombination around the sex-determining SNP in pufferfish continues at the same rate as autosomes (Kamiya *et al.*, 2012). In contrast to pufferfish, the non-recombining region around human *SRY* spans almost the entire Y chromosome, leaving only small pseudoautosomal regions at each end that recombine as usual with the X chromosome (Freije *et al.*, 1992). Despite this differentiation, homology between sex chromosomes supports the theory that pairs of sex chromosomes originate from autosomes. While concepts surrounding the origin of sex chromosomes have been theoretically defined, it is difficult to practically distinguish MSD genes, SARs, and SDRs due to individual variation and incomplete co-occurrence with the sex-determining locus in a species. Here, I will refer to the region associated to sex as the SAR because it was surveyed on both X and Y chromosomes simultaneously in a representative sample, thus fixation of loci on the Y chromosome to the MSD throughout the species is unclear.

Importantly, the homogametic sex chromosome (*i.e.* the X chromosome in humans) continues to recombine over generations. Recombination suppression acts only on the heterogametic chromosome, which is usually mediated by epigenetic mechanisms such as DNA methylation (Metzger *et al.*, 2018). Because the heterogametic sex chromosome does not recombine, the homomorphic sex chromosome only "has the chance" to recombine half as often as autosomes (*e.g.* in humans, the X only recombines during meiosis in females but not during meiosis in males). Because of this, recombination suppression is often described on both sex chromosomes, which is true to some extent, but recombination is only suppressed directly on the SAR of one (heterogametic) chromosome. In the case of most mammals and sablefish, recombination suppression only occurs on the Y chromosome.

The favored hypothesis for recombination suppression on the heterogametic chromosome is sexual antagonism, which occurs when alleles that favour one sex but not the other gradually become trapped with the MSD as a result of selection (Charlesworth & Charlesworth, 2000). The hypothesis of sexual antagonism as an agent for recombination suppression is supported by models that imply ancestral selection for sexually antagonistic alleles based on enhanced divergence at distant positions from the sex-determining locus (Kirkpatrick & Guerrero, 2014). However, it has been difficult to formally capture the acquisition of sexually antagonistic alleles into the SAR. It is also possible that sexual antagonism only partially explains gene trapping in the SAR and there are other pressures that result in its acquisition of neutral alleles. Outside of sexual antagonism, inversions and other chromosomal rearrangements are sources of

recombination suppression and result in evolutionary strata along sex chromosomes (Lahn & Page, 1999). These chromosomal changes generally cause abrupt expansion of the SAR and encompass many genes. Alternatively, recombination suppression may also spread via accumulation of transposable elements (Srikulnath *et al.*, 2022). However, it is difficult to tell whether transposable elements have accumulated as a result of recombination suppression or recombination suppression has allowed for the accumulation of transposable elements. Recombination suppression remains a particularly interesting driver of sex chromosome evolution because recombination contributes so strongly to spreading beneficial alleles in a population whilst avoiding the cost of deleterious alleles.

The rate of SAR expansion and turnover (movement of the MSD to re-establish the sex-determining system) varies. Turnover occurs when a new MSD appears on the same sex chromosome or an autosome. Alternatively, a fragment from the original sex chromosome that includes the MSD can be translocated onto an autosome. One extreme example of this diversity is in East African cichlids that have developed over twelve SDR's within the last 45 million years (Gammerdinger & Kocher, 2018). Sex-determining systems in cichlids includes species with male heterogametic systems and female heterogametic systems. Each species has one MSD on one of ten linkage groups, each of which is associated with a small SAR and homomorphic sex chromosomes (Gammerdinger & Kocher, 2018). However, homomorphic sex chromosomes do not indicate the youth of a system. One genus of tree frog, *Hyla*, has 'ever-young' sex chromosomes that do not recombine but have not degraded (Stöck *et al.*, 2011). All three *Hyla* species use the same sex chromosomes and operate with a male heterogametic system (Stöck *et al.*, 2011). Therefore, while there is a trend from homomorphic to heteromorphic sex chromosomes, there are exceptions, and causes of variation in the rate of sex chromosome evolution are not well understood.

In some cases, the Y chromosome is degraded until it is completely lost. In mammals, species that function with a XX/X0 system include the Japanese spinous country rat (Honda *et al.*, 1977) and the Transcaucasian mole vole (Mulugeta *et al.*, 2016). If the Y chromosome is not lost, the MSD may also undergo turnover. This restores recombination on the heterogametic chromosome. By 'resetting' recombination suppression, rapid turnover of a MSD may allow a species to maintain homomorphic sex chromosomes. The evolution of sex chromosomes, including turnover, is summarized in Figure 1.

There is a remarkable diversity of sex-determining mechanisms in teleosts (Mank *et al.*, 2006). Unlike the conservation of a single sex determining system as observed in birds and mammals, teleosts operate with male heterogametic systems such as sablefish (Rondeau *et al.*, 2013), female heterogametic systems such as sole (Chen *et al.*, 2009), sequential hermaphroditism such as the bluehead wrasse (Warner & Swearer, 1991), and environmental sex determination such as sockeye salmon (Azuma *et al.*, 2004). To study the evolution of sex determining mechanisms, it is especially informative to focus on recently developed, homomorphic sex chromosome systems to uncover forces that promote recombination suppression and modulate the fate of sex determining systems. Therefore, teleosts are suitable models for studying sex chromosome evolution due to rapid turnover and appearance of sex-determining systems. While the age of male heterogametic system in sablefish is unknown, the homomorphic nature of these sex chromosomes suggest this system is in an early stage of SAR development. Here, we characterize the sex chromosomes of sablefish (*Anoplopoma fimbria*), a long-lived teleost with a nascent SAR.

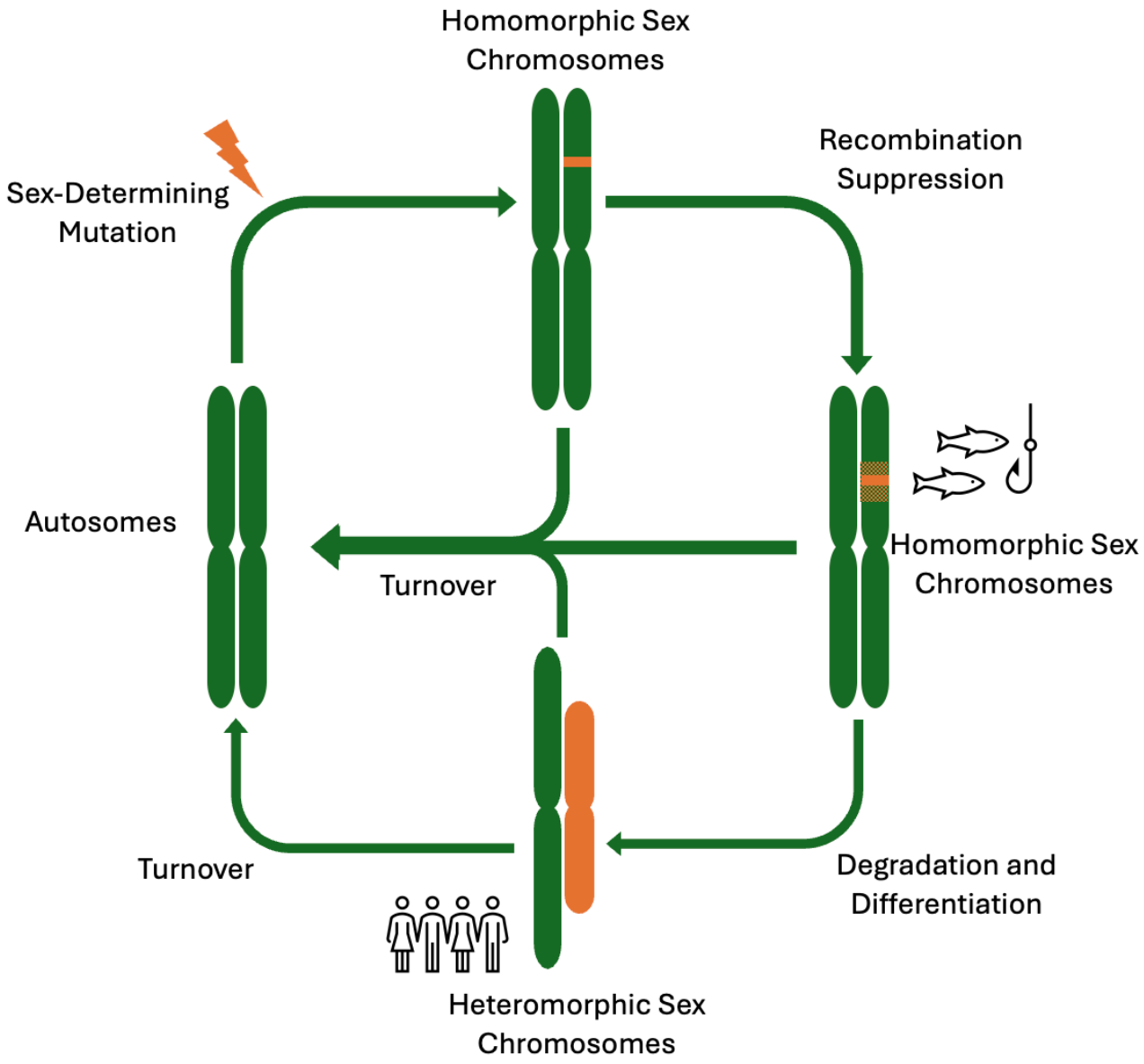


**Figure 1**. Classic theory of sex chromosome evolution. Sex chromosomes begin as autosomes (left) which recombine with minimal restrictions. Once a sex-determining mutation occurs in a MSD on an autosome (top), recombination suppression spreads outward from that locus (right). While recombination suppression spreads, genes within the sex-associated region on the heterogametic chromosome differentiate and degrade until they are karyotypically distinct (heteromorphic sex chromosomes, bottom). At any point, turnover of the MSD may occur if it is lost from the original chromosome via movement to a different autosome or replaced by a new MSD on the same or a different chromosome. Adapted from Furman *et al.* (2020).

**Sablefish as a Model for Recombination Suppression**

Sablefish is a species of teleost native to the Northern Pacific Ocean. Sex determination in this species functions in a male heterogametic system (XX/XY) due to a male-specific MSD on chromosome 14 (Hayman *et al.*, 2021). The MSD is *gonadal-soma derived factor* (*gsdf*) gene, which encodes *transforming growth factor, beta receptor 1a* (*tgfbr1a*), a protein involved in the transforming growth factor β system (Herpin *et al.,* 2021). This gene also determines sex in medaka species *Oryzias luzonensis* and Atlantic halibut *Hippoglossus hippoglossus* (Edvardsen *et al.*, 2022; Myosho *et al.*, 2012). In the medaka *O. luzonensis*, five SNP mutations and one 4-bp insertion upstream of *gsdf* cause maleness (Myosho *et al.*, 2012). In the Atlantic halibut, similar to sablefish, a transposable element (TE) insertion upstream of *gsdf* upregulates expression and initiates male development (Edvardsen *et al.*, 2022). Based on 20 sablefish sequences, the male (*gsdfY*) and female (*gsdfX*) alleles of *gsdf* differ by two nonsynonymous SNPs and four SNPs in the untranslated regions (Herpin *et al.*, 2021). Additionally, there is a 936 bp insertion 412 bp upstream of the start codon of *gsdfY* and a 412 bp insertion 1298 bp upstream of the *gsdfX* start codon (Herpin *et al.*, 2021). Neither exonic SNP between *gsdfY* and *gsdfX* seem to affect gene regulation or biochemical function of the protein (Herpin *et al.*, 2021). Instead, the upstream insertion of *gsdfY* enhances expression during fish development compared to *gsdfX*, which leads to male development (Herpin *et al.*, 2021).

Taxonomically, the closest relative of sablefish is the skilfish (*Erilepis zonifer*), which is the only other species in superfamily *Anoplopomatoidea*. Sex determination in *E. zonifer* has not been characterized. However, other related species such as three-spine stickleback, lumpfish, and lingcod have characterized SDRs. All three species operate under a male heterogametic system with *amhy* as the MSD in sticklebacks (Treaster *et al.*, 2025), *amh1* or *amh2* in lumpfish (Chaiyasut *et al.*, 2023), and *amh* in lingcod (Rondeau *et al.*, 2016).

A common result of recombination suppression is sexual dimorphism, which is when traits such as size, colour, or morphology differ between sexes due to sex-specific alleles in the SAR. Female sablefish are larger than males and weigh approximately 3 kg at age three years, whereas 3-year-old males weigh 2 kg (Goetz *et al.*, 2021). Genotypes associated with size are clustered around *gsdf* (Liu *et al.*, 2026). Size has a positive effect on reproductive success in females of other fish species (Morita & Takashima, 1998). Therefore, this might be an example of sexual antagonism in which size is beneficial to females but not males (Furman *et al.*, 2020),

or conversely that small size is beneficial to males but not females. Genes that influence growth rate might be linked to *gsdf* within the SAR, or size may be a downstream effect of *gsdf*. Sexual antagonism has historically been difficult to formally investigate due to difficulty in identifying alleles that benefit only one sex before they become fixed to the SAR. While *gsdf* has been identified as the MSD in sablefish, the size of the SAR is unknown. In this project, I examined the SAR for Y-specific alleles and genes in sablefish that have been trapped into LD with *gsdf* via recombination suppression. The goal of this project was to determine whether recombination is suppressed around the MSD and characterize genes in the nascent SAR.

## Methods

Genomes of 1,280 adult sablefish (approximately 2.5 years old) from Golden Eagle Sablefish were sequenced at 1x (paired end reads, Element AVITI). Imputation used a reference panel of 65 wild-caught fish sequenced at approximately 15x (Oxford Nanopore Sequencing, Illumina). Imputation to confirm or correct genotypes called from 1x data filled in missing data using haplotypes (SNPs inherited together) from our reference panel. When this project started, we had size data and genotypes for 745 sablefish. A size-determining region established by a GWAS was then used as a proxy for a SAR. Later in the project, size, sex, and genotype data for 602 fish became available. Sexes of the first group (n = 745) could then be determined using the data from the second dataset (n = 602).

### Genome-Wide Association Studies

Two genome-wide association studies (GWAS) were applied to size and sex data (credit: Zhaoze Liu). First, the size-related GWAS (n = 745) was used to estimate a SAR using a threshold of $p \leq 1 \times 10^{-20}$. Later, a GWAS with sex (n = 602) yielded two peaks in the Manhattan plot. The second SAR on chromosome 8 was investigated using a synteny analysis (Appendix A).

### Linkage Disequilibrium Analysis

LDBlockShow (version 2.2) was used to visualize recombination suppression in the chromosome 14 SAR using a 200,000 bp window that covered the size-associated region defined by the first GWAS (Dong *et al.*, 2021). Because calculations of LD are independent of phenotype and instead indicate the likelihood of SNPs being inherited together, this analysis sought to determine whether recombination suppression was present around the sex-determining locus, *gsdf*. The 200,000 bp region of chromosome 14 was surveyed for LD with both sexes (n = 1280), the reference panel (n = 65), males alone (n = 42), and females alone (n = 493). As mentioned above, all fish were eventually sexed, but at this stage only 602 fish had been sexed. Some individuals from this group were filtered out from the sex-specific LD plots due to lack of confidence during phenotype sexing. The boundary of the SAR was further defined based on additional LD analyses around *gsdf*. To test whether LD ($R^2$) in the putative SAR was unusual, two 35,388 bp fragments from each of nine chromosomes (chromosomes 4, 8, 10, 12, 15, 17, 18,

19, 20) were randomly selected and surveyed for LD in comparison to the SAR. Standardization of centromere distance and recombination hotspots (Choo, 1998) were achieved through sampling of the end and middle of these chromosomes.

**Characterization of Genes in the SAR**

SNPs that differed in the LD blocks between X and Y chromosomes were identified using SAMtools v.1.20 and Interactive Genome Viewer v2.19.7. Seven individuals (5672A, 8083, 8533, 2794A, 5645A, 9080, and 8823) from the reference panel were randomly selected. Reads from all seven individuals were mapped to a female reference genome. Using the SNP most significantly linked to sex from the GWAS with size, each sample was assigned male (heterozygous) or female (homozygous reference). Sexes were then confirmed from a previous study on the same individuals (Rubi *et al.*, 2022). Once sex was assigned, SNPs in the LD blocks were filtered based on whether they were heterozygous only in male individuals. Heterozygous SNPs that were present in any of the females were not considered. From these putative Y-specific SNPs, each was characterized based on gene, position in the gene, and type of mutation (nonsynonymous or synonymous).

## Results

### Genome-Wide Association Studies

We initially defined the SAR as a region that included *gsdf* and was associated with size (Fig. 2). This SAR was 35,388 bp long and spanned between positions 12,093,795 and 12,129,183 on chromosome 14. This peak on chromosome 14 includes *gsdf*, the MSD gene in sablefish according to the study by Herpin *et al.* (2021). After acquiring sex data from 602 fish, a GWAS with sex yielded a significant peak on chromosome 14 and to a lesser extent chromosome 8 as shown in Figure 3A. The region on chromosome 8 exceeding the threshold of -log(p) < 20 was tested for misassembly (Appendix A). After generating a new GWAS that took relatedness into account, the peak on chromosome 8 disappeared (Fig. 3B, credit: Zhaoze Liu). We attributed this observation to the inclusion of all-female broods generated from neo-male crosses in the sample. From these data, I established an approximate SAR of 35,388 bp on chromosome 14.

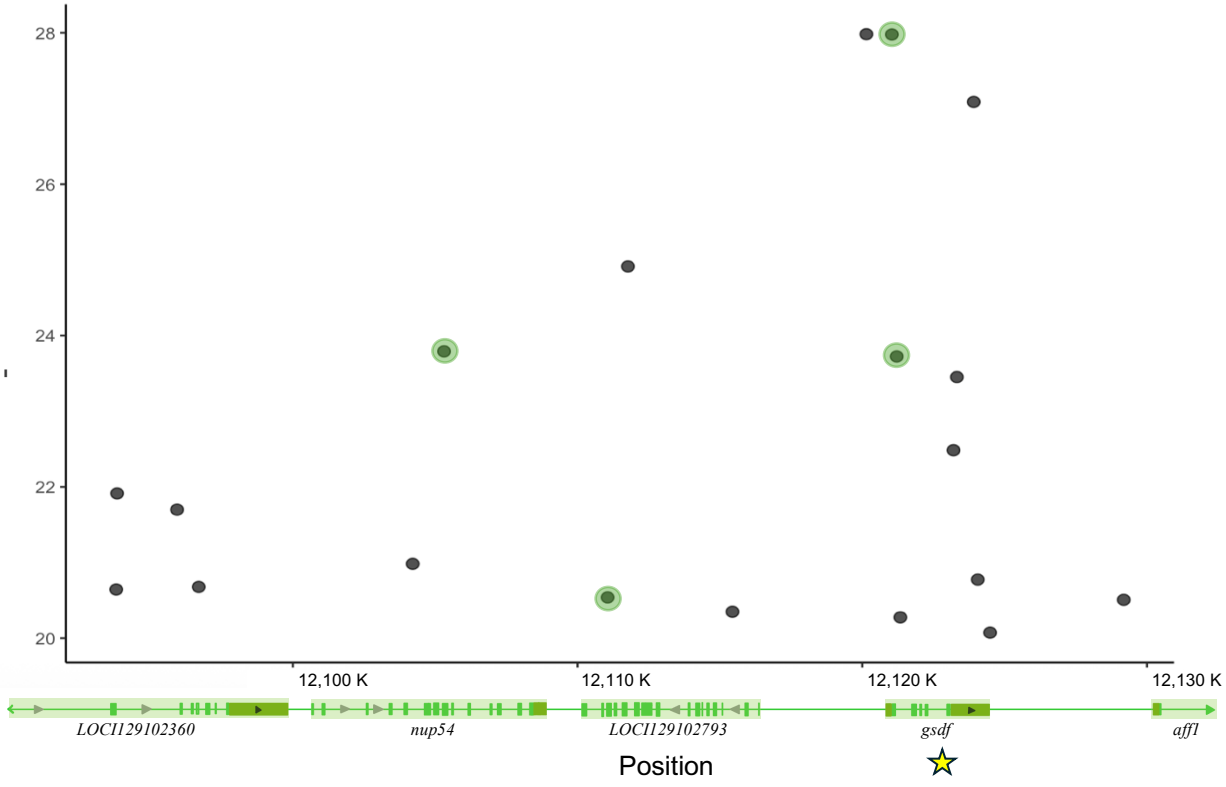


**Figure 2**. SNPs significantly associated with size located on chromosome 14 according to a threshold of $-\log_{10}(p)$ values greater than 20 ($p \leq 1x10^{-20}$) from a GWAS of 745 fish. Four exonic SNPs are highlighted in green; one exonic SNP in *nup54*, one in *LOCI129102793*, and two in *gsdf*. The MSD of sablefish, *gsdf*, is marked with a yellow star. These data were used to define an approximate SAR.

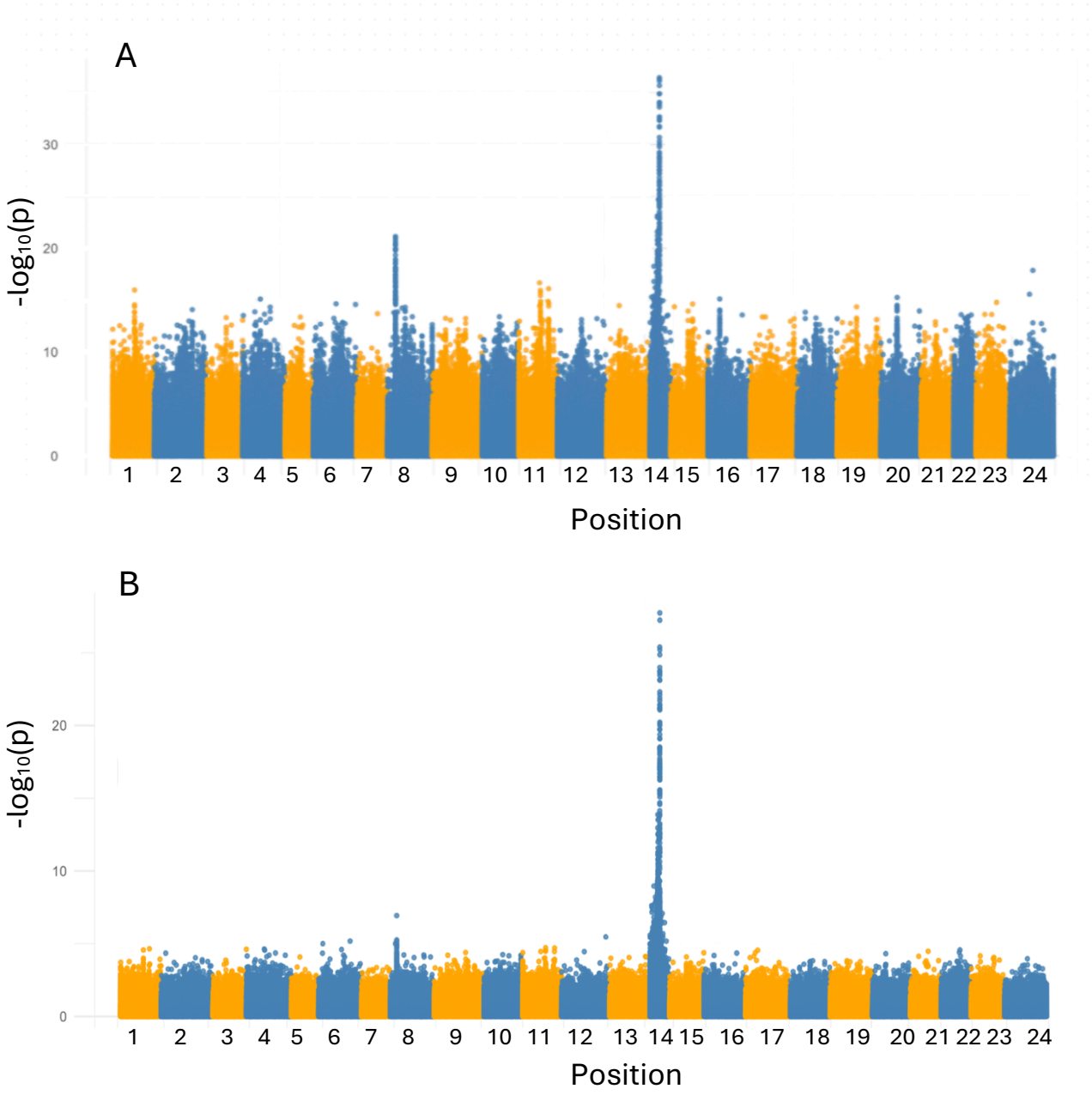


**Figure 3**. Genome-wide association study (GWAS) measuring significant association of 1 million SNPs to sex in 602 sablefish. (A) GWAS with sex. (B) GWAS with sex after relatedness was considered. (credit: Zhaoze Liu)

**Results from Linkage Disequilibrium Analysis**

Linkage disequilibrium (LD) values between SNPs were used as a proxy for recombination suppression in the region defined by the size-GWAS (n = 745). Using LDBlockShow, a 200,000 bp window around the SAR was surveyed for LD in the imputed data from both sexes, reference panel from both sexes, male-only imputed data, and female-only imputed data. A block of LD around *gsdf* spanned a region of approximately 38,000 bp, which I refer to here as the SAR. This region covered the same five genes from our size-GWAS. Imputed data from both sexes (n = 1,280) showed background LD around *gsdf* with a peak around *gsdf* of approximately 38,000 bp. The reference panel yielded less background LD compared to the imputed data but showed high LD between SNPs in and around *gsdf* (Fig. 4B). The male-only plot (Fig. 4C) displayed weak background LD similar to that observed in pooled sexes with enhanced LD around *gsdf*. Furthermore, a second male-specific block of LD beside the defined SAR was evidenced by the comparison between male and female plots. This second block is approximately 40,000 bp long and includes *aff1* and *LOCI129102365*, which is shown more clearly in Figure 5. The female-only plot (Fig. 4D) showed background LD comparable to that of both sexes (Fig. 4A) and lacked a peak near *gsdf*. The difference in size of SARs defined by GWAS (35,388 bp) and LD (38,000 bp) analyses can be explained by the arbitrary selection of a significance threshold ($p \leq 1 \times 10^{-20}$) in the GWAS and is further discussed in Appendix B. From this analysis, a SAR was determined based on evidence of recombination suppression around *gsdf*.

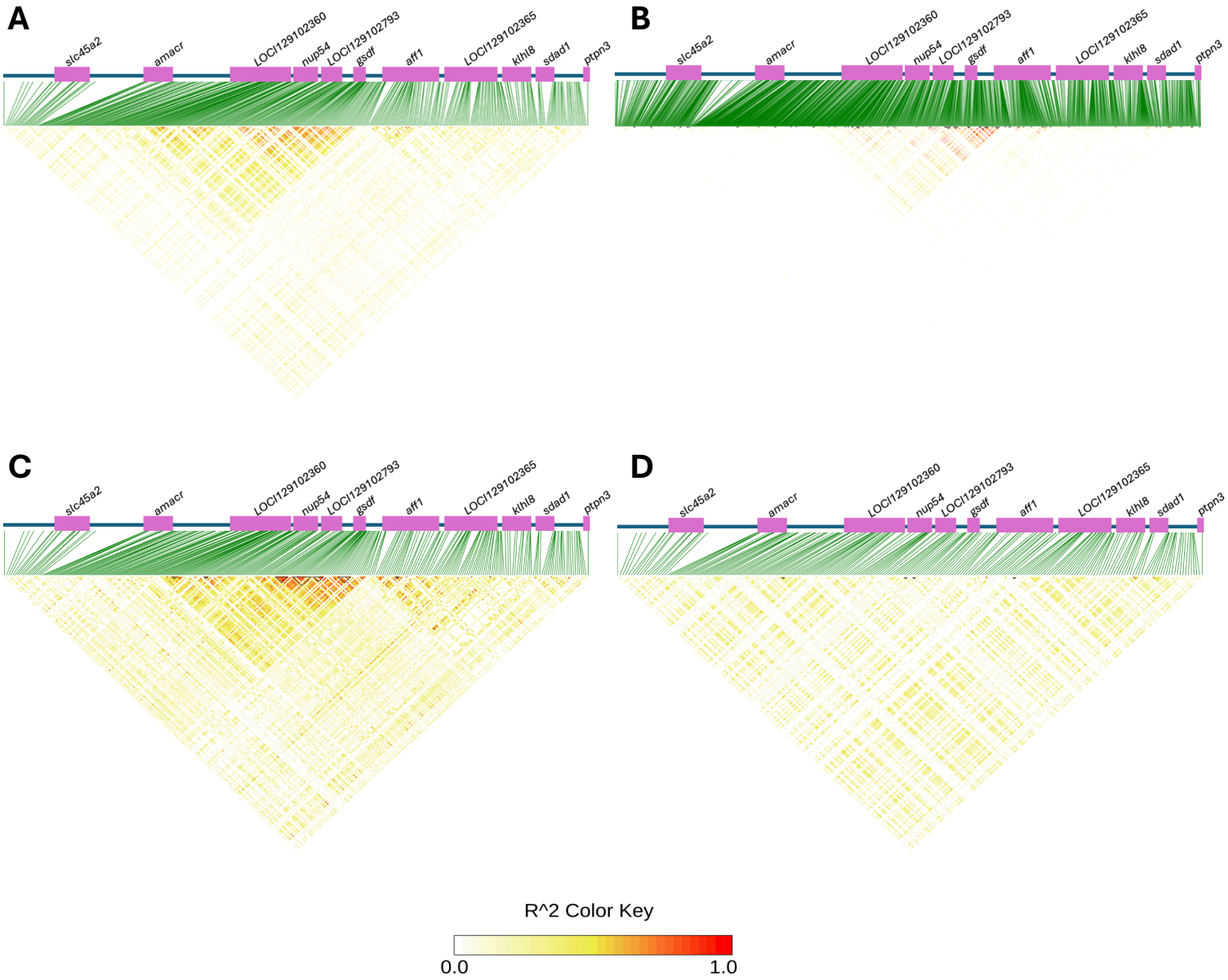


**Figure 4.** LD heatmaps of a 200,000 bp fragment of chromosome 14 framing the SAR (n = 1,280). Genes are labelled at the top of each heatmap, and gene positions are indicated in pink. Each pairwise calculation between SNPs indicates the tendency of two genotypes to have correlated genotypes; red boxes (high $R^2$) indicate genotypes are inherited together frequently, whereas white boxes (low $R^2$) indicate genotypes are unlikely to be inherited together. (A) Both sexes (n = 1,280). (B) Both sexes, reference panel (n = 65). (C) Male-only (n = 42). (D) Female-only (n = 493).

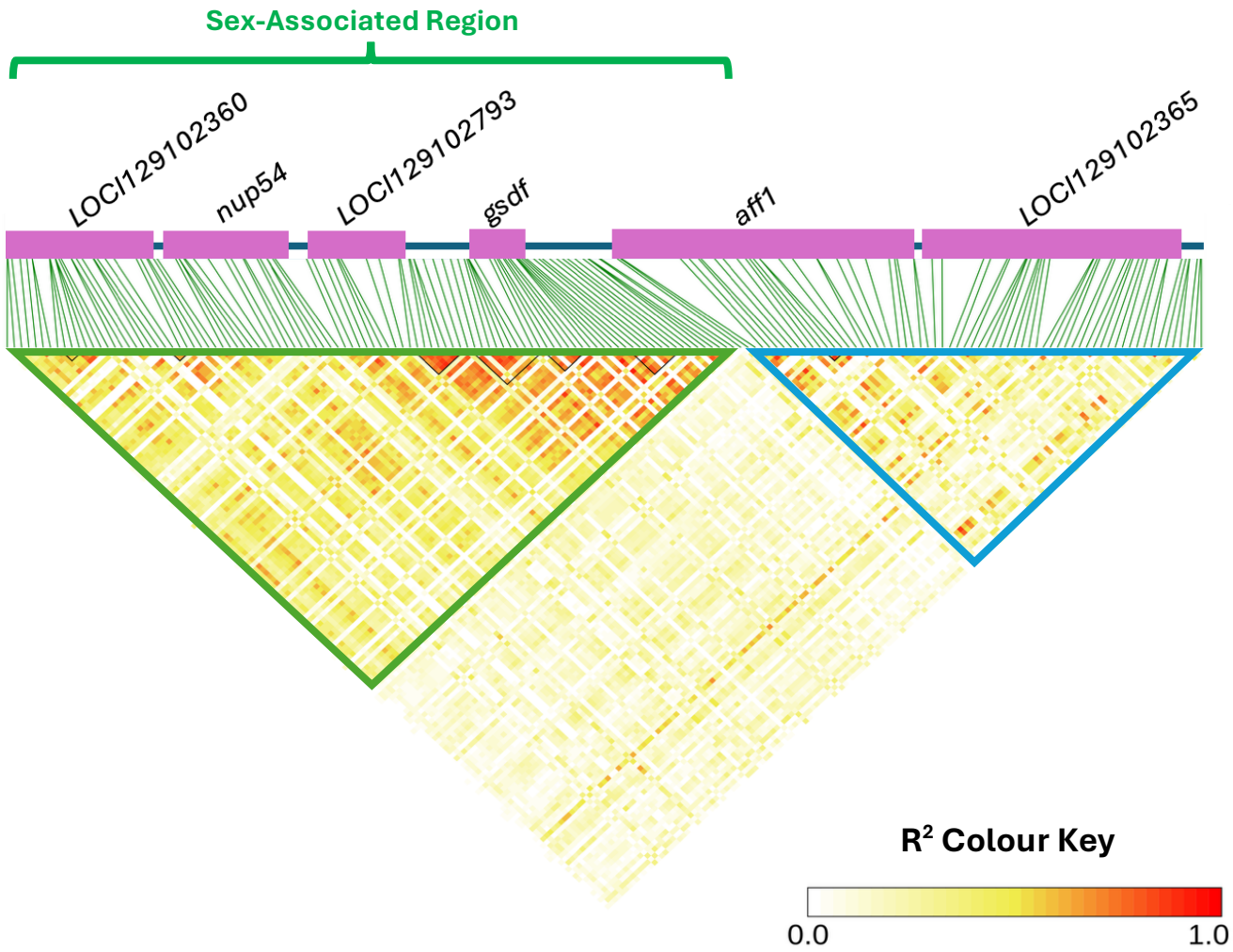


**Figure 5.** LD heatmap of a 78,000 bp region between positions 12,090,700 and 12,169,700 on chromosome 14. LD was estimated using data from 42 male sablefish. Pairwise calculations are based on $R^2$ values of SNP genotypes. Red squares (high $R^2$ values) indicate strong linkage between SNPs, whereas white squares (low $R^2$ values) indicate weak linkage between SNPs. The SAR is outlined in a green triangle (left), while the second male-specific LD block is shown in blue (right). Small segments of the chromosome in which average $R^2 > 0.85$ are outlined in black. LD from both blocks is absent in the female-only plot.

Using the SAR defined by the GWAS with size, LD was compared to the rest of the genome to determine whether the linkage observed was unusual. From the reference panel, I noted that LD in the SAR was significantly greater ($p < 2.0 \times 10^{-16}$) than LD from any other surveyed fragment of the genome (Fig. 6). In each region there were comparisons between SNPs that yielded high $R^2$ values, but the average LD in the SAR was significantly higher ($0.32 \pm 0.01$) than average $R^2$ values from any other region. These data provide evidence for recombination suppression around *gsdf*.

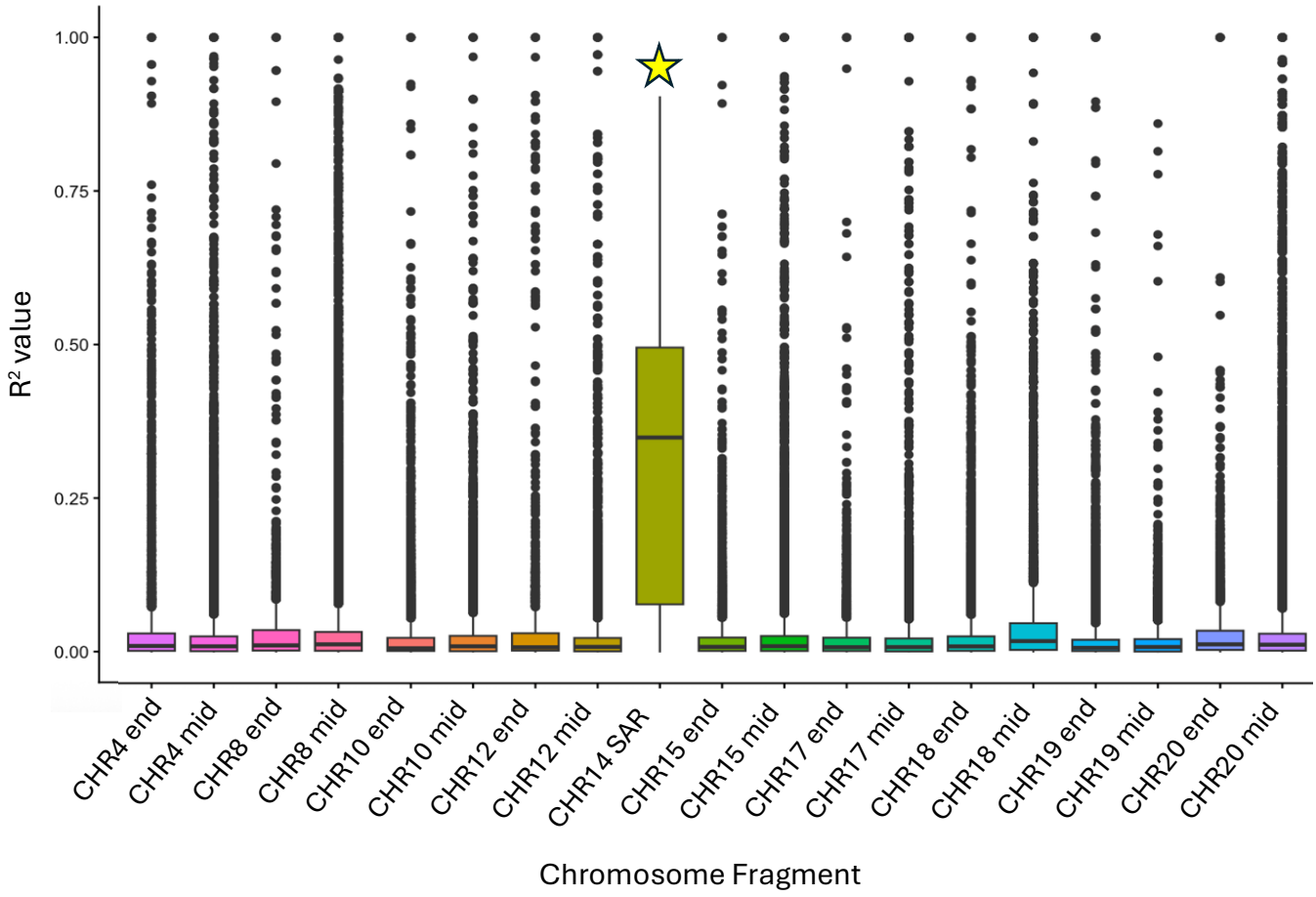


**Figure 6.** Distribution of $R^2$ values representing linkage in eighteen 35,388-bp fragments from nine chromosomes compared to the SAR (CHR14 SAR, marked by the yellow star). LD was estimated from 15x reference panel genotypes (n = 65). Data from 1x individuals were imputed (Methods), which has the potential to inflate LD. Chromosomes 4, 8, 10, 12, 15, 17, 18, 19, and 20 were selected randomly and are listed on the x-axis. Each $R^2$ value represents the likelihood of two SNPs to be inherited together. Fragments from the end or middle of each chromosome are included. Each box covers 50% of the $R^2$ values in that region. Horizontal bars indicate the median of each region and outliers are indicated by points.

### Characterization of Genes in the SAR

Once a SAR was defined using LD, I sought to characterize Y-specific alleles trapped by recombination suppression. The SAR of 38,000 bp spans four genes; *LOCI129102360*, *nup54*,

*LOCI129102793*, and *gsdf*. These genes function in protein tagging, nucleoporin structure, and stimulus perception (Table 1) (NCBI Genome, 2025). I looked for Y-specific SNPs in these four genes and two others (*aff1* and *LOCI129102365*) from the second, weaker LD block identified from the male-specific analysis.

**Table 1**. Four genes from the 38,000 bp SAR defined by LD on chromosome 14 of sablefish and their probable functions ascribed using NCBI Gene.

| Gene Name | Protein Description | Probable Function |
|---|---|---|
| *LOCI129102360* | Lysosome membrane protein 2-like | Protein targeting to lysosome, receptor-mediated endocytosis |
| *nup54* | Nucleoporin 54 | Nuclear pore organization, protein import into nucleus |
| *LOCI129102793* | Serine/threonine-protein phosphatase with EF hands 2-like | Detection of stimulus, triggers sensory cascade, MAPK cascade |
| *gsdf* | Gonadal somatic derived factor | Growth factor and cytokine activity |

To identify Y-specific alleles, seven individual sablefish were isolated from the reference panel and sexed based on *SNP12121039* in *gsdf* (Herpin *et al.,* 2021). Based on *SNP12121039*, samples 5672A, 8083, 8533 were classified as males and 2794A, 5645A, 9080, and 8823 were marked as females. These reference panel individuals happened to have been used in a parental analysis study (Rubi *et al.*, 2022), so I was able to confirm these genotype-based classifications. The study used to confirm these classifications did not included sample 8533 but affirmed all other classifications.

Using these seven individuals, I identified Y-specific alleles in genes from the SAR and adjacent male-specific block of LD (six genes in total). In the SAR, a total of 200 SNPs were recorded. From these SNPs, I identified 40 Y-specific SNPs, including 35 intronic, four exonic, and one nonsynonymous mutation. In *gsdf* alone*,* 16 SNPs were identified. Two were exonic and

one was nonsynonymous for a phenylalanine to leucine mutation. Both exonic SNPs in *gsdf* identified by this analysis correspond to those found in the study by Herpin *et al.* (2021). Then, genes in the second block of LD (Fig. 5, outlined in blue) were characterized. One nonsynonymous Y-linked SNP was identified in *aff1* (AF4, FMR2 family, member 1) at position 12,143,189 and causes a serine to leucine change in exon 11. Using the sexes of individuals from the reference panel, I identified positions of Y-linked SNPs in and around the SAR.

## Discussion

Rondeau *et al.* (2013) mapped the MSD of sablefish, *gsdf*, to a region of chromosome 14. In a 300-locus genotype panel, Liu *et al.*, (2026) showed that SNPs in the region were significantly associated with growth rate. At this point, there was little information on the surrounding region and development of a SAR. These analyses were expanded here by the analyses of SNP data generated from whole-genome sequencing. Sequencing, sex, and size data were used to define a SAR around *gsdf* and characterize Y-specific alleles that have been taken up by the SAR via recombination suppression.

The analyses of approximately 1 million SNPs allowed us to define a chromosome 14 size-associated region that was used as a proxy for a SAR. The size-associated region spanned 35,388 bp and contained four genes (*LOC129102360*, *nup54*, *LOC129102793*, and *gsdf*). Other than *gsdf*, none of these genes are known to act as a MSD in other species. *Gsdf* acts as a MSD in other fish species including medaka (*Oryzias luzonensis*) (Myosho *et al.*, 2012; Zhang *et al.*, 2016) and Atlantic halibut. In medaka, a 4-bp insertion in the male copy of *gsdf* cause a frameshift and result in a non-functional protein product (Zhang *et al.*, 2016). Interestingly, the loss of *gsdfY* triggers femaleness in the medaka *O. luzonensis*, whereas the overexpression of *gsdfY* in sablefish causes maleness (Herpin *et al.*, 2021; Zhang *et al.*, 2016). In Atlantic halibut, a TE insertion upstream of *gsdf*, similar to sablefish, upregulates expression in males during development (Edvardsen *et al.*, 2022).

LD analyses of the SAR carried out for all samples and then separately for males and females provided stronger evidence for recombination suppression on the Y chromosome. As shown in Figure 4, when sexes were evaluated separately, LD was enhanced for the male-only cohort and decreased in the female-only cohort. This aligns with the hypothesis that direct recombination suppression is restricted to the Y chromosome and has a smaller effect on the X chromosome. Furthermore, separating males and females revealed male-specific LD beyond the 38,000 bp region. A second block of LD adjacent to the defined SAR of approximately 40,000 has stronger LD in the male-only plot (Fig. 4C) than the female-only plot (Fig. 4D) but was not included in the SAR due to weak linkage to the first block of LD. This fragment that includes *aff1* and *LOC129102365* may indicate a fragment of chromosome 14 that is currently integrating into the SAR or recombination being restored in a larger LD block.

The imputed dataset and reference panel yielded 84 and 200 SNPs in the SAR, respectively, due to higher read depth and genetic diversity of wild-caught samples in the reference panel. While the GWAS can be used to tentatively define a SAR, a GWAS alone should not inform the size of a SAR. Significant association to sex might be an artifact of the initial sex-determining mutation, which will immediately have linkage to all alleles on the new sex chromosome. It is therefore important to use LD in conjunction with GWAS to clarify biological processes around the MSD. This is further discussed in Appendix B.

To identify Y-specific alleles, five genes around *gsdf* were searched for Y-specific SNPs in seven individuals with high-quality genomes. The genotype at *SNP12121039* was used to sex individuals. While nonsynonymous, this SNP is not directly responsible for sex determination. The insertion most strongly associated with sex determination is 936 bp upstream of *gsdf*. Therefore, while there was significant association in the GWAS between the SNP and size, recombination suppression in the area, and association of the SNP with *gsdfY* (Herpin *et al.*, 2021), it may not be an absolute indicator for sex. Confirmation of sexes for all but one individual improved confidence in sexing (Rubi *et al.*, 2022). Another assumption was that Y-specific SNPs were indicated by heterozygosity in males were absent (homozygous reference) in females. The SNPs identified using this method, particularly the nonsynonymous SNP in *aff1*, are not fixed to *gsdfY* or would display a stronger association on both GWAS. This being said, the nature of a nascent SAR is that alleles in the proximity of the MSD are not immediately fixed to each other but become fixed as recombination suppression progresses. From these data, we can monitor the SAR for suppressed recombination via fixation of Y-specific SNPs to *gsdfY* and appearance of new Y-specific SNPs at more distant loci.

**Association to Sex via Neo-Males**

From our LD analysis, a 38,000 bp SAR was identified on chromosome 14 and a smaller peak was identified on chromosome 8. The second SAR was explored as a genome assembly error or a second sex-determining locus but was attributed to the use of neo-males in production of some samples. Neo-males are genetically female (XX) fish that are treated using methyltestosterone hormones early in development and thus develop male reproductive organs (Luckenbach *et al.*, 2016). When neo-males are crossed with wild type females (XX), they produce all female offspring (Fig. 7). Our samples did not include neo-male fish. The

commercial basis of preferential female breeding is the larger size of females. Crosses of this method produced many of the samples at Golden Eagle Sablefish and likely explain the observed linkage to sex on chromosome 8. If an allele was present in either parent of a cross between a neo-male and female, only females would ever inherit that allele. This creates artificial association with sex in a GWAS because the allele only appears in females, but is not sex-determining and has no bearing on femaleness. The synteny analysis (Appendix A) showed that an assembly error was unlikely and characterization of genes (*i.e. mylk4b*) in the chromosome 8 SAR revealed no history in sex determination. The lack of association with sex determination in current literature supported the conclusion the chromosome 8 SAR was an artifact of neo-male breeding (Fig. 7).

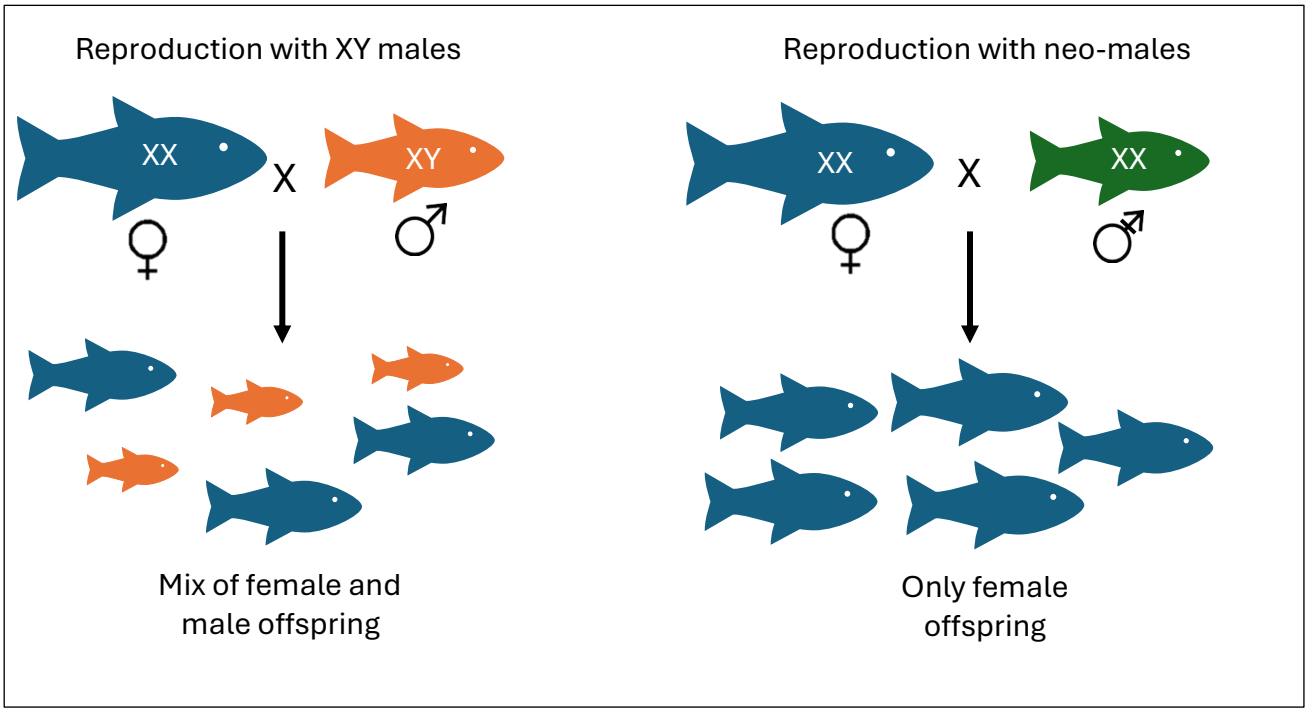


**Figure 7**. Sex association due to breeding of neo-males. A cross between a XX wildtype female and a XY male (left) results in offspring that are 50% males and 50% females. A cross between a XX wildtype female and hormone-treated neo-male (XX) with male reproductive organs results in only female offspring (right). If a unique allele appears in either parent in the latter case, that allele will be associated with sex (females) in the progeny.

**Characterization of Genes in the SAR**

Although none of the three genes in the SAR defined by LD have known roles as MSDs in other species, *nup54* is notably expressed in developing ovaries of zebrafish and medaka (Gautier *et al.*, 2011). Furthermore, genes around *gsdf* are conserved phylogenetically in fishes and are considered an example of functional clustering in the genome (Gautier *et al.*, 2011). Of genes in the cluster, *gsdf* is the only gene that is upregulated in testis. Other genes in the cluster, including *nup54,* are highly expressed in previtellogenic oocytes and associated with ovary development (Gautier *et al.*, 2011). This being said, some genes identified in our survey such as *LOCI129102360* and *LOCI129102793* are not mentioned by Gautier *et al.* (2011) as genes involved in the development of sexual organs (Gautier *et al.*, 2011). Genes just outside the SAR, *aff1* and *LOCI129102365*, may be integrating into the SAR, evidenced by male-specific LD and a Y-specific nonsynonymous SNP found in *aff1*. *Aff1* was mentioned in the study by Gautier *et al.* (2011), along with *nup54*, as a gene directing ovary development in zebrafish. Furthermore, the functional cluster of sex-related genes in zebrafish and medaka includes others such as *klhl8*, *sdad1*, and *ptpn3* (Gautier *et al.*, 2011), which are just beyond male-specific linkage observed here. Interspecific genes involved in sexual development are functionally similar and are likely integrated in signalling cascades associated with sexual development. The fragment of chromosome 14 that includes *aff1*, *LOCI129102365*, *klhl8*, *sdad1*, and *ptpn3* should be monitored for stronger association to the SAR.

The possibility that sexual antagonism has promoted recombination suppression cannot be excluded and is worthy of further inquiry. That is, it is possible that alleles in the SAR (*LOCI129102360*, *nup54*, *LOCI129102793* and *gsdf*) serve sex-specific functions that benefit one sex but not the other. While there were few nonsynonymous SNPs identified in the SAR, intronic SNPs (35 of which were identified) may affect transcriptional regulation and favor one sex, thus allowing sexual antagonism to suppress recombination. However, the converse is also possible that recombination suppression is occurring independently of sexual antagonism and another selective pressure is driving expansion of the SAR, such as accumulation of transposable elements. If the LD observed here reflects a growing SAR, the hotspot of recombination between the two LD blocks implies that recombination suppression may spread in a discontinuous pattern rather than a gradual, stepwise manner.

The nonsynonymous, Y-specific SNP identified on *aff1* might be an example of gradual incorporation of distant alleles into the SAR via recombination suppression. The family of proteins that includes *aff1* (AF4/FMR2) is known to regulate RNA transcript elongation through a RNA super elongation complex. Interestingly, deficiencies of these genes in humans are deleterious and result in disease including intellectual disability (Voisin *et al.*, 2021) and acute leukemia (Tamai *et al.*, 2017). Furthermore, knockouts of other proteins expressed in the RNA super elongation complex reduce stress tolerance in zebrafish (Liu *et al.*, 2023). While the function of *aff1* in sablefish may be different from that observed in humans or zebrafish, it is possible that an amino acid change is deleterious. Biochemically, it would be informative to determine what effect, if any, the leucine/serine change has on the protein function and survey homologous genes to determine whether this is a highly conserved position.

Beyond functional analyses of these Y-specific alleles, it would also be interesting to repeat a sex-related GWAS now that sexes of all 1,280 fish are known. Using this GWAS, we would be able to assign a threshold for significance and compare the sex-linked region and the size-linked region to determine whether they encompass different genes. Furthermore, it is now possible to repeat LD analyses using larger sample groups to affirm the SAR boundaries.

Characterization of SARs in non-model species contributes to our understanding of sex chromosome evolution. In this project, a 38,000 bp SAR in sablefish was defined and characterized using GWAS and LD analyses. The SAR spans four genes, all of which were surveyed for allelic association with the Y-chromosome. A second block of male-specific LD was identified that covers two genes, which is separated from the SAR by a recombination hotspot. Y-specific SNPs from all six genes were characterized based on their position. Notably, one Y-linked nonsynonymous SNP in *aff1* was identified. This, and intronic SNPs, should be further explored as sexually antagonistic alleles. Furthermore, genes associated with sexual development such as *nup54*, *aff1*, *klhl8*, *sdad1*, and *ptpn13* should be monitored for integration into the SAR. If expansion of the SAR should continue with *gsdf* as the MSD, recombination suppression may eventually cover most of the Y chromosome and allow degradation of genes on the Y chromosome. Using the sex-linked alleles and SAR characterized in this study, the rate of expansion and fixation of the SAR around *gsdf* can be quantified. Furthermore, it will be interesting to explore the selective basis for recombination suppression in sablefish, along with the ecological and epigenetic factors that influence the evolution of sex chromosomes.

## References


Aida, T. 1930. Further genetical studies of *Aplocheilus latipes*. *Genetics*, *15*(1), 1–16. https://doi.org/10.1093/genetics/15.1.1

Azuma, T., K. Takeda, T. Doi, K. Muto, M. Akutsu, M. Sawada & S. Adachi. 2004. The influence of temperature on sex determination in sockeye salmon *Oncorhynchus nerka*. *Aquaculture,* 234: 461–473.

Benassi, M., Stark, L. R., Brinda, J. C., McLetchie, D. N., Bonine, M., & Mishler, B. D. 2011. Plant size, sex expression and sexual reproduction along an elevation gradient in a desert moss. *The Bryologist*, *114*(2), 277–288. https://doi.org/10.1639/0007-2745-114.2.277

Brant, J. W. A. 1952. A review of literature on chromosome studies of the fowl. *Poultry Science,* 31: 409–417.

Bulnheim, H. P. 1978. Interaction between genetic, external and parasitic factors in sex determination of the crustacean amphipod *Gammarus duebeni*. *Helgoländer Wissenschaftliche Meeresuntersuchungen*, *31*(1–2), 1–33. https://doi.org/10.1007/BF02296989

Chaiyasut, K., P. Merviel, P. Palma, J. Perschthaler, E. Jimenez-Fernandez, A. Davie & A. P. Gutierrez. 2023. Development of an accurate polymerase chain reaction (PCR) assay for genetic sex identification in lumpfish (*Cyclopterus lumpus*) based on male-specific anti-Müllerian hormone (*amh*) gene. *Fishes,* 8: 327.

Ciofi, C., & Swingland, I. R. 1997. Environmental sex determination in reptiles. *Applied Animal Behaviour Science*, *51*(3–4), 251–265. https://doi.org/10.1016/S0168-1591(96)01108-2

Charlesworth, B. 1991. The evolution of sex chromosomes. *Science* 251: 1030-1033.

Charlesworth, B. & D. Charlesworth. 2000. The degeneration of Ychromosomes. *Philosophical Transactions of the Royal Society B,* 355: 1563–1572.

Charlesworth, D., Charlesworth, B., & Marais, G. 2005. Steps in the evolution of heteromorphic sex chromosomes. *Heredity*, *95*(2), 118–128. https://doi.org/10.1038/sj.hdy.6800697

Chen, S.-L., Y.-S. Tian, J.-F. Yang, C.-W. Shao, X.-S. Ji, J.-M. Zhai, X.-L. Liao, Z.-M. Zhuang, P.-Z. Su, J.-Y. Xu, *et al.* 2009. Artificial gynogenesis and sex determination in half-smooth tongue sole (*Cynoglossus semilaevis*). *Marine Biotechnology,* 11: 243–251.

Choo, K. H. A. 1998. Why is the centromere so cold? *Genome Research*, *8*(2), 81–82. https://doi.org/10.1101/gr.8.2.81

Dong, S. S., He, W. M., Ji, J. J., Zhang, C., Guo, Y., & Yang, T. L. 2021. LDBlockShow: A fast and convenient tool for visualizing linkage disequilibrium and haplotype blocks based on variant call format files. *Briefings in Bioinformatics*, *22*(4), bbaa227. https://doi.org/10.1093/bib/bbaa227

Edvardsen, R. B., O. Wallerman, T. Furmanek, L. Kleppe, P. Jern, A. Wallberg, E. Kjærner-Semb, S. Mæhle, S. K. Olausson, E. Sundström, T. Harboe, R. Mangor-Jensen, M. Møgster, P. Perrichon, B. Norberg, & C. J. Rubin. 2022. Heterochiasmy and the establishment of *gsdf* as a novel sex determining gene in Atlantic halibut. *PLoS Genetics,* 18: e1010011.

Eisbrenner, W. D., N. Botwright, M. Cook, E. A. Davidson, S. Dominik, N. G. Elliott,

J. Henshall, S. L. Jones, P. D. Kube, K. P. Lubieniecki, S. Peng, & W. S. Davidson. 2014. Evidence for multiple sex-determining loci in Tasmanian Atlantic salmon (*Salmo salar*). *Heredity*, *113*(1), 86–92. https://doi.org/10.1038/hdy.2013.55

Erdman, S. E., & Burtis, K. C. 1993. The *Drosophila* doublesex proteins share a novel zinc finger related DNA binding domain. *The EMBO Journal*, *12*(2), 527–535. https://doi.org/10.1002/j.1460-2075.1993.tb05684.x

Freije, D., C. Helms, M. S. Watson & H. Donis-Keller. 1992. Identification of a second pseudoautosomal region near the XQ and YQ telomeres. *Science,* 258: 1784–1787.

Furman, B.L.E., D. C. H. Metzger, I. Darolti, A. E. Wright, B. A. Sandkam, P. Almeida, J. J. Shu, & J. E. Mank. 2020. Sex chromosome evolution: so many exceptions to the rules. *Genome Biology and Evolution,* 12: 750-763.

Gammerdinger, W. J., & T. D. Kocher. 2018. Unusual diversity of sex chromosomes in African cichlid fishes. *Genes*, *9*(10), 480. https://doi.org/10.3390/genes9100480

Gautier, A., F. Le Gac, & J. J. Lareyre. 2011. The *gsdf* gene locus harbors evolutionary conserved and clustered genes preferentially expressed in fish previtellogenic oocytes. *Gene*, *472*(1–2), 7–17. https://doi.org/10.1016/j.gene.2010.10.014

Genome [Internet]. Bethesda (MD): National Library of Medicine (US), National Center for Biotechnology Information; 2004 – [cited 2025 12 10]. Available from: https://www.ncbi.nlm.nih.gov/genome/

Gene [Internet]. Bethesda (MD): National Library of Medicine (US), National Center for Biotechnology Information; 2004 – [cited 2026 03 30]. Available from: https://www.ncbi.nlm.nih.gov/genome/

Goetz, F. W., B. F. Anulacion, M. R. Arkoosh, M. A. Cook, W. W. Dickhoff, J. P. Dietrich, W. T. Fairgrieve, E. S. Hayman, M. B. R. Hicks, C. Jensen, R. B. Johnson, J. S. F. Lee, J. A. Luckenbach, K. C. Massee, & T. H. Wade. 2021. Status of sablefish, *Anoplopoma fimbria*, aquaculture. *Journal of the World Aquaculture Society,* 52: 607-646.

Gregg, K. B. (1982). Sunlight-enhanced ethylene evolution by developing inflorescences of Catasetum and Cycnoches and its relation to female flower production. *Botanical Gazette*, *143*(4), 466–475. https://doi.org/10.1086/337322

Hayman, E. S., W. T. Fairgrieve & J. A. Luckenbach. 2021. Molecular and morphological sex differentiation in sablefish (*Anoplopoma fimbria*), a marine teleost with XX/XY sex determination. *Gene,* 764: 145093.

Herpin, A., M. Schartl, A. Depincé, Y. Guiguen, J. Bobe, A. Hua-Van, E. S. Hayman, A. Octavera, G. Yoshizaki, K. M. Nichols, G. W. Goetz, & J. A. Luckenbach. 2021. Allelic diversification after transposable element exaptation promoted *gsdf* as the master sex of sablefish. *Genome Research,* 31: 1366-1380.

Honda, T., Suzuki, H., & Itoh, M. 1977. An unusual sex chromosome constitution found in the Amami spinous country-rat, *Tokudaia osimensis osimensis*. *The Japanese Journal of Genetics*, *52*(3), 247–249. https://doi.org/10.1266/jjg.52.247

Jeffries, D. L., G. Lavanchy, R. Sermier, M. J. Sredl, I. Miura, A. Borzée, L. N. Barrow, D. Canestrelli, P. A. Crochet, C. Dufresnes, J. Fu, W. J. Ma, C. M. Garcia, K. Ghali, G. A. Nicieza, R. P. O'Donnell, N. Rodrigues, A. Romano, Í. Martínez-Solano, I. Stepanyan, S. Zumbach, A. Brelsford, & N. Perrin. 2018. A rapid rate of sex-chromosome turnover and non-random transitions in true frogs. *Nature Communications*, *9*(1), 4088. https://doi.org/10.1038/s41467-018-06517-2

Kamiya, T., Kai, W., Tasumi, S., Oka, A., Matsunaga, T., Mizuno, N., Fujita, M., Suetake, H.,

Suzuki, S., Hosoya, S., Tohari, S., Brenner, S., Miyadai, T., Venkatesh, B., Suzuki, Y., & Kikuchi, K. 2012. A trans-species missense SNP in *amhr2* is associated with sex determination in the tiger pufferfish, *Takifugu rubripes* (Fugu). *PLoS Genetics*, *8*(7), e1002798. https://doi.org/10.1371/journal.pgen.1002798

Kirkpatrick, M., & Guerrero, R. F. 2014. Signatures of sex-antagonistic selection on recombining sex chromosomes. *Genetics*, *197*(2), 531–541. https://doi.org/10.1534/genetics.113.156026

Koury, S. A. 2023. Predicting recombination suppression outside chromosomal inversions in *Drosophila melanogaster* using crossover interference theory. *Heredity*, *130*(4), 196–208. https://doi.org/10.1038/s41437-023-00593-x

Lahn, B. T., & Page, D. C. 1999. Four evolutionary strata on the human X chromosome. *Science*, *286*(5441), 964–967. https://doi.org/10.1126/science.286.5441.964

Li, M., Sun, Y., Zhao, J., Shi, H., Zeng, S., Ye, K., Jiang, D., Zhou, L., Sun, L., Tao, W., Nagahama, Y., Kocher, T. D., & Wang, D. 2015. A tandem duplicate of anti-Müllerian hormone with a missense SNP on the Y chromosome is essential for male sex determination in Nile tilapia, *Oreochromis niloticus*. *PLOS Genetics*, *11*(11), e1005678. https://doi.org/10.1371/journal.pgen.1005678

Liu, W., Lin, S., Li, L., Tai, Z., & Liu, J.X. 2023. Zebrafish ELL-associated factors Eaf1/2 modulate erythropoiesis via regulating *gata1a* expression and WNT signaling to facilitate hypoxia tolerance. *Cell Regeneration*, *12*(1), 10. https://doi.org/10.1186/s13619-022-00154-3

Liu, Z., N. Mokariasl, B. Raeisi, M. Inch, C. L. Rihela, H. V. Barnes, B. Campbell, X. Zhang & J. S. Taylor. 2026. Predicting growth rate in sablefish (*Anoplopoma fimbria*) using a 300-locus SNP panel. *Aquaculture*, 613: 743313.

Lee, C. L. 1954. Sex chromosomes in ginkgo biloba. *American J of Botany* 41: 545–549.

Luckenbach, J. A. & W. T. Fairgrieve. 2016. Gonadal sex differentiation and effects of dietary methyltestosterone treatment in sablefish (*Anoplopoma fimbria*). *Fish Physiology and Biochemistry,* 42: 233-248.

Mank, J. E., Promislow, D. E. L., & Avise, J. C. 2006. Evolution of alternative sex-determining mechanisms in teleost fishes. *Biological Journal of the Linnean Society*, *87*(1), 83–93. https://doi.org/10.1111/j.1095-8312.2006.00558.x

Matsuda, M., Nagahama, Y., Shinomiya, A., Sato, T., Matsuda, C., Kobayashi, T., Morrey, C. E., Shibata, N., Asakawa, S., Shimizu, N., Hori, H., Hamaguchi, S., & Sakaizumi, M. 2002. *DMY* is a Y-specific DM-domain gene required for male development in the medaka fish. *Nature*, *417*(6888), 559–563. https://doi.org/10.1038/nature751

Miura, I., Y. Hasegawa, M. Ito, T. Ezaz, & M. Ogata. 2024. Disruption of sex-linked *SOX3* causes ZW female-to-male sex reversal in the Japanese frog *Glandirana rugosa*. *Biomolecules*, *14*(12), 1566. https://doi.org/10.3390/biom14121566

Metzger, D. C. H., & Schulte, P. M. 2018. The DNA methylation landscape of stickleback reveals patterns of sex chromosome evolution and effects of environmental salinity. *Genome Biology and Evolution*, *10*(3), 775–785. https://doi.org/10.1093/gbe/evy034

Morita, K. & Y. Takashima. 1998. Effect of female size on fecundity and egg size in white-spotted char: comparison between sea-run and resident forms. *Journal of Fish Biology,* 53: 1140-1142.

Mulugeta, E., Wassenaar, E., Sleddens-Linkels, E., Van IJcken, W. F. J., Heard, E., Grootegoed,

J. A., Just, W., Gribnau, J., & Baarends, W. M. 2016. Genomes of *Ellobius* species provide insight into the evolutionary dynamics of mammalian sex chromosomes. *Genome Research*, *26*(9), 1202–1210. https://doi.org/10.1101/gr.201665.115

Myosho, T., Otake, H., Masuyama, H., Matsuda, M., Kuroki, Y., Fujiyama, A., Naruse, K., Hamaguchi, S., & Sakaizumi, M. 2012. Tracing the emergence of a novel sex-determining gene in medaka, *Oryzias luzonensis*. *Genetics*, *191*(1), 163–170. https://doi.org/10.1534/genetics.111.137497

Rondeau, E. B., C. V. Laurie, S. C. Johnson & B. F. Koop. 2016. A PCR assay detects a male-specific duplicated copy of Anti-Müllerian hormone (*amh*) in the lingcod (*Ophiodon elongatus*). *BMC Research Notes,* 9: 230.

Rondeau, E. B., A. M. Messmer, D. S. Sanderson, S. G. Jantzen, K. R. Von Schalburg, D. R. Minkley, J. S. Leong, G. M. Macdonald, A. E. Davidsen, W. A. Parker, *et al.* 2013. Genomics of sablefish (*Anoplopoma fimbria*): expressed genes, mitochondrial phylogeny, linkage map and identification of a putative sex gene. *BMC Genomics,* 14: 452.

Rubi, T. L., A. G. McPherson, N. Mokariasl, B. F. Koop, B. Campbell, & J. S. Taylor, J. S. 2022. Sablefish (*Anoplopoma fimbria*) parentage analyses in aquaculture. *Aquaculture Research*, *53*(5), 1890–1895. https://doi.org/10.1111/are.15716

Shearwin-Whyatt, L., Fenelon, J., Yu, H., Major, A., Qu, Z., Zhou, Y., Shearwin, K., Galbraith, J., Stuart, A., Adelson, D., Zhang, G., Pyne, M., Johnston, S., Smith, C., Renfree, M., & Grützner, F. 2025. *AMHY* and sex determination in egg-laying mammals. *Genome Biology*, *26*(1), 144. https://doi.org/10.1186/s13059-025-03546-1

Sinclair, A. H., P. Berta, M. S. Palmer, J. R. Hawkins, B. L. Griffiths, M. J. Smith, J. W. Foster, A.-M. Frischauf, R. Lovell-Badge & P. N. Goodfellow. 1990. A gene from the human sex-determining region encodes a protein with homology to a conserved DNA-binding motif. *Nature,* 346: 240–244.

Smith, C. A., Roeszler, K. N., Ohnesorg, T., Cummins, D. M., Farlie, P. G., Doran, T. J., & Sinclair, A. H. 2009. The avian Z-linked gene *DMRT1* is required for male sex determination in the chicken. *Nature*, *461*(7261), 267–271. https://doi.org/10.1038/nature08298

Stöck, M., A. Horn, C. Grossen, D. Lindtke, R. Sermier, C. Betto-Colliard, C. Dufresnes, E. Bonjour, Z. Dumas, E. Luquet, *et al.* 2011. Ever-young sex chromosomes in European tree frogs. *PLoS Biology,* 9: e1001062.

Strunnikov, V. 1975. Sex control in silkworms. *Nature*, *255*(5504), 111–113. https://doi.org/10.1038/255111a0

Srikulnath, K., S. F. Ahmad, W. Singchat & T. Panthum. 2022. Do Ty3/Gypsy transposable elements play preferential roles in sex chromosome differentiation? *Life,* 12: 522.

Takehana, Y., M. Matsuda, T. Myosho, M. L. Suster, K. Kawakami, T. Shin-I, Y. Kohara, Y. Kuroki, A. Toyoda, A. Fujiyama, S. Hamaguchi, M. Sakaizumi, & K. Naruse. 2014. Co-option of *SOX3* as the male-determining factor on the Y chromosome in the fish *Oryzias dancena*. *Nature Communications*, *5*(1), 4157. https://doi.org/10.1038/ncomms5157

Tamai, H., H. Yamaguchi, K. Miyake, M. Takatori, T. Kitano, S. Yamanaka, S. Yui, K. Fukunaga, K. Nakayama & K. Inokuchi. 2017. Amlexanox Downregulates S100A6 to sensitize *KMT2A/AFF1*-positive acute lymphoblastic leukemia to TNFα treatment. *Cancer Research,* 77: 4426–4433.

The Tree of Sex Consortium. Tree of Sex: A database of sexual systems. *Sci Data* **1**, 140015 2014. https://doi.org/10.1038/sdata.2014.15

Thorgaard, G. H. 1977. Heteromorphic sex chromosomes in male rainbow trout. *Science* 196: 900–902.

Treaster, M. J., J. McCann, K. S. Solovei, R. J. Palmieri & M. A. White. 2025. A Y-linked duplication of anti-Müllerian hormone is the sex determination gene in threespine stickleback. *PLoS Genetics,* 21: e1011932.

Vergnaud, G., D. C. Page, M. C. Simmler, L. Brown, F. Rouyer, B. Noel, D. Botstein, A. D. La Chapelle, & J. Weissenbach. 1986. A deletion map of the human Y chromosome based on DNA hybridization: *Obstetrical & Gynecological Survey*, *41*(10), 641–643. https://doi.org/10.1097/00006254-198610000-00015

Voisin, N., R. E. Schnur, S. Douzgou, S. M. Hiatt, C. F. Rustad, N. J. Brown, D. L. Earl, B. Keren, O. Levchenko, S. Geuer, *et al*. 2021. Variants in the degron of AFF3 are associated with intellectual & disability, mesomelic dysplasia, horseshoe kidney, and epileptic encephalopathy. *The American Journal of Human Genetics*, 108: 857–873.

Warner, R. R. & S. E. Swearer. 1991. Social control of sex change in the bluehead wrasse, *Thalassoma bifasciatum* (Pisces: Labridae). *The Biological Bulletin,* 181: 199–204.

Waters, P. D., Wallis, M. C., & Graves, J. A. M. 2007. Mammalian sex—Origin and evolution of the Y chromosome and *SRY*. *Seminars in Cell & Developmental Biology*, *18*(3), 389–400. https://doi.org/10.1016/j.semcdb.2007.02.007

Xirocostas, Z. A., S. E. Everingham, & A. T. Moles. 2020. The sex with the reduced sex chromosome dies earlier: A comparison across the tree of life. *Biology Letters*, *16*(3), 20190867. https://doi.org/10.1098/rsbl.2019.0867

Zhang, X., G. Guan, M. Li, F. Zhu, Q, Liu, K. Naruse, A. Herpin, Y. Nagahama, J. Li, & Y. Hong. 2016. Autosomal *gsdf* acts as a male sex initiator in the fish medaka. *Scientific Reports*, *6*(1), 19738. https://doi.org/10.1038/srep19738

**Appendix A**. Synteny Analysis of the SAR on Chromosome 8

**Synteny Analysis**

Positional conservation of genes on and around the chromosome 8 SAR across species was analyzed to determine whether the significant peak on chromosome 8 was caused by misassembly. Genes within the chromosome 8 SAR (*mylk4b*, *serpinb1*, and *pycr3*), outside the chromosome 8 SAR but on chromosome 8 (*gmds*, *exoc2*, *infr4l*, *nos1apa*, *rgs16*, *brinp3a.1*, *aspm*), and in the chromosome 14 SAR (*gsdf, aff1, nup54*) were surveyed in eight species (3 spine stickleback, Atlantic salmon, amur stickleback, giant grouper, Atlantic cod, zebrafish, lumpfish, and rockfish) to determine if positions were conserved on respective chromosomes. Initially, positions of genes in chromosome 8 and chromosome 14 SARs were compared to identify if they co-occurred (on the same chromosome, indicates misassembly) or separately (on separate chromosomes, indicates no misassembly). Then, genes inside and outside the chromosome 8 sex-linked region were surveyed based on whether they occurred together.

For our initial analysis (Table A1), chromosomal locations of genes in the chromosome 8 and 14 SARs were compared to determine if a fragment of chromosome 14 had been misassembled onto chromosome 8 and ancestral linkage was maintaining the peak on chromosome 8. Genes in the chromosome 8 SAR (*mylk4b*, *serpinb1*, and *pycr3*) appeared together in all species except for Atlantic salmon, in which *mylk4b* is found on chromosome 13 and *pycr3* on chromosome 10. Genes surveyed from the chromosome 14 region (*gsdf*, *nup54*, and *aff1*) all appeared on the same chromosome. While there may be exceptions to observed synteny for cases in which search results were sparse (for example, two of three genes in the CHR14 SAR were not in Atlantic salmon or Atlantic cod), these data were considered sufficient to consider misassembly unlikely.

Our second analysis (Table A2) compared genes inside the chromosome 8 sex-linked region (*mylk4b*, *serpinb1*, and *pycr3*) to genes surrounding the region (*gmds*, *exoc2*, *nos1apa*, and *rgs16*) to determine if the chromosome 8 fragment had been misassembled. Homologs under different names were located using NCBI BLAST. All genes surveyed from outside the sex-linked region on chromosome 8 co-occurred with genes inside the region with the exception of *nos1apa* and *rgs16* homologs in the giant grouper. Therefore, an assembly error involving the misassembly of the chromosome 8 SAR was unlikely.

**Table A1.** Synteny analysis of genes in the chromosome 8 SAR and chromosome 14 SAR in lumpfish, rockfish, amur stickleback, three-spine stickleback, Atlantic salmon, giant grouper, Atlantic cod, and zebrafish. The label "-" indicates no result from a NCBI gene search using the name indicated (as labeled in the sablefish genome), and multiple results from the gene search are indicated using "/".

| | **CHR8** SAR | | | **CHR14** SAR | | |
|---|---|---|---|---|---|---|
| **Species** | *mylk4b* | *serpinb1* | *pycr3* | *gsdf* | *nup54* | *aff1* |
| Sablefish | 8 | 8 | 8 | 14 | 14 | 14 |
| Lumpfish | 4 | 4 | 4 | 12 | 12 | 12 |
| Rockfish | 5 | 5 | 5 | 19 | 19 | 19 |
| Amur stickleback | 3/8 | 8 | 8 | 14 | 14 | 14 |
| Three-spine stickleback | 8 | - | 8 | 14 | 14 | 14 |
| Atlantic salmon | 13 | - | 10 | - | 1 | - |
| Giant grouper | 21 | 13/21 | 21 | 17 | 17 | 17 |
| Atlantic cod | 12 | 12 | 12 | - | 4 | - |
| Zebrafish | 20 | 20 | 20 | 21 | 21 | 21 |

**Table A2.** Synteny analysis between genes in the chromosome 8 SAR and genes outside the chromosome 8 SAR in lumpfish, rockfish, three-spine stickleback, amur stickleback, Atlantic salmon, giant grouper, Atlantic cod, and zebrafish. Genes with no result are marked by "-" and genes that had paralogs on separate chromosome are indicated with "/".

| | **Inside** CHR8 SAR | | | **Outside** CHR8 SAR | | | |
|---|---|---|---|---|---|---|---|
| **Species** | *mylk4b* | *serpinb1* | *pycr3* | *gmds* | *exoc2* | *nos1apa* | *rgs16* |
| Sablefish | 8 | 8 | 8 | 8 | 8 | 8 | 8 |
| Lumpfish | 4 | 4 | 4 | 4 | 4 | 4 | 4 |
| Rockfish | 5 | 5 | 5 | 5 | 5 | - | - |
| Three-spine stickleback | 8 | - | 8 | 8 | 8 | 8 | - |
| Amur stickleback | 3/8 | 8 | 8 | 8 | 8 | 8 | 8 |
| Salmon | 13 | - | 10 | 10 | - | 10 | 10 |
| Giant grouper | 21 | 13/21 | 21 | 5 | 5 | 6 | 6 |
| Atlantic cod | 12 | 12 | 12 | 12 | 12 | 12 | 12 |
| Zebrafish | 20 | 20 | 20 | 20 | 20 | 6 | 6 |

**Appendix B:** Theoretical use of GWAS alone to determine SAR

While a SAR is defined by linkage to a MSD gene, it is important to supplement GWAS data with LD data as evidence for a SAR. Sex chromosome evolution begins with the occurrence of a mutation in a MSD gene on an autosome. Any allele on that chromosome will immediately become linked to sex, with especially strong linkage to the MSD gene in rare alleles. If the new MSD gene persists, it spreads through the population with hitchhiking alleles on the rest of the chromosome. Recombination gradually separates physically distant loci, but LD with nearer alleles is maintained. If a GWAS with sex were applied to the population at this point, there would be a wide peak of significance around the MSD gene. This peak would not be a true SAR as recombination suppression would not be present. However, recombination suppression is impossible to elucidate using a GWAS alone and introduces the need for LD analysis, which indicate chromosomal linkage independent of phenotype. The evolution of LD in the first stages of sex chromosome evolution are illustrated in Figure B1.

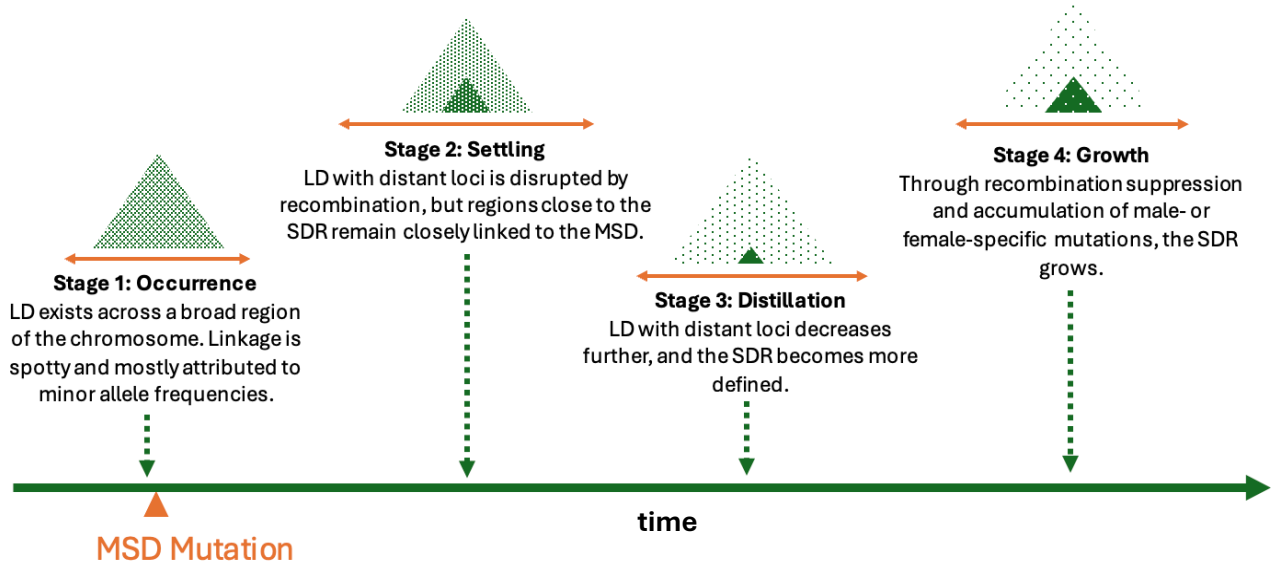


**Figure B1.** Theoretical progression of LD about a sex-determining locus immediately after the appearance of an MSD gene. One result from this project is that a peak in a GWAS does not imply a nascent sex-determining region. Instead, GWAS data must be used in conjunction with LD to identify a young sex-associated region.

After the MSD gene has covered the population, the vast majority of loci outside the MSD will be unlinked to sex. This is the case in pufferfish, in which there are no loci linked to sex other than the MSD gene itself (Kamiya *et al.*, 2012). For our study, GWAS data was used with LD analysis to determine a SAR. Due to the narrow peak of significantly linked SNPs in the SAR, high $R^2$ values between SNPs in the region, and isolation of significant association to males, we determined that sablefish have a nascent SAR instead of association as an artifact of recent acquisition of a MSD gene.