Research Paper

Identification of Sex-associated SNPs of Greater Amberjack (*Seriola dumerili*)

Junya Kawase[1✉], Jun-ya Aoki[2], Kazuhisa Hamada[3], Akiyuki Ozaki[2], and Kazuo Araki[2✉]

1. Department of Life Sciences, Graduate School of Bioresources, Mie University, 1577 Kurimamachiya-cho, Tsu City, Mie 514-8507, Japan
2. Research Center for Aquatic Breeding, National Research Institute of Aquaculture, Fisheries Research Agency, 224 Hiruda, Tamaki-cho, Watarai, Mie 519-0423, Japan
3. Marine Farm Laboratory Limited Company, 309 Takahiro, Tachibaura, Otsuki-cho, Hata-gun, Kochi 788-0352, Japan

✉ Corresponding authors: Junya Kawase, 515d303@m.mie-u.ac.jp, Tel.: +81 59-231-9626, ext. 2531, fax: +81 59-231-9634 and Kazuo Araki, arakin@affrc.go.jp, Tel.: +81 596-58-6411, fax: +81 596-58-6413

Abstract

The sex determination systems of fish are highly diverse compared with those of mammals. Thus, performing investigations using nonmodel fish species helps to understand the highly diverse sex determination systems of fish. Because greater amberjack (*Seriola dumerili*) is one of the most important edible fish globally and knowledge of its sex determination system is economically important in the field of aquaculture, we are interested in the mechanisms of sex determination of *Seriola* species. In this study, we identified sex-associated SNPs of greater amberjack using SNP information of 10 males and 10 females by an association test. We determined that the sex-associated SNPs were on chromosome 12 and mainly covered with two scaffolds (about 7.1 Mbp). Genotypes of sex-associated SNPs indicated that females are the heterogametic sex (ZZ/ZW). Furthermore, we compared the genomic structure of greater amberjack with those of Japanese amberjack (*Seriola quinqueradiata*), California yellowtail (*Seriola dorsalis*), and medaka (*Oryzias latipes*). Whole-genome alignments and synteny analysis indicated that the sex determination system of greater amberjack is markedly different from that of medaka and implied that the sex determination system is conserved in the *Seriola* species.

Key words: sex determination system, genome-wide association study, heterogametic female, synteny, *Seriola dumerili*

Introduction

The sex determination systems of fish are highly diverse, while sex is determined by the *Sry* gene on the Y chromosome in most mammals [1]. Although sex is determined genetically in many fish species, it is also affected by environmental factors (e.g., exogenous steroids and temperature) and some fish change sex at a certain growth stage. Various genes have been identified as master sex genes in fish, for example, *Dmy* of Japanese medaka (*Oryzias latipes*) [2], *GsdfY* of *Oryzias luzonensis* [3], *Sox3* of *Oryzias dancena* [4], *Amhy* of *Odontesthes hatcheri* [5], *Amhr2* of torafugu (*Takifugu rubripes*) [6], *Dmrt1* of tongue sole (*Cynoglossus semilaevis*) [7], and *SdY* of rainbow trout (*Oncorhynchus mykiss*) [8]. Therefore, researching these systems of various fish species is valuable to

boost our understanding of sex determination systems in general and their evolution in fish in particular.

We are interested in whether the same gene acts as an SD gene in the genus *Seriola*. We already know that the yellowtail sex-determining gene is at the end of linkage group 12 [9]. In addition, the sex-determining region of California yellowtail (*Seriola dorsalis*) has been identified and it has been hypothesized that estradiol 17-beta-dehydrogenase is the putative sex-determining gene [10].

Japanese amberjack (*Seriola quinqueradiata*) and greater amberjack (*Seriola dumerili*) are the main species for aquaculture production in Japan. Knowledge of their sex determination process is

valuable given the importance of controlling the sex ratio in aquaculture.

These two *Seriola* species are taxonomically close and their genome and transcriptome sequences are available [11–13]. We focused on this issue and also examined specifically whether the same gene acts as an SD gene in the *Seriola* species.

In this study, we identified sex-associated SNPs of greater amberjack and indicated that most of these SNPs are present on chromosome 12. Furthermore, we compared the genomic structure of greater amberjack with those of medaka, Japanese amberjack, and California yellowtail, and indicated that the sex determination system of greater amberjack is markedly different from that of medaka. However, the findings show that the SD region is conserved in the *Seriola* species.

Materials and Methods

Data collection

Genomic sequences of greater amberjack and yellowtail were retrieved from the DNA Data Bank of Japan (DDBJ) (greater amberjack: BDQW01000001–BDQW01034655, yellowtail: BDMU01000001–BDMU01000384). Genome scaffold sequences of greater amberjack and yellowtail were combined into each chromosome based on a previously published physical genetic map [12]. We also obtained transcriptome data of greater amberjack from DDBJ (greater amberjack: IACO01000001–IACO01045109, yellowtail: IACH01000001–IACH01013125). The sequence variation data of greater amberjack were obtained in a previous study [13].

The complementary DNA (cDNA) sequences of medaka were obtained from Ensembl [14]. The chromosome-level assembly of medaka was obtained from NCBI Assembly (ASM31367v1) [15]. The genome assembly of California yellowtail was obtained from GenBank (GCA_002814215.1) [10].

Synteny analysis

To identify orthologs, a nucleotide-to-nucleotide BLAST (blastn) search was performed between greater amberjack cDNA sequences and yellowtail cDNA sequences using an e-value cut-off of 1e−10 and reciprocal best hits. A translated-nucleotide-to-translated-nucleotide BLAST (tblastx) search was also performed between greater amberjack cDNA sequences and Japanese medaka cDNA sequences using an e-value cut-off of 1e−5 and reciprocal best hits. The orthologs were aligned to the genetic map of each species to identify syntenic relationships. Then, we constructed Circos plots [16] to illustrate the syntenic relationships.

Whole-genome alignments

Interspersed repeats and low-complexity DNA sequences of each genome assembly were masked using Tandem repeats finder [17] and Repeatmasker [18] with the following option: species "teleost fish." Then, whole-genome alignments of each orthologous chromosome between greater amberjack (Sdu) and yellowtail (Squ) were performed using LASTZ [19] with the following options: no transition, step=20, and chain.

Whole-genome alignments of each orthologous chromosome between greater amberjack and Japanese medaka (Ola) were also performed using LASTZ [19] with the following options: transition, step=20, and chain.

Genomic sequence alignments between Sdu chromosome 12 and the scaffold sequence (PEQF01098998.1) that contains the *Hsd17b1* gene of California yellowtail (Sdo) were performed using LASTZ with the following options: no transition, step=20, and chain.

After removing alignments that were shorter than the thresholds (Sdu vs. Squ and Sdu vs. Sdo: 400, Sdu vs. Ola: 150) from Rdotplot files generated from LASTZ, the alignment blocks were plotted by R [20].

Sex-determining SNP identification

Biallelic SNP information of 10 males and 10 females was extracted from the sequence variation data of greater amberjack. Those SNPs with a minor allele frequency of less than 0.2 or missing data for more than two individuals were filtered out. An association test was performed by GWASpoly [21] with the 1-dom model and false discovery rate level = 0.01. Haplotypes were estimated by PHASE v2.1.1 [22,23] from the genotype data included in the result of GWASPoly. To identify the proteins into which the transcripts in greater amberjack translate, a translated nucleotide-to-protein BLAST search (blastx) of the greater amberjack transcriptome sequences against protein sequences of RefSeq vertebrate other (release 69) [24] was performed with an e-value cut-off of 0.01.

The study that identified the sex-determining locus of *Seriola dorsalis* showed that there is a female-specific 61-nucleotide deletion and that estradiol 17-beta-dehydrogenase 1 (*Hsd17b1*) is the putative sex-determining gene [10]. To check whether greater amberjack has the female-specific deletion, we extracted indel variations near *Hsd17b1* (between 50 kb upstream and downstream) from the variation data of greater amberjack and investigated the existence of a female-specific deletion.
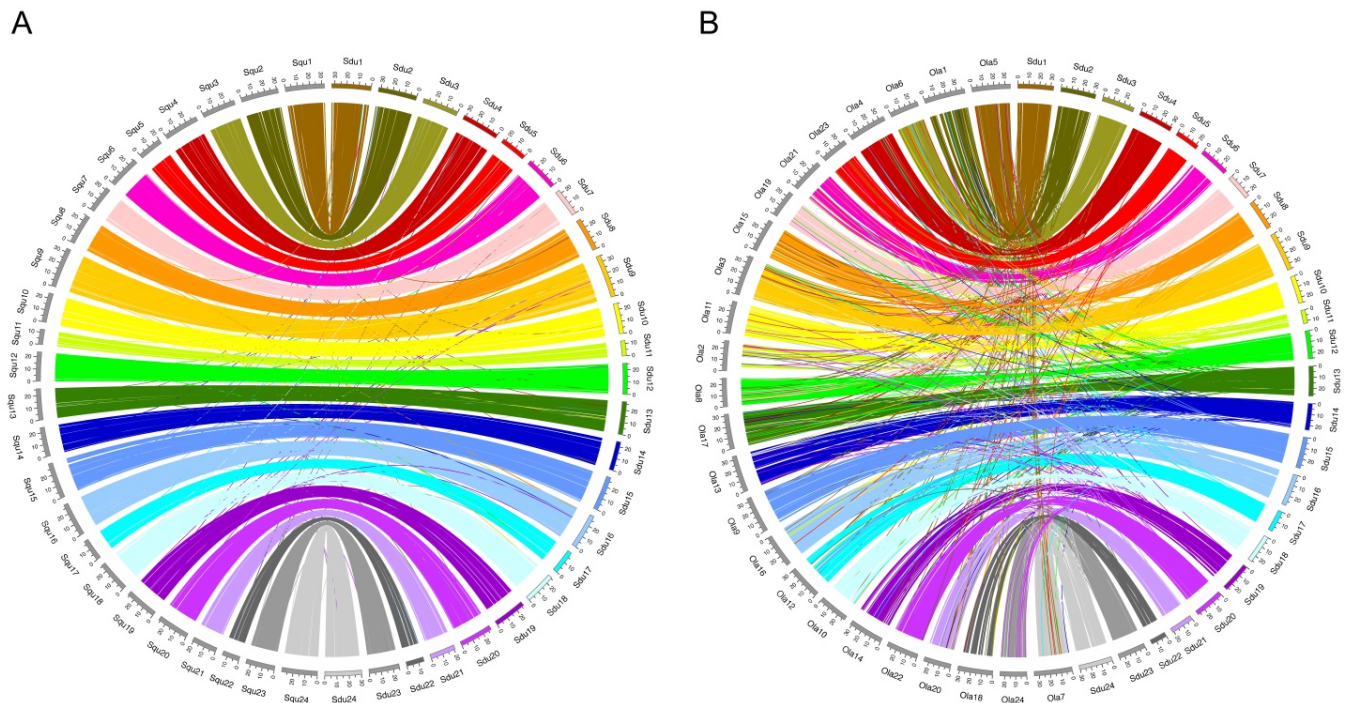
Figure 1. Circos plots showing conservation of synteny between (A) greater amberjack (right side, multicolored) and yellowtail (left side, gray), and (B) greater amberjack (right side, multicolored) and medaka (left side, gray). Lines linking two chromosomes indicate the location of orthologs.

Results

Synteny analysis

Between greater amberjack and yellowtail, 10,258 entries were identified as orthologs. The chromosomes that are thought to be orthologous chromosomes shared the orthologs and retained the order of those orthologs; in addition, inter-chromosomal translocations were observed at low frequency (Figure 1A). Between greater amberjack and medaka, 14,037 entries were identified as orthologs. The chromosome structure is generally conserved in these two species, but inter-chromosomal rearrangements have occurred more frequently than between greater amberjack and yellowtail (Figure 1B).

Whole-genome alignments

Pairwise whole-genome alignments of greater amberjack were performed against yellowtail and Japanese medaka with LASTZ [19]. The plots of alignments are shown in Figure 2 (greater amberjack vs. yellowtail) and Figure 3 (greater amberjack vs. Japanese medaka). The alignments between greater amberjack and yellowtail indicated that the chromosome structure was well conserved in these two species and homology of the genome sequences was retained in almost all regions of the genome (Figure 2, Figure S1). Hereafter, we assign the same

number to the orthologous chromosomes of these two species using the numbers assigned according to linkage group numbers in a previous genetic map study [11]. In greater amberjack and Japanese medaka, the sequences could be aligned in a large range in some regions (e.g., Sdu2: 4.5–17.8 Mb vs. Ola1: 25.4–8.4 Mb, Sdu3: 7.0–16.7 Mb vs. Ola6: 5.6–16.3 Mb and Sdu4: 17.8–26.0 Mb vs. Ola4: 14.0–23.6 Mb), but chromosomal inversions and intra-chromosomal translocations were frequently observed (Figure 3, Figure S2). Genomic sequence alignments between Sdu12 and the scaffold sequence (PEQF01098998.1) of California yellowtail showed completely linear correspondence (Figure 4).

Sex-associated SNP identification

After SNP pruning, 1,797,184 SNPs were used for subsequent analysis. The 1-dom model in GWASpoly was considered to be an appropriate statistical model because minor alleles control sex in both XX/XY and ZW/ZZ sex determination systems. The loci significantly associated with sex in the 1-dom test by GWASpoly are shown in Table 1 and Manhattan plots of the 1-dom-alt and 1-dom-ref models are shown in Figure 5. In 1-dom-alt and 1-dom-ref, it is assumed that the alternative allele and reference allele are dominant, respectively. SNPs associated with sex were mainly covered with two scaffolds, contig43 and contig168, and we found one

high-scoring locus, c168_4648290, associated with sex on contig168. The genotypes of sex-associated SNPs are shown in Table 2. Although eight loci (c16_12746979, c43_1018703, c43_1944077, c43_4403 196, c168_3069620, c168_4823950, c168_5414906, and c387_222985) were heterozygous in the male and homozygous in the female, or XX/XY system, the other loci were heterozygous in the female and homozygous in the male, or ZZ/ZW system (Table 2), and localized on chromosome 12. We define the SD region as the region with significant SNPs on chromosome 12 in this paper. This region ranges from nucleotide positions 56,399 bp to 4,405,455 bp on contig43 and from 338,405 bp to 5,457,861 bp on contig168 (Table 1). In this SD region on chromosome 12, there are genes that encode proteins associated with sex hormones, such as G-protein-coupled estrogen receptor 1 (GPER), estradiol 17-beta-dehydrogenase 1, 17-beta-hydroxysteroid dehydroge-nase 14, and transcription factors SOX8 and SOX9 (Table S1).

Table 1. Output from GWASpoly. The softoware was run with 1-dom model and FDR level = 0.01

Trait	Model	Threshold	Marker	Chrom	Position	Ref	Alt	Score	Effect
sex	1-dom-alt	5.25	c43_4405455	12	6857868	0	1	7.36	0.91
sex	1-dom-alt	5.25	c43_3763669	12	7499654	0	1	7.36	0.91
sex	1-dom-alt	5.25	c43_3697080	12	7566243	0	1	7.36	0.91
sex	1-dom-alt	5.25	c43_3027521	12	8235802	0	1	7.36	0.91
sex	1-dom-alt	5.25	c43_1782220	12	9481103	0	1	7.36	0.91
sex	1-dom-alt	5.25	c43_1229848	12	10033475	0	1	7.36	0.91
sex	1-dom-alt	5.25	c43_1221491	12	10041832	0	1	7.36	0.91
sex	1-dom-alt	5.25	c43_1018703	12	10244620	0	1	7.36	-0.91
sex	1-dom-alt	5.25	c43_826937	12	10436386	0	1	7.36	0.91
sex	1-dom-alt	5.25	c43_573705	12	10689618	0	1	7.36	0.91
sex	1-dom-alt	5.25	c43_470587	12	10792736	0	1	7.36	0.91
sex	1-dom-alt	5.25	c43_358726	12	10904597	0	1	7.36	0.91
sex	1-dom-alt	5.25	c43_330973	12	10932350	0	1	7.36	0.91
sex	1-dom-alt	5.25	c43_317236	12	10946087	0	1	7.36	0.91
sex	1-dom-alt	5.25	c43_316756	12	10946567	0	1	7.36	0.91
sex	1-dom-alt	5.25	c43_190404	12	11072919	0	1	7.36	0.91
sex	1-dom-alt	5.25	c43_96832	12	11166491	0	1	7.36	0.91
sex	1-dom-alt	5.25	c43_94819	12	11168504	0	1	7.36	0.91
sex	1-dom-alt	5.25	c43_56399	12	11206924	0	1	7.36	0.91
sex	1-dom-alt	5.25	c168_5414906	12	11661323	0	1	7.36	-0.91
sex	1-dom-alt	5.25	c168_5007150	12	12069079	0	1	7.36	0.91
sex	1-dom-alt	5.25	c168_5006505	12	12069724	0	1	7.36	0.91
sex	1-dom-alt	5.25	c168_4823950	12	12252279	0	1	7.36	-0.91
sex	1-dom-alt	5.25	c168_4823685	12	12252544	0	1	7.36	0.91
sex	1-dom-alt	5.25	c168_4648349	12	12427880	0	1	7.36	0.91
sex	1-dom-alt	5.25	c168_4648290	12	12427939	0	1	275.24	1
sex	1-dom-alt	5.25	c168_4536097	12	12540132	0	1	7.36	0.91
sex	1-dom-alt	5.25	c168_4147190	12	12929039	0	1	7.36	0.91
sex	1-dom-alt	5.25	c168_3641661	12	13434568	0	1	7.36	0.91
sex	1-dom-alt	5.25	c168_3376458	12	13699771	0	1	7.36	0.91
sex	1-dom-alt	5.25	c168_3287014	12	13789215	0	1	7.36	0.91
sex	1-dom-alt	5.25	c168_3160230	12	13915999	0	1	7.36	0.91
sex	1-dom-alt	5.25	c168_3109625	12	13966604	0	1	7.36	0.91
sex	1-dom-alt	5.25	c168_338405	12	16737824	0	1	7.36	0.91
sex	1-dom-alt	5.25	c82_71361	24	5556130	0	1	7.36	0.91
sex	1-dom-alt	5.25	c82_71360	24	5556131	0	1	7.36	0.91
sex	1-dom-alt	5.25	c82_71357	24	5556134	0	1	7.36	0.91
sex	1-dom-ref	5.25	c65_3467901	1	29094626	0	1	7.36	-0.91
sex	1-dom-ref	5.25	c43_4403196	12	6860127	0	1	7.36	0.91
sex	1-dom-ref	5.25	c43_4287549	12	6975774	0	1	7.36	-0.91
sex	1-dom-ref	5.25	c43_2438162	12	8825161	0	1	7.36	-0.91
sex	1-dom-ref	5.25	c43_1944077	12	9319246	0	1	7.36	0.91
sex	1-dom-ref	5.25	c43_827115	12	10436208	0	1	7.36	-0.91
sex	1-dom-ref	5.25	c43_362303	12	10901020	0	1	7.36	-0.91
sex	1-dom-ref	5.25	c168_5457861	12	11618368	0	1	7.36	-0.91
sex	1-dom-ref	5.25	c168_4996348	12	12079881	0	1	7.36	-0.91
sex	1-dom-ref	5.25	c168_3444410	12	13631819	0	1	7.36	-0.91
sex	1-dom-ref	5.25	c168_3116350	12	13959879	0	1	7.36	-0.91
sex	1-dom-ref	5.25	c168_3069620	12	14006609	0	1	7.36	0.91
sex	1-dom-ref	5.25	c16_12746979	15	4013786	0	1	7.36	0.91
sex	1-dom-ref	5.25	c387_222985	22	4833093	0	1	7.36	0.91

Table 2. Genotype data of SNPs detected by GWASpoly. Homozygous is indicated by one letter. M1, M2, .. M10 are male IDs. F1, F2, .. F10 are female IDs.

Marker	chr	position	M1	M2	M3	M4	M5	M6	M7	M8	M9	M10	F1	F2	F3	F4	F5	F6	F7	F8	F9	F10
			*Genotype of each sample																			
c43_4405455	12	6857868	A	A	A	A	A	A	AG	A	A	A	G	AG	G	AG	AG	G	AG	AG	AG	AG
c43_3763669	12	7499654	T	T	T	T	T	T	TG	T	T	T	G	TG	G	TG	TG	G	TG	TG	TG	TG
c43_3697080	12	7566243	G	G	G	G	G	G	GA	G	G	G	A	GA	A	GA	GA	A	GA	GA	GA	GA
c43_3027521	12	8235802	G	G	G	G	G	G	GA	G	G	G	GA	GA	GA	GA	GA	A	GA	GA	GA	GA
c43_1782220	12	9481103	G	G	G	G	G	G	GT	G	G	G	T	GT	GT	GT	GT	T	GT	GT	GT	GT
c43_1229848	12	10033475	CT	C	C	C	C	C	C	C	C	C	CT	T	CT	CT	CT	CT	CT	CT	CT	CT
c43_1221491	12	10041832	T	TC	T	T	T	T	T	T	T	T	TC	TC	TC	TC	TC	TC	TC	TC	TC	TC
c43_1018703	12	10244620	AG	G	AG	AG	G	AG	AG	AG	AG	AG	A	A	A	A	A	A	A	A	AG	A
c43_826937	12	10436386	T	T	T	T	T	T	T	T	T	T	TA	TA	T	TA	TA	TA	TA	TA	TA	TA
c43_573705	12	10689618	TA	T	T	T	T	T	T	T	T	T	TA	TA	A	TA	TA	TA	TA	TA	TA	TA
c43_470587	12	10792736	A	A	A	A	A	A	A	A	A	A	AC	A	C	AC	AC	AC	AC	AC	AC	AC
c43_358726	12	10904597	C	C	C	C	C	C	C	C	C	C	C	CA	CA	CA	CA	CA	CA	CA	CA	CA
c43_330973	12	10932350	CA	C	C	C	C	C	C	C	C	C	CA	CA	CA	CA	CA	CA	CA	CA	CA	CA
c43_317236	12	10946087	A	A	A	A	A	A	A	A	A	A	AG	AG	AG	AG	A	AG	AG	AG	AG	AG
c43_316756	12	10946567	T	T	T	T	T	T	T	T	T	T	TC	TC	TC	T	TC	TC	TC	TC	TC	TC
c43_190404	12	11072919	A	A	A	AT	A	A	A	A	A	A	T	AT	AT	AT	AT	AT	AT	AT	AT	AT
c43_96832	12	11166491	C	C	C	C	C	C	C	C	C	C	CT	CT	CT	C	CT	CT	CT	CT	CT	CT
c43_94819	12	11168504	T	T	T	T	T	T	T	T	T	T	TC	TC	TC	T	TC	TC	TC	TC	TC	TC
c43_56399	12	11206924	T	T	T	T	T	T	T	T	T	T	TC	TC	TC	TC	TC	TC	TC	TC	TC	TC
c168_5414906	12	11661323	TC	TC	TC	TC	C	TC	TC	TC	TC	TC	T	T	T	T	T	T	T	T	TC	T
c168_5007150	12	12069079	A	G	G	G	G	G	G	G	G	G	GA	GA	GA	GA	GA	GA	GA	GA	GA	GA
c168_5006505	12	12069724	G	A	A	A	A	A	A	A	A	A	AG	AG	AG	AG	AG	AG	AG	AG	AG	AG
c168_4823950	12	12252279	A	GA	GA	GA	A	GA	GA	GA	GA	GA	G	G	G	G	G	G	G	G	GA	G
c168_4823685	12	12252544	T	T	T	T	T	T	T	T	T	T	TC	C	TC	T	TC	TC	TC	TC	TC	TC
c168_4648349	12	12427880	G	G	G	G	G	G	G	G	G	G	GA	G	GA	GA	GA	GA	GA	GA	GA	GA
c168_4648290	12	12427939	A	A	A	A	A	A	A	A	A	A	AT	AT	AT	T	AT	AT	AT	AT	AT	AT
c168_4536097	12	12540132	A	A	A	A	A	A	A	A	A	A	A	AG	AG	AG	AG	AG	AG	AG	AG	AG
c168_4147190	12	12929039	C	C	C	C	C	C	C	C	C	C	C	CT	CT	CT	CT	CT	CT	CT	CT	CT
c168_3641661	12	13434568	G	G	G	G	G	G	G	G	G	G	GA	GA	GA	G	GA	GA	GA	GA	GA	GA
c168_3376458	12	13699771	A	A	A	A	A	A	A	A	A	A	AT	AT	AT	A	AT	AT	AT	AT	AT	AT
c168_3287014	12	13789215	C	C	CT	C	C	C	C	C	C	C	CT	CT	CT	CT	CT	CT	CT	CT	CT	CT
c168_3160230	12	13915999	C	CT	C	C	C	C	C	C	C	C	CT	CT	CT	CT	CT	CT	CT	CT	CT	CT
c168_3109625	12	13966604	G	G	G	G	G	G	G	G	G	G	G	GA	GA	A	GA	GA	GA	GA	GA	GA
c168_338405	12	16737824	C	C	C	C	C	C	C	C	T	C	CT	CT	CT	CT	T	CT	CT	CT	CT	CT
c82_71361	24	5556130	A	A	A	A	AG	A	A	A	A	A	AG	AG	AG	AG	AG	AG	AG	AG	G	AG
c82_71360	24	5556131	A	A	A	A	AC	A	A	A	A	A	AC	AC	AC	AC	AC	AC	AC	AC	C	AC
c82_71357	24	5556134	A	A	A	A	AG	A	A	A	A	A	AG	AG	AG	AG	AG	AG	AG	AG	G	AG
c65_3467901	1	29094626	G	G	G	G	G	G	C	G	G	G	C	CG	C	CG	CG	CG	CG	C	CG	CG
c43_4403196	12	6860127	A	AG	AG	AG	A	AG	AG	AG	AG	AG	G	G	G	G	G	G	G	G	AG	G
c43_4287549	12	6975774	C	C	C	C	C	C	TC	C	C	C	T	TC	TC	TC	TC	T	TC	TC	TC	TC
c43_2438162	12	8825161	G	G	G	G	G	G	AG	G	G	G	A	AG	AG	AG	AG	A	AG	AG	AG	AG
c43_1944077	12	9319246	TC	T	TC	TC	T	TC	TC	TC	TC	TC	C	C	C	C	C	C	C	C	TC	C
c43_827115	12	10436208	A	A	A	A	A	A	A	A	A	A	GA	GA	A	GA	GA	GA	GA	GA	GA	GA
c43_362303	12	10901020	A	A	A	TA	A	A	A	A	A	A	T	TA	TA	TA	TA	TA	TA	TA	TA	TA
c168_5457861	12	11618368	C	C	T	C	C	C	C	C	C	C	T	TC	TC	T	TC	TC	TC	TC	TC	TC
c168_4996348	12	12079881	TA	A	A	A	A	A	A	A	A	A	TA	TA	T	T	TA	TA	TA	TA	TA	TA
c168_3444410	12	13631819	GC	C	C	C	C	C	C	C	C	C	GC	GC	G	GC	GC	GC	GC	GC	GC	GC
c168_3116350	12	13959879	C	C	C	C	C	C	C	C	C	C	GC	C	GC	GC	GC	GC	GC	GC	GC	GC
c168_3069620	12	14006609	TG	TG	TG	TG	T	TG	TG	TG	TG	TG	G	G	G	G	G	G	G	G	TG	G
c16_12746979	15	4013786	TA	TA	TA	TA	TA	T	T	TA	TA	TA	A	A	A	A	A	A	A	A	TA	A
c387_222985	22	4833093	GA	GA	G	GA	GA	GA	GA	GA	GA	GA	A	A	A	A	A	A	A	A	A	GA

The reconstructed haplotypes estimated by PHASE from the genotype data are shown in Table 3. All females have haplotypes that are unique to females, while five male individuals (M5, M6 M7, M8, and M10 in Table 4) have two haplotypes that appear in both males and females.

Indel variations near *Hsd17b1* were extracted from the sequence variation data of greater amberjack and we investigated whether a female-specific deletion was present. However, the female-specific deletion could not be found (Table S2).

Discussion

Sex determinant location

By genome-wide sex association analysis, the sex-associated markers were detected on chromosomes (chr)12, 24, 1, 15, and 22 in decreasing order of the number of detected markers on each chromosome. A

total of 31 of the 51 sex-associated SNPs are localized on chr12. Sex-associated linkage analysis of yellowtail showed that sex determinants are present on the tip of linkage group 12 [9]. Thus, part of chr12 acts as a sex chromosome in greater amberjack, as well as in yellowtail. Greater amberjack's SD loci were present at scaffolds 168 and 43. These sequences are located in the middle of chr12 and the range of the SD region is approximately 7.1 Mb, if ignoring the gaps between scaffolds.

Comparative analysis

Whole-genome alignments and synteny analysis between yellowtail and greater amberjack showed a very well-conserved chromosome structure (Figure 2). Yellowtail has the ZW/ZZ system and its SD segment is located on Squ12 [9]. The sex-associated marker of greater amberjack is also located at Sdu12, which is a chromosome orthologous to Squ12. The plots of alignment of Sdu12 vs. Squ12 show a linear pattern (Figure S1). Research on the sex-linked SNPs of yellowtail indicated that *Gipc1* and *Sox9* are present near the SD region [9], which is also the case in greater

amberjack. The SD region is present on the tip of the linkage map in yellowtail [9], whereas it is present in the middle of the chromosome in greater amberjack. This is because recombination is suppressed in the sex chromosome and thus the SD region is very narrow on the linkage map of yellowtail.

The alignment of Sdu12 vs. the scaffold sequence (PEQF01098998.1) of California yellowtail also demonstrated well-conserved sequences. Hence, it is inferred that the three *Seriola* species share sex determination mechanisms because these *Seriola* species exhibit female heterogamety and have the same genes in orthologous genomic regions.

Medaka has the SD gene on chromosome 1 (Ola1). However, whole-genome alignment showed remarkable homology between Ola8 and Sdu12 (Figure 6). Medaka exhibits male heterogamety and its master SD gene is *Dmy*. Hence, the *Seriola* species and medaka clearly have different SD mechanisms and it is thought that SD systems rapidly evolved after the divergence of the common ancestor of those species.
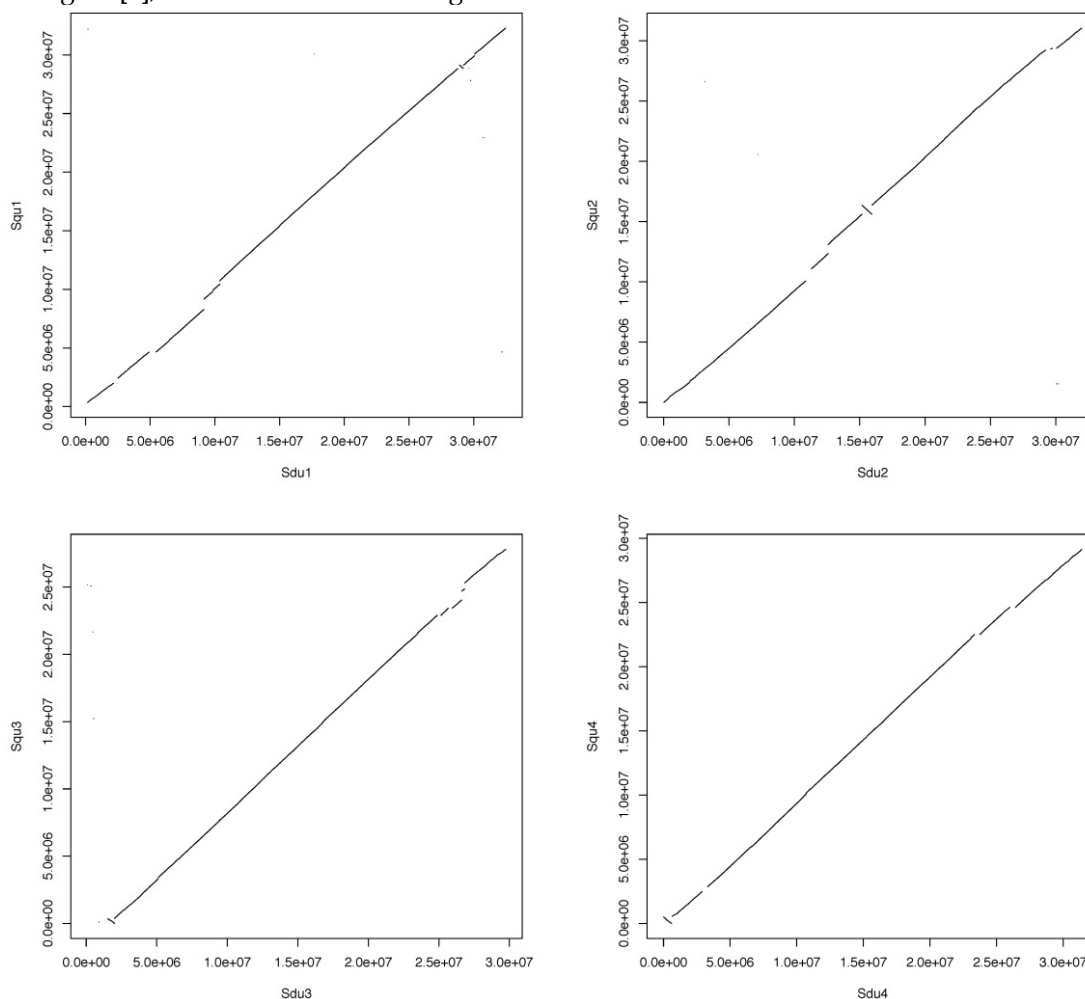

Figure 2. Plots showing the alignment of each orthologous chromosome pair (1-4). The horizontal axis and the vertical axis indicate the positions (base pair) of greater amberjack (Sdu) and yellowtail (Squ), respectively.
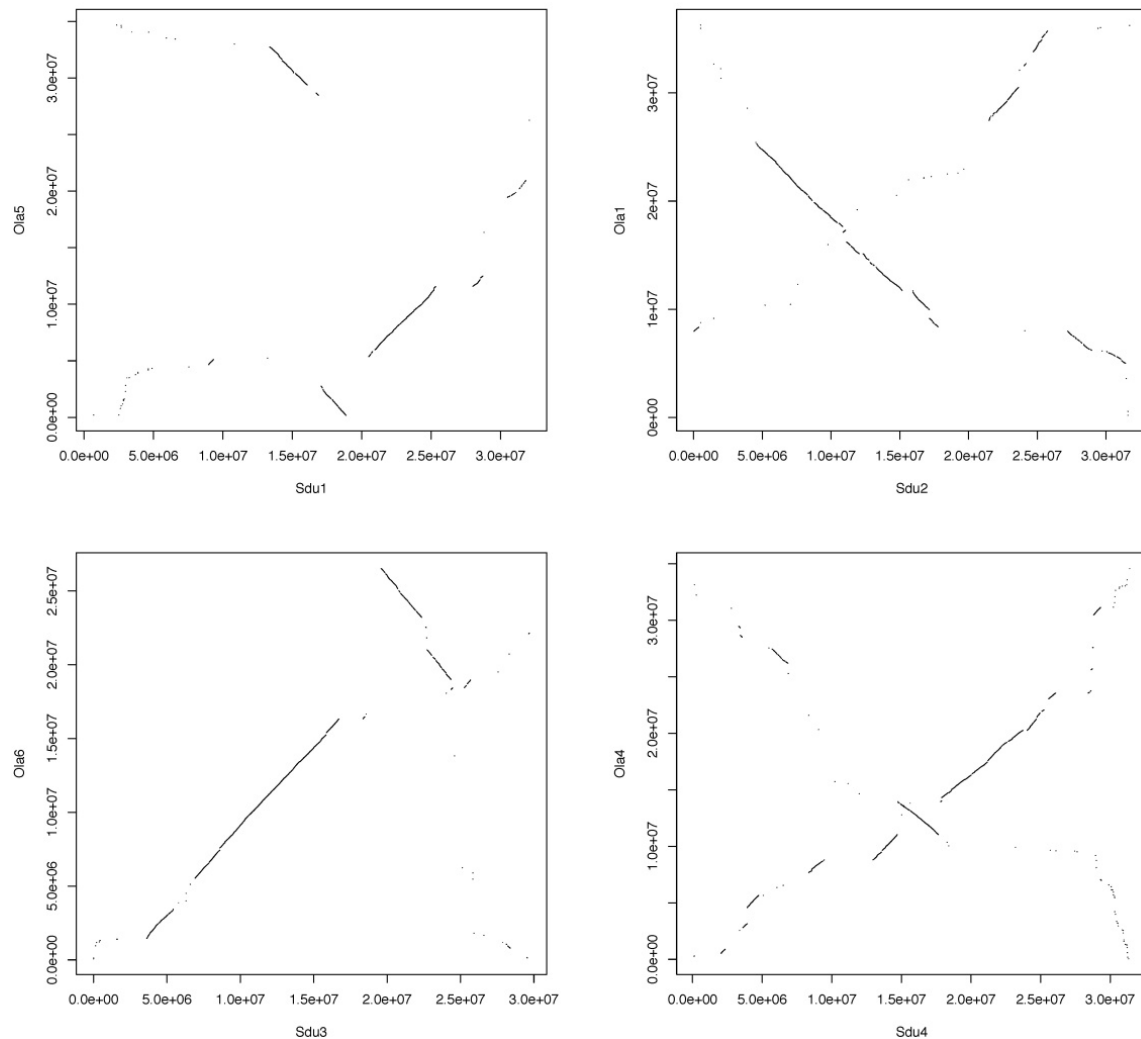
Figure 3. Plots showing the alignment of each orthologous chromosome pair. The horizontal axis and the vertical axis indicate the positions (base pair) of greater amberjack (Sdu) and medaka (Ola), respectively.
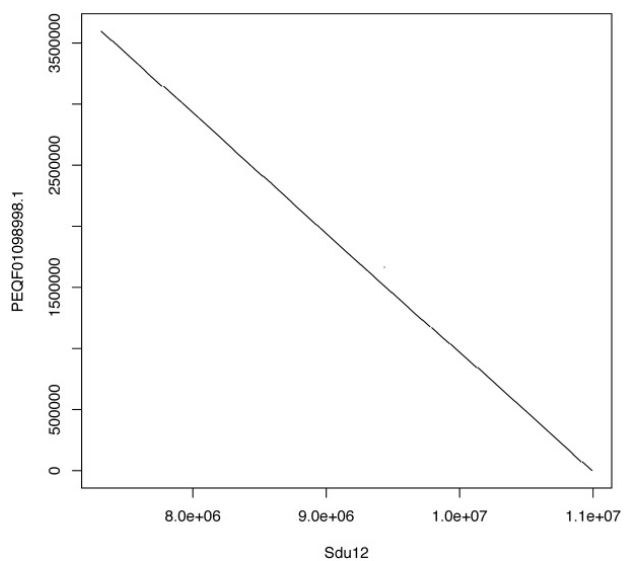

Figure 4. Plots showing the alignment of Sdu12 vs. the scaffold sequence (PEQF01098998.1) that contains the *Hsd17b1* gene of California yellowtail.

Amberjack has the ZW system

In this study, almost all sex-associated SNPs of greater amberjack demonstrated that females are heterozygous (Table 2) and all females have haplotypes that are unique to females (Table 3). Therefore, it is likely that greater amberjack females are the heterogametic sex (ZW/ZZ). SNPs between loci c43_4405455 and c43_1782220 were heterozygous in the male individual (M7 in Table 2) and homozygous in the female individual (F6 in Table 2), so the segment might have undergone recombination. Hence, this SD region may tolerate small-scale recombination and has not yet completely evolved into a sex chromosome.

In the ZW/ZZ system, sex differentiation is caused by the W or Z chromosome. The presence of a W chromosome acts dominantly to lead to the development of a female, while the presence of a Z chromosome leads to the development of a male by a dosage-sensitive mechanism.
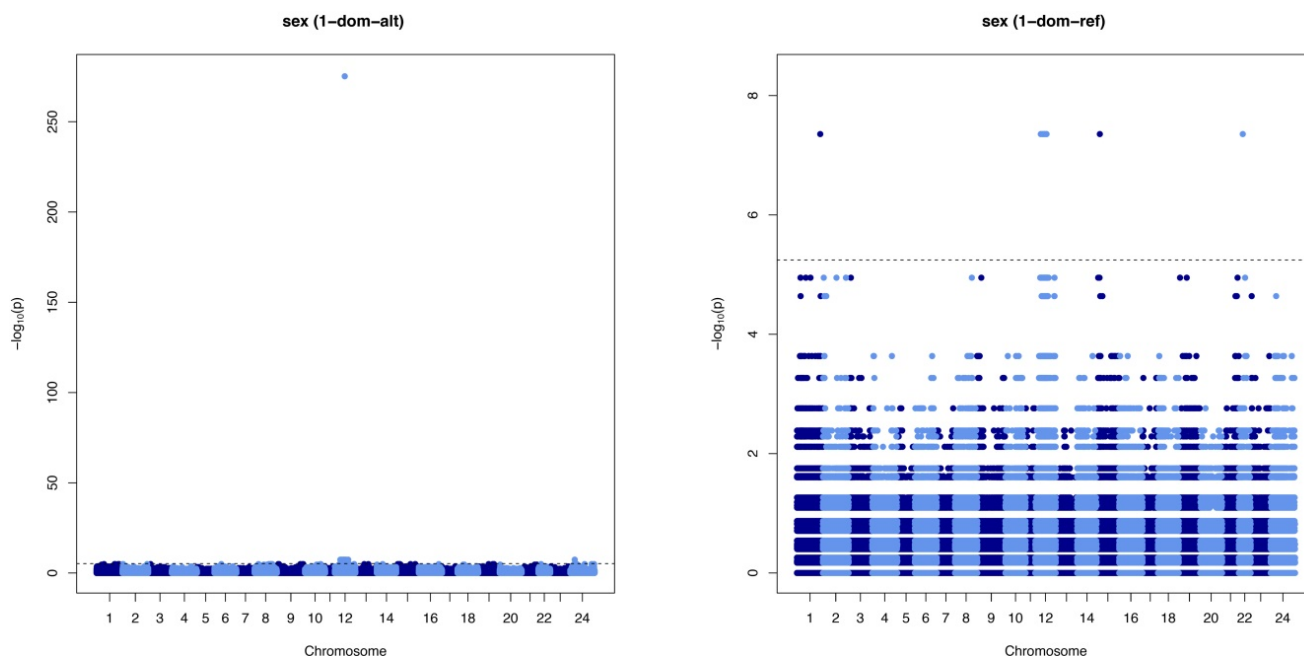
Figure 5. Manhattan plots displaying the result of the 1-dom test by GWASpoly. The horizontal axis indicates the chromosome number and the position of each SNP. The vertical axis indicates the negative logarithm of the P-value for each SNP. Each dot signifies an SNP. The broken line indicates the threshold FDR level of 0.01.
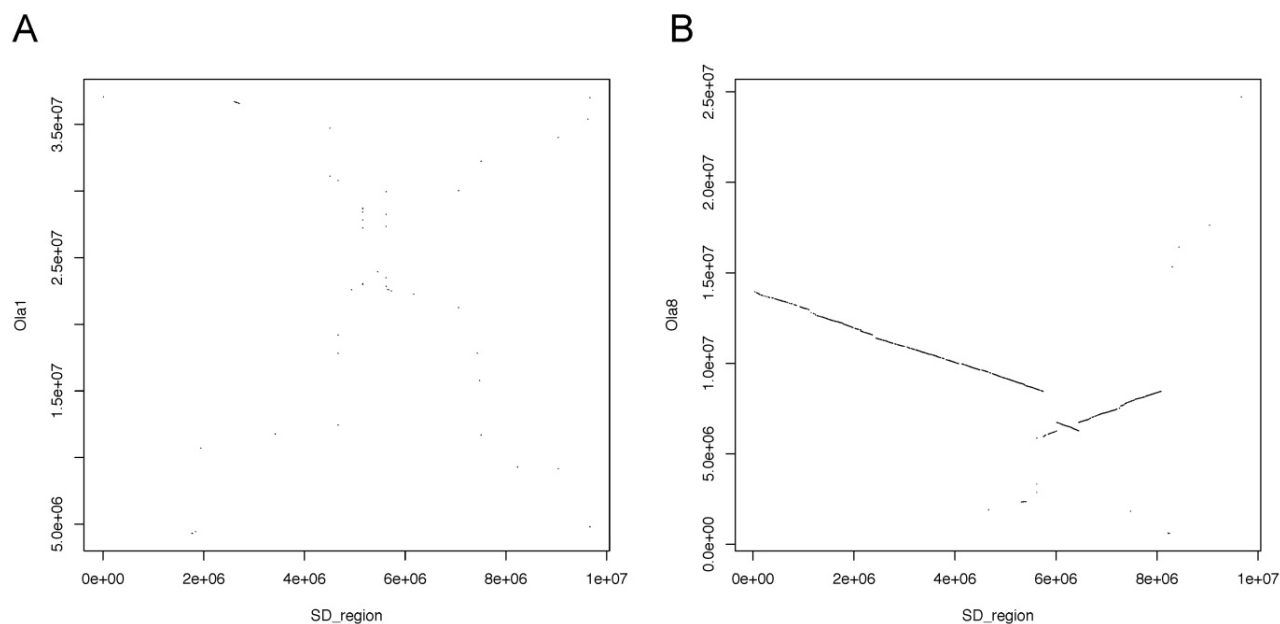

Figure 6. Plots showing the alignments between the sequence of the SD region in greater amberjack and (A) medaka chromosome 1 (Ola1) and (B) medaka chromosome 8 (Ola8).

Various sex determination mechanisms have been reported in fish species, and in most fish species, heteromorphic sex chromosomes have not been revealed [25]. In the SD region of greater amberjack, there are genes that encode proteins associated with sex hormones, such as G-protein-coupled estrogen receptor 1 (GPER), estradiol 17-beta-dehydrogenase 1, and 17-beta-hydroxysteroid dehydrogenase 14 (Table S1). *Sox9* and *Sox8* are present in the SD region, so they might be sex determinants. We also found

several genes that are expected to be involved in sex differentiation in the SD region.

GPER is activated by estrogen, the main female sex hormone, and plays an important role in female development. In some teleosts, 17-beta estradiol plays a critical role in ovarian differentiation [26–29]. Hence, if the *Gper* gene or the estradiol 17-beta-dehydroge-nase 1 gene in the SD region affects the female pathway and is involved with promoting sex differentiation in greater amberjack, the W chromo-

some is supposed to induce female differentiation. Recently, Purcell et al. identified the sex-determining region of California yellowtail using genome assembly and re-sequences and hypothesized that the estradiol 17-beta-dehydrogenase 1 gene was the putative sex-determining gene [10]. Thus, it is assumed that the three *Seriola* species may retain the same SD region that their common ancestor acquired, and that same gene might be involved in sex determination.

Table 3. Haplotypes estimated by PHASE software.

No.	haplotypes
1	AACTGGGTGCTGATTAACCATACTTCCGAAATGAACGCACCCGTC
2	AACTGGGTGCTGATTAACCATACTTCCGAAATGAACGCACCCGTT
3	AACTGGGTGCTGATTAACCATACTTTCGAAATGAACGCATCCGTC
4	AACTGGGTGCTGATTAACAATACTTCCAGAATGAACGCACCCGTC
5	AACTGGGCGTTAATAAACCATACTTCTAGTATGAACGGACCCGGC
6	AGCTGGGTGCCGATTAACCATACTTCTGAAGTGAACGCACTCGGC
7	AGCTGGGCGCTAATTAACCATACTTCTGAAGTGAACGCACCCGGC
8	AGCTGGGCGCTAATTAACCATACTTCTGAAGTGAACGCACCCGGT
9	AGCTGGGCGCTAATTAACCATACTTTTGATGTGTACGCACCCAGC
10	AGCTGGGCGCTAATTAACCATACTTTTGAAGTGAACGCACCCGGC
11	AGCTGGGCGCTAATTATCCATTCTTCTGAAGTGAACGCACCCGGC
12	AGCTGGGCGTTAATTAACCATACTTCTGAAGCGAACGCACCCGGC
13	GGCGAAACTTCAATACTAAGCTTCCTTAGTGCATGTAGTTTGAGT
14	GGTGAGGCGCTAATACACCATACTTCTGATGTGAACGGACCCGGC
15	GGTGAGACTTCAGAACTCAGCTTCCTTAGTGCATACAGTTTGGGT
16	GGTGAAACTCTAATTAACCATACTTCTGAAGTGAACGCACCCGGC
17	GGTGAAACTCTAATTATCCATTCTTTTGAAGTGAACGCACCCGGC
18	GGTGAAACTTCAGAAATAAGCTTCCTTAGTGCGTGTAGTTTCAGT
19	GGTGAAACTTCAGAACTAAACTTCCTTAGTGCATGTAGTTTGAGT
20	GGTGAAACTTCAGAACTAAGTTCTTTAGTGTATGTGGATTGAGT
21	GGTGAAACTTCAGAACTAAGCTTCCTTAGTGCATGTAGTTTGAGT

Table 4. Haplotype combinations of each individual estimated by PHASE software. The number of haplotype corresponds to table 3.

ID	haplotype combinations
M1	(4,5)
M2	(1,6)
M3	(3,10)
M4	(1,11)
M5	(1,1)
M6	(1,7)
M7	(1,16)
M8	(1,7)
M9	(2,8)
M10	(1,7)
F1	(15,17)
F2	(12,18)
F3	(13,14)
F4	(9,20)
F5	(8,19)
F6	(16,21)
F7	(7,21)
F8	(7,21)
F9	(1,21)
F10	(7,21)

GIPC (PDZ domain-containing protein, which interacts specifically with the C terminus of RGS-GAIP) was originally identified as a protein that binds to the C terminus of the RGS (G protein signaling regulator) protein GAIP (RGS19), a GTPase-activating protein (GAP) for Gαi subunits [30]. Endoglin, which is one of the PDZ ligands, interacts with GIPC and specifically enhances the TGF-β1-induced phosphorylation of Smad1/5/8 [31]. Testicular TGF-β1 modulates Leydig cell steroidogenesis, the organization of peritubular myoid cells, testis development, and spermatogenesis [32]. Therefore, although the role of endoglin in the reproductive system has not been shown, GPIC1 might play a role in sex determination in greater amberjack.

SOX9 and SOX8 are group E SOX proteins. These transcription factors contain, besides a DNA-binding HMG domain and a transactivation domain, a DNA-dependent dimerization domain, unique among SOX proteins. In mammals, *Sox9* is activated by SRY in pre-Sertoli cells and induces Sertoli cell and testis cord differentiation [33]. SOX9 binds to a SOX binding site within the *Amh* promoter and interacts with SF1 to synergistically activate *Amh* expression [34]. SOX8 has the same mechanism of action as SOX9, but acts less efficiently [35]. In many teleosts such as medaka, zebrafish, three-spine stickleback, and rice field eel, *Sox9a* and/or *Sox9b* are expressed in gonad [36–41]. In greater amberjack, *Sox9* and *Sox8* are present in the SD region, so they might be sex determinants.

Although we found several genes that are expected to be involved in sex differentiation in the SD region, the molecular mechanisms of sex differentiation need to be investigated. We will attempt to determine which among these candidate genes is key for the sex determination of greater amberjack.

Supplementary Material

Supplementary figures.
http://www.jgenomics.com/v06p0053s1.pdf
Supplementary tables.
http://www.jgenomics.com/v06p0053s2.xlsx

Acknowledgments

Competing Interests

The authors have declared that no competing interest exists.

References

1. Waters PD, Wallis MC, Graves JAM. Mammalian sex—Origin and evolution of the Y chromosome and SRY. *Semin Cell Dev Biol*. 2007;18(3):389-400.
2. Matsuda M, Nagahama Y, Shinomiya A, et al. DMY is a Y-specific DM-domain gene required for male development in the medaka fish. *Nature*. 2002;417(6888):559-563.
3. Myosho T, Otake H, Masuyama H, et al. Tracing the emergence of a novel sex-determining gene in medaka, *Oryzias luzonensis*. *Genetics*. 2012;191(1):163-170.
4. Takehana Y, Matsuda M, Myosho T, et al. Co-option of Sox3 as the male-determining factor on the Y chromosome in the fish *Oryzias dancena*. *Nat Commun*. 2014;5:4157.
5. Hattori RS, Strüssmann CA, Fernandino JI, Somoza GM. Genotypic sex determination in teleosts: insights from the testis-determining amhy gene. *Gen Comp Endocrinol*. 2013;192:55-59.
6. Kamiya T, Kai W, Tasumi S, et al. A trans-species missense SNP in Amhr2 is associated with sex determination in the tiger pufferfish, *Takifugu rubripes* (fugu). Peichel CL, ed. *PLoS Genet*. 2012;8(7):e1002798.
7. Cui Z, Liu Y, Wang W, et al. Genome editing reveals dmrt1 as an essential male sex-determining gene in Chinese tongue sole (*Cynoglossus semilaevis*). *Sci Rep*. 2017;7:42213.
8. Yano A, Guyomard R, Nicol B, et al. An immune-related gene evolved into the master sex-determining gene in rainbow trout, *Oncorhynchus mykiss*. *Curr Biol*. 2012;22(15):1423-1428.
9. Koyama T, Ozaki A, Yoshida K, et al. Identification of sex-linked SNPs and sex-determining regions in the yellowtail genome. *Mar Biotechnol*. 2015;17(4):502-510.
10. Purcell CM, Seetharam AS, Snodgrass O, Ortega-García S, Hyde JR, Severin AJ. Insights into teleost sex determination from the Seriola dorsalis genome assembly. *BMC Genomics*. 2018;19(1):1-11.
11. Aoki J, Kai W, Kawabata Y, et al. Second generation physical and linkage maps of yellowtail (*Seriola quinqueradiata*) and comparison of synteny with four model fish. *BMC Genomics*. 2015;16(1):406.
12. Kawase J, Aoki J, Araki K. Constructing a "Chromonome" of yellowtail (*Seriola quinqueradiata*) for comparative analysis of chromosomal rearrangements. *J Genomics*. 2018;6:9-19.
13. Araki K, Aokic J, Kawase J, et al. Whole genome sequencing of greater amberjack (*Seriola dumerili*) for snp identification on aligned scaffolds and genome structural variation analysis using parallel resequencing. *Int J Genomics*. 2018;2018:1-12.
14. Aken BL, Achuthan P, Akanni W, et al. Ensembl 2017. *Nucleic Acids Res*. 2017;45(D1):D635-D642.
15. Kasahara M, Naruse K, Sasaki S, et al. The medaka draft genome and insights into vertebrate genome evolution. *Nature*. 2007;447(7145):714-719.
16. Krzywinski M, Schein J, Birol I, et al. Circos: An information aesthetic for comparative genomics. *Genome Res*. 2009;19(9):1639-1645.
17. Benson G. Tandem repeats finder: A program to analyze DNA sequences. *Nucleic Acids Res*. 1999;27(2):573-580.
18. [Internet] Smit A, Hubley R, Green P. RepeatMasker Open-4.0. http://www.repeatmasker.org/
19. Harris RS. Improved Pairwise Alignment of Genomic DNA. TPhD thesis, Pennsylvania State University; 2007.
20. [Internet] R_Core_Team. R: A language and environment for statistical computing. https://www.r-project.org/
21. Rosyara UR, De Jong WS, Douches DS, Endelman JB. Software for genome-wide association studies in autopolyploids and its application to potato. *Plant Genome*. 2016;9(2):1-10.
22. Stephens M, Smith NJ, Donnelly P. A new statistical method for haplotype reconstruction from population data. *Am J Hum Genet*. 2001;68(4):978-989.
23. Stephens M, Scheet P. Accounting for decay of linkage disequilibrium in haplotype inference and missing-data imputation. *Am J Hum Genet*. 2005;76(3):449-462.
24. O'Leary NA, Wright MW, Brister JR, et al. Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. *Nucleic Acids Res*. 2016;44(D1):D733-D745.
25. Devlin RH, Nagahama Y. Sex determination and sex differentiation in fish: An overview of genetic, physiological, and environmental influences. *Aquaculture*. 2002;208(3-4):191-364.
26. Kobayashi T, Kajiura-Kobayashi H, Nagahama Y. Induction of XY sex reversal by estrogen involves altered gene expression in a teleost, tilapia. *Cytogenet Genome Res*. 2003;101(3-4):289-294.
27. Miyata S, Kubo T. In vitro effects of estradiol and aromatase inhibitor treatment on sex differentiation in *Xenopus laevis* gonads. *Gen Comp Endocrinol*. 2000;119(1):105-110.
28. Rashid H, Kitano H, Hoon Lee K, et al. Fugu (*Takifugu rubripes*) sexual differentiation: CYP19 regulation and aromatase inhibitor induced testicular development. *Sex Dev*. 2007;1(5):311-322.
29. Piferrer F, Zanuy S, Carrillo M, Solar II, Devlin RH, Donaldson EM. Brief treatment with an aromatase inhibitor during sex differentiation causes chromosomally female salmon to develop as normal, functional males. *J Exp Zool*. 1994;270(3):255-262.
30. De Vries L, Lou X, Zhao G, Zheng B, Farquhar MG. GIPC, a PDZ domain containing protein, interacts specifically with the C terminus of RGS-GAIP. *Proc Natl Acad Sci U S A*. 1998;95(21):12340-12345.
31. Lee NY, Ray B, How T, Blobe GC. Endoglin promotes transforming growth factor β-mediated Smad 1/5/8 signaling and inhibits endothelial cell migration through its association with GIPC. *J Biol Chem*. 2008;283(47):32527-32533.
32. Rocio GC, Saul CR, Silvia Ines G-C. Testicular expression of the TGF-β1 system and the control of Leydig cell proliferation. *Adv Biosci Biotechnol*. 2013;04(10):1-7.
33. Barrionuevo F, Scherer G. SOX E genes: SOX9 and SOX8 in mammalian testis development. *Int J Biochem Cell Biol*. 2010;42(3):433-436.
34. De Santa Barbara P, Bonneaud N, Boizet B, et al. Direct interaction of SRY-related protein SOX9 and steroidogenic factor 1 regulates transcription of the human anti-Müllerian hormone gene. *Mol Cell Biol*. 1998;18(11):6653-6665.
35. Schepers G, Wilson M, Wilhelm D, Koopman P. Sox8 is expressed during testis differentiation in mice and synergizes with SF1 to activate the Amh promoter in vitro. *J Biol Chem*. 2003;278(30):28101-28108.
36. Chiang EF-L, Pai CI, Wyatt M, Yan YL, Postlethwait J, Chung B. Two sox9 genes on duplicated zebrafish chromosomes: expression of similar transcription activators in distinct sites. *Dev Biol*. 2001;231(1):149-163.
37. Cresko WA, Yan Y-L, Baltrus DA, et al. Genome duplication, subfunction partitioning, and lineage divergence: Sox9 in stickleback and zebrafish. *Dev Dyn*. 2003;228(3):480-489.
38. Nakamoto M, Suzuki A, Matsuda M, Nagahama Y, Shibata N. Testicular type Sox9 is not involved in sex determination but might be in the development of testicular structures in the medaka, *Oryzias latipes*. *Biochem Biophys Res Commun*. 2005;333(3):729-736.
39. Ijiri S, Kaneko H, Kobayashi T, et al. Sexual dimorphic expression of genes in gonads during early differentiation of a teleost fish, the Nile tilapia *Oreochromis niloticus*. *Biol Reprod*. 2008;78(2):333-341.
40. Zhou R, Liu L, Guo Y, et al. Similar gene structure of two Sox9a genes and their expression patterns during gonadal differentiation in a teleost fish, rice field eel (*Monopterus albus*). *Mol Reprod Dev*. 2003;66(3):211-217.
41. Paul-Prasanth B, Nakamura M, Nagahama Y. Sex Determination in Fishes. In: Norris DO, Lopez KH, eds. *Hormones and Reproduction of Vertebrates*. Cambridge, MA: Academic Press; 2011:1-14.