Switching mechanism of the master sex-determining gene in medaka fishes

Taijun Myosho

Doctoral Program in Environmental Science and Technology

Graduate School of Science and Technology, Niigata University,

2012

Contents

ABSTRACT

Three sex-determining genes, *SRY* (mammals), *Dmy* (medaka, *Oryzias latipes*) and *DM-W* (*Xenopus laevis*), have been identified to date in vertebrates. However, how and why a new sex-determining gene appears remains unknown, as do the switching mechanisms of the master sex-determining gene. In this study, I identified two novel sex determining gene, *GsdfY* in *O. luzonensis* and *GsdfbY* in *O. pectoralis*, from which the appearance process differs although it evolved from same ancestral gene.

First, I used positional cloning to search for the sex-determining gene in *O. luzonensis* and found that $Gsdf^{Y}$ (*gonadal soma derived growth factor* on the Y chromosome) has replaced *Dmy* as the master sex-determining gene in this species. I found that $Gsdf^{Y}$ showed high expression specifically in males during sex differentiation. Furthermore, the presence of a genomic fragment that included $Gsdf^{Y}$ converts XX individual into fertile XX males. Luciferase assays demonstrated that the upstream sequence of $Gsdf^{Y}$ contributes to the male-specific high expression. *Gsdf* is downstream of *Dmy* in the sex-determining cascade of *O. latipes*, suggesting that emergence of the *Dmy*-independent *Gsdf* allele led to the appearance of this novel sex-determining gene in *O. luzonensis*.

Next, by identifying the sex determination system and the sex chromosomes in seven *Oryzias* species, I examined a possibility whether *Gsdf* is the SD gene in these species. Linkage analysis demonstrated that sex determination system of six species were XX-XY and that the sex chromosomes of these species as follows; LG 4 in *O.sp* (Laos), LG 24 in *O. celebensis*, LG 24 in *O. woworae*, LG 24 in *O. matanenssis*, LG 10 in *O. profundicola* and LG 10 in *O. marumoratus*. Since sex chromosomes homologous to LG 12 have not been found, *Gsdf* homologue cannot be the SD gene in these species.

Finally, in *O. pectoralis*, I found two copies of *Gsdf*, one was located on LG 12 (*Gsdf* orthologue), another on LG 1. *Gsdf* on LG 1 showed male-specific inheritance. Since *Gsdf* is mapped on LG 12 in *O. latipes* and its close relatives, I regarded *Gsdf* on LG 1 was a duplicated

copy of *Gsdf* (*GsdfbY*). The Sex-determining region including *GsdfbY* was mappped on the long arm near the telomere of LG 1. This region is equivalent to about 500 kbp in the *O. latipes* genome. I examined the expression profiles of *Gsdf* and *GsdfbY* during the sex differentiation stages. RT-PCR revealed that in XY embryo, *GsdfbY* showed high expression from 2 days before hatching (dbh) to the hatching day just before the period when the sex difference in the gonad became manifest. Expression of *Gsdf* was detected at very low levels at 2 dbh and became a similar high level to *GsdfbY* at the hatching day. In XX embryo, expression of *Gsdf* was low from 2 dbh to the hatching day. These results suggest that *GsdfbY* has acquired a novel earlier expression pattern during sex-determining period. Finally, I performed gain-of-function experiment. Sex-reversed XX males were obtained by microinjection of a genomic fragment that included *GsdfbY*. These results demonstrated that *GsdfbY* is the prime candidate for the sex-determining gene of *O. pectoralis*.

Part 1: Change in a regulatory element of *Gsdf* leads to the creation of new sex chromosome in *Oryzias luzonensis*

Introduction

In most vertebrates, sex is determined genetically. Mammals and birds with cytogenetically well-differentiated sex chromosomes have sex determination systems that differ between the taxonomic classes but not within them (Solari 1994). In mammals, for example, the sex-determining gene *SRY/Sry* on the Y chromosome has a universal role in sex determination (Gubbay *et al.* 1990; Sinclair *et al.* 1990; Koopman *et al.* 1991; Foster *et al.* 1992). By contrast, some fish groups, such as salmonids, sticklebacks, and *Oryzias* fishes, have sex chromosomes that differ among closely related species (Devlin and Nagahama 2002; Woram *et al.* 2003; Takehana *et al.* 2007a; Ross *et al.* 2009).

A DM-domain gene, *Dmy,* was the first sex-determining (SD) gene identified in a non-mammalian vertebrate, the fish medaka *Oryzias latipes* (Matsuda *et al.* 2002, 2007). In this species, the term Y chromosome is employed to refer to a recombining chromosome that carries the male-determining gene *Dmy*, and X is used for the homologous chromosome; these are not a heteromorphic pair. This gene is conserved among all wild populations of *O. latipes* examined to date (Shinomiya *et al.* 2004). The closely related species *O. curvinotus* also has *Dmy* on its Y chromosome, which is orthologous to the *O. latipes* Y chromosome (Matsuda *et al.* 2003). However, *Dmy* has not been detected in any other type of fish, including other *Oryzias* fishes (Kondo *et al.* 2003). Analysis of the Y-specific region of the *O. latipes* sex chromosome has demonstrated that *Dmy* arose from duplication of the autosomal *Dmrt1* gene (Nanda *et al.* 2002; Kondo *et al.* 2006). This *Dmrt1* duplication is estimated to have occurred within the last 10 million years in a common ancestor of *O. latipes*, *O. curvinotus* and *O. luzonensis.* In *O. luzonensis*, however, no functional duplicated copy of *Dmrt1* has been detected (Kondo *et al.* 2003) (Figure 6A).

Oryzias luzonensis possesses an XX–XY system, which is homologous to an autosomal linkage group (LG 12) in *O. latipes* (Hamaguchi *et al.* 2004; Tanaka *et al.* 2007). This species, like *O. latipes*, has homomorphic sex chromosomes without recombination suppression between

them. This supports the hypothesis that *Dmy* lost its SD function and disappeared after a new SD gene appeared in *O. luzonensis*. *Oryzias luzonensis* may, therefore, be very informative for studying the evolution of master SD genes and of the early stages of sex-chromosome differentiation.

Materials and Methods

Fish

Oryzias luzonensis was collected by M. J. Formacion and H. Uwa at Solsona, Ilocos Norte, Luzon, The Philippines, in 1982, and has been maintained as a closed colony (Formacion and Uwa 1985). In the d-rR strain, the wild-type allele R of r (a sex-linked pigment gene) is located on the Y chromosome. The body of the female was white, whereas that of the male was orange-red (Hyodo-Taguchi and Sakaizumi 1993). These fishes were supplied by a subcentre (Niigata University) of the National BioResource Project (medaka) supported by the Ministry of Education, Culture, Sports, Science, and Technology of Japan.

DNA and RNA extraction

Total RNA and genomic DNA were extracted from each hatched embryo after homogenization in a 1.5-ml tube with 350 ml RLT buffer supplied with the RNeasy Mini Kit (Qiagen). The homogenized lysates were centrifuged and supernatants were used for RNA extraction with the RNeasy Mini Kit and the RNase-Free DNase set protocol (Qiagen). Precipitated material was used for DNA extraction by using the DNeasy tissue Kit (Qiagen) according to the manufacturer's protocol.

Genotypic sexing

Genotyping of the SD region was conducted by using genomic PCR of fin clip DNA. Genomic PCR was performed by using four sets of primers designed in the SD region (Table 1). PCR conditions were as follows: 5 min at 95°C, followed by 35 cycles of 20 s at 95°C, 30 s at 58°C, and 1 min at 72°C, followed by 5 min at 72°C.

Construction of genomic libraries and chromosome walking

A BAC genomic library (LMB1) was constructed from cultured cells that were derived from an embryo that was produced by mating an XX female with a sex-reversed XX male. The cultured cells were embedded in the agarose gel, and then, partially digested with *Sac*I. The fragments in

the size range of 150–225 kb were selected. The size-selected DNA fragments were ligated to pKS145 vector and used to transform DH10B. A total of 36,864 BAC clones were picked and arrayed to 384 microtiter plates and then 3D DNA pools were constructed for clone screening by PCR. Chromosome walking started at OluX2-8. Inserted end fragments of positive BAC clones were amplified by using vectorette PCR and then used to assemble positive clones (Arnold and Hodgson 1991). An amplified end fragment at the far end of the SD side was used in the subsequent screening of the BAC library.

A fosmid library was constructed from a YY individual obtained by hybridizing an XY male and a sex-reversed XY female with pCC1FOS vecter (Epicentre Technology) following the manufacturer's protocol. We made 48 fosmid clone pools, which contained 2000 fosmid clones per tube. Seven fosmid clones that correspond to the SD region in the X chromosome were isolated by the PCR screening method, which decreased the number of fosmid clones in the pool by three PCR steps.

Shotgun sequencing

BAC DNA was hydrodynamically sheared to average sizes of 1.5 and 4.5 kb, and the DNA was ligated into a pUC18 vector. We sequenced each BAC to have a genome coverage of 13 by using dye-terminator chemistry. Individual BACs were assembled from the shotgun sequences by using phred Version 0.000925.c, crossmatch Version 0.990319 and phrap Version 0.990319 (Codon Code), as well as PCP Version 2.1.6 and Cap4 Version 2.1.6 (Paracel). The gaps in each BAC were closed with a combination of BAC walking, directed PCR, and resequencing of individual clones. The sequence of the fosmid clones was determined by using the same method.

Phylogenetic analysis

The predicted mature domain of GSDF, additional members of the transforming growth factor (TGF)-β superfamily, and another human cystine knot cytokine (brain-derived neurotrophic factor [BDNF]) were aligned by using Molecular Evolutionary Genetics Analysis (MEGA) Ver. 3.1 software (http://www.megasoftware.net). The GenBank accession numbers of the aligned

amino acid sequences are as follows: human TGF-β1, NP_000651.3; mouse TGF-β1, NP_035707.1; zebrafish TGF-β1a, NP_878293.1; rainbow trout TGF-β1, CAA67685.1; medaka TGF-β1, ENSORLP00000001563; human growth-differentiation factor 5, NP_000548.1; mouse GDF5, NP_032135.2; zebrafish GDF5, CAA72733.1; medaka GDF5, ENSORLP00000003714 human inhibin α(INHα), NP_002182.1; mouse INHα, NP_034694.3; zebrafish INHα, CAK11253.1; rainbow trout INHα, BAB19272.1; medaka INHα, ENSORLP00000002713; human anti-Mullerian hormone (AMH/MIS), NP_000470.2; mouse AMH, NP_031471.2; zebrafish AMH, NP_001007780.1; medaka Amh, NP_001098198.1; zebrafish GSDF, NP_001108140.1; rainbow trout GSDF, ABF48201.1; medaka GSDF, NP_001171213.1; human BDNF, NP_001137277.1; mouse BDNF, NP_031566.4, zebrafish BDNF, NP_571670.2; and rainbow trout BDNF, ACY54685.1.

RT-PCR

RT–PCR was performed by using a One Step RT–PCR kit (Qiagen). Aliquots (20 ng) of total RNA samples were used as templates in 25-μl reaction volumes.

The PCR conditions were: 30 min at 55°C; 15 min at 95°C; cycles of 20 s at 96°C, 30 s at 55°C, and 60 s at 72°C; and 5 min at 72°C. The number of cycles for each gene was adjusted to be within the linear range of amplification, specifically 35 cycles for *predicted genes* (*PGs*) and 24 cycles for β-actin. Specific primers for PGs were designed in each exon (Table 1).

Real-time PCR

Expression levels were quantified by using RNA from the body trunks of fry from -2 to 10 days after hatching (dah). Concentrations were adjusted to a total of 5 ng for each real-time assay. Using base substitutions between $Gsdf^X$ and $Gsdf^Y$, primers were designed to examine the expression profiles of $Gsdf^X$ and $Gsdf^Y$ (Supplementary Table 1). Quantitative gene expression analysis was performed on an ABI PRISM 7000 (ABI) using a One Step SYBR Prime-Script RT-PCR kit (Takara Bio). The PCR conditions were 5 min at 42°C, 10 s at 95°C, then 40 cycles of 5 s at 95°C and 30 s at 65°C.

In situ hybridization

Fry at 5 dah and adult gonads were fixed in 4% paraformaldehyde in PBS at 4°C overnight. Digoxigenin (DIG)-labelled RNA probes were generated by *in vitro* transcription with a DIG RNA labelling kit (Roche, Basalm, Switzerland) from a *Gsdf*^{Y} cDNA plasmid. Sections were deparaffinised, hydrated, treated with proteinase K (10 μg/ml) at 37°C for 5 min, and hybridized with the DIG-labelled antisense RNA probes at 60°C for 18–24 h. Hybridization signals were detected by using an alkaline phosphatase-conjugated anti-DIG antibody (Roche) with NBT/BCIP (Roche) as the chromogen.

Transgenic constructs

I made two constructs to obtain transgenic lines. First, I inserted a fluorescent reporter gene into the fosmid clone containing *Gsdf*^{Y} (OluFY3-1); this was a crystal-lens-specific crystalline-γM2 promoter driving red fluorescent protein (RFP). The fluorescent reporter was inserted into the fosmid vector by using a Quick and Easy BAC Modification kit (Gene Bridges GmbH, Dresden, Germany), which relies on homologous recombination in *Escherichia coli*. This construct (Construct 1) contained 3.5 kb of the coding sequence, 20 kb of the upstream region, and 13 kb of the downstream region. By removing other PGs, I obtained the second construct (Construct 2), which contained only *Gsdf*^{Y}, by using In-Fusion (Takara Bio) methods. I amplified two fragments: a 7.3-kb genomic sequence containing 3.5 kb of the coding region, 1.8 kb of the upstream region and 2 kb of the downstream region, and a fosmid vector sequence. I then cloned the *Gsdf*^{Y} fragments into vectors by using an In-Fusion Advantage PCR Cloning Kit (Takara Bio). Reporter gene integration was similarly achieved.

Microinjection

Fertilized eggs were collected within 20 min of spawning and were microinjected. I used DNA at 10 ng/μl in Yamamoto's solution (133 mM NaCl, 2.7 mM KCl, 2.1 mM $CaCl_2$, 0.2 mM $NaHCO_3$; pH 7.3). The injected eggs were incubated at 27°C until hatching.

Luciferase assay

The *Gsdf*^{Y} and *Gsdf*^{X} luciferase reporter plasmids (Luc Y and Luc X) were generated by cloning the 3-kb upstream region of each *Gsdf* into the vector (pGL.4.14; Promega) by using the In-fusion Advantage PCR Cloning Kit (Takara Bio) with designed primers (Table 1). Modified reporter plasmids (Luc 1–6) were generated on the basis of Luc Y and Luc X by using In-Fusion methods. HEK293 cells were cultured at 37°C in DMEM (Invitrogen) supplemented with 10% foetal bovine serum (Gibco, U.S.A); 2.5×10^4 cells were plated in each well of 96-well plates 24 h before transfection. The cells were transfected with 100 ng of the *Gsdf*^{Y} luciferase reporter, *Gsdf*^{X} luciferase reporter or modified luciferase reporters, and 100 ng of TK-*Renilla* luciferase plasmid (pGL.4.79; Promega) with Lipofectamine 2000 reagent (Invitrogen) and opti-MEM (Invitrogen). After 40 h, luciferase assay was performed with the Dual-Glo Luciferase Reporter Assay System (Promega, Madison, WI) and a Wallac 1420 ARVO-SX multilabel counter (Perkin Elmer). The levels of firefly luciferase activity were normalized against *Renilla* luciferase activity. At least three independent experiments were performed.

Results and Discussion

Nine genes were predicted in the SD region

The SD region of *O. luzonensis* maps between *eyeless* and 171M23F on LG 12 (Tanaka *et al.* 2007). I performed further linkage analysis and obtained two male recombinants for this region. One male had a recombination breakpoint between OluX2-8 and OluX2-25, and the other had a breakpoint between OluX3-34 and OluX4-6, refining the SD region to between OluX2-8 and OluX4-6 (Figure 1A).

I constructed a BAC library of an XX fish and a fosmid library of a YY fish, and I made physical maps of the SD region of the X and Y chromosomes. This region was covered with two BAC clones (OluBXKN2 and OluBXKN1) on the X chromosome and with seven fosmid clones (OluFY13-1, OluFY24-1, OluFY18-1, OluFY3-1, OluFY8-1, OluFY7-1, and OluFY29-1) on the Y chromosome (Figure 1A). The entire nucleotide sequence was determined by using shotgun sequencing, except for a repetitive region in OluFY3-1 and OluBXKN1. Restriction analysis of both clones demonstrated that the length of the repetitive region was the same for both chromosomes (data not shown). The SD region is about 180 kb for the X and Y chromosomes, and both chromosomes exhibit high sequence identity with no large deletions or insertions. The gene-prediction program Genscan identified nine genes in this region; all are found on both the X and Y chromosomes (Figure 1A).

$Gsdf^Y$ is responsible to male specific high expression during sex differentiation

To examine whether the *predicted genes* (*PGs*) are expressed during sexual differentiation, I performed RT-PCR for each *PG*. The first difference in germ cell number is seen 3 dah in *O. luzonensis* (Nakamoto *et al.* 2009). Given that expression of the SD gene *Dmy* precedes the first morphological gonadal difference in *O. latipes*, the SD gene of *O. luzonensis* should function sometime before 3 dah. RT-PCR detected the expression of seven of the nine genes at 0 dah (Figure 1B). Only one gene, *PG5*, shows higher expression in XY embryos than in XX embryos.

I determined the full-length mRNA sequence of *PG5* on the X and Y chromosomes using 5' and 3' RACE. The longest open reading frame (ORF) spanned five exons and encodes a putative protein of 215 amino acids (Figure 2, A and B). The N-terminal regions are rich in hydrophobic amino acid residues and are followed by a potential cleavage site comprising Ala and Phe (amino acid residues 19 and 20; Figure 2B). Phylogenetic analysis of the mature domain of the cystine-knot cytokines revealed that the *PG5* sequence is found in the same clade as *Gsdf*, which is a member of the TGF-β superfamily (Figure 2C). When I compared *Gsdf* on the X chromosome ($Gsdf^X$) with that on the Y chromosome ($Gsdf^Y$), I found 12 base substitutions in the full-length mRNA, including two synonymous substitutions in the ORF; however, the amino acid sequences of $Gsdf^X$ and $Gsdf^Y$ are the same.

Using the base substitutions between $Gsdf^X$ and $Gsdf^Y$, I examined the expression profiles of $Gsdf^X$ and $Gsdf^Y$ using real-time PCR. Expression of *Gsdf* was higher in XY embryos than in XX embryos from 2 days before hatching (dbh) to 10 dah (Figure 3A). In the XY embryo, $Gsdf^Y$ expression was higher than $Gsdf^X$ expression at 0 dah, whereas it was similar to $Gsdf^X$ expression at 5 and 10 dah.

At 5 dah in the developing gonads, supporting cells surrounding the germ cells expressed *Gsdf* in both XY and XX embryos, although *Gsdf* expression was much higher in XY embryos (Figure 3, B and C). In the adult testis, *Gsdf* was detected in the Sertoli cells around type A spermatogonia (Figure 3D); in the adult ovary, *Gsdf* was expressed in the granulosa cells surrounding well-developed oocytes (Figure 3E).

$Gsdf^Y$ induced fertile XX male in O. luzonensis

I performed overexpression experiments using a $Gsdf^Y$ genomic clone. First, I used a fosmid clone (OluFY3-1) that spans 20 kb upstream and 13 kb downstream of $Gsdf^Y$. Construct 1, containing $Gsdf^Y$, *PG3,* and *PG4*, was injected into one-cell-stage embryos of *O. luzonensis* (Figure 4A). In generation zero (G0), I obtained 54 adult fish with the transgene, one of which

was a sex-reversed XX male (Table 2). I mated the XX male with a normal female to obtain G1 progeny, and G2 progeny were obtained from an XX male of the G1 progeny. All fish bearing the transgene developed as males in both the G1 and G2 progeny, whereas all fish without the transgene developed as females. Consequently, I established a transgenic strain (Strain 1) whose sex was determined by the transgene Construct 1. Next, I made a construct (Construct 2) that contained 3.5 kb of *Gsdf*^{Y}, as well as 1.8 kb of its upstream region and 2 kb its downstream region, but no other predicted genes (Figure 4B). As with the previous transgenic experiment, I established a strain (Strain 2) whose sex was determined by the transgene (Table 2). To confirm the mRNA expression of both strains, I examined embryos at 0 dah by using real-time PCR. XX embryos carrying the transgene expressed higher levels of *Gsdf* than did XX embryos without the transgene in both strains (data not shown).

Gsdf^{Y}*-specific mutations involved in high expression*

I hypothesized that there were *Gsdf*^{Y} sequences specific for the high expression within Construct 2. According to Gautier *et al.* (2011), the *Gsdf* proximal gene promoter harbours evolutionarily conserved cis-regulatory motifs among fish species. To find these sequences, I compared 1.8 kb upstream and 2.0 kb downstream of *Gsdf*^{Y} with those of *Gsdf*^{X} and *Gsdf* in *O. latipes*. I found 13 substitutions between the X and Y in the upstream region, 9 of them *Gsdf*^{Y}-specific mutations, and 31 between the X and Y in the downstream region (including 20 *Gsdf*^{Y}-specific mutations) (Figure 5, A and B). I used a luciferase assay to assess the 9 *Gsdf*^{Y}-specific upstream mutation sites. The *Gsdf*^{Y} reporter plasmid with all mutations in the Y-type allele (Luc Y) showed higher luciferase activity than the *Gsdf*^{X} reporter plasmid (Luc X) (Figure 5C). Luciferase activity was significantly decreased in recombinant constructs Luc 3, 4, 5, and 6, whereas two constructs (Luc 1 and 2) showed high luciferase activity, equal to that of Luc Y. Because the constructs yielding high expression all had Y-type mutations 1, 2 or 3-6 in addition to mutations 6-9, I conclude that Y-type mutations 6-9 are necessary for the high expression and that either 1, 2 or 3-6 Y-type

mutations are also required.

Gsdf^{Y} induced sex-reversal in O. latipes

In *O. latipes*, the orthologue of *Gsdf$^{X/Y}$* is located on an autosome (LG 12). *Gsdf* in XY fish shows significantly higher expression levels compared with that in XX fish during sex differentiation, suggesting that expression levels of *Gsdf* is directly or indirectly controlled by *Dmy* (Shibata *et al.* 2010). To examine whether *Dmy*-independent expression of *Gsdf^{Y}* induces sex reversal in *O. latipes*, we injected Construct 1 into one-cell-stage embryos of the d-rR strain of *O. latipes*. Consequently, I established an *O. latipes* strain (Strain 3) whose sex was determined by *Gsdf^{Y}* from *O. luzonensis* (Table 2). Real-time PCR revealed that this strain showed high expression of *Gsdf^{Y}* in an XX embryo at 0 dah (data not shown).

The evolutionary process leading to a novel SD gene

Present results strongly suggest that *Gsdf^{Y}* is the SD gene in *O. luzonensis* and represents a new SD gene in vertebrates. Three SD genes, *SRY*, *Dmy,* and *DM-W,* had been identified to date (Yoshimoto *et al.* 2008). These genes encode transcription factors, whereas *Gsdf* encodes a secretory protein belonging to the TGF-β superfamily and was originally identified as a factor controlling the proliferation of primordial germ cells and spermatogonia in rainbow trout (Sawatari *et al.* 2007). Since homologous sequences with high similarity to *Gsdf* have not been found in non-piscine species, *Gsdf* is likely unique to teleosts. The three SD genes are not allelic. *Dmy* and *DM-W* might have emerged by duplication of *DMRT1* and are located on the Y and W chromosomes, respectively (Yoshimoto *et al.* 2008). *SRY* is believed to have arisen from *SOX3* 130–170 million years ago (mya), suggesting that it was formerly allelic to *SOX3* (Marshall-Graves 2002). Although *Gsdf^{Y}* appeared in the same way as *SRY,* it remains allelic to *Gsdf^{X}* likely because of its more recent origin (within 5 mya) (Tanaka *et al.* 2007).

Expression analysis and the reporter assay suggest that *cis*-regulatory sequences of *Gsdf^{Y}* are

involved in higher expression of the gene in males (Figures, 3 and 5). *In silico* analysis of the regulatory motif suggested that the sequences containing 6–9 mutations are a steroidogenic factor 1 (SF1) binding site (i.e., SF1 can bind upstream of $Gsdf^X$ but not of $Gsdf^Y$). $Gsdf^Y$ may have evolved from ancestral *Gsdf* by acquiring high expression during an earlier stage of sex determination via a change in the SF1 binding site. In *O. latipes*, *Gsdf* shows high expression specifically in males during sex differentiation (Shibata *et al.* 2010). Since *Dmy* determines sex in *O. latipes*, the sex-specific high expression of *Gsdf* should be triggered by *Dmy* in this species. However, the transgene expressing $Gsdf^Y$ in *O. latipes* is sufficient to induce fertile XX males (Table 2). During *O. luzonensis* sex differentiation, other genes, such as *Sox9a2*, *Dmrt1*, and *Foxl2*, which are presumably downstream of *Gsdf*, show expression patterns similar to those in *O. latipes* (Nakamoto *et al.* 2009). Taken together, these results imply that *O. luzonensis* and *O. latipes* share a common sex differentiation pathway downstream of *Gsdf* and that, if high *Gsdf* expression can be achieved during sex differentiation, then the XX embryo will develop as a male without *Dmy*.

Wilkins proposed that sex-determination pathways grow by the successive addition of upstream control elements to an ancient conserved downstream module (Wilkins 1995). For example, in *Drosophila*, *double sex* determined the sex in the ancestral state. Then, sex-related genes were added in succession upstream of *double sex* to give the present SD cascade (Pomiankowski *et al.* 2004). In *O. luzonensis*, the scenario is somewhat different (Figure 6, A and B). *Gsdf* was downstream of *Dmy* in the ancestor of *O. luzonensis*. Mutations involved in high expression of *Gsdf* without the *Dmy* signal then accumulated, until the expression exceeded the threshold which determines male development, leading to the new SD gene $Gsdf^Y$. If these mutations induced high expression independently of *Dmy*, individuals with either *Dmy* or $Gsdf^Y$ would develop as males, and those with neither *Dmy* nor $Gsdf^Y$ would develop as females. Since mating occurs only between males (with either *Dmy* or $Gsdf^Y$) and females (with neither *Dmy* nor $Gsdf^Y$), the sex ratio did not become skewed toward males. In this population, two SD genes

(*Dmy* and $Gsdf^{Y}$) could temporarily coexist. Finally, if the chromosome with *Dmy* is lost from this population, the master SD gene *Dmy* would be replaced by $Gsdf^{Y}$. I conclude that SD cascades can also evolve by expression of a downstream gene becoming independent of an existing sex-determining gene, and usurping control of the downstream cascade.

Part 2: A duplicated copy of *Gsdf* on Y chromosome is the sex-determining gene in *Oryzias pectoralis*

Introduction

As I discussed in part 1, in *O. luzoensis*, the SD gene $Gsdf^{Y}$ evolved from *Gsdf* by an appearance of a neofunctionalized allele without duplication. On the other hand, the SD gene *Dmy* that only found in *O. latipes* and *O. curvinotus* may have evolved from *Dmrt1* as a case of neofunctionalization with duplication (Kondo *et al.* 2006). These facts indicate that a novel SD gene can arise by neofunctionalization with or without duplication. In *Oryzias* fishes, more than 20 extant species are recognized and can be divided into three phylogenetic groups based on nuclear *tyrosinase* and mitochondrial 12S and 16S rRNA genes (*latipes*, *javanicus* and *celebensis* groups) (Figure 13) (Takehana *et al.* 2005). The sex chromosomes of seven species have been identified; *O. latipes* (LG 1), *O. curvinotus* (LG 1), *O. luzonensis* (LG 12), *O. minutillus* (LG 8), *O. dancena* (LG 10), *O. hubbsi* (LG 5) and *O, javanicus* (LG 16) (Nagai *et al.* 2008; Takehana *et al.* 2007a, 2007b, 2008). The sex chromosomes homologous to LG 9 have not been found, suggesting that there is no case from which *Dmrt1* evolved into the SD gene by neofunctionalization without duplication. In *Xenopus laevis*, like *O. latipes*, a duplicated copy of *dmrt1* acquired the SD function. This result demonstrated that the SD gene independently evolved twice from *Dmrt1* in vertebrates.

In this part, by identifying the sex chromosomes of six *Oryzias* species, *O. sp (Laos), O. celebensis, O. woworae, O. matanensis, O. profundicola and O. marmoratus*, I examined the possibility that the SD gene could be derived by neofunctionalization with or without duplication. Although I could not find additional species that have the sex chromosome homologous to LG 12 of *O. latipes*, a duplicated copy of *Gsdf*, *GsdfbY*, was demonstrated to be the SD gene in *O. pectoralis*.

Materials and Methods

Fish and genetic cross

Oryzias pectoralis, *O. sp.* (Laos), *O. celebensis*, *O. woworae*, *O. matanensis*, *O. profundicola* and *O. marmoratus* have been maintained as a closed colony in Niigata University. In all fishes, I obtained more than 40 F1 progeny by crossing a female with a male.

DNA and RNA extraction

Total RNA and genomic DNA were extracted from each hatched embryo after homogenization in a 1.5-ml tube with 350 ml RLT buffer supplied with the RNeasy Mini Kit (Qiagen). The homogenized lysates were centrifuged and supernatants were used for RNA extraction with the RNeasy Mini Kit and the RNase-Free DNase set protocol (Qiagen). Precipitated material was used for DNA extraction by using the DNeasy tissue Kit (Qiagen) according to the manufacturer's protocol.

Sex linkage analysis

By crossing a female with a male, I obtained F1 progeny for linkage analysis. Phenotypic sex was determined by secondary sex characteristics of adult fish, namely the shapes of dorsal and anal fins. PCR amplification was performed as follows: 35 cycles at 95°C for 30 sec, 55°C for 30 sec, and 72°C for 2 min. Each amplified marker was genotyped by indel, heteroduplex or RFLPs. Primers for amplifying ESTs were listed in Table 3.

Mapping of GsdfbY and Gsdfa

Genotyping of the SD region was conducted by using genomic PCR of fin clip DNA. Genomic PCR was performed by using four sets of primers designed in the SD region (Table 3). PCR conditions were as follows: 5 min at 95°C, followed by 35 cycles of 20 s at 95°C, 30 s at 58°C, and 1 min at 72°C, followed by 5 min at 72°C.

Construction of a genomic library and chromosome walking

A fosmid library was constructed from an XY individual with pCC1FOS vecter (Epicentre Technology) following the manufacturer's protocol. I made 48 fosmid clone pools, which

contained 2000 fosmid clones per tube. Seven fosmid clones were isolated by the PCR screening method by three PCR steps.

RT-PCR

RT–PCR was performed by using a One Step RT–PCR kit (Qiagen). Aliquots (20 ng) of total RNA samples were used as templates in 25 μl reaction volumes.

The PCR conditions were: 30 min at 55°C; 15 min at 95°C; cycles of 20 s at 96°C, 30 s at 55°C, and 60 s at 72°C; and 5 min at 72°C. The number of cycles for each gene was adjusted to be within the linear range of amplification, specifically 32 cycles for *GsdfbY* and *Gsdfa*, and 24 cycles for EF1alpha. Specific primers were designed in each exon (Table 3).

Transgenic constructs

I amplified two fragments: a 7.5 kb genomic sequence containing 4.5 kb of *GsdfbY* coding region, 2.8 kb of the upstream region and 0.2 kb of the downstream region, and a fosmid vector sequence. I then cloned the *GsdfbY* fragments into vectors by using an In-Fusion HD PCR Cloning Kit (Takara Bio). Finally, I inserted a fluorescent reporter gene into the fosmid clone containing $Gsdf^{Y}$ (OluFY3-1); this was a crystal-lens-specific crystalline-γM2 promoter driving green fluorescent protein (RFP). The fluorescent reporter was inserted into the fosmid vector using an In-Fusion HD PCR Cloning Kit (Takara Bio).

Microinjection

Fertilized eggs were collected within 20 min of spawning and were microinjected. I used DNA at 10 ng/μl in Yamamoto's solution (133 mM NaCl, 2.7 mM KCl, 2.1 mM $CaCl_2$, 0.2 mM $NaHCO_3$; pH 7.3). The injected eggs were incubated at 27°C until hatching.

Phylogenetic analysis

The mRNA sequences of *Gsdf* in *Oryzias* fishes (*O. latipes, O. curvinotus*, *O. luzonensis*, *O. mekogensis*, *O. songkramensis*, *O. dancena*, *O. hubbsi*, *O. celebensis* and *O. nigrimas*) were obtained by RT-PCR using RNAs extracted from ovary. Specific primers were designed in 1st exon and 5th exon (Table 3). GSDFbY and GSDFa in *O. pectoralis* and GSDF in other *Oryzias*

fishes were aligned by using Molecular Evolutionary Genetics Analysis (MEGA) Ver. 4 software (http://www.megasoftware.net).

Results and Discussion

Identification of sex chromosomes

First, I investigated whether sex chromosomes are homologous to LG 1 in *O. sp* (Laos), most closely related to *O. curvinotus* in the *latipes* group. Ten EST markers on LG 1 that including *Dmy* locus in *O. latipes* and *O. curvinotus* were not linked to sex, suggesting that *O. sp* (Laos) did not have *Dmy* on LG 1. To isolate sex-linked markers, I searched for polymorphic EST markers between the parents, and genotyped F1 progeny. I screened 183 ESTs, and identified 31 polymorphic markers. One of these markers, 13.63 on LG 4 showed male-specific inheritance in the F1 progeny. To isolate additional sex-linked markers, I examined other ESTs on LG 4, and identified a total of twelve sex-linked markers, demonstrating that, sex determination system is XX-XY type and the sex chromosomes are homologous to LG 4 in *O. sp*. (Laos) (Figure 7). The SD region was located between OLc0504d and 31.31. This region is equivalent to about 25 Mb in *O. laipes*, suggesting that this species may have recombination suppression between the sex chromosomes.

Second, I focused on *O. celebensis* in the *celebensis* group. To isolate sex-linked markers, I screened 479 ESTs, and identified 42 polymorphic markers between the parents. I found that an *O. latipes* EST (SOX7) yielded male specific banding patterns, suggesting that *O. celebensis* had an XX-XY sex determination system. I identified additional three sex-linked markers and constructed a sex linkage map. Four sex-linked markers were mapped on LG 24 in *O. latipes* (Figure 8). The SD region was located between HNRNPU and PCMT1. On the basis of the draft genome sequence, this region was calculated to be about 4.9 Mb in *O. latipes*. Then, I examined whether sex chromosomes were homologous to LG 24 in other four species in the *celebensis* group. In *O. woworae* and *O. matanensis,* like *O. celebensis*, *O. latipes* EST (SOX7) yielded male specific banding patterns, suggesting that the sex chromosomes were homologous to LG 24 in *O. latipes* and that these species have an XX-XY system (Figures 9 and 10). By contrast, in *O. profundicola* and *O. marmoratus*, the EST maker (SOX7) did not link to sex, suggesting that the

sex chromosomes are not homologous to LG 24.

Finally, to isolate sex-linked markers in *O. profundicola*, I screened 272 EST markers except for markers on LG 24, and identified 25 polymorphic markers. One of these markers, MF01SSA75C05 showed male-specific inheritance in the F1 progeny. This result revealed that *O. profundicola* had an XX-XY sex determination system, and that the sex chromosomes were homologous to LG 10 in *O. latipes*. Although I identified additional four sex-linked markers that is located from 3.6 Mb to 25 Mb, SD region could not be narrowed (Figure 11). *Oryzias* fishes in *celebensis* group has some large metacentric pairs which assumed to have been formed by centric fusion (Parenti 2008), suggesting that if LG 10 is fused chromosome, another LG in addition to LG 10 function as sex chromosomes in *O. profundicola*. Although I can't perform linkage analysis across all LG, I revealed that EST markers on LG 12 were not linked to sex. Since in *O. profundicola* sex chromosomes were homologous to LG 10 in *O. latipes*, I examined whether EST makers on LG 10 could be linked to sex in *O. marmoratus*. MF01SSA075C05 showed male-specific inheritance in the F1 progeny, suggesting that that *O. marmoratus* had an XX-XY sex determination system, and that the sex chromosomes are homologous to LG 10 in *O. latipes* (Figure 12).

Although I identified sex determination system and the sex chromosomes in six species, sex chromosomes homologous to LG 12 have not been found (Figure 13). This result suggests that the *Gsdf* orthologue cannot be the SD gene in these species. However, given that duplicated copy of *Dmrt1* has evolved into SD gene like *Dmy* in *O. latipes* and *DM-W* in *X. laevis*, there remains the possibility that a duplicated copy of *Gsdf* functions as the SD gene in these species.

Gsdf showed male specific inheritance in O. pectoralis

To examine whether *Gsdf* is linked to sex in *O. pectoralis*, I performed sex linkage analysis using some primer sets that were designed on *Gsdf* exons. One of these primer sets (Gsdf-f and Gsdf-r) amplified two bands in the male parent and one band in the female. Lower band in the

male parents showed male-specific inheritance in the F1 progeny, suggesting that *O. pectoralis* has an XX-XY system (Figure 14). Sequences of two bands in the parents showed high homology to *Gsdf* in *O. latipes* and a deletion of about 100 bp was found in the lower band. The upper band in the parents had a single nucleotide polymorphism. These results suggest that all bands are *Gsdf* product, and that upper bands can be autosomal, and that the lower band a duplicated copy of *Gsdf*. To confirm this inference, I conducted linkagae analysis of two copies of *Gsdf*. The upper bands was mapped on LG 12 (*Gsdf* orthologue), another on the long arm near the telomere of LG 1 (Figure 15, A and B). Since *Gsdf* is mapped on LG 12 in *O. latipes* and its close relatives, I regarded *Gsdf* on LG1 as a duplicated copy of *Gsdf* (*GsdfbY*).

I performed further linkage analysis in F2 and F3 progeny (N=806), refining the SD region to between PMM2 and HOGA1 (Figure 16). This region is equivalent to about 500 kb in *O. laipes*. I constructed a fosmid library of an XY fish, and made physical maps around *GsdfbY* in the SD region of the X and Y chromosomes. I isolated three fosmid clones (Opec20-1, Opec22-1 and Opec25-1) on the X chromosome, three fosmid clones (Opec1-1, Opec3-1 and Opec8-1) on the Y chromosome and a fosmid clone including *Gsdf* on LG 12 (Figure 16). The nucleotide sequence of each clone was determined by subcloning and primerwalking. These sequences revealed that *GsdfbY* was inserted between *Gene 1* and *Gene 4* on LG 1, and that homologous region between Y chromosome and LG 12 is only 8 kb including *GsdfbY* and *Gene 5*.

GsdfbY has acquired an earlier high expression than _Gsdfa_ during sex differentiation

I performed RT-PCR for *GsdfbY* during sexual differentiation. A *Hinf*I site polymorphism was found between *Gsdf* and *GsdfbY*. RT-PCR products of *GsdfbY* can be distinguished from *Gsdf* after *Hinf*I digestion. RT-PCR revealed that in XY embryos, *GsdfbY* showed high expression from 2 days before hatching (dbh) to the hatching day just before the period when the sex difference in the gonad became manifest (Figure 17). Expression of *Gsdf* was detected at very low levels at 2 dbh and became a similar high level to *GsdfbY* at the hatching day. In XX

embryos, expression of *Gsdf* was low from 2 dbh to the hatching day. These results suggest that *GsdfbY* has acquired a novel earlier expression pattern during sex-determining period.

GsdfbY induced fertile XX males in O. pectoralis

I performed overexpression experiments using a *GsdfbY* genomic clone (Opec3-1). I made a construct (Construct 3) that contained 4.2 kb of *GsdfbY*, as well as 2.8 kb of its upstream region and 0.3 kb its downstream region, but no other genes (Figure 18). Construct 3 was injected into one-cell-stage embryos of *O. pectoralis*. In generation zero (G0), I obtained 30 adult fishes with the transgene, 15 of which were sex-reversed XX males (Table 4). I mated an XX male with a normal female to obtain G1 progeny. All fish bearing the transgene developed as males, whereas all fish without the transgene developed as females. Consequently, I established a transgenic strain (Strain 1) whose sex was determined by the transgene Construct 3.

Similarly, I established Strain 2 from another XX male in G0 progeny (Table 4). These results suggest that *GsdfbY* is the SD gene in *O. pectoralis* and represents a new SD gene in vertebrates.

GsdfbY arose from the common ancestor of the mekongensis subgroup

I determined the mRNA sequences of *Gsdf* and *GsdfbY* using RT-PCR product in *O. pectoralis*. Similarly, those of *Gsdf* were determined in nine *Oryzias* fishes. The longest open reading frame (ORF) of *Gsdf* and *GsdfbY* spanned five exons and encodes a putative protein of 217 amino acids. The N-terminal regions are rich in hydrophobic amino acid residues and are followed by a potential cleavage site comprising Ala and Phe. Phylogenetic analysis of the complete amino acid sequences demonstrated that the GSDFbY sequence is found in a clade with GSDF sequence in *O. mekongensis* and *O. songkramensis*. Since the bootstrap values supporting the clade was low (72 %), I regarded that the clade supported by low bootstrap values is a trifurcation, suggesting that duplication of *Gsdf* was occurred in common ancestor of the mekongensis subgroup. This inference means that species in the mekongensis subgroup may employ *GsdfbY* as the SD gene.

General Discussions

In this study, I identified two novel SD genes, *Gsdf*^{Y} in *O. luzonensis* and *GsdfbY* in *O. pectoralis*, differing in the appearance process from the same ancestral gene *Gsdf*. These SD genes keep high expression of total *Gsdf* in males during sex differentiation period (Figure 20). In *O. luzonensis*, *Gsdf*^{Y} expresses at high level and also induces high expression of *Gsdf*^{X}. In *O. pectoralis, GsdfbY* expresses earlier than *Gsdf* and leads to high expression of *Gsdf*. Likewise, in *O. latipes*, *Dmy* give rise to high expression of *Gsdf* (Shibata *et al.* 2010). This common feature suggests that a gene which keeps high expression of *Gsdf* during sex differentiation period can be the master SD gene.

Models of sex chromosome and SD gene evolution state that a pair of autosomes became sex chromosomes when a new SD allele evolved on one partner (Ohno 1967). Then, since advantageous alleles were accumulated, the sex chromosomes became to suppress recombination between them. The non-recombination region with mutations, deletion and insertion led to morphological difference of the sex chromosomes. The SD region of X and Y in *O. latipes*, *O. luzonenis* and *O. pectoralis* is almost same except for the region around the SD gene. These facts indicate that sex chromosome in these species are in the early stage of evolution. For this reason, SD gene would have high turnover rates. Transgenic insertion of *Dmy* has made an autosome into sex chromosome (Otake *et al.* 2010). I established four transgenic strains in which either *Gsdf*^{Y} or *GsdfbY* were inserted into the genome. In these strains, autosomes would also be change into sex chromosome. These results indicate that autosomes can quite easily become sex chromosomes. Given that *Gsdf* and *Dmrt1* repeatedly evolved into SD gene, sex-related genes including *Gsdf* and *Dmrt1* would be easy to become SD gene by neofunctinalization with or without duplication.

Acknowledgements

I thank C. L. Peichel for constructive comments on the manuscript. I am grateful to Dr. Y. Takehana and A. shinomiya for helpful advice. We thank the Sequencing Technology Team at the RIKEN Genomic Sciences Center (GSC) for BAC and fosmid clone sequencing. This work was supported by KAKENHI via Grants-in-Aid for Scientific Research on the Priority Area "Comparative Genomics" from the Ministry of Education, Culture, Sports, Science, and Technology of Japan (MEXT). And this work also supported by KAKENHI via Grants-in-Aid for Scientific Research on the Priority Area Scientific Research B (16370094 and 20370086).

References

Arnold, C., and I. J. Hodgson, 1991 Vectorette PCR: a novel approach to genomic walking. Genome Res. 1: 39-42.

Devlin, R. H., and Y. Nagahama, 2002 Sex determination and sex differentiation in fish: an overview of genetic, physiological, and environmental influences. Aquaculture 208: 191–364.

Formacion, M. J., and H. Uwa, 1985 Cytogenetic studies on the origin and species differentiation of the Philippine medaka, *Oryzias luzonensis*. J. Fish Biol. 27: 285–291.

Foster, J. W., F. E. Brennan, G. K. Hampikian, P. N. Goodfellow, A. H. Sinclair, *et al.*, 1992 Evolution of sex determination and the Y chromosome: *SRY*-related sequences in marsupials. Nature 359: 531–533.

Gautier, A., F. Sohm, J. S. Joly, F. Le Gac and J. J. Lareyre, 2011 The proximal promoter region of the zebrafish gsdf gene is sufficient to mimic the spatio-temporal pattern of the endogenous gene in Sertoli and granulosa cells. Biol. Reprod. 85: 1240-51.

Gubbay, J., J. Collignon, P. Koopman, B. Capel, A, Economou, *et al.*, 1990 A gene mapping to the sex-determining region of the mouse Y chromosome is a member of a new family of embryonically expressed genes. Nature 346: 245–250.

Hamaguchi, S., Y. Toyazaki, A. Shinomiya, and M. Sakaizumi, 2004 The XX-XY sex-determination system in *Oryzias luzonensis* and *O. mekongensis* revealed by the sex ratio of the progeny of sex-reversed fish. Zool. Sci. 21: 1015–1018.

Hyodo-Taguchi, Y., and M. Sakaizumi, 1993 List of inbred strains of the medaka, *Oryzias latipes,* maintained in the Division of Biology, National Institute of Radiological Sciences. Fish Biol. J. Medaka 5: 29–30.

Kondo, M., I. Nanda, U. Hornung, S. Asakawa, N. Shimizu, *et al.*, 2003 Absence of the candidate male sex-determining gene dmrt1b(Y) of medaka from other fish species. Curr. Biol. 13: 416–420.

Kondo, M. I. Nanda, U. Hornung, S. Imai, T. Sasaki, *et al.*, 2006 Genomic organization of the

sex-determining and adjacent regions of the sex chromosomes of medaka. Genome Res. 16: 815–826.

Koopman, P., J. Gubbay, N. Vivian, P. Goodfellow, and R. Lovell-Badge, 1991 Male development of chromosomally female mice transgenic for *Sry*. Nature 351: 117–121.

Marshall-Graves, J. A., 2002 The rise and fall of *SRY*. Trend. Genet. 18: 259–264.

Matsuda, M. Y. Nagahama, A. Shinomiya, T. Sato, C. Matsuda, *et al.*, 2002 *DMY* is a Y-specific DM-domain gene required for male development in the medaka fish. Nature 417: 559–563.

Matsuda, M., T. Sato, Y. Toyazaki, Y. Nagahama, S. Hamaguchi, *et al.*, 2003 *Oryzias curvinotus* has *DMY*, a gene that is required for male development in the medaka, *O. latipes*. Zool. Sci. 20: 159–161

Matsuda, M., A. Shinomiya, M. Kinoshita, A. Suzuki, T. Kobayashi, *et al.*, 2007 The *DMY* gene induces male development in genetically female (XX) medaka fish. Proc. Natl. Acad. Sci. USA 104: 3865–3870.

Nagai, T., Y. Takehana, S. Hamaguchi, and M. Sakaizumi, 2008 Identification of the sex-determining locus in the Thai medaka, *Oryzias minutillus*. Cytogenet Genome Res. 121: 137-42.

Nakamoto, M., S. Muramatsu, S. Yoshida, M. Matsuda, Y. Nagahama, *et al.*, 2009 Gonadal sex differentiation and expression of *Sox9a2*, *Dmrt1*, and *Foxl2* in *Oryzias luzonensis*. Genesis 47: 289–299.

Nanda, I., M. Kondo, U. Hornung, S. Asakawa, C. Winkler, *et al.*, 2002 A duplicated copy of *DMRT1* in the sex-determining region of the Y chromosome of the medaka, *Oryzias latipes*. Proc. Natl. Acad. Sci. USA 99: 11778–11783.

Ohno, S. 1967 Sex Chromosomes and Sex Linked Genes. Springer-verlag

Otake, H., H. Masuyama, Y. Mashima, A. Shinomiya, T. Myosho, *et al.*, 2010 Heritable artificial sex chromosmes in the medaka, *Oryzias latipes*. Heredity 105: 247-56.

Parent, L. R. 2008 A phylogenetic analysis and taxonomic revision of ricefishes, *Oryzias* and

relatives (Beloniformes, Adrianichthyidae). Zoo. J. Linn. Soc. 154: 494-610

Pomiankowski, A., R. Nothiger, and A. Wilkins, 2004 The evolution of the Drosophila sex-determination pathway. Genetics 166: 1761-1773.

Ross, J. A., J. R. Urton, J. Boland, M. D. Shapiro, and C. L. Peichel, 2009 Turnover of sex chromosomes in the Stickleback fishes (Gasterosteidae). PLos. Genet. 5: e1000391.

Sawatari, E., S. Shikina, T. Takeuchi, and G. Yoshizaki, 2007 A novel transforming growth factor-b superfamily member expressed in gonadal somatic cells enhances primordial germ cell and spermatogonial proliferation in rainbow trout (*Oncorhynchus mykiss*). Dev. Biol. 301: 266–275.

Shibata, Y., B. Paul-Prasanth, A. Suzuki, T. Usami, M. Nakamoto, *et al.*, 2010 Expression of *gonadal soma derived factor* (*GSDF*) is spatially and temporally correlated with early testicular differentiation in medaka. Gene. Expr. Patterns 10: 283–289.

Shinomiya, A., H. Otake, K. Togashi, S. Hamaguchi, and M. Sakaizumi, 2004 Field survey of sex-reversals in the medaka, *Oryzias latipes*: Genotypic sexing of wild populations. Zool. Sci. 21: 613–619.

Sinclair, A. H., P. Berta, M. S. Palmer, J. R. Hawkins, B. L.Griffiths, *et al.*, 1990 A gene from the human sex-determining region encodes a protein with homology to a conserved DNA-binding motif. Nature 346: 240–244.

Solari, A. J., 1994 Sex Chromosomes and Sex Determination in Vertebrates. CRC Press, Boca Raton, FL.

Takehana, Y., K. Naruse, and M. Sakaizumi, 2005 Molecular phylogeny of the medaka fishes genus Oryzias (Beloniformes: Adrianichthyidae) based on nuclear and mitochondrial DNA sequences. Mol. Phylogenet. Evol. 36: 417–428.

Takehana, Y., D. Demiyah, K. Naruse, S. Hamaguchi, and M. Sakaizumi, 2007a Evolution of different Y chromosomes in two medaka species, *Oryzias dancena* and *O. latipes*. Genetics 175: 1335–1340.

Takehana, Y., K. Naruse, S. Hamaguchi, and M. Sakaizumi, 2007b Evolution of ZZ/ZW and XX-XY sex-determination systems in the closely related species, *Oryzias hubbsi* and *O. dancena*. Chromosoma. 2007b 116: 463-70

Takehana, Y., S. hamaguchi, and M. Sakaizumi 2008 Different origins of ZZ/ZW sex chromosomes in closely related medaka fishes, *Oryzias javanicus* and *O. hubbsi*. 2008 Chromosome Res. 16: 801-11.

Tanaka, K., Y. Takehana, K. Naruse, S. Hamaguchi, and M. Sakaizumi, 2007 Evidence for different origins of sex chromosomes in closely related *Oryzias* fishes: substitution of the master sex-determining gene. Genetics 177: 2075–2081.

Wilkins, A. S., 1995 Moving up the hierarchy: a hypothesis on the evolution of a genetic sex determination pathway. BioEssays 17: 71–77.24.

Woram, R. A., K. Gharbi, T. Sakamoto, B. Hoyheim, L. E. Holm, *et al.*, 2003 Comparative genome analysis of the primary sex-determining locus in salmonid fishes. Genome Res. 13: 272–280.

Yoshimoto, S., E. Okada, H. Umemoto, K. Tamura, Y. Uno, *et al.*, 2008 A W-linked DM-domain gene, *DM-W*, participates in primary ovary development in *Xenopus laevis*. Proc. Natl. Acad. Sci. USA 105: 2469–2474.

Table 1 Sequences of primers

method	name	primer (5'-3')
genotypic sexing	OluX2-8f	AAGTGAACAGATGAATTAAAAATA
	OluX2-8r	GTTCTTGAGGAACTACGTCC
	OluX2-25f	CTGCCTGAGGCAAATGATGAA
	OluX2-25r	TTTTGACAGGTTGGATCTGT
	OluX3-34f	ACGGTGTTGCCGTTTCTCATTTC
	OluX3-34r	CGTACCTTCCTGGGGTCTCATTT
	OluX4-6f	TAATTAAACGTGGGTGCATGTGG
	OluX4-6r	GGGCCATGTTCAAACATTGT
RT-PCR	OluPG1f	CCAAGGAAAAGCGTACAGGA
	OluPG1r	ATTCCGACAGTTCAGGTTGC
	OluPG2f	CACCCTATCGGAGAAGTTGG
	OluPG2r	TGGAAACGTGTGGAACTCTG
	OluPG3f	CCATCCGGTTGGTCAAGA
	OluPG3r	AAGCGCAGTTCAGAGTCCAT
	OluPG4f	TCCATGAAACTGCCAACTGA
	OluPG4r	GTTCGCCATTGTCGGTAGAT
	OluPG5f	TGGAAACTTGTGGGGGAAT
	OluPG5r	GCAATCCCTCGGATCAAACAG
	OluPG6f	AGCCACAGTGATCAACGACA
	OluPG6r	GGAGGCGGAAAATATTGTGA
	OluPG7f	GTGTGAAATCCCTGCCAAAT
	OluPG7r	CGCAGCAGCTCTTTAAATCC

	OluPG8f	ATGACACGGAGATCCTACGC
	OluPG8r	TTTCCACCATGGTCACATTG
	OluPG9f	GCTGAAGCAAAATCCTCGTC
	OluPG9r	TGCTGGCATCAATGATCTGT
	B-actin3b	CMGTCAGGATCTTCATSAGG
	B-actin4	CACACCTTCTACAATGAGCTGA
real-time PCR	qGsdfY_F	TCGAATAGAATGAAATCTGACCACTCT
	qGsdfY_R	GGTAAACCCAGGTTGGTCAAAT
	qGsdfX_F	TCGAATAGAATGAAATCTGACCACTTC
	GsdfX_R	GGTAAACCCAGGTTGGTCAAAG
Luciferase assay	Luc In-fusion 1f	TAGGTACCTTAGGCAAGGCCCTTAACGCTAC
	Luc In-fusion 1r	AACAAGCTTCAATGAGTGCCAAAGACATGGTGGA

Table 2 Phenotypic sex of transgenic strains

Strain No.	Host species / construct	Generation	Exogenous *Gsdf*^{tr}	XX	
				♀	♂
1	*O. luzonensis* / Construct 1	G0	+	53	1
			−	46	0
		G1	+	0	28
			−	28	0
		G2	+	0	49
			−	33	0
2	*O. luzonensis* / Construct 2	G0	+	35	5
		G1	+	1	15
3	*O. latipes* / Construct 1	G0	+	95	7
		G1	+	0	13
		G2	+	0	34
			−	21	0

Table 3 Sequences of primers

method	name	primer (5'-3')
Sex linkage analysis in *O.sp.* (Laos)	OLc0504df	TACAAAAGCAGCCGAAGAAGGAGCTC
	OLc0504dr	CCAGCAAGACGTTGTCACTTTTGATGTC
	13.63f	GTCTCTGAGGCGGGACTG
	13.63r	CCCGTTTCATGGCTGAATA
	AU169256f	CTACCTTTACCTGGCCAACCTGAGGG
	AU169256f	CGAGGCCGTAAGCAAAGAAGTACTCCTT
	28.53f	CGGCTCTGGAGATCTTTCAC
	28.53r	ACCAATGGATTCCTGCTTGA
	31.31f	ACTGGAAGGAGTGGAAGCAG
	31.31r	CGTTGGTCAGCTGTTCCTG
	32.4f	GCAGATTGTCTTTTCCTGTGC
	32.4r	TCTTCTCTCTGGGCTACTGTCC
Sex linkage analysis in *O. celebensis, O. woworae and O.matanensis*	HNRNPU-2f	AGAAGAAGGGGACGAAGACG
	HNRNPU-2r	TGATGTTTTGTGGGGTGAAA
	SOX7-f	ATCAGCGCGTACTCGTCGT
	SOX7-r	GTTGGGGTAGTCCTGCATGT
	HSDL1f	CTTCCAGCTTTTGTACAGAGAAAT
	HSDLr1	CGTGGCGAGCGTAGACCT
	PCMT1f	CCTACATGGACTCGCCTCAGT
	PCMT1r	CCTTGGCTCCCTCGTACAG
	EHD3-1f	GATTCTTTCATTGCAGTGATGC
	EHD3-1r	GTCGATCACGCTGATGCTCT

	FDFT1-f	GTGGGATCCATGAAGGAGTG
	FDFT1-r	GCATGGCAGCAGACAGGTACAT
	RPF-f	ACCCATGCTGGTGTTTGC
	RPF-r	AAGTCCTGCTTCTGCATGTG
	MED23-1f	GCAGTTTAATATCAGACTTTGTGG
	MED23-1r	AGTCCTAACATGTTGCACACC
	CLU-1f	AAGGCTCTGTGGGAGGAGTG
	CLU-1r	GTTGACCCAAATGGAGAAGG
Sex linkage analysis in *O. profundicola and O. marmoratus*	OLc3106af	CTCGGTCCTCCGGACGTTTATCCCCAA
	OLc3106ar	TCTGCGACCTTGCTGTGACGAGCCA
	VEGFRf	GGCTCATGCTGATCCTCTTC
	VEGFRr	ATGGACGGACCCATCTTGTA
	MF01SSA075C05f	AGCATATTTAGGGTCTGAGGGA
	MF01SSA075C05r	GGCAGAAGAGAGGCTAAATCG
	OLb2511ef	ACACCGAAAAGTCGCAGCCATGG
	OLb2511er	GTGACCCGAGATGTGACATCAAACATG
	SPARCf	ATGAGGGACTGGCTGAAGAA
	SPARCr	CCAGTGGACAGGGAAGATGT
Sex linkage analysis in *O. pectoralis*	Gsdf-f	AATTGGCATCATCTGAAGGTT
	Gsdf-r	CTCAGCTTTGCATTCCTGCT
Mapping of *GsdfbY* and *Gsdf*	MF01SSA011H06f	CTCCCCAATGACGACGGATCTC
	MF01SSA011H06r	CAGCCGAAGCGTGATTTGACCC
	MF01SSA043H02f	TCAGCGTGCTACTTGTACTGCTTCCTCC
	MF01SSA043H02r	GGAGCTGCTGACGCTCAGAAATATTGA

	SNP12-1f	ACGACATCATCAACTCAGGCT
	SNP12-1r	TGGGTTTCGAATGTACCCATA
	OLd2509ef	TGCAGGTCTGGAGGAGAAGATCAAGC
	OLd2509er	TGAAGATCTTCACCACCTTCGGGG
	OLb1004ff	CACCCGTTTCAAGCTGGTCGTTT
	OLb1004fr	CCAGCTCGTCGAACTTCACCTTCTC
	OlMA1f	CAGGGAGAAAATGACCCAGA
	OlMA1r	CCTCGTAGATGGGGACATTG
	MF01SSA009B03f	ACCACCAGCTCCTCCAAGATGT
	MF01SSA009B03r	CGCTCACATCGACCACTTCACT
	SNP1-3f	GGAATTGATTGTTTGGTCCCT
	SNP1-3r	GGTCGAGGACGTCAATGATAA
	PMM2-f	CAGGAGGAGCGGATCGAGTTCTA
	PMM2-r	CCTCTTATCCCAGCCGTCAGG
	HOGA1f	CCAGCTTCTACAAGGGCAAGAT
	HOGA1r	CCGGAACGCTGTACAGAACCAC
RT-PCR	OpecGsdf-f	ACCACAGGTGTCAGGATGAG
	OpecGSdf-r	CCCTTGCCTACTTTTTGCTG

Table 4 Phenotypic sex of transgenic strains in *O. pectoralis*

Host species / construct	Generation	Strain No.	Presence of transgene	XX ♀	XX ♂
O. pectoralis / Construct 3	G0		+	15	15
	G1	1	+	0	7
			-	5	0
		2	+	0	5
			-	6	0

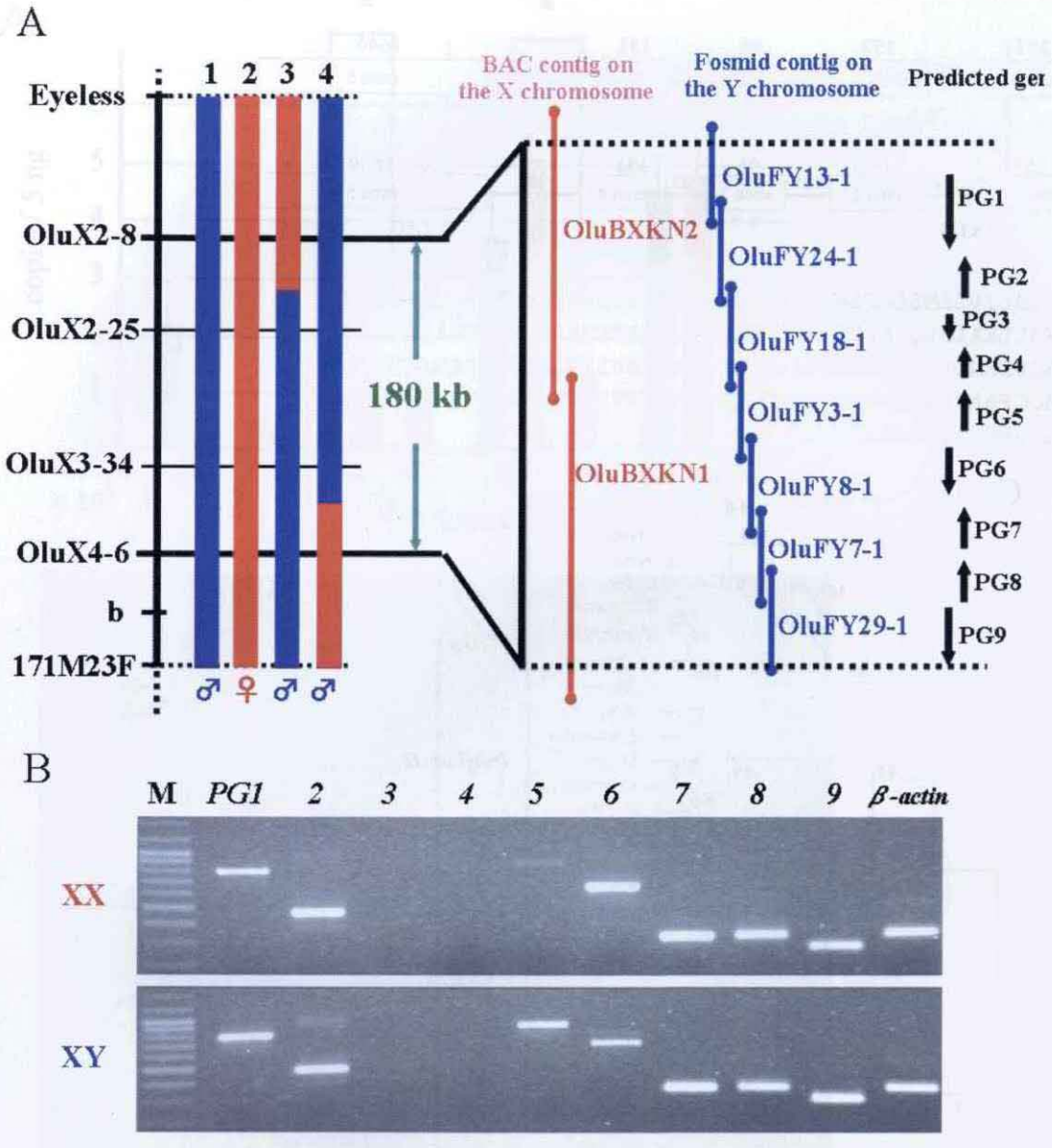

Figure 1 (A) Genetic and physical map of the sex-determining regions on the X and Y chromosomes. 1, normal XY male, 2, XX female. 3 and 4, recombinant male. Blue column, regions derived from the Y chromosome. Red, from the X chromosome. Red bar, BAC clones. Blue, fosmid clones. Black arrows indicate *predicted genes*. (B) RT-PCR products of *PGs* (*PG1–9*) in the XX and XY body trunk at 0 days after hatching. M represents the size marker.

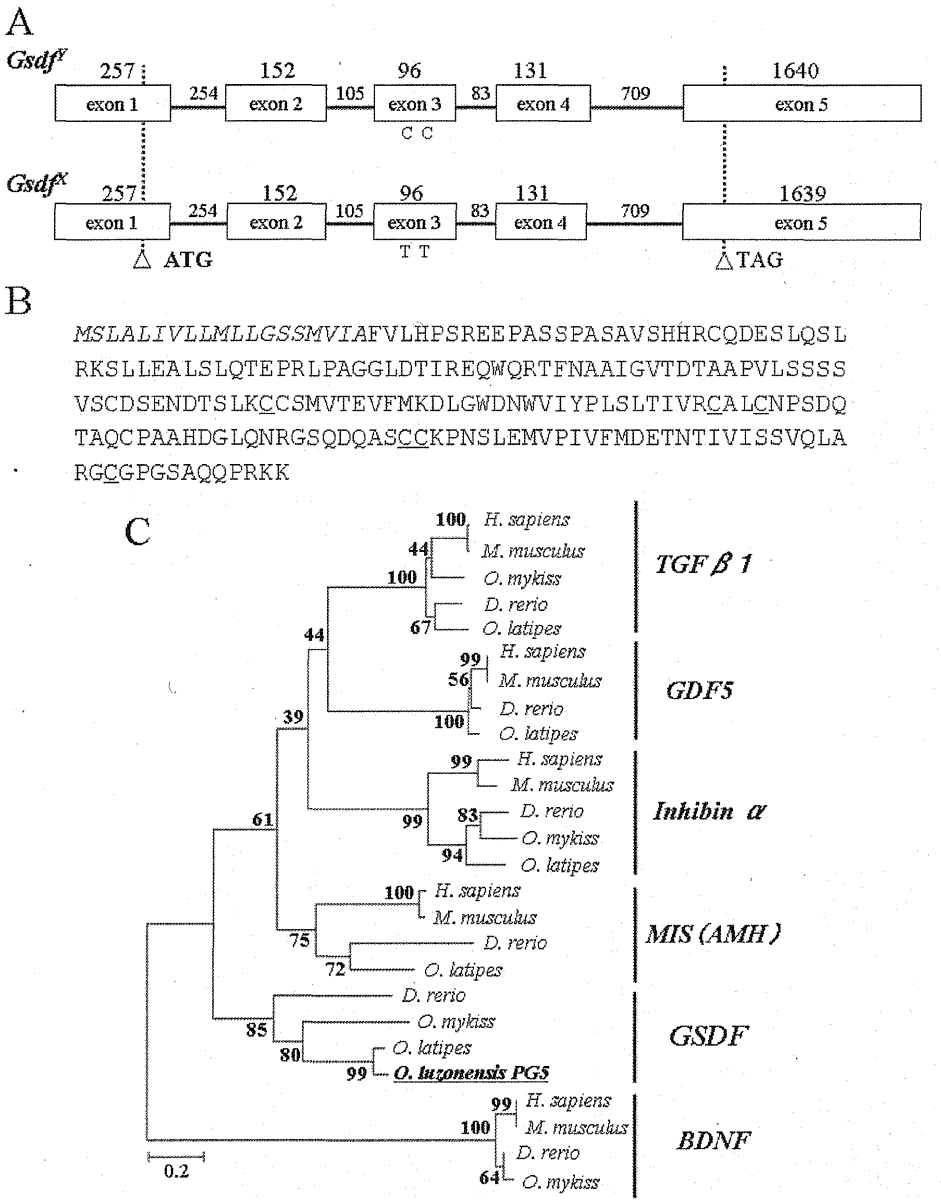

Figure 2 (A) Genomic structure of *Gsdf*^{Y} and *Gsdf*^{X}. Two synonymous substitutions are present in exon 3. Open boxes, exons. Horizontal bars, introns. Numbers represent nucleotide sequence length (bp). Open arrowheads indicate the translation start (ATG) and stop (TGA) sites. (B) Amino acid sequences of *Gsdf*. *Gsdf*^{Y} and *Gsdf*^{X} are the same. Signal peptide is indicated in italics. The six conserved cysteine residues are underlined. (C) Neighbour-joining (NJ) tree for the TGF-β superfamily using the amino acid sequence of the mature domain. The tree was rooted by using BDNF.

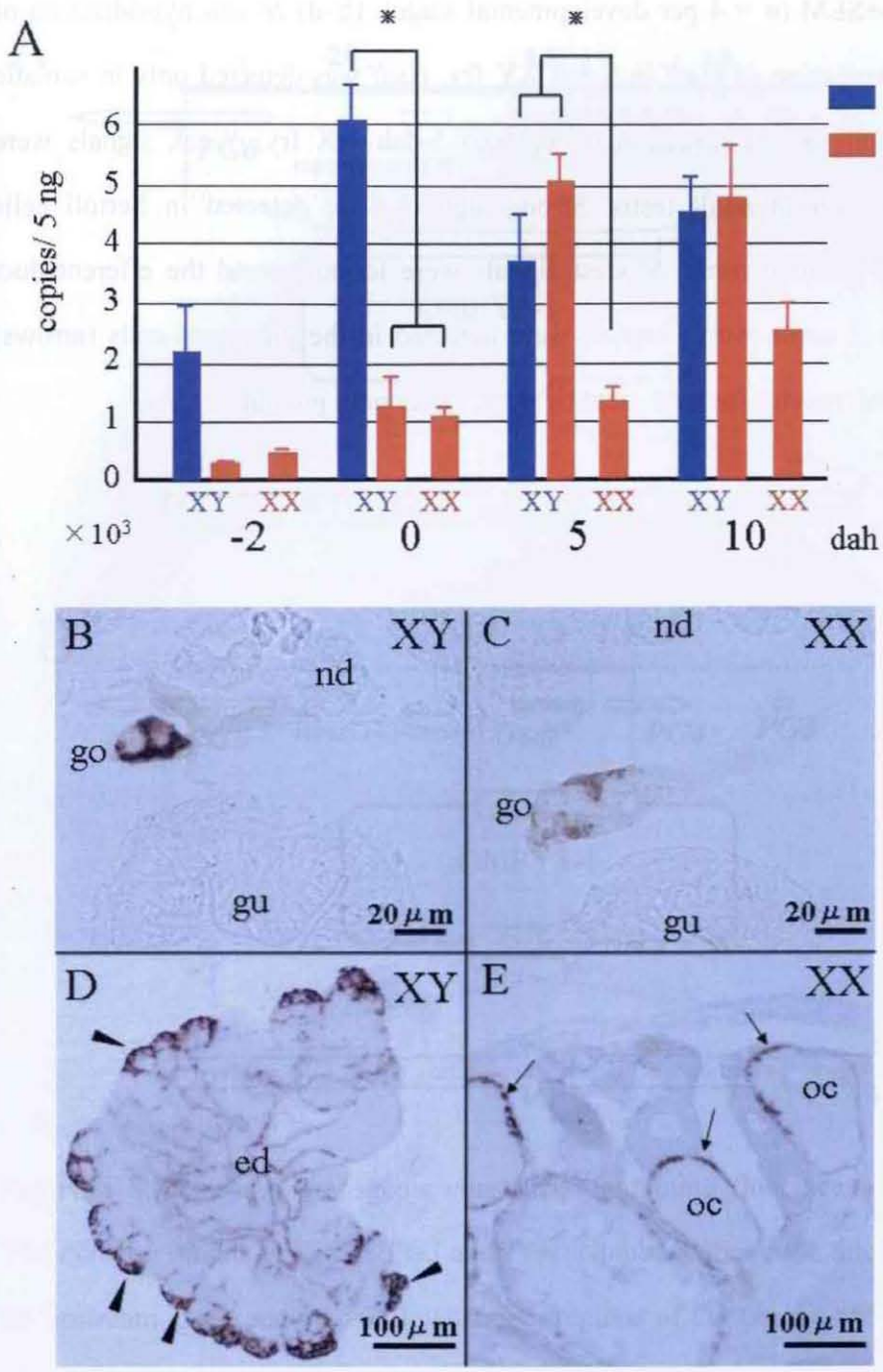

Figure 3 Expression of *Gsdf*. (A) Real-time PCR of *Gsdf* in XX and XY fry from 2 days before hatching (dbh) to 10 days after hatching (dah). Blue bars, expression of *PG5* on the Y chromosome ($Gsdf^Y$). Red, the X chromosome ($Gsdf^X$). Significant differences were analyzed by using two-way analysis of variance (ANOVA) followed by Bonferroni post-tests. Columns and

error bars represent mean±SEM (n = 4 per developmental stage). (b–d) *In situ* hybridization of *Gsdf* in the gonad. (B) Expression of *Gsdf* in 5-dah XY fry. *Gsdf* was detected only in somatic cells surrounding germ cells. (C) Expression of *Gsdf* in 5-dah XX fry. Weak signals were detected. (D) *Gsdf* expression in adult testis. Strong signals were detected in Sertoli cells (arrowheads) surrounding spermatogonia. Modest signals were found around the efferent duct (ed). (E) *Gsdf* expression in adult ovary. Signals were detected in the granulosa cells (arrows) surrounding well-developed oocytes (oc). nd, nephric duct; go, gonad; gu, gut.

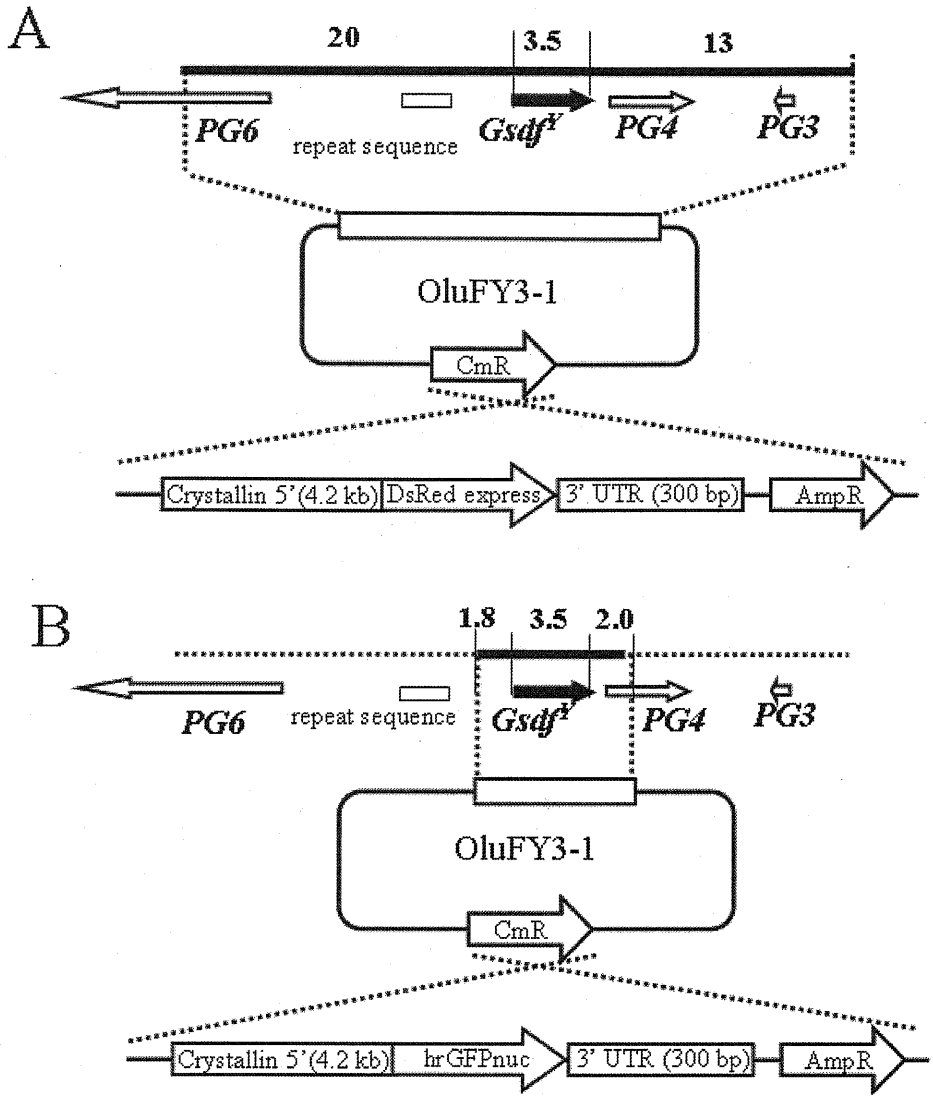

Figure 4 Structures of transgenic constructs containing fluorescent markers. (A) Construct 1. This construct includes *Gsdf*^{Y}, *PG3* and *PG4*. Numbers above the thick bar indicate the length of the upstream, gene coding and downstream regions of the *Gsdf*^{Y} gene in the fosmid construct. A fragment encoding crystallin-RFP and an ampicillin resistance gene was integrated into the ORF of the chloramphenicol resistance gene. (B) Construct 2. This construct includes only *Gsdf*^{Y}. Numbers above the thick bar indicate the length of the upstream, gene coding and downstream regions of the *Gsdf*^{Y} gene. A fragment encoding crystallin-GFP and an ampicillin resistance gene was integrated into the ORF of the chloramphenicol resistance gene.

B

upstream	1	2	3	4	5	6-9
Gsdf^{Y}	C	T	T	G	C	AATG
Gsdf^{X}	T	G	A	T	A	TCA
Gsdf(*O. latipes*)	T	G	A	T	A	TCA

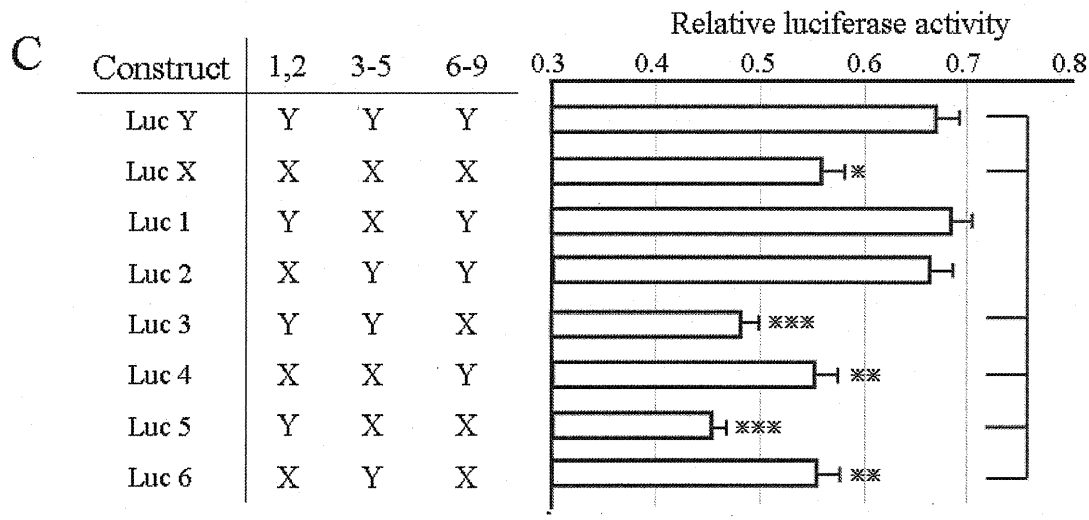

Figure 5 Mutations in the *Gsdf* cis-regulatory element contribute to *Gsdf*^{Y}-specific high expression. (A) Illustration of *Gsdf*^{X}-specific mutations in the 1.8 kb upstream and 2 kb downstream regions of *Gsdf*^{Y}, comparing with *Gsdf*^{X} and *Gsdf* (*O. latipes*). Numbers represent positions of *Gsdf*^{Y} specific mutations. (B) *Gsdf*^{X}-specific sequences in the upstream region of *Gsdf*^{Y}. (C) Luciferase assay analysis of sequences responsible for *Gsdf*^{Y}-specific high expression. * $P < 0.05$; ** $P < 0.01$; *** $P < 0.001$, one-way ANOVA, post-hoc comparisons, Turkey's test. Columns and error bars represent mean±SEM (n = 12).

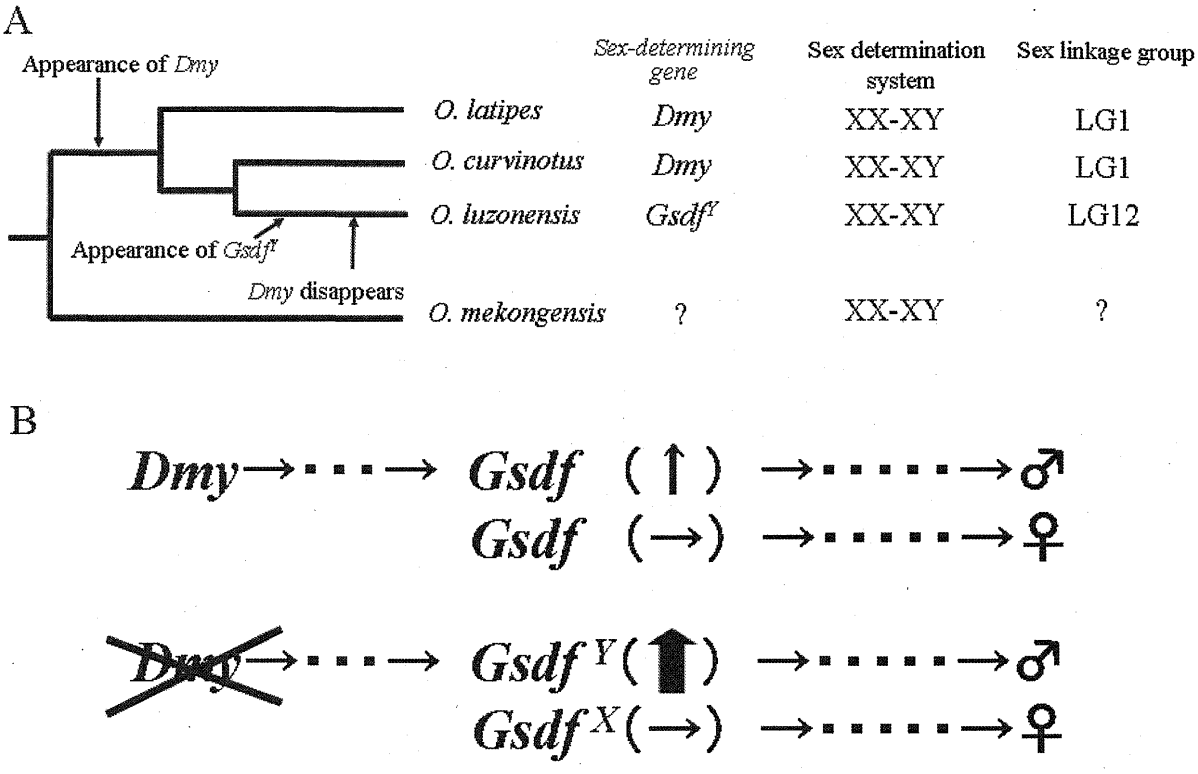

Figure 6 Changes in the sex-determining gene of *O. luzonensis*. (A) Evolutionary history up to the appearance of $Gsdf^{Y}$. *Dmy* appeared in a common ancestor of *O. latipes*, *O. curvinotus,* and *O. luzonensis*. $Gsdf^{Y}$ appeared in, and *Dmy* disappeared from, the ancestor of *O. luzonenesis*. (B) Top, sex-determining cascade in an ancestor of *O. luzonensis*. Bottom, current sex-determining cascade in *O. luzonensis*. *Gsdf* was downstream of *Dmy*. A mutation then occurred in *Gsdf*, allowing its expression without *Dmy*. $Gsdf^{Y}$ then became the new sex-determining gene.

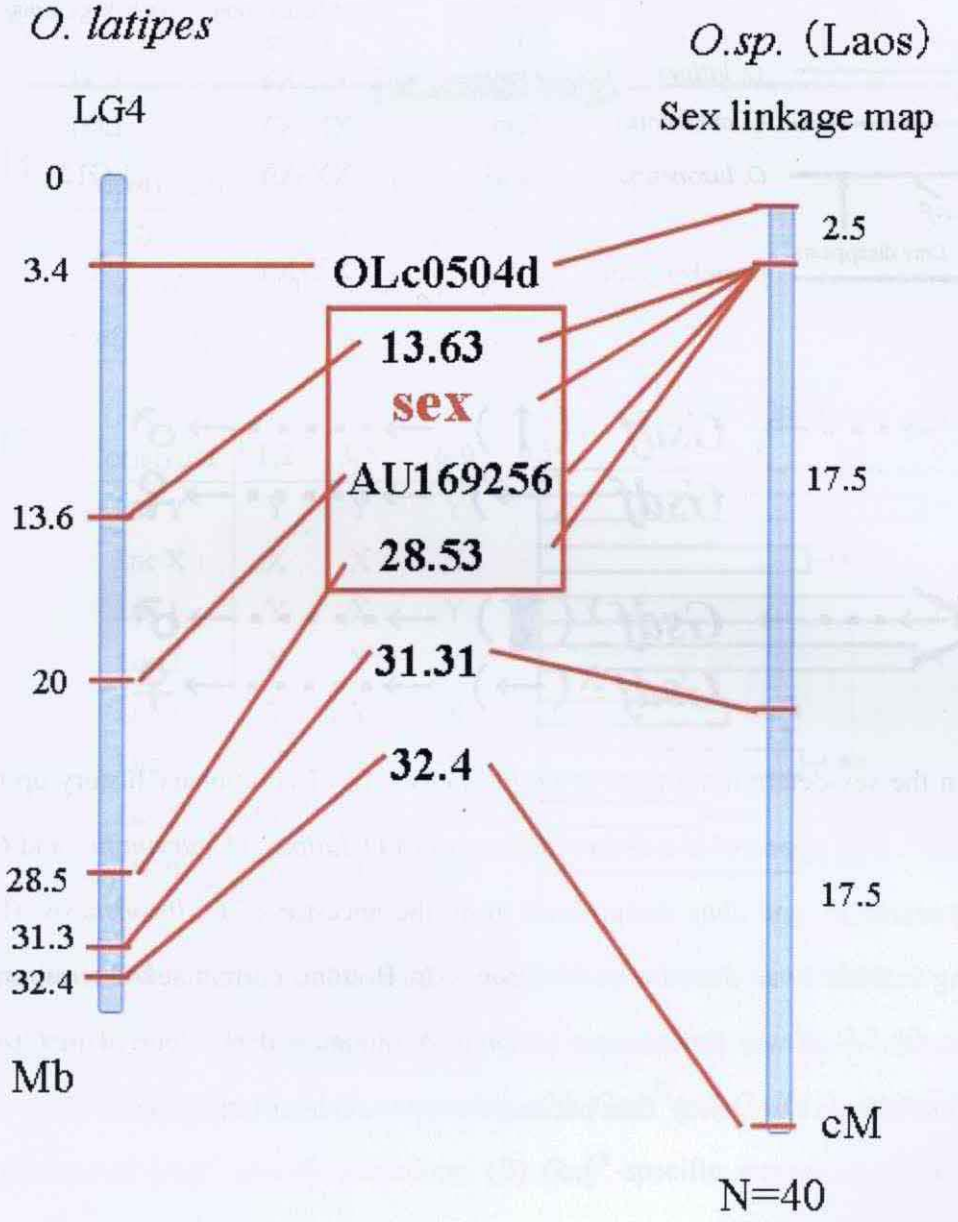

Figure 7 Comparison of the gene orders between the sex linkage map in *O. sp* (Laos) and the physical map of LG 4 in *O. latipes*. Lines between the compared chromosomes connect the positions of orthologous gene pairs in these two species. The distances between flanking markers are shown as physical length (left) or in cM (right).

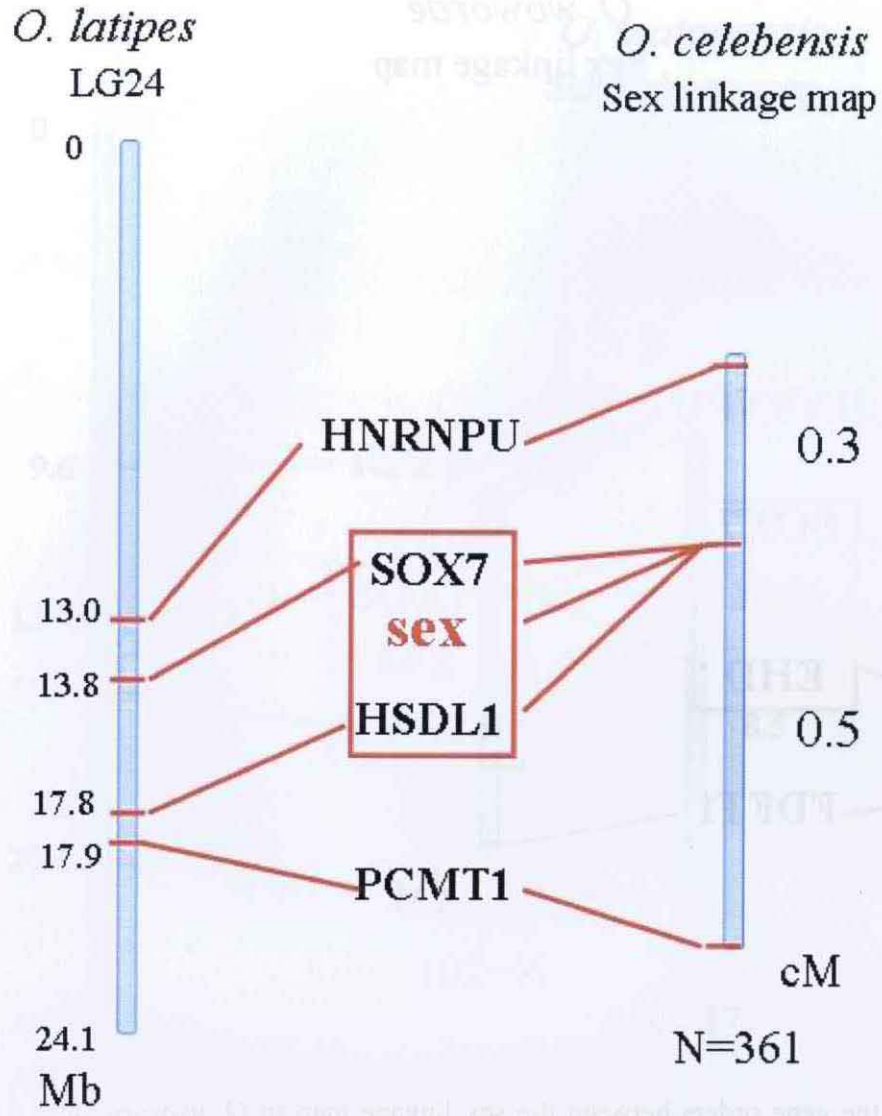

Figure 8 Comparison of the gene orders between the sex linkage map in *O. celebensis* and the physical map of LG 24 in *O. latipes*. Lines between the compared chromosomes connect the positions of orthologous gene pairs in these two species. The distances between flanking markers are shown as physical length (left) or in cM (right).

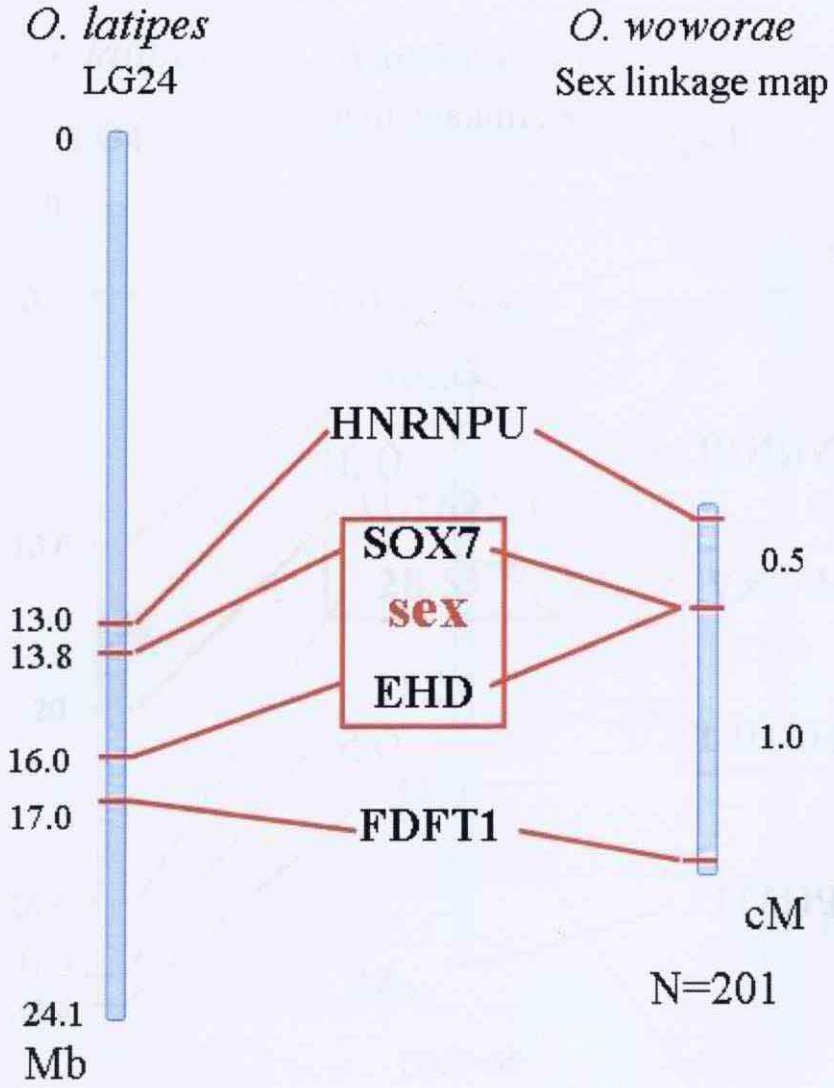

Figure 9 Comparison of the gene orders between the sex linkage map in *O. woworae* and the physical map of LG 24 in *O. latipes*. Lines between the compared chromosomes connect the positions of orthologous gene pairs in these two species. The distances between flanking markers are shown as physical length (left) or in cM (right).

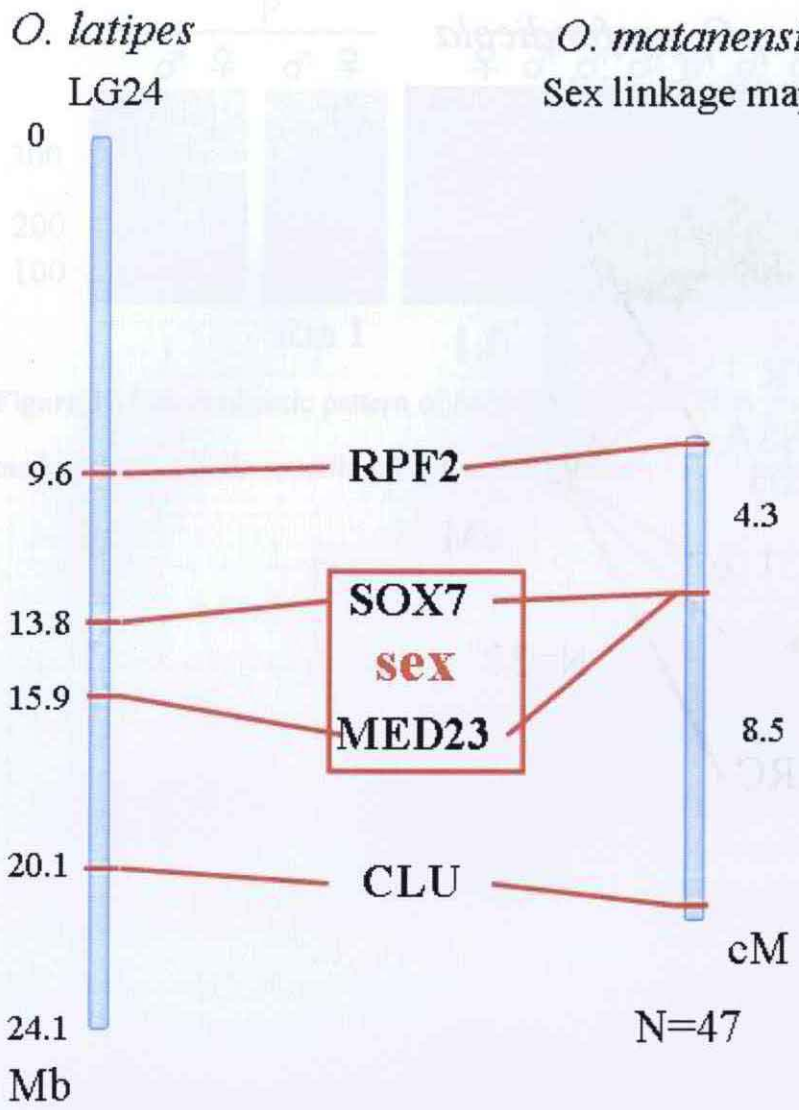

Figure 10 Comparison of the gene orders between the sex linkage map in *O. matanensis* and the physical map of LG 24 in *O. latipes*. Lines between the compared chromosomes connect the positions of orthologous gene pairs in these two species. The distances between flanking markers are shown as physical length (left) or in cM (right).

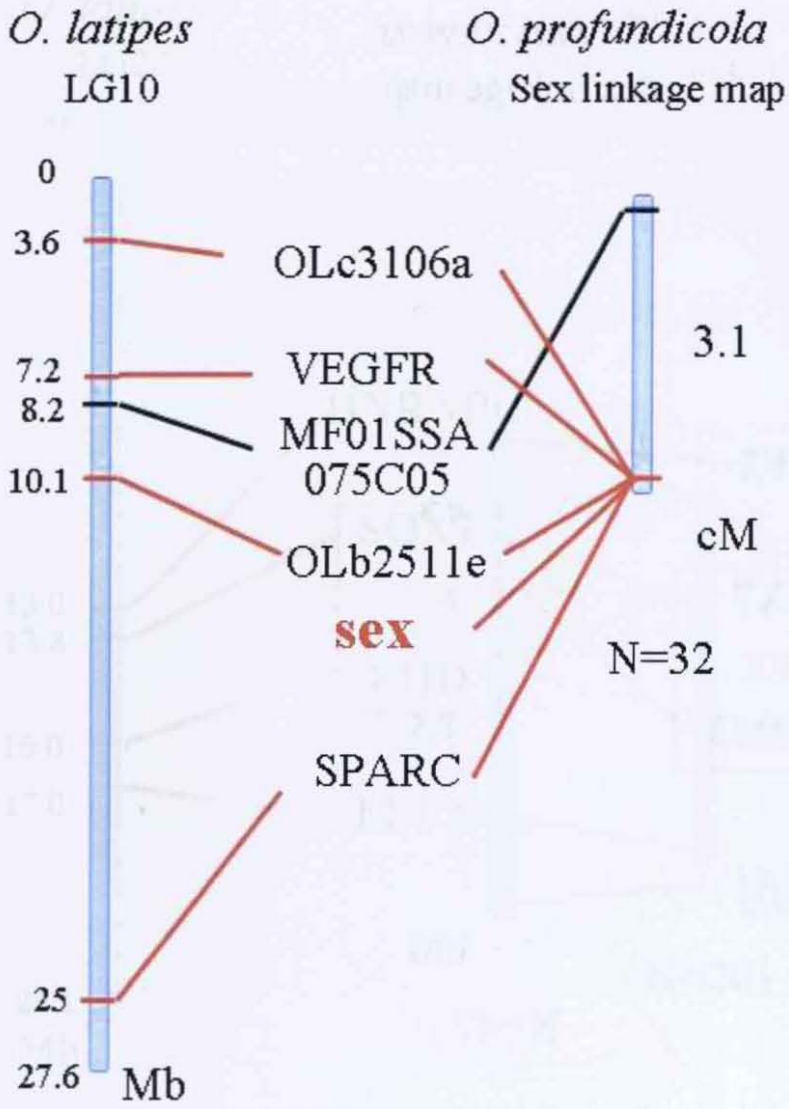

Figure 11 Comparison of the gene orders between the sex linkage map in *O. profundicola* and the physical map of LG 10 in *O. latipes*. Lines between the compared chromosomes connect the positions of orthologous gene pairs in these two species. The distances between flanking markers are shown as physical length (left) or in cM (right).

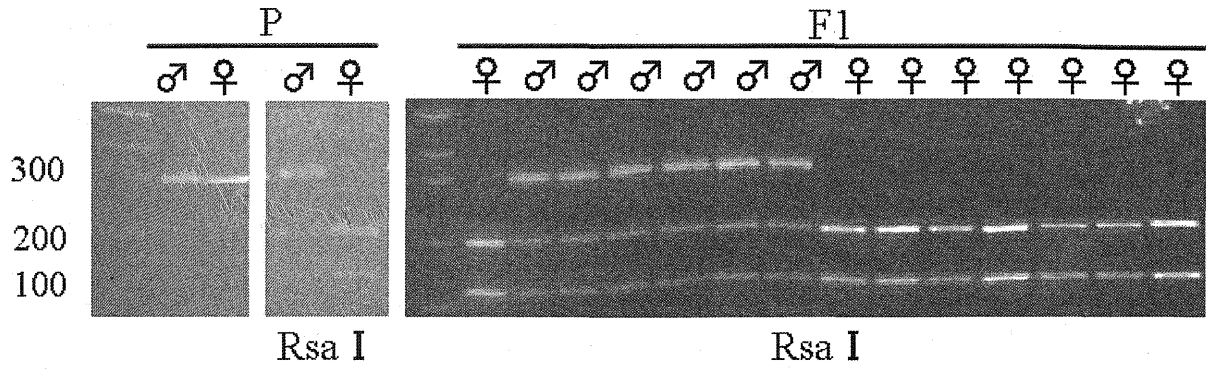

Figure 12 Electrophoetic pattern of MF01SSA075C05 PCR products after Rsa 1 treatment. This marker showed male-specific inheritance. P; parents, F1; F1 progeny from the parents.

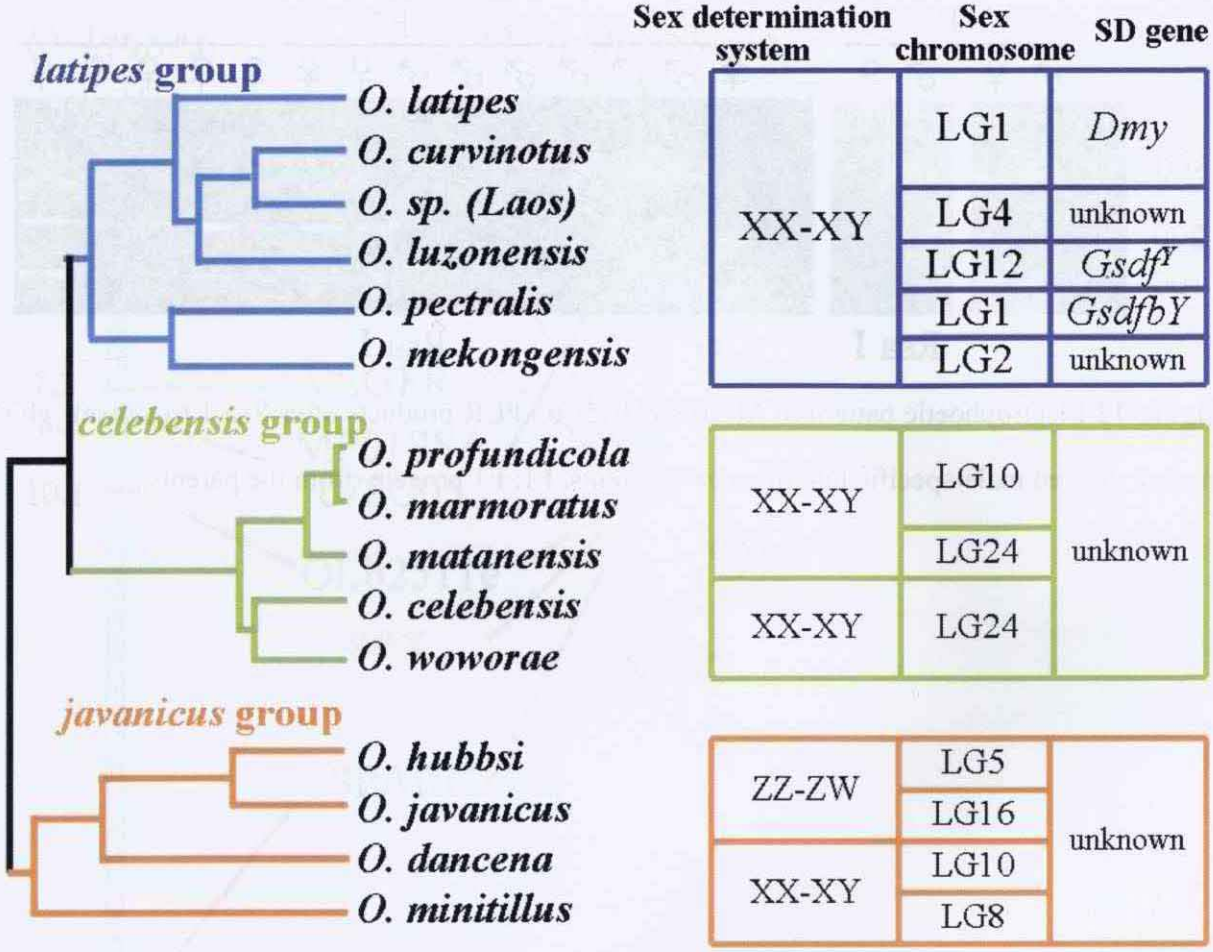

Figure 13 Phylogenetic relationship and sex determination mechanism in *Oryzias* fishes. The phylogenetic information was taken from Takehana *et al.* (2005).

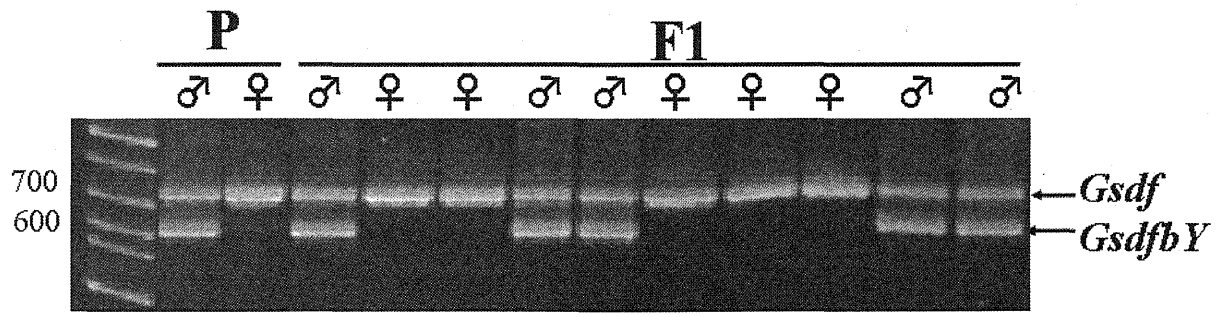

Figure 14 Electrophoetic pattern of PCR products of *Gsdf* in *O. pectoralis*. This marker showed male-specific inheritance. P; parents, F1; F1 progeny from the parents.

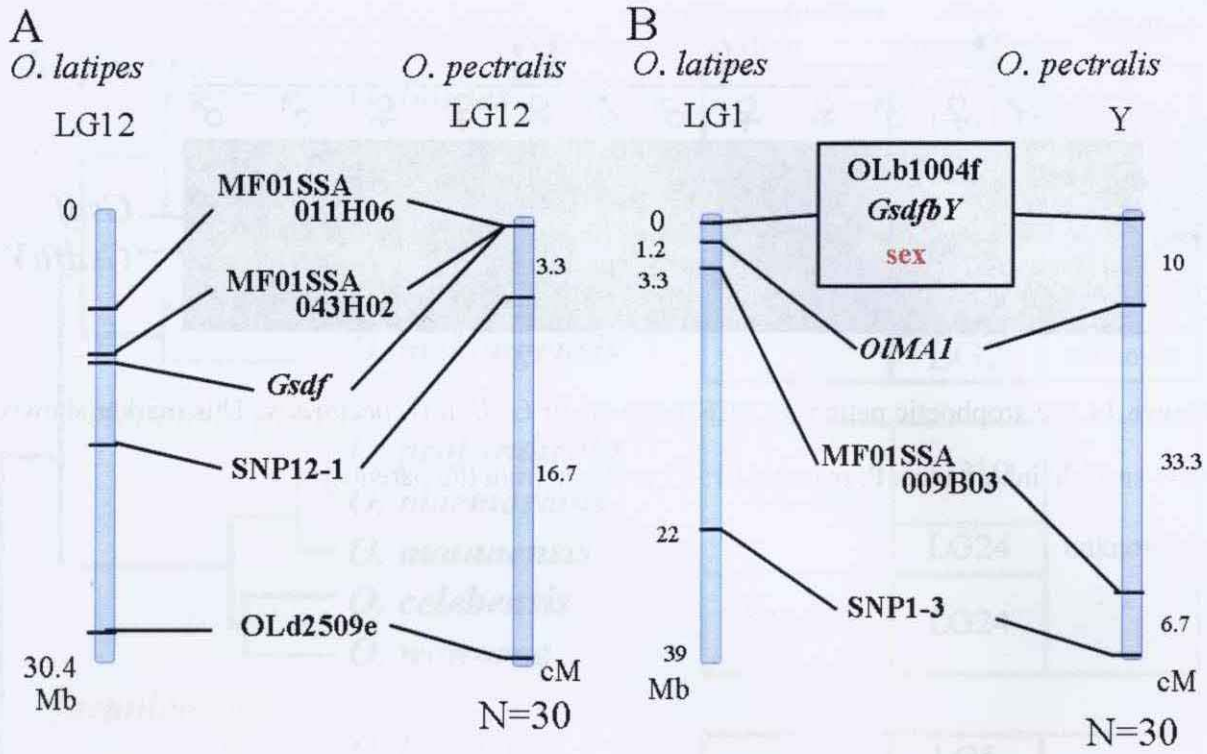

Figure 15 Linkage analysis *Gsdf* and *GsdfbY* (A) Comparison of the gene orders between *Gsdf* in *O. pectoralis* and the physical map of LG 12 in O. latipes. (B) Comparison of the gene orders between *GsdfbY* in *O. pectoralis* and the physical map of LG 1 in *O. latipes*.

Lines between the compared chromosomes connect the positions of orthologous gene pairs in these two species. The distances between flanking markers are shown as physical length (left) or in cM (right).

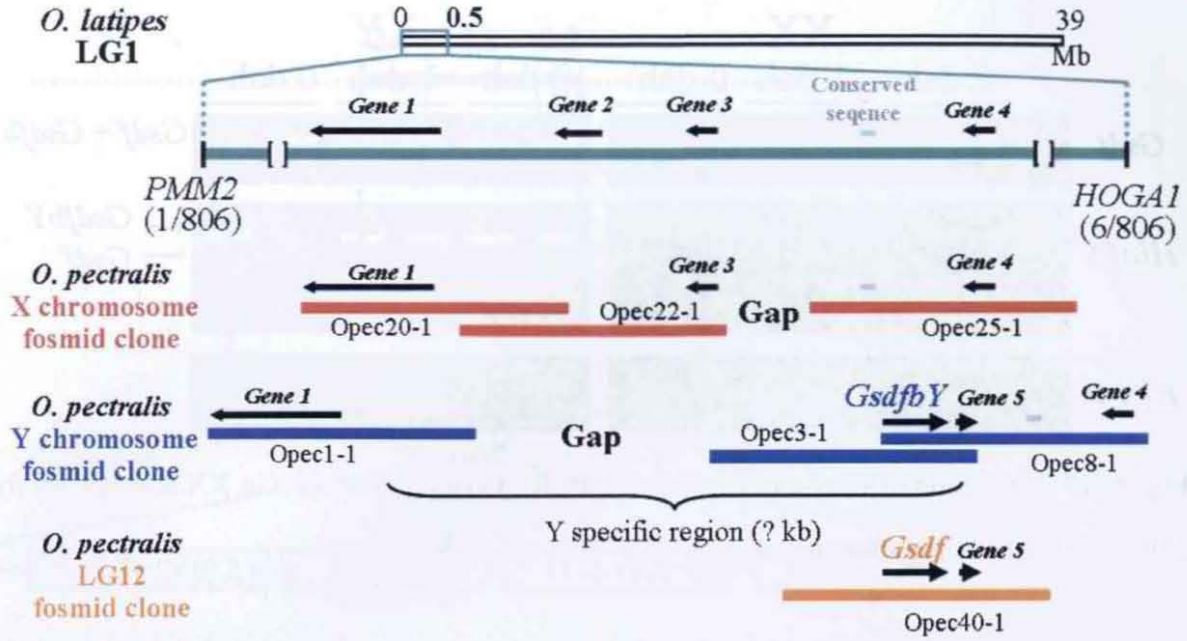

Figure 16 Physical map of the sex-determining regions on the X and Y chromosomes in *O. pectoralis*. Red, blue and orange bar indicate fosmid clones.

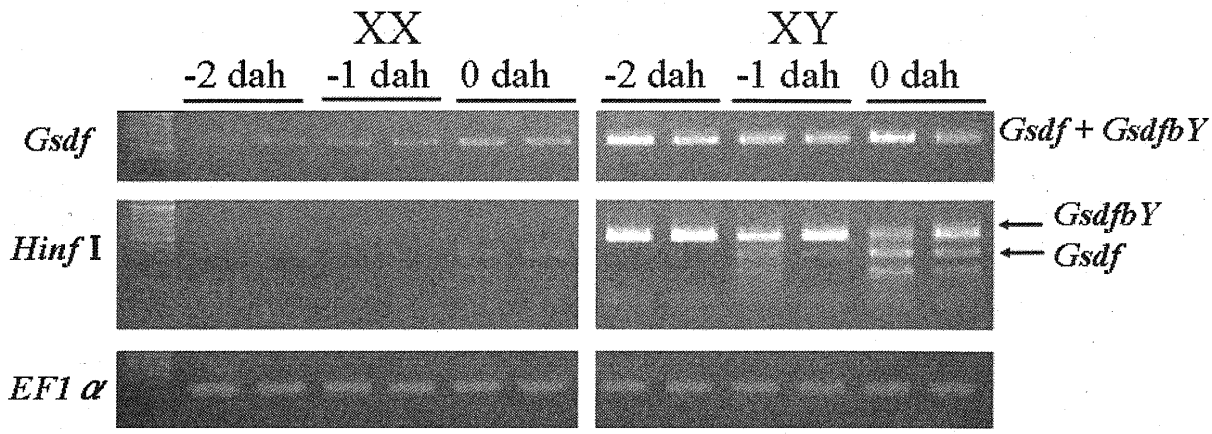

Figure 17 Expression of *Gsdf and GsdfbY.* RT-PCR of *Gsdf* and *GsdfbY* in XX and XY fry from 2 days before hatching (dbh) to hatching day.

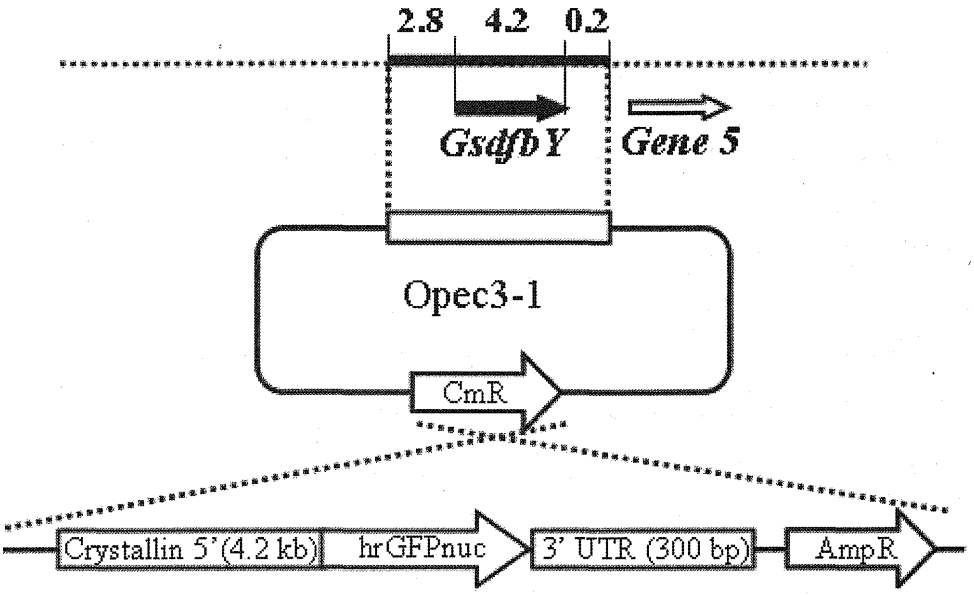

Figure 18 Structures of transgenic construct containing fluorescent markers in Construct 3. This construct includes *GsdfbY*. Numbers above the thick bar indicate the length of the upstream, gene coding and downstream regions of the *GsdfbY* gene in the fosmid construct. A fragment encoding crystallin-GFP and an ampicillin resistance gene was integrated into the ORF of the chloramphenicol resistance gene.

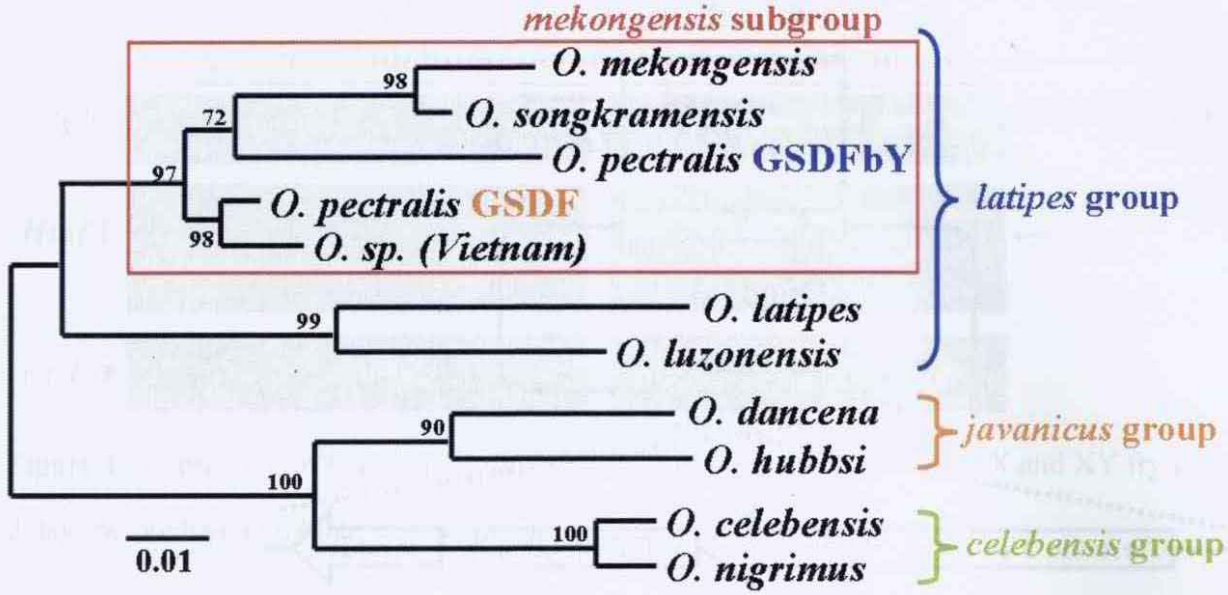

Figure 19 Neighbour-joining (NJ) tree for the *Gsdf* based on the amino acid sequence.

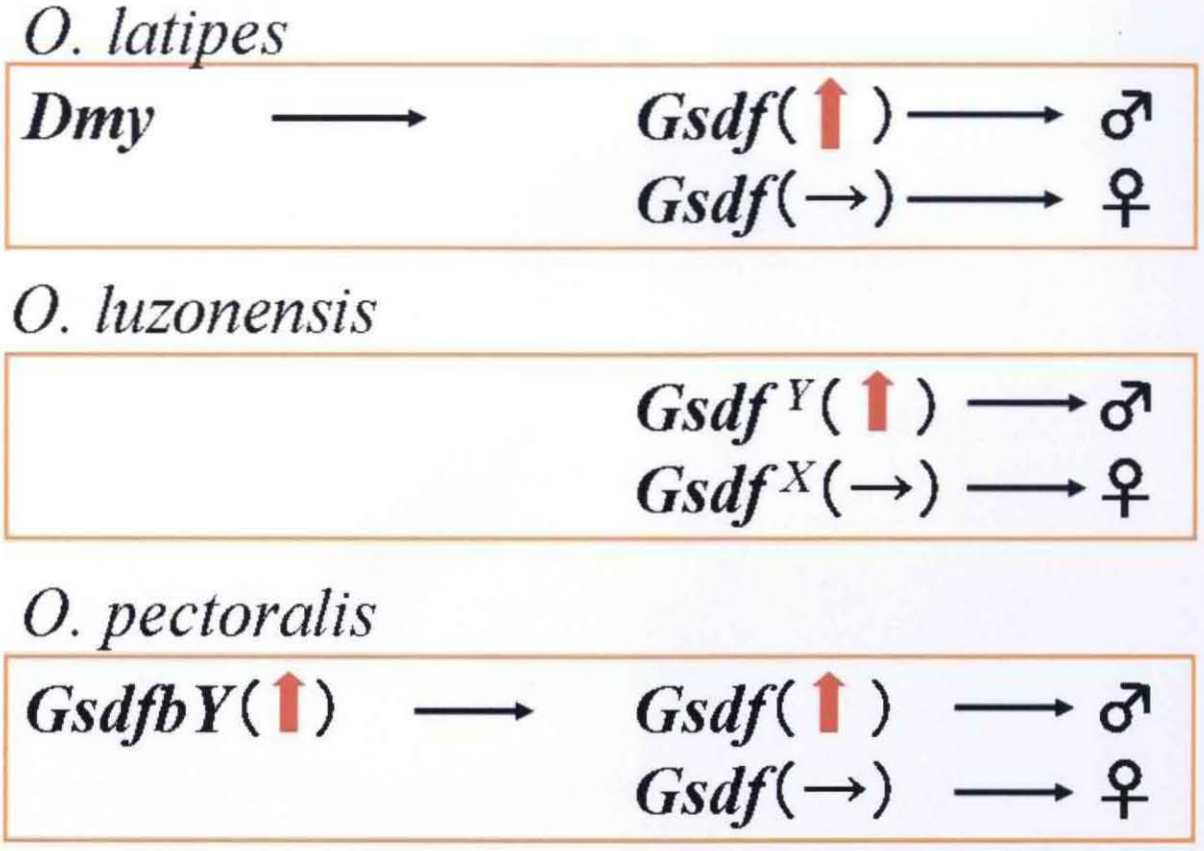

Figure 20 The relationship between three SD genes, *Dmy*, *Gsdf*^{Y} and *GsdfbY*, and autosomal *Gsdf* on SD casucade.