Tracing the Emergence of a Novel Sex-Determining Gene in Medaka, *Oryzias luzonensis*

Taijun Myosho,*,[1] Hiroyuki Otake,* Haruo Masuyama,[†] Masaru Matsuda,[†] Yoko Kuroki,[‡] Asao Fujiyama,[§,] Kiyoshi Naruse,[††] Satoshi Hamaguchi,* and Mitsuru Sakaizumi***

*Institute of Science and Technology, Niigata University, Niigata 950-2181, Japan, [†]Center for Bioscience Research and Education, Utsunomiya University, Utsunomiya 321-8505, Japan, [‡]Laboratory for Immunogenomics, Research Center for Allergy and Immunology, RIKEN Yokohama Institute, Tsurumi, Yokohama, Kanagawa 230-0045, Japan, [§]Comparative Genomics Laboratory, National Institute of Genetics, Mishima, Shizuoka 411-8540, Japan, **Department of Informatics, Graduate University for Advanced Studies, and National Institute of Informatics, Chiyoda-ku, Tokyo 101-8430, Japan, and [††]Laboratory of Bioresources, National Institute for Basic Biology, Okazaki 444-8585, Japan

ABSTRACT Three sex-determining (SD) genes, *SRY* (mammals), *Dmy* (medaka), 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. Here, we used positional cloning to search for the sex-determining gene in *Oryzias 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. We 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 individuals 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*.

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 (SD) 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 SD gene identified in a nonmammalian 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 5A).

O. 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

doi: 10.1534/genetics.111.137497
Manuscript received December 5, 2011; accepted for publication February 1, 2012
Supporting information is available online at http://www.genetics.org/content/suppl/2012/02/23/genetics.111.137497.DC1.
[1]Corresponding author: Institute of Science and Technology, Niigata University, Niigata 950-2181, Japan. E-mail: f09m004a@mail.cc.niigata-u.ac.jp

hypothesis that *Dmy* lost its SD function and disappeared after a new SD gene appeared in *O. luzonensis*. *O. 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

O. 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 subcenter (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 (Supporting Information, Table S1). PCR conditions were as follows: 5 min at 95°, followed by 35 cycles of 20 s at 95°, 30 s at 58°, and 1 min at 72°, followed by 5 min at 72°.

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 (ver.) 0.000925.c, Crossmatch ver. 0.990319, and Phrap ver. 0.990319 (Codon Code), as well as PCP ver. 2.1.6 and Cap4 ver. 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°; 15 min at 95°; cycles of 20 s at 96°, 30 s at 55°, and 60 s at 72°; and 5 min at 72°. 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 S1).

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 $Gsdf^Y$ (gonadal soma derived growth factor on the Y chromosome), primers were designed to examine the expression profiles of $Gsdf^X$ and $Gsdf^Y$ (Table S1). 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°, 10 s at 95°, and then 40 cycles of 5 s at 95° and 30 s at 65°.

In situ hybridization

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

Transgenic constructs

We made two constructs to obtain transgenic lines. First, we 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, 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, we obtained the second construct (construct 2), which contained only $Gsdf^Y$, by using In-Fusion (Takara Bio) methods. We 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. We 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. We 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° 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 S1). 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° in DMEM (Invitrogen) supplemented with 10% fetal bovine serum (Gibco); 2.5×10^4 cells were plated in each well of 96-well plates 24 hr 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 hr, luciferase assay was performed with the Dual-Glo Luciferase Reporter Assay system (Promega) 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). We 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).

We constructed a BAC library of an XX fish and a fosmid library of a YY fish, and we 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 ∼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 for male-specific high expression during sex differentiation

To examine whether the *predicted genes* (*PGs*) are expressed during sexual differentiation, we performed RT–PCR for

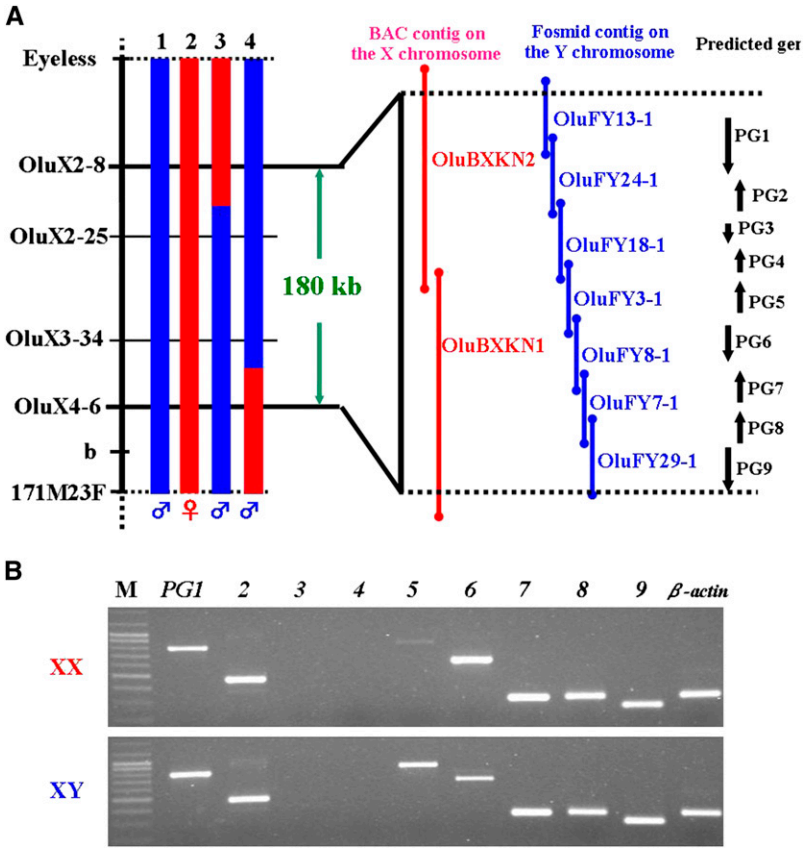

Figure 1 (A) Genetic and physical map of the sex-determining regions on the X and Y chromosomes. 1, normal XY male; 2, XX female; 3 and 4, recombinant male. Blue column, regions derived from the Y chromosome; red, from the X chromosome. Red bar, BAC clones; blue bar, fosmid clones. Black arrows indicate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

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.

We 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 we compared *Gsdf* on the X chromosome ($Gsdf^X$) with that on the Y chromosome ($Gsdf^Y$), we 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 $Gsdf^Y$, we 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 4E).

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

We performed overexpression experiments using a $Gsdf^Y$ genomic clone. First, we 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 S1A). In generation zero (G0), we obtained 54 adult fish with the transgene, one of which was a sex-reversed XX male (Table 1). We 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, we established a transgenic strain (strain 1) whose sex was determined by the transgene construct 1. Next, we made a construct (construct 2) that contained 3.5 kb of $Gsdf^Y$, as

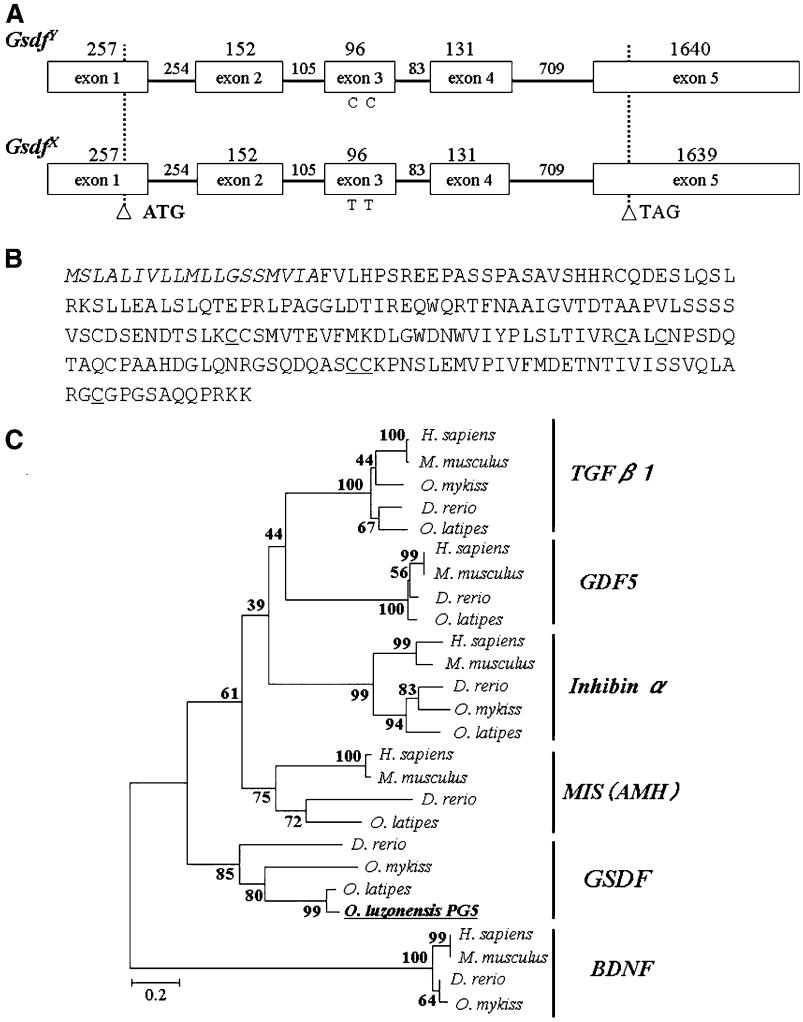

Figure 2 (A) Genomic structure of *Gsdf*^{Y} and *Gsdf*^{X}. Two synonymous substitutions are present in exon 3. Open boxes, exons. Horizontal bars, introns. Numbers represent nucleotide sequence length (bp). Open arrowheads indicate the translation start (ATG) and stop (TGA) sites. (B) Amino acid sequences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

well as 1.8 kb of its upstream region and 2 kb of its downstream region, but no other predicted genes (Figure S1B). As with the previous transgenic experiment, we established a strain (strain 2) whose sex was determined by the transgene (Table 1). To confirm the mRNA expression of both strains, we 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*

We 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 harbors evolutionarily conserved *cis*-regulatory motifs among fish species. To find these sequences, we compared 1.8 kb upstream and 2.0 kb downstream of *Gsdf*^{Y} with those of *Gsdf*^{X} and *Gsdf* in *O. latipes*. We 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 4, A and B). We 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 4C). Luciferase activity was significantly decreased in recombinant constructs Luc 3–6, 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, we 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 ortholog of *Gsdf*$^{X/Y}$ is located on an autosome (LG12). *Gsdf* in XY fish shows significantly higher expression levels compared with that in XX fish during sex differentiation, suggesting that expression levels of *Gsdf* are 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, we established an *O. latipes* strain (strain 3)

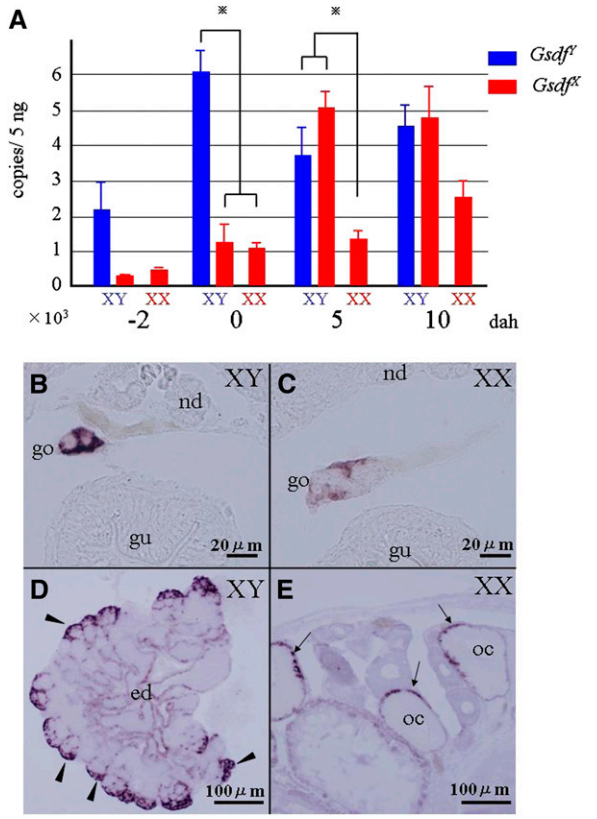

Figure 3 Expression of *Gsdf*. (A) Real-time PCR of *Gsdf* in XX and XY fry from 2 days before hatching (dbh) to 10 days after hatching (dah). Blue bars, expression of *PG5* on the Y chromosome (*Gsdf*^{Y}); red, the X chromosome (*Gsdf*^{X}). Significant differences were analyzed by using two-way analysis 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.

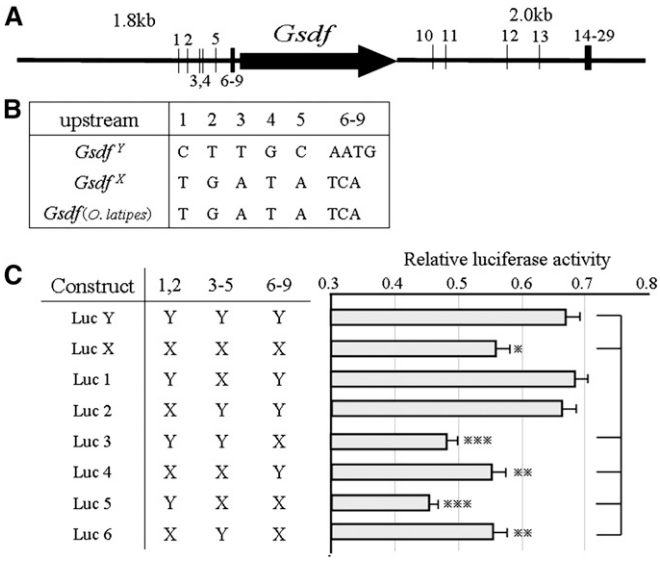

Figure 4 Mutations in the *Gsdf cis*-regulatory element contribute to *Gsdf*^{Y}-specific high expression. (A) Illustration of *Gsdf*^{Y}-specific mutations in the 1.8-kb upstream and 2-kb downstream regions of *Gsdf*^{Y}, comparing with *Gsdf*^{X} and *Gsdf* (*O. latipes*). Numbers represent positions of *Gsdf*^{Y}-specific mutations. (B) *Gsdf*^{Y}-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 ($n = 12$).

whose sex was determined by *Gsdf*^{Y} from *O. luzonensis* (Table 1). 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

Our 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*, have been identified (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 nonpiscine 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 (Sawatari *et al.* 2007; 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 our reporter assay suggest that *cis*-regulatory sequences of *Gsdf*^{Y} are involved in higher expression of the gene in males (Figures 3 and 4). *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 1). 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

Table 1 Phenotypic sex of transgenic strains

Strain no.	Host species/construct	Generation	Presence of transgene	XX ♀	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

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*.

Willkins 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 5, 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 $Gsdf^Y$. We 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.

In *Oryzias* fishes, >20 extant species are recognized, and the sex chromosomes of 7 species have been identified; *O. latipes* (LG1), *O. curvinotus* (LG1), *O. luzonensis* (LG12), *O. minutillus* (LG8), *O. dancena* (LG10), *O. hubbsi* (LG5), and *O. javanicus* (LG16) (Takehana *et al.* 2007a,b, 2008; Nagai *et al.* 2008). *Dmy* has been detected only in *O. latipes* and *O. curvinotus*. Since sex chromosomes homologous to LG12 have not evolved repeatedly, *Gsdf* cannot be the SD gene in the 4 remaining species. We are now examining the role of *Gsdf* on the SD cascade in these species.

Acknowledgments

We thank C. L. Peichel for constructive comments on the manuscript, Y. Takehana and A. Shinomiya for advice, and the Sequencing Technology Team at the RIKEN Genomic Sciences Center for BAC and fosmid clone sequencing. This work was supported by KAKENHI via Grants-in-Aid for

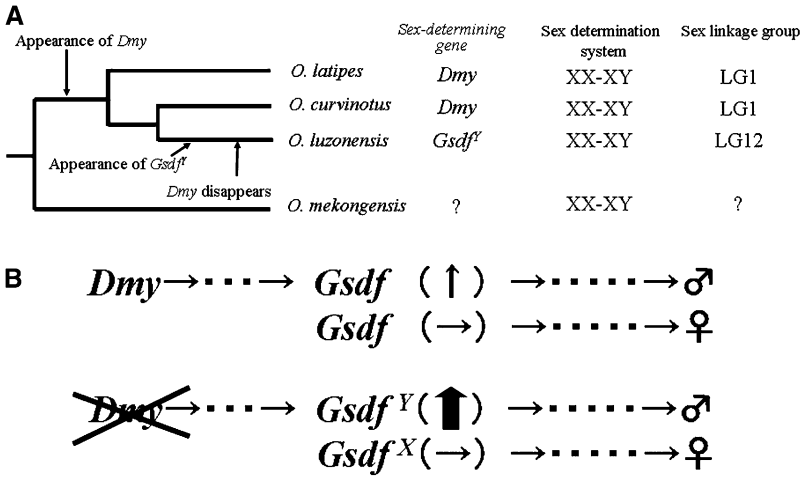

Figure 5 Changes in the sex-determining gene of *O. luzonensis*. (A) Evolutionary history up to the appearance of $Gsdf^Y$. *Dmy* appeared as a common ancestor of *O. latipes*, *O. curvinotus*, and *O. luzonensis*. $Gsdf^Y$ appeared in, and *Dmy* disappeared from, the ancestor of *O. luzonenesis*. (B) Change in the sex-determining cascade. (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.

Scientific Research on the priority area "Comparative Genomics" from the Ministry of Education, Culture, Sports, Science, and Technology of Japan and via Grants-in-Aid for Scientific Research on the priority area "Scientific Research B" (16370094 and 20370086).

Literature Cited

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: 1240–1251.

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. Zoolog. 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., U. Hornung, I. Nanda, 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*. Trends 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*. Zoolog. 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: 137–142.

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.

Pomiankowski, A., R. Nothiger, and A. Willkins, 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. Zoolog. 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., 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*. 2007b. Chromosoma 116: 463–470.

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*. 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.

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.

Communicating editor: D. Charlesworth

GENETICS

Supporting Information
http://www.genetics.org/content/suppl/2012/02/23/genetics.111.137497.DC1

Tracing the Emergence of a Novel Sex-Determining Gene in Medaka, *Oryzias luzonensis*

Taijun Myosho, Hiroyuki Otake, Haruo Masuyama, Masaru Matsuda, Yoko Kuroki, Asao Fujiyama, Kiyoshi Naruse, Satoshi Hamaguchi, and Mitsuru Sakaizumi

DOI: 10.1534/genetics.111.137497

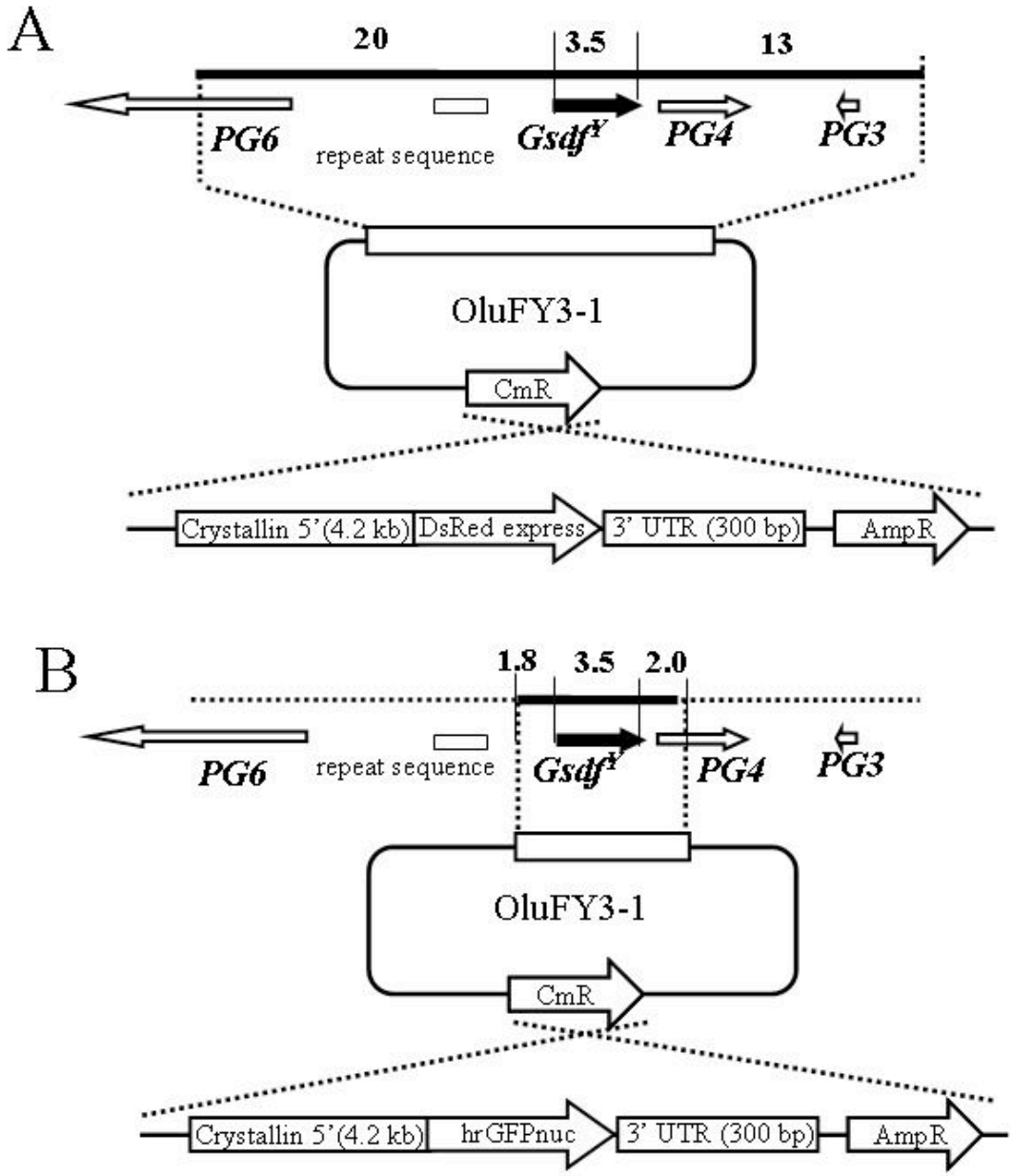

Figure S1 Structures of transgenic constructs containing fluorescent markers. (A) Construct 1. This construct includes *Gsdf*^{Y}, *PG3* and *PG4*. Numbers above the thick bar indicate the length of the upstream, gene coding and downstream regions of the *Gsdf*^{Y} gene in the fosmid construct. A fragment 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.

Table S1 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