Functional analysis evolution history of gonadal soma-derived factor (*gsdf*) in black rockfish, *Sebastes schlegelii*

Xuemei Li[1], Xuliang Wang[1], Zhigang Wang[1]*

[1]*MOE Key Laboratory of Marine Genetics and Breeding, College of Marine Life Sciences, Ocean University of China, 5 Yushan Road, Qingdao 266003, China*

Correspondence: zgwang@ouc.edu.cn ; Tel.: +86-532-82032720

Abstract

As a teleost- and gonad-specific growth factor, gsdf has been indicated to play an important role in sex differentiation and determination. In this study, the complete open reading frame (ORF) of *gsdf* was isolated from black rockfish, *Sebastes schlegelii*. Bioinformatics analysis showed that there is a conserved transforming growth factor-β (TGF-β) domain located on the C-terminus of Gsdf. Multiple sequence alignments revealed that fish *gsdf* were highly conserved in TGF-β domain which suggested their functional conservatism. Synteny analysis provided evidence for the hypothesis that gsdf was originated from fish-specific genome duplication (FSGD). To further explore its function, the expression pattern was examined based on the RNA-seq data and the result showed that significantly sexually dimorphic expression existing between male and female individuals. These results suggested that *gsdf* might play an important role in maintenance of male characteristics in *Sebastes schlegelii*.

Keywords: *gsdf*, black rockfish, evolution history, gene expression

1. Introduction

Over the past decades, many studies have focused on sex differentiation/determination systems in vertebrates. It has been revealed that *SRY/sry*, the sex-linked testis-determining gene, triggers male differentiation in most mammals (1, 2). *Dmrt1* was found as an important sex-determination gene in several species such as *Gallus gallus* and *Cynoglossus semilaevis* (3, 4). In *Oryzias latipes*, the sex-determination gene is *DMY* (5, 6). Besides, the members of transforming growth factor-β (TGF-β) family also contribute to sex-determination in fish (7-9). For example, *Amhy*, the Y-linked replication of *Amh*, function as the sex-determination gene in *Odontesthes hatcheri* and *Oreochromis niloticus* which demonstrated the significant role of *Amh/AmhR2* signaling pathway in sex determination in fish (8, 10). The *gsdf* gene belongs to TGF-β family, which was found only in teleost fish (11-13). In general, it is predominantly expressed in Sertoli cells and surrounding cells in mature gonads, probably with some lineage-specific function (9, 11, 12). Recent study has confirmed that *gsdf* play an important role in the process of fish reproduction and development. In *Oncorhynchus mykiss*, *gsdf* could enhance primordial germ cell and spermatogonial proliferation (11). Moreover, *gsdf* transcription is activated directly by *dmy*, which established the autosomal *gsdf* as the first male sex initiator in *Oryzias latipes* (14). And in *Cynoglossus semilaevis*, the autosomal *gsdf* gene play a positive role in germ differentiation and proliferation via influencing genes related to sex

differentiation (15).

Black rockfish (*Sebastes schlegelii*), an economic fish species, is cultured worldwide. During the grow-out period, the growth rate of females is substantially faster than that of males under the same culturing condition. Therefore, sex-controlled breeding is very important and it is very meaningful to study mechanisms of gonadal sex differentiation in this species. In the present study, a *gsdf* ortholog was isolated from black rockfish and its distribution pattern in tissues was detected. Also, synteny analysis was carried to discuss the origin of *gsdf*. Our results will facilitate to understand the function of *gsdf* in black rockfish and help to the sex-controlled breeding in the future.

2. Methods

2.1 Ethics statement

Black rockfish was obtained from a commercial hatchery. This study was conducted in accordance with the Institutional Animal Care and Use Committee of the Ocean University of China and the China Government Principles for the Utilization and Care of Vertebrate Animals Used in Testing, Research, and Training (State science and technology commission of the People's Republic of China for No. 2, October 31, 1988. http://www.gov.cn/gongbao/content/2011/content_1860757.htm).

2.2 Sampling and RNA extraction

In this study, all healthy fishes were anesthetized and killed by severing spinal cord.

Organs, including heart, liver, spleen, kidney, brain, gill, muscle, intestine, and gonad, were collected from each fish. Samples were snap-frozen in liquid nitrogen and stored at -80 ° C until use.

Total RNA was extracted using TRIzol Reagent (Invitrogen, USA) according to the manufacturer's instructions, A total of 1 μg RNA from each sample was reverse-transcribed according to the instructions of the PrimeScript™ RT reagent kit with the gDNA Eraser (Takara, Dalian, China). The final volume was set at 20 μL. The total RNA was evaluated qualitatively and quantitatively by 1.5% agarose gel electrophoresis and spectrophotometry with the NanoPhotometer Pearl. RNA was used for RNA-seq.

2.3 Sequence identification and bioinformatics analysis

The cDNA sequence of *Ssgsdf* was obtained using local BLAST from transcriptome. DNASTAR was used to analyze putative amino acid sequence, calculated molecular weight, and theoretical isoelectric point. The signal peptide was analyzed by SignalP v4.0 program. Protein domains were predicted using the Simple Modular Architecture Research Tool (16, 17).

2.4 Multiple sequence alignments and phylogenetic analysis

All the other sequences were downloaded from NCBI websites (Table S1). Alignment of putative amino acid sequences of black rockfish and other known vertebrates was carried out by clustalX2 with the default parameters (18). Phylogenetic tree was constructed by neighbor-joining method and a bootstrap test with 1000 replicates

carried by MEGA 7.0 (19).

2.5 Synteny analysis

Synteny comparisons of the fragments harboring *gsdf* and flanking genes were performed to test the genes' syntenic conservation. Flanking genes of *gsdf* used in the synteny analysis were extracted from online genome databases. The genes were mapped according to their relative locations in the chromosome for the synteny analysis.

2.6 Expression analysis

To study the potential functions of *Ssgsdf*, the expression of *Ssgsdf* mRNA was analyzed. Based on the RNA-seq data, the counts of mapped reads were used to compute expression values as Transcript Per Million (TPM), to provide a reliable comparison of highly heterogeneous samples. The expression of *Ssgsdf* mRNA in different tissues were analyzed by TPM scores calculated from RNA-seq data.

3. Results

3.1 Sequence identification and analysis of Ssgsdf

The *gsdf* cDNA sequence of black rockfish was retrieved from the transcriptome library. As shown in Fig. 1A, *gsdf* contains a 648 bp open reading frame (ORF) encoding 216 amino acid residues. Black rockfish Gsdf (SsGsdf) has a calculated molecular mass of 22.77 kDa and a theoretical pI of 5.06. An N-terminal signal peptide formed by residues 1 to 19 was predicted by SignalP v4.0 program. Protein

106 domain prediction by SMART showed that the mature protein of *Ssgsdf* contains a
107 conserved TGF-β domain formed by residues 114-205 (Fig. 1B).
108 *3.2 Multiple sequence alignments and phylogenetic analysis of gsdf orthologs*
109 Multiple alignments of amino acid sequences of *Ssgsdf* and other teleost *gsdf* showed
110 that SsGsdf also presented seven conserved cysteine residues (Cys^{114}, Cys^{115}, Cys^{143},
111 Cys^{146}, Cys^{171}, Cys^{172}, Cys^{204}) that could be involved in the formation of the
112 characteristic cysteine knot motif, which is involved in intrachain disulfide bonds or
113 dimerization (20, 21). The phylogenetic tree was constructed by was performed by
114 Bayesian method to show the relationship among different species. Two distinct
115 groups were separated in the phylogenetic tree. SsGsdf was clustered in the same
116 clade with teleost Gsdf. The clade formed by fish Gsdf was seperated from the other
117 clade formed by other members of the TGF-β superfamily, which suggested that gsdf
118 was a unique member of this superfamily.
119 *3.3 Synteny analysis of gsdf orthologs*
120 As shown in Fig. 4, the *gsdf* genes and adjoining genes of several teleost species were
121 placed according to their relative locations on the scaffold or chromosome. All genes
122 near *gsdf* were highly conserved and shared the same direction in Tilapia, Amazon
123 molly, spotted gar and black rockfish. Comparison between three species revealed that
124 eight upstream genes (*gng10*, *btc*, *rxfp*, *slc45a2*, *amacr*, *sacrb2*, *nup54* and *ppef2*) and
125 eight downstream genes (*aff1*, *mrc1*, *klhl8*, *sdad1*, *ptpn13*, *mapk10*, *arhgap24* and
126 *esm1*) were conserved. In black rockfish, the downstream genes, *aff1* and *ptpn13*,

were replicated to produce two copies. Interestingly, only a few genes near *gsdf* gene were found in cave fish, zebrafish and spotted gar. In these three species, the transcriptional direction and the relative position relationship of some adjoining genes were changed.

3.4 Tissue distribution pattern of gsdf mRNA

To analyze the presence of *gsdf* mRNA in black rockfish, we examined the expression levels in ten different tissues of two-year post-hatching female and male individuals by TPM scores. As shown in Fig. 5, large amounts of *gsdf* mRNA was observed in adult testis, whereas the adult ovary expressed small amounts. Almost negligible amplification for *gsdf* was detected in the non-gonadal tissues such as heart, liver, spleen, kidney and gill.

4. Discussion

As a fish specific TGF-β family gene, *gsdf* is well known for its function in male sex determination and differentiation (9, 12, 22-24). In the present study, a *gsdf* ortholog was isolated from *Sebastes schlegelii*, which encoded a 215-residue protein with a single TGF-β domain. The deduced amino acid sequence indicated that it is relatively well conserved in TGF-β domain. The cysteine residues of the TGF-β domain are very conserved, indicating that Gsdf has similar functions in different species. As an important signaling molecule, the actual form of Gsdf is a homo- or heterodimer of a small carboxy-terminal. In TGF-β domain of SsGsdf, three disulfide bonds were

148 formed by Cys^{114}-Cys^{171}, Cys^{115}-Cys^{172} and Cys^{143}-Cys^{204}, and the intrachain disulfide
149 bond was formed by Cys^{146} and a cysteine residue of another molecule (20). Its N-
150 terminal signal peptide indicates that SsGsdf is secreted, suggesting that SsGsdf may
151 be secreted out of the cell and interact with the specific receptor molecules on the
152 surfaces of target cells.
153 During the long history of evolutionary process, many events occurred in fish genome,
154 such as whole-genome duplication (WGD), loss of chromosome fragments, gene
155 rearrangement and others (25, 26). Previous studies revealed that *gsdf* is specific to
156 fish, which suggests that *gsdf* may originate from the fish-specific genome
157 duplication (FSGD) and be selectively dropped in the evolutionary process from fish
158 to tetrapods. Highly conserved genes around *gsdf* suggested that the functions of *gsdf*
159 and its adjoining genes tend to be constant. The different upstream and downstream
160 genes around *gsdf* between fishes suggested that gene rearrangement occurred in fish
161 genome. All these findings indicate the diversity of evolutionary events during the
162 evolutionary process of fish.
163 Up to date, the functions of gsdf were not elucidate clearly. Numerous studies in fish
164 revealed that *gsdf* only expressed in gonadal tissues, which suggested that its
165 functions may be restricted to the gonads. In this study, *Ssgsdf* mRNA was only
166 expressed in the gonads, and the expression level in testis was significantly higher
167 than in the ovary. The expression pattern has obvious gender dimorphism. This result
168 was consistent with those in *Oncorhynchus mykiss*, *Oryzias latipes* and *Cynoglossus*

semilaevis (11, 12, 15). The distinct gonad expression pattern in fish indicated that *gsdf* may have a particular impact on gonad during the period of sex differentiation and maintenance male characteristics.

In this study, *gsdf* was isolated and verified from *Sebastes schlegelii*. Bioinformatic analysis revealed that gsdf was conserved in terms of potential domains and synteny relationship. The expression pattern showed *Sebastes schlegelii gsdf* existed obvious sexual dimorphism in adult. These results indicate that *Sebastes schlegelii gsdf* may have essential functions in maintenance of male characteristics. Further studies are necessary to illustrate the initiation mechanism of testis differentiation, which is related to the regulatory mechanisms of gsdf expression.

References

1. Berta P, Hawkins JR, Sinclair AH, Taylor A, Griffiths BL, Goodfellow PN, et al. Genetic evidence equating SRY and the testis-determining factor. Nature. 1990;348(6300):448-50. doi: 10.1038/348448A0.

2. Sinclair AH, Berta P, Palmer MS, Hawkins JR, Griffiths BL, Smith MJ, et al. A gene from the human sex-determining region encodes a protein with homology to a conserved DNA-binding motif. Nature. 1990;346(6281):240-4. doi: 10.1038/346240a0.

3. Smith CA, Roeszler KN, Ohnesorg T, Cummins DM, Farlie PG, Doran TJ, et al. The avian Z-linked gene DMRT1 is required for male sex determination in the chicken. Nature. 2009;461(7261):267-71.

4. Chen S, Zhang G, Shao C, Huang Q, Liu G, Zhang P, et al. Whole-genome sequence of a flatfish

provides insights into ZW sex chromosome evolution and adaptation to a benthic lifestyle. Nat Genet. 2014;46(3):253-60. doi: 10.1038/ng.2890.

5. Matsuda M, Nagahama Y, Shinomiya A, Sato T, Matsuda C, Kobayashi T, et al. DMY is a Y-specific DM-domain gene required for male development in the medaka fish. Nature. 2002;417(6888):559-63. doi: 10.1038/nature751.

6. Nanda I, Kondo M, Hornung U, Asakawa S, Winkler C, Shimizu A, et al. A duplicated copy of DMRT1 in the sex-determining region of the Y chromosome of the medaka, *Oryzias latipes*. Proc Natl Acad Sci U S A. 2002;99(18):11778-83. doi: 10.1073/pnas.182314699.

7. Kamiya T, Kai W, Tasumi S, Oka A, Matsunaga T, Mizuno N, et al. A trans-species missense SNP in Amhr2 is associated with sex determination in the tiger pufferfish, *Takifugu rubripes* (fugu). PLoS Genet. 2012;8(7):e1002798. doi: 10.1371/journal.pgen.1002798.

8. Hattori RS, Murai Y, Oura M, Masuda S, Majhi SK, Sakamoto T, et al. A Y-linked anti-Mullerian hormone duplication takes over a critical role in sex determination. Proc Natl Acad Sci U S A. 2012;109(8):2955-9. doi: 10.1073/pnas.1018392109.

9. Myosho T, Otake H, Masuyama H, Matsuda M, Kuroki Y, Fujiyama A, et al. Tracing the emergence of a novel sex-determining gene in medaka, *Oryzias luzonensis*. Genetics. 2012;191(1):163-70. doi: 10.1534/genetics.111.137497.

10. Li M, Sun Y, Zhao J, Shi H, Zeng S, Ye K, et al. A Tandem Duplicate of Anti-Mullerian Hormone with a Missense SNP on the Y Chromosome Is Essential for Male Sex Determination in Nile Tilapia, *Oreochromis niloticus*. PLoS Genet. 2015;11(11):e1005678. doi: 10.1371/journal.pgen.1005678.

11. Sawatari E, Shikina S, Takeuchi T, Yoshizaki G. A novel transforming growth factor-beta

superfamily member expressed in gonadal somatic cells enhances primordial germ cell and spermatogonial proliferation in rainbow trout (*Oncorhynchus mykiss*). Dev Biol. 2007;301(1):266-75. doi: 10.1016/j.ydbio.2006.10.001.

12. Shibata Y, Paul-Prasanth B, Suzuki A, Usami T, Nakamoto M, Matsuda M, et al. Expression of gonadal soma derived factor (GSDF) is spatially and temporally correlated with early testicular differentiation in medaka. Gene Expr Patterns. 2010;10(6):283-9. doi: 10.1016/j.gep.2010.06.005.

13. Gautier A, Le Gac F, Lareyre JJ. The gsdf gene locus harbors evolutionary conserved and clustered genes preferentially expressed in fish previtellogenic oocytes. Gene. 2011;472(1-2):7-17. doi: 10.1016/j.gene.2010.10.014.

14. Zhang X, Guan G, Li M, Zhu F, Liu Q, Naruse K, et al. Autosomal gsdf acts as a male sex initiator in the fish medaka. Sci Rep. 2016;6:19738. doi: 10.1038/srep19738 (2016).

15. Zhu Y, Meng L, Xu W, Cui Z, Zhang N, Guo H, et al. The autosomal Gsdf gene plays a role in male gonad development in Chinese tongue sole (*Cynoglossus semilaevis*). Sci Rep. 2018;8(1):17716. doi: 10.1038/s41598-018-35553-7.

16. Letunic I, Doerks T, Bork P. SMART: recent updates, new developments and status in 2015. Nucleic Acids Res. 2015;43(Database issue):D257-60. doi: 10.1093/nar/gku949.

17. Letunic I, Bork P. 20 years of the SMART protein domain annotation resource. Nucleic Acids Res. 2017;46(D1):D493-D6. doi: 10.1093/nar/gkx922.

18. Chenna R, Sugawara H, Koike T, Lopez R, Gibson TJ, Higgins DG, et al. Multiple sequence alignment with the Clustal series of programs. Nucleic Acids Res. 2003;31(13):3497-500. doi: 10.1093/nar/gkg500.

19. Kumar S, Stecher G, Tamura K. MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. Mol Biol Evol. 2016;33(7):1870-4. doi: 10.1093/molbev/msw054.

20. Daopin S, Piez KA, Ogawa Y, Davies DR. Crystal structure of transforming growth factor-beta 2: an unusual fold for the superfamily. Science. 1992;257(5068):369-73. PMID: 1631557.

21. Kingsley DM. The TGF-beta superfamily: new members, new receptors, and new genetic tests of function in different organisms. Genes & development. 1994;8(2):133-46. PMID: 8299934.

22. Chen Y, Hong WS, Wang Q, Chen SX. Cloning and pattern of gsdf mRNA during gonadal development in the protogynous *Epinephelus akaara*. Anim Reprod Sci. 2016;165:46-55. doi: 10.1016/j.anireprosci.2015.12.004.

23. Kaneko H, Ijiri S, Kobayashi T, Izumi H, Kuramochi Y, Wang DS, et al. Gonadal soma-derived factor (gsdf), a TGF-beta superfamily gene, induces testis differentiation in the teleost fish *Oreochromis niloticus*. Mol Cell Endocrinol. 2015;415:87-99. doi: 10.1016/j.mce.2015.08.008.

24. Liu Y, Zhang W, Du X, Zhao J, Liu X, Li X, et al. Sexually dimorphic expression in developing and adult gonads shows an important role of gonadal soma-derived factor during sex differentiation in olive flounder (*Paralichthys olivaceus*). Comp Biochem Physiol B Biochem Mol Biol. 2017;210:1-8. doi: 10.1016/j.cbpb.2017.05.003.

25. Meyer A, Schartl M. Gene and genome duplications in vertebrates: the one-to-four (-to-eight in fish) rule and the evolution of novel gene functions. Curr Opin Cell Biol. 1999;11(6):699-704. PMID: 10600714.

26. Meyer A, Van de Peer Y. From 2R to 3R: evidence for a fish-specific genome duplication (FSGD). Bioessays. 2005;27(9):937-45. doi: 10.1002/bies.20293.

Figures

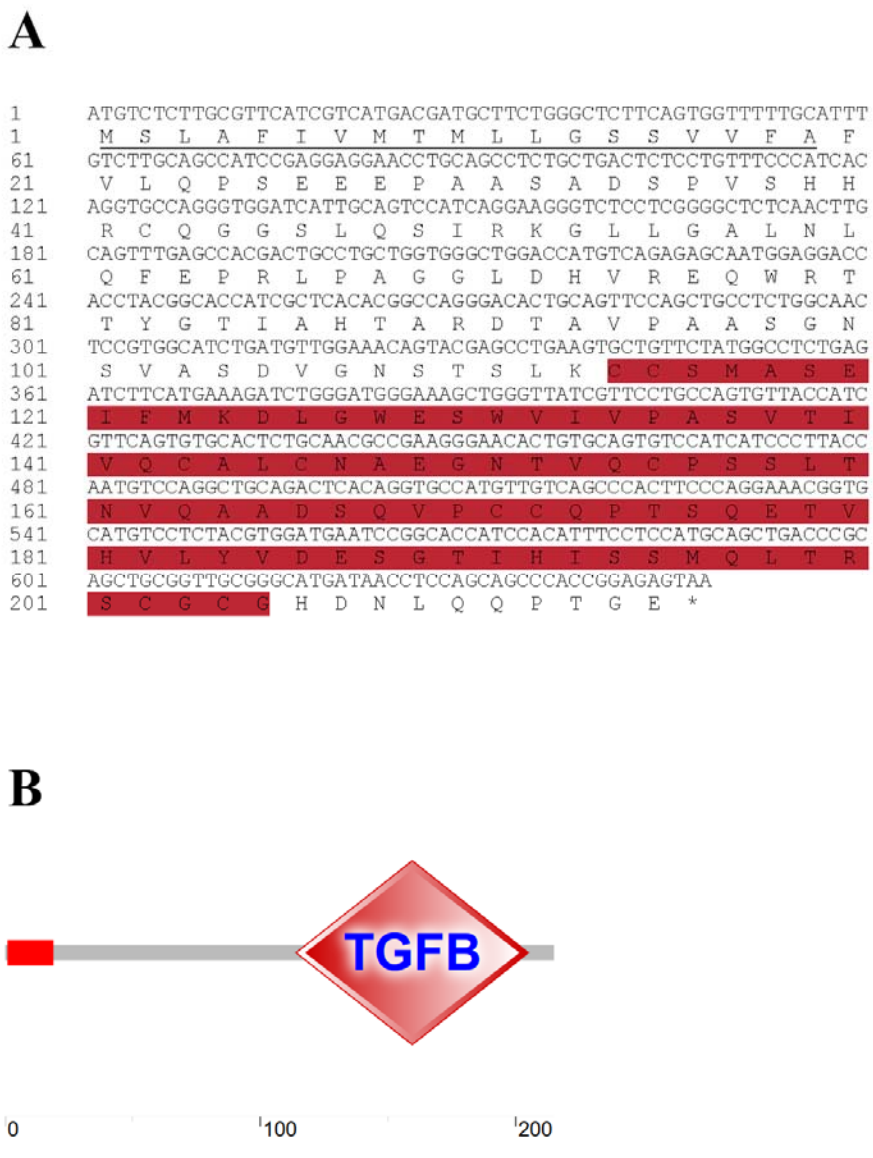

Fig. 1. A: Nucleotide sequence of *Ssgsdf* cDNA and deduced amino acid sequence.

The N-terminus signal peptide is underlined. The TGF-β region is shown in brownish red. The numbers represent nucleotides and amino acids, respectively. B: The domains of SsGsdf predicted by SMART program. The red box represents the signal peptide. The TGF-β region is shown in brownish red.

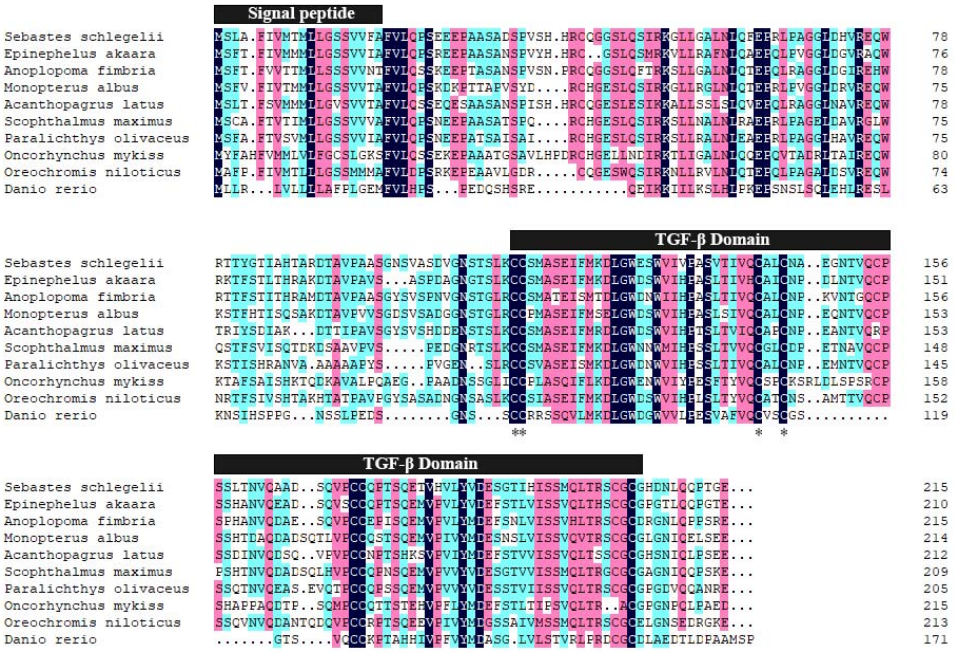

Fig. 2. Multiple alignment of deduced amino acid sequences of *gsdf* from *Sebastes schlegelii* and other fishes. The conserved cysteine residues are marked with asterisks. Accession numbers are as follows: *Epinephelus akaara* (AIW52566.1), *Anoplopoma fimbria* (AGR33990.1), *Monopterus albus* (ALG62631.1), *Acanthopagrus latus* (AIW52571.1), *Scophthalmus maximus* (AJO67894.1), *Paralichthys olivaceus* (ARH56437.1), *Oncorhynchus mykiss* (NP_001118051.1), *Oreochromis niloticus* (BAJ78985.1), *Danio rerio* (ABZ01522.1).

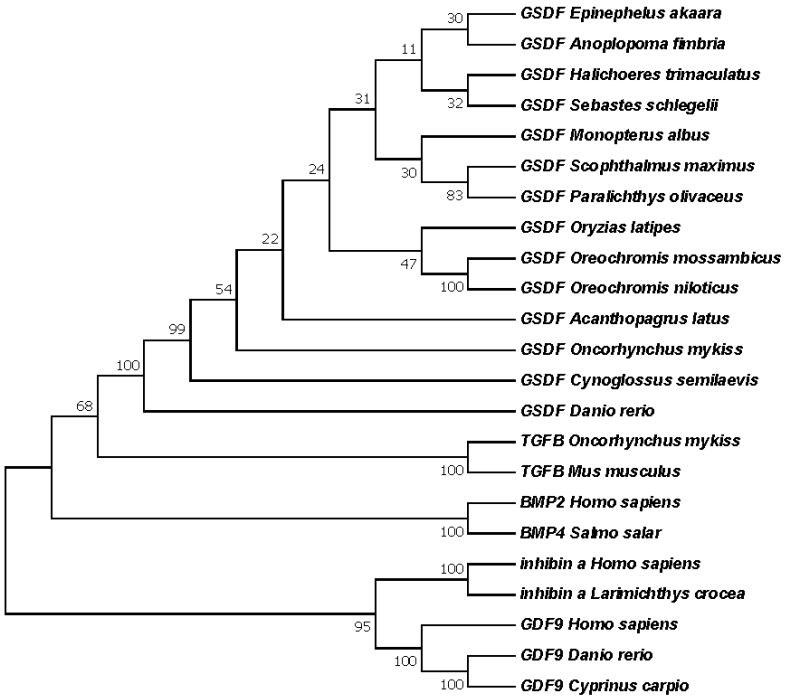

Fig. 3. Construction of phylogenetic tree with the protein sequences of *gsdf* and other genes of TGF-β family. The phylogenetic tree was drawn by neighbor-joining method replicates based on multiple sequence alignment by ClustalW and a bootstrap test with 1000.

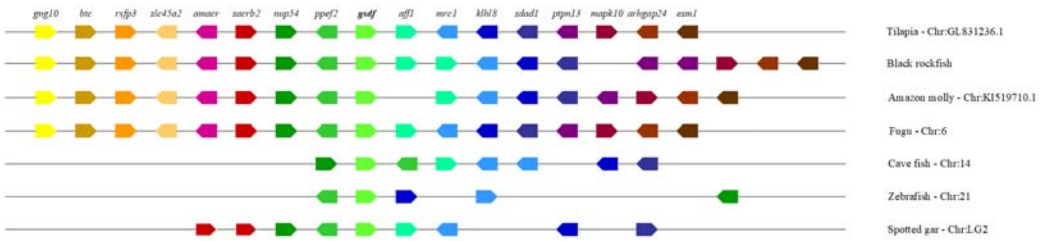

Fig. 4. Chromosomal segments showing the synteny of *gsdf* in teleost. Different genes are represented by different colored pentagons and gene order is determined according to their relative positions in the chromosome or scaffold; the gene names are placed on top of the pentagons. The direction of pentagons indicate the gene direction, the vertical lines represent noncontiguous regions on the scaffold or chromosome.

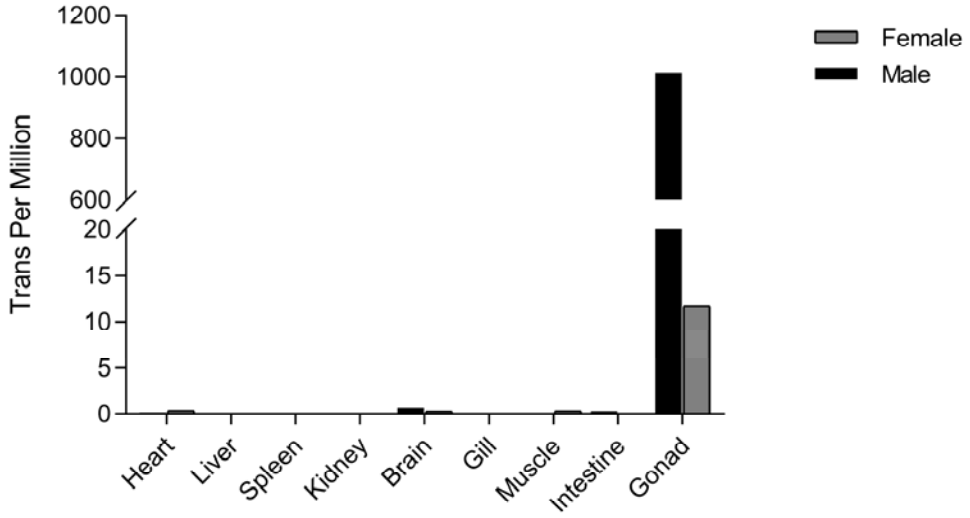

Fig. 5. *Ssgsdf* gene expression in different tissues. The expression level was measured with TPM scores. Numbers represent the mean value of TPM calculated with RNA-seq data of six individuals.

Table S1 Sequences information downloaded from NCBI

Gene name	Species	Accession number
GSDF	*Epinephelus akaara*	AIW52566.1
GSDF	*Anoplopoma fimbria*	AGR33990.1
GSDF	*Monopterus albus*	ALG62631.1
GSDF	*Halichoeres trimaculatus*	BAM75186.1
GSDF	*Acanthopagrus latus*	AIW52571.1
GSDF	*Scophthalmus maximus*	AJO67894.1
GSDF	*Oreochromis mossambicus*	ALO18792.1
GSDF	*Paralichthys olivaceus*	ARH56437.1
GSDF	*Cynoglossus semilaevis*	AYP19379.1
GSDF	*Oryzias latipes*	NP_001171213.1

GSDF	*Oncorhynchus mykiss*	NP_001118051.1
GSDF	*Oreochromis niloticus*	BAJ78985.1
GSDF	*Danio rerio*	ABZ01522.1
BMP2	*Homo sapiens*	NP_001191.1
BMP4	*Salmo salar*	NP_001133316.1
inhibin a	*Homo sapiens*	CAA01158.1
inhibin a	*Larimichthys crocea*	XP_027147478.1
GDF9	*Homo sapiens*	EAW62309.1
GDF9	*Danio rerio*	AAV91155.1
GDF9	*Cyprinus carpio*	AOW71519.1

315