

Supplementary Materials: Evolutionary Conservation of *pou5f3* Genomic Organization and Its Dynamic Distribution during Embryogenesis and in Adult Gonads in Japanese Flounder *Paralichthys olivaceus*

Jinning Gao, Quanqi Zhang and Xubo Wang

Table S1. The Genbank accession numbers or Ensembl IDs of the POUV proteins analyzed in Figure 1C.

Species	Genbank Accession No./ Ensembl IDs	Species	Genbank Accession No./ Ensembl IDs
<i>Acipenser sinensis</i>	AEK81554.1	<i>Pan troglodytes</i>	NP_001238970.1
<i>Anolis carolinensis</i>	ENSACAP00000017001	<i>Ornithorhynchus anatinus</i>	NP_001229656.1
<i>Bos taurus</i>	NP_777005.1	<i>Oryzias latipes</i>	NP_001098339.1
<i>Carassius auratus</i>	AET79963.1	<i>Oreochromis niloticus</i>	XP_003444455.1
<i>Danio rerio</i>	NP_571187.1	<i>Pagrus major</i>	BAH08689.1
<i>Gadus morhua</i>	ENSGMOP00000005053	<i>Paralichthys olivaceus</i>	KJ522774
<i>Gasterosteus aculeatus</i>	ENSGACP00000023125	<i>Rattus norvegicus</i>	NP_001009178.1
<i>Homo sapiens</i>	NP_002692.2	<i>Takifugu rubripes</i>	XP_003965650.1
<i>Labeo rohita</i>	ADC96616.1	<i>Tetraodon nigroviridis</i>	ENSTNIP00000016222
<i>Latimeria chalumnae</i>	XP_005994419.1	<i>Xenopus tropicalis</i>	ENSXETP00000008656
<i>Mus musculus</i>	NP_038661.2	<i>Xiphophorus maculatus</i>	XP_005799711.1

Table S2. The Ensembl IDs of the analyzed sequences in Figure 3B.

Species	Ensembl IDs
Human	<i>Homo sapiens</i> ENST00000259915
Cow	<i>Bos taurus</i> ENSBTAT00000028122
Mouse	<i>Mus musculus</i> ENSMUST00000025271
Stickleback	<i>Gasterosteus aculeatus</i> ENSGACT00000023169
Tilapia	<i>Oreochromis niloticus</i> ENSONIT00000004857
Zebrafish	<i>Danio rerio</i> ENSDART00000065817

gtgccacgaaaacctgtggagccttgaggagagcagcatttgagcttgacggtggagcggcgtgaaggcaaggttcatggagtagccggaggcgt 100
 gttggtgtaaatgtggcaccacacagcagctgtaaggccttttgcaaacctctcccaaacggactttacgcaagcaaacagttgatggaattttgga 200
 ggcgctccctccgctgcttctctttgacttggcgtgtaaacctgaaacggatgagcagctttaaactttatccattaaactttctttaaactttcta 300
 tctcggactgttctgttaataatctcccacagcgtttggaag**ATG**TCTGAAAGATCTCAGAGTCCGGAGTGCCAAAGCAGGCCGTAGACTTCAGCC 400
 M S E R S Q S P E C Q S R P Y D F S 18
 GCGCGGCCATTGCGCTCAGGTTTTGGGTCAAGATGGTTTTGGGCAACCGCTGCGTCATTCACGCTTCCTCACGGCCTTTTGCCAGACCCGACCTCTCTA 500
 R A G H C A Q V L G Q D G L G N A A S F Q L P H G V L P D P S L L Y 52
 TAACAAAACGGCGTATGGTGGCATCACGTGGCCCTTGGCGAGACTTTCTCCATTCCACCATGGCCAGTGACTACAGGGGAACCGACCTGCAGGCT 600
 N K T A Y G G I T S A S A Q T F F P P P M A S D Y R G T D L Q A 85
 GGGGAGTTTGGGCAACCCAAACACTGGTATCCCTTGGCTGCGCCGAGTACACCGGCCAGGTACCGGCGTAAACAGCAGCTACCCAGCTACTAACCTGA 700
 G E F G Q P K H W Y P F A A P E Y T G Q V P G V T A A T Q P T N L 118
 GTCCCTTATGGCCGAGACCCGGGAGCAAAATCAAAATGCCGAGATCAAAACCGAGAAGGACACCGGCGATGACTACTCAACGGAGATCAAGGTCACAA 800
 S P P M A E T R E Q I K M P E I K T E K D T G D D Y S T E I K G Q Q 152
 GTATCTGACACCGTACGCTCCGCGCCATGGCTCATGGAGTCTTACCCCGCGCTGGAACCCGTCCTTCTGGCCTGGAAATCGCCACATCACCCCG 900
 Y L T P S A S A A M A H G V F Y H A A W N P S F W P G I A H I T P 185
 CCCGATCAGTAATCAAAATCTCCACATCTCTGCACTGCGCGCTCTATGTCCTCCACCCGAGCAACGGGCTCCAGGAACCGCTTTTCA 1000
 P G I S N Q N P S T S A S S P S M S P S P S N G L P G N A F F 218
 GCGTGAACCAACCCGCGCCCGGCGCAGACGAGAACCCGCTCTCCACCGGAGCAGCGGTTCTCCAGCGCGGATCGAGCAGCTCAGA 1100
 S V N P T Q A A P G P Q T Q N P A S S T R S S G S S S G G C S D S E 252
 GGAAGtaagattagaatcccttcaaccaaccatttaataccaacagttgtaatatgagcctcaagtcaaacacacatggttcaactctgtgaagttcta 1200
 E 253
 catctatatggttttcaactgtctcacagGAGAACCTTCCACTGAGGAACCTGGAGCAGTTTGGCAAGGAACCTGAAACACAAACGCATCACTT 1300
 E N L S T E E L E **Q F A K E L K H K R I T** 274
 TGGGTTTCACTCAAGCTGATGTTGGCCTTGGCCTGGGTAACCTCTATGtaagtttggttcccatggtgaaacgctctattgggtcatattgcaactccct 1400
L G F T Q A D V G L A L G N L Y 290
 ccctaatgctttcaagattcttatttgaagtaataatgatggactttctagtattggtttgtatggttcacagttcagattcttctacagctcagtag 1500
 aaggttactgttattgtttgtggttgaggcaggaggtcaactggtttggcaagatctctctgattgccactcctgaaactctgctgaaatcatcagccat 1600
 ccaatggttgagattttggacagatacatgtgcatgagcttctgtctattcagagataatggcagtgaaactctgatgatttatatttgcgtgcagGT 1700
G 291
 AAGATGTTCAAGCAGACGACCATTGGCCGCTTCGAGGCTTCGACGCTGAGCTTAAAGACATGTGCAAGCTGAAACCCCTTCTCAGAGATGGCTGAATG 1800
K M F S Q T T I C R F E A L Q L S F K N M C K L K P L L Q R W L N 324
 AGGCAGAGACCTCAGACAAATCCCGAGGATgtaagctgcaaacactgacatctgccagctagttacaagacaacagatgatccttgagcctgatgggatc 1900
E A E T S D N P Q D 334
 agttggtctttcagaactgagctgactttaaccaagcactttgttctcaggaagtgtgacctgccatctagttggcaacattaatttggctctcttagag 2000
 atcaatgaacattatttgaaccataacctccaatacaaggcttaaaccaaaaggacacaaatgtctcggtacataggtagctcctagttgggctgtg 2100
 tcatacctttttgtacatctacatatttccaattatgtgttattctcagATGTACAAGATCGAGAGAGTGTGTTGACACCAAGAGGAGGAGAG 2200
 M Y K I E R V F V D T R **K R K R** 350
 AAGGACTAGTCTGGAGGGGGCAGTGGCTCTGCTCTGAGTCTACTTTATCAAGTGTCCCAACCCAAACACCCAGGAGATTACACACATTTAGATGAC 2300
R T S L E G A V R S A L E S Y F I K C P K F N T Q E I T H I S D D 383
 CTGGGCTTGGAGAGACGtaagaacaaaagtgcagtagtttgctctcaagatgtgtctgttgggttgaacagagaatctgttcttggcagG 2400
L G L E R D 389
 TGCTGCGTGTGGTTCTGCAACCCGGAGACAGAAAGGAAAACGACTGGCTTGGCCCTGGATGAGGAGTGGCGCGCCAGTATTACGAGCAGAGCCPTT 2500
V V R V W F C N R R Q K G K R **L A L P L D E E C D G Q Y E C D S P S** 423
 CCCACTGAACATGGCCCTTCCCCCATCCCGAGTCAAGACTACCCCTGTGTCAGCTACCCCGGAGCCACTCTACACACTGTACATGCCCCCTTCCAC 2600
 P L N M A P S P I P S Q S Y P V S S Y P G A T P T T L Y M P P L H 456
 CGGCCGATGCTGAAACAGGCCCTTCCACCCGGACTTATTGGTCACATGACTGGAT**TAA**acacgctcgggttgccctagcgaatcccgcagataacct 2700
 R P D V L K Q A L H P G L H M T G * 475
 cccaagctttctgtcaaaaggccaactcaactggagaggagcagcagtgatcacaggacaacccgttcaaggaggaggatcttttctgttatct 2800
 tgaactacacaaatgtcattgtccacatcctcaaacatgaccagaaacccacaccaacaattttggatcaagctctggatgaagcatagtaaatcaa 2900
 tccactggcttattcagttcatttgaacaatttaagatttagctttgtacagcaaaaggcaagttatggaatgatgtaagaacctgaacgttctct 3000
 ccaattatactaaatcctggtaggttttctgggggtgcagtgagaaacagcttaggttatcaagagaaccaaaatgtgagtaataagaattatttttta 3100
 gagttggctcagttatttggccatactgtccaccaactgtgaaaagagcagctacggattttgtcttccatttgaatggagaacagtttcttattgggg 3200
 ttaattttttccatattattttgcacaaagttgcttttgaatgtgaaactcttttcaaaatgttcccggtttgtttacaaacaaatatttttctc 3300
 ttatagttgtggtgtcatcgacatcttgagcaagatttttggcagcagtgaggagattcaagtaataattttagatgaatgtgagaatcttttgtt 3400
 cagtgcttagctgctaaatctgcaattgcaactctgcaatatttctctctcatgctcaggtgagtgccagtgagataaatatccgacctgactcagc 3500
 tatcgcaa (aaaaaaaaaaaaaaaaaaaaa) tgtgactgactgttcttatttcccaccctattgtaagggttttatgttaagtgtgaatta 3573
 caccgtgtcaagtgctcactgaaatataatttaagaagttgctttttagtggaatgatcatagttttagaagaagacataaatgtctcattttgcttc 3673
 aaaaactccatcagactgaagcttaagtgaccatcaaatcaactggagatgacattaacattggatatactttaaactttataaacacgctt 3773
 gtctgcagagattgaaatggattcattataatttgaatgttggtaagctgagatgtttgactgtaaccacctgagacctaaatgtaaaacgga 3873
 cactttattgaaaaattgatatgcaaacacaaaattgttcttctccagtaaatctttgcacatgtcccacaaaa (aaaaaaaaaaaaaaaa 3973
 aaaaaa) 3979

Figure S1. Nucleotide and protein sequences of Japanese flounder *pou5f3* gene (accession number: KJ522774). Coding sequences (CDS) are denoted in uppercase letters, whereas the UTRs or introns are denoted in lowercase letters. The deduced amino acids are indicated in one-letter code and exhibited below the coding regions. The GT/AG intron/exon boundaries are shown in italic and the start and termination codons are highlighted in bold. The POU-specific domain and POU homeodomain are shadowed with light and dark gray boxes, respectively. The red letters indicate the poly (A) tails of the two isoforms of *Popou5f3* transcripts.

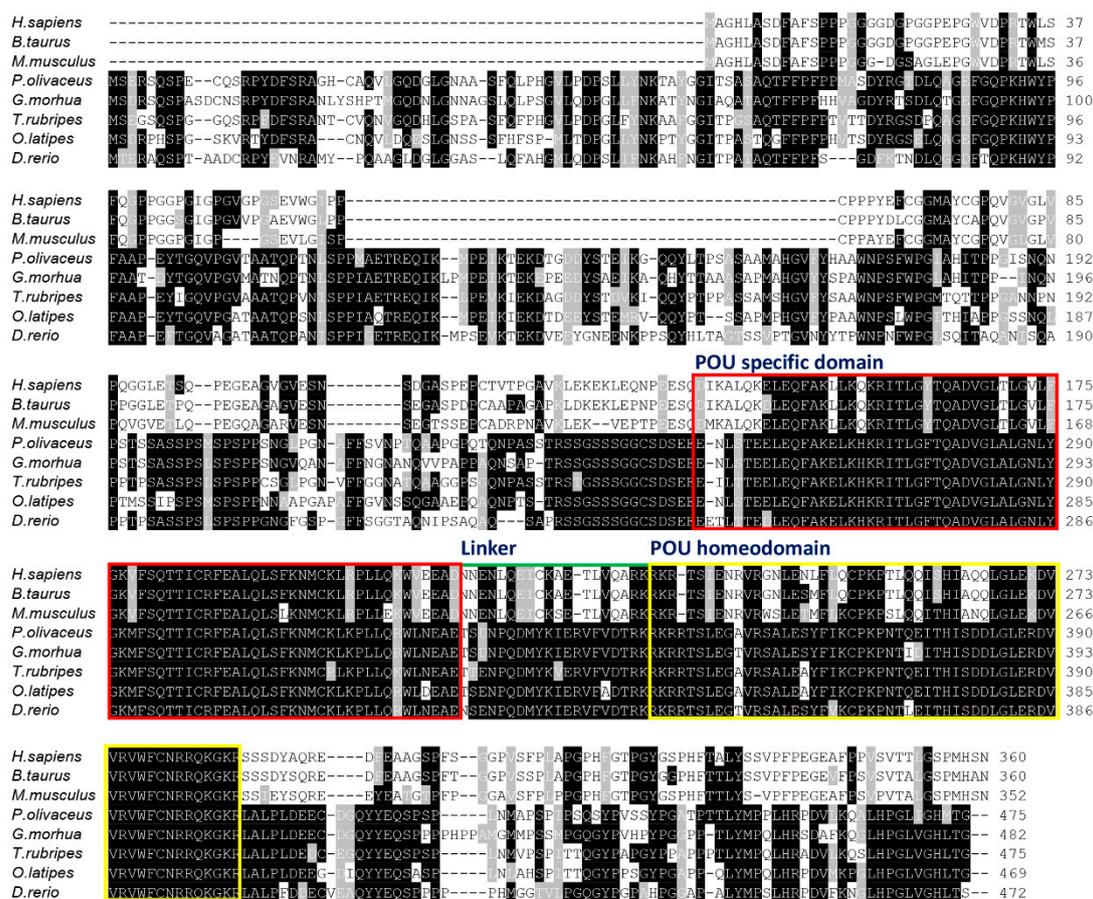


Figure S2. Comparison of the full-length amino acid sequences of Japanese flounder Pou5f3 protein with other vertebrate orthologues. Identical and similar residues are highlighted in black and gray, respectively. POU specific and homeodomains are indicated with red and yellow frames, respectively. The linker regions between two domains are shown by green line. The alignment is generated with ClustalW and shaded with BOXSHADE 3.21. The hyphens (-) indicate gaps that were inserted to maximize sequence similarity.

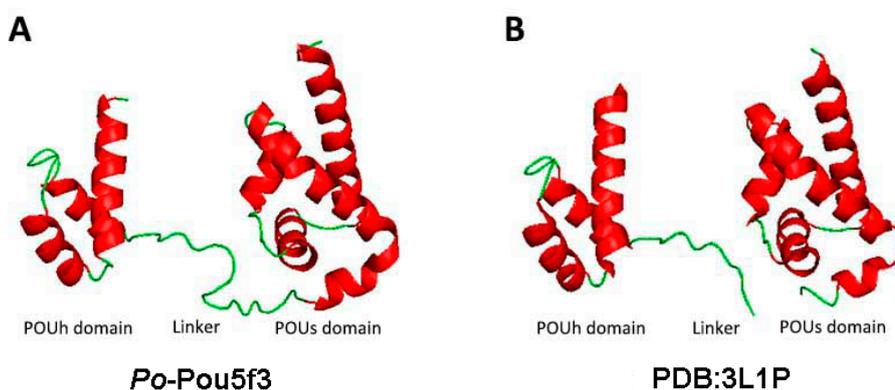


Figure S3. The predicted three-dimensional model of *PoPou5f3* POU domain based on *M. musculus* Pou5f1 (PDB ID: 3L1P) by homology modelling method using SWISS-MODEL online server. The α -helices and loop regions are shown as red and green ribbons, respectively.