PNAS
www.pnas.org

Supplementary Information for

Evolution of vertebrate gill covers via shifts in an ancient Pou3f3 enhancer

Lindsey Barske, Peter Fabian, Christine Hirschberger, David Jandzik, Tyler Square, Pengfei Xu, Nellie Nelson, Haoze Vincent Yu, Daniel M. Medeiros, J. Andrew Gillis, J. Gage Crump

Lindsey Barske
Email: lindsey.barske@cchmc.org

Gage Crump
Email: gcrump@usc.edu

This PDF file includes:

Figures S1 to S8
Legends for Movie S1
Tables S1 to S4
SI References

Other supplementary materials for this manuscript include the following:

Movie S1

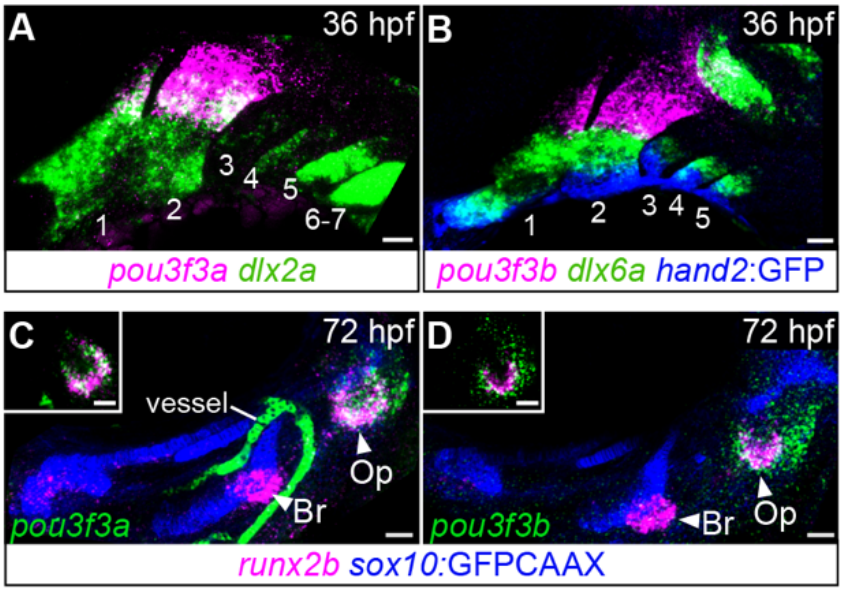
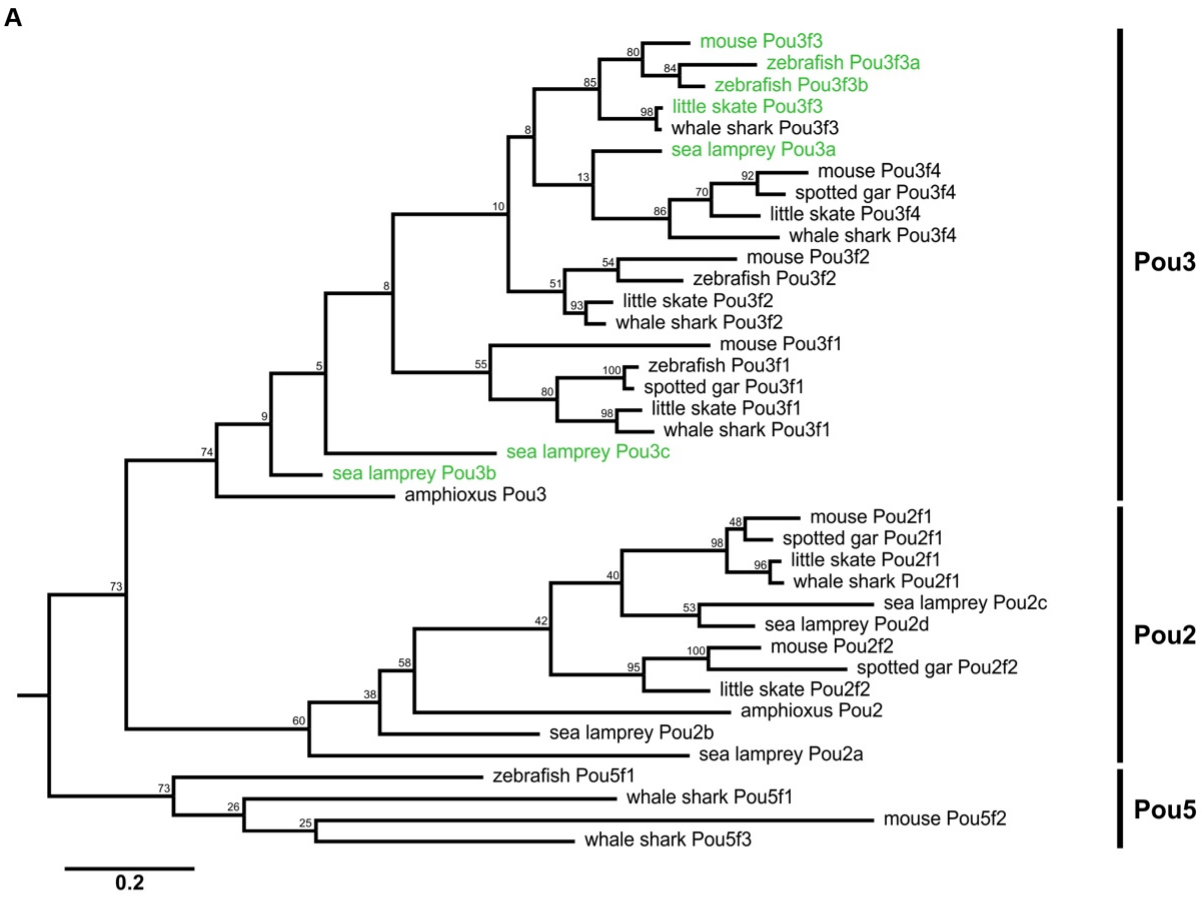


Fig. S1. Zebrafish *pou3f3a/b* expression is restricted to the mandibular and hyoid arches and later to the opercle-forming domain

A-B, Unlike most dorsoventral patterning genes, such *dlx2a*, *dlx6a* and *hand2*, which are upregulated in each arch as they form from anterior to posterior, *pou3f3a* and *pou3f3b* remain restricted to arches 1 and 2. Visible arches are indicated in each panel. **C-D**, By 72 hpf, *pou3f3a/b* expression is further restricted to the hyoid arch-derived operculum, overlapping the semicircle of *runx2b*$^{+}$ pre-osteoblasts at the ventral end of the forming opercle (Op) bone but not the branchiostegal ray (Br) bone. Insets in **C-D** are single optical sections; all others are maximum intensity projections. *sox10:*GFPCAAX marks the facial cartilages. Scale bars = 20 μm.


A
mouse Pou3f3
zebrafish Pou3f3a
zebrafish Pou3f3b
little skate Pou3f3
whale shark Pou3f3
sea lamprey Pou3a
mouse Pou3f4
spotted gar Pou3f4
little skate Pou3f4
whale shark Pou3f4
mouse Pou3f2
zebrafish Pou3f2
little skate Pou3f2
whale shark Pou3f2
mouse Pou3f1
zebrafish Pou3f1
spotted gar Pou3f1
little skate Pou3f1
whale shark Pou3f1
sea lamprey Pou3c
sea lamprey Pou3b
amphioxus Pou3
Pou3
mouse Pou2f1
spotted gar Pou2f1
little skate Pou2f1
whale shark Pou2f1
sea lamprey Pou2c
sea lamprey Pou2d
mouse Pou2f2
spotted gar Pou2f2
little skate Pou2f2
amphioxus Pou2
sea lamprey Pou2b
sea lamprey Pou2a
Pou2
zebrafish Pou5f1
whale shark Pou5f1
mouse Pou5f2
whale shark Pou5f3
Pou5
0.2

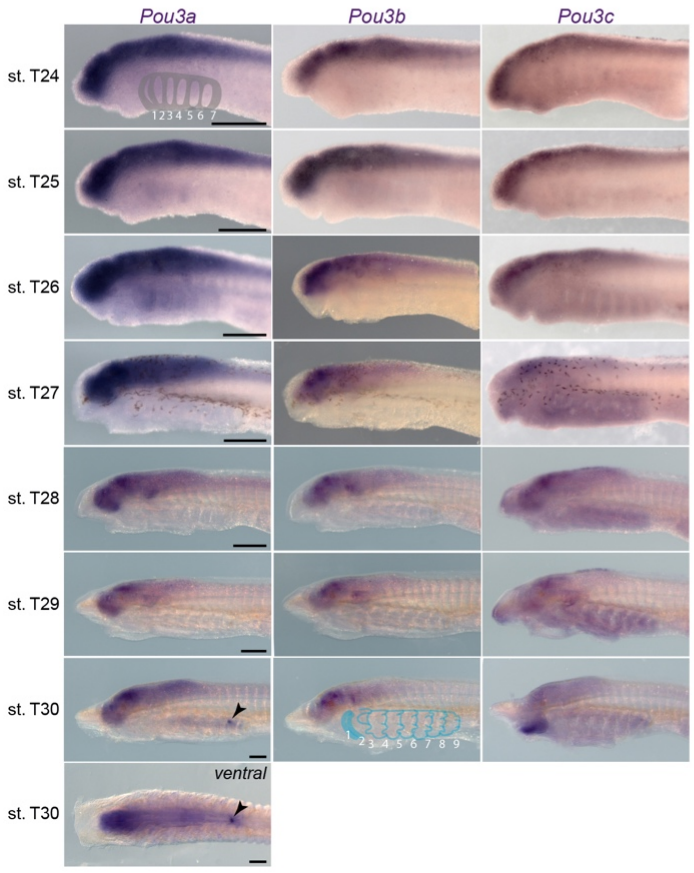

B
Pou3a
Pou3b
Pou3c
st. T24
st. T25
st. T26
st. T27
st. T28
st. T29
st. T30
ventral
st. T30

Fig. S2. Phylogenetic analysis of POU domain homeobox genes reveals three divergent Pou3 homologs in the lamprey genome

A, To identify possible lamprey Pou3f3 homologs, we retrieved lamprey (*Petromyzon marinus*) Pou-domain containing sequences from a larval transcriptome and determined their similarity to gnathostome Pou sequences. Two transcript translation products clustered at the base of the gnathostome Pou3 class (termed *Pou3b* and *Pou3c*), while one showed an affinity to Pou3f4 (*Pou3a*). The results of a bootstrap analysis with 100 replications are shown at every node as the percentage of replicates that supported the branch position. Although this bootstrap analysis did not provide strong support for the specific positions of these short lamprey Pou3 fragments, these sequences are the only lamprey sequences we identified that show a specific affinity for the Pou3 group, to the exclusion of the closely related lamprey Pou2 orthologs we also identified. The *Pou3a* and *Pou3c* genes appear to be incompletely assembled in the 2017 petMar3 germline genome and are located on short contigs lacking other genes, preventing us from verifying their orthology by synteny. *Pou3b* is located on contig 54v1 near the *Klhl32* gene, a gene that is also present at the gnathostome *Pou3f2* locus. GenBank accession numbers for lamprey Pou sequences are as follows: *Pou2a*, MN877652; *Pou2b*, MN877653; *Pou2c*, MN877654; *Pou2d*, MN877655; *Pou3a*, MN877656; *Pou3b*, MN877657; *Pou3c*, MN877658. **B**, In situ hybridizations for lamprey *Pou3a* and *Pou3b* show expression in the CNS but not the pharyngeal arches at stages T24-30. Schematics of the pharyngeal arches and arch-derived visceral skeleton are overlaid on the images of *Pou3a* at stage T24 and *Pou3b* at stage T30, respectively. The small *Pou3a* expression domain in the posterior pharynx at stage T30 (arrowhead) is located at the midline (see ventral view in bottom panel), not in the arch mesenchyme to either side (see schematic in Fig. 1). Weakly diffuse *Pou3c* expression is detected in the pharyngeal region at later stages. Stage T28-30 larvae were bleached to reveal staining in areas normally covered by pigment cells. Scale bars = 250 μm.

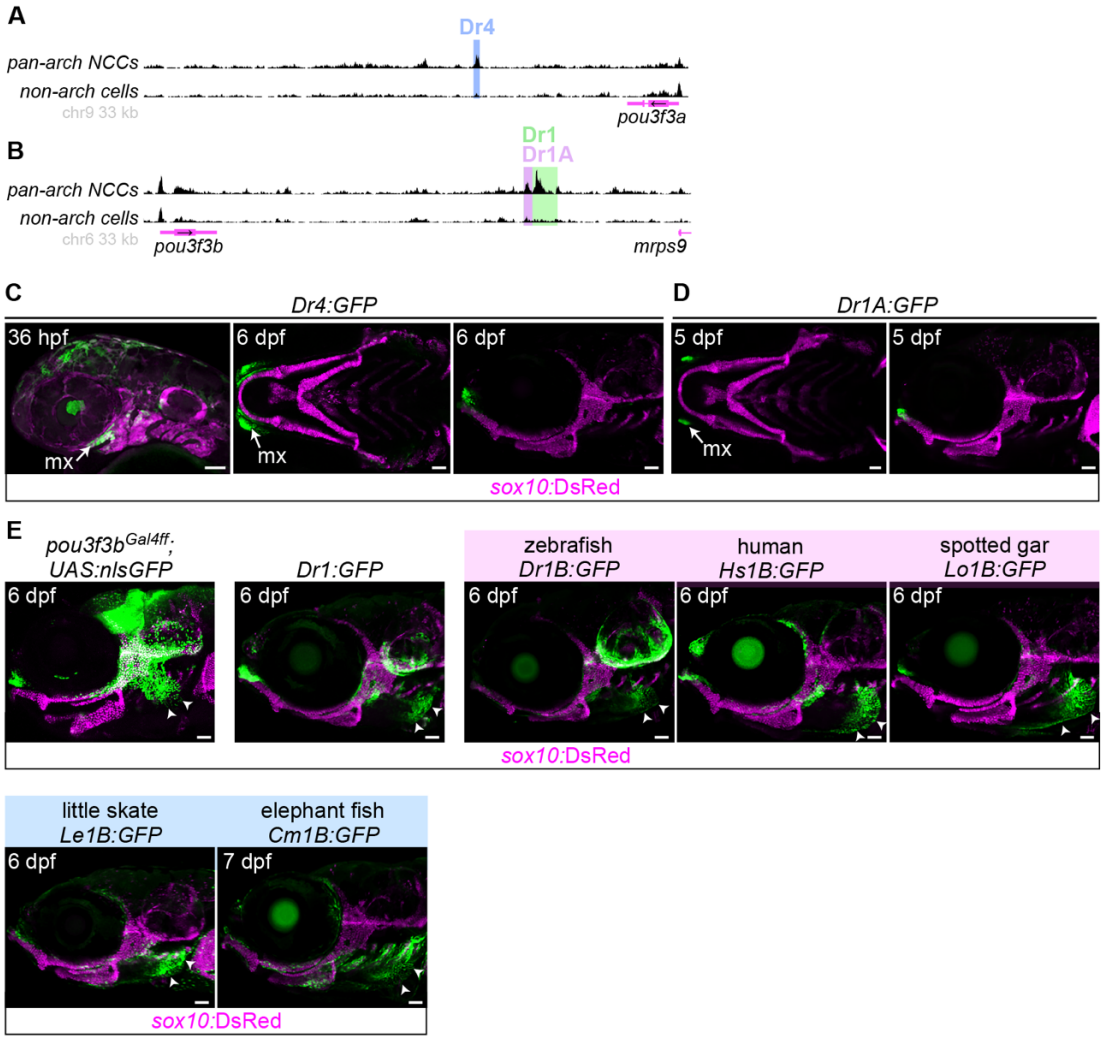


Fig. S3. Zebrafish Dr4 and Dr1A enhancers do not drive strong opercular expression compared with 1B enhancers

A-B, The Dr4 (12.3 kb downstream of the *pou3f3a* TSS) and Dr1A (21.7 kb downstream of *pou3f3b*) enhancers were identified via enrichment for open chromatin in FACS-purified *fli1a:*EGFP; *sox10*:DsRed double-positive (arch NCCs) versus double-negative (non-arch) cells. ATAC-seq data are visualized with the UCSC Genome Browser. **C**, Transgenic reporter lines show that Dr4 drives maxillary, mandibular and hyoid expression at 36 hpf, but minimal opercular expression at 6-7 dpf. **D**, *Dr1A:*GFP is restricted to the maxillary domain at 5 dpf. **E**, Lateral views of the *pou3f3b* Gal4ff knockin line compared with the Dr1 and 1B enhancer lines at 6 dpf. Note the enrichment of GFP expression at the leading edge of the growing operculum and lack of CNS expression in all enhancer lines. Scale bars = 50 μm.

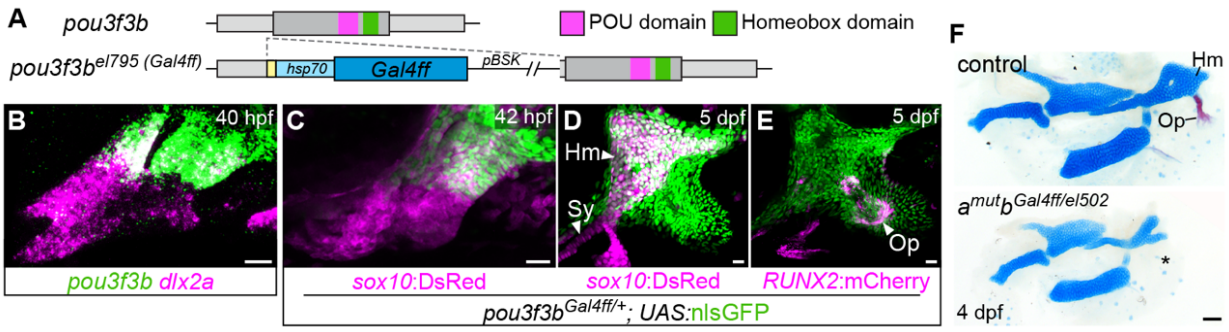


Fig. S4. *pou3f3b* knockin line recapitulates endogenous expression and behaves as a null allele

A, A linearized pBluescriptSK vector containing a CRISPR target site (yellow) and the *hsp70l* minimal promoter driving Gal4ff was knocked into the *pou3f3b* 5'UTR to create allele el795 (*pou3f3b*^{Gal4ff}). **B**, When crossed to a *UAS:*nlsGFP reporter, expression is observed in the CNS, periotic mesenchyme, and arches 1-2 (**B'**), mimicking endogenous *pou3f3b* mRNA expression at 40 hpf (**C**). *sox10:*DsRed and *dlx2a* label arch neural crest cells. **D**, At 5 dpf, the reporter marks the dorsal hyoid-derived hyomandibula (Hm) but not the symplectic (Sy) cartilage (*sox10:*DsRed^{+}) as well as the opercular flap and opercle bone (Op, *RUNX2:*mCherry^{+}) (**E**). Scale bars = 20 μm. **f**, *pou3f3a*$^{-/-}$*; pou3f3b*$^{Gal4ff/el502}$ larval skeletons lack the opercle bone (asterisk) and have a forked hyomandibula, similar to *pou3f3a*$^{-/-}$*; pou3f3b*$^{el502/el502}$ mutants. Scale bar = 50 μm.

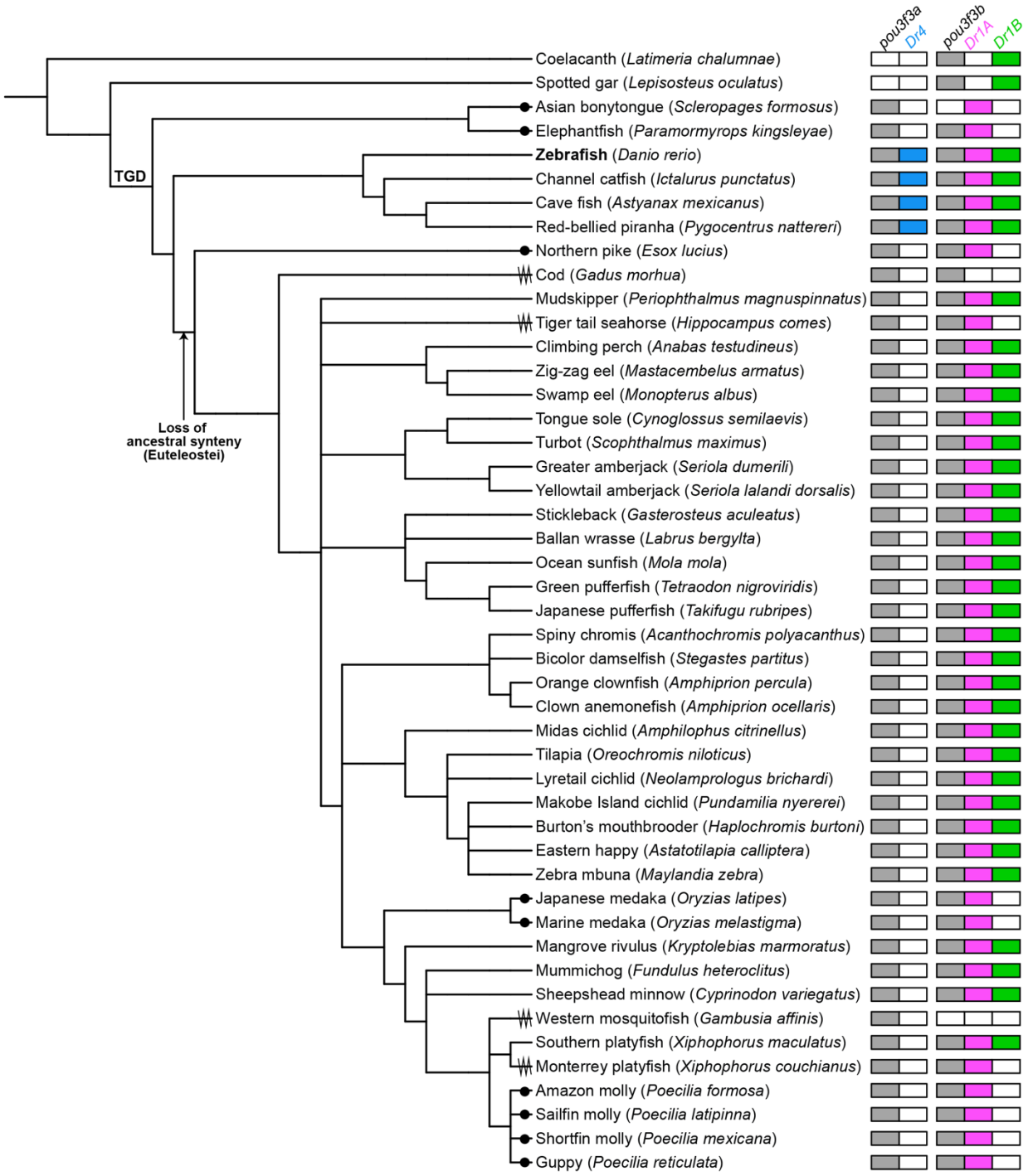


Fig. S5. Conservation of *pou3f3* genes and associated regulatory sequences in fishes

Colored boxes next to each species name indicate that homologs of the *pou3f3a* or *pou3f3b* coding sequences or the Dr4, Dr1A, or Dr1B enhancers were successfully retrieved by BLASTing with either the zebrafish or turbot homolog. TGD = teleost genome duplication. A genomic rearrangement leading to loss

of ancestral regional synteny in the Euteleostei lineage (arrow) preserved the small *pou3f3b-mrps9* block and associated intergenic sequence where the 1B element is located. Subsequent sequence modifications and rearrangements led to loss of 1B in nine species that retain the *pou3f3b-mrps9* block (black circles) and in four species in which the two genes have been dissociated (zig-zags). Tree generated using iTOL (1).

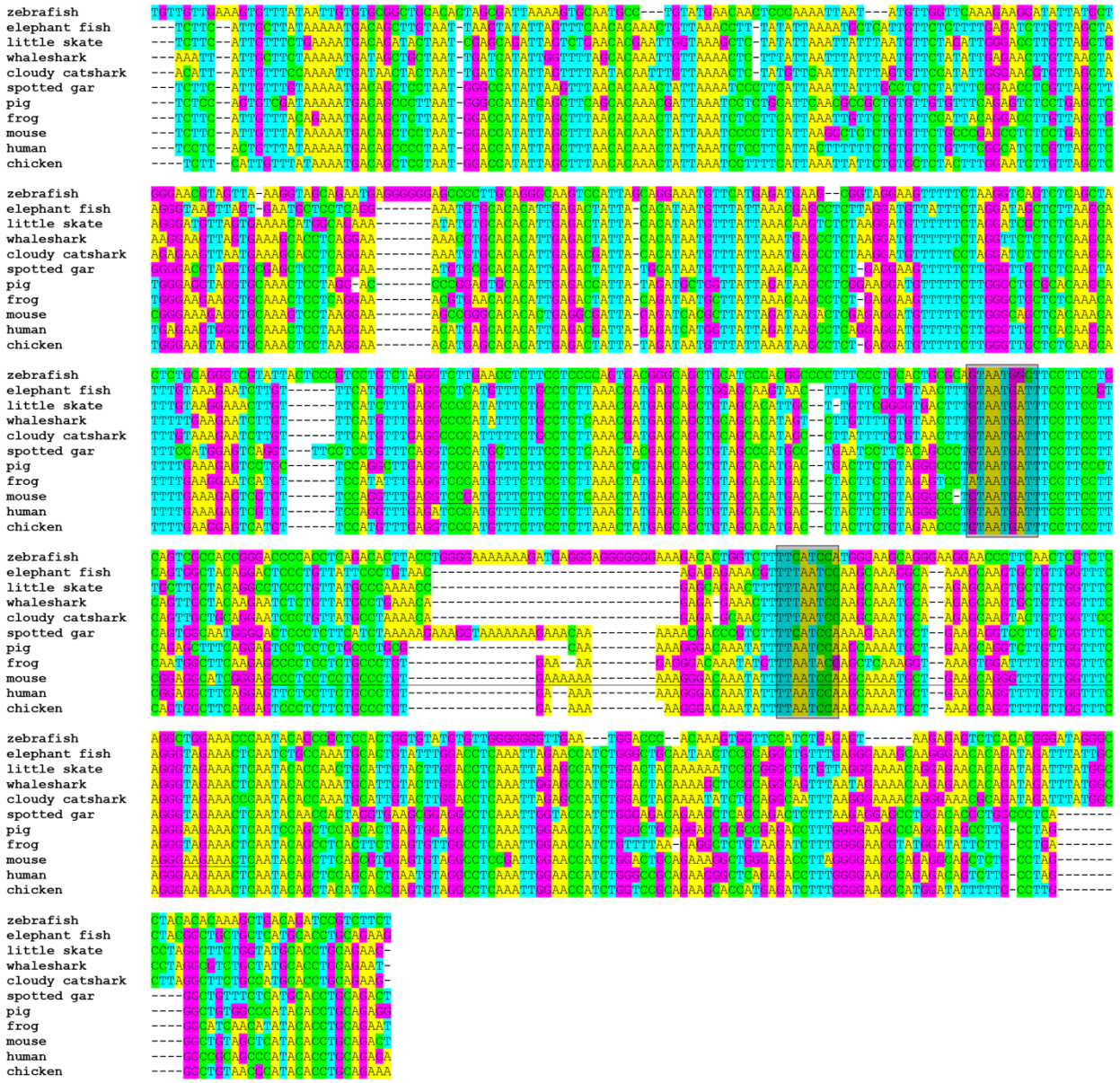


Fig. S6. Multiple sequence alignment of eleven vertebrate *Pou3f3* 1B enhancers

1B enhancer sequences retrieved from representative bony fish (zebrafish, spotted gar, frog, chicken, pig, mouse, human) and cartilaginous fish (little skate, cloudy catshark, whaleshark, elephant fish) genomes were aligned by Clustal Omega. Sequence identity values ranged from 51-76% for bony vs. cartilaginous species, 50-89% across bony fishes, and 82-90% across cartilaginous fishes. Underlined mouse sequence corresponds to the summit of a HOXA2 ChIP-Seq peak (2), and grey boxes denote predicted homeobox binding motifs.

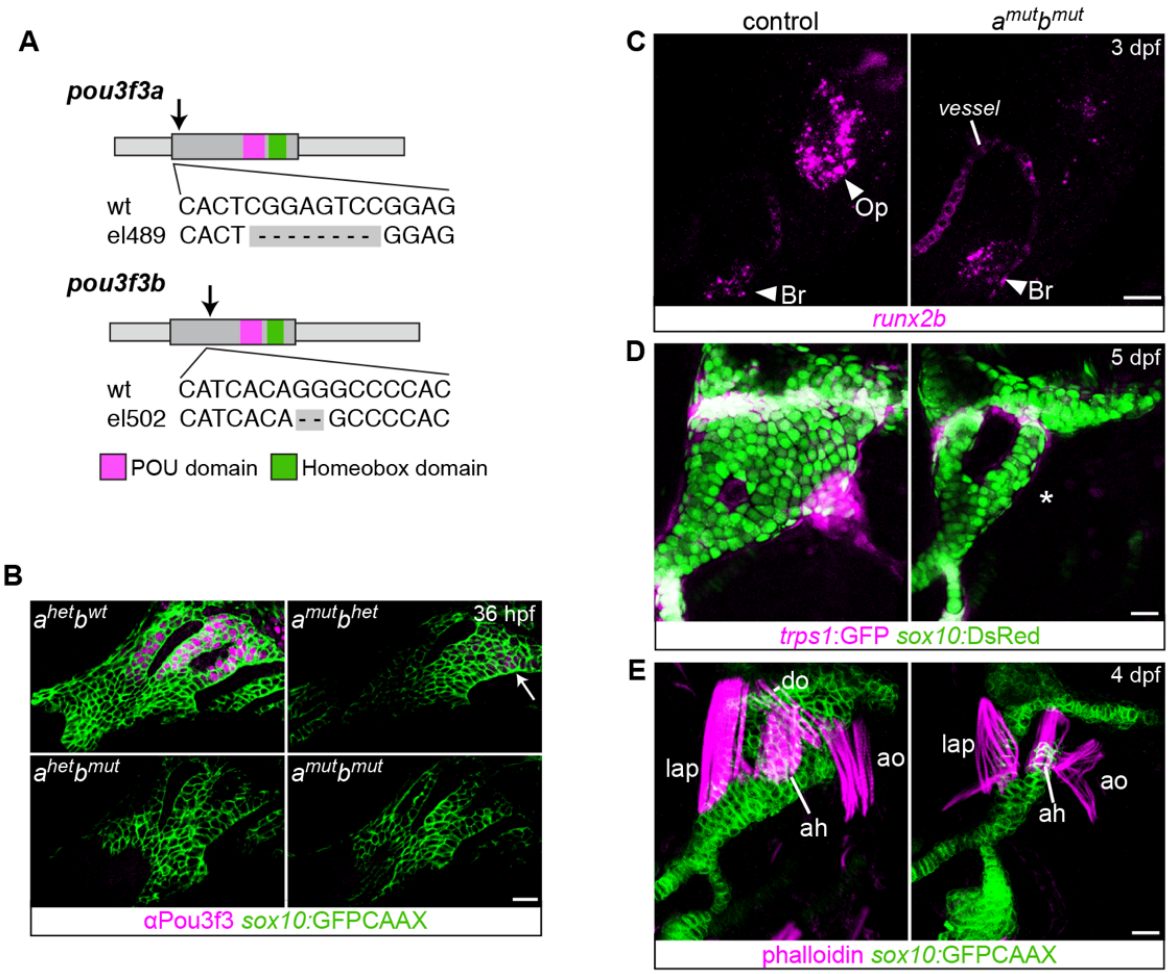


Fig. S7. Zebrafish *pou3f3a* and *pou3f3b* TALEN mutants

A, Mutations were targeted to the 5' end of each gene, upstream of the POU and homeobox domains (black arrows). **B**, Reduced pan-Pou3f3 protein expression in *pou3f3a*$^{-/-}$; *pou3f3b*$^{+/-}$ embryos (white arrow), and no expression in *pou3f3a*$^{+/-}$; *pou3f3b*$^{-/-}$ or double mutants at 36 hpf. The antibody recognizes the C-terminus of the protein. *sox10:*GFPCAAX labels arch NCCs. **C**, *runx2b*$^{+}$ osteoblasts are specifically lost in the dorsal opercle (Op) but not the ventral branchiostegal ray (Br) domains in *pou3f3a*$^{-/-}$; *pou3f3b*$^{-/-}$ mutants at 72 hpf. **D**, The *trps1:*GFP^{+} joint connecting the Op to the Hm cartilage does not form in the mutant (asterisk). **e**, Phalloidin staining reveals that the muscles that abduct and adduct the Op bone (dilator operculi (do) and adductor operculi (ao)) and Hm cartilage (levator arcus palatini (lap) and adductor hyoideus (ah)) are absent (do) or disorganized (ao, lap, ah) in the mutant. *sox10:*DsRed and *sox10:*GFPCAAX mark chondrocytes. Images in **B** are single optical sections; **C**-**E** are maximum intensity projections. Scale bars = 20 μm.

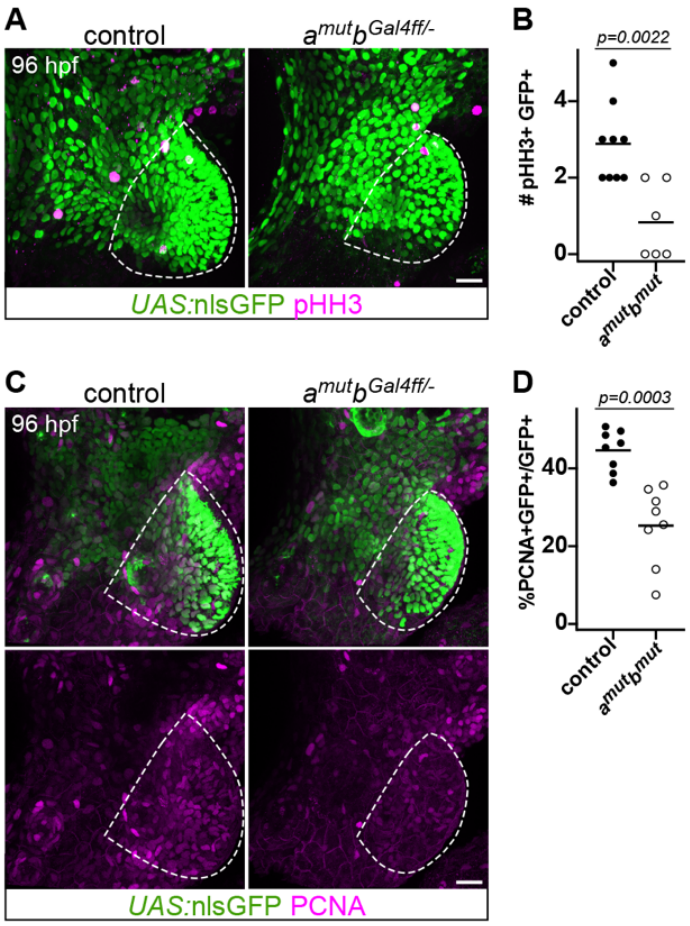


Fig. S8. Cell proliferation is reduced in the *pou3f3a/b* mutant operculum

Significantly fewer pHH3^{+}/GFP^{+} (**A-B**) and PCNA^{+}/GFP^{+} cells (**C-D**; counted relative to total GFP^{+} cells) are observed in the mutant opercular flap at 96 hpf (unpaired t-test; horizontal lines indicate the mean). The quantified opercular region is outlined. Scale bars = 20 µm.

Movie S1. Posterior-directed growth of *pou3f3b*-expressing hyoid mesenchyme

Time-lapse imaging of a control *pou3f3b*^{Gal4ff}*; UAS:*nlsGFP; *sox10*:DsRed embryo from 56 to ~72 hpf. Note the posterior migration of GFP^{+} hyoid cells to form the opercular flap. Anterior is to the left. *sox10:*DsRed (magenta) is expressed by chondrocytes in the facial skeleton (middle) and the pectoral fin (top right) as well as the otic vesicle (top center).

Table S1. Pou3f3 enhancers

species	name	build	length (bp)	coordinates
zebrafish	Dr4	GRCz10	380	chr9:6671401-6671780
zebrafish	Dr1	GRCz10	2033	chr6:14844738-14846770
zebrafish	Dr1A	GRCz10	515	chr6:14844758-14845272
zebrafish	Dr1B	GRCz10	610	chr6:14845220-14845829
spotted gar	Lo1B	LepOcu1	588	LG17:22752624-22753211
human	Hs1B	hg38	568	chr2:104936400-104936967
mouse	Mm1B	mm10	569	chr1:42771807-42772375
chicken	-	GRCg6a	565	chr1:136308706-136309270
pig	-	Sscrofa11.1	565	chr3:49738243-49738807
frog	-	JGI_4.2	566	GL173358:190090-190655
little skate	Le1B	LSb2	572	ctg1625584:439-1010
elephant fish	Cm1B	calMil1	573	KI635860:11302302-11302874
whale shark	-	ASM164234v2	572	ASM164234v2tig00293572:77873-78444
cloudy catshark	-	Storazame_v1.0	572	scf_scyto00007021:102120-102692

Table S2. Primers used to amplify cDNA for in situ probes

gene	forward primer (5'-3')	reverse primer (5'-3')
Dr pou3f3b-2	AGTAACCTCAGTGTCCGGTG	AACGGCTTCAGTTTGCACAT
Mm Pou3f3	AGACCACTATCTGCCGCTTC	CTCATCGGTCTCCAGCTGTC
Le Pou3f3	CTTTCTCTTCCTGCCCTGTG	GTCCAGCCGCTCATCTACTC
Cm Pou3f3	AGACCACCATCTGCAGGTTC	ATTCTATGCAATCGCCAACC
Pm Pou3a	AGGAGCTCGACCAGCATCAC	TGCAACCACTTGCGGAGCAG
Dr bmp7b	CTTCGTTCCAGAGCACTTCG	TTCAGGATGACGTTGGAGCT
Dr fgf24	GATCATCTATCCTGACAACTTTAGG	CACTAGTTCCTTACAGGT

Table S3. Synthesized cDNA sequences for lamprey in situ probes

gene	synthesized fragment (5'-3')
Pm Pou3b	ACTCAACAAGTGGCTCGAGGAGGCGGACTCGGCCACGGGAAGCCCCAGCAGCATCGACAAG ATCGCTGCGCAGGGCCGCAAGAGGAAGAAACGGACGTCGATCGAGGTGACGGTGAAGGGCG CACTGGAGAATCACTTCCTCAAGTGTCCCAAGCCCGCCGCGCAGGAGATAACAGCGCTGGCG GACAGCCTGCAGCTGGAGAAGGAGGTGGTGCGCGTTTGGTTCTGCAACAGGAGGCAGAAGG AGAAGAGAATGACGCCCCCTCTGAGCCTCACGCCCGGTGGGGAGGAGAATGGACC
Pm Pou3c	ACGGCCCCGGTGTTGGTGGTGGAGGAGGAGGAGGAGGGGGAGGCAGCATCTCGGACACGG AGCTTGAGGAGCCTGCGCCGCCCACCTCGGATGACCTCGAGCAATTCGCCAAGGAGTTCAAG CAGAGGCGGATCAAGCTGGGCTTCACGCAGGCCGACGTGGGCCTGGCGCTGGGCACCCTGT ACGGCAACGTCTTCTCGCAGACGACCATCTGCCGCTTCGAGGCGCTGCAGCTGAGCTTCAAG AACATGTGCAAGCTCAAGCCGCTGCTCAGCAAGTGGCTGGAGGAGGCCGACTCGGCCAACG GCTGCGCGGGCGGCGCCTGCGGCCTCGATAAGATGGCAGCGCAGGGTCGCAAGCGCAAGA AGAGGACCTCGATCGAGGTGGCGGTGAAGGGGGCCCTGGAGAAGCACTTCCTGCAGAGCCC CAAGCCCTCGGCGCAGGAGATAACGGCCATAGCCGAGGCGCTGCAGCTCGAGAAGGAGGTG GTGCGCGTGTGGTTCTGCAACCGACGCCAAAAGGAGAAGCGGATGACCGTGCAGGGAGTAG GAGGCCCTGAGGATCAGGGTTTGCAGGGCTCCATGGCTGGAGGGTGCAACA

Table S4. Experimental numbers

panel	number
Fig. 1B-D	NOTE: very precious samples; analyzed at least 1 embryo per stage for skate/holocephalan; n ≥ 6 per stage for lamprey
Fig. 1E	n > 10 wild-type embryos evaluated per stage
Fig. 1F	n > 2-3 wild-type embryos evaluated per stage
Fig. 2D-E	> 2 independently derived lines per construct evaluated where possible; ~10 transgenic embryos/larvae evaluated visually and 2-3 imaged per stage
Fig. 3C	n = 3 adult pou3f3b mutant fish stained and imaged
Fig. 3D	n = 31 *pou3f3a+/-; pou3f3b-/-* sides and n = 47 *pou3f3a-/-; pou3f3b-/-* sides stained and imaged
Fig. 3E	n = 3 *pou3f3a-/-; pou3f3b-/-* mutants imaged
Fig. 3F	n = 11 *pou3f3a-/-; pou3f3b-/-* mutants imaged
Fig. 3G-H	n = 4 controls and 4 mutants re-imaged at 6 developmental stages
Fig. 3I-J	n = 8 controls and 8 double mutants imaged and quantified at 72 hpf; n = 18 controls, 9 *pou3f3a+/+; pou3f3b-/-* mutants, 8 *pou3f3a+/-; pou3f3b-/-* mutants, 21 double mutants at 96 hpf
Fig. 3K	n = 2-3 *pou3f3a-/-; pou3f3b-/-* mutants imaged per probe
Fig. 3L	n = 2 transplanted *pou3f3a+/-; pou3f3b-/-* larvae
Fig. 4B-C	n = 16 *pou3f3a+/-; pou3f3b+/+*, *pou3f3a+/-; pou3f3b*$^{el502/-}$, or *pou3f3a+/+; pou3f3b*$^{el502/+}$ control sides; n = 21 *pou3f3a-/-; pou3f3b+/+ sides; n = 10 pou3f3a-/-; pou3f3b*$^{el502/ci3010}$ sides; 18.5-22 mm standard length
Fig. 4D	n = 5 *pou3f3b*$^{ci3010/ci3010}$ mutants imaged
Fig. 5A	n = 19 *fli1a:Gal4VP16; UAS:pou3f3b* transgenics imaged
Fig. 5B	n = 10 *kat6a/moz* mutants imaged
Fig. 5C	n = 5 *fli1a:Gal4VP16; UAS:pou3f3b* transgenics imaged
Fig. 5D	n = 5 *fli1a:Gal4VP16; UAS:pou3f3b* transgenics imaged
Fig. 5E	n = 7 *kat6a/moz* mutants imaged
Fig. 5F	n = 2 *kat6a/moz* mutants imaged
Fig. S2B	n ≥ 6 embryos per stage
Fig. S3C	> 2 independently derived lines per construct evaluated where possible; ~10 transgenic embryos/larvae evaluated visually and 2-3 imaged per stage
Fig. S3D	> 2 independently derived lines per construct evaluated where possible; ~10 transgenic embryos/larvae evaluated visually and 2-3 imaged per stage
Fig. S3E	> 2 independently derived lines per construct evaluated where possible; ~10 transgenic embryos/larvae evaluated visually and 2-3 imaged per stage
Fig. S4F	n = 10 *pou3f3a-/-; pou3f3b^Gal4ff/-* mutant sides stained and imaged
Fig. S7C	n = 9 *pou3f3a+/-; pou3f3b-/-* or *pou3f3a-/-; pou3f3b-/-* mutants imaged at 56 or 72 hpf
Fig. S7D	n = 5 *pou3f3a-/-; pou3f3b-/-; trps1:GFP* mutants imaged
Fig. S7E	n = 3 *pou3f3a-/-; pou3f3b-/-* mutants imaged
Fig. S8A-B	n = 9 controls and 6 double mutants imaged and quantified
Fig. S8C-D	n = 8 controls and 8 double mutants imaged and quantified

SI References

1. I. Letunic, P. Bork, Interactive tree of life (iTOL) v3: an online tool for the display and annotation of phylogenetic and other trees. *Nucleic Acids Res* **44**, W242-245 (2016).
2. I. J. Donaldson *et al.*, Genome-wide occupancy links Hoxa2 to Wnt-beta-catenin signaling in mouse embryonic development. *Nucleic Acids Res* **40**, 3990-4001 (2012).