Supplementary Information

Supplementary Methods

Embryo Collection. *P. marinus* embryos were collected at Hammond Bay Biological Station, Millersburg, MI, USA. Spawning males and females were removed from nests and returned to the laboratory. Eggs from gravid females were manually stripped into a beaker and sperm from a mature male was expressed onto the eggs. Fertilized eggs were washed through several changes of filtered lake water and reared at 18°C in a constant temperature bath and were changed into fresh water on alternate days to prevent fungal growth. Alternatively, gravid adults shipped to the laboratory were maintained at 10°C and 18°C. Embryos were collected as described and maintained in 0.1X Marc's Modified Ringers solution at 18°C.

Antibody synthesis and immunohistochemistry. A *PmSoxE1* coding sequence corresponding to 82 amino acids at the amino terminus of the predicted protein was subcloned in frame between EcoRI/XhoI polylinker sites on the pGEX-KG vector and transformed into BL-21 cells (Stratagene, La Jolla, CA, USA) for in vitro translation. A glutathione-S-transferase (GST) tagged fusion protein product was isolated using the B-PER bacterial protein extraction reagent (Pierce, Rockford, IL, USA) according to manufacturers instruction, and used for rabbit polyclonal antibody production (Zymed Laboratories, South San Francisco, CA, USA). Immune serum was affinity purified and used at a 1:1000 dilution for whole mount immunohistochemistry.

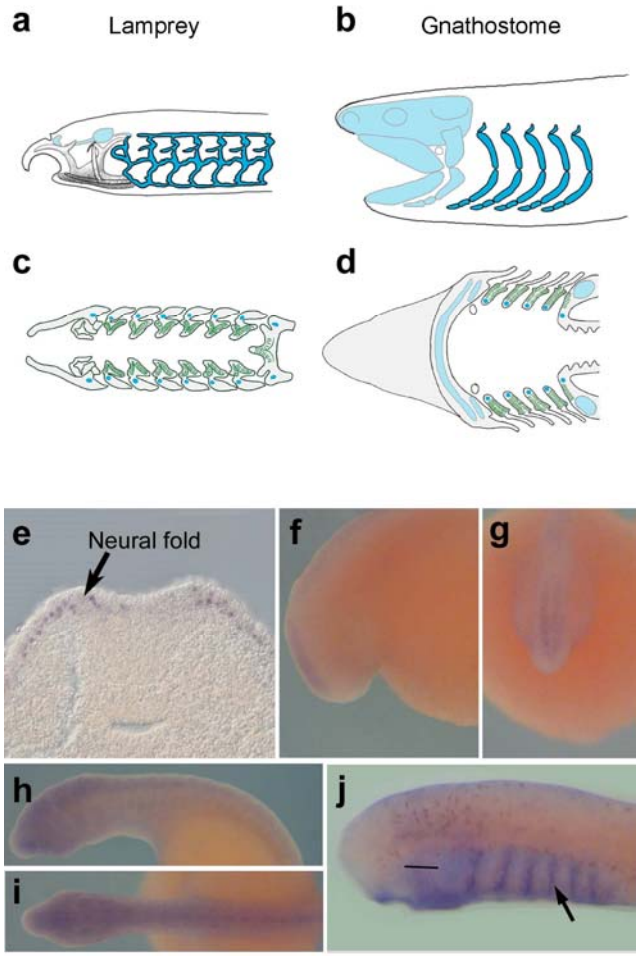
Morpholino Injection. Fluorescein-labeled (FITC) morpholino antisense oligonucleotides (MO; Gene Tools, Philomath, OR, USA) were designed to recognize the *PmSoxE1* flanking sequence 5' to the start codon (MO1, 5'-CTCGCTCGAATCGCTCGCTTCCTCT-3') and immediately 3' to the predicted initiation codon (MO2, 5'-GTTGTGCAGTCGCTTCTCTGCCATC-3'). FITC-MO's were resuspended in deionized Millipore-filtered water and pressure injected (10-100ng) into newly fertilized zygotes, or into a single blastomere following the first cleavage. Fluorescence microscopy was used to check for presence of morpholino following injection. A

5-base mismatch morpholino or a *G.galllus* Pax7 MO (10-100ng) was injected as a control for non-specific deleterious effects of morpholino injection. MO designed against *P.marinus* Pax2/5/8 (10-100ng) was injected and embryos inspected for effects on development of the branchial arches. Injected embryos were allowed to develop to day 15 when control embryos had reached stage 26.

Gene Cloning and Phylogenetic Analysis. Degenerate forward (5'-TACGAYTGGWCIYTNGTNCCIATGCC-3') and reverse (5'-GGCTGRTAYTTRTAI TCIGGRTRRTC-3') oligonucleotides were used in the polymerase chain reaction to amplify a 278 base pair sequence that was cloned into the pCR2.1 TOPO vector (Invitrogen, Carlsbad, CA, USA), and used to screen an embryonic cDNA library as described previously[1]. Library screening yielded a 6kb clone that was partially sequenced on both strands, and an 1805bp clone that was fully sequenced (Davis Sequencing, Davis, CA, USA). The Lasergene analysis software package (DNAstar, Madison, WI, USA) was used to construct a 2434 bp contiguous sequence that was found to include a putative 617 amino acid open reading frame. The *PmSoxE1* full length coding sequence is available from Genbank under accession number AY830453. The 1805bp library clone included a putative 482 amino acid open reading frame that aligned to known SoxF family members and was used for outgroup comparison in the phylogenetic analysis.. *PmSoxE2* (3857bp) and *PmSoxE3* (2156) clones were obtained as described[1] in subsequent library screens using the *PmSoxE1* sequence and a SoxB sequence as probes. Clones were fully sequenced along both strands. ClustalW[2] and Bioedit 5.0.9[3] were used to align the translated *SoxE* putative amino acid sequences to known gnathostome SoxE members to create an alignment that includes 353 (*SoxE1*), 356 (*SoxE2*) and 357 (*SoxE3*) amino acid positions in the lamprey sequence (Fig. 1). A neighboring-joining[4] phylogenetic tree was

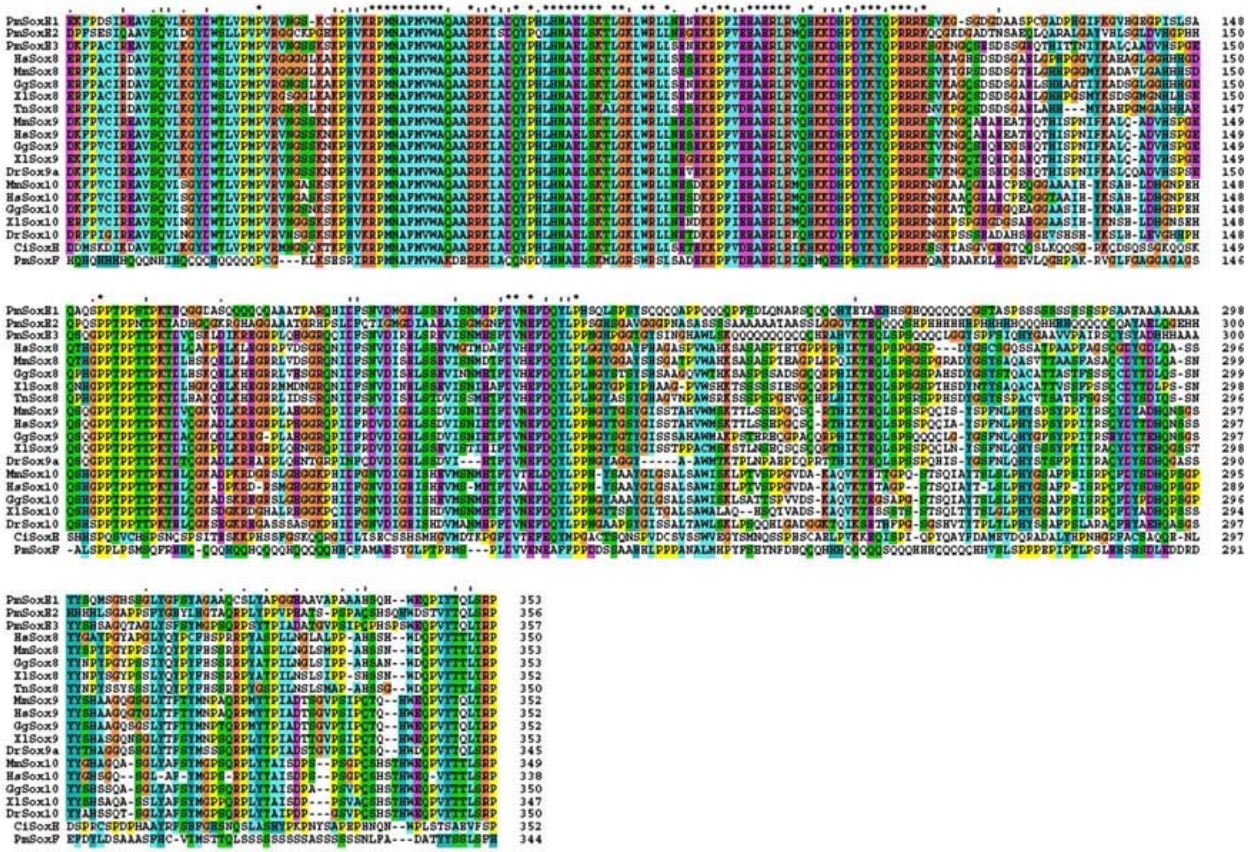
constructed from the aligned sequences using MEGA2[5]. Accession numbers for all genes included in the analysis are provided next to the respective gene on the tree (Fig. 2).

Immunoblot Analysis. Crude embryonic extract was made by homogenizing 2 stage 25 and 2 stage 26 embryos with an Eppendorf pestle 100µl m-per protein extraction reagent (Pierce). Homogenate was centrifuged 25 min at 14,000 xg to yield 1µg/µl crude extract. 5µg and 10µ samples were run on 10% SDS-PAGE, blotted onto Millipore Immobilon-P PVDF membrane, and probed with a 1:1000 dilution of antibody in phosphate buffered saline containing 0.1% Tween 20 (PTW) after blocking 20min. with 5%nonfat milk in PTW. The blot was probed overnight at 4C, rinsed 3x in PTW, followed by 3x15 min washes in PTW. A secondary antibody (goat antirabbit conjugated to horse radish peroxidase) was diluted 1:2500 and probed against the blot 1 hr at room temperature. Blot was washed 3 x 15min. and ECL detection was according to manufacturers instructions (Amersham).

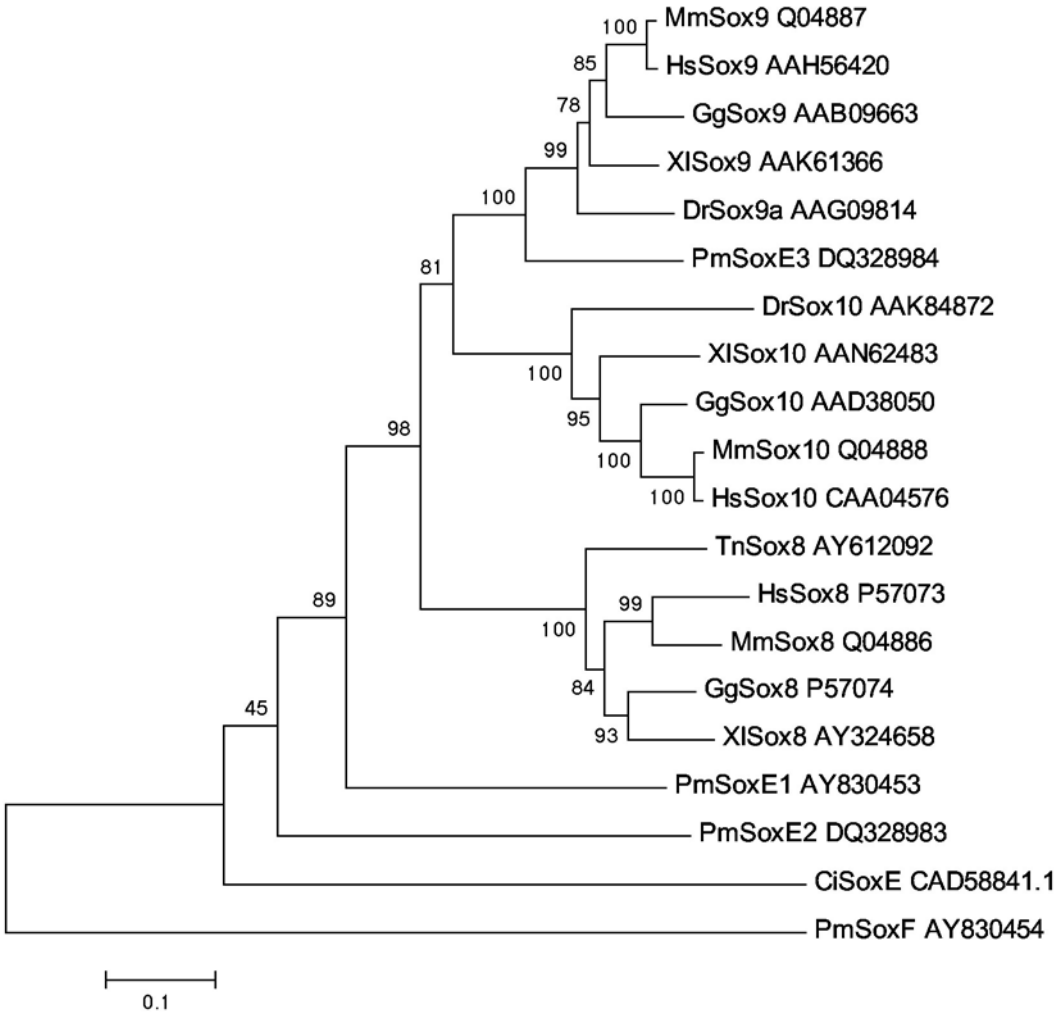


Supplementary Figure 1. Comparison of branchial arches between lamprey and an idealized gnathostome. Non-branchial arch chondrogenic elements of the head are shown in light blue; branchial arches in dark blue. **a**. Lateral view of lamprey cranial skeleton. The branchial basket is composed of each branchial arch fused to the adjacent arch. Larval mucocartilage shaded gray. **b**. An idealized gnathostome. Branchial arches are separate with articulations between each arch element. **c**. Frontal section through lamprey pharynx. Branchial arches (blue) are lateral to gill filaments (green) in each arch. **d**. Frontal section through the pharynx of a shark. Branchial arches are medial to the gill filaments present on each arch. **e-j**. *SoxE1* expression during embryogenesis. *SoxE2* expression (not shown) appears in the same population of cells as *SoxE1*

during early development **e**. *SoxE1* expression in presumptive neural crest within the open neural folds. **f** (lateral) and **g** (dorsal) expression in premigratory neural crest at stage 22. Stage 23 migratory neural crest expression of *SoxE1* in lateral (**h**) and dorsal (**i**) views. **j,** Stage 25 lateral view shows SoxE1 expression in branchial arches caudal to the mandibular arch. Drawings adapted from refs[6-9]

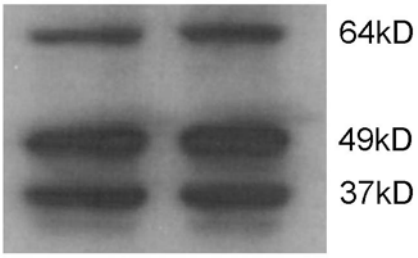


Supplementary Figure 2. Alignment of *SoxE1*, *SoxE2*, and *SoxE3* to gnathostome *Sox8*, *Sox9,* and *Sox10* representative genes. ClustalW was used to create a full length alignment that was then edited to include 353, 356 and 357 positions of the putative lamprey amino acid sequences aligned with gnathostome paralogues. A lamprey SoxF gene is included for outgroup comparison.



Supplementary Figure 3. Neighbor-joining[4] tree constructed in Mega2 from Clustal alignment of amino acids from each of the putative lamprey protein sequences to gnathostome SoxE genes positions *PmSoxE1* basal to all gnathostome SoxE paralogs, *PmSoxE2* is more distantly related to vertebrate SoxE genes, whereas *PmSoxE3* shows strong sequence similarity to gnathostome Sox9. Lamprey SoxF (PmSoxF) is included for outgroup comparison. Accession numbers for sequences used in the phylogenetic analysis are listed next to each corresponding gene. Species abbreviation, name: Ci, *Ciona intestinalis*; Dr, *Danio rerio*; Gg, *Gallus gallus*; Hs, *Homo*

sapiens; Mm, *Mus musculus*; Pm, *Petromyzon marinus*; Tn, *Tetraodon nigroviridis*; Xl, *Xenopus laevis*.



Supplementary Figure 4. AntiPmSoxE1 was probed against whole embryonic extract of stage 25 and stage 26 embryos. The antibody recognized three protein bands, suggestive of crossreactivity with multiple SoxE proteins.

Supplementary Discussion.

The SoxE tree topology (supplementary figure 3) suggests a complex history of *SoxE* duplication in early vertebrate history. The positioning of both *SoxE1* and *SoxE2* basal to all gnathostome SoxE members suggests an independent duplication event specific to the lamprey lineage. However, the position and strong support for *SoxE3* at the base of gnathostome *Sox9* genes suggests that *Sox9* may be most like the ancestral *SoxE* gene. Among gnathostomes, an early duplication of the ancestral *SoxE* gene may have resulted in the presence of the existing *Sox8* and *Sox9* genes. A more recent duplication in gnathostomes may have resulted in existing *Sox9* and *Sox10* branches that show higher sequence homology to each other than to *Sox8*. Regulatory constraints could have been important for the high homology observed between gnathostome *Sox9* and *SoxE3* in the lamprey. Along the lamprey lineage, *SoxE3* would have the same regulatory constraints imposed on gnathostome *Sox9* as suggested by their high affinity. Early independent duplications in the lamprey led to the current existence of *SoxE1* and *SoxE2* and lack of constraint on these newly duplicated genes may have allowed for rapid sequence divergence and developmental roles specific to lamprey (e.g., development of non-collagenous cartilage). Rapid sequence divergence would also cause a long branch artefact in our sequence analysis, positioning *SoxE1* and *SoxE2* at the base of the SoxE family but distant from lamprey SoxE3.

Our expression data support a scenario for lamprey specific *SoxE* gene evolution as outlined above. *SoxE3* is expressed in the velar mesenchyme of the first arch, and in the perichondrial cells surrounding the chondrocytes of the posterior branchial arches, whereas *SoxE1* and *SoxE2* are expressed in the lamprey specific cartilage cells of the branchial basket (Fig. 2 and supplementary movies). Thus, *SoxE3* and *Sox9* expression in branchial arches may reflect an

ancestral SoxE role in a homologous series of arches, while *SoxE1* and *SoxE2* expression in pharyngeal arches represent roles specific to lamprey cartilage. This scenario also has implications regarding neural crest evolution. *SoxE3* is not expressed either in premigratory or migratory neural crest whereas both *SoxE1* and *SoxE2* are expressed in these cells (Fig. 1). Such a hypothetical scenario, supported by our results, would suggest that the neural crest role of a *SoxE* gene arose subsequent to the initial *SoxE* duplication event. In gnathostomes, *Sox9* and *Sox10* play roles in neural crest specification and differentiation, but in the lamprey *SoxE1* and *SoxE2* assumed the *SoxE* role in neural crest development.

Supplementary References

1. McCauley, D. W. & Bronner-Fraser, M. Conservation of Pax gene expression in ectodermal placodes of the lamprey. Gene 287, 129-39 (2002).
2. Thompson, J. D., Higgins, D. G. & Gibson, T. J. Clustal-W - Improving the Sensitivity of Progressive Multiple Sequence Alignment through Sequence Weighting, Position-Specific Gap Penalties and Weight Matrix Choice. Nucleic Acids Research 22, 4673-4680 (1994).
3. Hall, T. A. Bioedit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucl. Acids Symp. Ser. 41, 95-98 (1999).
4. Saitou, N. & Nei, M. The Neighbor-Joining Method - a New Method for Reconstructing Phylogenetic Trees. Japanese Journal of Genetics 61, 611-611 (1986).
5. Kumar, S., Tamura, K., Jakobsen, I. B. & Nei, M. MEGA2: molecular evolutionary genetics analysis software. Bioinformatics 17, 1244-1245 (2001).
6. Hardisty, M. W. & Potter, I. C. The Biology of Lampreys (Academic Press, London, 1981).
7. Jarvik, E. Basic structure and evolution of vertebrates (Academic Press, London, 1980).
8. Liem, K. F., Bemis, W. E., Walker, W. F. J. & Grande, L. Functional anatomy of the vertebrates: An evolutionary perspective (Thomson, Belmont, CA, 2001).
9. Morrison, S. L., Campbell, C. K. & Wright, G. M. Chondrogenesis of the branchial skeleton in embryonic sea lamprey, Petromyzon marinus. Anat Rec 260, 252-67 (2000).