Cartilaginous Fishes Provide Insights into the Origin, Diversification, and Sexually Dimorphic Expression of Vertebrate Estrogen Receptor Genes

Grant L. Filowitz,[†,1] Rajendhran Rajakumar,[†,1] Katherine L. O'Shaughnessy,[†,1] and Martin J. Cohn*[,1,2]

[1]Department of Molecular Genetics and Microbiology, UF Genetics Institute, University of Florida, Gainesville, FL

[2]Department of Biology, University of Florida, Gainesville, FL

[†]These authors contributed equally to this work.

***Corresponding author:** E-mail: mjcohn@ufl.edu.

Associate editor: John Parsch

Abstract

Vertebrate estrogen receptors (ERs) perform numerous cell signaling and transcriptional regulatory functions. ERα (*Esr1*) and ERβ (*Esr2*) likely evolved from an ancestral receptor that duplicated and diverged at the protein and *cis*-regulatory levels, but the evolutionary history of ERs, including the timing of proposed duplications, remains unresolved. Here we report on identification of two distinct ERs in cartilaginous fishes and demonstrate their orthology to ERα and ERβ. Phylogenetic analyses place the ERα/ERβ duplication near the base of crown gnathostomes (jawed vertebrates). We find that ERα and ERβ from little skate (*Leucoraja erinacea*) and mammals share key subtype-specific residues, indicating conserved protein evolution. In contrast, jawless fishes have multiple non-orthologous *Esr* genes that arose by parallel duplications. *Esr1* and *Esr2* are expressed in subtype-specific and sexually dimorphic patterns in skate embryos, suggesting that ERs might have functioned in sexually dimorphic development before the divergence of cartilaginous and bony fishes.

Key words: **estrogen receptor, chondrichthyan, vertebrate evolution, sex steroid evolution, sexual dimorphism.**

Estrogen receptors (ERs) regulate an array of physiological and developmental processes in male and female vertebrates. Most gnathostomes (jawed vertebrates) possess two distinct ERs, ERα and ERβ, which are encoded by the *Esr1* and *Esr2* genes, respectively (Thornton 2001; Katsu et al. 2008; Katsu, Kohno, et al. 2010; Katsu, Taniguchi, et al. 2010). Ligand binding activates ERs either by triggering translocation to the nucleus, where ERs bind estrogen response elements (EREs) to regulate transcription, or by initiating nongenomic signaling at extranuclear sites, such as the plasma membrane (Bjornstrom and Sjoberg 2005; Levin and Hammes 2016). The evolutionary origin of ERα and ERβ is not well resolved, and previous phylogenetic studies have led to debate over the relationships of ERs in cyclostomes (lampreys and hagfishes, the extant jawless fishes) and gnathostomes. Over the last two decades, four alternative models have been proposed for the evolution of ERs in vertebrates. The first model (fig. 1A) suggests that cyclostomes possess a pro-ortholog of the gnathostome ERα/ERβ (Thornton 2001). The second model (fig. 1B) places cyclostome ER at the base of the gnathostome ERα clade, implying that the ERα/ERβ duplication occurred at the base of vertebrates, that cyclostomes have a *bona fide* ERα, and that ERβ was subsequently lost (Baker and Chandsawangbhuwana 2008). The third model (fig. 1C) places one cyclostome ER at the base of the gnathostome ERβ clade, whereas the other cyclostome ER falls out as a sister to the ERα/ERβ clade (Baker et al. 2014; Nishimiya et al. 2017). The fourth model (fig. 1D), similar to the trees published by Thornton (2001), places cyclostome ERs outside of the gnathostome ERα/ERβ clades (Katsu et al. 2016).

The lack of resolution between the jawed and jawless vertebrate *Esr* sequences has been confounded, in part, by a paucity of data from chondrichthyans (cartilaginous fishes, including sharks, skates, rays, chimaeras, and their relatives), which occupy a critical phylogenetic position. In an effort to resolve the evolutionary history of vertebrate ER genes, we enriched taxonomic sampling of chondrichthyans and performed molecular phylogenetic analyses, protein homology modeling, and in situ hybridization studies. Here we report on the identification of true orthologs of osteichthyan ERα and ERβ in elasmobranchs and in a holocephalan, indicating that ERα and ERβ arose prior to the divergence of chondrichthyans and osteichthyans.

Results

We identified, cloned, and sequenced two potential ER orthologs in the little skate *Leucoraja erinacea* (Chondrichthyes: Elasmobranchii: Batoidea; see supplementary materials and methods, Supplementary Material online). Based on initial protein motif prediction, both ER orthologs possessed the canonical ER motifs, including the ligand-independent transactivation domain (AF-1), a DNA-binding domain (DBD), a hinge region, and a ligand-binding domain (LBD/AF-2) (supplementary fig. S1, Supplementary Material online). Furthermore, we found high percent identity and sequence conservation of these critical motifs between the *L. erincacea*

Letter

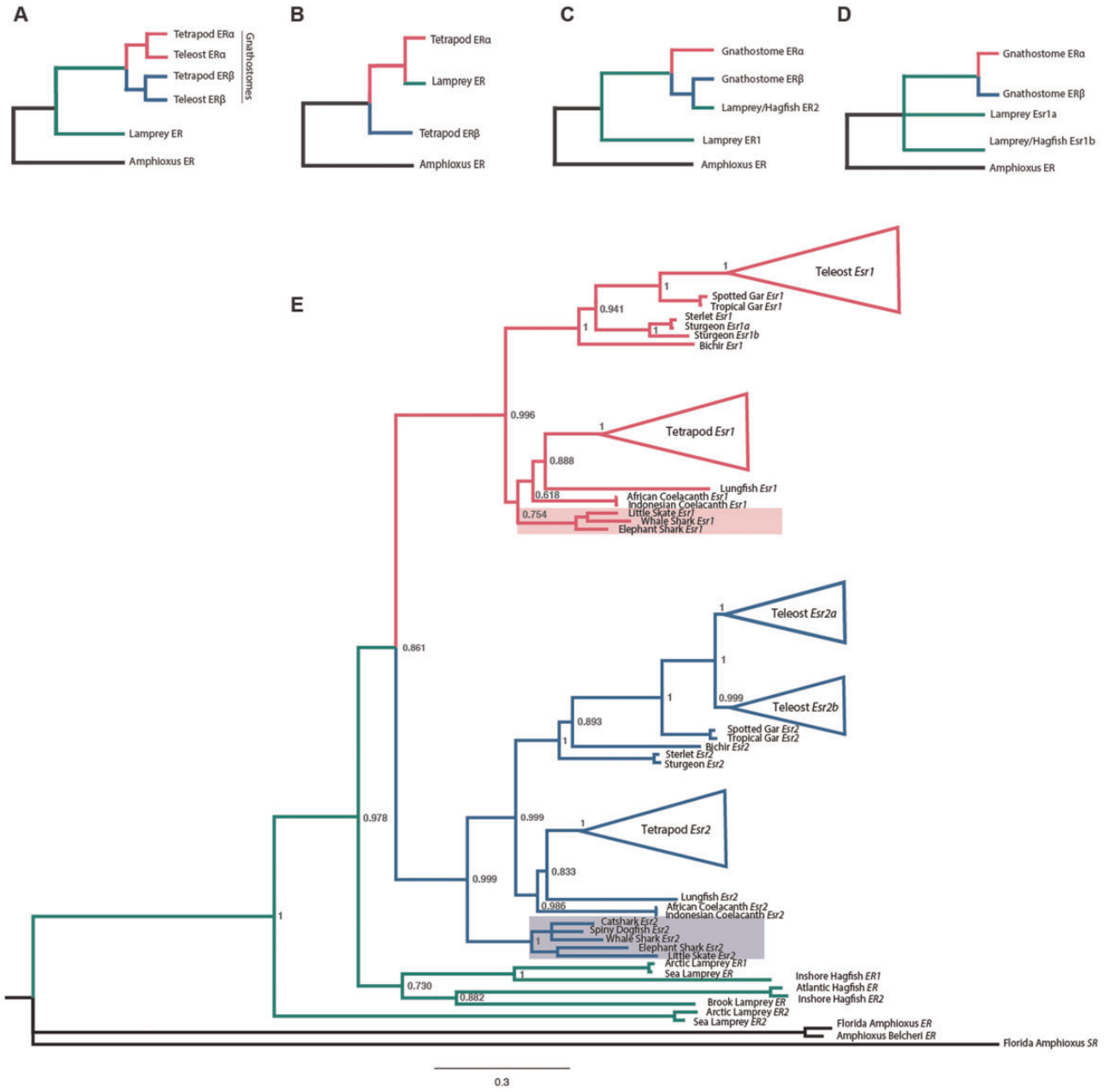

Fig. 1. Evolution of vertebrate ERs. Gnathostome ERα clade is highlighted in pink, ERβ clade in blue, and lamprey/hagfish (cyclostome) species in green. Trees from previous studies of ER evolution show contrasting topologies and different patterns of gene duplication. (*A*) Model 1 from Thornton (2001). (*B*) Model 2 from Baker and Chandsawangbhuwana (2008). (*C*) Model 3 from Baker *et al.* (2014) and Nishimiya *et al.* (2017). (*D*) Model 4 from Katsu *et al.* (2016). (*E*) Nucleotide tree generated from the new data reported in this study (full version shown in supplementary fig. S4, Supplementary Material online). All node values are posterior probabilities. Shaded boxes show the positions of chondrichthyan species.

ERs and the ERs of other vertebrates (supplementary fig. S1*A* and *B*, Supplementary Material online), suggesting orthology to ERα and ERβ.

To test the hypothesis that the two putative ER sequences from *L. erinacea* are orthologs of vertebrate ERα and ERβ, we performed phylogenetic reconstructions using ERα and ERβ sequences from osteichthyans, putative ER sequences from elephant shark (*Callorhinchus milii*; Chondrichthyes: Holocephali), whale shark (*Rhincodon typus;* Chondrichthyes: Elasmobranchii: Selachii), and catshark (*Scyliorhinus torazame*; Chondrichthyes: Elasmobranchii: Selachii), and all available ER sequences from cyclostomes and cephalochordates (Venkatesh et al. 2014; Read et al. 2017; Katsu et al. 2016) (supplementary table S1, Supplementary Material online). The steroid receptor (SR) of *Amphioxus floridae* served as an outgroup (supplementary table S1, Supplementary Material online). Bayesian phylogenetic analysis using full-length amino acid sequences (see supplementary materials and methods, Supplementary Material online) placed a single skate ER within each of the two gnathostome ER clades (supplementary fig. S2*A*, Supplementary Material online). The tree topology was similar to the third model (fig. 1C), with some cyclostome sequences falling within the respective ERβ clade, whereas

the remaining sequences grouped independently of the ERα/β clades. Though this topology was generally well supported by posterior probabilities (major clades ≥0.73), long branches within cyclostomes suggested long-branch attraction (LBA) bias. Analysis of the alignment revealed cyclostome-specific insertion/deletions. To test for LBA, we removed suspect sequences in different combinations and re-evaluated tree topology. These subsequent analyses yielded different tree topologies depending on which sequence(s) were eliminated (supplementary fig. S2*B*–*E*, Supplementary Material online), suggesting that our initial tree (and those of other studies) was likely affected by LBA artifacts.

To limit the influence of LBA, we next restricted our analysis specifically to the LBD of the receptors, which is highly conserved across species but, unlike the DBD, retains subtype-specific residues at key positions that may provide adequate informative characters for phylogenetic resolution (see supplementary materials and methods, Supplementary Material online). The LBD-specific analysis resulted in a tree topology in which all cyclostome sequences formed clades independent of gnathostome ERα/β (supplementary fig. S3, Supplementary Material online; major node support ≥0.89), and identified skate sequences as ERα and ERβ (supplementary fig. S3, Supplementary Material online, highlighted taxa). Although promising, limiting the data set to only the LBD resulted in some taxa (e.g., cyclostomes, amphibians) having identical sequences, and, as a consequence, these were unresolvable. Therefore, we analyzed the corresponding nucleotide sequences of the LBD (supplementary table S1 and supplementary materials and methods, Supplementary Material online). In contrast to the protein-level analysis, this reconstruction had increased resolution, particularly with respect to the cyclostomes (fig. 1*E*; supplementary fig. S4, Supplementary Material online). To confirm that LBA was not affecting these results, we systematically removed hagfish and lamprey sequences and evaluated the data set under equivalent parameters. In all cases, the tree topology was unchanged and support remained high at major nodes (compare supplementary fig. S5*A*–*C* with supplementary fig. S4, Supplementary Material online). Taken together, these results show that chondrichthyans have two genes that encode ERs, and these are orthologs of *Esr1*/ERα and *Esr2*/ERβ, consistent with Thornton's original hypothesis (fig. 1*A* and *E*; Thornton 2001).

In order to infer ligand-binding properties of the skate receptors, we used protein homology modeling of the LBD to estimate ER structure and function in *L. erinacea* (supplementary materials and methods, Supplementary Material online). Our results show the secondary and tertiary structures of the LBD of both skate and human ERα and ERβ are strikingly similar, consisting of 12 antiparallel α-helices (supplementary fig. S6*A* and *B*, Supplementary Material online). As in humans, 11 of these helices fold into a 3-layered "wedge-shaped" molecular scaffold that maintains a ligand-binding cavity (Brzozowski et al. 1997; Ascenzi et al. 2006). The remaining secondary structural elements are a small two-stranded antiparallel β-sheet and a final α-helix, which are located at the ligand-binding portion of the molecule and flank the main three-layered motif (Brzozowski et al. 1997; Ascenzi et al. 2006). The accuracy of these models was supported by global model quality estimation (GMQE) scores of 0.95 and 0.87, respectively (see supplementary materials and methods, Supplementary Material online). In addition to the strong structural conservation of the skate ER orthologs, they also possess the amino acid residues in the ligand-binding pocket that are known to facilitate binding of estradiol and other receptor agonists (pink boxes in supplementary fig. S6*A* and *B*, Supplementary Material online; see also Kuiper et al. 1997). Finally, the LBD of human ERα and ERβ possesses amino acid substitutions that alter the conformation of their ligand-binding pocket and configure subtype-specific binding of certain molecules (Paech et al. 1997; Barkhem et al. 1998; Paige et al. 1999). These substitutions were also identified in the skate predictive model (black boxes in supplementary fig. S6*A* and *B*, Supplementary Material online). Therefore, the conformation of the skate and human ERα and ERβ LBD are highly similar, and the predicted skate structures possess vital subtype-specific residues, suggesting that functional divergence of the LBD in these two receptors is conserved in skates.

To determine whether the predicted structures of skate ERα and ERβ are conserved in other chondrichthyan lineages, we performed homology modeling of holocephalan and shark ER proteins. Like the skate proteins, ERα and ERβ from *C. milii* (elephant shark) shared strong structural conservation with the human orthologs (GMQE = 0.87 and 0.82, respectively; supplementary figs. S7*A* and *B*, Supplementary Material online), as did ERβ from *Rhincodon typus* (whale shark; GMQE = 0.78; supplementary fig. S7*C*, Supplementary Material online). Interestingly, an amino acid substitution within the LBD of ERβ was found in all chondricthyans but not in humans (supplementary figs. S6 and S7, Supplementary Material online, ERβ, alignment position 76). This amino acid change, M76L, also has been described in lungfish and some amphibians and it does not confer any appreciable differences in ERβ-specific activity (Katsu et al. 2008; Katsu, Taniguchi, et al. 2010). These results suggest that the features of skate ERα and ERβ proteins are broadly applicable to chondrichthyans.

Although our LBD homology modeling suggests that skate ERs can bind estrogenic ligands, we also examined the AF-1 N-terminal domain, which is associated with transcriptional activity in response to ligand-dependent and ligand-independent activation of the ERs (Pettersson et al. 2000; Metivier et al. 2001; Zwart et al. 2010; Arao et al. 2012). The presence of a predicted α-helical domain in ERα, but not ERβ, is associated with the different transactivation potentials of the receptors in several vertebrates (Metivier et al. 2000, 2001; Zwart et al. 2010; Fuchs et al. 2013). Hydrophobicity cluster analysis of the AF-1 region in skates revealed an α-helix present in a similar position within the skate ERα subtype only (supplementary fig. S8*A*, Supplementary Material online; boxed). Thus, skate ERα possesses an important secondary structure that has been characterized as a key difference associated with functional divergence of the two ER subtypes.

To predict whether skate ERα and ERβ are capable of DNA interaction, we analyzed the DBD of both receptors.

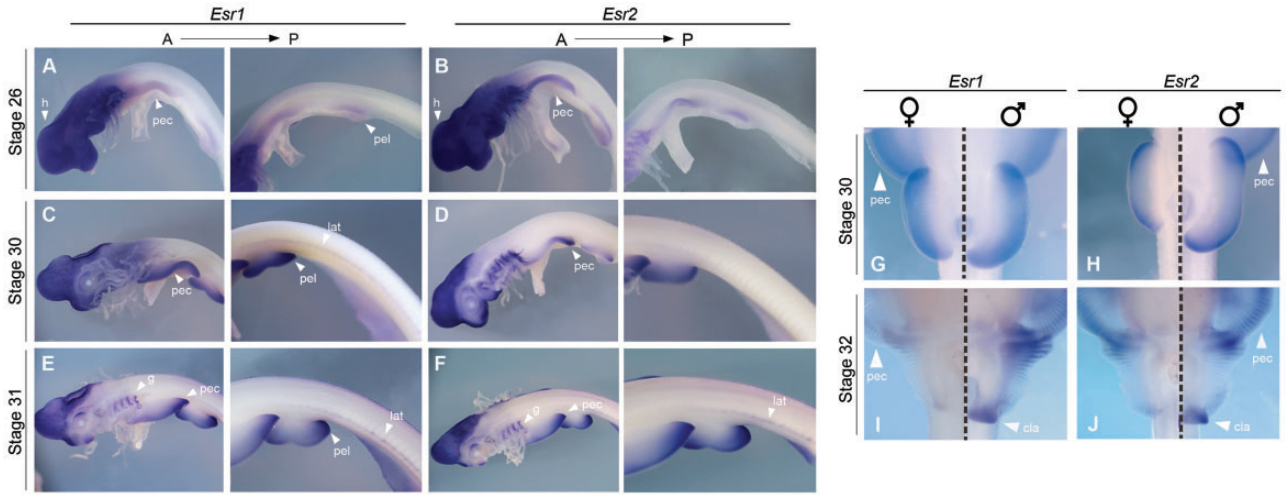

Fig. 2. Expression of *Esr1* and *Esr2* in *Leucoraja erinacea* embryos. Whole mount in situ hybridizations; anterior is to left in (*A*)–(*F*) and top in (*G*)–(*J*). (*A*) Stage 26 embryo showing expression of *Esr1*. Arrowheads point to expression domains in early fin buds and head. (*B*) Stage 26 embryo showing *Esr2* expression in domains similar to *Esr1* (compare with *A*). Stage 30 embryos showing sustained expression of *Esr1* (*C*) and *Esr2* (*D*) in paired fins, and in the lateral line sensory system. Note that *Esr2* expression domains are more restricted than at stage 26. Stage 31 embryos showing expression of *Esr1* (*E*) and *Esr2* (*F*) in fins, gill arches, and the lateral line. Pelvic fins of male (right) and female (left) skates at stages 30 (*G*, *H*) and 32 (*I*, *J*) showing expression of *Esr1* (*G*, *I*) and *Esr2* (*H*, *J*). pec, pectoral fins; pel, pelvic fins; h, head; lat, lateral line; g, gill arches.

Alignment of this highly conserved region of ERα and ERβ showed that the "P" and "D" boxes of the DBD, the two structural elements that confer DNA binding capabilities (Schwabe et al. 1990, 1993), are fully conserved in skates (supplementary fig. S8*B*, Supplementary Material online). Indeed, this conservation extended to all of the ER orthologs that we examined, with the notable exceptions of arctic lamprey and sea lamprey ER2 (discussed below). Collectively, these structural models suggest that skate ERα and ERβ: 1) are capable of binding estrogenic ligands in a subtype-specific manner; 2) possess transcriptional responses that are subtype-specific; and 3) have the potential to bind ERE sequences.

We next investigated whether *Esr1* and *Esr2* have subtype-specific expression patterns during skate embryonic development. In situ hybridization analysis of embryos at stage 26 showed that *Esr1* and *Esr2* are expressed in the developing head, gill arches and early pectoral and pelvic fin buds; however, no expression was detected in the flank region between the emerging paired fins (fig. 2*A* and *B*). At stage 30, expression of both genes persisted in the cranial region (fig. 2*C* and *D*), but they showed divergent expression patterns elsewhere. Specifically, *Esr1 was* detected throughout the fins buds, whereas *Esr2* showed restriction to the posterior region of the pectoral fin and the anterior and posterior regions of the pelvic fin (compare fig. 2*C* and *D*). Furthermore, *Esr1* was expressed throughout the gill arches at stage 30, whereas *Esr2* showed stronger expression in posterior arches (fig. 2*C* and *D*; supplementary fig. S9, Supplementary Material online). We also detected expression of both receptors in the lateral line, the superficial chain of mechanosensory organs (fig. 2*C* and *D*). At stage 31, *Esr1* began to show more intense staining in the posterior gill arches and *Esr2* expression was even more posteriorly restricted (fig. 2*E* and *F*; supplementary fig. S9, Supplementary Material online). Thus, *Esr1* and *Esr2* show subtle differences in their temporal dynamics of expression.

We previously demonstrated that *androgen receptor* (*AR*) is expressed in developing pelvic fins of skates, and that AR signaling regulates sexually dimorphic development of claspers, which are the copulatory organs on male pelvic fins (O'Shaughnessy et al. 2015). Furthermore, in tetrapod limbs, sexually dimorphic development of the digits is regulated by both AR and ER signaling (Zheng and Cohn 2011). We therefore investigated whether ER subtypes show sexually dimorphic expression patterns in skate fin buds. At early stages of pelvic fin development (stages 28 and 29), *Esr1* and *Esr2* are expressed throughout the fin bud mesenchyme and in the cloaca (supplementary fig. S10*A*–*D*, Supplementary Material online). At stage 30, when sexual differentiation of the pelvic fins begins (O'Shaughnessy et al. 2015), *Esr1* and *Esr2* begin to show enhanced staining in the clasper-forming region of male pelvic fins (fig. 2*G* and *H*). In female pelvic fins, *Esr1* and *Esr2* expression patterns are similar to males, but the *Esr1* domain appeared broader than *Esr2* (fig. 2*G* and *H*). By stage 32, *Esr1* and *Esr2* showed strong expression anteriorly and posteriorly (in the clasper bud) in male pelvic fins. In contrast, *Esr1* showed little staining in the female pelvic fin at stage 32, but *Esr2* was strong anteriorly and weaker posteriorly (fig. 2*I* and *J*). Taken together, these results show that skate *Esr1* and *Esr2* have subtype-specific and sexually dimorphic expression patterns during embryonic development.

Discussion

Our results show that chondrichthyans have *bona fide* orthologs of *Esr1* and *Esr2*, the genes that encode ERα and ERβ. Comparisons of the predicted ERα and ERβ structures to those of other gnathostomes showed conservation of residues that characterize the α and β subtypes, which are necessary for ligand binding. Furthermore, the presence of an α-helical domain in the AF-1 region of the ERα ortholog, but

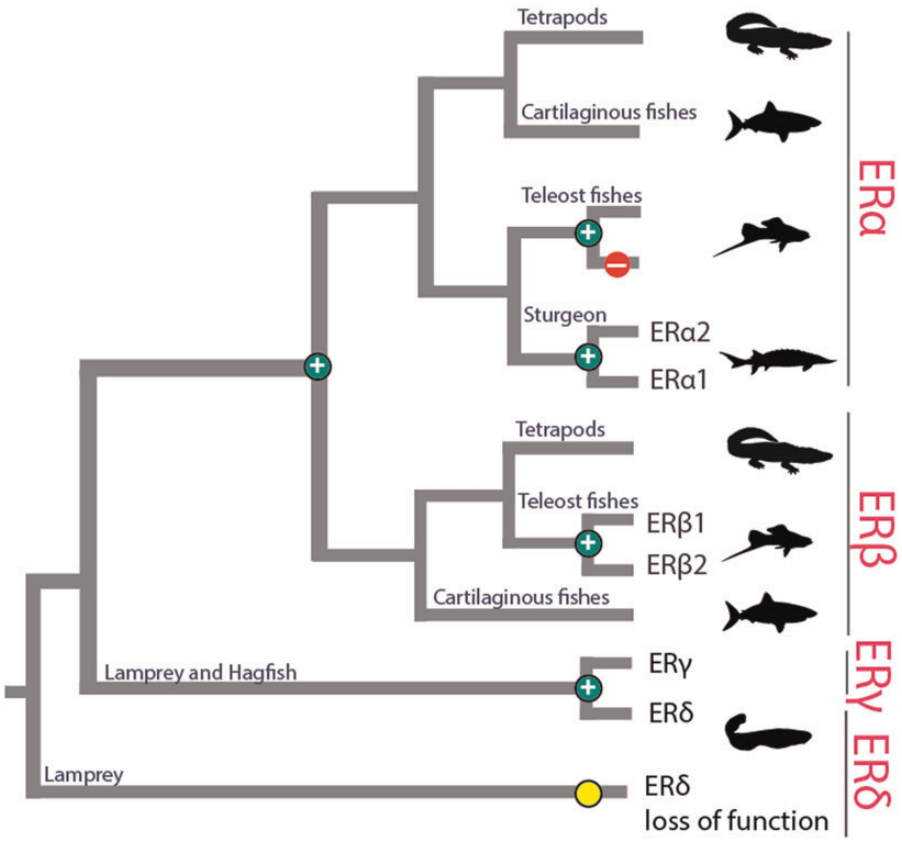

Fig. 3. The results presented here show that chondrichthyans have two ER subtypes, ERα and ERβ, encoded by the *Esr1* and *Esr2* genes, respectively. The duplication that gave rise to ERα and ERβ occurred near the base of gnathostomes, prior to the divergence of chondrichthyans and osteichthyans. Cyclostome ERs underwent parallel duplications and, because they lack strict orthology to gnathostome ERα and ERβ, we propose that they are named ERδ/γ. Black circles represent gene duplication; white circle indicates gene loss; gray circle indicates degeneration (loss of ligand and DNA binding activity; Katsu et al. 2016) It is noted the lamprey ERδ represented here is either in the process of degrading, or has acquired ligand-independent activity.

not ERβ, is consistent with the subtype-specific secondary structures that are found in teleosts and humans (Metivier et al. 2000, 2001). Collectively, these data suggest that *Esr1* and *Esr2* arose by duplication of the ancestral *Esr1/2* gene before the chondrichthyan and osteichthyan lineages diverged, and they had evolved subtype-specific characteristics (including sequence, protein structure, and spatiotemporal expression dynamics) associated with development of sexually dimorphic morphology.

Based on phylogenetic evidence, Thornton (2001) suggested that the ancestral *Esr* gene duplicated in gnathostomes to produce *Esr1* and *Esr2*. At the time, only a single lamprey ER had been identified, and the conclusion that the ER duplication occurred in gnathostomes was consistent with a WGD (1R) before the vertebrate radiation, followed by a gnathostome-specific WGD (2R). As additional ERs were identified in cyclostomes, the resolution of their phylogenetic positions was confounded by a deficiency in sampling and by the LBA artifacts that we described above. Our phylogenetic reconstruction, which includes the newly cloned *L. erinacea* *Esr1*/ERα and *Esr2*/ERβ, supports the existence of two separate paralogy groups of vertebrate ER genes, gnathostome *Esr1* and *Esr2*, and cyclostome genes that we designate *Esrγ* and *Esrδ*. This new proposed naming scheme is intended to resolve the dissonance in the literature, as these cyclostome sequences have been called various other names based on conflicting phylogenetic placements (e.g., ERα and ERβ, ER1 and ER2, and *Esr1a* and *Esr1b* in Baker et al. 2014; Katsu et al. 2016; Nishimiya et al. 2017, respectively). Thus, the results presented here validate the hypothesis that *Esr1* and *Esr2* arose by a gnathostome-specific duplication of an ancestral *Esr* gene, as originally proposed by Thornton (2001).

Our tree places *Esrδ* from arctic lamprey and sea lamprey as an outgroup to *Esr1*, *Esr2*, *Esrɣ*, and the other *Esrδ* (fig. 1*E*). Although this could imply that *Esrδ* is a pro-ortholog of the gnathostome and cyclostome ERs, an alternate explanation is that this positioning reflects the degeneration or functional divergence of *Esrδ* (Katsu et al. 2016). Arctic lamprey Esrδ was shown to be incapable of binding estrogen-related ligands, which is primarily thought to be due to an insertion of 4 amino acids, in addition to other destabilizing forces on the ligand binding pocket (Katsu, et al. 2016) (refer to supplementary fig. S6C, Supplementary Material online). Presumably, sea lamprey Esrδ groups with the arctic lamprey Esrδ, because they share this insertion in their LBD, as well as unique mutations within their DBD including the D-box

(supplementary fig. S8B, Supplementary Material online). Teleosts have two *Esr2* genes (*Esr2a* and *Esr2b*; refer to supplementary table S1, Supplementary Material online) due to the teleost-specific WGD (3R), and the single copy of *Esr1* suggests that its paralog was lost. A duplicated *Esr1* has been reported in sturgeon, which may be the result of an *Acipenseriform*-specific duplication (Katsu et al. 2008). Thus, several independent duplications of ERs have occurred throughout vertebrates (fig. 3).

Although vertebrates shared a WGD (1R), a separate parallel WGD (2R) in cyclostomes and gnathostomes (Mehta et al. 2013) could have led to *Esrγ* and *Esrδ* in the former and *Esr1* and *Esr2* in the latter. Alternatively, if vertebrates shared a 1R/2R WGD (Smith et al. 2013), then extensive independent sequence evolution would be required for gnathostome ERs to group separately from those of cyclostomes. A third possibility is that the two ER paralogs in gnathostomes resulted from a gnathostome-specific WGD (2R), whereas cyclostome ER paralogs arose by tandem or chromosomal-scale duplications. Support for each of these three scenarios can be found in other studies, and the timing of the WGD events in vertebrates remains a topic of debate (Smith and Keinath 2015; Smith et al. 2018). Hemoglobin and myoglobin duplicated independently in cyclostomes and gnathostomes (Schwarze et al. 2015), as did clade A fibrillar collagen genes (Zhang and Cohn 2006), which can explain the lack of obvious 1:1 orthology. Based on the recently assembled sea lamprey germline genome (and the associated scaffolds), *Esrγ* and *Esrδ* are on different large scaffolds and, therefore, are not likely syntenic (Smith et al. 2018). Furthermore, the lack of related neighboring genes suggests that large chromosomal segment duplication is unlikely; in contrast, the *Esr1* and *Esr2* regions in jawed vertebrates do contain paralogous neighboring genes.

In addition to the localized expression of both ERs in skate paired appendages (fins and male claspers), we also found polarized expression in the gill arches. The structure of the cartilaginous gill bars of sharks led Gegenbauer to posit that pectoral fins may have evolved from posterior gill arches (Gegenbauer 1878). Although this hypothesis has received little support from the fossil record (Coates and Cohn 1998), it has been suggested that molecular similarities between chondrichthyan gill arches and fins, such as polarized expression of sonic hedgehog (*Shh*) and fibroblast growth factors (*Fgfs*), may reflect shared ancestry of these structures (Gillis et al. 2009; Gillis and Hall 2016). Although our data does not necessarily support or refute Gegenbauer's hypothesis, we find it intriguing that *Esr1* and *Esr2* are expressed (i) in the developing gill arches in spatial domains that are strikingly similar to *Shh* and its receptor *Patched1* (*Ptch1*); (ii) during fin bud initiation; (iii) throughout paired fin development; and (iv) within the developing claspers of male skates. We interpret these findings as evidence for deep conservation of the ancient gene regulatory network (GRN) that governs development of appendages, including paired fins and gills (Shubin et al. 2009; Pieretti et al. 2015) rather than evidence for an actual embryonic morphological transformation of gills to fins. Moreover, a recent analysis of skate pelvic fin development showed that another SR, the AR, controls transcription of *Hand2*, an upstream regulator of *sonic hedgehog*, to initiate sexually dimorphic fin (clasper) development in males (O'Shaughnessy et al. 2015). This led to the proposal that AR may have played a role in the evolution and development of vertebrate paired appendages through its cooperation with the appendage development GRN. Our new data on *Esr1* and *Esr2* in skate fin development, together with previous findings that AR and ER control sexually dimorphic development of tetrapod digits (Zheng and Cohn 2011), further support the hypothesis that sex steroids played a role in the evolution of vertebrate appendages. We propose that the *Esr1/2* duplication in gnathostomes allowed further modularization of sex hormone signaling and contributed to the evolution of sexually dimorphic development of vertebrate morphology.

Supplementary Material

Supplementary data are available at *Molecular Biology and Evolution* online.

Acknowledgments

The authors thank Scott Bennett and David Remsen of the Marine Biological Laboratory (Woods Hole, MA) for access to embryos, and members of our lab for stimulating discussions and critical comments on this work. This project was supported by a Howard Hughes Medical Institute Early Career Scientist award to M.J.C. K.L.O. was supported in part by a University of Florida Alumni Fellowship. R.R. was supported by an Natural Sciences and Engineering Research Council of Canada postdoctoral fellowship.

References

Arao Y, Hamilton KJ, Goulding EH, Janardhan KS, Eddy EM, Korach KS. 2012. Transactivating function (AF) 2-mediated AF-1 activity of estrogen receptor alpha is crucial to maintain male reproductive tract function. *Proc Natl Acad Sci U S A.* 109(51): 21140–21145.

Ascenzi P, Bocedi A, Marino M. 2006. Structure-function relationship of estrogen receptor alpha and beta: impact on human health. *Mol Aspects Med.* 27(4): 299–402.

Baker ME, Chandsawangbhuwana C. 2008. Motif analysis of amphioxus, lamprey and invertebrate estrogen receptors: toward a better understanding of estrogen receptor evolution. *Biochem Biophys Res Commun.* 371(4): 724–728.

Baker ME, Nelson DR, Studer RA. 2014. Origin of the response to adrenal and sex steroids: roles of promiscuity and co-evolution of enzymes and steroid receptors. *J Steroid Biochem Mol Biol.* (121): 12–24.

Barkhem T, Carlsson B, Nilsson Y, Enmark E, Gustafsson J, Nilsson S. 1998. Differential response of estrogen receptor alpha and estrogen receptor beta to partial estrogen agonists/antagonists. *Mol Pharmacol.* 54(1): 105–112.

Bjornstrom L, Sjoberg M. 2005. Mechanisms of estrogen receptor signaling: convergence of genomic and nongenomic actions on target genes. *Mol Endocrinol.* 19(4): 833–842.

Brzozowski AM, Pike AC, Dauter Z, Hubbard RE, Bonn T, Engstrom O, Ohman L, Greene GL, Gustafsson JA, Carlquist M. 1997. Molecular basis of agonism and antagonism in the oestrogen receptor. *Nature* 389(6652): 753–758.

Coates MI, Cohn MJ. 1998. Fins, limbs, and tails: outgrowths and axial patterning in vertebrate evolution. *BioEssays* 20(5): 371–381.

Fuchs S, Nguyen HD, Phan TT, Burton MF, Nieto L, de Vries-van Leeuwen IJ, Schmidt A, Goodarzifard M, Agten SM, Rose R, et al.

2013. Proline primed helix length as a modulator of the nuclear receptor-coactivator interaction. *J Am Chem Soc.* 135(11): 4364–4371.
Gegenbauer C. 1878. Elements of comparative anatomy. London: Macmillan and Company.
Gillis JA, Dahn RD, Shubin NH. 2009. Shared developmental mechanisms pattern the vertebrate gill arch and paired fin skeletons. *Proc Natl Acad Sci U S A.* 106(14): 5720–5724.
Gillis JA, Hall BK. 2016. A shared role for sonic hedgehog signalling in patterning chondrichthyan gill arch appendages and tetrapod limbs. *Development* 143(8): 1313–1317.
Katsu Y, Cziko PA, Chandsawangbhuwana C, Thornton JW, Sato R, Oka K, Takei Y, Baker ME, Iguchi T. 2016. A second estrogen receptor from Japanese lamprey (*Lethenteron japonicum*) does not have activities for estrogen binding and transcription. *Gen Comp Endocrinol.* 236:105–114.
Katsu Y, Kohno S, Hyodo S, Ijiri S, Adachi S, Hara A, Guillette LJ Jr, Iguchi T. 2008. Molecular cloning, characterization, and evolutionary analysis of estrogen receptors from phylogenetically ancient fish. *Endocrinology* 149(12): 6300–6310.
Katsu Y, Kohno S, Narita H, Urushitani H, Yamane K, Hara A, Clauss TM, Walsh MT, Miyagawa S, Guillette LJ Jr, et al. 2010. Cloning and functional characterization of Chondrichthyes, cloudy catshark, Scyliorhinus torazame and whale shark, Rhincodon typus estrogen receptors. *Gen Comp Endocrinol.* 168(3): 496–504.
Katsu Y, Taniguchi E, Urushitani H, Miyagawa S, Takase M, Kubokawa K, Tooi O, Oka T, Santo N, Myburgh J, et al. 2010. Molecular cloning and characterization of ligand- and species-specificity of amphibian estrogen receptors. *Gen Comp Endocrinol.* 168(2): 220–230.
Kuiper GG, Carlsson B, Grandien K, Enmark E, Haggblad J, Nilsson S, Gustafsson JA. 1997. Comparison of the ligand binding specificity and transcript tissue distribution of estrogen receptors alpha and beta. *Endocrinology* 138(3): 863–870.
Levin ER, Hammes SR. 2016. Nuclear receptors outside the nucleus: extranuclear signalling by steroid receptors. *Nat Rev Mol Cell Biol.* 17(12): 783–797.
Mehta TK, Ravi V, Yamasaki S, Lee AP, Lian MM, Tay BH, Tohari S, Yanai S, Tay A, Brenner S, et al. 2013. Evidence for at least six Hox clusters in the Japanese lamprey (Lethenteron japonicum). *Proc Natl Acad Sci U S A.* 110(40): 16044–16049.
Metivier R, Penot G, Flouriot G, Pakdel F. 2001. Synergism between ERalpha transactivation function 1 (AF-1) and AF-2 mediated by steroid receptor coactivator protein-1: requirement for the AF-1 alpha-helical core and for a direct interaction between the N- and C-terminal domains. *Mol Endocrinol.* 15(11): 1953–1970.
Metivier R, Petit FG, Valotaire Y, Pakdel F. 2000. Function of N-terminal transactivation domain of the estrogen receptor requires a potential alpha-helical structure and is negatively regulated by the A domain. *Mol Endocrinol.* 14(11): 1849–1871.
Nishimiya O, Katsu Y, Inagawa H, Hiramatsu N, Todo T, Hara A. 2017. Molecular cloning and characterization of hagfish estrogen receptors. *J Steroid Biochem Mol Biol.* 165(Pt B): 190–201.
O'Shaughnessy KL, Dahn RD, Cohn MJ. 2015. Molecular development of chondrichthyan claspers and the evolution of copulatory organs. *Nat Commun.* 6:6698.
Paech K, Webb P, Kuiper GG, Nilsson S, Gustafsson J, Kushner PJ, Scanlan TS. 1997. Differential ligand activation of estrogen receptors ERalpha and ERbeta at AP1 sites. *Science* 277(5331): 1508–1510.
Paige LA, Christensen DJ, Gron H, Norris JD, Gottlin EB, Padilla KM, Chang CY, Ballas LM, Hamilton PT, McDonnell DP, et al. 1999. Estrogen receptor (ER) modulators each induce distinct conformational changes in ER alpha and ER beta. *Proc Natl Acad Sci U S A.* 96(7): 3999–4004.
Pettersson K, Delaunay F, Gustafsson JA. 2000. Estrogen receptor beta acts as a dominant regulator of estrogen signaling. *Oncogene* 19(43): 4970–4978.
Pieretti J, Gehrke AR, Schneider I, Adachi N, Nakamura T, Shubin NH. 2015. Organogenesis in deep time: a problem in genomics, development, and paleontology. *Proc Natl Acad Sci U S A.* 112(16): 4871–4876.
Read TD, Petit RA, Joseph SJ, Alam MT, Weil R, Ahmad M, Bhimani R, Vuong JS, Haase CP, Webb DH, Tan M. 2017. Draft sequencing and assembly of the genome of the world's largest fish, the whale shark: rhincodon typus Smith 1828. *BMC genomics*. 18(1): 532.
Schwabe JW, Chapman L, Finch JT, Rhodes D. 1993. The crystal structure of the estrogen receptor DNA-binding domain bound to DNA: how receptors discriminate between their response elements. *Cell* 75(3): 567–578.
Schwabe JW, Neuhaus D, Rhodes D. 1990. Solution structure of the DNA-binding domain of the oestrogen receptor. *Nature* 348(6300): 458–461.
Schwarze K, Singh A, Burmester T. 2015. The full globin repertoire of turtles provides insights into vertebrate globin evolution and functions. *Genome Biol Evol.* 7(7): 1896–1913.
Shubin N, Tabin C, Carroll S. 2009. Deep homology and the origins of evolutionary novelty. *Nature* 457(7231): 818–823.
Smith JJ, Keinath MC. 2015. The sea lamprey meiotic map improves resolution of ancient vertebrate genome duplications. *Genome Res.* 25(8): 1081–1090.
Smith JJ, Kuraku S, Holt C, Sauka-Spengler T, Jiang N, Campbell MS, Yandell MD, Manousaki T, Meyer A, Bloom OE, et al. 2013. Sequencing of the sea lamprey (Petromyzon marinus) genome provides insights into vertebrate evolution. *Nat Genet.* 45:415–421, 421e411-412.
Smith JJ, Timoshevskaya N, Ye C, Holt C, Keinath MC, Parker HJ, Cook ME, Hess JE, Narum SR, Lamanna F, et al. 2018. The sea lamprey germline genome provides insights into programmed genome rearrangement and vertebrate evolution. *Nat Genet.* 50(2): 270–277.
Thornton JW. 2001. Evolution of vertebrate steroid receptors from an ancestral estrogen receptor by ligand exploitation and serial genome expansions. *Proc Natl Acad Sci U S A.* 98(10): 5671–5676.
Venkatesh B, Lee AP, Ravi V, Maurya AK, Lian MM, Swann JB, Ohta Y, Flajnik MF, Sutoh Y, Kasahara M, et al. 2014. Elephant shark genome provides unique insights into gnathostome evolution. *Nature* 505(7482): 174–179.
Zhang G, Cohn MJ. 2006. Hagfish and lancelet fibrillar collagens reveal that type II collagen-based cartilage evolved in stem vertebrates. *Proc Natl Acad Sci U S A.* 103(45): 16829–16833.
Zheng Z, Cohn MJ. 2011. Developmental basis of sexually dimorphic digit ratios. *Proc Natl Acad Sci U S A.* 108(39): 16289–16294.
Zwart W, de Leeuw R, Rondaij M, Neefjes J, Mancini MA, Michalides R. 2010. The hinge region of the human estrogen receptor determines functional synergy between AF-1 and AF-2 in the quantitative response to estradiol and tamoxifen. *J Cell Sci.* 123(Pt 8): 1253–1261.

Supplementary Materials

Title: Cartilaginous fishes provide insights into the origin, diversification, and sexually dimorphic expression of vertebrate estrogen receptor genes

Authors: Filowitz, Rajakumar, O'Shaughnessy, and Cohn

Supplementary Figures S1-S10
Supplementary Table S1
Supplementary Materials and Methods

Supplementary Figures S1-S10

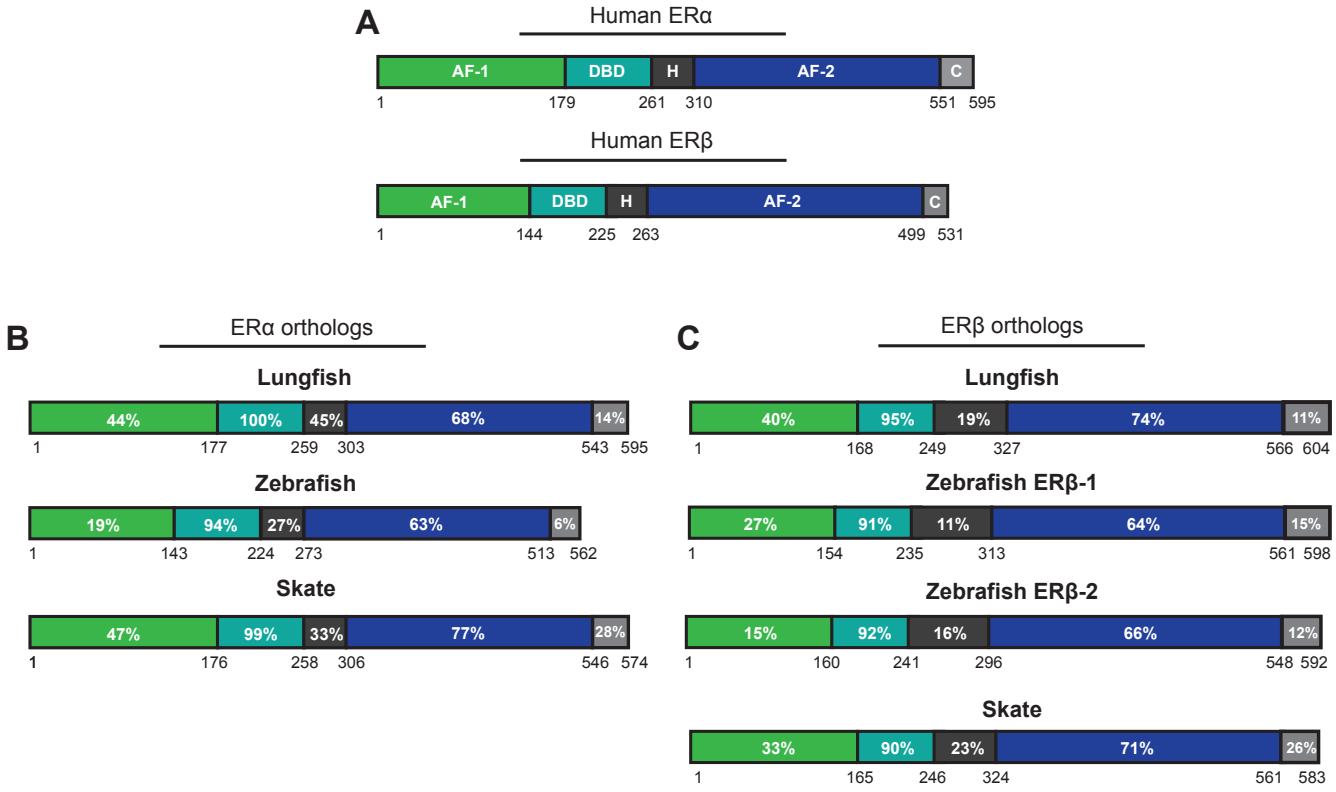

Figure S1. Structural domains of ERα and ERβ. Each color block corresponds to a specific motif, as denoted in the human proteins. Light green represents the ligand-independent transactivation function (AF-1), teal is the DNA-binding domain (DBD), dark gray is the hinge region (H), dark blue is the ligand-binding domain (AF-2), and light gray is the C-terminus (C). **(A)** Human ERα and ERβ. **(B)** Domains of ERα in lungfish, zebrafish, and the predicted skate ortholog. Denoted within each domain is the percent identity to the human sequence. **(C)** Domains of ERβ, with percent identity denoted within each functional domain. Note that, overall, the skate sequences display higher percent identity with human than do the zebrafish orthologs.

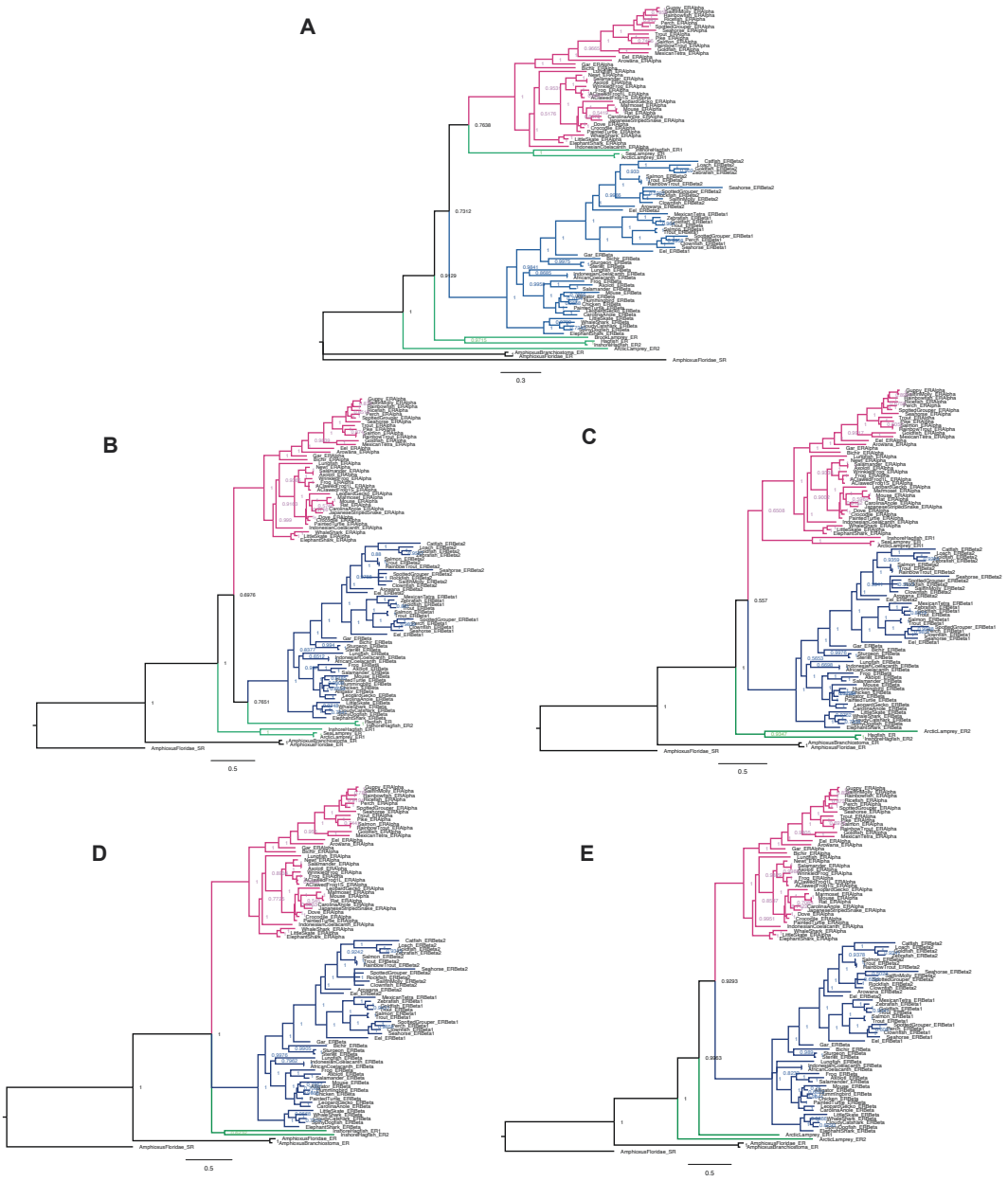

Figure S2. Identification of long branch attraction (LBA) using full length amino acid sequences with Bayesian inference. Gnathostome ERα clade is highlighted in pink, ERβ clade in blue, and lamprey/hagfish (cyclostome) species in green. **(A)** Tree includes all species in this study. Note that some lamprey ER sequences fall within the gnathostome ERα clade. All major nodes show high support. **(B)** Removal of the arctic lamprey ER2 sequence changed the tree topology, causing some cyclostome sequences to group with the gnathostome ERβ clade. This topology was also well supported (posterior probabilities ≥0.69). **(C-E)** To further test for LBA, we removed brook lamprey and sea lamprey ER sequences (C), removed all lamprey sequences, leaving only hagfish sequences (D), and used only arctic lamprey sequences (E). These iterations each resulted in different but well supported tree topologies, demonstrating that the ER dataset likely suffered from LBA artifacts. See Supplementary Table S1 for sequence information.

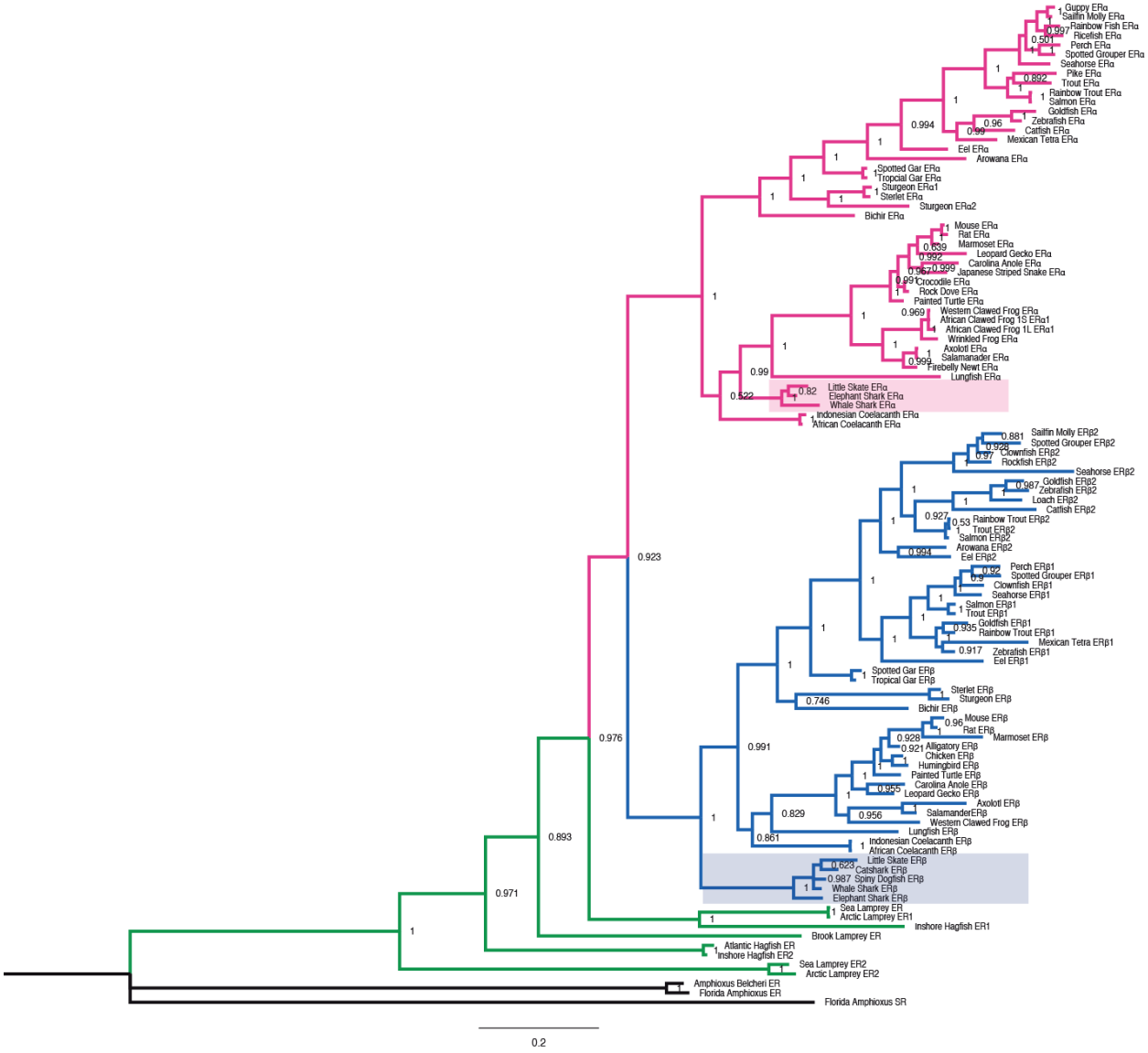

Figure S3. Full, uncollapsed amino acid tree using only the ligand-binding domain. All node values are posterior probabilities. Green branches denote cyclostome species; blue indicates gnathostome ERα; pink indicates gnathostome ERβ. Shaded boxes show the positions of chondrichthyan species.

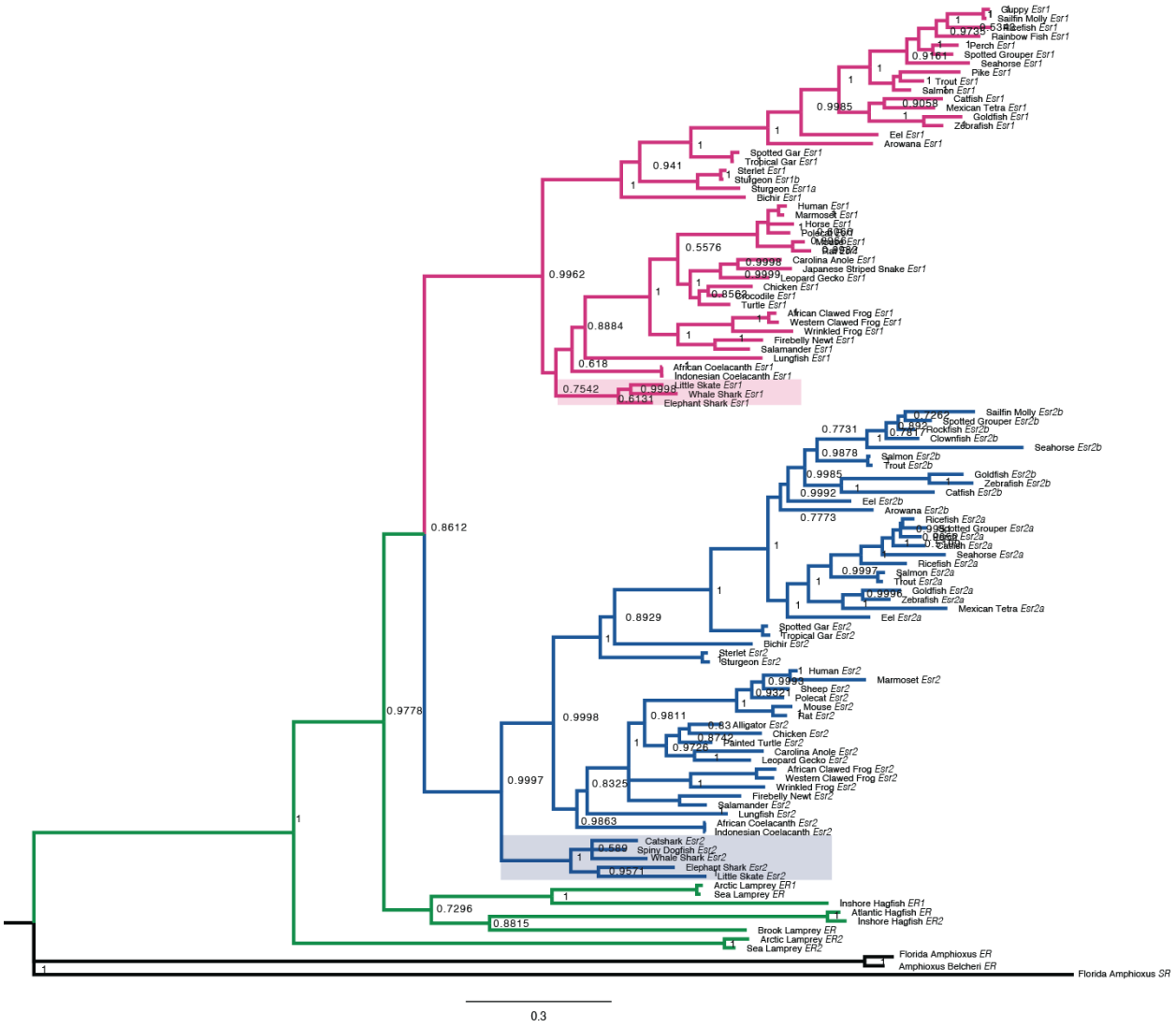

Figure S4. Full, uncollapsed nucleotide tree using only the ligand-binding domain. All node values are posterior probabilities. Green branches denote cyclostome species; blue indicates gnathostome *Esr1*; pink indicates gnathostome *Esr2*. Shaded boxes show the positions of chondrichthyan species.

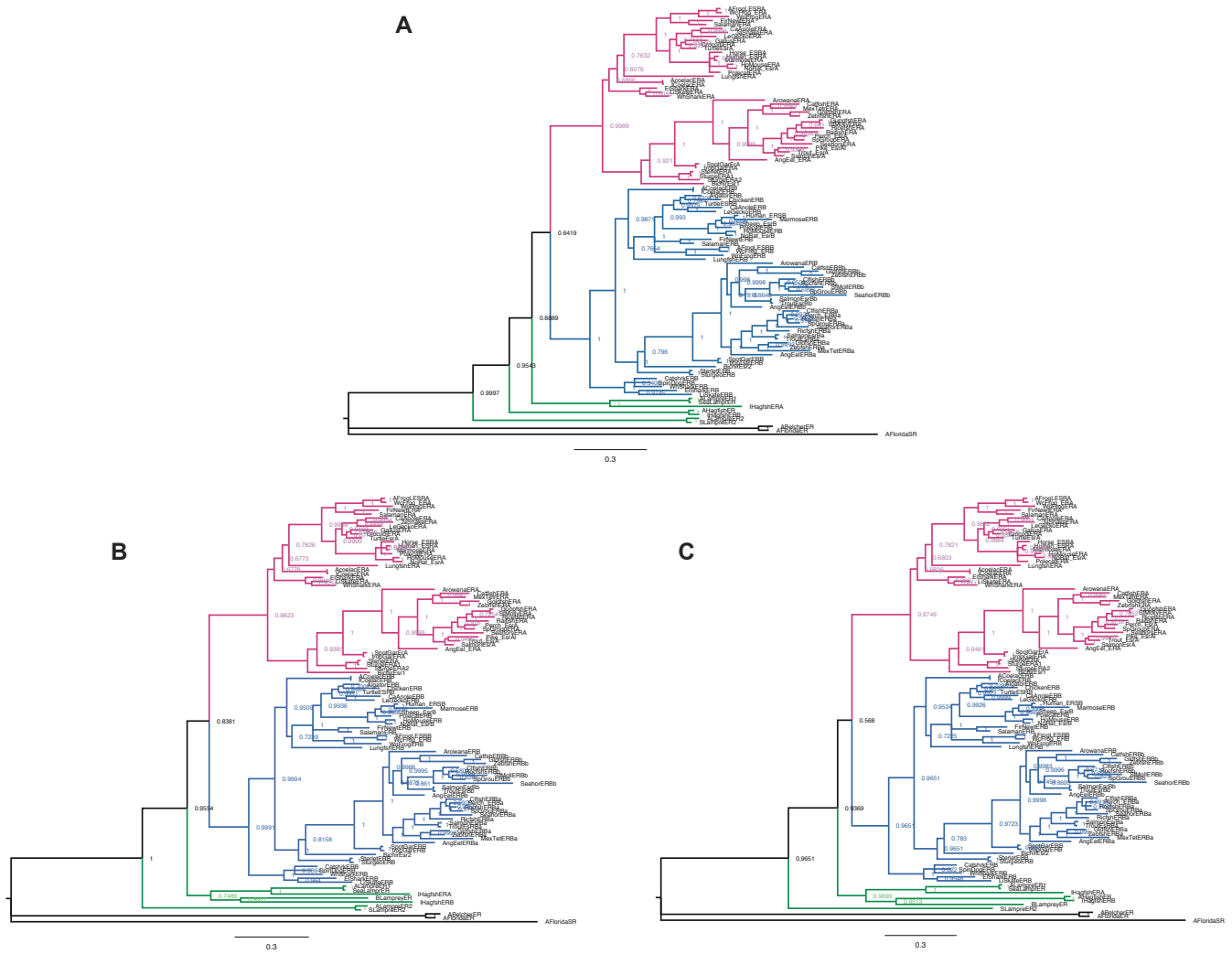

Figure S5. Phylogenetic reconstructions using only the ligand-binding domains of ERs eliminate long branch attraction artifacts. Nucleotide trees were analyzed by Bayesian inference (see Supplementary Materials and Methods). Generation of well-supported topologies, regardless of inclusion or exclusion of cyclostome sequences in subsequent reconstruction runs, demonstrates robustness of trees. **(A)** Brook lamprey ER was excluded from the dataset. **(B)** Atlantic hagfish ER was excluded. **(C)** Arctic lamprey ER2 sequence was excluded. See Supplementary Table S1 for details of species and sequences.

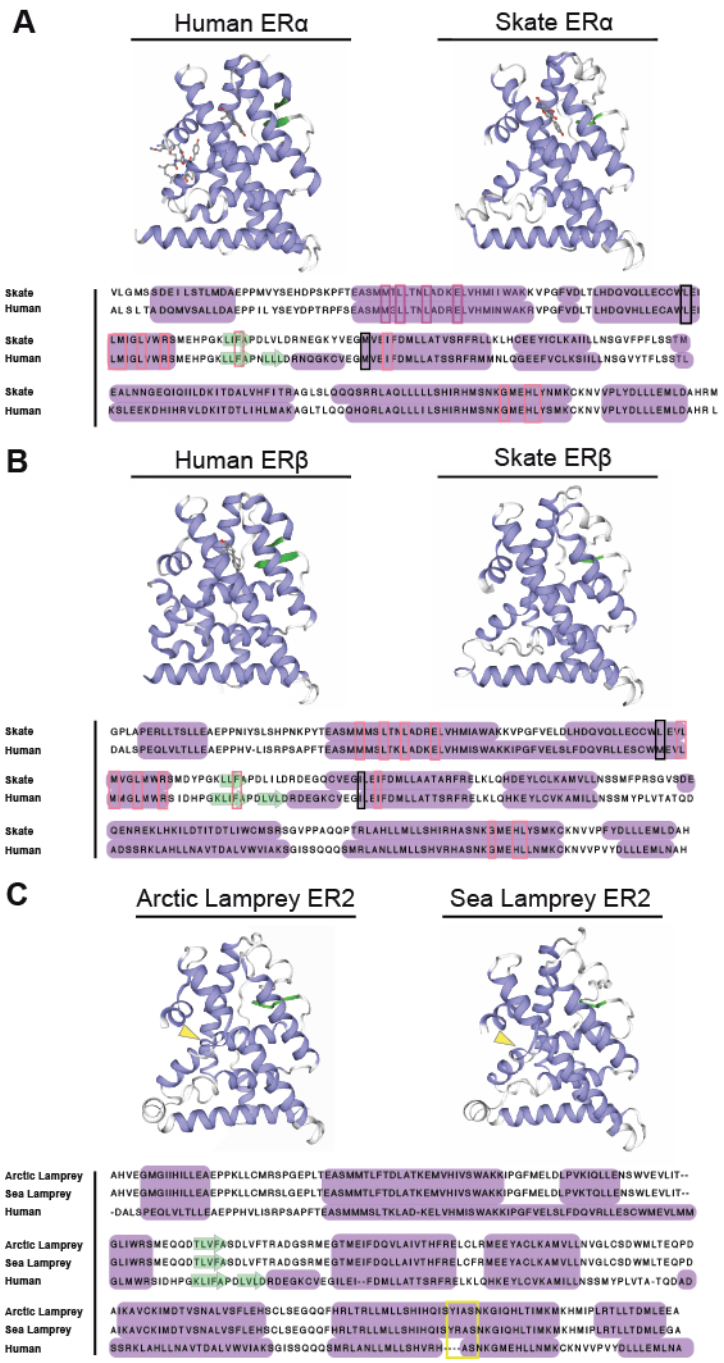

Figure S6. Homology modeling of skate and lamprey ER protein structures. ER structures were generated with SWISS-MODEL utilizing human ER crystal structures as templates. **(A, B)** Skate ERα and ERβ are highly similar to their human orthologs. Residues necessary for ligand binding are conserved (pink boxes in alignments). Black boxes show conservation of subtype-specific residues in the ligand-binding domain. **(C)** Models of the arctic lamprey ER2 (left) using human ERβ as the template. Our predicted structure of the LBD is similar to that described by Katsu et al., 2016. Specifically, the insertion (yellow box in alignment) constitutes a destabilizing looping of the alpha helix (yellow arrow) that renders this ER unresponsive to estrogen ligand (Katsu, et al. 2016). Sea lamprey ER2 contains a similar mutation within the LBD, which results in a similar destabilizing loop (yellow arrow in C) to that observed in arctic lamprey. This suggests that sea lamprey ER2 also has a loss of ligand-binding function, potentially mediated by accumulation of this unique protein sequence (yellow box in alignment; compare lampreys to human).

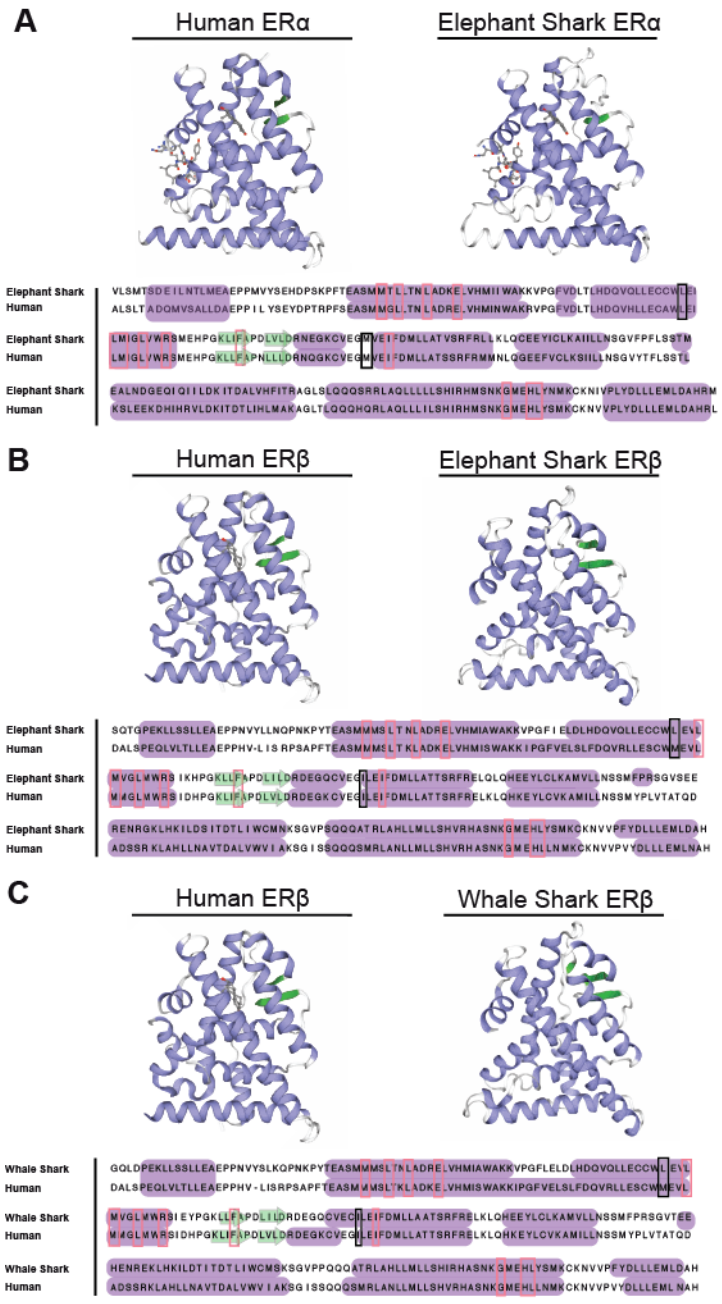

Figure S7. Homology modeling of holocephalan and shark ER protein structures. As described in Figure S6, ER models were generated with SWISS-MODEL utilizing human ER crystal structures as templates. **(A, B)** Elephant shark ERα and ERβ, like the skate proteins, are highly similar to human ER orthologs. **(C)** The structure of whale shark ERβ is also highly similar to the human ERβ protein. Whale shark ERα was not estimated due to the availability of only a partial LBD sequence. In all models, the residues necessary for ligand binding are conserved and are highlighted in pink boxes. The black boxes show conservation of subtype-specific residues within the ligand-binding domain.

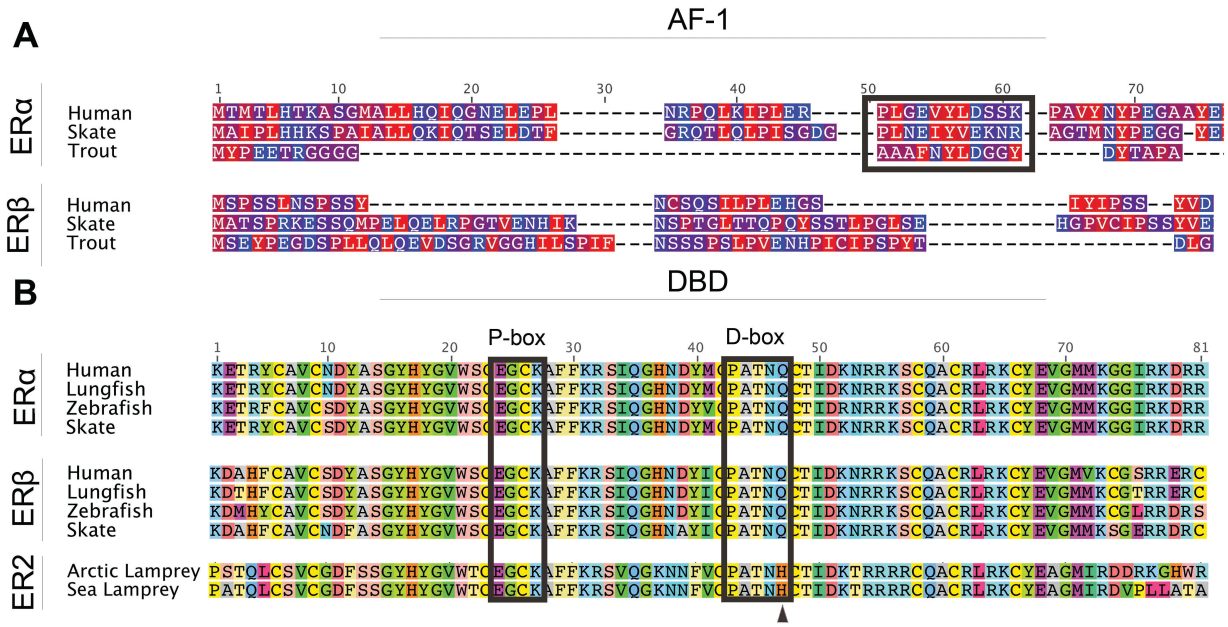

Figure S8. AF-1 N-terminal region and DNA binding domain. (A) Alignment of ERα and ERβ AF-1 region in human, skate, and trout. Alignment is colored by hydrophobicity, with red residues being hydrophobic, blue hydrophilic, and purple intermediate. Previous work has identified a potential α-helix in ERα by hydrophobic cluster analysis (HCA) in human and trout, and showed that this region was important to ERα transcriptional activity (Metivier, et al. 2000; Petit, et al. 2000). We used HCA to corroborate those findings, and also identified a similar structure in skate (boxed). The predicted α-helix in skate has an iterative hydrophobic chain more similar to human than trout. Additionally, the secondary structure is in a closer relative position between skate and human (note gap in trout ERα). **(B)** DBD region of ERα and ERβ in multiple vertebrates. The black rectangles indicate the P-box and D-box, the two regions that confer ER-specific DNA recognition and constitute the protein dimer interface. There is complete conservation of amino acids in the P-box and D-box of the gnathostome ERs.

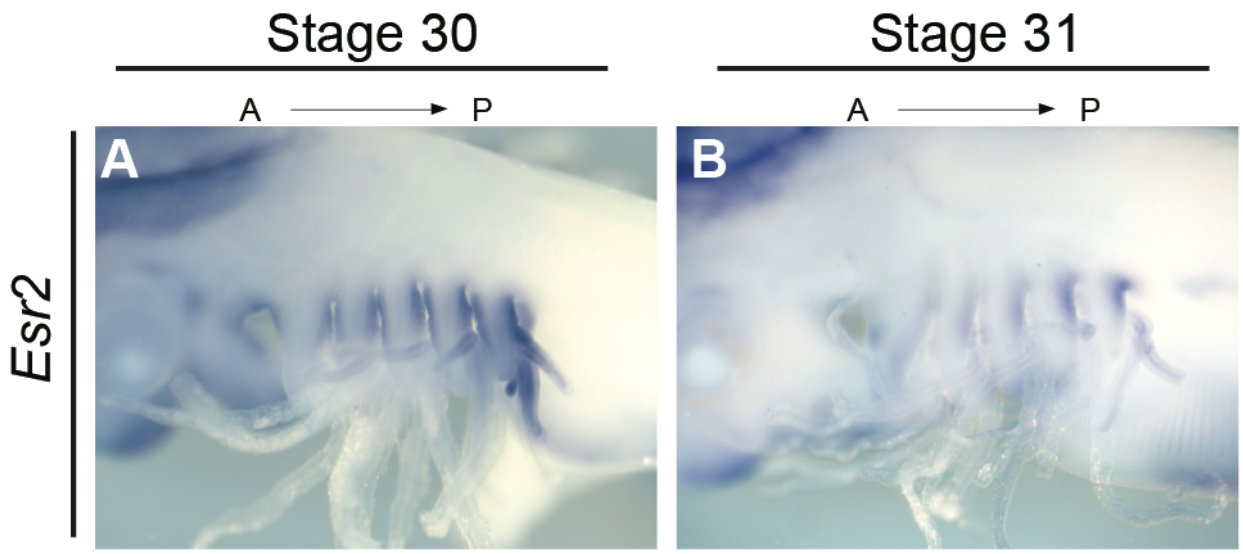

Figure S9. *Esr2* is expressed in the pharyngeal arches of skate embryos. Whole mount *in situ* hybridizations showing *Esr2* mRNA localization (purple signal). The anterior-posterior axis of the embryos is indicated by A→ P. *Esr2* is expressed in the anterior and posterior regions of the developing gill arches, though staining appears to be stronger posteriorly.

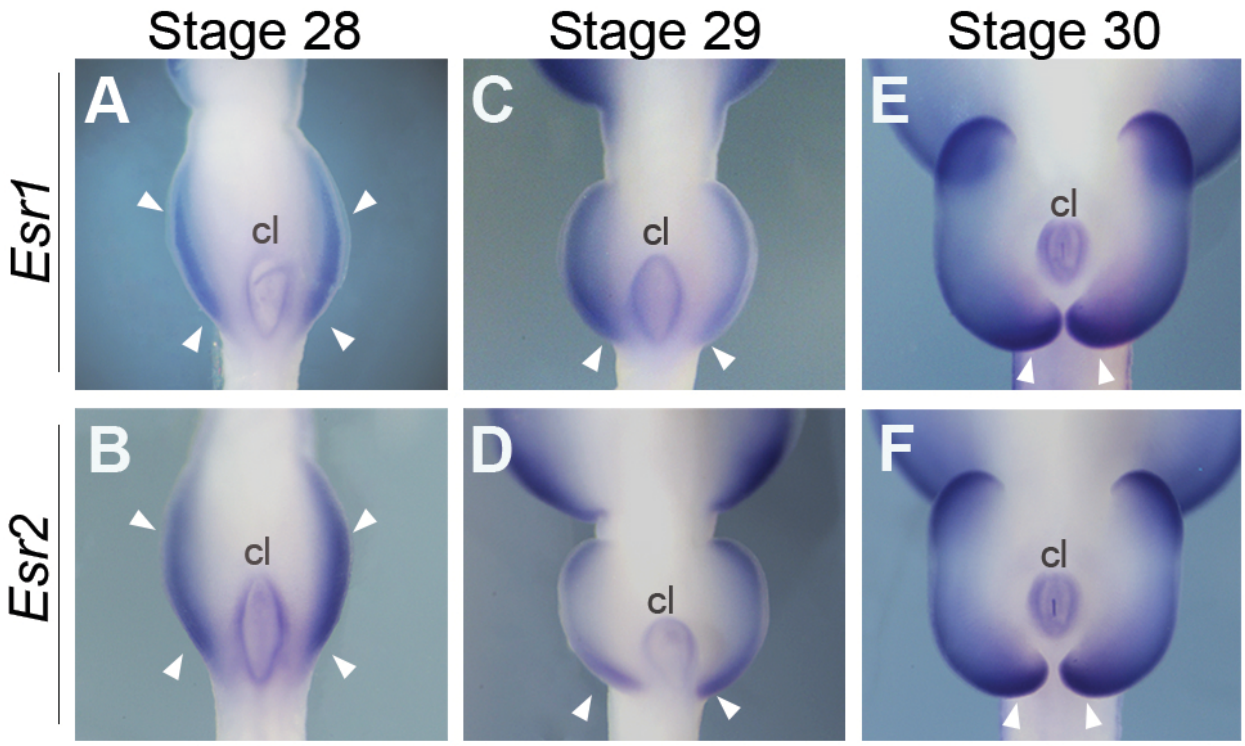

Figure S10. *Esr1* and *Esr2* are expressed in skate pelvic fin buds. Whole mount *in situ* hybridizations showing *Esr1* and *Esr2* mRNA localization (purple). The cloacal opening is marked by (cl) at the anterior margin. **(A, B)** *Esr1* and *Esr2* are expressed broadly throughout the early pelvic fin bud mesenchyme at stage 28**. (C, D)** At stage 29, expression of both genes begins to show stronger expression in the posterior region of the pelvic fin buds (arrowheads). **(E, F)** *Esr1* and *Esr2* expression in the pelvic fins persists through stage 30, when sexual differentiation of the fins is initiated. Note dark staining in the regions where the male copulatory organs (claspers) begin to appear as posterior expansions of the pelvic fin buds (white arrows)

Supplementary Table S1. Complete list of sequences used in phylogenetic analyses

Common Name	Scientific name	NCBI annotation	Proposed annotation	NCBI accession code
Chinese Amphioxus	*Branchiostoma belcheri*	ER	ER	AB510027
Florida Amphioxus	*Branchiostoma floridae*	ER	ER	EF554313
Arctic Lamprey	*Lethenteron camtschaticum*	ER1	Erγ	AB626148
Inshore Hagfish	*Eptatretus burgeri*	ER1	Erγ	KP987796.1
Sea Lamprey	*Petromyzon marinus*	ER1	Erγ	AY028456
Atlantic Hagfish	*Myxine glutinosa*	ER	Erδ	EU439936
Arctic Lamprey	*Lethenteron camtschaticum*	ER2	Erδ	AB626149
Inshore Hagfish	*Eptatretus burgeri*	ER2	Erδ	KP987797.1
Northern Brook Lamprey	*Ichthyomyzon fossor*	ER2	Erδ	GBEL01000002
Sea Lamprey	*Petromyzon marinus*	ER2	Erδ	APA19937.1
Whale Shark	*Rhincodon typus*	ER-like	Esr1	XM_020532945.1
Atlantic Salmon	*Salmo salar*	Esr1	Esr1	NM_001123592.1
Axolotl	*Ambystoma mexicanum*	Esr1	Esr1	AB524912
Carolina Anole	*Anolis carolinensis*	Esr1	Esr1	NM_001290517
Channel Catfish	*Ictalurus punctatus*	Esr1	Esr1	NM_001200074.1
Crocodile	*Crocodylus niloticus*	Esr1	Esr1	AB209933
Dove	*Columba livia*	Esr1	Esr1	NM_001282825
Elephant Shark	*Callorhinchus milii*	Esr1	Esr1	XM_007894403.1
European Eel	*Anguilla anguilla*	Esr1	Esr1	LN879034.1
Goldfish	*Carassius auratus*	Esr1	Esr1	JX440380
Guppy	*Poecilia reticulata*	Esr1	Esr1	NM_001297487.1
Horse	*Equus ferus caballus*	Esr1	Esr1	NM_001081772
Human	*Homo sapiens*	Esr1	Esr1	NM_000125
Indonesian Coelacanth	*Latimeria menadoensis*	Esr1	Esr1	HF562327
Japanese Puffer	*Takifugu rubripes*	Esr1	Esr1	XM_003971746.2
Japanese Striped Snake	*Elaphe quadrivirgata*	Esr1	Esr1	AB548295
Killifish	*Fundulus heteroclitus*	Esr1	Esr1	AY571785.1
Leopard Gecko	*Eublepharis macularius*	Esr1	Esr1	AB240528
Little Skate	*Leucoraja erinacea*	Esr1	Esr1	******
Lungfish	*Protopterus annectens*	Esr1	Esr1	AB435636
Marmoset	*Callithrix jacchus*	Esr1	Esr1	XM_008995270
Medaka	*Oryzias latipes*	Esr1	Esr1	XM_020714493.1
Mexican Tetra	*Astyanax mexicanus*	Esr1	Esr1	XM_007253897
Mouse	*Mus musculus*	Esr1	Esr1	NM_007956
Newt	*Cynops pyrrhogaster*	Esr1	Esr1	AB524908
Painted Turtle	*Chrysemys picta*	Esr1	Esr1	NM_001282246
Polecat	*Mustela putorius*	Esr1	Esr1	XM_004784330
Rainbowfish	*Melanotaenia fluviatilis*	Esr1	Esr1	GU319956.1
Rat	*Rattus norvegicus*	Esr1	Esr1	NM_012689

Salamander	*Hynobius tokyoensis*	Esr1	Esr1	AB524910
Silver Arowana	*Osteoglossum bicirrhosum*	Esr1	Esr1	LC057258.1
Spotted Gar	*Lepisosteus oculatus*	Esr1	Esr1	XM_006625845.2
Spotted Grouper	*Epinephelus coioides*	Esr1	Esr1	GU721076.1
Stickleback	*Gasterosteus aculeatus*	Esr1	Esr1	LC006094.1
Western Clawed Frog	*Xenopus tropicalis*	Esr1	Esr1	AY310902
Trout	*Oncorhynchus mykiss*	Esr1b	Esr1	NM_001124558.1
Sturgeon	*Acipenser schrenckii*	Esr1a	Esr1a	BAG82650.1
Sturgeon	*Acipenser schrenckii*	Esr1b	Esr1b	BAG82651.1
African Coelacanth	*Latimeria chalumnae*	Esr2	Esr2	XM_005986391
Alligator	*Alligator mississippiensis*	Esr2	Esr2	AB548298
Atlantic Salmon	*Salmo salar*	Esr2	Esr2	NM_001123577.1
Axolotl	*Ambystoma mexicanum*	Esr2	Esr2	AB524913
Carolina Anole	*Anolis carolinensis*	Esr2	Esr2	XM_008125840
Chicken	*Gallus gallus*	Esr2	Esr2	NM_204794
Cloudy Catshark	*Scyliorhinus torazame*	Esr2	Esr2	AB551715
Elephant Shark	*Callorhinchus milii*	Esr2	Esr2	XM_007910258.1
Firebelly Newt	*Cynops pyrrhogaster*	Esr2	Esr2	AB524909
Human	*Homo sapiens*	Esr2	Esr2	NM_001291723
Hummingbird	*Calypte anna*	Esr2	Esr2	XM_008501049
Indonesian Coelacanth	*Latimeria menadoensis*	Esr2	Esr2	HF562328
Japanese Sturgeon	*Acipenser schrenckii*	Esr2	Esr2	AB435633.1
Leopard Gecko	*Eublepharis macularius*	Esr2	Esr2	AB240529
Little Skate	*Leucoraja erinacea*	Esr2	Esr2	*********
Lungfish	*Protopterus annectens*	Esr2	Esr2	AB435637
Mouse	*Mus musculus*	Esr2	Esr2	NM_207707
Painted Turtle	*Chrysemys picta*	Esr2	Esr2	XM_005285890
Rat	*Rattus norvegicus*	Esr2	Esr2	NM_012754
Salamander	*Hynobius tokyoensis*	Esr2	Esr2	AB524911
Senegal Bichir	*Polypterus senegalus*	Esr2	Esr2	LC057257
Silver Arowana	*Osteoglossum bicirrhosum*	Esr2	Esr2	LC057259.1
Spiny Dogfish	*Squalus acanthias*	Esr2	Esr2	AF147746
Spotted Gar	*Lepisosteus oculatus*	Esr2	Esr2	XM_006632189.2
Western Clawed Frog	*Xenopus tropicalis*	Esr2	Esr2	NM_001040012
Whale Shark	*Rhincodon typus*	Esr2	Esr2	AB551716
Goldfish	*Carassius auratus*	Esr2	Esr2a	AF061269.1
Catfish	*Ictalurus punctatus*	Esr2-1-like	Esr2a	XM_017456575
Seahorse	*Hippocampus comes*	Esr2-like	Esr2a	XM_019870757.1
Clownfish	*Amphiprion melanopus*	Esr2a	Esr2a	HM185180
Killifish	*Fundulus heteroclitus*	Esr2a	Esr2a	AY570922.1
Medaka	*Oryzias latipes*	Esr2a	Esr2a	NM_001104702.1
Perch	*Perca flavescens*	Esr2a	Esr2a	DQ984125.1

Rockfish	*Sebastes schlegelii*	Esr2a	Esr2a	FJ646610
Spotted Grouper	*Epinephelus coioides*	Esr2a	Esr2a	GU721077.1
Zebrafish	*Danio rerio*	Esr2a	Esr2a	NM_180966.2
Eel	*Anguilla anguilla*	Esr2b	Esr2a	LN879036.1
European Eel	*Anguilla anguilla*	Esr2b	Esr2a	LN879036.1
Sailfin Molly	*Poecilia latipinna*	Esr2-1	Esr2a	KT022998.1
Rockfish	*Sebastes schlegelii*	Esr2a	Esr2a	FJ646610.3
Channel Catfish	*Ictalurus punctatus*	Esr2	Esr2b	NM_001200083.1
Japanese Puffer	*Takifugu rubripes*	Esr2	Esr2b	XM_003978635.2
Mexican Tetra	*Astyanax mexicanus*	Esr2-like	Esr2b	XM_007230932
Seahorse	*Hippocampus comes*	Esr2-like1	Esr2b	XM_019877457.1
European Eel	*Anguilla anguilla*	Esr2a	Esr2b	LN879035.1
Arowana	*Osteoglossum bicirrhosum*	Esr2b	Esr2b	LC057259.1
Atlantic Salmon	*Salmo salar*	Esr2b	Esr2b	JF798871.1
Clownfish	*Amphiprion melanopus*	Esr2b	Esr2b	HM185178.1
Goldfish	*Carassius auratus*	Esr2b	Esr2b	AF177465.1
Killifish	*Fundulus heteroclitus*	Esr2b	Esr2b	AY570923.1
Medaka	*Oryzias latipes*	Esr2b	Esr2b	NM_001128512.1
Rainbow Trout	*Oncorhynchus mykiss*	Esr2b	Esr2b	NM_001124570.1
Rockfish	*Sebastes schlegelii*	Esr2b	Esr2b	HQ452829
Spotted Grouper	*Epinephelus coioides*	Esr2b	Esr2b	GU721078.1
Zebrafish	*Danio rerio*	Esr2b	Esr2b	NM_174862.3
Florida Amphioxus	*Branchiostoma floridae*	SR	SR	EU371729.1

Supplementary Materials and Methods

Data Mining for Chondrichthyan ER Sequences

Initial searches for little skate (*Leucoraja erinacea*) ER sequences were performed on the National Center for Biotechnology Information (NCBI) sequence databases using BLASTn under the discontiguous megablast algorithm. West African lungfish (*Protopterus annectens*) full-length nucleotide sequences for *Esr1* (AB435636) and *Esr2* (AB435637) were used as queries to mine whole genome shotgun sequences of *L. erinacea* (taxid: 7782). Several hits for potential orthologs of *Esr1* (AESE012519307 and AESE010097393) and *Esr2* (AESE011520100) were recovered, but these were either short fragments or had poor overall query coverage. Therefore, these sequences were used as templates to clone *Esr* genes from *L. erinacea* by polymerase chain reaction (PCR). Predicted mRNA sequences for elephant shark (*Callorhinchus milii*) *Esr*1 (XM_007894403) and *Esr2* (XM_007910258) and a partial fragment for whale shark (*Rhincodon typus*) *Esr1* (XM_020512638.1) were recovered from NCBI RefSeq (see Supplementary Table S1).

Little Skate Husbandry and Tissue Collection

Leucoraja erinacea eggs were obtained from the Marine Biological Laboratory (Woods Hole, MA). Eggs were maintained in Fluval marine salt adjusted to 32 ppt at ambient temperature in tanks with both mechanical and biological filtration systems. For tissue collection in preparation for RNA extraction, embryos were removed from their egg cases and staged (Maxwell, et al. 2008). Male and female animals of developmental stages 31 and 32 were euthanized in MS-222 and the liver, gonads, and pelvic fins were dissected and preserved in RNAlater (Qiagen).

Cloning of Leucoraja erinacea ER orthologs

RNA was isolated using RNAeasy Mini Plus Kits (Qiagen) and quantified on a Nanodrop-1000 (Thermo Scientific). RNA integrity was evaluated by using a Bioanalyzer 2100 (Agilent Technologies). For cDNA synthesis, only RNA with a 260/280 >1.9 and a RNA integrity number (RIN) of $\geq$9.0 was used. For initial cloning experiments, 1 μg of RNA was used for cDNA synthesis using the Maxima cDNA kit (Thermo Scientific). Two potential *Esr* fragments were

amplified by PCR. Amplicons were separated on a 1.2% gel stained with ethidium bromide, and bands of ~500 bp were excised and purified using the Wizard SV Gel Clean-up System (Promega). The fragments were then ligated into pGEM-T Easy Vectors (Promega) and transformed into NEB Turbo Competent cells (New England Biolabs). Plasmids were harvested using the Wizard Plus SV minipreparation system (Promega) and then sequenced in both directions. Preliminary BLASTn searches showed that the two clones had high query coverage for *Esr1* and *Esr2,* respectively. Rapid amplification of cDNA ends (RACE)-PCR and mining of the *L. erinacea* genome were used to recover full-length transcripts of the putative *Esr1* and *Esr2* orthologs. To confirm the full-length sequences, RACE-PCR was carried out using the SMARTer 5'/3' system (Clontech) following the manufacturers protocol. Amplicons were ligated into either StrataClone PCR Vectors (Agilent) or pGEM-T Easy Vectors (Promega), which were introduced into NEB Turbo Competent cells by heat-shock. Plasmids were then isolated and sequenced as described above.

Analyses of ER Structural Domains

For multi-species comparisons of ERα and ERβ, skate *Esr1* and *Esr2* were translated and the functional domains for human (AAD52984 and AAC05985), lungfish (BAG82655 and BAG82656), and zebrafish (AAK16740 and AAK16742) were retrieved from NCBI. Sequences were then aligned manually in Se-Al v2.0a11 (courtesy of Andrew Rambaut). Percent identity (PID) was calculated relative to human sequences as follows: PID=100 (Identical residues/average length of sequence).

Phylogenetic Analyses

Preliminary phylogenetic reconstruction of the *Esr* family was conducted using the skate sequences reported here and the annotated *Esr* sequences from cyclostomes and gnathostomes. Cephalochordate steroid receptor (SR) was used as an outgroup representative. All sequences were imported into Se-Al v2.0a11, frame-corrected, translated, and trimmed at the 5' and 3' untranslated region. Sequences were aligned using TranslatorX (Abascal, et al. 2010) and then imported into GENEIOUS (Kearse, et al. 2012) for manual adjustment to compensate for large insertions/deletions. For analysis of the ligand-binding domain (LBD) after discovery of long branch attraction (LBA; see below), the LBD was isolated from each species

and then manually examined and aligned using GENEIOUS. Phylogenetic reconstructions of amino acid sequences were performed by Bayesian inference using MRBAYES 3.2.2 (Huelsenbeck and Ronquist 2001) via the Cipres Science Gateway (Miller 2010). We utilized the Poisson distribution as an amino acid substitution model, with four estimated gamma categories, and we performed each run for 5,000,000 generations with a burn-in of 25%. Run convergence was confirmed when the average standard deviation of split frequencies was $\leq$0.01. For phylogenetic analysis of nucleic acid sequences, we first used jModelTest 2.1.6 (Darriba, et al. 2012) to estimate an appropriate model of DNA substitution by AICc and BIC criteria. In both scenarios, the GTR+G model was selected with four estimated gamma categories, and we performed each run for 10,000,000 generations with a burn-in of 25%. As with the amino acid analysis described above, nucleotide sequence run convergence was confirmed when the average standard deviation of split frequencies was $\leq$0.01. All trees were read and edited in FigTreev1.4 (courtesy of Andrew Rambaut).

Testing for Long Branch Attraction (LBA)

We tested for LBA as previously described (Siddall and Whiting 1999; Pol and Siddall 2001). In short, potentially problematic sequences were identified and removed individually and in alternative combinations to allow for reevaluation of tree topologies by equivalent phylogenetic methods (described above). We considered an analysis to be affected by a LBA phenomenon when removal of a sequence altered the tree topology.

Protein Homology Modeling

Homology modeling was done using the SWISS-MODEL server (Guex, et al. 2009; Biasini, et al. 2014) utilizing human crystal structures as templates for ERα and ERβ (RCSB Protein Database ID 1gwr.2 (alpha) and 4j26.1 (beta). Human template structures were originally obtained at resolution by X-ray crystallography (2.40 Å for ERα and 2.30 Å for ERβ)(Warnmark, et al. 2002; Fuchs, et al. 2013). Our alignments yielded 77% and 71% conservation of amino acid residues between human and skate ERα and ERβ, respectively. Model quality estimations were evaluated using Global Model Quality Estimation and QMEAN scores (Benkert, et al. 2009), which were performed directly on the workspace. Both skate homology models received strong quality scores, indicating high accuracy of the predicted structure from the input

sequence. ERα for elephant shark (*Callorhinchus milii*) and ERβ for elephant shark and whale shark (*Rhincodon typus*) were modeled using the same crystal structure templates as described for skate. For lamprey homology modeling, the LBD of arctic lamprey (*Lethenteron camtschaticum*, BAM48574.1) and sea lamprey (*Petromyzon marinus*, APA19937.1) were estimated using human ERβ as the template (RCSB Protein Database 4j26.1). Our estimated structure for arctic lamprey was compared against previously published homology models for this species (Katsu, et al. 2016).

Secondary Structure Prediction

Secondary structure of the AF-1 domain was estimated using hydrophobicity cluster analyses (HCA)(Gaboriaud, et al. 1987), which was performed on the Mobyle server (Neron, et al. 2009). To compare our data with previous analyses of the AF-1 functional core (Metivier, et al. 2000; Metivier, et al. 2001), AF-1 domains of ERα and ERβ in human and trout were analyzed in addition to skate. Our analyses predicted an α-helix within the AF-1 region of ERα for all species tested. For clarity of presentation, we generated an alignment of ERα and ERβ (colored by hydrophobicity) and boxed the predicted secondary structure.

In Situ Hybridization

RNA *in situ* hybridization was performed with digoxigenin-labeled probes for *Esr1* and *Esr2* in *L. erinacea* embryos as previously described (Freitas and Cohn 2004; O'Shaughnessy, et al. 2015) with the following modifications: Proteinase K (PK) concentrations were 10 μg/ml for embryos younger than stage 28, and 20 μg/ml for embryos at stage 29 and older; embryos were not agitated during PK digestion and the reaction was stopped using 2 mg/ml glycine for five minutes. The color reaction was completed in BM Purple (Roche) at 4°C overnight, and embryos were dehydrated in a graded methanol series before imaging. Expression of each gene was examined in a minimum of 3 males and 3 females at each stage.

Bibliography for Supplementary Materials

Abascal F, Zardoya R, Telford MJ. 2010. TranslatorX: multiple alignment of nucleotide sequences guided by amino acid translations. Nucleic Acids Res 38:W7-13.

Benkert P, Kunzli M, Schwede T. 2009. QMEAN server for protein model quality estimation. Nucleic Acids Res 37:W510-514.

Biasini M, Bienert S, Waterhouse A, Arnold K, Studer G, Schmidt T, Kiefer F, Cassarino TG, Bertoni M, Bordoli L, et al. 2014. SWISS-MODEL: modelling protein tertiary and quaternary structure using evolutionary information. Nucleic Acids Res 42:W252-258.

Darriba D, Taboada GL, Doallo R, Posada D. 2012. jModelTest 2: more models, new heuristics and parallel computing. Nat Methods 9:772.

Freitas R, Cohn MJ. 2004. Analysis of EphA4 in the lesser spotted catshark identifies a primitive gnathostome expression pattern and reveals co-option during evolution of shark-specific morphology. Dev Genes Evol 214:466-472.

Fuchs S, Nguyen HD, Phan TT, Burton MF, Nieto L, de Vries-van Leeuwen IJ, Schmidt A, Goodarzifard M, Agten SM, Rose R, et al. 2013. Proline primed helix length as a modulator of the nuclear receptor-coactivator interaction. J Am Chem Soc 135:4364-4371.

Gaboriaud C, Bissery V, Benchetrit T, Mornon JP. 1987. Hydrophobic cluster analysis: an efficient new way to compare and analyse amino acid sequences. Febs Letters 224:149-155.

Guex N, Peitsch MC, Schwede T. 2009. Automated comparative protein structure modeling with SWISS-MODEL and Swiss-PdbViewer: a historical perspective. Electrophoresis 30 Suppl 1:S162-173.

Huelsenbeck JP, Ronquist F. 2001. MRBAYES: Bayesian inference of phylogenetic trees. Bioinformatics 17:754-755.

Katsu Y, Cziko PA, Chandsawangbhuwana C, Thornton JW, Sato R, Oka K, Takei Y, Baker ME, Iguchi T. 2016. A second estrogen receptor from Japanese lamprey (Lethenteron japonicum) does not have activities for estrogen binding and transcription. Gen Comp Endocrinol 236:105-114.

Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, Buxton S, Cooper A, Markowitz S, Duran C, et al. 2012. Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics 28:1647-1649.

Maxwell EE, Froebisch NB, Heppleston AC. 2008. Variability and conservation in late chondrichthyan development: Ontogeny of the winter skate (Leucoraja ocellata). Anatomical Record-Advances in Integrative Anatomy and Evolutionary Biology 291:1079-1087.

Metivier R, Penot G, Flouriot G, Pakdel F. 2001. Synergism between ERalpha transactivation function 1 (AF-1) and AF-2 mediated by steroid receptor coactivator protein-1: requirement for the AF-1 alpha-helical core and for a direct interaction between the N- and C-terminal domains. Mol Endocrinol 15:1953-1970.

Metivier R, Petit FG, Valotaire Y, Pakdel F. 2000. Function of N-terminal transactivation domain of the estrogen receptor requires a potential alpha-helical structure and is negatively regulated by the A domain. Mol Endocrinol 14:1849-1871.

Miller MA, Pfeiffer, W., and Schwartz, T. editor. Proceedings of the Gateway Computing Environments Workshop (GCE). 2010 14 Nov. 2010: New Orleans, LA.

Neron B, Menager H, Maufrais C, Joly N, Maupetit J, Letort S, Carrere S, Tuffery P, Letondal C. 2009. Mobyle: a new full web bioinformatics framework. Bioinformatics 25:3005-3011.

O'Shaughnessy KL, Dahn RD, Cohn MJ. 2015. Molecular development of chondrichthyan claspers and the evolution of copulatory organs. Nat Commun 6:6698.

Petit FG, Valotaire Y, Pakdel F. 2000. The analysis of chimeric human/rainbow trout estrogen receptors reveals amino acid residues outside of P- and D-boxes important for the transactivation function. Nucleic Acids Res 28:2634-2642.

Pol D, Siddall ME. 2001. Biases in Maximum Likelihood and Parsimony: A Simulation Approach to a 10-Taxon Case. Cladistics 17:266-281.

Siddall ME, Whiting MF. 1999. Long-Branch Abstractions. Cladistics 15:9-24.

Warnmark A, Treuter E, Gustafsson JA, Hubbard RE, Brzozowski AM, Pike AC. 2002. Interaction of transcriptional intermediary factor 2 nuclear receptor box peptides with the coactivator binding site of estrogen receptor alpha. Journal of Biological Chemistry 277:21862-21868.