J Mol Evol (1999) 49:609–614

JOURNAL OF MOLECULAR EVOLUTION

Phylogenetic Analyses Reveal Ancient Duplication of Estrogen Receptor Isoforms

Scott T. Kelley,[1,*] Varykina G. Thackray[2,*]

[1] Department of Molecular, Cellular and Developmental Biology, University of Colorado, Boulder, CO 80309, USA
[2] Program in Molecular Biology, University of Colorado Health Sciences Center, Denver, CO 80262, USA

Received: 19 January 1999 / Accepted: 26 May 1999

Abstract. To determine the origin and evolutionary significance of a recently discovered isoform of the estrogen receptor (ERβ), we examined the phylogenetic relationship of ERβ to the well-known α isoform (ERα) and other steroid receptors. Our phylogenetic analyses traced the origin of ERβ to a single duplication event at least 450 million years ago. Since this duplication, the evolution of both ER isoforms has apparently been constrained such that 80% of the amino acid positions in the DNA binding domain (DBD) and 53% of the ligand binding domain (LBD) have remained unchanged. Using the phylogenetic tree, we determined the amount of evolutionary change that had occurred in two ER isoforms. The DBD and the LBD had lower rates of evolutionary change compared to the NH_2 terminal domain. However, even with strong selective constraints on the DBD and LBD, our phylogenetic analyses demonstrate two clearly separate phylogenetic histories for ERα and ERβ dating back several hundred million years. The ancient duplication of ER and the parallel evolution of the two ER isoforms suggest that, although ERα and ERβ share a substantial degree of sequence identity, they play unique roles in vertebrate physiology and reproduction.

Key words: Estrogen — Estrogen receptor β — Steroid receptors — Phylogeny — Selection

The estrogen receptor (ER) regulates the expression of genes involved in the growth, proliferation and differentiation of numerous bodily tissues including those in the bone, heart, brain, and reproductive organs (Ciocca and Roig 1995; Green et al. 1986). The estrogen receptor is organized into structural domains, which include a divergent NH_2 terminal domain, a DNA binding domain (DBD), and a ligand binding domain (LBD). The binding of estrogen to ER induces a conformational change that allows the hormone–receptor complex to dimerize and bind to DNA in order to modulate the transcription of specific genes. Two distinct regions of ER are involved in the activation of transcription. The NH_2 terminal domain contains a region with a transcriptional activation function (AF-1) that does not require the presence of hormone in order to be active, whereas the LBD's transcriptional activation function (AF-2) operates in a ligand-dependent manner (Kumar et al. 1986; Mangelsdorf et al. 1995; Tora et al. 1989).

The recent discovery of a distinct form of ER (Kuiper et al. 1996; Mosselman et al. 1996), termed ERβ, generated questions about the biological importance of this newly discovered gene. Alignments of amino acid sequences have uncovered a high degree of similarity between ERα and ERβ, particularly in the DBD and LBD regions, and experimental evidence has shown that ERα

*These authors contributed equally to this work

Correspondence to: S. Kelley; *e-mail:* Scott.Kelley@Colorado.edu

Table 1. Organism names and accession numbers for estrogen receptor amino acid sequences used in this study

Common name	Taxonomic name	Accession No.
	Estrogen receptor β	
Eel	*Anguilla japonica*	AB003356
Human	*Homo sapiens*	X99101
Marmoset[a]	*Callithrix jacchus*	Y09372
Mouse	*Mus musculus*	U81451
Quail[a]	*Coturnix japonica*	AF045149
Rat	*Rattus norvegicus*	U57439
	Estrogen receptor α	
Camel[a]	*Camelus dromedarius*	X98107
Chicken	*Gallus gallus*	X03805
Cichlid	*Oreochromis aureus*	P50240
Cow[a]	*Bos taurus*	P49884
Frog	*Xenopus laevis*	625330
Horse[a]	*Equus caballus*	AF007799
Human	*Homo sapiens*	72114
Lizard[a]	*Anolis carolinensis*	AF095911
Medaka	*Oryzias latipes*	P50241
Mouse	*Mus musculus*	P19785
Pig	*Sus scrofa*	Q29040
Rat	*Rattus norvegicus*	P06211
Rhesus monkey[a]	*Macaca mulatta*	P49886
Salmon	*Salmo salar*	P50242
Seabream, gilthead[a]	*Sparus aurata*	AF013104
Seabream, red	*Chrysophrys major*	AB007453
Sheep[a]	*Ovis aries*	U30299
Trout	*Oncorhynchus mykiss*	103903
Turtle[a]	*Trachemys scripta*	1703692
Zebra finch	*Taeniopygia guttata*	L79911

[a] Organisms for which only partial sequences were available.

Table 2. Organism names and accession numbers for other steroid receptor sequences used in this study

Common name	Taxonomic name	Accession No.
	Androgen receptor (AR)	
Eel	*Anguilla japonica*	BAA75464
Frog	*Xenopus laevis*	AAC97386
Chimpanzee	*Pan troglodytes*	AAC73048
Human	*Homo sapiens*	P10275
Mouse	*Mus musculus*	P19091
Rat	*Rattus norvegicus*	AAD13349
Trout	*Oncorhynchus mykiss*	BAA32784
	Glucocorticoid receptor (GR)	
Frog	*Xenophus laevis*	1730255
Human	*Homo sapiens*	4504133
Mouse	*Mus musculus*	121073
Rat	*Rattus norvegicus*	P06536
Trout	*Oncorhynchus mykiss*	1730254
	Mineralocorticoid receptor (MR)	
Human	*Homo sapiens*	4505199
Rat	*Rattus norvegicus*	P22199
	Progesterone receptor (PR)	
Chicken	*Gallus gallus*	P07812
Human	*Homo sapiens*	QRHUP
Mouse	*Mus musculus*	Q00175
Rat	*Rattus norvegicus*	I53280

and ERβ have similar DNA binding, ligand binding, and transactivation properties (Cowley et al. 1997; Kuiper et al. 1997; Pace et al. 1997; Pettersson et al. 1997). To clarify the relationship of ERβ to ERα, we performed phylogenetic analyses using the available ER sequences and other steroid receptor sequences. With these phylogenetic analyses, we asked whether or not ERβ had evolved convergently from the ancestor of a related steroid receptor to resemble ERα. Because our initial results indicated that the two ER isoforms were closely related, we then asked whether or not these isoforms arose from a single duplication event.

Steroid receptor sequences, including ER, glucocorticoid receptor (GR), progesterone receptor (PR), androgen receptor (AR), and mineralocorticoid receptor (MR), were accessed through the National Center for Biotechnology Information (NCBI). The organism names and the accession numbers for the ER sequences and the other steroid receptor sequences are presented in Tables 1 and 2. Multiple sequence alignments of all the sequences were constructed using the default settings of Pileup in the GCG software package version 9.1 for Unix (Genetics Computer Group, Madison WI, USA). The final alignment used in the phylogenetic analyses contained approximately 562 aligned amino acid positions.

All phylogenetic tree-building analyses were implemented using PAUP* (Swofford 1998). We used the maximum-parsimony (MP) criterion to determine the relationships among the various sequences. Of the 562 positions, 492 were parsimony-informative. Most parsimonious trees, for both the unrooted and the rooted analyses, were estimated using a heuristic search strategy with 100 random addition sequence replicates. Bootstrap resampling procedures were again performed using MP, and 1000 replications were completed in each analysis, with 10 heuristic random addition sequence searches performed in each replicate. Decay index values for the rooted tree were calculated using the program Autodecay 4.0 (Eriksson 1998). (Decay indices measure the number of steps separating the particular node from the next most-parsimonious tree or set of trees.)

To test whether ERβ might have evolved convergently to resemble ERα from other similar receptors, we used an unrooted phylogenetic analysis to determine the closest relatives of each of the ER isoforms. In these analyses, we included the 16 complete ER sequences (Table 1) along with 18 other steroid receptor sequences (Table 2). If the ERβ isoform arose convergently from a related steroid receptor ancestor to resemble ERα, we would not expect the ER isoforms to be each other's closest relatives. Figure 1 shows the results of the phylogenetic analyses for all the full-length steroid receptor sequences. Note that regardless of the placement of the root, the two ER isoforms are closely related, indicating that ERβ did not arise convergently from the ancestor of another steroid receptor.

Next we asked whether the ERα and ERβ sequences form two distinct monophyletic groups, which would in-

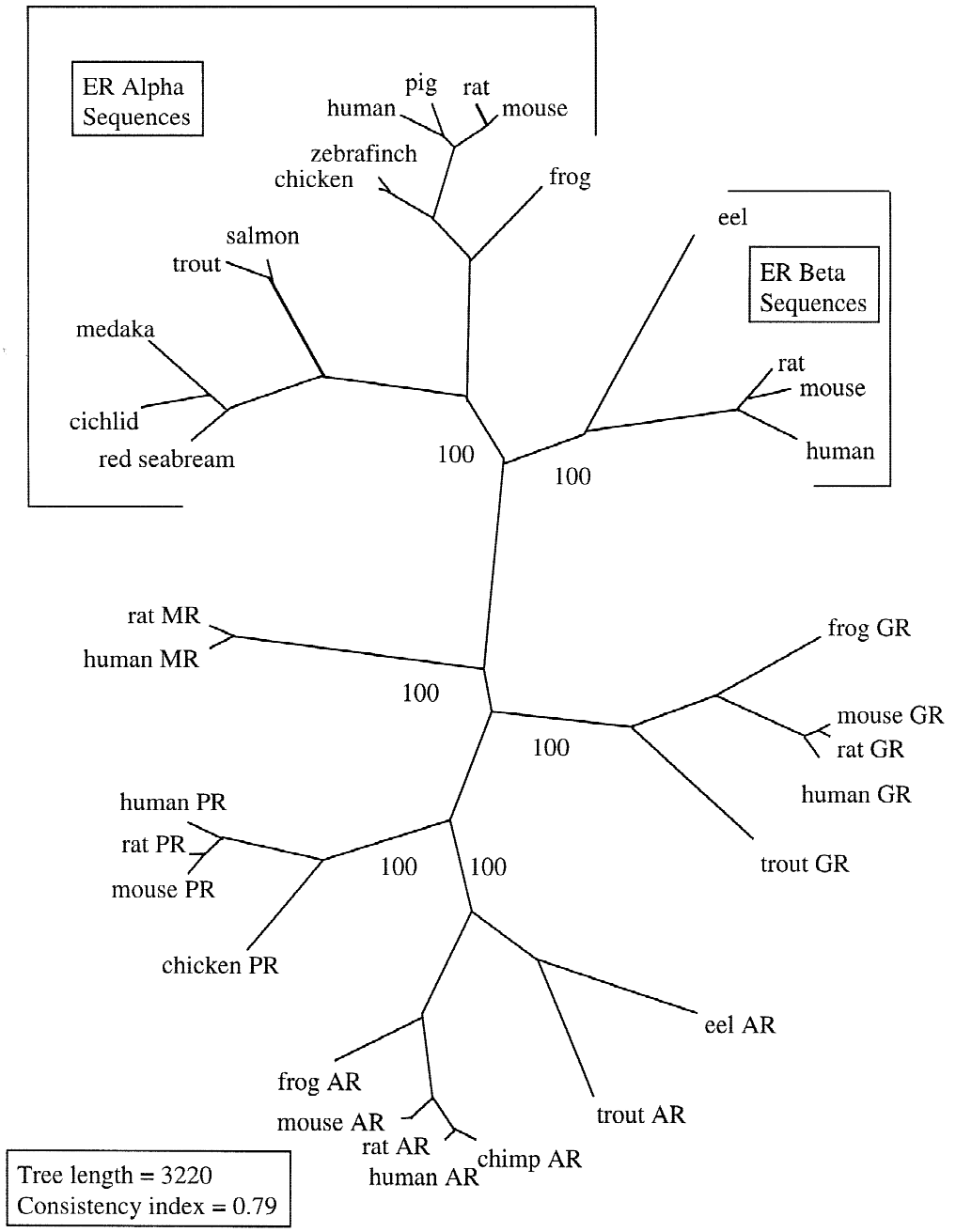

Fig. 1. Unrooted phylogenetic tree of the evolutionary relationships of the estrogen receptor sequences to other steroid receptors based on amino acid sequences. The tree is a consensus of two equally parsimonious trees. *Numbers* represent bootstrap values for the various steroid receptor groups.

dicate that they arose as a result of a single duplication event. Previous sequence alignments have shown that ERβ sequences share common elements that are distinct from ERα sequences (Pettersson et al. 1997), which in turn suggests that the ERβ sequences may belong to a separate monophyletic clade with respect to ERα. To answer these questions, we performed MP analyses using additional steroid receptor sequences as "outgroups." We considered the other steroid receptors to be appropriate outgroups based on several pieces of evidence. First, all of these receptors bind steroids. Second, all of the members of the steroid receptor family share a similar structure composed of modular domains, with the DBD placed in the middle, flanked by a NH_2 terminal domain and a LBD. Third, the DBD of all these receptors contain a double zinc-finger motif involved in receptor-specific DNA binding and receptor dimerization. GR, PR, AR, and MR recognize the same DNA consensus sequence, whereas ER recognizes a slightly different, but still very similar, consensus sequence (Mangelsdorf et al. 1995).

Phylogenetic analyses of 12 complete ERα and 5 ERβ sequences, using the steroid receptors as outgroups,

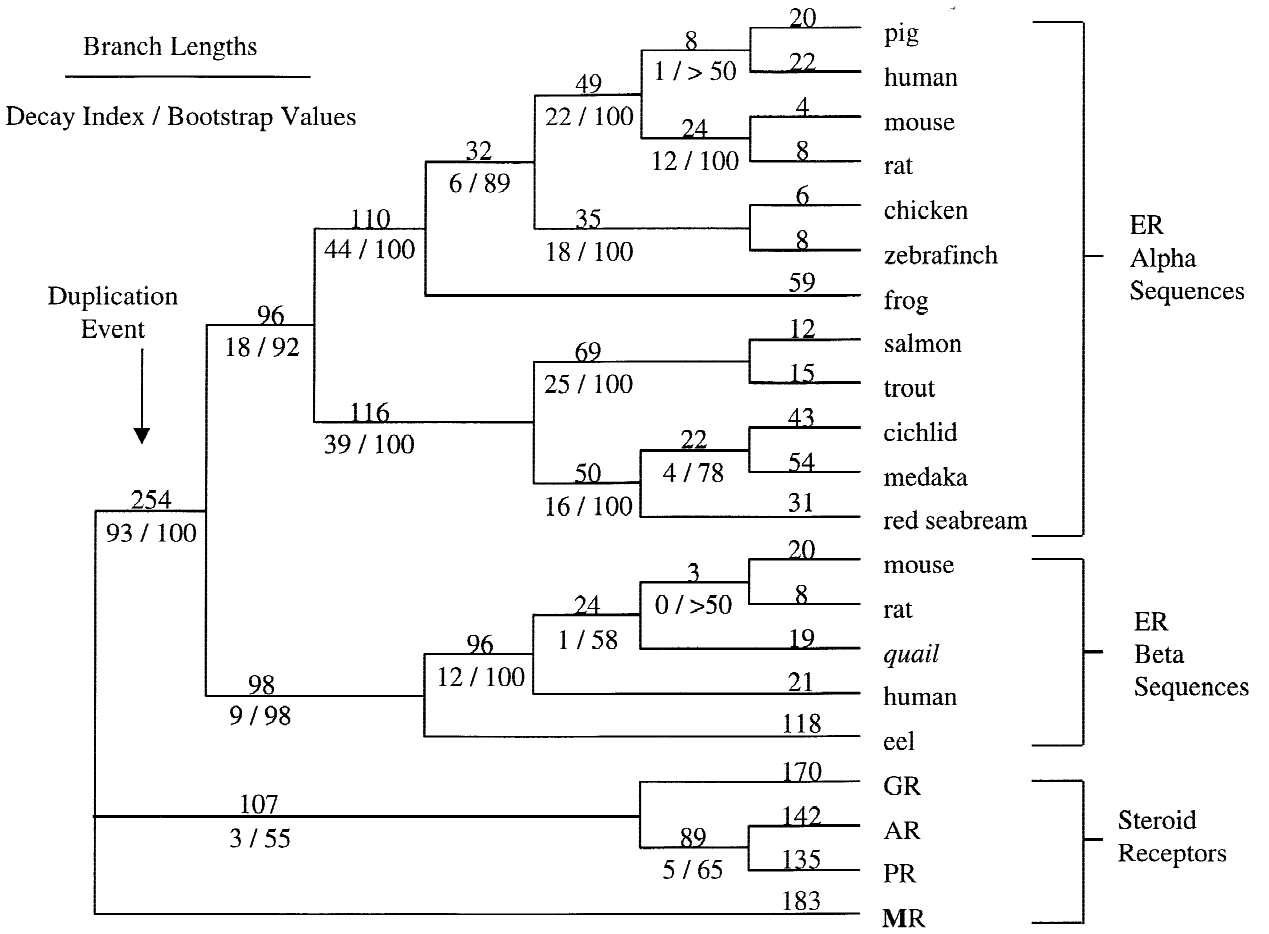

Fig. 2. Consensus of two equally parsimonious trees for the relationships among ERα and ERβ sequences using the steroid receptor sequences as outgroups. The length of the shortest trees was 2380 steps, with a consistency index of 0.86. Branch lengths are presented *above the branches*, while decay indices and bootstrap values are presented *below the branches.*

found that ERα and ERβ formed two separate and strongly supported monophyletic groups (92 and 98% bootstrap values, respectively; Fig. 2). The topology of the ERα clade was also concordant with the accepted relationships among vertebrates based on fossil evidence (Naylor and Brown 1998), supporting the notion that ERα and ERβ have evolved in parallel. (The topology of the ERβ clade did not agree with the accepted relationships among vertebrates probably because only a few, distantly related ERβ sequences were available for analysis). We also performed a number of MP analyses with the partial ER sequences listed in Table 1. Because the partial sequences did not overlap one another, we performed a separate phylogenetic analysis (with the same MP search parameters indicated above) for each of the partial sequences along with the full-length sequences. In every case, the partial ERα and ERβ sequences always grouped with their respective clades (data not shown).

The robust support for ERα and ERβ comprising two distinct, but clearly related, gene lineages suggests that the two isoforms arose from a "primordial" receptor in an ancient duplication event. Because both of the ER isoform clades contained sequences from bony fish, the duplication event must have occurred prior to the diversification of these fish, approximately 450 million years ago (Kumar and Hedges 1998; Naylor and Brown 1998). The duplication event may have occurred even earlier, though more extensive sampling of ERα and ERβ in basal vertebrate lineages will be necessary to determine just how long ago this may have happened.

During this long period of time, much of the DBD and LBD regions of ERα and ERβ remained unchanged. Using the phylogenetic character analysis program MacClade 3.0 (Maddison and Maddison 1998), we traced the evolutionary changes at every amino acid position and discovered that 80% of the amino acids in the DBD and 53% in the LBD were completely invariant between ERα and ERβ, with the exception of a few unique (autapomorphic) changes in some of the taxa. On the other hand, 97% of the amino acid positions in the NH_2 terminal domain have changed at least once over this same period of time.

Statistical analyses revealed significantly different rates of change among the three domains. For each amino acid position of ERα and ERβ, we used MacClade to trace the minimal number of mutations (steps) that must have occurred at that position given the phyloge-

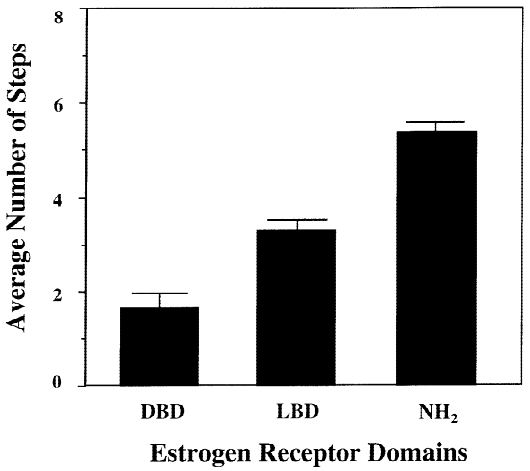

Fig. 3. Average number of steps (mutations) calculated for amino acid positions in three of the functional domains of the estrogen receptor. The number of steps at each amino acid position was calculated using parsimony over the entire tree in Fig. 2 (i.e., the analysis included both ERα and ERβ sequences). There were significant differences in the average amount of amino acid change occurring both among (nonparametric one-way ANOVA; $p < 0.0001$) and between each of the domains (Tukey's HSD test on ranked data; $p < 0.05$ for all pairwise comparisons). Standard errors are given *above each bar* of the graph.

netic tree (Fig. 2). As an example, if at amino acid position 100 of the ER multiple sequence alignment, the ERα fish sequences have a valine, while all the other sequences have an isoleucine (i.e., isoleucine is the ancestral state), we would infer that this position had changed one time on the branch leading to the fishes. Thus, the number of steps (the "treelength") for this position would be 1.

Using MacClade, we determined the tree lengths for all amino acid positions in the three main ER functional domains: the DBD, the LBD, and the NH_2 terminal domains. The average number of steps at all positions in each of the domains is presented in Fig. 3. Domains with a higher rate of overall evolutionary change would have a greater average number of steps than other domains. We found significant differences among the various ER domains in the average amount of change that had occurred (nonparametric one-way ANOVA; $p < 0.0001$), and the mean rates of change among the three domains were all significantly different (Tukey's HSD test on ranked data; $p < 0.05$). The NH_2 terminal domain experienced the greatest rate of change, followed by the LBD, and the DBD changed the least of the three (Fig. 3).

The low rates of change and the conservation of critical residues in the DBD and LBD, especially compared with the NH_2 terminal domain, imply that there has been strong selective pressure to maintain these functions in both ERα and ERβ. Alternatively, the NH_2 terminal domain may have experienced more diversifying selection compared with the DBD and LBD. However, there are both functional and structural reasons to suspect that the DBD and LBD have been under strong selective constraints. For instance, the amino acids in the "P-box" of the DBD, which are involved in the sequence-specific recognition of a hormone response element, are conserved among the ERα and ERβ isoforms (Umesono and Evans 1989). In addition, the amino acids that make direct contacts with the DNA in the crystal structure of the DBD (i.e., the amino acids that bind DNA) are completely conserved (Schwabe et al. 1993). Also, our analyses showed that the amino acids of the ligand binding cavity, identified in the crystal structure of the LBD to be involved in direct and indirect hydrophobic interactions with the ligand (Brzozowski et al. 1997), are conserved with only a few autapomorphic changes.

Although the NH_2 terminal domain appears relatively unconstrained compared with the DBD and LBD, this region of ER plays an important role in the transactivation of gene expression. Experiments have shown that, in most circumstances, transcriptional activation functions in the NH_2 terminal domain and the LBD are both required for full receptor activity (Kumar et al. 1987; Tzukerman et al. 1994). In addition, the NH_2 terminal region of ER contains serine residues which have been implicated in cross-talk with various cell signaling pathways (Katzenellenbogen 1996; Tremblay et al. 1997; Weigel 1996).

Indeed, the divergence in the NH_2 terminal domain may well be the source of the apparent functional differences between ERα and ERβ. The serine phosphorylation sites in the NH_2 terminal domain of ERα and ERβ are not conserved, suggesting that ERα and ERβ may be regulated differently by cell signaling pathways. Additionally, there are two instances where differences in function between ERα and ERβ may involve the NH_2 terminal domain. First, Webb et al. (1995) showed an apparent interaction between the NH_2 terminal domain of ERα and the transcription factor Jun. ERα can activate the transcription of a reporter gene containing an AP-1 element in response to estrogen, whereas ERβ cannot (Paech et al. 1997). Second, the partial agonism of the antiestrogen tamoxifen was mapped to the NH_2 terminal domain of ERα (McInerney and Katzenellenbogen 1996). Tamoxifen only displays antagonism to ERβ on a consensus estrogen response element, as opposed to partial agonism and antagonism to ERα (Barkhem et al. 1998). These experimental observations, and the significant sequence divergence observed between ERα and ERβ in the NH_2 terminal domain, suggest that any distinct biological functions found between ERα and ERβ may be traceable to differences in this domain.

The fact that ERβ comprises a separate genetic lineage dating back at least 450 million years argues that this gene performs distinct biological functions that have been maintained by natural selection for this long period of time. Our analyses also indicate that the ERβ gene is quite widespread among chordates. Since ERα and ERβ appear to have originated from a single duplication event predating the diversification of the bony fish, we predict

that copies of both isoforms will be discovered in all higher vertebrates. Indeed, Le Roux et al. (1993) detected the presence of multiple ER isoforms in trout, prior to the discovery of ERβ, which further supports the notion that both isoforms are ubiquitous in vertebrates. In summary, the phylogenetic approach described in this paper enabled us to date the minimum time of divergence of the two ER isoforms, describe the selective constraints on three domains of these molecules, and predict what organisms possess these genes.

Acknowledgments. We thank A. de Queiroz, J. Mitton, S. Martin, S. Nordeen, and two anonymous reviewers for helpful discussions and comments on the manuscript.

References

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 54:105–112

Brzozowski AM, Pike AC, Dauter Z, et al. (1997) Molecular basis of agonism and antagonism in the oestrogen receptor. Nature 389: 753–758

Ciocca DR, Roig LM (1995) Estrogen receptors in human nontarget tissues: biological and clinical implications. Endocr Rev 16:35–62

Cowley SM, Hoare S, Mosselman S, Parker MG (1997) Estrogen receptors alpha and beta form heterodimers on DNA. J Biol Chem 272:19858–19862

Eriksson T (1998) AutoDecay version 4.0 (program distributed by the author). Department of Botany, Stockholm University, Stockholm

Green S, Walter P, Kumar V, et al. (1986) Human oestrogen receptor cDNA: Sequence, expression and homology to v-erb-A. Nature 320:134–139

Katzenellenbogen BS (1996) Estrogen receptors: Bioactivities and interactions with cell signaling pathways. Biol Reprod 54:287–293

Kuiper GG, Enmark E, Pelto-Huikko M, Nilsson S, Gustafsson JA (1996) Cloning of a novel receptor expressed in rat prostate and ovary. Proc Natl Acad Sci USA 93:5925–5930

Kuiper GG, Carlsson B, Grandien K, et al. (1997) Comparison of the ligand binding specificity and transcript tissue distribution of estrogen receptors alpha and beta. Endocrinology 138:863–870

Kumar S, Hedges SB (1998) A molecular timescale for vertebrate evolution. Nature 392:917–920

Kumar V, Green S, Staub A, Chambon P (1986) Localization of the oestradiol-binding and putative DNA-binding domains of the human oestrogen receptor. EMBO J 5:2231–2236

Kumar V, Green S, Stack G, Berry M, Jin JR, Chambon P (1987) Functional domains of the human estrogen receptor. Cell 51:941–951

Le Roux MG, Theze N, Wolff J, Le Pennec JP (1993) Organization of a rainbow trout estrogen receptor gene. Biochim Biophys Acta 1172:226–230

Maddison WP, Maddison DR (1993) MacClade: Analysis of phylogeny and character evolution. Sinauer Associates, Sunderland, MA

Mangelsdorf DJ, Thummel C, Beato M, et al. (1995) The nuclear receptor superfamily: the second decade. Cell 83:835–839

McInerney EM, Katzenellenbogen BS (1996) Different regions in activation function-1 of the human estrogen receptor required for antiestrogen- and estradiol-dependent transcription activation. J Biol Chem 271:24172–24178

Mosselman S, Polman J, Dijkema R (1996) ER beta: Identification and characterization of a novel human estrogen receptor. FEBS Lett 392:49–53

Naylor GJP, Brown WM (1998) Amphioxus mitochondrial DNA, chordate phylogeny and the limits of inference based on comparisons of sequences. Syst Biol 47:61–76

Pace P, Taylor J, Suntharalingam S, Coombes RC, Ali S (1997) Human estrogen receptor beta binds DNA in a manner similar to and dimerizes with estrogen receptor alpha. J Biol Chem 272:25832–25838

Paech K, Webb P, Kuiper GG, et al. (1997) Differential ligand activation of estrogen receptors ERalpha and ERbeta at AP1 sites. Science 277:1508–1510

Pettersson K, Grandien K, Kuiper GG, Gustafsson JA (1997) Mouse estrogen receptor beta forms estrogen response element-binding heterodimers with estrogen receptor alpha. Mol Endocrinol 11: 1486–1496

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 75:567–578

Swofford D (1998) PAUP*: Phylogenetic analysis using parsimony (*and other methods). Sinauer Associates, Sunderland, MA

Tora L, White J, Brou C, et al. (1989) The human estrogen receptor has two independent non-acidic transcriptional functions. Cell 59:477–487

Tremblay GB, Tremblay A, Copeland NG, et al. (1997) Cloning, chromosomal localization, and functional analysis of the murine estrogen receptor beta. Mol Endocrinol 11:353–365

Tzukerman MT, Esty A, Santiso-Mere D, et al. (1994) Human estrogen receptor transactivational capacity is determined by both cellular and promoter context and mediated by two functionally distinct intramolecular regions. Mol Endocrinol 8:21–30

Umesono K, Evans RM (1989) Determinants of target gene specificity for steroid/thyroid hormone receptors. Cell 57:1139–1146

Webb P, Lopez GN, Uht RM, Kushner PJ (1995) Tamoxifen activation of the estrogen receptor/AP-1 pathway: Potential origin for the cell-specific estrogen-like effects of antiestrogens. Mol Endocrinol 9:443–456

Weigel NL (1996) Steroid hormone receptors and their regulation by phosphorylation. Biochem J 319:657–667