

# Evolutionary Genomics of Nuclear Receptors: From Twenty-Five Ancestral Genes to Derived Endocrine Systems

Stéphanie Bertrand, Frédéric G. Brunet, Hector Escriva, Gilles Parmentier, Vincent Laudet, and Marc Robinson-Rechavi<sup>1</sup>

\*Laboratoire de Biologie Moléculaire de la Cellule, Ecole Normale Supérieure de Lyon, Lyon, France

Bilateria animals are notably characterized by complex endocrine systems. The receptors for many steroids, retinoids, and other hormones belong to the superfamily of nuclear receptors, which are transcription factors regulating many aspects of development and homeostasis. Despite a diversity of regulatory mechanisms and physiological roles, nuclear receptors share a common protein organization. To obtain the broad picture of bilaterian nuclear hormone receptor evolution, we have characterized the complete set of nuclear receptor genes from nine animal genome sequences and analyzed it in a phylogenetic framework. In addition, expressed sequence tags from key lineages with no available genome sequence were also searched. This allows us to date the evolutionary events that led from an ancestral nuclear receptor gene, in an early metazoan, to present day diversity. We show that there were ~25 nuclear receptor genes in Urbilateria, the ancestor of bilaterians, at which point the fundamental diversity of the subfamily was already established. Surprisingly, differential gene loss played an important role in the evolution of different nuclear receptor sets in bilaterian lineages. The nuclear receptor distribution was also shaped by periods of gene duplication, essentially in vertebrates, as well as a lineage-specific duplication burst in nematodes. Our results imply that the genes for major receptors such as steroid receptors or thyroid hormone receptors were present in Urbilateria.

## Introduction

One of the striking features of bilaterian animal evolution is the development of complex endocrine systems, which allow the organism to coordinate its reaction to the environment, to regulate its development, and to maintain homeostasis. Among the players in these complex systems, the superfamily of nuclear receptors (NRs) is specific to animals and performs an abundance of functions, from embryonic development to metamorphosis and from homeostasis of various physiological functions to the control of metabolism (for a review, see Laudet and Gronemeyer 2002). Nuclear receptors are ligand-activated transcription factors. Many members of the superfamily thus bind major hormones, such as steroids, thyroid hormones, or retinoids. These occupy a special position in gene regulation, since they provide a direct link between the ligand, which they bind, and the target gene, whose expression they regulate. There are also many nuclear receptors that are not known to bind any ligand and are thus called “orphan nuclear receptors” (Gustafsson 1999; Kliewer, Lehmann, and Willson 1999). Some nuclear receptors which were originally discovered as orphans have since been shown to bind small hydrophobic molecules, such as fatty acids, but other nuclear receptors appear to be true orphans (Ruse, Privalsky, and Sladek 2002; Wang et al. 2003; H. Gronemeyer, J. A. Gustafsson, and V. Laudet, in preparation).

Nuclear receptors share a common organization in structural domains, which notably includes a very conserved DNA binding domain (DBD) and a moderately conserved ligand-binding domain (LBD). These domains allow relatively easy identification of nuclear receptors in

genomic sequences (Robinson-Rechavi et al. 2001a) and also robust phylogenetic reconstruction at the scale of the superfamily (Laudet et al. 1992; Escriva, Laudet, and Robinson-Rechavi 2003). This phylogenetic reconstruction shows that receptors for similar ligands do not group in the tree, but are interspersed with receptors for totally different ligands, while orphans are widely distributed in the tree. This led to the hypothesis that the superfamily evolved from an orphan receptor that acquired several times independently the capacity to bind ligands (Escriva et al. 1997; Laudet 1997; Escriva, Delaunay, and Laudet 2000). Thus, the capacity to bind, for example, retinoic acids, would have been acquired several times by mutation of an orphan nuclear receptor in some ancestor of vertebrates, allowing the establishment of nuclear retinoic acid signaling pathways (RARs, RXRs). The recent identification of fatty acids that play a structural role inside the 3-D structure of some orphan receptors suggests that bona fide nuclear receptor ligands may have evolved from such structural ligands (review in Auwerx, Drouin, and Laudet 2003). The isolation of an ortholog of the vertebrate estrogen receptor (ER) from a mollusk has shown the interest of phylogenetic analysis of nuclear receptors from distant organisms in defining ancestral endocrinology signaling pathways (Thornton, Need, and Crews 2003).

The evolution of nuclear receptors (or other genes) can only be understood on the background of a reference phylogeny of the species analyzed. One could hope that dealing with two nematode worms, two insects, and five chordates, we would not encounter any problems defining this reference. But the phylogeny of animals has been the object of molecular studies that have resulted in contradictory results concerning the relative positions of these three lineages. Studies of ribosomal RNA suggest that nematodes and insects form a clade, “Ecdysozoa” (Aguinaldo et al. 1997; Mallatt and Winchell 2002), which is also consistent with information from developmental genes (Adoutte et al. 2000). But recent studies, using more genes and less species, have given renewed support to the

<sup>1</sup> Present address: Joint Center for Structural Genomics, UCSD, La Jolla, California.

Key words: bilaterian, development, endocrinology, gene duplication, gene loss, phylogeny.

E-mail: vincent.laudet@ens-lyon.fr.

*Mol. Biol. Evol.* 21(10):1923–1937. 2004

doi:10.1093/molbev/msh200

Advance Access publication June 30, 2004

traditional view of a clade grouping insects and chordates, “Coelomata” (Blair et al. 2002; Dopazo, Santoyo, and Dopazo 2004; Wolf, Rogozin, and Koonin 2004). In this work, we have chosen to consider all results in light of both possibilities, the “ectoderm hypothesis” and the “coelomate hypothesis.” Thus, we hope our results to be robust to the outcome of this interesting debate.

In this work we take advantage of the wealth of data provided by complete genome projects, spanning several degrees of divergence among animals, to redefine relationships inside the superfamily and to characterize nuclear receptor evolution since the divergence of major bilaterian lineages. We are also in the process of characterizing experimentally all nuclear receptors from the zebrafish *Danio rerio* (unpublished data), and we took these sequences into account in our interpretations. Using a phylogenetic approach, and thanks to the completeness of the information available, we can for the first time define overall nuclear receptor evolution in terms of gene gain and loss and show which mechanisms were instrumental in establishing modern endocrine systems.

## Materials and Methods

A standard set of nuclear receptor sequences was established for querying genome sequences by retrieving one complete protein sequence per group (Nuclear Receptors Nomenclature Committee 1999) from Nurebase (Ruau et al. 2004). These sequences were compared to predicted peptides from each of the queried genomes by BlastP (Altschul et al. 1990). The data used for each queried genome is presented in table S3 in the online Supplementary Material. The same methodology was used to identify and classify potential nuclear receptor genes as in Robinson-Rechavi et al. (2001a), including checks against genomic DNA sequences to detect annotation errors and pseudogenes.

All new nuclear receptor protein sequences were aligned by eye using Seaview (Galtier, Gouy, and Gautier 1996) on the reviewed alignment of Nurebase. The alignments used for phylogenetic analysis included all available sequences and were not limited to the species discussed in the text. Sequences with less than 200 amino acids including the DBD were excluded, and only complete sites (no gap, no X) were used for phylogenetic reconstruction. Phylogenetic trees were built using (1) Neighbor-Joining (Saitou and Nei 1987) with distances corrected for rate heterogeneity between sites, with a gamma law of parameter alpha estimated in Tree-Puzzle (Schmidt et al. 2002) with eight categories, and using (2) PHYML (Guindon and Gascuel 2003), a fast and accurate maximum likelihood heuristic, under the JTT substitution model (Jones, Taylor, and Thornton 1992), with a gamma distribution of rates between sites (eight categories, parameter alpha estimated by PHYML). Trees were built for the whole superfamily, for each subfamily, and for each group. In case of conflict, trees built using groups were considered more reliable because they use more complete sites and less divergent sequences. All phylogenetic relations that are discussed were tested by a comparison of likelihoods between alternative topologies, using the SH test (Shimodaira

and Hasegawa 1999) as implemented in Tree-Puzzle, correcting for rate heterogeneity between sites with a gamma law.

Use of taxonomic and common names: We use the word “fish” for Actinopterygii, or ray-finned fishes; this usage does not include cartilaginous fishes (sharks and rays) or flesh-finned fishes (lungfishes and coelacanths). We use the word “pufferfish” for Tetraodontiformes, which include *Takifugu rubripes* and *Tetraodon nigroviridis*. We use the word “nematode worm” for the two *Caenorhabditis* species studied. We use the word “trematode” for *Schistosoma mansoni*.

## Results

### How Does the Number of Nuclear Receptor Genes Vary Among Bilaterian Lineages?

The original reports of several bilaterian genome projects include accounts on the nuclear receptor genes found. While the result originally reported for the fruit fly was correct (21 NR genes; Adams et al. 2000), the results reported for the human genome were not: original reports suggested 60 NR genes, whereas further examination only found 48 (Maglich et al. 2001; Robinson-Rechavi et al. 2001a). As for the nematode *C. elegans*, it has a very divergent nuclear receptor complement, with more than 250 NR genes, which have been extensively studied elsewhere (Sluder et al. 1999; Maglich et al. 2001; Sluder and Maina 2001; unpublished data). Here we repeated the search for each available genome sequence, whatever the published results. This yields consistent results with those reported for the sea squirt (17 NR genes; Dehal et al. 2002; Yagi et al. 2003) and *C. briggsae* (>250 NR genes; Stein et al. 2003), but we found two more than reported for the fugu (70 vs. 68 NR genes; Maglich et al. 2003). A recent report analyzed nuclear receptor genes from the rat and mouse genomes (Zhang et al. 2004). We confirm the nuclear receptor genes found in that study (47 in rat, 49 in mouse). Because of the incomplete nature of the rat genome (Rat Genome Sequencing Project Consortium 2004; Zhang et al. 2004), and because it does not add any new insights into the evolution of superfamily (i.e., no new genes discovered), we limit our discussion to the more complete mouse genome (Waterston et al. 2002). We present the first analysis of nuclear receptor genes for the mosquito (21 NR genes; Holt et al. 2002) and tetraodon (66 NR genes; unpublished genome). All genes we found are presented in table S1 (see online Supplementary Material), and an overall view of homology relations is presented in figure 1. We have also analyzed EST data from several key lineages with no available genome sequence, which allowed us to clarify some unresolved phylogenies. Genes found by Expressed Sequence Tag (EST) analysis are presented in table S2 (online Supplementary Material). All protein alignments are available as online Supplementary Material and will be included in future versions of Nurebase (Ruau et al. 2004).

A striking pattern in the results obtained by complete genome analysis is that species that have diverged since the Mesozoic (i.e., in the last 248 Myr: the diversification of closely related animals species) have conserved very

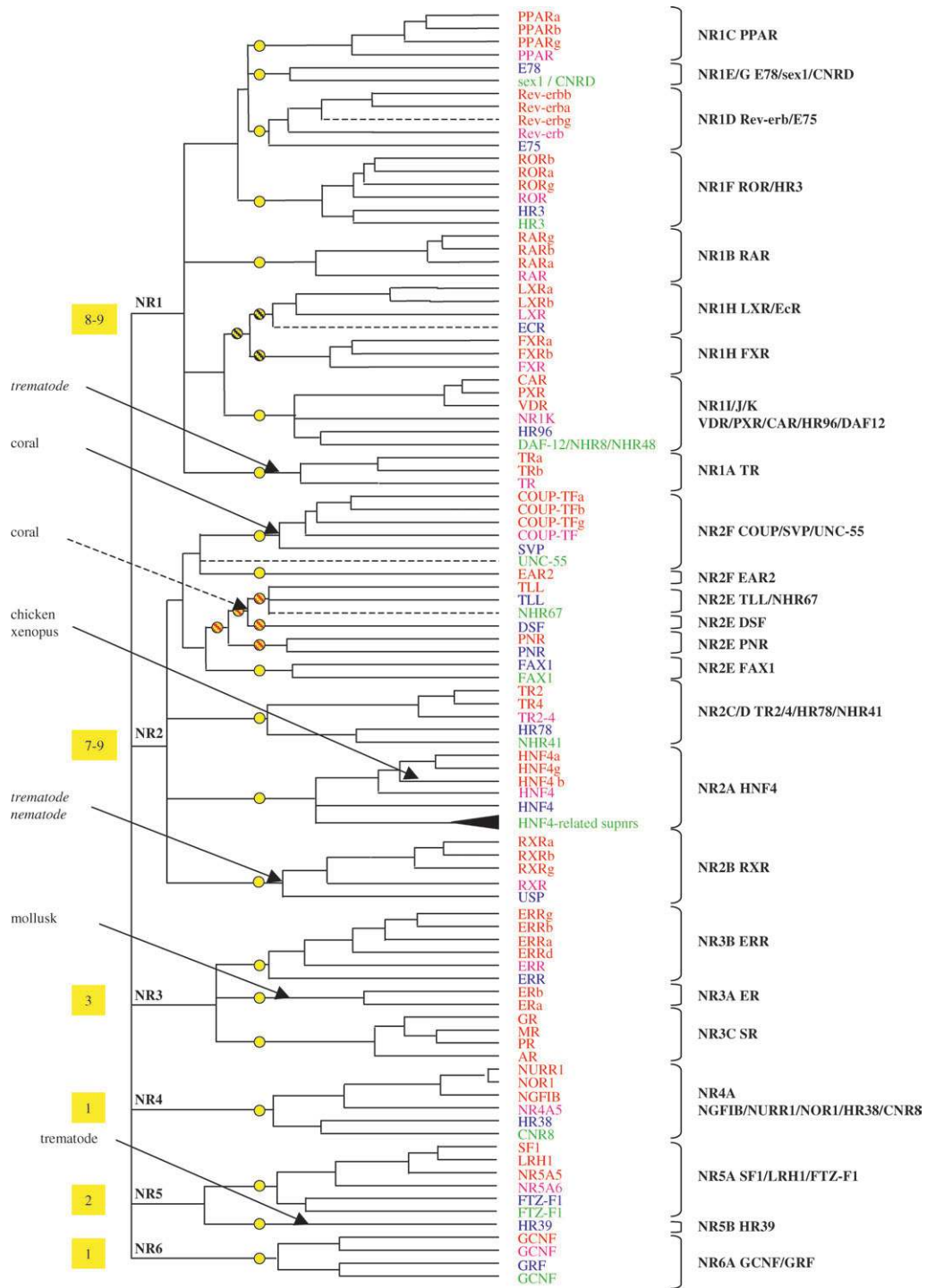


FIG. 1.—Summarized phylogenetic tree of the nuclear receptor superfamily. This represents a summary of information from gene trees of each group and subfamily of nuclear receptors. Branches that are not significantly supported were collapsed into polytomies. Although only taxonomic groups with genome sequences available are represented, all available genes were used for each gene tree. Nuclear receptor gene names follow table S1 (online Supplementary Material). Genes are color-coded according to taxonomic group: red for vertebrates, purple for sea squirt, blue for dipteran insects, and green for nematodes. For readability, species phylogeny inside each of these groups is not indicated, nor are evolutionary events inside each of these groups: secondary gene loss, fish-specific gene duplications, etc. Branch lengths are arbitrary. On the right, gene groups with official nomenclature (Nuclear Receptors Nomenclature Committee 1999) are listed. Yellow background circles indicate genes that are inferred to have existed in the common ancestor of insects, nematodes, and chordates; yellow background numbers at the root of each subfamily indicate the number of such ancestral genes inferred to have existed in this subfamily. Hatched background circles represent cases where the ancestral gene state cannot be inferred with confidence; black hatching represents alternative ancestral genes for the NR1H group; red hatching represents alternative ancestral genes for the NR2E group. Arrows represent key contributions of taxa for which we do not have complete genomes; taxon names are in italics if only EST sequences were used. The broken lines leading to EcR, UNC-55, Rev-erb, NHR67, and to the coral TLL/DSF indicate the lack of significant resolution of phylogenetic methods to position these genes.

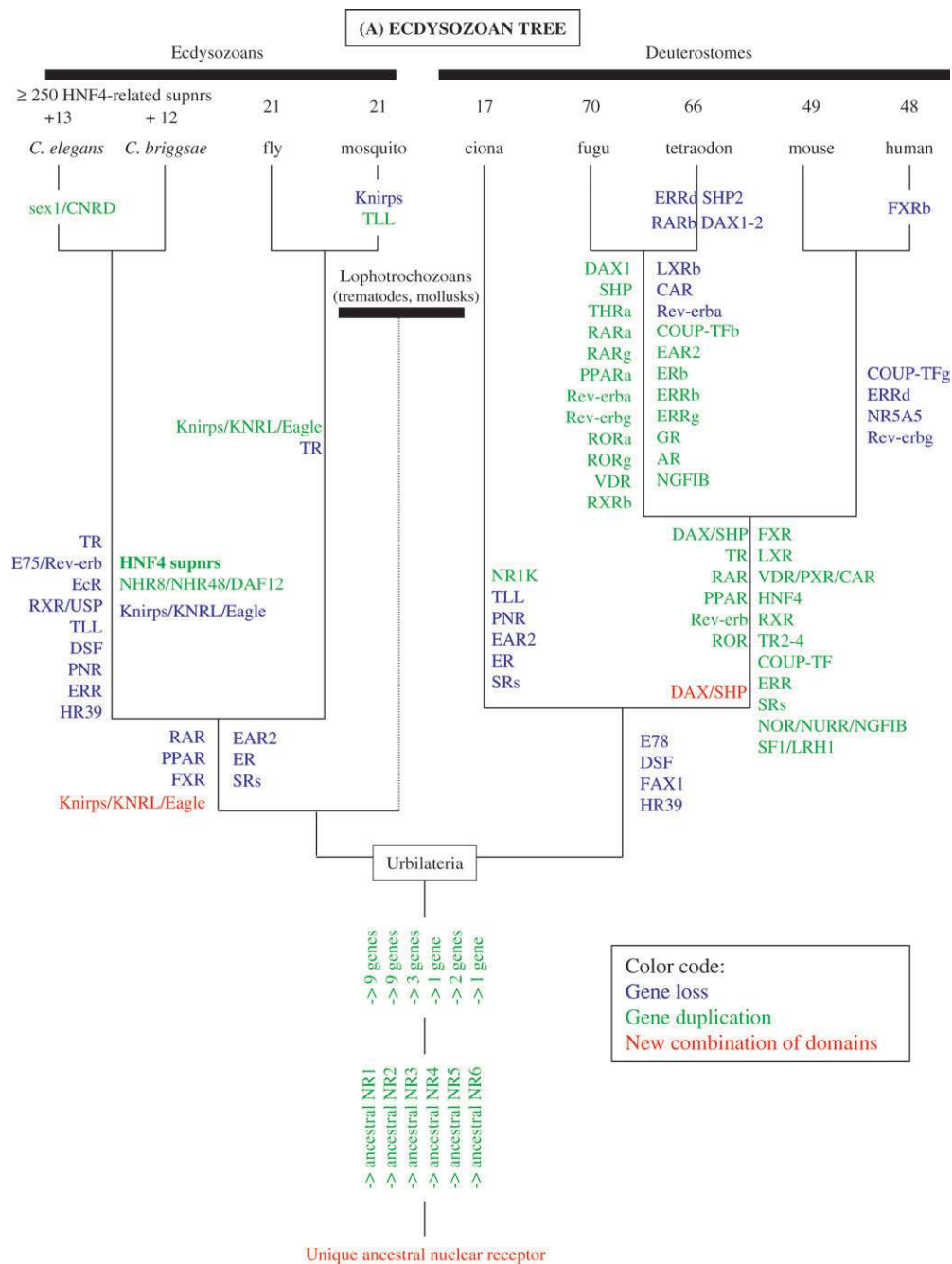


FIG. 2.—Summary of nuclear receptor gene gain and loss in animal evolution. Branch lengths are arbitrary. Urbilateria is the common ancestor of all bilaterians. Naming as in figure 1. Events are mapped to a branch of the tree, but order of events on each branch is not known. The broken lines leading to trematodes and mollusks indicate the key phylogenetic positions of these groups (placed according to fig. 1 of Adoutte et al. [2000]), although we do not have any complete genome sequence. Above each species is the number of nuclear receptor genes found in its genome. “Supnrs” are the supplementary, divergent, nuclear receptors found in nematode genomes. “SRs” stands for “steroid receptors” and cover the four close paralogs AR, MR, GR, and PR. In two cases where the phylogeny was ambiguous, the favored phylogeny was used even though it was not significant (see hatched background circles in fig. 1). (A) The “ecdysozoan” species tree is assumed to map gene duplication and loss (Aguinaldo et al. 1997). (B) The “coelomate” species tree is assumed to map gene duplication and loss (Wolf, Rogozin, and Koonin 2004).

similar nuclear receptor complements (fig. 2): the two dipterans (250 MYA) both have 21 NR genes, the two mammals (80 MYA) have 48 and 49, the two pufferfishes (25 MYA) have 70 and 66, and the two nematodes (40 MYA) have 13 and 12 conserved NRs, plus more than 250 supplementary divergent NRs each (which form one

clade, unpublished data). Of note, a recent study reports 15 conserved NRs in *C. elegans* (Gissendanner et al. 2004) by classifying the two least divergent supplementary nuclear receptors as conserved orthologs of HNF4; because these two genes belong to the monophyletic clade of supplementary NRs, we consider them separately. The similarities are not limited to the number of genes; we indeed find orthologous nuclear receptors in these pairs of species. The



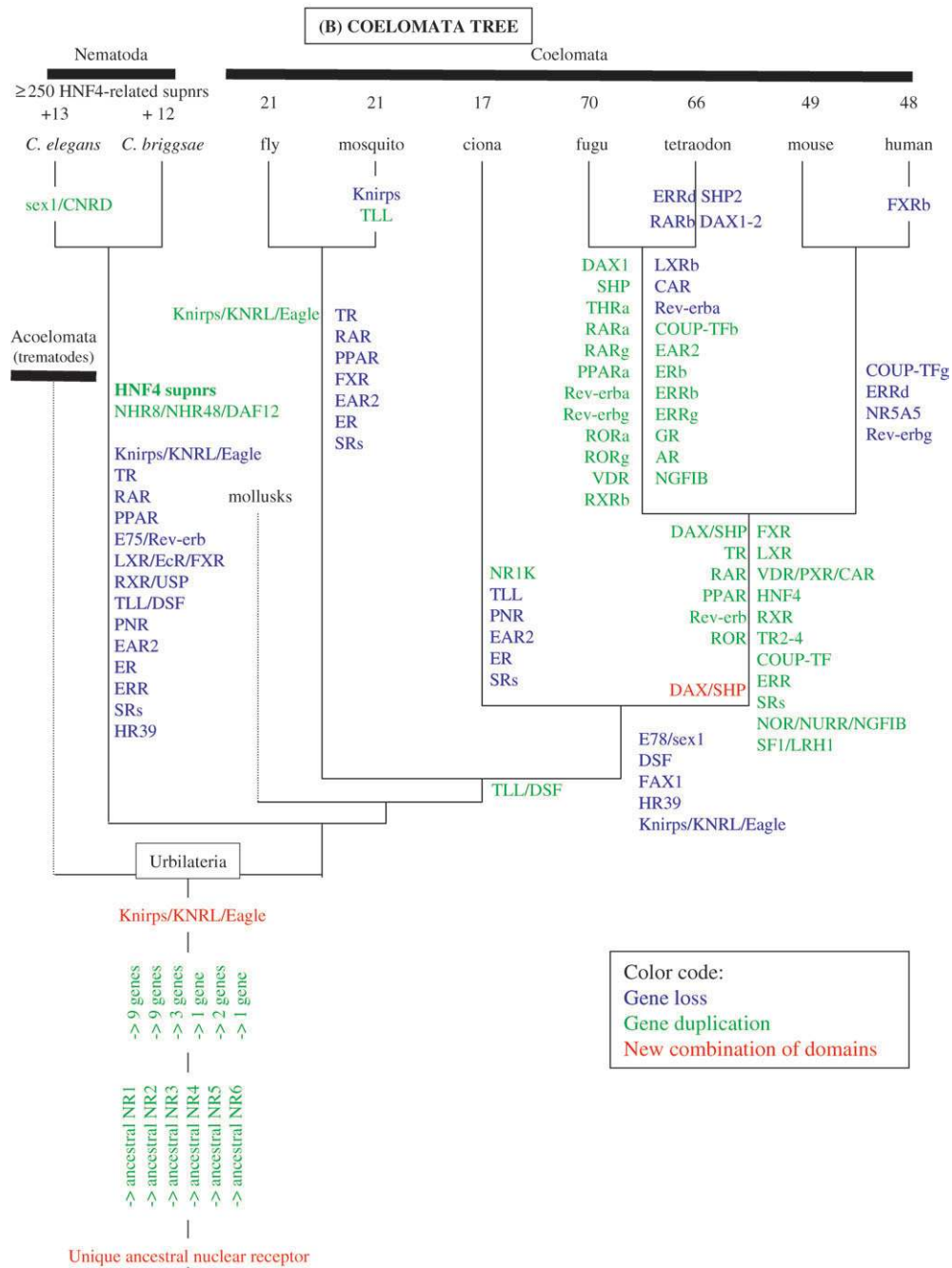


FIG. 2.—Continued.

fly and the mosquito are separated by only one gene loss (of Knirps in the mosquito) and one gain (duplication of TLL-related in the mosquito), human and mouse (and rat) are separated by one gene loss (FXRb), the two pufferfishes are separated by four gene losses (SHP2, DAX1-2, RARb, and ERRd, all lost in tetraodon), and the two nematodes are separated by one duplication (plus the ongoing diversification of HNF4-derived NRs). These figures should be compared to a total of more than 75 estimated events of duplication or gene loss since the divergence of bilaterian phyla (fig. 2). Species that diverged in the Vendian (650 to 543 MYA: first evidence

of major animal phyla) or the Paleozoic (543 to 248 MYA: diversification inside major animal phyla), on the other hand, seem to have accumulated quite divergent complements: nematodes have five to ten times more nuclear receptors than other triploblastic animals (600 MYA), pufferfishes have 38% more than mammals (400 MYA), and the sea squirt has considerably less than any vertebrate (555 MYA). Although the divergence dates are debated, most gene duplication or loss in nuclear receptors clearly occurred relatively early in animal evolution, more than 400 MYA. An interesting exception to the divergence between animal phyla is the GCNF group (NR6A), which

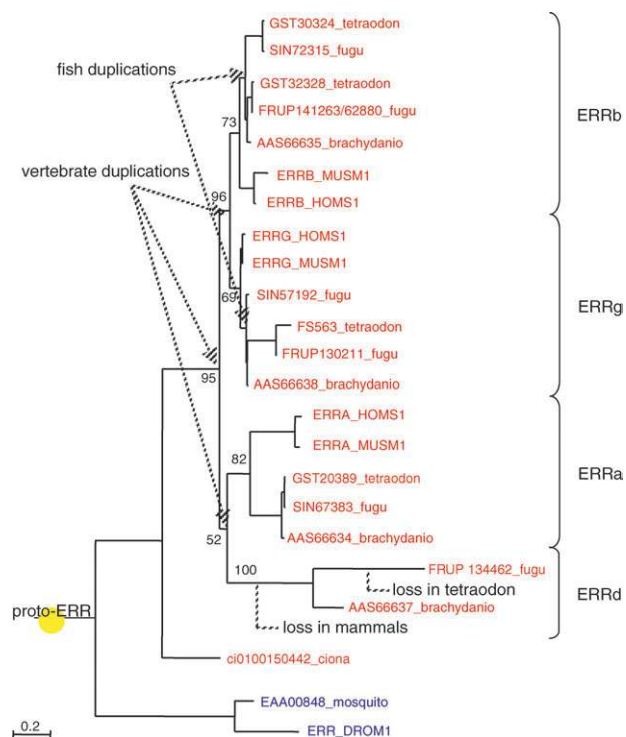


FIG. 3.—A simple case: estimation of the number of ancestral genes for estrogen related receptors. Maximum-likelihood tree of sequences from Nurebase and predicted from genomes; zebrafish sequences which were not yet in Nurebase were added from GenBank. Two ERRg sequences with partial DBDs were excluded from the alignment (GST32848 from Tetraodon and AAS66636 from zebrafish). Bootstrap support, in percentage of 2,000 replicates in a Neighbor-Joining analysis, is indicated on the branches that define the vertebrate duplications. The alternative placing of FRUP134462 + AAS66637 as fish-specific duplicates of ERRa is significantly less likely (SH test:  $P = 0.042$ ). Branch length is proportional to estimated evolutionary change; the measure bar represents 0.2 substitutions/site. Naming and color codes as in figure 1. Arrows indicate gene duplications or gene losses, some of which are not visible on the simplified tree of figure 1. Thick-hatched branches indicate secondary gene loss. The yellow background circle indicates the gene inferred to have existed in the common ancestor of bilaterians.

is the only group of nuclear receptor genes to have remained apparently stable since the ancestor of bilaterians, with one gene per genome in all animals studied.

#### Number of Nuclear Receptor Genes in Urbilateria

More widely, we can infer the minimum NR complement of the common ancestor of chordates, nematodes, and insects (i.e., the common ancestor of bilaterians, Urbilateria; fig. 2) from the phylogenetic relationships of their NRs. The principle of this inference is that for each group of NRs we can locate the speciation event that led to the different bilaterian lineages by the divergence between chordate, insect, and nematode orthologs. Then, all genes resulting from duplications that occurred before this speciation were probably present in Urbilateria. This reasoning remains correct even after secondary loss in some lineages. For example, although there are only three ERR paralogs in the human genome, we can infer that there were four in the ancestor of vertebrates, due to the

phylogenetic position of fish ERRd (fig. 3; Bardet et al. 2004); the fourth ERR was secondarily lost in mammals. ERRs are orphan receptors closely related to estrogen receptors (ERs). Moreover, these ERs include a mollusk ortholog, which implies that the proto-ERR existed in the ancestor of bilaterians. So the absence of ERR in nematodes must again correspond to secondary loss, and we can infer that there were three nuclear receptor genes of subfamily 3 in the urbilaterian (fig. 1; see also Thornton, Need, and Crews 2003).

The example represented in figure 4 is more complex. The placement of nematode DAF-12 implies that the split between the groups of VDR/PXR, FXR, and LXR (hatched black arrow) occurred before the speciation of bilaterian lineages (orange arrow). However, further inference is dependent on our understanding of the species phylogeny. If the ecdysozoan hypothesis is correct (Aguinaldo et al. 1997), then the phylogenetic position of a *Drosophila* gene represents protostomes as well as a nematode gene does. Thus, any subfamily that contains either of these species can be dated back to the ancestor of bilaterians. In that case, FXRs, EcR and LXRs derive from two paralogous genes in the urbilaterian (fig. 4B). However, if the coelomate hypothesis is correct (Wolf, Rogozin, and Koonin 2004), then an ambiguity arises: it is possible that there were a proto-FXR and a proto-LXR/EcR in Urbilateria, both being lost in nematodes (fig. 4C). It is also possible that there was only one ancestral gene in Urbilateria, that it was this proto-FXR/EcR/LXR which was lost in nematodes and that it was duplicated in the ancestor of coelomates before the speciation of chordates and arthropods (fig. 4D). This is the only subfamily for which we have no further evidence to choose between the different solutions. However, the relative rarity of gene duplication in the putative ancestor of coelomates (fig. 2B) and the frequency of gene loss in the nematode lineage favor the hypothesis of separate proto-LXR/EcR and proto-FXR paralogs in Urbilateria (fig. 4C).

In several other subfamilies, where similar problems occur, the phylogenetic position of nuclear receptor genes from coral, other nematodes, or trematode, notably from EST data (online Supplementary table S2), allows us to position the common ancestor of animals (arrows in fig. 1). The most notable result from trematode EST data is the first evidence for a thyroid hormone receptor ortholog in an invertebrate, proving that there was a proto-TR in the urbilaterian. The phylogenetic position of these EST sequences inside the TR clade is strongly supported both when nuclear receptors from all groups of the superfamily are included in the analysis (65 complete aa sites; SH test:  $P = 0.0030$ ) and when only closely related sequences are used (90 complete aa sites; fig. 5).

Applying this method to the entire superfamily, it appears that there were at least 22 to 25 nuclear receptor genes in the common ancestor of nematodes, insects, and chordates (fig. 2); given the data discussed above, the most probable number seems to be 25 ancestral nuclear receptor genes (yellow background points in fig. 1; table 1). Of note, the lack of phylogenetic resolution in some groups does not appear to be due simply to higher evolutionary rates: NR2E sequences do evolve significantly faster than

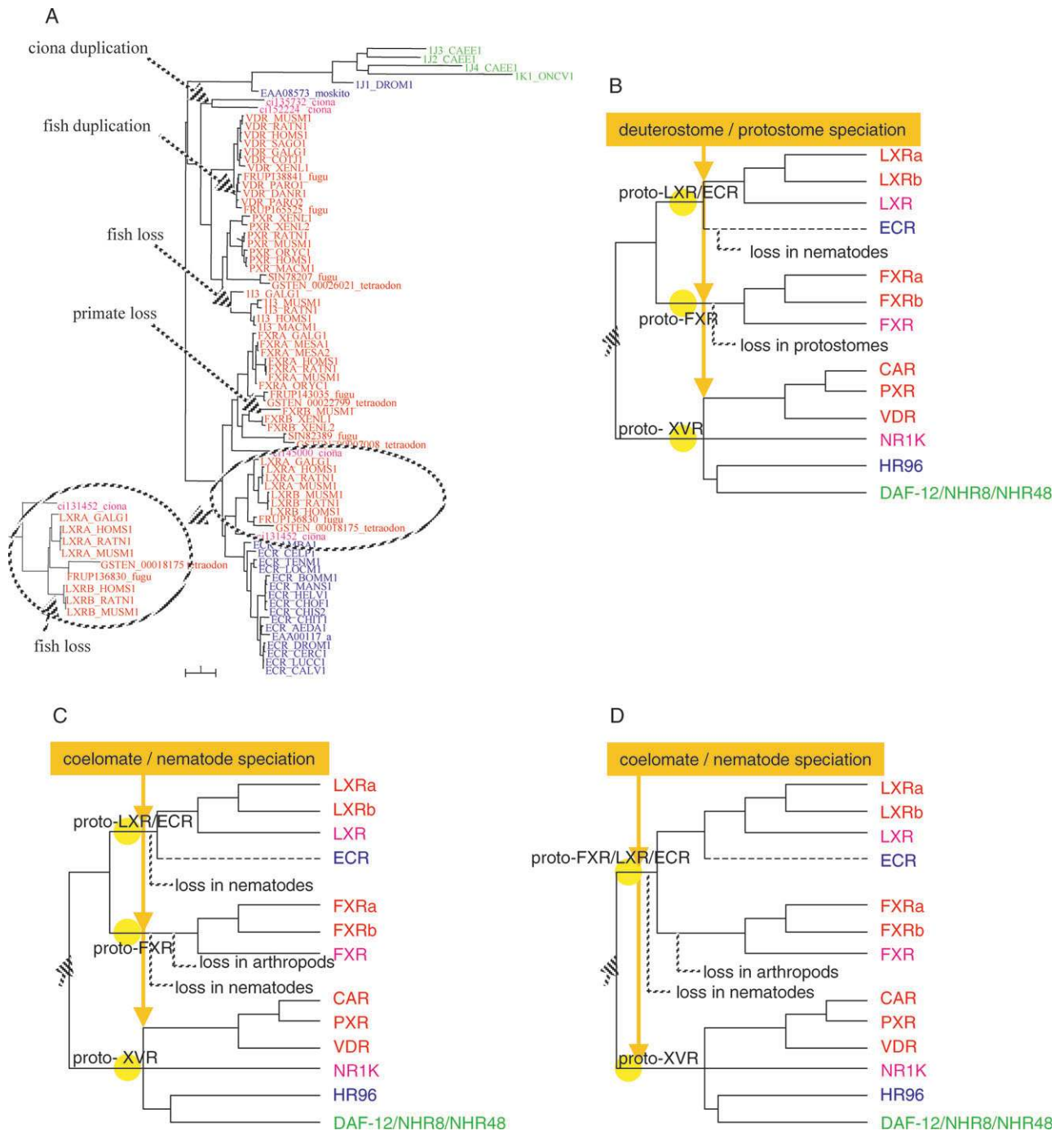


FIG. 4.—A problematic case: estimation of the number of ancestral genes for ecdysone receptors and closely related genes. Naming and color codes as in figure 1. (A) Maximum-likelihood tree; short predicted sequences are not included because they diminish dramatically the number of complete sites in the alignment. In the circle, subtree built with only LXR sequences, without interference by the long branches of the other sequences. All other conventions as in figure 3. (B–D) Simplified trees, as in figure 1, with evolutionary reasoning superimposed. Orange arrows indicate the first speciation separating these lineages. Yellow background circles indicate genes that are inferred to have existed in the common ancestor of insects, nematodes, and chordates. “Proto-XVR” = proto-xenobiotic vitamin D receptor. Thick-hatched branches indicate secondary gene loss. The arrow indicates the last common ancestor gene of all nuclear receptors in this tree. The broken line indicates the nonsignificant support for the phylogenetic position of EcR (SH test:  $P = 0.45$ ). (B) The “ecdysozoan” species tree is assumed to map gene duplication and loss (Aguinaldo et al. 1997); the orange arrows separate the deuterostome (chordates) and protostome (insects and nematodes) lineages. (C and D) The “coelomate” species tree is assumed to map gene duplication and loss (Wolf, Rogozin, and Koonin 2004); the orange arrows separate the coelomate and nematode lineages. (C) In the absence of further evidence, this speciation is assumed on the latest branch possible. (D) In the absence of further evidence, this speciation is assumed on the earliest branch possible.

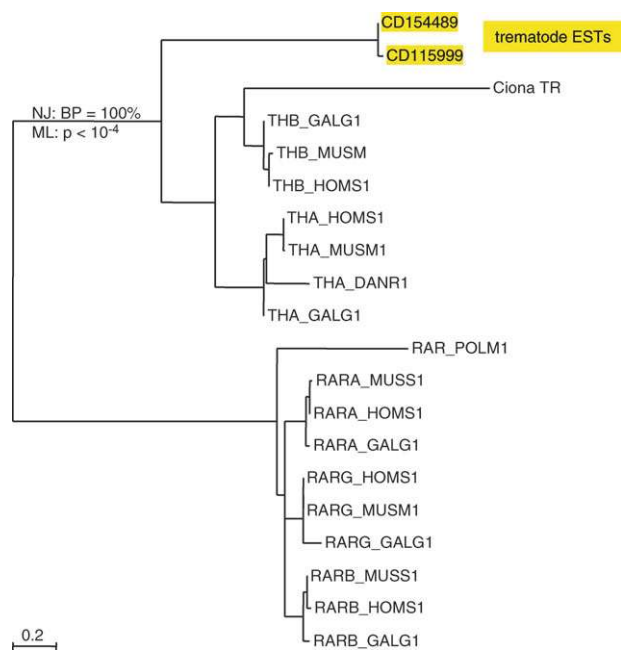


FIG. 5.—Phylogenetic position of the trematode thyroid hormone receptor. Maximum-likelihood (ML) tree; the sequences predicted from trematode EST sequences are on a yellow background. At the branch grouping trematode EST sequences with TR sequences, statistical support: above the branch, percentage of 2,000 bootstrap replicates in a Neighbor-Joining analysis; under the branch, *P*-value of an SH likelihood test. Due to short EST sequences (90 complete amino acid sites used), phylogenetic relationships among TR sequences are not well resolved, but the species tree is almost as likely as the ML tree: SH test, *P* = 0.61. All other conventions as in figure 3.

other NR2 sequences (relative-rate test:  $dK = 0.23 \pm 0.076$ ; *P* = 0.002), but NR1H sequences do not evolve significantly faster than other NR1 sequences ( $dK = 0.20 \pm 0.11$ ; *P* = 0.07). Moreover, there are other fast-evolving receptors whose phylogenetic classification is not problematic (i.e., PPARs, PXR/CAR).

#### Nuclear Receptor Subfamilies that Distinguish Vertebrates and Insects

While vertebrates have more NR genes than their urbilaterian ancestor (fig. 2), dipteran insects have fewer. In addition, there are groups of receptors that are specific to each lineage. Where do these differences come from? A first factor in the divergence of these two lineages is specific loss of different genes: seven nuclear receptors were lost in the lineage leading to dipteran insects since their divergence with deuterostomes (including chordates) and four were lost in the ancestor of chordates (fig. 2). Strikingly, all nuclear receptor genes missing in insects are also absent from nematodes, which may either mean that they were lost in the ancestor of ecdysozoans, with little latter loss in the insect lineage (fig. 2A), or that there was an extremely high parallelism in gene loss between the two lineages (fig. 2B). Data on lophotrochozoans would be especially useful to understand this phenomenon, as shown by the isolation of a mollusk ortholog of ER (Thornton, Need, and Crews 2003) or the finding of a trematode TR in

**Table 1**  
Nuclear Receptor Genes Inferred to Have Existed in the Uribilaterian

Urbilaterian NR Gene	Group	Comments
Proto-TR	1A	ESTs in a trematode
Proto-RAR	1B	
Proto-PPAR	1C	
Proto-Rev-erb/E75	1D	
Proto-E78/sex1	1E+G	
Proto-ROR/HR3	1F	
Proto-LXR/EcR	1H	Low phylogenetic resolution; dependent on animal phylogeny
Proto-FXR		
Proto-XVR <sup>a</sup>	1I+J+K	Descendants include VDR, PXR, CAR, DHR96, DAF-12, NHR8, NHR48
Proto-HNF4	2A	Descendants include nematode supnrs
Proto-RXR/USP	2B	ESTs in a trematode and a nematode
Proto-TR2-4/HR78/NHR41	2C+D	
Proto-TLL	2E	Low phylogenetic resolution; known gene in a coral
Proto-DSF		
Proto-PNR		
Proto-FAX-1		
Proto-COUP/SVP	2F	Unclear relation to unc-55
Proto-EAR2		
Proto-ER	3A	Known gene in a mollusk
Proto-ERR	3B	
Proto-SR	3C	
Proto-NR4	4A	Descendants include NURR1, NOR1, NGFIB, DHR38, CNR8
Proto-FTZF1/SF1/LRH1	5A	
Proto-HR39	5B	
Proto-GCNF	6A	

NOTE.—The Uribilaterian NR genes are the nuclear receptor genes inferred to have existed in the common ancestor of insects, nematodes, and chordates. They are represented by yellow background circles in figure 1.

<sup>a</sup> Proto-XVR = Proto-xenobiotic vitamin D receptor.

the EST databases (fig. 5), both genes that are missing in insects and nematodes.

The insect and nematode lineages (ectdysozoans) notably lost genes that code for major liganded receptors in vertebrates. The lethal or sterile phenotypes of knockout mice for these genes (for a review, see Laudet and Gronemeyer 2002) and their conservation among vertebrates suggest that the ancestral organism(s) lost genes which did not yet have their present function. Contradictory to this, a reconstructed ancestor of steroid receptors (ERs and SRs) appears to be regulated by estrogen (Thornton, Need, and Crews 2003). However, the mollusk ER is not regulated by estrogen, which implies that either the result on the ancestral receptor is an artifact or that it had insufficient functional importance in the urbilaterian to prevent loss of the gene (nematodes, insects) or of the function (mollusk). These essential hormone binding functions were probably either acquired or integrated into signaling pathways secondarily, together with the development of the vertebrate endocrine system; thyroid hormone signaling is thought to be specific to chordates (see Dehal et al. 2002), as is nuclear retinoic acid signaling, although other forms of retinoid signaling may



exist in insects (Mansfield et al. 1998; Adam, Perrimon, and Noselli 2003). Concerning steroid signaling, the only subfamily 3 receptor in insects is the orphan ERR (Adams et al. 2000; Ostberg et al. 2003), whereas the only steroid hormone known in insects, ecdysone, is mediated by EcR, a subfamily 1 nuclear receptor closely related to other vertebrate steroid receptors (FXRs and LXRs; fig. 4) but not to subfamily 3. The functional study of more orthologs of liganded nuclear receptors from diverse lineages, such as trematode TR (fig. 5), is needed to clarify this issue.

It should be emphasized that the data unambiguously support the hypothesis that these nuclear receptors were lost in nematodes and insects, as opposed to vertebrate “innovations.” This result is consistently found across the different nuclear receptors cited in this paragraph and across phylogenetic methods, with strong support, notably from maximum likelihood. For example, if subfamily 3 steroid receptors were a chordate or vertebrate invention, both insect ERR and mollusk ER should branch at the base of the subfamily 3 gene tree, a possibility strongly rejected by a likelihood ratio test (SH test:  $P = 0.0030$ ). This view contradicts our previous conclusion that most liganded nuclear receptors are chordate-specific (Escriva et al. 1997; Laudet 1997).

In a similar manner, the chordate lineage lost four genes that mediate insect-specific responses (fig. 2). Unlike genes lost in insects and nematodes, all are orphan receptors. FAX1 and DSF are both members of the TLL/PNR group, which bind DNA as strict homodimers (for a review Laudet and Gronemeyer 2002), which may have favored their loss in chordates (see Krylov et al. 2003). Two other genes lost in chordates, *Drosophila* HR39 and E78 are both active in ecdysone signaling, which is absent in chordates. It should be noted that an alternative scenario, with E78 as the result of an arthropod-specific duplication of E75, cannot be significantly excluded (likelihood SH test:  $P = 0.069$ ). Data from more diverse arthropods would probably help us reach a significant conclusion. Finally, if the coelomate tree of animals is correct (fig. 2B), chordates also lost the ancestor of Knirps, KNRL, and Eagle, a monophyletic group of proteins with a typical nuclear receptor DNA-binding domain but without any detectable similarity to a ligand-binding domain. These unusual members of the superfamily were previously only known in dipteran insects, which made them appear to be a dipteran innovation. However, we have found clear representatives of this group in trematode EST data, which pushes back significantly their appearance by domain recombination. In dipterans, they are active in the regulation of early developmental genes.

More generally, the only members of the superfamily known to lack the typical DBD-hinge-LBD structure are always different between insects and vertebrates, forming another source of divergence. The common functional point of these atypical nuclear receptors in dipterans and vertebrates seems to be their role in the regulation of the derived developmental features of dipterans and vertebrates (Laudet 1997). It may be noted that the diversity of nuclear receptors in nematodes also includes several proteins lacking the LBD or the DBD (Sluder et al. 1999), notably Odr-7 (Sengupta, Colbert, and Bargmann 1994),

and that these genes may also regulate derived features of nematodes. As a side note, there is an ortholog of fly Trithorax in the mosquito; it is not considered a nuclear receptor, although it includes a highly divergent nuclear receptor like DBD (Stassen et al. 1995).

As opposed to gene loss, gain by gene duplication is frequent in vertebrates, but almost absent in insects, and absent in early evolution of bilaterian lineages (fig. 2). While there are ten nuclear receptors with one ortholog in insects versus two or more in vertebrates, there is only one receptor with one ortholog in vertebrates versus two in a dipteran insect, and it is a recent duplication of TLL specifically in the mosquito. This abundance of vertebrate specific duplications is not restricted to nuclear receptors (Escriva, Laudet, and Robinson-Rechavi 2003), which reflect a general pattern of abundant gene (or genome) duplication at the origin of vertebrates that is discussed elsewhere (Holland et al. 1994; Dehal et al. 2002; McLysaght, Hokamp, and Wolfe 2002; Panopoulou et al. 2003; Robinson-Rechavi, Boussau, and Laudet 2004).

#### Derived Nuclear Receptor Complements for Derived Anatomies: Nematodes and the Sea Squirt

The highest rates of loss of nuclear receptor groups are found in two very distant groups of organisms: eight in nematodes and five in the sea squirt (fig. 2). Both are characterized by comparatively simple anatomy, brought to the extreme in *C. elegans* adult morphology, which has fewer than 1,000 cells. The apparent contradiction between this description of nematodes as having lost nuclear receptors and the abundance of more than 250 nuclear receptor genes in *C. elegans* (Sluder et al. 1999) comes from our phylogenetic reasoning. There are eight independent events of loss of genes, which most probably already had an established functional role at that point in evolution. On the other hand, the supplementary nuclear receptor genes in nematodes come from a unique burst of lineage-specific duplications (unpublished data), concerning only one group of the superfamily (HNF4). Thus, we obtain the apparent paradox that an evolutionary history dominated by loss of genes of known function also produced a huge diversity of novel nuclear receptors, whose function remains unclear.

Given the importance of nuclear receptors in development and gene regulation, it is tempting to establish a relationship between loss of regulatory genes and derived, simple anatomies and developmental pathways, especially because there are several cases of parallel gene loss. All nuclear receptors lost in the sea squirt except TLL were also lost in parallel in the nematode lineage (fig. 2). Concerning TLL, phylogenetic resolution is not significant, but nematode NHR67 seems to be the nematode ortholog of TLL. The parallel losses with nematodes (or ecdysozoans) notably include estrogen and “classical” steroid hormone receptors. Similarly, at least HR39 was lost in parallel in nematodes and in the ancestor of chordates; DSF was probably also lost in nematodes and chordates, although phylogenetic support is not significant. Both these genes are instrumental in the development of insect specialized structures, which nematodes lack. This suggests that both sea squirt and nematode lineages went through a process

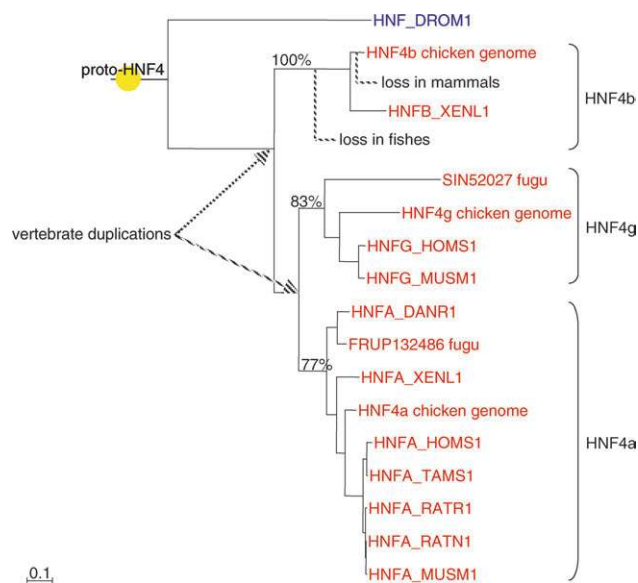


FIG. 6.—Phylogeny of HNF4 genes, including chicken genome sequences. Maximum-likelihood tree. The prereleased version “wash-uc1” was queried at [http://pre.ensembl.org/Gallus\\_gallus/](http://pre.ensembl.org/Gallus_gallus/). All other conventions as in figure 3.

of anatomical simplification, losing the corresponding regulatory genes. It is consistent with data that suggest that the anatomy both of nematodes (Aboobaker and Blaxter 2003) and sea squirt (Holland and Gibson-Brown 2003) are derived, and not ancestral. Notably, it would be interesting to investigate whether regulation cascades downstream of nuclear receptors have also been simplified in these organisms (Gissendanner et al. 2004).

#### Recent Evolution of Nuclear Receptor Subfamilies

The only functional difference between the two mammalian genomes is that the mouse (and rat, Zhang et al. 2004) genome contains FXRb, a paralog of FXR, which was lost in primates. Similarly, four genes (DAX1-2, SHP2, RARb, and ERRd) found in the fugu were lost specifically in tetraodon. When compared to *Drosophila*, the mosquito genome includes a specific duplication of TLL, but has lost Knirps. TLL is a key gene in the establishment of nonmetameric units of the embryo as well as in nervous system development in *Drosophila*, while Knirps is a transcription repressor also involved in the regulation of developmental genes during early development, some of them the same as TLL (i.e., Ftz). It is thus possible that this loss and this gain represent a divergence in developmental regulation pathways between flies and mosquitoes.

Most differences between pufferfish and mammal genomes are due to gene duplications in the ray-finned fish lineage (as already observed by Maglich et al. 2003), consistent with the general pattern of abundant gene/genome duplication in fishes (Robinson-Rechavi et al. 2001b; Christoffels et al. 2004; Vandepoele et al. 2004). Differences between the two lineages are also due to differential gene loss: four genes were lost in the mammalian (or tetrapode) lineage, and three were lost in pufferfishes (figs.

2, 3, and 4). Of note, we found duplications for two pufferfish NR genes which were not reported previously (Maglich et al. 2003): TRa1/2 and RARg1/2. It is also worth mentioning that RARb had never been isolated in a fish and was thought to be amniote-specific, but it is found in the pufferfish genomes, as well as in *Paralichthys olivaceus* (accession numbers BAB71757, BAB71755). Out of a total of 53 NR genes that can be estimated to have existed in the ancestor of vertebrates (48 in human + FXRb + 4 lost in mammals), 20 are secondarily duplicated in fishes (38%). Due to their number, we will not detail these 20 duplications here; most are described in Maglich et al. (2003).

A special case of gene loss is HNF4b, which appears to have been lost in parallel in both the pufferfish and mammalian lineages. We were also unable to clone it from zebrafish (unpublished data). Thus, HNF4b is absent from all the completed vertebrate genomes. The only previously known HNF4b was cloned in *Xenopus* (Holewa et al. 1997). A search on the prerelease of the chicken genome shows three HNF4 genes, clear orthologs of human HNF4a and HNF4g plus a clear ortholog of *Xenopus* HNF4b (fig. 6). This sequence clearly establishes that HNF4b is indeed a vertebrate-specific duplicate, lost independently in fishes and mammals.

All genes that were lost in the mammalian (or tetrapode) lineage or in the fish lineage are paralogs formed by vertebrate-specific duplications. The case of Rev-erbbs is interesting because of probable parallel loss in pufferfishes and mammals. The favored phylogeny (by likelihood and distance methods) shows a parallel loss of Rev-erbg in tetrapodes and of Rev-erba in tetraodontiformes. However, an alternative phylogeny, in which the “Rev-erbg” genes would be fast evolving Rev-erba genes, is not significantly less likely (SH test:  $P = 0.62$ ). Although the alternative phylogeny is appealing because it does not necessitate the parallel gene losses, we have cloned orthologs of both potential Rev-erbg genes in the zebrafish, which already has a Rev-erba (unpublished data). This supports the hypothesis that Rev-erba was lost in tetraodontiformes, which is consistent with the most likely phylogeny. Thus, we have retained this hypothesis of parallel loss of Rev-erbbs, although data from more diverse vertebrates (notably amphibians) will be needed to get a solid answer. The other cases of gene loss are well supported (likelihood tests,  $P < 0.05$ ).

#### Discussion

##### Periods of Gene Duplication or Loss

Pregenome studies of nuclear receptor evolution brought evidence for two “waves” (or periods) of gene duplication in the superfamily (Laudet et al. 1992; Laudet 1997): one before the diversification of major bilaterian lineages and the other specifically in vertebrates. Evidence from complete genomes confirms this picture, but refines it to a great extent. The most striking characteristic of the evolution of the superfamily, in fact, is that gene duplication and loss are not at all regularly distributed during evolution of species (fig. 2); there are branches of the phylogeny that represent “duplication periods” and there

are distinct branches that represent “gene loss periods,” with very little overlap between these two phenomena.

Branches dominated by gene duplication include the two originally identified (Laudet et al. 1992), plus the branch leading to fishes. At the origin of bilaterians, duplications can be divided into at least two “subperiods” (fig. 2). The first subperiod leads from an ancestral nuclear receptor gene to six paralogs, which represent the ancestors of the currently defined six subfamilies of nuclear receptors (Nuclear Receptors Nomenclature Committee 1999). The second leads to further diversification of each subfamily, in a very nonequibrated manner: while subfamilies 1 and 2 gave rise to approximately nine paralogs each, subfamilies 4 and 6 did not preserve any duplicates from this period (or did not experience any duplications). This diversification at the origin of bilaterians does not seem to have been linked to events in genome evolution, but these may be difficult to detect at such evolutionary distances. A similar trend has been observed for other gene families involved in regulation and development (Miyata and Suga 2001). On the other hand, the periods of nuclear receptor gene duplication at the origin of vertebrates and of teleost fishes are clearly part of more general patterns of genome evolution (Escriva, Laudet, and Robinson-Rechavi 2003); in both cases many gene families experienced gene duplication. This observation has given rise to widely discussed genome duplication hypotheses (Holland et al. 1994; Meyer and Schartl 1999; Postlethwait et al. 2000; Hughes, da Silva, and Friedman 2001; Robinson-Rechavi et al. 2001*b*, 2001*c*; Dehal et al. 2002; McLysaght, Hokamp, and Wolfe 2002; Hughes and Friedman 2003; Panopoulou et al. 2003; Christoffels et al. 2004; Robinson-Rechavi, Boussau, and Laudet 2004; Vandepoele et al. 2004). It seems that when nuclear receptor gene duplication was frequent, gene loss in the superfamily was rare; LXRb, Rev-erba, and CAR in the teleost fish lineage are exceptions, although gene loss in the ancestor of bilaterians would be harder to establish. Also, there is the special case of nematodes, with both high loss rate over the superfamily and explosive duplication specific to HNF4. There may be a link between low rates of gene loss and conservation of duplicate genes over long periods of evolutionary time. Such a link may be selection for gene diversity or neutral change in genome dynamics (as in Lynch and Conery 2003).

There are several branches of the species tree characterized by abundant gene loss. As previously mentioned, two of these events may be related to anatomical simplification: the sea squirt and nematode lineages. In the case of the nematodes, this loss is paralleled by duplications of two genes, proto-DAF12 and HNF4, the latter giving rise to a greater diversity of nuclear receptors than in any other animal. Thus, lineage-specific expansion (Lepinet et al. 2002) of one gene can more than compensate for an overall trend to gene loss. Another period of gene loss is at the origin of chordates (e.g., E78, DSF, FAX1, HR39). A third group of gene loss is more difficult to place, because of its dependence on the reference animal phylogeny used: if the cdysozoan hypothesis (Aguinaldo et al. 1997) is correct, then ecdysozoans paralleled chordates in important nuclear receptor gene losses at the

origin of the group; if the coelomate hypothesis is correct (Blair et al. 2002; Wolf, Rogozin, and Koonin 2004), then these genes were lost in parallel in nematodes and in insects. In either case, gene loss apparently played the major role in differentiating the different nuclear receptor complements of nematodes, insects, and chordates, not duplication. Finally, there is a minor period of gene loss apparent in our data, which is most unexpected in the lineage leading to mammals (e.g., COUP-TFg, ERRd, NR5A5, and probably Rev-erbg). The *Xenopus* and chicken genomes should be instrumental in determining whether these are mammal-specific or general to tetrapods. None of these four genes is found in the prerelease of the chicken genome, but a more complete sequence should be analyzed before we can draw final conclusions. These losses concern genes which had been “recently” duplicated at the origin of vertebrates and which can be supposed to have been relatively redundant at the time of loss. In a similar manner, Hox genes appear to have been prone to loss after duplication in fishes (Amores et al. 2004).

It may be noted that the coelomate hypothesis of animal phylogeny necessitates six additional parallel gene losses in the dipteran and nematode lineages (fig. 2*B*), which in the ecdysozoan hypothesis would be lost in the common ancestor of ecdysozoans (fig. 2*A*). Probably due to the fast evolution of nematode nuclear receptors (unpublished data), phylogenetic signal for the relative position of nematodes and insects is low in the phylogeny based on nuclear receptor sequences (fig. 1), but the gene loss pattern of nuclear receptor genes supports the ecdysozoan hypothesis.

#### Nuclear Receptor Genes, the Evolution of Endocrine Systems, and the “Cambrian Explosion”

Previous studies have noted that patterns of the presence/absence of nuclear receptor genes correspond to the evolution of major developmental and endocrine pathways, such as thyroid hormone regulation or “classical” steroid hormones (testosterone, glucocorticoids, mineralocorticoids, progesterone, and estrogen) in vertebrates (Dehal et al. 2002; Baker 2003). The phylogenetic approach we have adopted allows us to delineate and correct this notion. The most notable observation is that the presence of steroid hormone receptor or thyroid hormone receptor genes cannot be properly called a “chordate innovation,” as confirmed by the recent cloning of an ER in a mollusk (Thornton, Need, and Crews 2003) and our finding of a TR in a trematode (fig. 5). Rather, these genes are absent in other lineages because they were secondarily lost, but they were present in the ancestral Urbilaterian genome. In that case, the proper innovation is the function, not the presence of the gene, and this will not be understood by counting genes but by functional experiments in species chosen for their key phylogenetic position, such as sea squirt, amphioxus (e.g., Escriva et al. 2002), lamprey, or hagfish. Indeed, the precise status of the mollusk ER and the trematode TR remain to be established by detailed functional characterization, notably in terms of ligand binding and transcriptional activation.



To get the full picture of the evolution of steroid hormone signaling we should consider other nuclear receptors that bind steroids; it is most parsimonious to assume that the ancestor of LXRs, FXRs, and EcR (NR1H) was regulated by steroids, because all genes in this monophyletic group are steroid receptors (discussed in Escriva, Delaunay, and Laudet 2000). This implies that steroid binding was lost secondarily in nematodes. Considering this, it is probable that other animal lineages than insects and chordates have steroid receptors from the NR1H group. Moreover, since proto-SR and proto-ER genes were present in the ancestor of bilaterians (fig. 4), it is possible that there are other invertebrate lineages than those sampled now in which they were not secondarily lost. It should be noted that since steroids are also found in plants (for a review, Clouse 2002), they might have been used in signaling through membrane receptors before the divergence of plants and animals. It should also be noted that orthology does not always imply conservation of function, and orthologs of SRs or ERs may or may not be regulated by steroid hormones in divergent lineages. Indeed, the mollusk ER is not regulated by estrogen (Thornton, Need, and Crews 2003), and the expression of the fly E75 gene is regulated by ecdysone, a steroid hormone, whereas the expression of its vertebrate orthologs, Rev-erbs, is regulated by circadian rhythm (although cross-regulation between these pathways may exist). In any case, the results on NR1H as well as NR3 receptors imply that steroid hormone regulation is widespread, probably in all bilaterian animals.

The one gene group that presents a true innovation in vertebrates is the DAX1/SHP group, which results from a domain recombination that occurred after the divergence between vertebrates and sea squirt. This echoes a similar innovation leading to the Knirps/KNRL/Eagle group earlier in animal evolution. In both cases these genetic innovations seem to have been used to regulate specialized developmental features, although the role of the Knirps homolog in trematode remains to be determined.

The most important episode in nuclear receptor evolution occurred before the diversification of bilaterian animals, in their common ancestral lineage. Understanding of these events will come with further characterization of sponge or cnidarian nuclear receptors (see Grasso et al. 2001; Wiens et al. 2003). This period was dominated by gene duplication, also observed in other gene families (Ono et al. 1999; Suga et al. 1999; Miyata and Suga 2001). In contrast, the divergence of the major bilaterian lineages (nematodes/arthropods/chordates) is characterized by little or no detectable gene duplication but a lot of gene loss. Miyata and Suga (2001) already noted the paradox that there was very little gene duplication during the diversification of animal lineages (the “Cambrian explosion”). Our results, although limited to one gene family, suggest that the diversification of these lineages may have been fueled by differential loss of genes acquired during a previous period of duplication.

#### Consequences of Genome Sequencing for Nuclear Receptor Research

The most obvious consequence of genome sequencing on nuclear receptor research is that we now know the

complete set of nuclear receptor genes for nine animals, which include major model organisms: the mouse *M. musculus*, the fly *D. melanogaster*, and the worm *C. elegans*, as well as human. To these we should add the draft rat genome (Zhang et al. 2004). For each of these we also know the genomic structure of the genes (intron positions) and have access to the promoter sequence, although this is of limited utility without further experimental work (discussed in Robinson-Rechavi and Laudet 2003). We have tried in this study to go beyond the simple enumeration of parts and specify the relations between nuclear receptors of these different animals. This allows us to precisely pinpoint the differences between them and to specify which experimental results can be generalized and which cannot.

Thus, we have observed very conserved sets of nuclear receptor genes over the last 250 MYA, whereas more divergent species have very different NR complements. From this observation we can predict similar nuclear receptor contents in other closely related animals (other dipterans, other mammals, etc.), but we should expect surprises in nuclear receptor characterization from more divergent species. For example, the genome sequencing projects of a sea urchin (an echinoderm) or a bee (an hymenopteran insect) may well yield new NR genes or new patterns of loss. Moreover, we should be very careful in extrapolating nuclear receptor results among such large groups as vertebrates or insects, while we can more safely extrapolate among mammals or among dipterans (although differences do exist). This is notably important for more distant models of human physiology and development, such as fishes. Indeed, more than a third of the genes present in the common ancestor of fishes and mammals were duplicated in the pufferfish lineage. Adding the genes lost in either lineage, 86% of NR genes were touched by events of gain or loss separating these two NR complements. Our preliminary data on RT-PCR in the zebrafish *Danio rerio* (unpublished data) indicate that in many cases the duplication predates the divergence with pufferfishes, as has already been reported for some nuclear receptors (Marchand et al. 2001; Robinson-Rechavi et al. 2001b; Bardet et al. 2002; Bury et al. 2003), predicting a similar complement of nuclear receptors in this major model of vertebrate development. This observation is consistent with other reports of duplicated development genes in zebrafish, which probably derives from a genome duplication event (Amores et al. 1998; Postlethwait et al. 2000; Robinson-Rechavi et al. 2001c; Christoffels et al. 2004; Vandepoele et al. 2004).

#### Conclusion

Complete genome sequences give us two major insights into nuclear receptor biology: (1) a larger picture of nuclear receptor evolution, and thus endocrine system evolution, and (2) a complete list of parts for functional studies. Concerning the evolutionary picture, we have shown that the diversity of nuclear receptor genes was already well established in the ancestor of bilaterians. Later divergence between lineages was done by periods of gene duplication, but also of gene loss, lineage-specific



expansion of one gene (HNF4 in nematodes), and probably secondary acquisition of ligand-binding specificity. Concerning function, the most important message is that some model species, such as fishes but also the mouse, have more nuclear receptors than humans. We have also been able to delimit among which groups of species (mammals, fishes, dipterans, nematodes) generalization of nuclear receptor characterization is straightforward; the conservation between *D. melanogaster* and *A. gambiae*, for example, is encouraging for our understanding of the endocrinology of the vector of malaria. Beyond these limits much more care should be taken and results should be put in proper context. The mouse is an excellent model of human for all NRs except FXRs, while results in the zebrafish should always take into account the probability of additional paralogs.

### Acknowledgments

We thank Hughes Roest Crolius, Jorge Duarte, and David Ruau for help with the data and François Bonneton for helpful comments. Work was supported by “Bio-informatique” and “Post-génome Anophèle” inter-EPST grants; we also thank CNRS, MENRT, and Région Rhone-Alpes for financial support. S.B. is supported by HRA Pharma, Fondation Mérieux, and Association de Recherche contre le Cancer. We thank genome sequencing centers for making their data publicly available before publication.

### Literature Cited

- Aboobaker, A. A., and M. L. Blaxter. 2003. Hox gene loss during dynamic evolution of the nematode cluster. *Curr. Biol.* **13**:37–40.
- Adam, G., N. Perrimon, and S. Noselli. 2003. The retinoic-like juvenile hormone controls the looping of left-right asymmetric organs in *Drosophila*. *Development* **130**:2397–2406.
- Adams, M. D., S. E. Celniker, R. A. Holt et al. (196 co-authors). 2000. The genome sequence of *Drosophila melanogaster*. *Science* **287**:2185–2195.
- Adoutte, A., G. Balavoine, N. Lartillot, O. Lespinet, B. Prud'homme, and R. de Rosa. 2000. The new animal phylogeny: reliability and implications. *Proc. Natl. Acad. Sci. USA* **97**:4453–4456.
- Aguinaldo, A. M., J. M. Turbeville, L. S. Linford, M. C. Rivera, J. R. Garey, R. A. Raff, and J. A. Lake. 1997. Evidence for a clade of nematodes, arthropods and other moulting animals. *Nature* **387**:489–493.
- Altschul, S. F., W. Gish, W. Miller, E. W. Myers, and D. J. Lipman. 1990. Basic local alignment search tool. *J. Mol. Biol.* **215**:403–410.
- Amores, A., A. Force, Y. L. Yan et al. (13 co-authors). 1998. Zebrafish Hox clusters and vertebrate genome evolution. *Science* **282**:1711–1714.
- Amores, A., T. Suzuki, Y.-L. Yan, J. Pomeroy, A. Singer, C. Amemiya, and J. H. Postlethwait. 2004. Developmental roles of pufferfish Hox clusters and genome evolution in ray-fin fish. *Genome Res.* **14**:1–10.
- Auwerx, J., J. Drouin, and V. Laudet. 2003. Récepteurs à la Provençale. EMBO workshop on the biology of nuclear receptors. *EMBO Rep.* **4**:1122–1126.
- Baker, M. E. 2003. Evolution of adrenal and sex steroid action in vertebrates: a ligand-based mechanism for complexity. *Bioessays* **25**:396–400.
- Bardet, P. L., B. Horard, M. Robinson-Rechavi, V. Laudet, and J. M. Vanacker. 2002. Characterization of oestrogen receptors in zebrafish (*Danio rerio*). *J. Mol. Endocrinol.* **28**:153–163.
- Bardet, P. L., S. Obrecht-Pflumio, C. Thisse, V. Laudet, B. Thisse, and J. M. Vanacker. 2004. Cloning and developmental expression of five estrogen-receptor related (ERR) genes in the zebrafish. *Dev. Genes. Evol.* **214**:240–249.
- Blair, J. E., K. Ikeo, T. Gojobori, and S. B. Hedges. 2002. The evolutionary position of nematodes. *BMC Evol. Biol.* **2**:7.
- Bury, N. R., A. Sturm, P. Le Rouzic, C. Lethimonier, B. Ducouret, Y. Guiguen, M. Robinson-Rechavi, V. Laudet, M. E. Rafestin-Oblin, and P. Prunet. 2003. Evidence for two distinct functional glucocorticoid receptors in teleost fish. *J. Mol. Endocrinol.* **31**:141–156.
- Christoffels, A., E. G. L. Koh, J.-M. Chia, S. Brenner, S. Aparicio, and B. Venkatesh. 2004. Fugu genome analysis provides evidence for a whole-genome duplication early during the evolution of ray-finned fishes. *Mol. Biol. Evol.* **21**:1146–1151.
- Clouse, S. D. 2002. Brassinosteroid signal transduction: clarifying the pathway from ligand perception to gene expression. *Mol. Cell* **10**:973–982.
- Dehal, P., Y. Satou, R. K. Campbell et al. (87 co-authors). 2002. The draft genome of *Ciona intestinalis*: insights into chordate and vertebrate origins. *Science* **298**:2157–2167.
- Dopazo, H., J. Santoyo, and J. Dopazo. 2004. Phylogenomics and the number of characters required for obtaining an accurate phylogeny of eukaryote model species. *Bioinformatics* **20**:i116–121.
- Escriva, H., F. Delaunay, and V. Laudet. 2000. Ligand binding and nuclear receptor evolution. *Bioessays* **22**:717–727.
- Escriva, H., N. D. Holland, H. Gronemeyer, V. Laudet, and L. Z. Holland. 2002. The retinoic acid signaling pathway regulates anterior/posterior patterning in the nerve cord and pharynx of amphioxus, a chordate lacking neural crest. *Development* **129**:2905–2916.
- Escriva, H., V. Laudet, and M. Robinson-Rechavi. 2003. Nuclear receptors are markers of animal genome evolution. *J. Struct. Funct. Genomics* **3**:177–184.
- Escriva, H., R. Safi, C. Hänni, M.-C. Langlois, P. Saumitou-Laprade, D. Stehelin, A. Capron, R. Pierce, and V. Laudet. 1997. Ligand binding was acquired during evolution of nuclear receptors. *Proc. Natl. Acad. Sci. USA* **94**:6803–6808.
- Galtier, N., M. Gouy, and C. Gautier. 1996. SEAVIEW and PHYLO\_WIN: two graphic tools for sequence alignment and molecular phylogeny. *Comput. Appl. Biosci.* **12**:543–548.
- Gissendanner, C. R., K. Crossgrove, K. A. Kraus, C. V. Maina, and A. E. Sluder. 2004. Expression and function of conserved nuclear receptor genes in *Caenorhabditis elegans*. *Dev. Biol.* **266**:399–416.
- Grasso, L. C., D. C. Hayward, J. W. H. Trueman, K. M. Hardie, P. A. Janssens, and E. E. Ball. 2001. The evolution of nuclear receptors: evidence from the coral *Acropora*. *Mol. Phyl. Evol.* **21**:93–102.
- Guindon, S., and O. Gascuel. 2003. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst. Biol.* **52**:696–704.
- Gustafsson, J. A. 1999. Seeking ligands for lonely orphan receptors. *Science* **284**:1285–1286.
- Holewa, B., D. Zapp, T. Drewes, S. Senkel, and G. U. Ryffel. 1997. HNF4beta, a new gene of the HNF4 family with distinct activation and expression profiles in oogenesis and embryogenesis of *Xenopus laevis*. *Mol. Cell. Biol.* **17**:687–694.
- Holland, L. Z., and J. J. Gibson-Brown. 2003. The *Ciona intestinalis* genome: when the constraints are off. *Bioessays* **25**:529–532.

- Holland, P. W., J. Garcia-Fernandez, N. A. Williams, and A. Sidow. 1994. Gene duplications and the origins of vertebrate development. *Development (Suppl.)*:125–133.
- Holt, R. A., G. M. Subramanian, A. Halpern et al. (123 co-authors). 2002. The genome sequence of the malaria mosquito *Anopheles gambiae*. *Science* **298**:129–149.
- Hughes, A. L., J. da Silva, and R. Friedman. 2001. Ancient genome duplications did not structure the human Hox-bearing chromosomes. *Genome Res.* **11**:771–780.
- Hughes, A. L., and R. Friedman. 2003. 2R or not 2R: testing hypotheses of genome duplication in early vertebrates. *J. Struct. Funct. Genomics* **3**:85–93.
- Jones, D. T., W. R. Taylor, and J. M. Thornton. 1992. The rapid generation of mutation data matrices from protein sequences. *Comput. Appl. Biosci.* **8**:275–282.
- Kliwer, S. A., J. M. Lehmann, and T. M. Willson. 1999. Orphan nuclear receptors: shifting endocrinology into reverse. *Science* **284**:757–760.
- Krylov, D. M., Y. I. Wolf, I. B. Rogozin, and E. V. Koonin. 2003. Gene loss, protein sequence divergence, gene dispensability, expression level, and interactivity are correlated in eukaryotic evolution. *Genome Res.* **13**:2229–2235.
- Laudet, V. 1997. Evolution of the nuclear receptor superfamily: early diversification from an ancestral orphan receptor. *J. Mol. Endocrinol.* **19**:207–226.
- Laudet, V., and H. Gronemeyer. 2002. The nuclear receptors factsbook. Academic Press, London.
- Laudet, V., C. Hänni, J. Coll, C. Catzeflis, and D. Stéhelin. 1992. Evolution of the nuclear receptor gene family. *EMBO J.* **11**:1003–1013.
- Lespinet, O., Y. I. Wolf, E. V. Koonin, and L. Aravind. 2002. The role of lineage-specific gene family expansion in the evolution of eukaryotes. *Genome Res.* **12**:1048–1059.
- Lynch, M., and J. S. Conery. 2003. The origins of genome complexity. *Science* **302**:1401–1404.
- Maglich, J. M., J. A. Caravella, M. H. Lambert, T. M. Willson, J. T. Moore, and L. Ramamurthy. 2003. The first completed genome sequence from a teleost fish (*Fugu rubripes*) adds significant diversity to the nuclear receptor superfamily. *Nucleic Acids Res.* **31**:4051–4058.
- Maglich, J. M., A. E. Sluder, X. Guan, Y. Shi, D. D. McKee, K. Carrick, K. Kamdar, T. M. Willson, and J. T. Moore. 2001. Comparison of complete nuclear receptor sets from the human, *Caenorhabditis elegans* and *Drosophila* genomes. *GenomeBiology.com* **2**:research0029.0021–0029.0027.
- Mallatt, J., and C. J. Winchell. 2002. Testing the new animal phylogeny: first use of combined large-subunit and small-subunit rRNA gene sequences to classify the protostomes. *Mol. Biol. Evol.* **19**:289–301.
- Mansfield, S. G., S. Cammer, S. C. Alexander, D. P. Muehleisen, R. S. Gray, A. Tropsha, and W. E. Bollenbacher. 1998. Molecular cloning and characterization of an invertebrate cellular retinoic acid binding protein. *Proc. Natl. Acad. Sci. USA* **95**:6825–6830.
- Marchand, O., R. Safi, H. Escriva, E. Van Rompaey, P. Prunet, and V. Laudet. 2001. Molecular cloning and characterization of thyroid hormone receptors in teleost fish. *J. Mol. Endocrinol.* **21**:51–65.
- McLysaght, A., K. Hokamp, and K. H. Wolfe. 2002. Extensive genomic duplication during early chordate evolution. *Nat. Genet.* **31**:200–204.
- Meyer, A., and M. Schartl. 1999. Gene and genome duplications in vertebrates: the one-to-four (-to-eight in fish) rule and the evolution of novel gene functions. *Curr. Opin. Cell Biol.* **1**:699–704.
- Miyata, T., and H. Suga. 2001. Divergence pattern of animal gene families and relationship with the Cambrian explosion. *Bioessays* **23**:1018–1027.
- Nuclear Receptors Nomenclature Committee. 1999. A unified nomenclature system for the nuclear receptor superfamily. *Cell* **97**:1–3.
- Ono, K., H. Suga, N. Iwabe, K. Kuma, and T. Miyata. 1999. Multiple protein tyrosine phosphatases in sponges and explosive gene duplication in the early evolution of animals before the parazoan-eumetazoan split. *J. Mol. Evol.* **48**:654–662.
- Ostberg, T., M. Jacobsson, A. Attersand, A. Mata de Urquiza, and L. Jendeberg. 2003. A triple mutant of the *Drosophila* ERR confers ligand-induced suppression of activity. *Biochemistry* **42**:6427–6435.
- Panopoulou, G., S. Hennig, D. Groth, A. Krause, A. J. Poustka, R. Herwig, M. Vingron, and H. Lehrach. 2003. New evidence for genome-wide duplications at the origin of vertebrates using an amphioxus gene set and completed animal genomes. *Genome Res.* **13**:1056–1066.
- Postlethwait, J. H., I. G. Woods, P. Ngo-Hazelett, Y. L. Yan, P. D. Kelly, F. Chu, H. Huang, A. Hill-Force, and W. S. Talbot. 2000. Zebrafish comparative genomics and the origins of vertebrate chromosomes. *Genome Res.* **10**:1890–1902.
- Rat Genome Sequencing Project Consortium. 2004. Genome sequence of the Brown Norway rat yields insights into mammalian evolution. *Nature* **428**:493–521.
- Robinson-Rechavi, M., B. Boussau, and V. Laudet. 2004. Phylogenetic dating and characterization of gene duplications in vertebrates: the cartilaginous fish reference. *Mol. Biol. Evol.* **21**:580–586.
- Robinson-Rechavi, M., A.-S. Carpentier, M. Duffraisse, and V. Laudet. 2001a. How many nuclear hormone receptors in the human genome? *Trends Genet.* **17**:554–556.
- Robinson-Rechavi, M., and V. Laudet. 2003. Bioinformatics of nuclear receptors. *Meth. Enz.* **364**:95–118.
- Robinson-Rechavi, M., C. V. Maina, C. R. Gissendar, V. Laudet, and A. Sluder. 2004. Explosive lineage-specific expansion of the orphan nuclear receptor HNF4 in nematodes. *J. Mol. Evol.* (in press).
- Robinson-Rechavi, M., O. Marchand, H. Escriva, P.-L. Bardet, D. Zelus, S. Hughes, and V. Laudet. 2001b. Euteleost fish genomes are characterized by expansion of gene families. *Genome Res.* **11**:781–788.
- Robinson-Rechavi, M., O. Marchand, H. Escriva, and V. Laudet. 2001c. An ancestral whole-genome duplication may not have been responsible for the abundance of duplicated fish genes. *Curr. Biol.* **11**:R458–R459.
- Ruau, D., J. Duarte, T. Ourjadal, G. Perriere, V. Laudet, and M. Robinson-Rechavi. 2004. Update of Nurebase: nuclear hormone receptor functional genomics. *Nucleic Acids Res.* **32** (Database):D165–167.
- Ruse, M. D., Jr., M. L. Privalsky, and F. M. Sladek. 2002. Competitive cofactor recruitment by orphan receptor hepatocyte nuclear factor 4 $\alpha$ 1: modulation by the F domain. *Mol. Cell. Biol.* **22**:1626–1638.
- Saitou, N., and M. Nei. 1987. The Neighbor-Joining method: a new method for reconstructing phylogenetic trees. *Mol. Biol. Evol.* **4**:406–425.
- Schmidt, H. A., K. Strimmer, M. Vingron, and A. von Haeseler. 2002. Tree-Puzzle: maximum likelihood phylogenetic analysis using quartets and parallel computing. *Bioinformatics* **18**:502–504.
- Sengupta, P., H. A. Colbert, and C. I. Bargmann. 1994. The *C. elegans* gene *odr-7* encodes an olfactory-specific member of the nuclear receptor superfamily. *Cell* **79**:971–980.
- Shimodaira, H., and M. Hasegawa. 1999. Multiple comparisons of log-likelihoods with applications to phylogenetic inference. *Mol. Biol. Evol.* **16**:1114–1116.

- Sluder, A. E., and C. V. Maina. 2001. Nuclear receptors in nematodes: themes and variations. *Trends Genet.* **17**:206–213.
- Sluder, A. E., S. W. Mathews, D. Hough, V. P. Yin, and C. V. Maina. 1999. The nuclear receptor superfamily has undergone extensive proliferation and diversification in nematodes. *Genome Res.* **9**:103–120.
- Stassen, M. J., D. Bailey, S. Nelson, V. Chinwalla, and P. J. Harte. 1995. The *Drosophila trithorax* proteins contain a novel variant of the nuclear receptor type DNA binding domain and an ancient conserved motif found in other chromosomal proteins. *Mech. Dev.* **52**:209–223.
- Stein, L. D., Z. Bao, D. Blasiar et al. (35 co-authors). 2003. The genome sequence of *Caenorhabditis briggsae*: a platform for comparative genomics. *PLoS Biol.* **1**:166–192.
- Suga, H., M. Koyanagi, D. Hoshiyama, K. Ono, N. Iwabe, K. Kuma, and T. Miyata. 1999. Extensive gene duplication in the early evolution of animals before the parazoan-eumetazoan split demonstrated by G proteins and protein tyrosine kinases from sponge and hydra. *J. Mol. Evol.* **48**:646–653.
- Thornton, J. W., E. Need, and D. Crews. 2003. Resurrecting the ancestral steroid receptor: ancient origin of estrogen signaling. *Science* **301**:1714–1717.
- Vandepoele, K., W. De Vos, J. S. Taylor, A. Meyer, and Y. Van De Peer. 2004. Major events in the genome evolution of vertebrates: paranome age and size differ considerably between ray-finned fishes and land vertebrates. *Proc. Natl. Acad. Sci. USA* **101**:1638–1643.
- Wang, Z., G. Benoit, J. Liu, S. Prasad, P. Aarnisalo, X. Liu, H. Xu, N. P. Walker, and T. Perlmann. 2003. Structure and function of Nurr1 identifies a class of ligand-independent nuclear receptors. *Nature* **423**:555–560.
- Waterston, R. H., K. Lindblad-Toh, E. Birney et al. (222 co-authors). 2002. Initial sequencing and comparative analysis of the mouse genome. *Nature* **420**:520–562.
- Wiens, M., R. Batel, M. Korzhev, and W. E. G. Muller. 2003. Retinoid X receptor and retinoic acid response in the marine sponge *Suberites domuncula*. *J. Exp. Biol.* **206**:3261–3271.
- Wolf, Y. I., I. B. Rogozin, and E. V. Koonin. 2004. Coelomata and not Ecdysozoa: evidence from genome-wide phylogenetic analysis. *Genome Res.* **14**:29–36.
- Yagi, K., Y. Satou, F. Mazet, S. M. Shimeld, B. Degnan, D. Rokhsar, M. Levine, Y. Kohara, and N. Satoh. 2003. A genome-wide survey of developmentally relevant genes in *Ciona intestinalis*. III. Genes for Fox, ETS, nuclear receptors and NFkappaB. *Dev. Genes. Evol.* **213**:235–244.
- Zhang, Z., P. E. Burch, A. J. Cooney, R. B. Lanz, F. A. Pereira, J. Wu, R. A. Gibbs, G. Weinstock, and D. A. Wheeler. 2004. Genomic analysis of the nuclear receptor family: new insights into structure, regulation, and evolution from the rat genome. *Genome Res.* **14**:580–590.

William Jeffery, Associate Editor

Accepted June 23, 2004