

# Letter to the Editor

## A Unified Nomenclature System for the Nuclear Receptor Superfamily

Nuclear hormone receptors (NRs) are important transcriptional regulators involved in widely diverse physiological functions such as control of embryonic development, cell differentiation, and homeostasis (Gronemeyer and Laudet, 1995; Mangelsdorf et al., 1995). In addition, these molecules are extremely important in medical research since a large number of them are implicated in diseases such as cancer, diabetes, or hormone resistance syndromes. Some of the NRs act as ligand-inducible transcription factors, while a large number of them have no defined ligand and are hence described as "orphan" receptors (Enmark and Gustafsson, 1996). Over the last decade, workers in the field have described more than 300 sequences of NRs using an increasingly complex and baroque nomenclature. The existence of several names for the same gene is an acute problem for the orphan receptors, which often cannot be described by their function, particularly at the moment of their discovery. As discussed during the Seventh International CBT Symposium on "Nuclear Orphan Receptors" in Huddinge, Sweden (September 9-12, 1995), this plethora of names has become more and more confusing and now constitutes a barrier for understanding of newly acquired knowledge to researchers outside as well as within the field. For that reason, four of us (V. L., J. A., J.-A. G., and W. W.) agreed to form a committee for the nomenclature of NRs. It is the purpose of this paper to recommend names for the subfamilies and groups of receptors based on a phylogenetic tree connecting all known NR sequences. This system, based on the evolution of the two well-conserved domains of NRs (the DNA-binding C domain and the ligand-binding E domain), offers a practical and significant framework to which subsequent genes can be easily added. The resulting nomenclature has now been endorsed by over 40 scientists<sup>1</sup> many of whom contributed to defining the nomenclature and to preparing this letter. This nomenclature has been discussed with the International Union of Pharmacology Committee on Receptor Nomenclature and Drug Classification (NC-IUPHAR). A subcommittee of NC-IUPHAR entitled "Nuclear Receptors" will be set up to further clarify receptor nomenclature to integrate structure and function.

A summary of the new nomenclature is presented in Table 1.

### Rationale of the System

A complete description of the alignment procedures, tree reconstruction methods, and evolutionary implications can be found in Laudet, 1997. A more complete description of our system can be found in a web site devoted to the regular implementation of this nomenclature (<http://www.ens-lyon.fr/LBMC/LAUDET/nomenc.html>). Briefly, subfamilies are defined as the last most internal branches of the evolutionary tree with robustness ("bootstrap") values above 90% (boxed in Figure 1). In the case of the NRs, the groups defined with

this criteria correspond to known functional groups of receptors. This procedure yields six subfamilies. All the unusual receptors that contain only one of the two conserved domains (C or E) were grouped into a separate subfamily (subfamily 0) irrespective of their evolutionary origin. Within subfamilies, groups of receptors are defined as the most internal branches with bootstrap values above 90%. In this nomenclature system, the number of a given receptor inside a group does not carry any specific information. In many cases these groups contain arthropod and vertebrate members. The various homologs of the same gene in invertebrates (e.g., *Drosophila* and *Caenorhabditis*) have the same name. It is

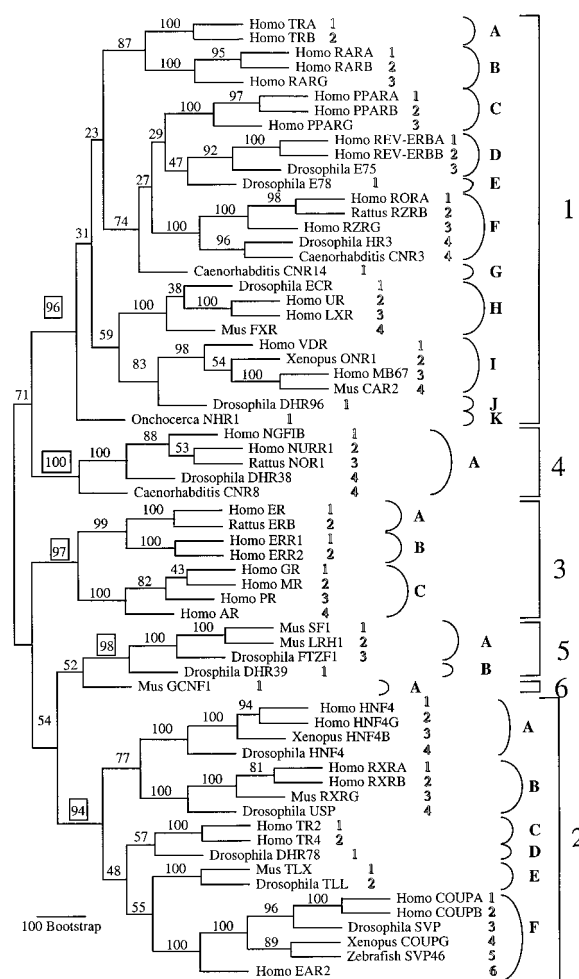


Figure 1. Phylogenetic Tree Connecting the 65 Known NR Genes in Vertebrates, Arthropods, and Nematodes

The length of the branches is proportional to the bootstrap value. Only one version of each individual gene was included in the tree. Subfamilies are indicated by arabic numerals at the extreme right of the figure, groups by capital letters and brackets, and individual genes by arabic numerals together with a representative name. The scale shows the length of the maximum possible bootstrap value (100). The bootstrap values defining the subfamilies are boxed, except in the case of GCNF1 defined by only one member.

Table 1. A Proposed Nomenclature for Nuclear Receptors

Subfamily and Group	NR/Gene	Trivial Names	Accession Number
1A	NR1A1	TR $\alpha$ , c-erbA-1, THRA	M24748
	NR1A2	TR $\beta$ , c-erbA-2, THRB	X04707
1B	NR1B1	RAR $\alpha$	X06538
	NR1B2	RAR $\beta$ , HAP	Y00291
	NR1B3	RAR $\gamma$ , RARD	M57707
1C	NR1C1	PPAR $\alpha$	L02932
	NR1C2	PPAR $\beta$ , NUC1, PPAR $\delta$ , FAAR	L07592
	NR1C3	PPAR $\gamma$	L40904
1D	NR1D1	REVERB $\alpha$ , EAR1, EAR1A	M24898
	NR1D2	REVERB $\beta$ , EAR1 $\beta$ , BD73, RVR, HZF2	L31785
	NR1D3	E75	X51548
1E	NR1E1	E78, DR-78	U01087
1F	NR1F1	ROR $\alpha$ , RZR $\alpha$	U04897
	NR1F2	ROR $\beta$ , RZR $\beta$	Y08639
	NR1F3	ROR $\gamma$ , TOR	U16997
	NR1F4	HR3, DHR3, MHR3, GHR3, CNR3, CHR3	M90806 U13075
1G	NR1G1	CNR14	U13074
1H	NR1H1	ECR	M74078
	NR1H2	UR, OR-1, NER1, RIP15, LXR $\beta$	U07132
	NR1H3	RLD1, LXR, LXR $\alpha$	U22662
	NR1H4	FXR, RIP14, HRR1	U09416
1I	NR1I1	VDR	J03258
	NR1I2	ONR1, PXR, SXR, BXR	X75163
	NR1I3	MB67, CAR1, CAR $\alpha$	Z30425
	NR1I4	CAR2, CAR $\beta$	AF00932
1J	NR1J1	DHR96	U36792
1K	NR1K1	NHR1	U19360
2A	NR2A1	HNF4	X76930
	NR2A2	HNF4G	Z49826
	NR2A3	HNF4B	Z49827
	NR2A4	DHNF4, HNF4D	U70874
2B	NR2B1	RXRA	X52773
	NR2B2	RXRB, H-2RIIBP, RCoR-1	M84820
	NR2B3	RXRG	X66225
	NR2B4	USP, Ultraspiracle, 2C1, CF1	X52591
2C	NR2C1	TR2, TR2-11	M29960
	NR2C2	TR4, TAK1	L27586
2D	NR2D1	DHR78	U36791
2E	NR2E1	TLL, TLX, XTLL	S72373
	NR2E2	TLL, Tailless	M34639
2F	NR2F1	COUP-TFI, COUPTFA, EAR3, SVP44	X12795
	NR2F2	COUP-TFII, COUPTFB, ARP1, SVP40	M64497
	NR2F3	SVP, COUP-TF	M28863
	NR2F4	COUP-TFIII, COUPTFG	X63092
	NR2F5	SVP46	X70300
	NR2F6	EAR2	X12794
3A	NR3A1	ER $\alpha$	X03635
	NR3A2	ER $\beta$	U57439
3B	NR3B1	ERR1, ERR $\alpha$	X51416
	NR3B2	ERR2, ERR $\beta$	X51417
3C	NR3C1	GR	X03225
	NR3C2	MR	M16801
	NR3C3	PR	M15716
	NR3C4	AR	M20132

continued

Table 1. Continued

Subfamily and Group	NR/Gene	Trivial Names	Accession Number
4A	NR4A1	NGFIB, TR3, N10, NUR77, NAK1	L13740
	NR4A2	NURR1, NOT, RNR1, HZF-3, TINOR	X75918
	NR4A3	NOR1, MINOR	D38530
	NR4A4	DHR38, NGFIB	U36762
		CNR8, C48D5	U13076
5A	NR5A1	SF1, ELP, FTZ-F1, AD4BP	D88155
	NR5A2	LRH1, xFF1rA, xFF1rB, FFLR, PHR, FTF	U93553
	NR5A3	FTZ-F1	M63711
5B	NR5B1	DHR39, FTZF1B	L06423
6A	NR6A1	GCNF1, RTR	U14666
0A	NR0A1	KNI, Knirps	X13331
	NR0A2	KNRL, Knirps related	X14153
	NR0A3	EGON, Embryonic gonad, EAGLE	X16631
	NR0A4	ODR7	U16708
	NR0A5	Trithorax	M31617
0B	NR0B1	DAX1, AHCH	S74720
	NR0B2	SHP	L76571

Note: subfamilies and groups are defined in the text. The groups contain highly related genes with often paralogous relationship in vertebrates (e.g., RARA, RARB, and RARG). The term isoform is reserved for different gene products originating from the same gene due to alternative promoter usage or splicing, or alternative initiation of translation.

not possible to define a strict percentage identity limit for belonging to a particular group but members of the same group in general share at least 80%–90% identity in the DNA-binding domain and at least 40%–60% in the ligand-binding domain.

Our system is based on the nomenclature system that was developed for cytochrome P450 by Nebert et al. (1987). This system has proven to be convenient and flexible, allowing for the inclusion of an ever increasing number of cytochrome P450 genes. We anticipate that our system will provide similar advantages.

Since editors of computerized nucleotide sequence databases discourage the use of Greek letters and hyphens in gene names, we have avoided their use except for the distinction between recent products of tetraploidization events in species such as zebrafish or *Xenopus*. We therefore recommend that gene subfamilies be designated by arabic numerals, groups by capital letters, and individual genes by arabic numerals. In the case of functionally and structurally distinct variants derived from the same gene (e.g., the isoforms SF1 and ELP, the two products of the mouse *Ftzh1* gene) a lower-case letter is added at the end of the name. For example, SF1 can be referred to as NR5A1a and ELP as NR5A1b. Only major isoforms can be considered at this stage, and implementation of the isoform nomenclature will be set up in the web page.

#### Use of the Nomenclature System

In each manuscript dealing with NRs, it is recommended that the receptor(s) be identified by the official name(s) at least once in the Summary and the Introduction. No hyphen is necessary between NR and the subfamily,

group, and gene numbers. Once the name has been established (e.g., "this paper describes GCNF1 [NR6A1], a member of the nuclear hormone receptor superfamily"), authors may use the trivial name for the remainder of the manuscript. When authors describe a new NR sequence they are kindly requested—after acceptance of the paper but before reading of galley proofs, to send the amino acid sequence of the receptor in complete confidentiality to V. L. (e-mail, [vincent.laudet@ens-lyon.fr](mailto:vincent.laudet@ens-lyon.fr); fax, 33 4 72 72 80 80; phone, 33 4 72 72 81 90). An acknowledgment of receipt will be immediately sent by e-mail or fax, and the official names will be returned within a week together with an evolutionary tree.

#### Nuclear Receptors Nomenclature Committee<sup>2,3</sup>

##### References

- Enmark, E., and Gustafsson, J.A. (1996). Orphan nuclear receptors—the first eight years. *Mol. Endocrinol.* 10, 1293–1307.
- Gronemeyer, H., and Laudet, V. (1995). Nuclear receptors. *Protein Profile* 2, 1173–1308.
- Laudet, V. (1997). Evolution of the nuclear receptor superfamily: early diversification from an ancestral orphan receptor. *J. Mol. Endocrinol.* 19, 207–226.
- Mangelsdorf, D.J., Thummel, C., Beato, M., Herrlich, P., Schütz, G., Umesono, K., Blumberg, B., Kastner, P., Mark, M., Chambon, P., and Evans, R.M. (1995). The nuclear receptor superfamily: the second decade. *Cell* 83, 835–839.
- Nebert, D.W., Adesnik, M., Coon, M.J., Estabrook, R.W., Gonzalez, F.J., Guengerich, F.P., Gunsalus, I.C., Johnson, E.F., Kemper, B., Levin, W., et al. (1987). The P450 gene superfamily: recommended nomenclature. *DNA* 6, 1–11.

---

<sup>1</sup> The following scientists have endorsed the use of this nomenclature: J. Auwerx, E. Baulieu, M. Beato, M. Becker-Andre, P. H. Burbach, G. Camerino, P. Chambon, A. Cooney, A. Dejean, C. Dreyer, R. M. Evans, F. Gannon, V. Giguere, H. Gronemeyer, J.-A. Gustafsson, V. Laudet, M. A. Lazar, D. J. Mangelsdorf, J. Milbrandt, E. Milgrom, D. D. Moore, B. O'Malley, M. Parker, K. Parker, T. Perlmann, M. Pfahl, M. G. Rosenfeld, H. Samuels, G. Schütz, F. M. Sladek, H. G. Stunnenberg, M. Spedding, C. Thummel, M.-J. Tsai, K. Umesono, B. Vennstrom, W. Wahli, C. Weinberger, T. M. Willson, and K. Yamamoto.

<sup>2</sup> The appropriate text citation for this letter in future publications will be: (Nuclear Receptors Nomenclature Committee, 1999).

<sup>3</sup> The committee consists of: Vincent Laudet,\* Johan Auwerx,† Jan-Ake Gustafsson,‡ and Walter Wahli.§

\*UMR5665 du CNRS, Ecole Normale Supérieure de Lyon, 46 Allée d'Italie, 69364, Lyon Cedex 07, France.

†Laboratoire de Biologie des Régulations chez les Eucaryotes, INSERM U325, SERLIA, Institut Pasteur, 1 rue Calmette, 59019, Lille Cedex, France.

‡Karolinska Institute, Department of Biosciences, NOVUM, 141 57, Huddinge, Sweden.

§Institut de Biologie Animale, Université de Lausanne, Bâtiment de Biologie, CH-1015 Lausanne, Switzerland.