The EMBO Journal vol.11 no.3 pp.1003-1013, 1992

Evolution of the nuclear receptor gene superfamily

Vincent Laudet, Catherine Hänni, Jean Coll, François Catzeflis[1] and Dominique Stéhelin

INSERM U186/CNRS URA 1160, Institut Pasteur, 1 Rue Calmette, 59019 Lille Cedex and [1]Laboratoire de Paléontologie, Institut des Sciences de l'Evolution (UA 327 CNRS), USTL, Place Eugène Bataillon, 34095 Montpellier Cedex, France

Communicated by D.Stéhelin

Nuclear receptor genes represent a large family of genes encoding receptors for various hydrophobic ligands such as steroids, vitamin D, retinoic acid and thyroid hormones. This family also contains genes encoding putative receptors for unknown ligands. Nuclear receptor gene products are composed of several domains important for transcriptional activation, DNA binding (C domain), hormone binding and dimerization (E domain). It is not known whether these genes have evolved through gene duplication from a common ancestor or if their different domains came from different independent sources. To test these possibilities we have constructed and compared the phylogenetic trees derived from two different domains of 30 nuclear receptor genes. The tree built from the DNA binding C domain clearly shows a common progeny of all nuclear receptors, which can be grouped into three subfamilies: (i) thyroid hormone and retinoic acid receptors, (ii) orphan receptors and (iii) steroid hormone receptors. The tree constructed from the central part of the E domain which is implicated in transcriptional regulation and dimerization shows the same distribution in three subfamilies but two groups of receptors are in a different position from that in the C domain tree: (i) the *Drosophila knirps* family genes have acquired very different E domains during evolution, and (ii) the vitamin D and ecdysone receptors, as well as the FTZ-F1 and the NGF1B genes, seem to have DNA binding and hormone binding domains belonging to different classes. These data suggest a complex evolutionary history for nuclear receptor genes in which gene duplication events and swapping between domains of different origins took place.

Key words: gene evolution/nuclear receptors/phylogeny/transcription factors/zinc fingers

Introduction

Nuclear receptors are ligand-activated transcription factors which regulate the expression of target genes by binding to specific *cis*-acting sequences (Evans, 1988; Green and Chambon, 1988; Beato, 1989). This family includes nuclear receptors for several hydrophobic ligands such as steroids, vitamin D, thyroid hormones, retinoic acid and also for dioxin and ecdysone. In addition to these receptors for known ligands, numerous authors have described 'orphan' receptors which are putative receptors for ligands still to be identified (Evans, 1988; Moore, 1990). The family of nuclear receptor genes which to date contains 32 genes has been classically divided into two subfamilies on the basis of structural homologies (Green and Chambon, 1988). One subfamily includes the genes encoding steroid hormone receptors (receptors for glucocorticoids, androgens, oestrogens, mineralocorticoids and progesterone) and the other comprises receptors for thyroid hormone, retinoic acid and vitamin D (Forman and Samuels, 1990) as well as the *ear1* group (the *ear1/Rev-erb* gene and the *Drosophila* E75 gene) (Miyajima *et al.*, 1989; Lazar *et al.*, 1989; Segraves and Hogness, 1990). The various orphan receptors so far described have not been classified among these subfamilies except for the recently discovered PPAR gene which is a close relative of *ear1* (Issemann and Green, 1990). In addition to these two subfamilies, Ryseck *et al.* (1989) have described a growth factor inducible orphan receptor gene named *nur77* (but also known as the NGF1B gene) which could define a third subfamily.

To perform their extremely diverse functions in homeostasis, reproduction, development and differentiation, nuclear hormone receptors link extracellular signals directly to transcriptional responses. All nuclear receptors are composed of at least four domains (Figure 1A) which are differentially conserved between the various products. The A/B domain implicated in transactivation and the hinge (D) domain are poorly conserved between the various subfamilies. The DNA binding (C) domain is composed of two zinc finger structures named CI and CII which play complementary roles in the DNA binding process (Green *et al.*, 1988). The C domain is the most conserved between nuclear receptors. The carboxy terminal E domain is the largest one and has a complex structure (Figure 1A). It is required for hormone binding, dimerization and regulation of transcription. Several authors have suggested that, at least for the thyroid hormone/retinoic acid receptor subfamily, the two extremities of this domain bind the ligand (Forman and Samuels, 1990; Glass *et al.*, 1989). In this model, the central part of the E domain is devoted to dimerization and to the so-called Ti domain implicated in the ligand-dependent regulation of transcription. This 'regulatory zipper' model (Forman and Samuels, 1990) seems suitable for an examination of the extent of the homologies between the various receptor sub-classes in this domain. In fact, the ligand binding part of E domain is nearly identical among receptors of the same binding specificity but does not have any apparent similarity among different receptor types. In contrast, the Ti domain exhibits 20–45% conservation among all receptors (Wang *et al.*, 1989; O'Donnell and Koenig, 1990).

The modular organization of nuclear receptors, the various degrees of conservation between their respective domains and the fact that the domains are encoded by different exons

(see Ponglikitmongkol *et al.*, 1988; Laudet *et al.*, 1991 and references therein) were prerequisites for investigating whether these genes had evolved by successive duplications from a common progenitor or if they had a chimeric structure originating from independent duplications of DNA binding and ligand binding domains from different origins.

In this study, we have taken advantage of the relative conservation of the DNA binding C domain and the Ti and dimerization domains (Ti-DM domain) among nuclear receptors to examine their evolutionary history. The phylogenetic trees derived from our study lead to the conclusion that the three subfamilies of nuclear receptor genes that we define have evolved through a simple duplication model except for some examples such as the *knirps* group or the receptors for ecdysone or vitamin D which may represent two types of 'evolutionary chimera'. Our work also shows that the three subfamilies of nuclear receptor genes emerged at an early stage during evolution.

Results

Alignment of nuclear receptor sequences in C and Ti-DM domains

The nuclear receptor genes included in this study are listed in Table I. To our knowledge this list is complete and all nuclear receptor genes identified to date are studied in this paper.

The C domain sequences were manually aligned (Figure 1B). The C domain is composed of a variable number of amino acids (aa): from 67 for PPAR to 70 for THRs, v-*erbA* and *tailless*. The majority of nuclear receptor genes possess 68 aa in this domain and it is likely that the

Table I. Sequences used in this study

Abbreviation	Name	Species	References
AR	Androgen receptor	Human	Tilley *et al.* (1989)
ARP-1	(Orphan receptor)	Human	Ladias and Karathanasis (1991)
COUP	= *ear3* (orphan receptor)	Human	Wang *et al.* (1989); Miyajima *et al.* (1988)
E75	(Orphan receptor)	*Drosophila*	Segraves and Hogness (1990)
EAR1	= *Rev-ErbA* (orphan receptor)	Human	Miyajima *et al.* (1989); Lazar *et al.* (1989)
EAR2	(Orphan receptor)	Human	Miyajima *et al.* (1988)
ECR	Ecdysone receptor	*Drosophila*	Koelle *et al.* (1991)
EGON	(Orphan receptor)	*Drosophila*	Rothe *et al.* (1989)
ER	Oestrogen receptor	Human	Green *et al.* (1986)
ERR1	(Orphan receptor)	Human	Giguère *et al.* (1988)
ERR2	(Orphan receptor)	Human	Giguère *et al.* (1988)
FTZ-F1	(Orphan receptor)	*Drosophila*	Lavorgna *et al.* (1991)
GR	Glucocorticoid receptor	Human	Hollenberg *et al.* (1985)
H2RIIBP	(Orphan receptor)	Mouse	Hamada *et al.* (1989)
HNF4	(Orphan receptor)	Rat	Sladek *et al.* (1990)
KNI	*knirps* (orphan receptor)	*Drosophila*	Nauber *et al.* (1988)
KNRL	*knirps-related* (orphan receptor)	*Drosophila*	Oro *et al.* (1988)
MR	Mineralocorticoid receptor	Human	Arriza *et al.* (1987)
NGF1B	= *nur77* = *N10* (orphan receptor)	Rat	Milbrandt (1988); Ryseck *et al.* (1989); Hazel *et al.* (1988)
PPAR	(Orphan receptor)	Mouse	Issemann and Green (1990)
PR	Progesterone receptor	Human	Misrahi *et al.* (1987)
RARA	Retinoic acid receptor α	Human	Petkovich *et al.* (1987); Giguère *et al.* (1987)
RARAX	Retinoic acid receptor α	*Xenopus*	Ragsdale *et al.* (1989)
RARB	Retinoic acid receptor β	Human	De Thé *et al.* (1987)
RARG	Retinoic acid receptor γ	Mouse	Zelent *et al.*, 1989
RARGX	Retinoic acid receptor γ	*Xenopus*	Ragsdale *et al.* (1989)
RXR	(Orphan receptor)	Human	Mangelsdorf *et al.* (1990)
SVP	*sevenup* (orphan receptor)	*Drosophila*	Mlodzik *et al.* (1990)
TLL	*tailless* (orphan receptor)	*Drosophila*	Pignoni *et al.* (1990)
TR2	(Orphan receptor)	Human	Chang and Kokontis (1988); Chang *et al.* (1989)
THRA	Thyroid hormone receptor α (= c-*erbA-1*)	Human	Thompson *et al.* (1987); Laudet *et al.* (1991)
THRAC	Thyroid hormone receptor α (= c-*erbA-1*)	Chicken	Sap *et al.* (1986)
THRAXA	Thyroid hormone receptor α (= c-*erbA-1*)	*Xenopus*	Yaoita *et al.* (1990)
THRAXB	Thyroid hormone receptor α (= c-*erbA-1*)	*Xenopus*	Yaoita *et al.* (1990)
THRB	Thyroid hormone receptor β (= c-*erbA-2*)	Human	Weinberger *et al.* (1986)
THRBC	Thyroid hormone receptor β (= c-*erbA-2*)	Chicken	Forrest *et al.* (1990)
THRBXA	Thyroid hormone receptor β (= c-*erbA-2*)	*Xenopus*	Yaoita *et al.* (1990)
THRBXB	Thyroid hormone receptor β (= c-*erbA-2*)	*Xenopus*	Yaoita *et al.* (1990)
USP	*Ultraspiracle* = 2C1-3 = CF1 (orphan receptor)	*Drosophila*	Oro *et al.* (1990); Henrich *et al.* (1990); Shea *et al.* (1990)
VERBA	v-*erbA* gene of avian erythroblastosis virus	Chicken	Debuire *et al.* (1984); Damm *et al.* (1987)
VDR	Vitamin D receptor	Human	Baker *et al.* (1988)

The names (and published synonyms when available) of the gene sequences are indicated, as well as their abbreviations used in the text and in the figures.

ancestral nuclear receptor C domain had the same number of amino acids. As most phylogenetic tree reconstruction procedures necessitate that all compared sequences be of the same size, it was necessary to introduce some gaps in the sequences. On recent structural observation it appears that the gaps introduced in numerous receptor sequences between C1 and C2 are localized in a loop structure (Schwabe and Rhodes, 1991). The cysteine residues of the zinc fingers (vertical arrows in Figure 1B) are absolutely conserved at an invariable position in the whole family, except for the first cysteine of the second finger CII in PPAR (shift of 2 aa downstream) and in the *Drosophila* gene *tailless* (shift of 2 aa upstream). Several other residues are conserved in the C domain, as indicated in the 'consensus' line in Figure 1B. Recently, Forman and Samuels (1990) have classified nuclear receptors into four classes (I−IV in Table II) on the basis of their DNA binding specificity. It is well known that DNA binding specificity is mediated, at least in part, by the three variable residues of a region called the P box which is situated at the end of the first finger. The alignment of Figure 1B gives an opportunity to complete this classification with three additional classes (NC, Table II). Interestingly, the orphan receptors all exhibit closely related P boxes (EGCKS or EGCKG) with the exception of the HNF4/*tailless* tandem and of the FTZ-F1 orphan receptor from *Drosophila*. The THRs and RARs also have such a P box (EGCKG). Indeed, orphan receptors, THRs, RARs, ECR and VDR all bind distantly related sequences, such as AGGTCATGACCT for THRs, RARs and VDR class II genes of Forman and Samuels (1990) or GTGTCaaAGGTCA for orphan receptors COUP/*knirps*/*ear2* class III genes (Green and Chambon, 1988). This is confirmed by several recent reports: for example, the ARP-1 orphan receptor is able to bind to a thyroid hormone responsive element (Ladias and Karathanasis, 1991). All these sequences are formed by partially conserved inverted

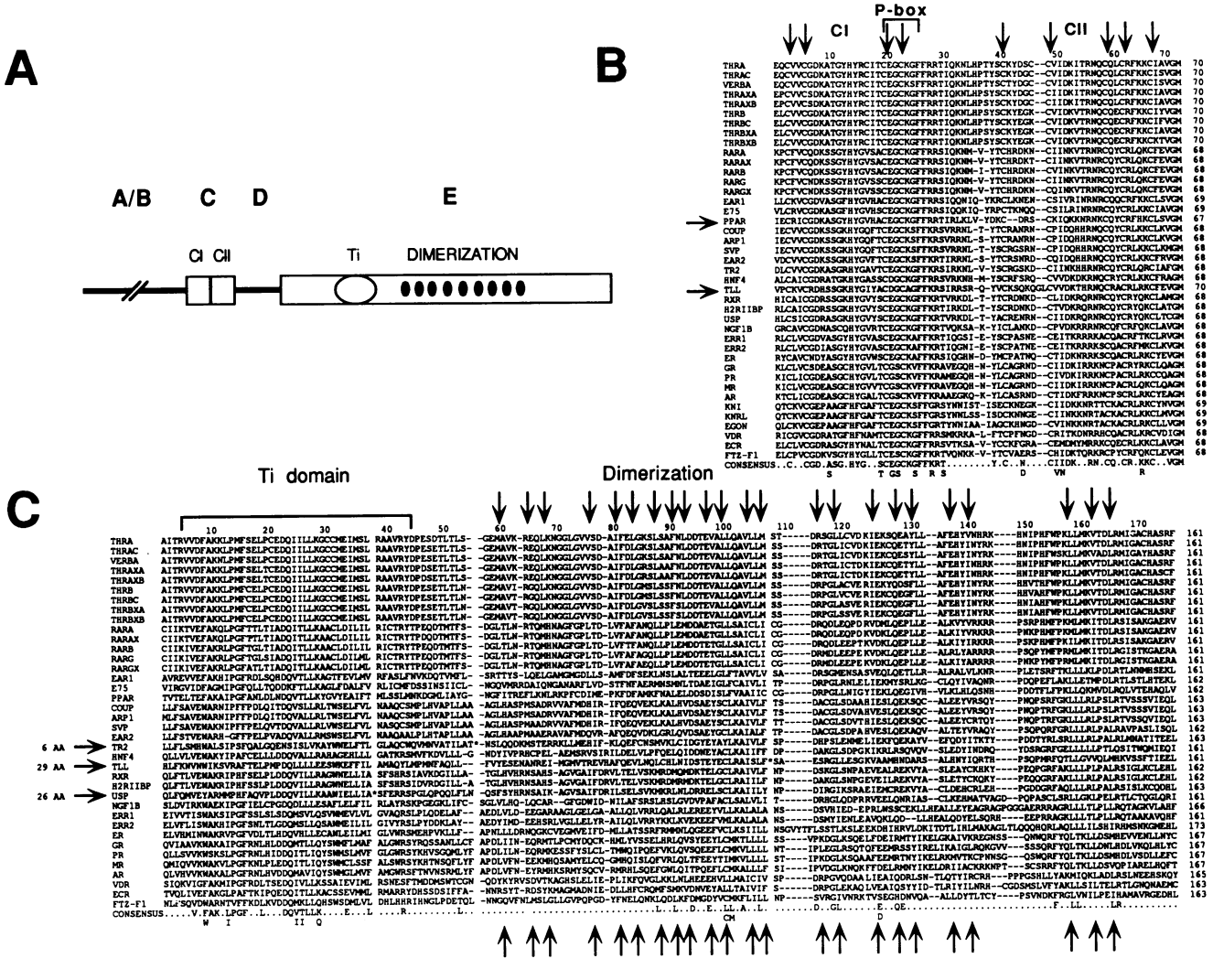

Fig. 1. A: Structure of a nuclear receptor gene product showing the A/B domain of variable size, the DNA binding domain (C), the hinge (D) and the ligand binding domain (E). The two zinc fingers (CI and CII) are indicated as well as the Ti and dimerization leucine-rich domains (Ti-DM) studied in this work. **B**: Alignment of nuclear receptor C domains. Vertical arrows point to the nine conserved cysteines of CI and CII. The region mediating the specificity of the DNA binding, known as the P box, is indicated. The first cysteine of CII in the PPAR and *tailless* genes are not strictly aligned with the others. **C**: Alignment of nuclear receptor Ti-DM domains. Vertical arrows show the hydrophobic amino acids pointed out by Forman and Samuels (1990) and which are presumably implicated in the dimerization process. For TR2, TLL and USP, the star shows the insertion of 6, 29 and 26 aa respectively specific for these three genes (Chang *et al.*, 1989; Pignoni *et al.*, 1990; Oro *et al.*, 1990). The KNI, KNRL and EGON genes show no homology in this domain with other nuclear genes and are not indicated in this panel. For B and C the bottom consensus line represents the strictly conserved amino acids or, more often, the amino acids present in the majority of the sequences for that position.

repeats of 5 bp half sites separated by a variable number of residues. Direct repeats and half sites alone are also possible recognition sites.

The alignment procedure for the Ti-DM domain was much more difficult due to the low similarity scores between the genes (see Materials and methods). We have chosen to favour the conserved hydrophobic residues (vertical arrows in Figure 1C) implicated in the heptad repeats described by Forman and Samuels (1990) and Glass *et al.* (1989). Moreover, we have attributed a high penalty score for gap insertions in the CLUSTAL alignment algorithm (Higgins and Sharp, 1988). This procedure is valid when the level of amino acid identity is low, especially when the sequences compared are from different sub-classes of nuclear receptor genes. For alignment of reverse transcriptase sequences, Xiong and Eickbush (1990) have also used a high gap penalty in a procedure they called 'conserved residues' alignment. As with the Ti-DM domain in nuclear receptor genes, reverse transcriptase sequences have a low similarity score.

The Ti-DM domain alignment presented in Figure 1C shows some well conserved positions, although only two amino acids are strictly conserved between all nuclear receptor genes: an F and a D in positions 15 and 22 of the Ti sub-domain. But even in the DM domain, ~20 residues are conserved in >80% of the genes (see consensus line in Figure 1C). Very often such conserved residues are hydrophobic and likely to be implicated in the dimerization process (Glass *et al.*, 1989).

Several papers have emphasized that in addition to the C domain, several regions of the ligand binding domain of the nuclear receptor genes are conserved. A detailed analysis of such conserved regions appears in Segraves and Hogness (1990) for the E75 gene, in Pignoni *et al.* (1990) for the *tailless* gene and recently in Koelle *et al.* (1991) for the ecdysone receptor gene. In all these cases, the conserved regions lie in the Ti-DM domain as defined here.

Numerous reports implicate this domain in dimerization for the THR/RAR/VDR proteins as well as the *ear1* gene product, such a detailed analysis was recently also done for the steroid hormone receptors (Fawell *et al.*, 1990). Our alignment shows that for these genes, only a fraction of the conserved residues of THR/RAR/VDR are conserved in the steroid hormone receptors. Nevertheless, it is possible to recognize the Ti-DM domain in the steroid receptor gene subfamily. A significant degree of homology exists in that region in all nuclear receptor genes (Pignoni *et al.*, 1990; Segraves and Hogness, 1990).

We were not able to align the *knirps* group of genes in the Ti-DM domains. Neither the Ti nor the DM leucine-rich sub-domains are visible in these molecules and their E domains have no homologies with E domains of other genes.

Generation of a phylogenetic tree for the C domain

Based on the alignment of C domains presented in Figure 1B, the Fitch least square method (Fitch, 1981; Swofford and Olsen, 1990 for review) was used to generate a phylogenetic tree of the 32 nuclear receptor genes, which is shown in Figure 2. This tree shows that the nuclear receptor gene family can be subdivided into three subfamilies: (i) the THR/RAR/*ear1* subfamily; (ii) the orphan receptor subfamily, which contains the COUP, RXR and HNF4/TLL type orphan receptors; and (iii) the steroid

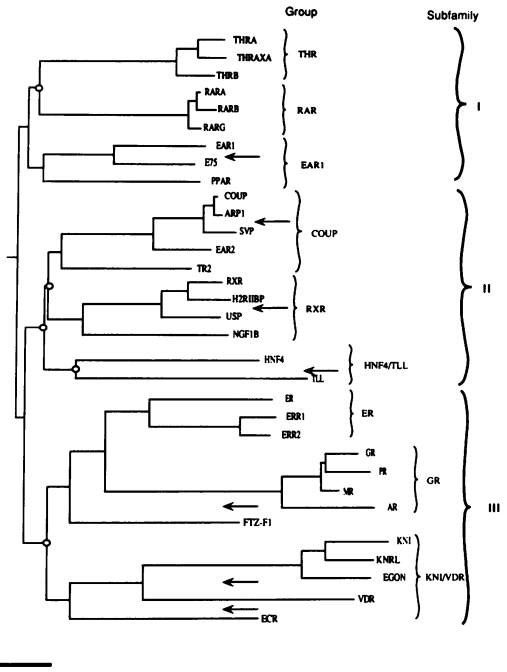

Fig. 2. Rooted phylogenetic Fitch tree for nuclear receptor genes based on C domain sequences. The bar represents a branch length of 10 units. The deepest branch between the internal nodes connecting all the nuclear receptor genes was divided at mid-length, in order to root the tree. Arrows point to the mammalian and *Drosophila* genes which cluster together. Open circles point out the differences between Fitch and NJ trees. Groups and subfamilies are indicated by brackets.

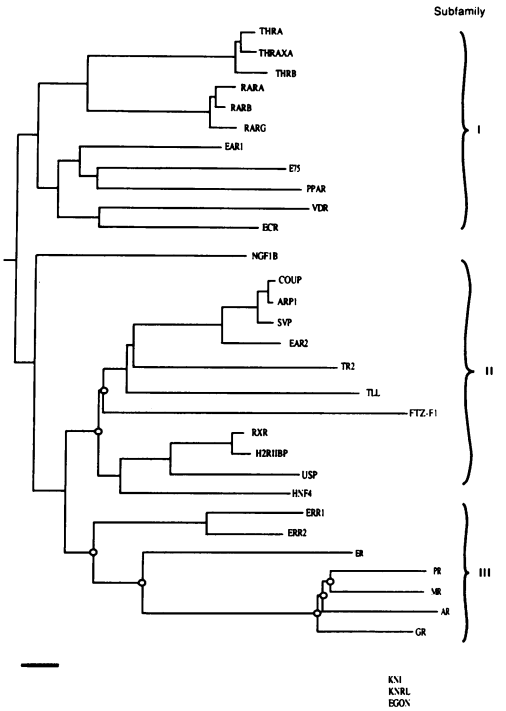

Fig. 3. Rooted phylogenetic Fitch tree for nuclear receptor genes based on Ti-DM domain sequences. The bar represents a branch length of 10 units. The deepest branch between the internal nodes connecting all the nuclear receptor genes was divided at mid-length in order to root the tree. Open circles point out the differences between Fitch and NJ trees. The three main subfamilies are bracketed. The KNI, KNRL and EGON genes which do not have recognizable Ti-DM domain are indicated at the bottom of the figure.

hormone receptor subfamily. All nuclear receptors published to date can be classified into one of these three subfamilies; this observation is in favour of the tree of Figure 2.

This Fitch tree was confirmed by a neighbour-joining (NJ) analysis of the 32 DNA binding domains (Saitou and Nei,1987). The NJ tree we obtained has the same topology as the Fitch tree except for the position of the VDR/*knirps* group, which clusters with the THR group in the NJ tree (Figure 2 and data not shown). In fact, ECR has been described as being close to THRB and to VDR (Koelle *et al.*, 1991). The position of the ECR gene near the THR group in the NJ tree, with very long branches, is poorly confirmed by bootstrap analysis: this association was found in only 18% of bootstrap trials (data not shown). In addition, in our previous tree analysis, done before the publication of ECR, the VDR/*knirps* group was always stably associated with the subfamily III genes as in our present Fitch tree. Thus, the rooting of this group with the THR group may be due to a convergence of ECR toward the THRs (see Discussion). Another difference between the Fitch and NJ trees is the precise position of HNF4 and *tailless*. These two genes form a clade in subfamily II in the Fitch tree (Figure 2) but root together with the COUP group in the NJ tree (data not shown). More information will be required in order to locate these two genes more precisely.

We have arbitrarily rooted this Fitch tree at mid-length of the deepest branch between subfamily I and subfamilies II and III. This order of first gene duplication is tentative, as an initial dichotomy between subfamilies I and II versus III cannot be ruled out (until an adequate outgroup can be included). Until more sequence data on these and additional genes are made available, the precise order of the first and second gene duplications which gave rise to the three ancestors of the nuclear receptor gene subfamilies will remain undetermined.

The finding that the majority of the orphan receptors (exceptions are *ear1*, E75, PPAR, ERR1, ERR2, FTZ-F1 and the *knirps* group) cluster together is surprising because there is no previous report of any particular similarity between these genes. This observation may have important repercussions for the study of the physiological roles of these molecules (see Discussion). Moreover, the Fitch tree in Figure 2 shows that grouping together the nuclear receptors based on similarity within their DNA binding domain sequences correlates with their ligand binding specificity, when this is known: receptors of the THR/RAR subfamily (class I) bind thyroid hormone and retinoic acid respectively and members of the steroid receptor subfamily (class III) bind diverse derivatives of the steroid hormones. This supports the hypothesis of a common evolution of DNA binding and ligand binding domains.

Each of the three subfamilies recognized in Figure 2 may, in turn, be divided into groups: THR, RAR and *ear1* groups for class I receptor genes, COUP, RXR and HNF4/*tailless* groups for orphan class II receptor genes, and ER, GR and VDR/*knirps* groups for class III genes.

In terms of nuclear receptor gene evolution, it is interesting to note that, in each of the three subfamilies described here, mammalian and *Drosophila* genes are grouped together: for example, this is the case for E75 and *ear1* class I genes. Moreover, each of the three groups of genes among the subfamily II orphan receptors has a cluster of human and *Drosophila* genes: *sevenup* for COUP, *ultraspiracle* for RXR and *tailless* for HNF4. This suggests that the three subfamilies of the nuclear receptor genes already existed before the divergence of the arthropod and vertebrate lineages.

A careful examination of Figure 2 leads to several additional observations: the most surprising is that the ecdysone and vitamin D receptors and the *knirps* group [(*knirps*, *knirps-related* and *egon* (Oro *et al.*, 1988; Nauber *et al.*, 1988)] cluster together. This observation was unexpected since there is no report of a close similarity between these genes. The fact that they are grouped together, after a rather long common ancestral branch length (see Figure 2), may be the result of a real ancient, common origin or, less probably, of a convergence phenomenon. In the NJ tree, the VDR and ECR genes also cluster with the *knirps* family (data not shown), confirming our observation based on the Fitch tree analysis, although the position of this whole group in the two trees is different, as mentioned previously.

Generation of a phylogenetic tree for the Ti-DM domain

To compare the evolution of DNA binding and Ti-DM domains we have constructed, using the same methods (i.e. Fitch least squares analysis confirmed by an NJ tree) a tree based on the sequence alignment of the Ti-DM domains. Figure 3 shows the topology of this tree which exhibits roughly the same distribution into three subfamilies as found in the C domain analysis, i.e. THRs/RARs (subfamily I), orphan receptors (II) and steroid hormone receptors (III). Thus, the three subfamilies we have defined with the C domain tree are also valid in the Ti-DM domain derived tree. It is interesting to note that the same topology is obtained when we construct a Fitch or an NJ tree based on the C and Ti-DM sequences treated together (data not shown).

Nevertheless, it is obvious from Figure 3 that some important differences exist between the two molecular phylogenies. The first point is that three genes cannot be included in the phylogenetic analysis of the other nuclear receptors: these are the three genes of the *Drosophila knirps* group (*knirps*, *knirps-related* and *egon*: reviewed in Pankratz and Jäckle, 1990). These genes have acquired completely new E domains that are unrelated to the Ti-DM domain of the other nuclear receptors. This observation fits well with the data of several authors who have described these genes (Oro *et al.*, 1988; Nauber *et al.*, 1988; Rothe *et al.*, 1989).

The other major difference concerns the VDR, ECR, FTZ-F1, NGF1B, *tailless* and HNF4 genes. In the C domain Fitch tree (Figure 2), VDR and ECR are associated with the *knirps* family in subfamily III, but in the Ti-DM domain tree (Figure 3) they appears to belong to subfamily I. Numerous data suggest that the VDR gene product has a physiological behaviour resembling that of the THR/RAR gene products (reviewed in Forman and Samuels, 1990). The fact that the VDR and the ECR genes belong to different subfamilies according to C and Ti-DM domain trees was confirmed by an NJ tree based on Ti-DM sequences (data not shown). This point will be developed further in the Discussion.

The FTZ-F1 gene belongs to subfamily III for the C domain, whereas in the Fitch tree constructed from the Ti-DM domain it belongs to the COUP group within the subfamily II with a very long branch (Figure 3). This position is not confirmed by the NJ tree of Ti-DM domain

where this gene clusters with all the subfamily II genes. Thus, although clearly differing in C and Ti-DM domain trees, the precise location of FTZ-F1 needs further work. The NGF1B gene also shows a different partition between C and Ti-DM domain trees. In the C domain tree, NGF1B belongs to the subfamily II orphan receptor genes in close association with the group of the retinoid responsive RXR genes. Surprisingly, in the Ti-DM domain-derived tree, NGF1B cannot be assigned to a particular subfamily since it branches early before the divergence of subfamilies II and III (see Figure 3). The *tailless* and HNF4 genes show a slightly different situation: these two genes cluster differently according to the tree construction procedure (NJ or Fitch) for C domain. For Ti-DM domains they are always separated: the *tailless* gene groups with the COUP group and the HNF4 gene clusters with the RXR group.

Finally, the comparison of the two trees shows a few additional, more subtle differences. In the C domain of subfamily III, the GR and PR genes form a rather young sister group and are then joined by MR and AR respectively, whereas for the Ti-DM domain, the PR and MR genes first cluster together and are then associated with AR and GR. A note of caution should be made since the internodal, ancestral, segments are very short (Figure 3), and the relationships between PR, MR and GR cannot be resolved unambiguously (trichotomy). The same type of discrepancies are seen within the *ear1* and ER groups of genes. Again, trichotomy events cannot be ruled out. These differences may possibly be attributed to different rates of evolution among and between the two domains involved in this comparison: for example, the C domain, which is smaller than the Ti-DM domain, is also much more conserved between the nuclear receptor genes.

Evolution of the THR/RAR gene complex

Figure 4 presents a Fitch tree obtained for the Ti-DM domains of all representatives of THR/RAR subgroups of genes. This is of interest since it is known that the genes encoding THRs and RARs have similar chromosomal locations (namely chromosome 17q21 for THRA and RARA and chromosome 3p24 for THRB and RARB), which suggests a common evolution for these genes. Furthermore, several authors have emphasized that the human and rat THRA genes partially overlap with the orphan receptor *ear1* gene (Lazar *et al.*, 1989; Miyajima *et al.*, 1989; Laudet *et al.*, 1991). Finally, the evolution of this gene subfamily is worth studying because the v-*erbA* oncogene represents a virally transduced and modified version of THRA which has evolved independently from its cellular progenitor (Debuire *et al.*, 1984).

The tree in Figure 4 confirms the pattern observed in the preceding trees (Figure 3), i.e. in the class I subfamily the individual genes of the *ear1* subgroup, namely ECR, VDR, *ear1*, PPAR and E75, have diverged at a very early stage, approximately at the same time as the THR and RAR subgroups. For the THR group, several additional observations can be made: first, the v-*erbA* oncogene, after its presumed divergence from the chicken THRA, has evolved more quickly than the other genes of its group. This was expected since this gene, transduced by the avian erythroblastosis virus (AEV), has been subjected to a high rate of replication by reverse transcriptase, which is known to have a very low fidelity. Curiously, the chicken THRA

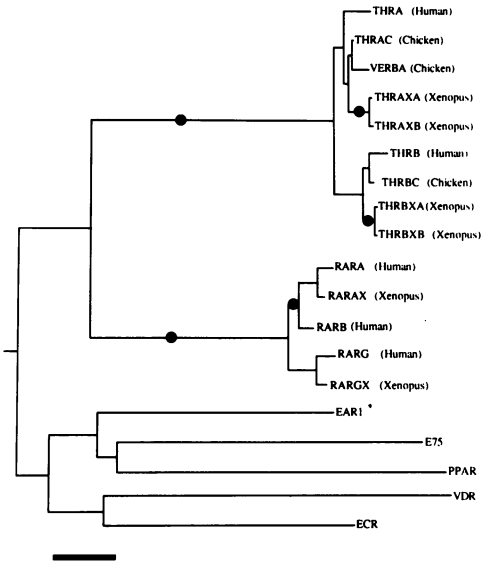

Fig. 4. Fitch tree for all sequences of subfamily I for Ti-DM domain, which shows the emergence of the various THRs and RARs genes. The bar represents a branch length of 10 units. For THR and RAR groups, when duplications appear unambiguous, they are marked by dots along ancestral branches.

gene does not root with the human THRA gene, but clusters with the *Xenopus* genes. Nevertheless, internodal segments defining the different clusters in this subgroup are very short, and a polytomy such as [(THRA, THRAC, v-*erbA*), (THRAXA, THRAXB)] cannot be ruled out. Secondly, it is worth emphasizing the relationships of the four THR genes described in *Xenopus*. These genes have, according to Figure 4, been duplicated rather recently in the *Xenopus* lineage; this recency represents a specific difference between *Xenopus* and other vertebrates. Similar gene duplication phenomena in *Xenopus* have been observed for many other genes (tetraploidization, see Kobel and Du Pasquier, 1986 and Bisbee *et al.*, 1977).

Discussion

The three nuclear receptor subfamilies

Our study allows classification of the nuclear receptor genes into three subfamilies: subfamily I joins together the THR, RAR and *ear1* subgroups of genes, subfamily II groups most of the orphan receptor genes and subfamily III contains the steroid hormone receptor genes and the *knirps* group. This observation, based on an evolutionary tree of C domain sequences, can be substantiated by structural data available on nuclear receptor genes (Baker *et al.*, 1988; Hughes *et al.*, 1988; Laudet *et al.*, 1991; Lehmann *et al.*, 1991; Miyajima *et al.*, 1989; Oro *et al.*, 1990; Ponglikitmongkol *et al.*, 1988; Rothe *et al.*, 1989; Ryseck *et al.*, 1989; Yaoita *et al.*, 1990; see also Green and Chambon, 1988 for review). It is well known that for all vertebrate nuclear receptor genes, except NGF1B, the A/B domain and each of the zinc fingers CI and CII are encoded by separate exons. Furthermore, the position of the intron lying between these two exons variable among the sub-classes; for the THR/RAR/*ear1* subfamily, this intron (number 4 for THRA; Laudet *et al.*, 1991) is located one amino acid after the last cysteine of CI, whereas for the steroid hormone receptor genes, this intron (number 2 for ER; Ponglikitmongkol *et al.*, 1988)

is situated 10 aa after this last cysteine of CI. The NGF1B orphan receptor gene is the only known exception to this observation since its CI finger is encoded by an exon which also encodes the A/B domain. Moreover, the intron separating CI and CII lies between the two positions known for THR/RAR and steroid receptors i.e. 5 aa after the last cysteine of CI (Ryseck *et al.*, 1989). Consequently, NGF1B—based on structural features—seems to define on its own a third subfamily of nuclear receptor genes. How do our data fit with these structural observations? The C domain tree does show clearly that orphan receptor genes, including NGF1B, cluster together (except for the *knirps*, *earl* and ERR groups). This interesting observation strongly supports the proposal that the nuclear receptor genes can be organized into three subfamilies as suggested by Ryseck *et al.* (1989) based on the study of only one gene (NGF1B). The members of subfamilies I and III (as defined by us, i.e. containing the subgroups THR, RAR, *earl* and steroid receptors including VDR respectively) have the same exon/intron structures within their own subfamily. Unfortunately, the *Drosophila* ECR gene has no intron between its two zinc fingers (Koelle *et al.*, 1991). For the vertebrate subfamily II genes, only the genomic structure of NGF1B is known and it is tempting to speculate that other orphan receptor genes may have the same structure in their C domain. To our knowledge, the only other gene of subfamily II whose genomic organization has been determined is the *ultraspiracle* gene, which has no introns. But it is well known that intron position and number are often more variable in *Drosophila* genes than in their vertebrate homologues. It is possible that the *ultraspiracle* gene has lost all its introns during the evolution of *Drosophila*. The only structural information which does not fit with our C domain tree also comes from *Drosophila* genes: the *knirps* group have the same exon−intron position between the two fingers as the THR/RAR/*earl* genes, although they cluster with class III steroid hormone receptors. Nevertheless, the *knirps* genes have only one intron and, again, this observation must be tempered by the fact that the *knirps* genes are *Drosophila* nuclear receptors with a very unusual evolutionary history (see below). Of course, there are other possible explanations for the actual intron positions in the various genes of the family, and especially in *Drosophila*. For example, some organisms such as *Drosophila* could have gained introns during evolution and these introns could have been inserted at putative 'proto-splice' sites as in the case of the tubulin and actin genes studied by Dibb and Newman (1989). The description of other genes from 'primitive' organisms in the future should enable us to test such a hypothesis.

The fact that the orphan receptor genes (except for the ERR, *earl* and *knirps* groups of genes) are all grouped together to form subfamily II strongly suggests that these genes have a common ancestor. The main problem to be solved regarding orphan receptor function is whether or not they bind a ligand and, if so, what is the identity of these ligands. Moore (1990) has hypothesized that terpenoids may be ligands of orphan receptors. *Drosophila* juvenile hormone may be such a ligand and the plant hormones gibberellic and abscissic acids are other examples of ligands of receptors yet to be found. Another possibility is that orphan receptors do not have ligands at all, but act as hormone-independent transcriptional regulators. These molecules, which all have very similar DNA binding specificities (see Table II), may bind to the same DNA sequences as the canonical nuclear receptors. As an illustration of this hypothesis, the THR/RAR proteins have the same DNA binding sequence (P box) as the majority of subfamily II orphan receptors. The fact that the ARP-1 orphan receptor, a very close relative of COUP, can bind the thyroid hormone responsive elements (T3REs) suggests that the majority of the orphan receptors, which possess the same P box, could also bind T3REs. As all subfamily II orphan receptors seem to have diverged from a common ancestor (Figures 2 and 3), it is tempting to speculate that these orphan receptors have diverged from THR/RAR-like ancestral genes to become hormone independent regulators of THR/RAR action by competing, dimerizing or synergizing with them. Indeed, numerous proteins are known to regulate the action of THRs and RARs (O'Donnell and Koenig, 1990). For example, the recently identified thyroid hormone receptor auxiliary protein (TRAP) (Beebe *et al.*, 1991; O'Donnell *et al.*, 1991; Darling *et al.*, 1991), which is able to bind a T3RE and to dimerize with THRs via the Ti-DM domain and the CII zinc finger, may indeed be an orphan receptor. The answer to this question awaits the cloning of the gene encoding TRAP. We believe that the common origin of members of subfamily II orphan receptors argues for this model and permits us to hypothesize a complex regulation of the action of T3 and retinoids.

Table II. Classification of nuclear receptors for their DNA binding specificity

Sequence of the P box	Class	Genes
GSCKV	I	AR, GR, MR, PR
EGCKG	II	E75, EAR1, ECR, RXR, H2RIIBP, USP, TR2, VDR, NGF1B, PPAR, RARA, RARB, RARG, THRA, THRB
EGCKS	III	ARP-1, COUP, EAR2, EGON, KNI, KNRL, SVP, VERBA
EGCKA	IV	ER, ERR1, ERR2
ESCKG	NC	FTZ-F1
DGCKG	NC	HNF4
DGCAG	NC	TLL

Classification of nuclear receptors according to their DNA binding specificity (sequence of the P box). This classification is based on that of Forman and Samuels (1990). Receptors which were not classified by these authors (namely HNF4, TLL and FTZ-F1) are indicated as NC (new class).

Comparison of C domain and Ti-DM domain trees

The comparison we have made between the C and Ti-DM domain trees (Figures 2 and 3) confirms the notion that the major mode of evolution of nuclear receptor genes has been by gene duplication. Nevertheless, we have identified some nuclear receptor genes which may have had a more complex history.

The first example is the *knirps* family (*Drosophila knirps*, *knirps-related* and *egon* genes). These genes lack the Ti-DM domain which is replaced by a unrelated domain. It is postulated that these genes arose from an incomplete duplication of an ancestor gene. This may result in nuclear receptor-like proteins which have no dimerization and

possibly no ligand binding motifs and which may act as classical transcription factors as suggested by others (Nauber *et al.*, 1988; Oro *et al.*, 1988; Rothe *et al.*, 1989). Alternatively, it is possible that these genes encode nuclear receptors for ligands that are chemically unrelated to steroids, retinoids and thyroid hormones. This hypothesis might explain the observation that their E domains are clearly different from other nuclear receptor genes.

The second example of an independent evolution between C and Ti-DM domains comes from genes such as VDR, ECR, FTZ-F1, NGF1B, *tailless* and HNF4, which belong to one subfamily when the C domain is considered and to another based on the Ti-DM domain analysis. The most dramatic examples of this are the VDR and ECR genes which were classified into subfamily III based on their C domains and into subfamily I based on their Ti-DM domains. This latter fact is not surprising and was already mentioned by others (see Forman and Samuels, 1990 and references therein). Furthermore, bootstrap analyses on the monophyletic character of the various groups which contain VDR and ECR give reliable results except, as previously mentioned, for the rooting of the VDR/*knirps* group with THR genes in the C domain NJ tree (data not shown). The placement of VDR in subfamily III from its C domain is confirmed by gene structure analysis since this gene has an exon−intron boundary between CI and CII that is diagnostic of the steroid hormone receptor from subfamily III. It is not surprising that the DNA binding specificity of VDR and ECR is like that of THRs and RARs (i.e. subfamily I members) because the specificity of DNA binding is encoded only by a very small number of amino acids (P box, see Table II). Thus, one can hypothesize an alteration (or 'correction') of binding specificity gained from a small number of mutations within the C domain of type III which has consequently been switched towards a subfamily I type physiological behavior. This gives a large evolutionary flexibility to the nuclear receptor superfamily of genes. It has to be noted, however, that our C domain tree is based on only 70 aa which have kept a high degree of homology. Phylogenetic trees based on this short domain are very sensitive to local phenomena of evolutionary convergence. This may explain the NJ tree pattern observed for the VDR/*knirps* group which may be driven to the THR group by the homology of ECR with the THR genes. It is important to note that our conclusion on the position of these genes in the C domain trees should be considered, for the moment, as tentative. Nevertheless, the fact that structural and phylogenetical observations fit together suggests that the ECR and VDR genes indeed had a complex evolutionary history. One can hypothesize that an incomplete gene duplication event, or a homologous recombination-like event, may explain the appearance of such nuclear receptor chimeras with C and Ti-DM domains arising from different ancestors.

The lack of structural and functional characterization of FTZ-F1, NGF1B and *tailless*/HNF4 limits the discussion of our observations. It should be noted, however, that the NGF1B gene was previously described as an exception since its C domain exon−intron structure is unique (Ryseck *et al.*, 1989). It is possible that the NGF1B gene represents an exceptional chimera among nuclear receptor genes; alternatively, for an unknown reason, its C and Ti-DM domains may have been subjected to very different rates of substitution. For the *tailless* and HNF4 genes, we strongly favour the hypothesis that they belong to different groups in the subfamily II of orphan receptors (COUP group for *tailless* and RXR group for HNF4). The fact that they cluster together in subfamily II in the C domain tree (Figure 2) might be the result of an evolutionary convergence between the C domains of these genes. Indeed for this domain, the NJ and Fitch trees do not group these genes in the same manner, suggesting that too much homoplasy prevents us from placing these genes correctly. Another possibility which cannot be excluded is that, again, these genes represent two types of chimera. More information, such as physiological data, will be necessary in order to infer the precise phylogenetic relationship between these genes. It is important to make clear that the C domain tree is based on rather short and highly conserved sequences and so is much more unstable than Ti-DM domain tree. The precise position of some deeply rooted sequences such as HNF4, *tailless* and FTZ-F1 may be poorly resolved by this tree.

In conclusion, our analysis suggests that in addition to canonical gene duplication, other events (recombination, translocation or exon shuffling) may have arisen during the evolution of the ancestral nuclear receptor genes in order to yield the complex gene family presently known. Some genes such as *knirps*, ECR, VDR, FTZ-F1 and NGF1B may represent different sorts of evolutionary chimera, increasing the sensitivity of the various hormone response pathways.

A hypothesis for the evolution of the nuclear receptor gene superfamily

It is obvious from our work that the different subfamilies of nuclear receptors appeared before the divergence of the arthropod and vertebrate lineages which occurred at least 500 million years ago (Hartland *et al.*, 1982). Indeed, numerous vertebrate nuclear receptor genes have a *Drosophila* homologue belonging to their subgroup clade, suggesting that all the preceding divergences took place before the separation of the arthropods and vertebrates. This is the case for the *ear1* group and for all the subfamily II orphan receptor groups (RXR, COUP and HNF4). From this observation alone, it is unclear whether THRs/RARs and steroid receptors co-existed, as they presently do, before the arthropod/vertebrate divergence. However, the recent cloning of the *Drosophila* ecdysone receptor showed that canonical nuclear receptors are functional in this fly (Koelle *et al.*, 1991). In *Drosophila*, other molecules which are potential ligands for nuclear receptors have been described (Segraves, 1991). Recently, several authors have hypothesized that the *Drosophila* E75 gene may be the juvenile hormone receptor (Ashburner, 1990; Segraves, 1991). It is clear that some of the *Drosophila* orphan receptors are probably receptors for molecules known to mediate numerous physiological responses in the fly. So we can conclude that the function of nuclear receptors (i.e transduction of cell−cell messages via small hydrophobic molecules) already existed before the divergence of the arthropod and vertebrate lineages.

Figure 5 presents an evolutionary hypothesis for the possible evolution of nuclear receptor genes, in which putative switches and incomplete duplication events are indicated. It is obvious from this figure that the nuclear receptor superfamily diverged (mainly by duplication) early on, and has since been well conserved with the exception of some minor events. This raises the intriguing question

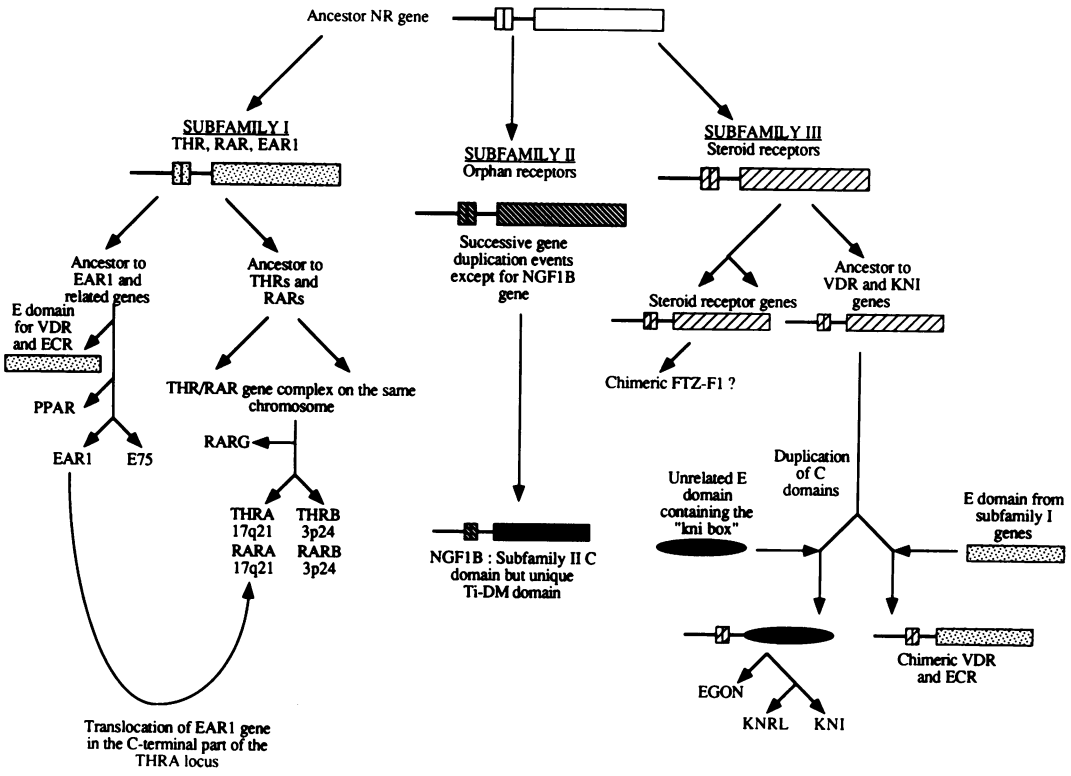

Fig. 5. An evolutionary scenario for the nuclear receptor genes, in which only important events and features are indicated. Each subfamily of nuclear receptor genes is represented by a different shading. The VDR and ECR genes may be formed with a subfamily III C domain and a subfamily I Ti-DM domain. The NGF1B gene has a subfamily II C domain and a Ti-DM domain which cannot be classified between subfamilies II and III. The FTZ-F1 gene is perhaps also an example of chimera. The *knirps* group of genes have an E domain unrelated to other nuclear genes but all the *knirps* genes have the same 'kni box' situated just after the second zinc finger CII. The order of duplication of THRs, RARs and EAR1 genes with VDR and ECR is taken from Ti-DM domain tree. The translocation event which has placed the EAR1 gene in the THRA locus is indicated: this event seems to have arisen just before the appearance of mammalian lineage (see text for details).

of the origin of this gene superfamily. O'Malley (1989) and Moore (1990) have speculated that nuclear receptor genes originated very early on by fusion of DNA binding sequences and genes implicated in steroid binding in the cytosol recently described in yeast (see Moore, 1990 for references). Nuclear receptor genes are not yet known to exist in organisms other than vertebrates and insects. However, it is obvious that, at least for worms, such genes do exist, as the identification of nuclear receptor genes has been announced recently in the nematode *Caenorhabditis elegans* (cited in Moore, 1990) and as ecdysteroids are known to trigger physiological responses in the helminth parasite *Schistosoma* (Nirdé *et al.*, 1983). Potential nuclear receptors found in primitive organisms are the receptor for the DIF factor, a hydrophobic morphogen, of the slime mould *Dictyostelium* which was recently described as a nuclear protein (Insall and Kay, 1990) and the receptor for the steroid hormone antheridiol from the filamentous fungus *Achlya ambisexualis* which like the vertebrate steroid receptors is able to bind heat shock proteins (Brunt *et al.*, 1990). Our data indicate that most of the evolution of this family was nearly complete 500 million years ago, i.e. at the arthropod/vertebrate split. We can imagine that there was a rapid evolution of these genes and then, when the different classes of receptors were fixed in their present function (i.e. transduction of a signal from other cells), only minor rearrangements occurred leading to an overall strong gene conservation. Numerous reports have recently emphazised the speed of early metazoan evolution in the early Cambrian (for review see Valentine, 1977). From our data it is tempting to speculate that, as nuclear receptor genes are implicated in cell−cell communication, a crucial need for metazoans, the rapid divergence of early nuclear receptor genes coincides with this 'burst' of metazoan evolution [see Erwin (1991) for a review].

The evolution of the THRs/RARs/*ear1* subfamily presented in Figure 5 needs further comments: THRA and RARA are situated on the same chromosome (17q21), and THRB and RARB are on the chromosomal segment, 3p24. Furthermore, the *ear1* gene partially overlaps with the mammalian THRA gene (Miyajima *et al.*, 1989; Lazar *et al.*, 1989; Laudet *et al.*, 1991). It was then of interest to study the history of that group of genes carefully . From Figure 4, it is clear that the *ear1* group diverged early from the THR/RAR group, before the arthropod/vertebrate divergence, since the *ear1* gene has a *Drosophila* homologue, E75, clustering with it. Then, the undifferentiated THR/RAR ancestor diverged to give THR and RAR progenitors likely to be on the same chromosome (probably an ancestor of chromosome 3 or 17). Except for the RARG gene, whose chromosomal location is still unknown, it is clear that this THR/RAR gene complex was finally duplicated to give the four present genes, which were already present in early vertebrates since they exist in *Xenopus* (Yaoita *et al.*, 1990; Ragsdale *et al.*, 1989). The overlap which exists between THRA and *ear1* genes seems to be a recent event which may be the result of a translocation which has placed *ear1* in the same locus as THRA in a tail-to-tail orientation. This event is likely to have created the THRA final exon (exon 10 for

the human THRA; Laudet *et al.*, 1991). It probably took place during the appearance of the mammals since the overlap structure (which is strictly conserved between rat and human) as well as the alternative final exon of THRA have not been detected in the chicken (Forrest *et al.*, 1990). It is then tempting to assume that the *ear1* gene was translocated in the THRA locus during the very first moments of mammalian evolution, giving rise to the complex gene tandem array we presently observe on human chromosome 17. Another possibility would be that the two genes were always linked in the same chromosome in a head-to-tail orientation and that an inversion of one gene could explain the present situation. It is interesting to note that in this case, an *ear-1* like gene would probably be found near the THRB locus since the duplication of the THR ancestor gene took place after the *ear-1*/THR separation. Finally, we have to note that, in our NJ tree for C domain, the VDR/*knirps* group containing ECR clusters with THR. Although we favour the Fitch tree pattern, we cannot definitively exclude an early relationship between at least some of these genes.

Although it is very difficult to speculate on the origin of nuclear receptor genes because of the lack of some data, our work gives some clues for understanding the evolution of the nuclear receptor family. The different events which gave rise to the present three subfamilies may reasonably be positioned within the evolutionary tree and, eventually, approximately dated. Additional work, including more sequence comparisons, genomic organization determination and studies of nuclear receptors from more 'primitive' organisms are needed to delineate more precisely this fascinating history.

Materials and methods

Sequence sources

Sequences used for this study are shown in Table I. For each gene the human or rodent sequence was used when available. We have checked that the introduction of various mammalian, and even avian, versions of these genes does not change the topology of the trees (data not shown). For the THR/RAR groups of genes, the chicken and *Xenopus* sequences as well as the v-*erbA* gene of avian erythroblastosis virus were used in order to survey the evolution of these genes among vertebrate lineages. All *Drosophila* sequences were used, even those that are closely related to mammalian nuclear receptor genes (such as E75 and *ear1*).

Sequence alignment

The procedures for the alignment were different for the C domain and the Ti-DM domain.

For the C domain the high homology score between sequences renders manual alignment possible and rapid. Gaps were introduced in the sequences at the same place just before the beginning of the second zinc finger CII. This is the region with the lowest homologies between receptors in the DNA binding domain. This alignment procedure was mainly confirmed by a computerized alignment using the CLUSTAL package (Higgins and Sharp, 1988).

For the Ti-DM domain the alignment of the sequences is much more tedious because of the low homology between sequences. Delimitation of this domain was made using the recently published alignment of the THR/RAR subfamily genes E domain (Forman and Samuels, 1990). The sequences were then used to conduct a computer alignment procedure using the CLUSTAL package available on the CITI-2/Bisance network (Dessen *et al.*, 1990). Because of the low sequence similarities in the dimerization domains we have chosen to confer a substantial penalty for the insertion of gaps. As some structural information is available on the importance of hydrophobic heptad repeats in this DM region (Forman and Samuels, 1990; Glass *et al.*, 1989) we have favoured the alignment of these heptad amino acids when it was possible.

Construction of phylogenetic trees

Our method was mainly identical to the one used by Xiong and Eickbusch (1988, 1990) for the construction of trees of reverse transcriptase sequences. The percent divergence values for all pairwise comparisons of the aligned sequences were calculated by dividing the number of different residues by the total number of residues compared. Gaps were treated as mismatches. Before tree construction all values were transformed into distances (d) with Poisson correction $d = \ln (1 - S)$ where S is the proportion of sites that differ (Nei, 1987).

These values were then used to construct phylogenetic trees by the Fitch least squares method (Fitch, 1981). In parallel to the Fitch algorithm, we used the neigbour-joining (NJ) method of Saitou and Nei (1987) which gave largely identical results (data not shown). We prefer NJ over UPGMA (Sneath and Sokal, 1973) because the UPGMA method assumes an equal rate of change along all sequences (Saitou and Nei, 1987; Swofford and Olsen, 1990), an assumption which might not hold for nuclear receptor genes. Bootstrap analyses were performed on NJ trees with the CLUSTAL V package available on the CITI-2/Bisance network (Dessen *et al.*, 1990).

Acknowledgements

We thank Philip Boucher, Jean Daniel Graf, Ray Pierce, Anne Bonnieu, Olivier Albagli, Veronique Fafeur and many colleagues at Lille for critical reading of the manuscript. We are also grateful to Nicole Devassine and Marie Christine Bouchez for patient typing. We thank Claire Valencien from the CITI-2/Bisance network for computer help. V.L. holds a fellowship from the French Ministère de la Recherche et de la Technologie. We thank the Association pour la Recherche contre le Cancer, the Centre National de la Recherche Scientifique, the Institut National de la Santé et de la Recherche Médicale and the Institut Pasteur de Lille for financial support.

References

Arriza,J.L., Weinberger,C., Cerelli,G., Glaser,T.M., Handelin,B.L., Housman,D.E. and Evans,R.M. (1987) *Science*, **237**, 268–275.
Ashburner,M. (1990) *Cell*, **61**, 1–3.
Baker,A.R., McDonnel,D.P., Hughes,M., Crisp,T.M., Mangelsdorf,D.J., Haussler,M.R., Pike,J.W., Shine,J. and O'Malley,B.W. (1988) *Proc. Natl. Acad. Sci. USA*, **85**, 3294–3298.
Beato, M. (1989) *Cell*, **56**, 325–344.
Beebe,J.S., Darling,D.S. and Chin,W.W. (1991) *Mol. Endocrinol.*, **5**, 85–93.
Bisbee,C.A., Baker,M.A., Wilson,A.C., Hadji-Azimi,I. and Fischberg,M. (1977) *Science*, **195**, 785–787.
Brunt,S.A., Riehl,R. and Silver,J.C. (1990) *Mol. Cell. Biol.*,**10**, 273–281.
Chang,C. and Kokontis,J. (1988) *Biomed. Biophys. Res. Commun.* **155**, 971–977.
Chang,C., Kokontis,J., Acakpo-Satchivi,L., Liao,S., Takeda,H. and Chang,Y. (1989) *Biomed. Biophys. Res. Commun.*, **165**, 735–741.
Damm,K., Beug,H., Graf,T. and Vennström,B. (1987) *EMBO J.*, **6**, 375–382.
Darling,D.S., Beebe,J.S., Burnside,J., Winslow,E.R. and Chin,W.W. (1991) *Mol. Endocrinol.*, **5**, 73–84.
De Thé,H., Marchio,A., Tiollais,P. and Dejean,A. (1987) *Nature*, **330**, 667–670.
Debuire,B., Henry,C., Benaissa,M., Biserte,G., Claverie,J.M., Saule,S., Martin,P. and Stéhelin,D. (1984) *Science* **224**, 1456–1459.
Dessen,P., Fondrat,C., Valencien,C. and Mugnier,C. (1990) *Comput. Appl. Biosci.*, **6**, 355–356.
Dibb,N.J. and Newman,A.J. (1989) *EMBO J.*, **8**, 2015–2021.
Erwin,D.H. (1991) *Trends Ecol. Evol.*, **6**, 131–134.
Evans,R.M. (1988) *Science*, **240**, 889–895.
Fawell,S.E., Lees,J.A., White,R. and Parker,M.G. (1990) *Cell*, **60**, 953–962.
Fitch,W.M. (1981) *J. Mol. Evol.*, **18**, 30–37.
Forman,B.M. and Samuels,H.H. (1990) *Mol. Endocrinol.* **4**, 1293–1300.
Forrest,D., Sjöberg,M. and Vennström,B. (1990) *EMBO J.*, **9**, 1519–1528.
Giguère,V., Yang,N., Segui,P. and Evans,R.M. (1988) *Nature*, **331**, 91–94.
Glass,C.K., Lipkin,S.M., Devary,O.V. and Rosenfeld,M.G. (1989) *Cell*, **59**, 697–708.
Green,S. and Chambon,P. (1988) *Trends Genet.*, **4**, 309–314.
Green,S., Walter,P., Kumar,V., Krust,A., Bornert,J.M., Argos,P. and Chambon,P. (1986) *Nature*, **320**, 134–139.

Green,S., Kumar,V., Theulaz,I., Wahli,W. and Chambon,P. (1988) *EMBO J.*, **7**, 3037–3044.
Hamada,K., Gleason,S.L., Levi,B.Z., Hirschfeld,S., Appella,E. and Ozato,K. (1989) *Proc. Natl. Acad. Sci. USA*, **86**, 8289–8293.
Hartland,W.B., Cox,A.V., Llewellyn,P.G., Pickton,C.A.G., Smith,A.G. and Walters,R. (1982) *A Geologic Time Scale*. Cambridge University Press, Cambridge.
Hazel,T.G., Nathan,D. and Lau,L.F. (1988) *Proc. Natl. Acad. Sci. USA*, **85**, 8444–8448.
Henrich,V.C., Sliter,T.J., Lubahn,D.B., Macintyre,A. and Gilbert,L.I. (1990) *Nucleic Acids Res.*, **18**, 4143–4148.
Higgins,D.G. and Sharp,P.M. (1988) *Gene*, **73**, 237–244.
Hollenberg,S.M., Weinberger,C., Ong,E.S., Cerelli,G., Oro,A., Lebo,R., Thompson,E.B., Rosenfeld,M.G. and Evans,R.M. (1985) *Nature* **318**, 635–641.
Hughes,M.R., Malloy,P.J., Kieback,D.G., Kesterson,R.A., Pike,J.W., Feldman,D. and O'Malley,B.W. (1988) *Science*, **242**, 1702–1705.
Insall, R. and Kay,R.R. (1990) *EMBO J.*, **9**, 3323–3328.
Issemann,I. and Green,S. (1990) *Nature*, **347**, 645–650.
Kobel,H.R. and Du Pasquier,L. (1986) *Trends Genet.*, **2**, 310–315.
Koelle,M.R., Talbot,W.S., Segraves,W.A., Bender,M.T., Cherbas,P. and Hogness,D.S. (1991) *Cell*, **67**, 59–77.
Ladias,J.A.A. and Karathanasis,S.K. (1991) *Science*, **251**, 561–565.
Laudet,V., Bègue,A., Henry-Duthoit,C., Joubel,A., Martin,P., Stéhelin,D. and Saule,S. (1991) *Nucleic Acids Res.*, **5**: 1105–1112.
Lavorgna,G., Ueda,H., Clos,J. and Wu,C. (1991) *Science*, **252**, 848–851.
Lazar,M.A., Hodin,R.A., Darling,D.S.and Chin,W.W. (1989) *Mol. Cell. Biol.*, **9**, 1128–1136.
Lehmann,J.M., Hoffmann,B. and Pfahl,M. (1991) *Nucleic Acids. Res.*, **19**, 573–578.
Mangelsdorf,D.J., Ong,E.S., Dyck,J.A. and Evans,R.M. (1990) *Nature*, **345**, 224–229.
Milbrandt,J. (1988) *Neuron*, **1**, 183–188.
Misrahi,M.,Atger,M., D'auriol,L., Loosfelt,H., Meriel,C., Fridlansky,F., Guiochon-Mantel,A., Galibert,F. and Milgrom,E. (1987) *Biochem. Biophys. Res. Commun.* , **143**, 740–748.
Miyajima,N., Kadowaki,Y., Fukusjoge,S.I., Shimizu,S.I., Semba,K., Yamanashi,Y., Matsubara,K.I., Toyoshima,K. and Yamamoto,T. (1988) *Nucleic Acids Res.*, **16**, 11057–11066.
Miyajima,N., Horiuchi,R., Shibuya,Y., Fukushige,S.I., Matsubara,K.I., Toyoshima,K. and Yamamoto,T. (1989) *Cell*, **57**, 31–39.
Mlodzik,M., Hiromi,Y., Weber,U., Goodman,C.S. and Rubin,G.M. (1990) *Cell*, **60**, 211–224.
Moore,D.D. (1990) *New Biol.*, **2**, 100–105.
Nauber,U., Pankratz,M.J., Kienlin,A., Seifert,E., Klemm,U. and Jäckle,H. (1988) *Nature* **336**, 489–492.
Nei,M. (1987) *Molecular Evolutionary Genetics*. Columbia University Press, New York.
Nirdé,P., Torpier,G.,De Reggi, M.L. and Capron, A. (1983) *FEBS Lett.*, **151**,223–227.
O'Donnell,A.L. and Koenig,R.J. (1990) *Mol. Endocrinol.*, **4**, 715–720.
O'Donnell,A.L., Rosen,E.D., Darling,D.S. and Koenig,R.J. (1991) *Mol. Endocrinol.*, **5**, 94–99.
O'Malley,B.W. (1989) *Endocrinology*, **125**, 1119–1120.
Oro,A.E., Ong,E.S., Margolis,J.S., Posakony,J.W., McKeown,M. and Evans,R.M. (1988) *Nature*, **336**, 493–496.
Oro,A.E., McKeown,M. and Evans,R.M. (1990) *Nature*, **347**, 298–299.
Pankratz,M.J. and Jäckle,H. (1990) *Trends Genet.*, **6**, 287–292.
Petkovitch,M., Brand,N.J., Krust, A. and Chambon,P. (1987) *Nature*, **330**, 444–450.
Pignoni,F., Baldarelli,R.M., Steingrimsson,E., Diaz,R.J., Patapoutian,A., Merriam,J.R. and Lengyel,J.A. (1990) *Cell*, **62**, 151–163.
Ponglikitmongkol,M., Green,S. and Chambon,P. (1988) *EMBO J.*, **7**, 3385–3388.
Ragsdale,C.W., Petkovich,M., Gates,P.B., Chambon,P. and Brockes,J.P. (1989) *Nature*, **341**, 654–657.
Rothe,M., Nauber,U. and Jäckle,H. (1989) *EMBO J.*, **8**, 3087–3094.
Ryseck,R.P., Macdonald-Bravo,H., Mattéi,M.G., Ruppert,S. and Bravo,R. (1989) *EMBO J.*, **8**, 3327–3335.
Saitou,N. and Nei,M. (1987) *Mol. Biol. Evol.* **4**, 406–425.
Sap,J., Munoz,A., Damm,K., Goldberg,Y., Ghysdael,J., Leutz,A., Beug,H. and VennstrÖm,B. (1986) *Nature*, **324**, 635–640.
Schawbe,J.W.R. and Rhodes,D. (1991) *Trends Biochem. Sci.*, **16**, 291–296.
Segraves,W.A. (1991) *Cell*, **67**, 225–228.
Segraves,W.A. and Hogness,D.S. (1990) *Genes Dev.*, **4**, 204–219.
Shea,M.J., King,D.L., Conboy,M.J., Mariani,B.D. and Kafatos,F.C. (1990) *Genes Dev.*, **4**, 1128–1140.
Sladek,F.M., Zhong,W., Lai,E. and Darnell Jr,J.E. (1990) *Genes Dev.*, **4**, 2353–2365.
Sneath,P.H.A. and Sokal. (1973) *Numerical Taxonomy*. W.H. Freeman, San Francisco.
Swofford,D.L. and Olsen G.J. (1990) In Hillis,D.M. and Moritz,C. (eds), *Molecular Systematics*. Sinauer Associates Inc., Sunderland, MA, pp.411–501.
Thompson,C.C., Weinberger,C., Lebo,R and Evans R.M. (1987) *Science*, **237**, 1610–1614.
Tilley,W.D., Marcelli,M., Wilson,J.D. and McPhaul,M.J. (1989) *Proc. Natl. Acad. Sci. USA*, **86**, 327–331.
Valentine,J.W. (1977) In Hallam,A. (ed.), *Pattern of Evolution*. Elsevier, Amsterdam, pp. 27–58.
Wang,L.H., Tsai,S.Y., Cook,R.G., Beattie,W.G., Tsai,M.J. and O'Malley,B.W. (1989) *Nature*, **340**, 163–166.
Weinberger,C., Thompson,C.C., Ong,E.S., Lebo,R., Gruol,D.J. and Evans,R.M. (1986) *Nature*, **324**, 641–646.
Xiong,Y. and Eickbush,T.H. (1988) *Mol. Biol. Evol.* **5**, 675–690.
Xiong,Y. and Eickbush,T.H. (1990) *EMBO J.* **9**, 3353–3362.
Yaoita,Y., Shi,Y.B. and Brown,D.D. (1990) *Proc. Natl. Acad. Sci. USA* **87**, 7090–7094.
Zelent,A., Krust,A., Petkovich,M., Kastner,P. and Chambon,P. (1989) *Nature*, **339**, 714–717.

Received on July 30, 1991; revised on November 22, 1991