



## The Nuclear Receptor Superfamily Has Undergone Extensive Proliferation and Diversification in Nematodes

Ann E. Sluder, Siuyien Wong Mathews, David Hough, et al.

*Genome Res.* 1999 9: 103-120

Access the most recent version at doi:[10.1101/gr.9.2.103](https://doi.org/10.1101/gr.9.2.103)

---

**References** This article cites 80 articles, 36 of which can be accessed free at:  
<http://genome.cshlp.org/content/9/2/103.full.html#ref-list-1>

### License

**Email Alerting Service** Receive free email alerts when new articles cite this article - sign up in the box at the top right corner of the article or [click here](#).

---

To subscribe to *Genome Research* go to:  
<https://genome.cshlp.org/subscriptions>

---

Cold Spring Harbor Laboratory Press

## Research

# The Nuclear Receptor Superfamily Has Undergone Extensive Proliferation and Diversification in Nematodes

Ann E. Sluder,<sup>1,3</sup> Siuyien Wong Mathews,<sup>1,4</sup> David Hough,<sup>2</sup> Viravuth P. Yin,<sup>2,5</sup> and Claude V. Maina<sup>2</sup>

<sup>1</sup>Department of Cellular Biology, University of Georgia, Athens, Georgia 30602 USA; <sup>2</sup>New England Biolabs, Beverly, Massachusetts 01915 USA

The nuclear receptor (NR) superfamily is the most abundant class of transcriptional regulators encoded in the *Caenorhabditis elegans* genome, with >200 predicted genes revealed by the screens and analysis of genomic sequence reported here. This is the largest number of NR genes yet described from a single species, although our analysis of available genomic sequence from the related nematode *Caenorhabditis briggsae* indicates that it also has a large number. Existing data demonstrate expression for 25% of the *C. elegans* NR sequences. Sequence conservation and statistical arguments suggest that the majority represent functional genes. An analysis of these genes based on the DNA-binding domain motif revealed that several NR classes conserved in both vertebrates and insects are also represented among the nematode genes, consistent with the existence of ancient NR classes shared among most, and perhaps all, metazoans. Most of the nematode NR sequences, however, are distinct from those currently known in other phyla, and reveal a previously unobserved diversity within the NR superfamily. In *C. elegans*, extensive proliferation and diversification of NR sequences have occurred on chromosome V, accounting for > 50% of the predicted NR genes.

[The sequence data described in this paper have been submitted to the GenBank data library under accession nos. AF083222–AF083225 and AF083251–AF083234.]

Nuclear receptors (NRs) are one of the most abundant classes of transcriptional regulators in metazoans and are involved in processes as diverse as sexual differentiation, metabolic regulation, insect metamorphosis, vertebrate limb development, and embryonic pattern formation (Gronemeyer and Laudet 1995; Kastner et al. 1995; Mangelsdorf et al. 1995; Thummel 1995). The defining members of the NR superfamily were first identified biochemically as receptors for steroid and thyroid hormones, and subsequent cloning of the genes revealed that these receptors constituted a class of related proteins. The signature motif of the superfamily is a DNA-binding domain (DBD) comprised of two Cys<sub>2</sub>–Cys<sub>2</sub> zinc coordinating modules (Rastinejad 1997). The ability of a nuclear receptor to regulate gene expression is modulated by binding of the cognate ligand to a domain positioned carboxy-terminal to the DBD. This ligand-binding domain (LBD) also participates in receptor homo- and heterodimerization, and contributes to transcriptional regulation. Although less highly conserved than the DBD, the LBDs of the classic nuclear receptors also contain regions of conserved similarity (Simons 1997).

Present addresses: <sup>4</sup>Biotechnology Facility, University of Hawaii at Manoa, Honolulu, Hawaii 96822 USA; <sup>5</sup>University of Pennsylvania School of Medicine, 320 CRB, Philadelphia, Pennsylvania 19104 USA.

<sup>3</sup>Corresponding author.

E-MAIL [asluder@cb.uga.edu](mailto:asluder@cb.uga.edu); FAX (706) 542-4271.

Additional members of the NR superfamily have been identified by sequence similarities to known NRs, primarily based on the conserved DBD motif (Mangelsdorf et al. 1995). The superfamily is now represented by >400 cloned sequences from >60 species. These sequences define at least 70 distinct NRs, several of which are represented by orthologs from multiple species (Gronemeyer and Laudet 1995). A number of NRs first recognized on the basis of sequence were subsequently shown to be receptors for known hormones (Arriza et al. 1987; Giguere et al. 1987; Petkovich et al. 1987; Evans 1988; Koelle et al. 1991). Efforts in many laboratories during the past decade have brought to 23 the number of NRs for which ligands or candidate activators are known (Mangelsdorf and Evans 1995; Mangelsdorf et al. 1995; Forman 1997; Forman et al. 1998). The majority of NRs remain orphan receptors for which the specific ligands, if any exist, have not yet been identified.

Extensive analysis of relationships within the NR superfamily by Laudet and coworkers defined six major NR families, most of which can be further subdivided into distinct classes of orthologs and paralogs (Laudet et al. 1992; Escriva et al. 1997; Laudet 1997). Five of these six NR families contain members from both vertebrates and insects (Laudet 1997, A. Sluder, unpubl.), indicating that these families arose before the evolu-

tionary divergence of protostomes and deuterostomes. The exception is the vertebrate steroid receptor family, members of which have thus far not been found in other phyla. In addition, the small *knirps* family of divergent orphan receptors (which was not included in Laudet's analysis) is currently represented only in dip- terans. Phylogenetic analysis of the six major families led to the proposal that current NR diversity arose from two distinct waves of gene duplication and divergence: one that occurred very early during the emergence of metazoans, and a second that expanded particular NR classes in the vertebrate phylogenetic lineage (Escriva et al. 1997; Laudet 1997).

Only five nematode sequences were included in the analysis described above. Using the conserved DBD zinc modules as the defining NR motif, we find >200 predicted NR genes in the genome of *Caenorhabditis elegans*, five-fold more than reported for any other species to date. Analysis of available genomic sequences from the related species *Caenorhabditis briggsae* indicates that it is likely to have a comparable number of NR genes. To establish a phylogenetic framework for the functional analysis of nematode NR genes, we undertook a comparative analysis of these sequences. Specifically, we addressed two questions: First, how many of the NR sequences correspond to functional genes? Second, what are the relationships of the nematode NRs to those from other organisms? Our experimental results in combination with data available in published reports and the public sequence databases demonstrate expression for 25% of the predicted *C. elegans* NR genes, and statistical arguments suggest that the majority of the 228 *C. elegans* NR sequences are likely to represent expressed genes. In addition, comparative analysis of the nematode NR DBD sequences shows that both highly divergent NR sequences and phylogenetically conserved NR classes are represented in the nematode genes. Our observations are consistent with the existence of primordial metazoan NR families and later extensive duplication and diversification of NR genes in nematodes.

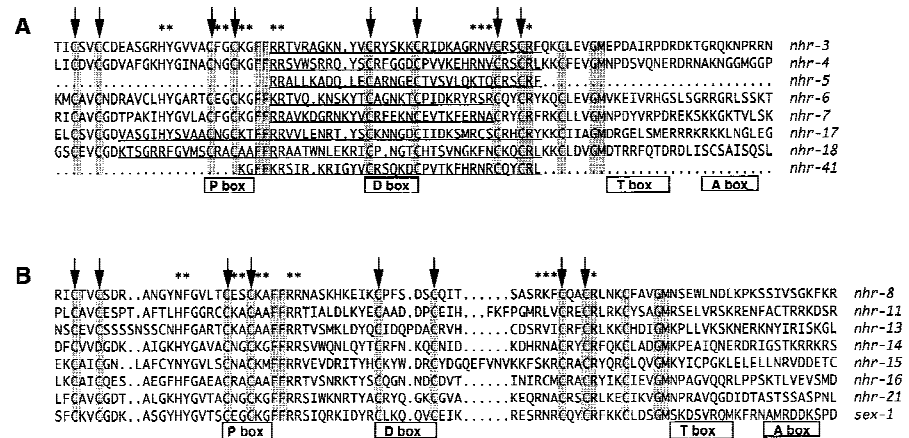
## RESULTS

Nematode NR genes were identified by three means — polymerase chain reaction (PCR) amplification of genomic sequences using NR-specific de-

generate primers, characterization of expressed sequence tags (ESTs), and analysis of genomic sequences released by the *C. elegans* Genome Sequencing Project. Below we first summarize our characterization of cDNA clones and the initial analysis of NR sequences revealed by the genome project. We then present the results of our expression studies and comparative analyses.

## Identification of cDNA Clones for 13 *C. elegans* NR Genes

*C. elegans* sequences encoding predicted members of the nuclear receptor superfamily were isolated by PCR amplification of genomic DNA using degenerate primers corresponding to highly conserved regions of the NR DBD (see Materials and Methods). Amplification products were cloned and sequenced to identify eight candidate NR genes (Fig. 1A). The cloned PCR products were used to probe a filter grid of ordered yeast artificial chromosomes (YACs) representing ~95% of the *C. elegans* genome (Coulson et al. 1995), verifying that these sequences are represented in the *C. elegans* genome. As each YAC represents a defined region of the genome, these hybridizations also mapped each sequence to a known physical location within the genome. *nhr-4* and *nhr-6* were also detected in a hybridization screen described elsewhere (Sluder et al. 1997). The predicted genes and their locations are summarized in Table 1A.



**Figure 1** DBD sequences encoded by *C. elegans* NR genes described in this study. Shaded amino acids are conserved in all NRs other than the nematode exceptions noted in the text. Arrows indicate the eight zinc-coordinating cysteines (four per zinc-binding module). Asterisks denote positions of residues found to make specific DNA contacts in the existing NR DBD X-ray crystal structures (Rastinejad 1997). Underlying rectangles in the core DBD indicate the P box sequences that are primary determinants of DNA-binding specificity and the D box dimerization interface (Gronemeyer and Laudet 1995). Also indicated are the T box and A box regions that contribute to dimerization and DNA binding, respectively, by some NRs (Gronemeyer and Laudet 1995; Rastinejad 1997). Alignments were generated using the GCG Pileup program (Devereux et al. 1984). (A) NR sequences identified by degenerate PCR amplification of genomic DNA. Amino acids encoded by the original amplification products are underlined. (B) NR sequences identified from expressed sequence tags. The *nhr-14* DBD sequences are from the corresponding genome project ORF prediction, as the cDNA sequenced was incomplete and did not extend to the DBD region.

**Table 1.** *C. elegans* NR Genes Described in this Study

A. Identified by degenerate PCR screen					
Chromosome	Gene	YACs	Genome ORF	cDNA (ESTs not sequenced)	Comments
II	<i>nhr-41</i>	Y42F9, Y27F6, Y19D2, Y39E3			
III	<i>nhr-6</i>	Y68C4, Y43E9	C48D5.1	pBS/nhr6, (yk446a12)	a.k.a. <i>ceb-1</i> (Wilson et al. 1992); <i>cnr-8</i> (Kostrouch et al. 1995)
IV	<i>nhr-4</i>	Y37F11, Y43F7, Y49F10, Y39E2	F32B6.1	pBS/nhr4, (yk411a10)	
	<i>nhr-5</i>	Y73F8			
	<i>nhr-7</i>	Y72H7, Y51B3, Y37G7	F54D1.4	yk396h7	cDNA splicing different from genome ORF prediction
V	<i>nhr-18</i>	Y54G1, Y5H2	F44C8.3a	pBS/nhr18	upstream gene in potential operon
X	<i>nhr-3</i>	Y52F5, Y52C11, Y43C6	H01A20.b	pBS/nhr3	cDNA splicing different from genome ORF prediction
	<i>nhr-17</i>	Y39F3, Y50C12, Y40C3	C02B4.2		
B. EST cDNA sequences completed					
			Genome ORF	ESTs	Comments
II	<i>nhr-16</i>		T12C9.1?	CEESU44	cDNA splicing different from ORF prediction
	<i>nhr-21</i>		F21D12.1	cm2h1, yk77h11	cDNAs reveal two isoforms
IV	<i>nhr-8</i>		F33D4.1	cm14e12	cDNA splicing different from ORF prediction
	<i>nhr-11</i>		ZC410.1	cm7a11	cDNA splicing different from ORF prediction
V	<i>nhr-13</i>			cm9g2	
	<i>nhr-15</i>		F33E11.1	CEESV11	
X	<i>nhr-14</i>		T01B10.4	cm15g5	cDNA splicing different from ORF prediction
	<i>nhr-24/sex-1</i>		F44A6.2	cm11f5	a.k.a. <i>cnr-14</i> (Kostrouch et al. 1995)

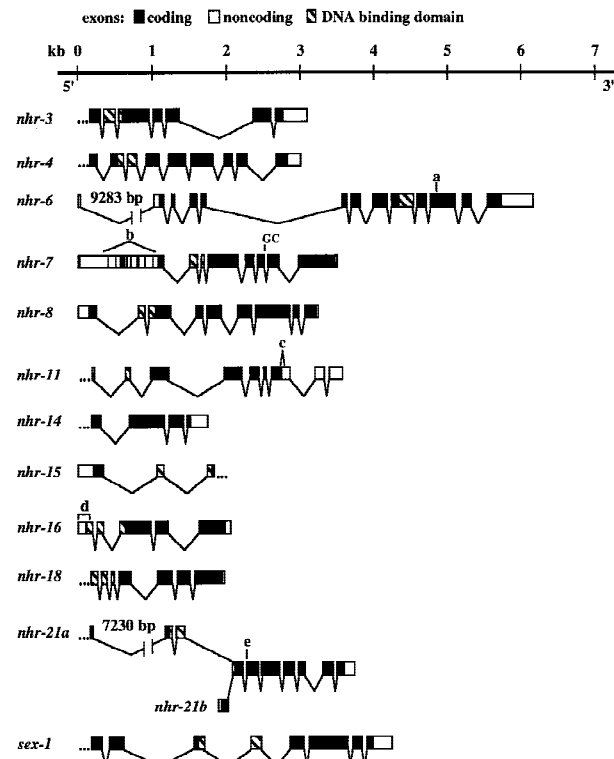
The cloned PCR products were also used as probes to screen a mixed stage *C. elegans* cDNA library. cDNA clones were recovered for four of the eight genes (*nhr-3*, *nhr-4*, *nhr-6*, and *nhr-18*), and the complete sequence of each of these cDNAs was determined. We also completed the sequences of cDNAs representing an additional eight *C. elegans* nuclear receptor genes (Fig. 1B and Table 1B) that were identified as ESTs in two different surveys (McCombie et al. 1992; Waterston et al. 1992). A recent, more extensive EST project (Y. Kohara, pers. comm.) has reported ESTs for *nhr-4*, *nhr-6*, *nhr-7*, and *nhr-21*, and we completed the sequences for the *nhr-7* and *nhr-21* ESTs. Through this combination of library screening and EST analysis, we have characterized 14 cDNA clones representing 13 different NR genes.

Two of the genes for which we report cDNA sequences, *nhr-6* and *nhr-24*, have also been described by others — *nhr-6* as *ceb-1* (Wilson et al. 1992) and *cnr-8* (Kostrouch et al. 1995) and *nhr-24* as *cnr-14* (Kostrouch et al. 1995). The *nhr* designation has been adopted as the standard nomenclature for nuclear receptor genes (Hodgkin 1997), and we use the *nhr-6* designation here. *nhr-24/cnr-14* has been shown to correspond to the sex determination gene *sex-1* (Carmi et al. 1998), and we use the *sex-1* designation in the remainder of this work.

### The cDNAs Reveal a Variety of Features in the NR Gene Structures

As of July 1, 1998, genomic sequences corresponding to all the cDNAs but that for *nhr-13* had been released by the *C. elegans* genome project (Wilson et al. 1994; *C. elegans* Genome Sequencing Project, pers. comm.). The corresponding predicted open reading frames (ORFs) are listed in Table 1 and the gene structures, except for *nhr-13*, are diagrammed in Figure 2. Although an exact match to most of the *nhr-16* cDNA sequence is found on cosmid T12C9 (chromosome II), the 5'-most 138 bp of the 154 bp exon 1 (Fig. 2, note d) are not yet represented in reported genomic sequence, which includes >25 kb of sequence upstream from the matching region. Thus, the exact gene structure and definitive location of *nhr-16* remain unclear. For *nhr-4*, *nhr-15*, and *sex-1* the splicing patterns predicted by the genome project are confirmed by the cDNA sequences. The splicing patterns, and thus the predicted protein products, of the remaining cDNAs differ from those predicted by the genome project.

None of the 14 cDNA clones sequenced contain the trans-spliced leader (SL) sequences found at the 5' ends of many *C. elegans* mRNAs (Krause and Hirsh 1987; Blumenthal and Steward 1997), so none unambiguously pinpoints the 5' end of the corresponding



**Figure 2** *C. elegans* NR gene structures defined by cDNA sequences. (Solid boxes) Coding; (open boxes) noncoding; (hatched boxes) DBD. Specific features of note (see text for additional discussion): (*nhr-6*) *a* marks site of possible alternate splice suggested by RT-PCR results. (*nhr-7*) vertical lines mark locations of AUG codons upstream of predicted initiation codon. Bracket *b* indicates possible splice sites that would remove upstream ORFs from the 5' UTR. Also indicated is the site of a rare GC 5' splice donor site (Blumenthal and Steward 1997). (*nhr-11*) bracket *c* marks predicted intron that would remove in-frame stop codon and extend the ORF. (*nhr-16*) bracket *d* indicates sequences without a current match in released genomic sequence. (*nhr-21*) *e* indicates the site of two splice acceptor sites separated by 6 bp; the 3' site is used in *nhr-21a* and the 5' in *nhr-21b*.

transcript. However, seven of the clones (cm2h1 and cDNAs for *nhr-3*, *nhr-4*, *nhr-6*, *nhr-7*, *nhr-8*, and *nhr-16*) have in-frame stop codons upstream of the apparent start codons and end in poly(A) tracts, indicating that these do contain complete coding regions. The EST clone CEESV11 (*nhr-15*) lacks the 3' portion of the cDNA sequence, as it does not contain a poly(A) tract; the cDNA was apparently truncated at an internal *Xho*I restriction site during library construction. The other six clones (yk77h11 and cDNAs for *nhr-11*, *nhr-13*, *nhr-14*, *nhr-18*, and *sex-1*) are also unlikely to be full length, as the ORFs continue to the 5' ends of the sequences. Although the *nhr-14* cDNA sequenced does not contain sequences encoding a NR DBD motif, this cDNA corresponds to a predicted ORF (T01B10.4) that does encode a DBD.

For several of the genes, the structures provide

hints of strategies that may contribute to their regulation and expression:

#### *nhr-6*

The first and fifth introns, with lengths of 9283 bp and 1815 bp, respectively, are unusually large for *C. elegans* (Blumenthal and Steward 1997), and the gene spans >15 kb of genomic sequence. The first five exons in our cDNA are not predicted by the genome project nor represented in a previously reported cDNA (Kostrouch et al. 1995). The large fifth intron could contain a second promoter that would allow differential expression of an alternate transcript, as has been observed for other NR genes (e.g., Talbot et al. 1993; Kozlova et al. 1998).

#### *nhr-7*

The cDNA contains an unusually large 5' untranslated region (UTR) of 1114 nucleotides. There are 15 AUG codons and 9 small ORFs upstream of the predicted initiator codon (Fig. 2). In addition, the 5' UTR has the potential to form several stable stem-loop structures (not shown). These features could contribute to the control of NHR-7 protein expression, as both upstream ORFs and 5' UTR secondary structures have been implicated in the regulation of translation in other systems (Kozak 1991; Damiani and Wessler 1993; Lohmer et al. 1993; Geballe and Morris 1994; Hinnebusch 1997). For example, upstream ORFs in the 5' UTRs of the *Drosophila Ultrabithorax* and *Antennapedia* genes and the mouse retinoic acid receptor  $\beta$  gene confer temporal and spatial translational control to downstream protein-coding regions (Zimmer et al. 1994; Ye et al. 1997). The ORFs in the *nhr-7* 5' UTR are flanked by reasonable consensus splice sites that were not used in the cDNA sequenced (Fig. 2, note b). Translation of prostaglandin synthase in chicken embryo fibroblasts is regulated at the level of mRNA splicing (Xie et al. 1991), and a similar regulated splicing event removing the ORF-laden region of the *nhr-7* 5' UTR could contribute to the control of NHR-7 expression.

#### *nhr-11*

The ninth predicted intron of the corresponding genome project ORF (ZC410.1) has not been spliced out in the cDNA, and a resulting in-frame stop codon leads to a much shorter carboxy-terminal domain than that predicted for ZC410.1 (Fig. 2, note c). For one NR gene in the dog heartworm *Dirofilaria immitis*, alternately spliced transcripts similar in structure to both cm7a11 and ZC410.1 are found (C. Maina, unpubl.), suggesting that both isoforms could also occur for *nhr-11*.

#### *nhr-18*

The large predicted ORF F44C8.3 contains two NR DBD motifs. The *nhr-18* cDNA corresponds to the 5' portion

of F44C8.3, spanning one of the two DBDs and ending in a poly(A) tail. Therefore, this predicted ORF is likely to represent two genes. The distance between the 3' end of the *nhr-18* cDNA and the next predicted in-frame AUG codon is only 99 bp, suggesting that these genes may be coexpressed as an operon (Zorio et al. 1994) encoding two NRs.

#### *nhr-21*

The two cDNAs (cm2h1 and yk77h11) exhibit different splicing patterns and encode different predicted protein isoforms that differ in their amino termini. The shorter isoform (*nhr-21b*, defined by cm2h1) has a truncated DBD bearing only the carboxy-terminal half of the second zinc module. This isoform is unlikely to bind DNA, but may retain an ability to heterodimerize with other NRs. Production of both DNA-binding and non-DNA-binding isoforms is reminiscent of the *Drosophila* E75 (Segraves and Hogness 1990) and E78 (Stone and Thummel 1993) orphan receptor genes. In the case of E75, the truncated isoform exerts a dominant negative effect on its heterodimer partner DHR3, modulating the ability of DHR3 to activate target genes (White et al. 1997). The NHR-21B isoform could perform a similar regulatory function.

### The *C. elegans* Genome Contains a Large Number of NR-Related Genes

Seven *C. elegans* NR genes have been described in other studies (Table 2). In addition, the predicted genes defined by the genome sequencing project (Wilson et al. 1994; *C. elegans* Genome Sequencing Project, pers. comm.) reveal an abundance of NR genes in the *C. elegans* genome, bringing the total as of July 1, 1998, to 228 predicted genes (a complete listing is available at <http://www.uga.edu/~cellbio/>). As 225 of these genes are found in the 85% of the genome for which sequencing has been completed, the genome may contain as many as ~260 NR genes. BLAST (Altschul et al. 1990) homology searches indicate that regions of the genome for which sequencing is in progress contain  $\geq 20$  additional NR genes. Thus far only five NR genes have been correlated definitively with mutationally de-

finied loci (*daf-12*, *fax-1*, *odr-7*, *sex-1*, and *unc-55*; Tables 1 and 2).

As noted above, 10 of the 14 cDNAs we sequenced required splicing patterns different from the initial genome project predictions. EST sequences have been reported for 34 additional NR genes (Table 3) and reveal splicing patterns different from the predicted structures for 8 genes. These observations indicated that the computer-aided assembly of ORF predictions by the genome project (Waterston et al. 1997) may frequently err in detail. Because inappropriately assembled sequences would not yield optimal multiple sequence alignments for use in the comparative analysis described below, we reviewed all 228 predicted gene structures, focusing primarily on the DBD motif.

In addition to F44C8.3 (see *nhr-18* above), four of the predicted ORFs (C28D4.1, C50B6.8, F44C8.2, and T19A5.4) contain two NR DBD motifs. Three considerations suggest that these five predicted ORFs most likely each represent two gene products inappropriately fused by the gene prediction algorithm used. First, in each case the paired DBD motifs are embedded in more extensive regions of repeated similarity, suggesting that they arose from gene duplication events. Second, the occurrence of operons, in which closely linked ORFs are cotranscribed and subsequently separated by trans-splicing, is well documented in *C. elegans* (Zorio et al. 1994; Blumenthal and Steward 1997). And finally, existing cDNAs corresponding to only one-half of a "double" NR ORF confirms the production of single ORF mRNA for two of these loci (*nhr-18* in Table 1 and *nhr-44* in Table 3). Therefore, in the analysis below we have treated each DBD motif as representing a separate gene product (designated, e.g., C28D4.1a for the 5' DBD and C28D4.1b for the 3').

A different situation is revealed by the EST clone yk271b11, which contains sequences from two tandem predicted genes (*nhr-51* and *nhr-52* in Table 3). These predicted genes may represent a single gene with the potential to produce multiple alternately spliced isoforms or to encode a novel NR bearing two DBDs. Elucidation of the exact gene structure will require additional analysis. However, at least two distinct DBDs are

**Table 2.** *C. elegans* NR Genes Described in Other Studies

Chromosome	Gene	Genome ORF	ESTs	References
I	<i>nhr-2</i>	C32F10.6	yk452h1	Sluder et al. (1997)
	<i>nhr-23</i>	C01H6.5	yk4d4, yk260c9	a.k.a. <i>cnr-3</i> , Kostrouch et al. (1995, 1998)
X	<i>unc-55</i>	F55D12.4		Zhou et al. (1997)
	<i>daf-12</i>	F11A1.3		Yeh (1991)
	<i>fax-1</i>	F56E3.4		B. Wightman (pers. comm.)
	<i>nhr-1</i>	R09G11.2	yk128c9, yk65f10, yk472h3, yk414f3	B. Honda (pers. comm.)
	<i>odr-7</i>	T18D3.2		Sengupta et al. (1994)

**Table 3.** *C. elegans* NR Genes Represented by ESTs

Chromosome	Gene	Genome ORF	ESTs	Comments
I	<i>nhr-49</i>	K10C3.6	yk61c12, yk26f4, yk242b12, yk212f7	
	<i>nhr-62</i>	K10C3.1	yk63h9, yk354c12	ORF truncated at cosmid end; ESTs reveal DBD
II	<i>nhr-64</i>	C45E1.1	yk384e7	
	<i>nhr-22</i>	K06A1.4	yk324h4, yk325g8	
	<i>nhr-61</i>	W01D2.2	yk442a3	EST splicing different than ORF prediction
IV	<i>nhr-31</i>	C26B2.3	yk221a5, yk296b12, yk192f3, yk226h10	EST splicing different than ORF prediction
	<i>nhr-34</i>	F58G6.5	yk201f9	
	<i>nhr-43</i>	C29E6.5	yk89g11	
	<i>nhr-46</i>	C45E5.6	yk5b9, yk82e1, yk36g3, yk327d10, yk412d10	
V	<i>nhr-66</i>	T09A12.4	yk195d2, yk295g11, yk374f3, yk432g1	EST splicing different than ORF prediction
	<i>nhr-67</i>	C08F8.a	yk273f2	
	<i>nhr-12</i>	R04B5.4	cm06h9	
	<i>nhr-42</i>	C33G8.6	yk482f11	EST splicing different than ORF prediction
	<i>nhr-44</i>	T19A5.4b	yk157h5	ORF has two DBDs, likely operon <i>csr-1</i> (GenBank accession no. U41296)
	<i>nhr-47</i>	T05C3.1		
	<i>nhr-50</i>	C06C6.5	yk311d1	
	<i>nhr-51</i>	K06B4.1	yk271b11 (5')	EST has sequences from two ORFs
	<i>nhr-52</i>	K06B4.2	yk271b11 (3')	EST has sequences from two ORFs
	<i>nhr-53</i>	K06B4.11	yk458f7, yk464a6	EST splicing different than ORF prediction
	<i>nhr-54</i>	F36D3.2	yk123a5	
	<i>nhr-55</i>	T01G6.7	yk255f5, yk412g5	
	<i>nhr-56</i>	F44C8.6	yk304h10	
	<i>nhr-57</i>	T05B4.2	yk370h8, yk203b8	
	<i>nhr-58</i>	R11G11.2	yk308e2	
<i>nhr-59</i>	T27B7.1	yk490h1		
<i>nhr-60</i>	F57A10.5	yk470e6		
<i>nhr-63</i>	C06C6.4	yk474f10		
<i>nhr-65</i>	Y17D7A.3	yk480h4	incomplete DBD	
X	<i>nhr-25</i>	F11C1.6	yk174f2, yk342d8	
	<i>nhr-28</i>	C11G6.4	yk158g9	EST splicing different than ORF prediction
	<i>nhr-35</i>	C07A12.3	yk215f10	
	<i>nhr-40</i>	T03G6.2	yk213d2	
	<i>nhr-45</i>	F16H11.5	yk347g3	EST 5' end has inverted duplication
	<i>nhr-48</i>	ZK662.3	yk226g12, yk333f9	EST splicing different than ORF prediction

possible, and we have treated them as separate for the purposes of this analysis.

Of the remaining 172 NR genes represented only by genome project ORF predictions, 48 exhibited either incomplete or unusually long DBD motifs. For each of these, the genomic sequence was examined for candidate alternate splicing strategies that would result in a more conventional DBD. Three ORFs were truncated at the end of a cosmid, and for two of these the DBD could be completed by sequences from the neighboring clone. Revised splicing strategies were identified for 27 additional ORFs, and the modified DBD motifs

were used in the comparative analysis below. (Revised splicing predictions are summarized at <http://www.uga.edu/~cellbio/>, and the details have been submitted to the Genome Project for inclusion in the ACeDB database.) The remaining 18 predicted ORFs do not appear to encode complete DBD motifs. For 8 of these 18, sequences capable of encoding the missing regions of the DBD are present in predicted introns or 5' untranslated sequences, but there are no candidate splice sites that would allow these to be spliced in frame with the remainder of the corresponding ORF. Some of these sequences encoding truncated DBDs

**Table 4.** RT-PCR Expression Profiles of NR Genes on Chromosome III

Gene	Genome ORF	Embryo	L1	L2	L3	L4	Adult
<i>No. of times cDNA product observed/no. of replicate amplifications</i>							
<i>nhr-6</i>	C48D5.1	3/3	3/3	3/3	2/3	3/3	3/3
<i>nhr-9</i>	ZK418.1	0/4	0/4	3/4	3/4	0/4	0/4
<i>nhr-10</i>	B0280.8	4/4	4/4	4/4	4/4	3/4	2/4
<i>nhr-20</i>	F43C1.4	2/2	2/2	2/2	2/2	2/2	0/2

may be pseudogenes, whereas others may encode NR-related proteins whose functions do not require DNA binding. Identification of an EST demonstrates that at least one gene incapable of encoding a complete DBD (*nhr-65*) is expressed as mRNA.

### Many of the *C. elegans* NR Sequences Are Expressed as mRNA

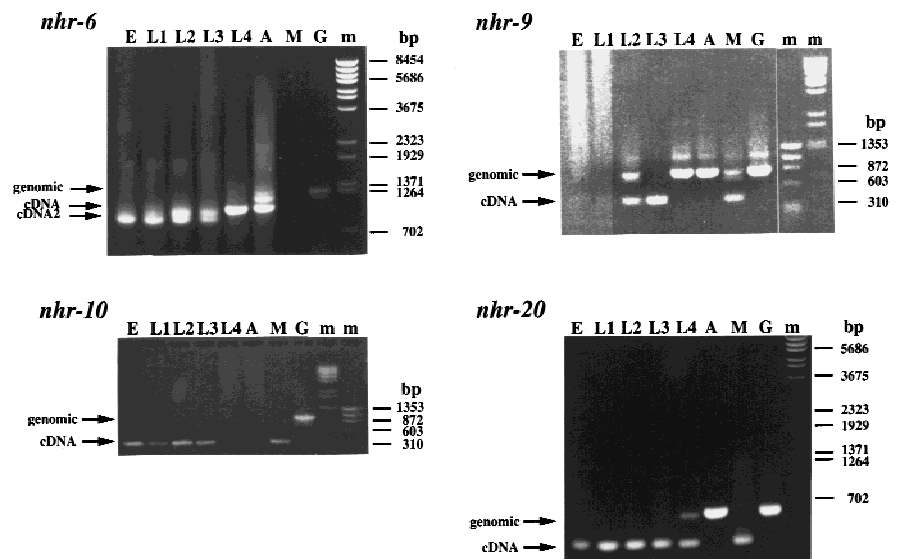
The number of predicted *C. elegans* NR genes is fivefold greater than that reported from any other species—at present *Homo sapiens* runs a distant second with 44 NR sequences in GenBank. One potential explanation for the large number of predicted *C. elegans* NR genes is that many of the sequences could be nonfunctional pseudogenes predicted as ORFs by the GeneFinder program and other computer algorithms used in the genome project's sequence analysis (Waterston et al. 1997). To assess how many of the NR sequences correspond to expressed genes, we used information available in published reports and the public databases, in addition to our own experimental results.

The EST database maintained by National Center for Biotechnology Information (NCBI) contained >70,000 *C. elegans* entries as of July 1, 1998, primarily from the large EST project under way at the National Institute of Genetics in Japan (Y. Kohara, pers. comm.). We searched the EST database with each of the predicted *C. elegans* NR genes, identifying 70 EST clones (~0.1% of *C. elegans* ESTs) that correspond to 48 NR sequences. These 48 genes include 14 of the 20 for which we or others have demonstrated expression, bringing to 54 the number of NR genes known to be expressed as mRNA (Tables 1–3). However, as predicted transcriptional regulators, many NRs are likely to be expressed at rela-

tively low levels, perhaps in only a few cells (e.g., *odr-7*, Sengupta et al. 1994) or during a restricted period of the life-cycle (e.g., *nhr-2*, Sluder et al. 1997), and thus may be poorly represented in the cDNA libraries used in the EST surveys. Notably, of the five NR genes with known genetic functions, only one (*sex-1*) is represented in the

EST database, suggesting that EST representation may underestimate the proportion of NR genes that are functional.

To probe further the proportion of expressed NR genes represented in the EST database, we assayed directly the expression of all known NR genes on Chromosome III, the first chromosome for which the nearly complete sequence was available (Genome Sequencing Project, pers. comm.). Of the four NR genes revealed by the genomic sequence of chromosome III (Table 4), *nhr-6* is the only one for which an EST sequence has been reported. However, expression of all four genes was demonstrated by RT-PCR amplification from staged mRNA preparations (Table 4 and Fig. 3). For *nhr-9*, *nhr-10*, and *nhr-20*, the amplification products obtained were of the sizes expected for transcripts spliced as predicted by the genome project, and the DNA sequences of the products corresponded to the appropriate gene sequences (see Materials and Methods). For *nhr-6* the predicted size RT-PCR product was



**Figure 3** RT-PCR temporal expression profiles of NR genes on Chromosome III. RT-PCR reactions were performed as described in Materials and Methods. RNA samples were from staged cultures of embryos (E), larval stages L1–L4, and adults (A) or from mixed stage cultures (M). PCR amplification of genomic DNA (G) was included for comparison. Sizes of DNA markers (m) are indicated. Products of the sizes predicted (see Materials and Methods) from amplification of genomic DNA or cDNA samples are designated. For *nhr-6*, cDNA2 indicates the smaller than expected product obtained from early stages.

obtained from L4 and adult, whereas an ~100-bp smaller product was amplified from embryo and L1. Both products were obtained from L2 and L3. DNA sequencing has verified that both products were derived from *nhr-6*. The differences between the two *nhr-6* RT-PCR products was not resolved by the single sequence reads used to verify the products, but analysis of the genomic sequence revealed a potential alternate splice site in exon 11 (see Fig. 2, note a) that would yield a transcript 108 nucleotides smaller than the cDNA while maintaining the reading frame. Developmental Northern blot analysis has confirmed the expression of *nhr-6* in all stages, although it would not have resolved a transcript size difference of 108 nucleotides (Wong 1997).

In summary, existing data demonstrate that at least 57 (25%) of the 228 predicted *C. elegans* NR genes are expressed as mRNA. Furthermore, the cDNA and EST sequences do not contain inappropriate stop codons as would appear for pseudogenes encoding nonfunctional mRNAs. Thus, mRNA expression may be a good indication that these NR sequences represent functional genes, and previously unnamed expressed genes have been assigned *nhr* designations (Tables 3 and 4). We will continue to refer to predicted genes for which no expression or genetic data yet exist by their genome project ORF designations.

Of the two groups of genes for which function or expression has been demonstrated by criteria other than representation in a cDNA library (genetically defined loci and the NR genes on chromosome III), only two of nine genes, or 22%, are represented in the EST database. If this proportion of EST representation holds for the NR genes in general, expression of 218 NR genes would be predicted from the existence of ESTs for 48 NR genes. We note, however, that this relatively small sample of nine NR genes may not be representative of *C. elegans* NR genes at large. In particular, recent progress in the sequencing of chromosomes other than III has revealed many apparent NR gene duplication events (see below), suggesting that the proportion of pseudogenes may be higher on the other chromosomes. A more comprehensive survey of NR gene expression will be needed for a definitive resolution of this issue.

### NR Genes Are Also Abundant in the *C. briggsae* Genome

Comparisons of orthologous gene sequences between *C. elegans* and the sibling species *C. briggsae* have been used to identify conserved elements of both protein coding and regulatory regions (e.g., Kennedy et al. 1993; Wightman et al. 1993; de Bono and Hodgkin 1996; Kuwabara 1996). To allow more extensive comparisons of genome structure, the Genome Sequencing Center at Washington University has begun sequenc-

ing the *C. briggsae* genome. We searched the ~5 Mb of available *C. briggsae* sequence (<http://genome.wustl.edu/gsc/gsc/gschmpg.html>) for candidate NR genes to determine whether these genes are also abundant in the *C. briggsae* genome. If *C. briggsae* also contains ~260 NR genes, the portion of the genome sequenced (~5%) should contain ~13 NR genes. Fourteen likely NR genes were identified, consistent with the prediction that NR sequences are also numerous in *C. briggsae*.

We used two criteria to assess orthology between the *C. elegans* and *C. briggsae* NR genes. First, we required >80% amino acid identity within the DBD sequence, as most pairs of *C. elegans* and *C. briggsae* homologs exhibit this degree of similarity (de Bono and Hodgkin 1996) and as apparent nematode/vertebrate or nematode/insect orthologous NR pairs (see below) share  $\geq 70\%$  amino acid identity within the DBD. Second, we looked for synteny between the two species for other potential coding regions on the clones bearing the NRs (Kuwabara and Shah 1994). All of the NR pairs with >80% amino acid identity in the DBD also occur in genomic regions exhibiting synteny, and no NR pairs with lower degrees of similarity are in syntenic regions. By these criteria, eight of the *C. briggsae* NRs have orthologs in *C. elegans*. A complete listing of the *C. briggsae* NRs and, when relevant, their *C. elegans* orthologs is available at <http://www.uga.edu/~cellbio/>.

### The Nematode NRs Exhibit Unprecedented Diversity in DBD Sequences

Several regions of NR DBDs are very highly conserved (see Fig. 1; Rastinejad 1997). Many of these conserved regions are known to be important for DNA binding or for the secondary structure of the domain (Rastinejad 1997). Whereas these elements are generally conserved in the nematode NRs, for each the sequence diversity observed in nematodes is much greater than that found in NRs currently known from other species. Among the *C. elegans* NRs, a number of changes are observed in amino acids that are otherwise absolutely conserved within the superfamily: replacement of one of the two adjacent conserved phenylalanine residues within the first zinc module by a leucine (F16H9.2) or a tyrosine (C17E7.5, F41B5.10); replacement of the methionine at the carboxyl boundary of the DBD by a leucine (T07C5.2, T07C5.3, T09E11.2, *nhr-43*); and replacement of one of the zinc-coordinating cysteines by a tyrosine (T19A5.4a). This latter change is likely to result in a nonfunctional DBD. Six of these amino acid changes could result from a single nucleotide change. To verify that these represent true diversity within the NR superfamily and not sequencing errors, each codon in question was double checked in the original sequence data by the Genome Sequencing Center, and all were found to be covered by unambiguous sequence reads (J. Spieth, pers. comm.). Whether these changes

indicate that the sequences in question encode divergent NR proteins or are pseudogenes remains to be determined, although we note that at least one (*nhr-43*) is expressed.

Of particular interest is the sequence diversity of the P-box region (see Fig. 1). These amino acids are primary contributors to NR DNA-binding specificity (Rastinejad 1997). To date 76 distinct P box sequences have been observed within the NR superfamily. Six of these (CDGCAG, CDGCKG, CEGCKG, CEGCKS, CESCKA, CESCKG) are found in NRs from a wide variety of species, including nematodes. Seven (CAGCKG, CDGCSG, CEACKA, CEACKV, CEACYA, CEGCKA, CG-SCKV) are limited to vertebrates, primarily in the steroid receptors. The remaining 63 P box sequences (complete list is available at <http://www.uga.edu/~cellbio/>) have thus far been observed only in nematodes. Therefore, many nematode NRs are likely to exhibit DNA-binding specificities distinct from those defined for members of the superfamily to date. One of these novel sequences (CRACAA) is found in nearly one-third of the predicted *C. elegans* NR genes.

#### Both Phylogenetically Conserved and Novel Divergent NR Classes Are Represented in the Nematode Genes

To examine the relationships of the *C. elegans* and *C. briggsae* NRs among themselves and with NRs from other species, we carried out a comparative analysis of the nematode core DBD sequences (as defined in Rastinejad 1997), which range in length from 64 to 82 amino acids. The 18 *C. elegans* NRs with truncated DBDs were omitted from this analysis. A neighbor-joining tree of selected NR sequences is shown in Figure 4. Bootstrap analysis was used to identify the most stable branches, many of which were also retained in a tree generated by maximum parsimony analysis (stable branches are noted in Fig. 4). Different selections of members from the various NR classes yield similar trees, and the topology of the tree shown is generally consistent with that of a much larger tree generated with >400 DBD sequences, including all 224 nontruncated nematode DBDs (not shown). The most stable NR groupings derived in our analysis using only DBD sequences are also consistent with those defined from a combined analysis of both DBD and LBD sequences (Laudet 1997). All eight of the *C. elegans/C. briggsae* orthologous pairs are supported by bootstrap confidence values >95%.

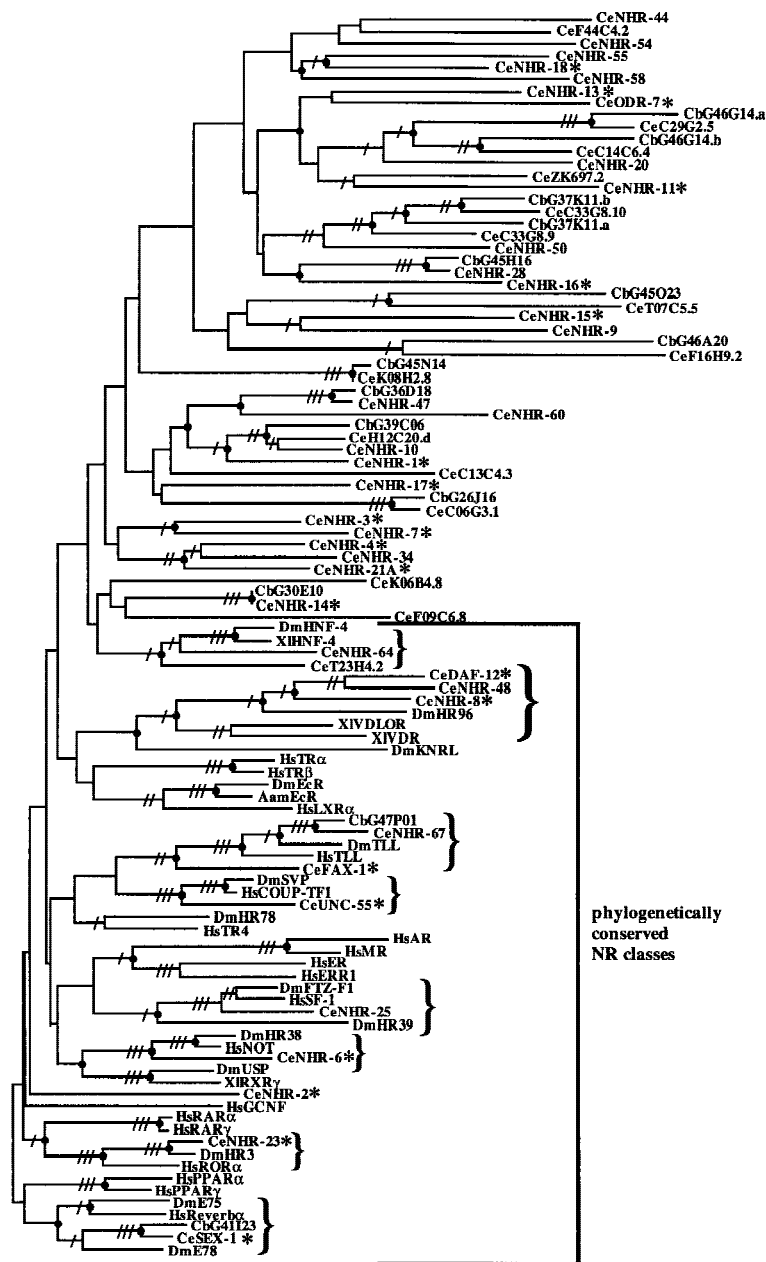
Twelve of the *C. elegans* and 2 of the *C. briggsae* NRs fall into eight conserved NR classes that also contain members from other phyla (Fig. 4). In addition, the *C. elegans* gene *nhr-2* consistently groups with the conserved classes, but, as noted previously (Sluder et al. 1997), it is not a clear member of any single NR class. Nematodes are generally thought to have diverged

from other metazoan phyla during the Cambrian expansion 750–650 million years ago (Fitch and Thomas 1997; Ayala et al. 1998). The conservation of particular NR classes in nematodes as well as in vertebrates and insects strongly supports the proposed ancient origin of these classes (Escriva et al. 1997; Laudet 1997). Furthermore, their continued conservation in nematodes suggests that they perform key biological functions. Consistent with this prediction, mutations in four of the *C. elegans* members of conserved NR classes affect key aspects of development (see Tables 1 and 2; Walthall and Plunkett 1995; Wightman et al. 1997; Antebi et al. 1998; Carmi et al. 1998).

The remaining 209 nontruncated nematode DBD sequences are divergent from all previously known NR classes. Neighbor-joining branch lengths, which are proportional to amino acid differences, are in general greater among these divergent NRs than among the conserved classes (Fig. 4), implying either a longer evolutionary history or a greater rate of evolutionary change for the divergent NRs. The latter currently seems more likely, as the divergent NRs appear to be specific to nematodes (see below) and thus to be more recently derived than the phylogenetically conserved NR classes. Detailed elucidation of the evolutionary relationships among these genes will require additional information, such as intron positions (e.g., Robertson 1998) and the relationships among their carboxy-terminal (ligand binding) domain sequences. Nevertheless, some relatively stable groups of related sequences can be defined among the divergent nematode NRs based on the DBD sequences. In general, similarities among members of a group extend beyond the core DBD through the T box but not the A box region (Fig. 1; data not shown). The T box region of the retinoid X and thyroid hormone receptors contribute to receptor dimerization (Gronemeyer and Laudet 1995; Rastinejad 1997 and references therein). However, the divergent “T boxes” differ from those of RXR and TR and their functional roles remain to be determined.

#### Gene Duplication and Divergence within the Nematode Evolutionary Lineage Contributed to the Abundance and Diversity of NR Genes

Most molecular screens to identify new NR genes with degenerate oligonucleotide probes (e.g., Escriva et al. 1997; Sluder et al. 1997, this work) would have been unlikely to detect many of the divergent NRs revealed by the nematode genome sequences. This raises the question of whether these divergent NRs define classes that are unique to nematodes, or represent classes not yet identified in other phyla because of the bias of the screens performed to date. Our comparative analysis of DBD sequences used all NR sequences from other species available in the main



**Figure 4** Neighbor-joining tree of selected NR core DBD sequences (as defined in Rastinejad 1997) generated by the GCG Growtree program. One thousand neighbor-joining bootstrap replicates were performed on the same data set using the Paup-search function of GCG 9.1. Bootstrap values for supported branches are indicated by hatch marks: (/) 50%-79%; (//) 80-94%; (///) 95-100%. Paup-search was also used to perform maximum parsimony analysis; branchpoints preserved in the resulting consensus tree are denoted by dots. *C. elegans* sequences are included for all NR genes characterized in this or other studies (\* and see Tables 1-3), all genes for which *C. briggsae* orthologs were identified, and selected representatives of other major groupings defined in a larger tree containing all 224 nontruncated nematode DBD sequences. Sequences not yet assigned *nhr* designations are indicated by genome project ORF numbers. Also included are all known *C. briggsae* NR sequences (designated by clone name) and both vertebrate and insect sequences representing the major classes defined by Laudet (1997). Curly brackets indicate phylogenetically conserved NR classes that contain nematode members. Sequence name prefixes denote species of origin: (Aam) the tick *Amblyomma americanum*, (Cb) *C. briggsae*, (Ce) *C. elegans*, (Dm) *Drosophila melanogaster*, (Hs) *Homo sapiens*, (Xl) *Xenopus laevis*. Nematode sequences are available through the Genome Sequencing Project servers and, for *C. elegans*, in GenBank, and can be retrieved by text searches for the cosmid designation. GenBank accession numbers for previously published nematode sequences and non-nematode sequences included are (AF020187), AamEcR; (U37424), CeNHR-2; (U13075), CeNHR-23; (X51548), DmE75; (U01087), DmE78; (M74078), DmEcR; (M63711), DmFTZ-F1; (pir:S36218), DmHNF-4; (M90806), DmHR3; (X89246), DmHR38; (U31517), DmHR78; (U36792), DmHR96; (X14153), DmKNRL; (M28863), DmSVP; (M34639), DmTLL; (X53417), DmUSP; X16155), HsCOUP-TFI; (L29496), HsAR; (X03635), HsER; (X51416), HsERR1; (U64876), HsGCNF; (U22662), HsLXRα; (M16801), HsMR; (X75918), HsNOT; (L02932), HsPPARα; (Z30972), HsPPARγ; (X06538), HsRARα; (M24857), HsRARγ; (X55066), HsReverba; (U04898), HsRORα; (U76388), HsSF-1; (Y13276), HsTL; (L27586), HsTR4; (X55005), HsTRα; (M26747), HsTRβ; (Z37526), XIHNF-4; (L11443), XIRXRγ; (X75163), XIVDLOR; (U91846), XIVDR.

GenBank and European Molecular Biology Laboratory (EMBL) databases, all of which are members of one of the major NR subfamilies defined by Laudet (1997). The separate EST database maintained by GenBank contains >1 million sequence entries and provides a source for identification of novel gene sequences. To determine whether the EST database entries for other species contained NR sequences that would be candidate members of the NR classes represented by the divergent nematode NRs, we searched the database with 25 different *C. elegans* NR peptide sequences representing at least 15 different groups of the divergent NRs. When these searches were per-

formed (April 1998), the EST database contained >900,000 entries for human, >300,000 for mouse, >70,000 for *C. elegans*, >30,000 for *Drosophila*, and smaller numbers of entries for a variety of other species. The only vertebrate NR sequences detected in these searches corresponded to genes already represented in GenBank, and are all members of previously known NR classes. One novel *Drosophila* NR sequence was detected; this appears to be most closely related to the vertebrate estrogen receptor-related orphan receptors (Giguere et al. 1988) and thus is also a member of a previously defined NR class. Therefore at present the divergent NR classes are found only in nematodes, al-

though this may change as genome-sequencing projects for other species progress.

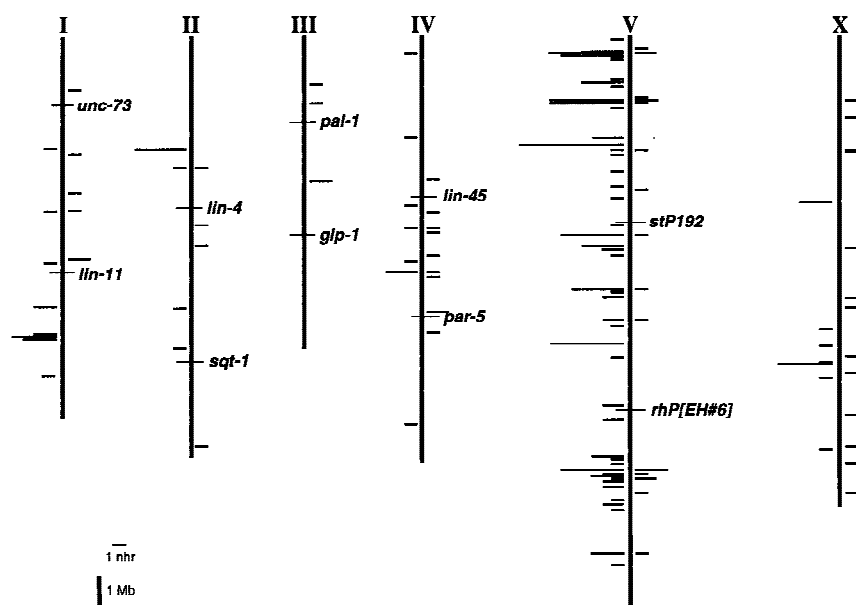
This analysis suggests that the majority of the nematode NRs may have arisen from an extraordinary proliferation of NR genes within the nematode phylogenetic lineage. Consistent with this proposal, comparisons of *C. elegans* and *C. briggsae* NRs indicate that some diversification of nematode NRs has occurred relatively recently. Molecular data suggest that many of the major nematode orders arose by 400 million years ago (Vanfleteren et al. 1994), whereas *C. elegans* and *C. briggsae* diverged between 180 and 40 million years ago (Kennedy et al. 1993), although in the absence of a nematode fossil record these evolutionary times remain uncertain. If the proliferation of NR genes occurred early within nematode evolution, the majority should be represented by orthologs in both *C. elegans* and *C. briggsae*. In this case, with ~85% of the *C. elegans* genome sequence completed, 12 of the 14 *C. briggsae* NRs should have orthologs among the known *C. elegans* genes. However, only eight (57%) of the *C. briggsae* NRs have clear orthologs in *C. elegans*. Furthermore, although the *C. briggsae* NR genes found in physical clusters (two clusters of two genes each) are related to specific *C. elegans* NR sequences, cluster structure is not conserved between the two species. On the basis of these observations, we suggest that although some of the diverged NRs were present in the common ancestor of *C. elegans* and *C. briggsae*, the process of NR proliferation and

diversification also continued after the divergence of the two species.

### The Genomic Distribution of *C. elegans* NR Genes Reveals Extensive Duplication and Diversification on Chromosome V

Over half of the predicted NR genes (145 of 228) lie on Chromosome V, whereas Chromosome III, in contrast, bears only four NR genes (Fig. 5). Although V is physically the largest chromosome (~21.7 Mb, calculated as described in Barnes et al. 1995) and III is the smallest (~11.6 Mb), the difference in physical size is less than twofold and not sufficient to account for the difference in NR gene numbers. The density of NR genes on V (6.7 NR genes/Mb) is ~5-fold greater than that on any other chromosome, and ~20-fold greater than that on III (0.3 NR gene/Mb). The overall densities of ORFs predicted by the genome project, however, vary by less than twofold among the chromosomes (J. Spieth, pers. comm.). Thus, Chromosome V is enriched for predicted NR ORFs relative to the other chromosomes. The presence of pseudogenes is one possible explanation for the abundance of NR sequences on Chromosome V. Consistent with this idea, the proportion of NR sequences on V known to be expressed is lower than on the other chromosomes (14% compared to 29%–100%). However, as noted above, the current cDNA/EST representation may be an unreliable measure of the proportion of NR genes that are expressed, and a more direct survey of NR expression will be needed to address this issue fully.

In addition to the large number of NR sequences on Chromosome V, a number of other observations also suggest that much of the diversification of *C. elegans* NRs has occurred on Chromosome V. All of the NR genes on V encode members of divergent classes, whereas the genes encoding members of phylogenetically conserved classes are distributed among the other chromosomes. Of the nine clusters of five or more tandemly arrayed NR ORFs, eight occur on V, indicating that frequent gene duplication events have contributed to the abundance of NR ORFs on V. Furthermore, two observations indicate that some of the expansion of NR sequences on V occurred after the evolutionary divergence of *C. elegans* and *C. briggsae*. First, both physical clusters of *C. briggsae* NRs are most closely related to *C. elegans* Chromosome V



**Figure 5** Chromosomal distribution of *C. elegans* NR genes. Vertical lines indicate relative physical lengths of chromosomes. Named loci designate the autosomal genetic cluster boundaries defined by Barnes et al. (1995). Histograms to the right represent NR genes with known genetic function or known to be expressed as mRNA. Those to the left represent NR sequences for which expression or genetic function has yet to be demonstrated. Physical distances were calculated as described in Barnes et al. (1995).

sequences, but, as noted above, cluster structure is not conserved between the two species. Second, only two (25%) of the eight *C. elegans* genes for which clear *C. briggsae* orthologs have been identified are on Chromosome V, compared to the 64% of all *C. elegans* NRs found on this chromosome. Although the current sample size is too small to be definitive, this suggests that *C. elegans* Chromosome V sequences may be more diverged from *C. briggsae* sequences than is true for the genome at large. It will be of interest to learn whether Chromosome V sequences other than NRs exhibit a similar diversity. If they do, one prediction is that as analysis of the *C. briggsae* genome progresses less synteny will be observed with Chromosome V than with other regions of the *C. elegans* genome.

## DISCUSSION

The nuclear receptors, with 228 predicted genes, constitute the largest family of transcriptional regulators encoded in the *C. elegans* genome (Waterston et al. 1997; J. Spieth, pers. comm.). This is also the largest number of NR genes yet described from a single species, revealing a previously unobserved breadth of variation within the NR superfamily. Members of both phylogenetically conserved and novel diverged NR classes are found in the 228 genes, although the majority are divergent NRs. Although members of the new NR classes defined by the *C. elegans* genes have not yet been reported from other metazoan phyla, many of these divergent classes are likely to have been missed in the molecular screens for NR sequences that have been performed. Thus, these or other as yet unobserved NR classes could remain to be discovered in other metazoans, and the nature of any new NR genes revealed as other metazoan genome sequencing projects progress will be of great interest.

The most thorough analysis of the evolution of the NR superfamily to date proposed that two distinct rounds of NR gene duplication and divergence have occurred—one early in metazoan evolution, before the divergence of protostomes and deuterostomes, and a second that expanded a subset of NR classes in the vertebrate lineage (Laudet 1997). Our observations on the NR superfamily in nematodes support the hypothesis of an early presence of several ancient NR classes shared among most, and perhaps all, metazoans. In addition, our analysis indicates that extensive duplication and diversification of NR genes has occurred within the nematode lineage, resulting in an abundance of divergent NR sequences that may be unique to nematodes. Although orthologs of some divergent nematode NRs may yet be identified in other phyla, two observations are consistent with the hypothesis that at least a portion of the observed NR diversity is nematode specific. First, at least two physical clusters

of divergent NR genes appear to have been expanded differentially in the genomes of *C. elegans* and the sibling species *C. briggsae*, suggesting that NR gene duplication and divergence continued after evolutionary divergence of these two species. Second, in *C. elegans* there is a striking concentration of divergent NRs on a single chromosome, and current data are consistent with a relatively rapid evolution of these NR sequences. A sequencing project that has begun recently for the parasitic nematode *Brugia malayi* (Blaxter et al. 1996) should provide additional information regarding the extent of NR diversification within different branches of the nematode lineage.

Phylogenetic analyses of the metazoans have traditionally concluded that the nematodes, along with other pseudocoelomates, separated from other metazoans before the divergence of the protostome and deuterostome lineages, although in the absence of a fossil record for nematodes, this placement remains under some debate (see Fitch and Thomas 1997 and references therein). A recent alternate model proposes molting as a synapomorphy for placing nematodes in an “Ecdysozoa” clade along with arthropods and other protostomes (Aguinaldo et al. 1997). If molting does represent a defining shared evolutionary trait, key elements of the genetic circuitry regulating the molting process should be found in both insects and nematodes. In *Drosophila melanogaster* the receptor for the molting hormone 20-hydroxyecdysone is a heterodimer of two NR proteins, EcR and ultraspiracle (Yao et al. 1992, 1993; Thomas et al. 1993). Strikingly, no apparent members of the EcR or ultraspiracle NR gene classes have yet been found in *C. elegans*, although vertebrate genes related to both EcR (FXR; Forman et al. 1995a) and ultraspiracle (RXR; Oro et al. 1990) are known. Such genes may occur in the ~15% of the *C. elegans* genome for which the sequence is yet to be completed, but no strong candidates were detected in the portions of sequence in progress available as of July 1, 1998, via the Genome Sequencing Project’s BLAST server. In addition, several direct screening strategies have failed to identify *C. elegans* EcR or ultraspiracle orthologs (C. Maina, unpubl.). The absence of a *C. elegans* ultraspiracle/RXR ortholog, if true, is particularly puzzling as the ultraspiracle/RXR class appears to be one of the more ancient NR classes (Escriva et al. 1997).

*C. elegans* does not produce ecdysteroids (Barker et al. 1990; Chitwood and Feldlaufer 1990), and any hormonal signal used in the regulation of molting is likely to be different from arthropod ecdysteroids. Thus if molting is a synapomorphy, the crucial conserved regulatory genes may be involved in execution of the molt rather than reception of a hormonal signal. The cellular response cascade activated by the ecdysone receptor includes additional NRs (Thummel 1995), and

*C. elegans* genes related to several of these have been identified (Fig. 4). *nhr-6*, *nhr-23*, and *nhr-25* are clear members of NR classes containing the *Drosophila* genes *DHR38* (Kozlova et al. 1998), *DHR3* (Koelle et al. 1992), and *ftz-f1* (Lavorgna et al. 1991, 1993), respectively. Three *C. elegans* genes — *daf-12*, *nhr-8*, and *nhr-48* — are most similar to the ecdysone responsive *DHR96* gene (Fisk and Thummel 1995). *sex-1* is the closest known *C. elegans* relative of two *Drosophila* ecdysone-inducible genes *E75* and *E78*, although it is not a clear ortholog of either gene (Laudet 1997). The *sex-1* mutant phenotype does not reveal any role in the regulation of molting (Carmi et al. 1998), but others of these genes are candidates to participate in the molting process. Notably, disruption of *nhr-23* function leads to defects in molting (Kostrouchova et al. 1998). *daf-12* mutants fail to progress properly through the later larval developmental stages, instead inappropriately repeating aspects of earlier stages (Antebi et al. 1998). This *daf-12* mutant phenotype is reminiscent of the reiteration of earlier stage-specific characteristics by lepidopterans exposed to exogenous juvenile hormone during larva-to-pupa or pupa-to-adult molts (for review see Riddiford 1994). Although *nhr-23* and *daf-12* provide intriguing parallels to the regulation of molting in insects, a full evaluation of the evolutionary significance of these parallels will require a more complete understanding of the control of molting in nematodes.

Two additional questions of considerable interest for understanding the roles of the NR superfamily in both evolution and nematode biology are whether any of the 228 *C. elegans* NRs are ligand-regulated and the identities of any ligands. As with the DBDs, the *C. elegans* LBDs are more diverse than those yet reported from other species, and a full analysis of the nematode LBD sequences will be presented elsewhere. Preliminary analysis indicates that ~30% of the nematode NRs, including all members of phylogenetically conserved NR classes except *nhr-48*, exhibit similarity to NR LBDs from other phyla, although the similarities are not sufficient to permit strong predictions regarding the identity of any ligands (A. Sluder, unpubl.). Strikingly, <10% of the divergent NR sequences on Chromosome V exhibit such LBD similarity.

The diversity of the nematode LBD sequences provides the potential for binding an equally diverse array of ligands. An organism with the small size and rapid life cycle of *C. elegans* seems an unlikely candidate to use enough hormones, in the classic sense, to use such a large number of receptors. One possibility is that many of the nematode NRs are not ligand-regulated, serving instead as “mere” transcriptional regulators. On the basis of currently known ligand/NR interactions, ligand binding appears to have evolved multiple independent times from a primordial NR that was not ligand regulated (Escriva et al. 1997), consistent with

the view that many orphan receptors do not bind ligands.

An alternative view of NR evolution, summarized by Yamamoto (1997), is that NRs “evolved in metazoans specifically to exploit simple lipophilic molecules for intercellular signaling,” and furthermore that each NR “will bind to a small metabolite, a nutrient, an environmental compound, that has acquired a signaling role.” Consistent with this prediction, a number of vertebrate NRs are specifically activated by metabolic products with previously undiscerned signaling functions (Forman et al. 1995a,b, 1998; Kliewer et al. 1995; Janowski et al. 1996; Lehmann et al. 1997; Blumberg et al. 1998). The ability to adjust its repertoire of active metabolic pathways for optimal utilization of whatever food source is at hand would be an advantage for *C. elegans* in its ecological niche of opportunistic soil-dwelling bacteriovore. NR proteins are well suited to contribute to such adjustment, with the capacity to regulate specific groups of genes in response to the presence of a ligand either encountered in the environment or produced from the catabolism of a particular food source. The diversity and number of NR genes, and also of predicted chemoreceptor genes (Robertson 1998), could be major components of the genetic sophistication permitting *C. elegans* to exploit its environment for successful growth and reproduction. One expectation arising from this model is that nematodes adapted to different life styles, such as parasitism, will have a subset of NRs specific for their environment and not shared with *C. elegans*.

To date only five mutationally defined loci have been shown to correspond to NR genes (Tables 1 and 2). Several factors could be contributing to this low representation of NRs among known genetic loci. First, as noted earlier, some of the predicted NR ORFs may be pseudogenes with no genetic function. Second, as transcription factors with the potential for heterodimerization, some NRs are likely to function in multiple developmental contexts. The mutant phenotypes resulting from disruption of the genes for these will be pleiotropic and are perhaps unlikely to have attracted the attention of workers interested in specific events. Many of these genes may be represented in collections of relatively uncharacterized lethal mutants (e.g., Hirsh and Vanderslice 1976; Meneely and Herman 1979; Cassada et al. 1981; Rogalski et al. 1982; Rosenbluth et al. 1988; Howell and Rose 1990; Johnsen and Baillie 1991; McKim et al. 1992). Third, more than one related NR may be capable of fulfilling some functions, therefore loss of a single gene may have little or no visible effect on the animal. Such genetic redundancy has been observed for the RXRs in mammals (Krezel et al. 1996). Fourth, NRs involved in the differentiation or function of a small number of cells may exhibit subtle mutant phenotypes unlikely to have been detected in

many of the phenotypic screens that have been done. Notably, mutations in three of the genetic loci known to encode NRs — *fax-1*, *odr-7*, and *unc-55* — each affect only a subset of neurons (Sengupta et al. 1994; Walthall and Plunkett 1995; Wightman et al. 1997). And finally, the mutant phenotypes of some NR genes, such as any involved in metabolic adaptation to environmental conditions, may not be apparent under laboratory culture conditions. Mutations in these genes may have little effect even under the majority of growth conditions in the wild, exhibiting a “latent selection potential” (Kimura 1990; Meagher 1995) that is realized only under specific circumstances.

Although they represent the largest family of predicted transcriptional regulators encoded in the *C. elegans* genome, the NRs are a largely untapped source of insight to nematode biology. Characterization of the expression and genetic function of these genes promises to contribute significantly to our understanding of many aspects of development and physiology. Furthermore, the diversity revealed by the *C. elegans* NR sequences provides an expanded context for evaluating the evolution of this family of metazoan transcription factors. In particular, identification of NRs shared among nematode species but not present in other phyla may provide a foundation for development of new strategies for combating parasitic nematode infections.

## METHODS

### General Methods

Unless otherwise noted, all molecular biology procedures used standard methods (Sambrook et al. 1989). DNA sequencing and oligonucleotide synthesis were carried out at the New England Biolabs core facility or the Molecular Genetics Instrumentation Facility at the University of Georgia.

### Databases Used

In addition to the GenBank and dbEST databases maintained by the NCBI, the analyses reported here used the genome sequence databases of the *C. elegans* Genome Project. These databases are accessible for both BLAST searching and sequence retrieval through servers maintained by the Genome Sequencing Centers at Washington University, St. Louis, MO (<http://genome.wustl.edu/gsc/gsc/gschmpg.html>) and at the Sanger Center, Cambridge, England ([http://www.sanger.ac.uk/Projects/C\\_elegans/](http://www.sanger.ac.uk/Projects/C_elegans/)). Genome sequence data for *C. briggsae* are also available through the Washington University server. Locations of clones on the physical map of the *C. elegans* genome were determined from the ACeDB database (Eeckman and Durbin 1995), data release WS4 4-26 (2/98). NR sequences were identified by BLAST homology searches (Altschul et al. 1990) of the databases using the amino acid sequence of the NHR-2 DBD (Sluder et al. 1997).

### Oligonucleotide Primers

Degenerate oligonucleotide primers were designed based on

the DBD sequences of the *Drosophila* EcR (Koelle et al. 1991) and Ultraspiracle (Oro et al. 1990) proteins or of an *ultraspiracle*-related gene from the dog heartworm *D. immitis* (primers 137–126 and 141–37; C. Maina, unpubl.). Degeneracies were accomplished either by incorporation of a mixture of two nucleotides at a given position or by use of inosine to approximate four-fold degeneracy. Additional sequences added at the 5' ends of the oligonucleotides provided restriction sites for cloning of amplified products. Sequences of the primers used were as follows (the corresponding amino acid sequences are noted in brackets): C1, 5'-ACAGAATTCTG(C/T)GA(A/G)GGITG(C/T)AA(A/G)GGITT(C/T)TT-3' [CEGCKGFF]; C2, 5'-ACAGGATCCATCICACIGCIA(A/G)(A/G)CA(C/T)TT(C/T)TT-3' [reverse KKCLAVGM]; C3, 5'-ACAGGATCCT(C/T)TG(A/G)TAIC(T/G)(A/G)CA(A/G)TA(C/T)TG-3' [reverse CQYCRYQK]; C4, 5'-ACAGAATTCTGA(A/G)(C/T)TITG(C/T)TIGTITG(C/T)GGIGA-3' [ELDLVDGC]; 137–126, 5'-ACAGGATCCCATICCCATI(G/C)(A/T)C(T)TG(A/G)CA(C/T)TTIC-3' [reverse RKCQSMGM]; 141–37, 5'-ACAGGATCCGGIGTITA(C/T)(A/T)G(C)ITG(C/T)GA(A/G)GGITG-3' [GVYCEGC].

Primers for reverse transcription–polymerase chain reaction amplification were designed based on known (*nhr-6*) or predicted (*nhr-9*, *nhr-10*, and *nhr-20*) splicing patterns. *nhr-6*–CD-1, 5'-CCAGAGGTAGGGTCAGGGACAACG-3'; *nhr-6*–CD-2, 5'-GGCGCGAATGCCTGAGGCTCCC-3'; *nhr-6*–CD-3, 5'-CCTGGTTCGACGAGAGGTTTGTGG-3'; *nhr-6*–CD-4, 5'-CCTGCAACCGCCGACCCGATTACG-3'. *nhr-9*–PEP9-5, 5'-CTCAACTTCTCTGAACGCC-3'; *nhr-9*–PEP9-6, 5'-AAGCACAACTGCCCAATATAC-3'; *nhr-9*–PEP9-7, 5'-GTTCCATAAATCCCATTCC-3'. *nhr-10*–PEP10-5, 5'-CGTCATCAAATAACTC-GTCTC-3'; *nhr-10*–PEP10-6, 5'-CTTCTTTTCATTTCAACAGCC-3'; *nhr-10*–PEP10-7, 5'-CACTCATCACTGGCTTCAAC-3'. *nhr-20*–PEP20-5, 5'-ATTCCACCCACATCCAATG-3'; *nhr-20*–PEP20-6, 5'-AATAGTTCGAAAAGTTCCGCC-3'; *nhr-20*–PEP20-7, 5'-TCCAGAGCATTCAAACCTGCAC-3'.

### Identification of *C. elegans* NR Sequences by PCR Amplification Using Degenerate Primers

Candidate NR sequences were amplified from *C. elegans* genomic DNA using the degenerate primers described above. Each 100- $\mu$ l amplification reaction included 200–400 ng *C. elegans* genomic DNA; 200  $\mu$ M each dATP, dCTP, dGTP, and dTTP; 1 unit *Taq* DNA polymerase; and 100 pmoles of each primer. After a 4-min denaturation at 94°C, reactions were cycled through 40 repetitions of 1 min at 94°C, 1 min at the selected annealing temperature, and 3 min at 72°C, followed by a final incubation at 72°C for 5 min. For each primer pair, annealing temperatures were varied in 5°C increments from 25°C to 55°C. Products that were reproducibly amplified at any specific temperature were gel purified, digested with appropriate restriction enzymes, cloned into the Litmus38 vector (New England Biolabs), and sequenced. *nhr-3*, *nhr-4*, and *nhr-5* were identified using the primer pair C1/C2, *nhr-6* and *nhr-7* with C1/C3, *nhr-17* and *nhr-18* with C4/C2, and *nhr-41* with 141-37/137-126.

Cloned products verified as NR sequences were used to probe filters of gridded YAC clones representing ~95% of the *C. elegans* genome (obtained from A. Coulson, Sanger Center, Cambridge, England) and to screen a mixed stage *C. elegans* cDNA library (Stratagene). cDNAs were identified for *nhr-3*, *nhr-4*, *nhr-6*, and *nhr-18*, and the DNA sequence of each clone was determined. The GenBank accession num-

bers are, respectively, AF083222, AF083223, AF083224, and AF083232.

### Expressed Sequence Tag Clones

NR ESTs were identified by searching the dbEST database maintained by NCBI. The corresponding cDNA clones (see Table 1) were obtained from The Institute for Genomic Research, Bethesda, MD (CEESV11 and CEESU44), the Genome Sequencing Center, Washington University, St. Louis, MO (all clones beginning with "cm"), or Y. Kohara at the National Institute of Genetics, Japan (clones beginning with "yk"). The complete DNA sequences of these clones was determined. We found that two separate samples of the clone cm06h9 did not yield sequence corresponding to that reported in dbEST or encode an NR. The cm11f5 cDNA sequence for *nhr-24/sex-1* revealed no differences from a previously reported cDNA (GenBank U13074; Kostrouch et al. 1995). Accession numbers for the other completed EST sequences are AF083225 through AF083231, AF083233, and AF083234.

### Isolation of RNA from Staged Cultures

Staged cultures of the *C. elegans* wild-type N2 strain (Brenner 1974) were grown as previously described (Sluder et al. 1997). Total RNA was purified by LiCl precipitation (Ausubel et al. 1991) or with Tri Reagent (Molecular Research Center, Inc., Cincinnati, OH). For the latter, worms were vortexed for 10 min in Tri Reagent with a small quantity of sand (Sigma catalog S-9887) to break the cuticles. The manufacturer's protocol for extraction of RNA from tissues was then followed. Selection of poly(A)<sup>+</sup> RNA was as described (Sambrook et al. 1989).

### RT-PCR Amplification

First strand cDNA was produced by reverse transcription of 1 µg poly(A)<sup>+</sup> RNA with MMLV reverse transcriptase (Epicentre Technologies, Madison, WI). After phenol extraction and ethanol precipitation, cDNA products were resuspended in 10 mM Tris (pH 7.4)/1 mM EDTA. PCR amplification of 1/50 of the cDNA preparation used standard reaction conditions (Sambrook et al. 1989). After an initial denaturation step at 94°C for 2 min, 35 amplification cycles of 93°C for 1 min, 60°C for 1 min, and 72°C for 4 min were performed, followed by a final incubation at 72°C for 3 min. Secondary amplifications were performed using 1/100 of the original reaction product as template.

The following primer pairs were used for amplification. For *nhr-6*, initial amplification was with CD-3 and CD-4, secondary amplification with CD-1 and CD-2; the final genomic amplification product is 1.0 kb and the predicted product from cDNA is 898 bp. For *nhr-9*, initial amplification was with PEP9-5 and PEP9-6, secondary amplification with PEP9-5 and PEP9-7; product sizes are 800 bp for genomic DNA and 420 bp for cDNA. For *nhr-10*, initial amplification was with PEP10-5 and PEP10-6, secondary amplification with PEP10-5 and PEP10-7; product sizes are 1.0 kb for genomic DNA and 330 bp for cDNA. For *nhr-20*, initial amplification was with PEP20-5 and PEP20-6, secondary amplification with PEP20-5 and PEP20-7; product sizes are 600 bp for genomic DNA and 340 bp for cDNA.

Several observations support the specificity of the RT-PCR reactions. The *nhr-6* cDNA was isolated and characterized at New England Biolabs, whereas the RT-PCR reactions were

performed at the University of Georgia, thus minimizing the potential for contamination. As a further guard against cross-contamination of samples, the pipettors used to set up the amplification reactions were never used to pipette cloned DNAs or PCR products. Because no cDNAs have been isolated for *nhr-9*, *nhr-10*, and *nhr-20*, the observed RT-PCR products must have arisen from sequences present in the mRNA preparations used. In addition, the amplification of genomic DNA sequences with each primer set confirmed the reaction conditions and controlled for any cDNA contamination of reagents. And finally, each RT-PCR amplification product was gel-purified and a single DNA sequencing reaction performed for each. All RT-PCR products corresponded to the appropriate gene sequences, although the single reactions performed did not resolve the nature of the differences between the two *nhr-6* products.

### Sequence Analysis and Comparisons

Sequence assembly and analysis were performed using the GCG program package (Devereux et al. 1984). Prediction of candidate splice sites in both *C. elegans* and *C. briggsae* genomic sequences was aided by the GCG fitconsensus program using consensus tables derived from the data presented by Fields (1990). Synteny between *C. briggsae* and *C. elegans* genomic clones was assessed by searching *C. elegans* genomic sequences in GenBank with the *C. briggsae* sequence using the gapped alignment function of BLAST 2.0 (<http://www.ncbi.nlm.nih.gov/BLAST/>). Regions of similarity yielding BLAST scores >80 were considered significant. Regions were not considered syntenic unless multiple blocks of aligned similarity occurred outside the NR sequences, and for six of the eight syntenic regions thus identified the alignments extended throughout the length of the *C. briggsae* fosmid clone being assessed.

Neighbor-joining trees of NR DNA-binding sequences were produced using the GCG Pileup, Distances, and Growtree programs. To obtain reliable alignments of the zinc-coordinating cysteine residues, we modified the rescaled Dayhoff scoring matrix (Gribskov and Burgess 1986) of the GCG software package to weight cysteine-cysteine alignments more heavily than other matches, assigning the arbitrary high score of 45 to cysteine-cysteine matches. This modified scoring matrix was used in all analyses. Neighbor-joining bootstrap analysis (1000 replicates) and maximum parsimony analysis were performed using the PAUPsearch feature of GCG version 9.1.

### ACKNOWLEDGMENTS

We gratefully acknowledge the *C. elegans* Genome Sequencing Project and Yuji Kohara for making much sequence information publicly available. John Spieth was particularly helpful in matters relating to the genome sequence. Adam Antebi, Barry Honda, Pam Larsen, Barbara Meyer, Don Riddle, Bill Walthall, and Bruce Wightman generously communicated results before publication. We thank Barton Slatko, Laurie Mazola, Jennifer Ware, and Mahul Ganatra for DNA sequencing, Marion Sibley for technical assistance in the cloning of *nhr-41*, Tim Lindblom for assistance with the sequencing of RT-PCR products, and Michael Weise for invaluable computer system support. A.E.S. thanks Judy Willis and Rich Meagher for many thought-provoking discussions concerning evolution and phylogeny. C.V.M. thanks Don Comb for encour-

agement and support and George Tzertzinis for many helpful discussions. This work was supported in part by grants to A.E.S. from the March of Dimes Birth Defects Foundation and the University of Georgia Research Foundation.

The publication costs of this article were defrayed in part by payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 USC section 1734 solely to indicate this fact.

## NOTE ADDED IN PROOF

The WWW address for supplementary information on nematode NRs has been changed to [http://cellmate.cb.uga.edu/cellbiology/nematode\\_nuclear\\_hormone\\_recepto.htm](http://cellmate.cb.uga.edu/cellbiology/nematode_nuclear_hormone_recepto.htm).

## REFERENCES

- Aguinaldo, A.M.A., 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.
- Antebi, A., J.G. Culotti, and E.M. Hedgecock. 1998. *daf-12* regulates developmental age and the dauer alternative in *Caenorhabditis elegans*. *Development* **125**: 1191–1205.
- Arriza, J.L., C. Weinberger, G. Cerelli, T.M. Glaser, B.L. Handelin, D.E. Housman, and R.M. Evans. 1987. Cloning of human mineralocorticoid receptor complementary DNA: Structural and functional kinship with the glucocorticoid receptor. *Science* **237**: 268–275.
- Ausubel, F.M., R. Brent, R.E. Kingston, D.D. Moore, J.G. Seidman, J.A. Smith, and K. Struhl, eds. 1991. *Current protocols in molecular biology*. John Wiley & Sons, New York, NY.
- Ayala, F.J., A. Rzhetsky, and F.J. Ayala. 1998. Origin of the metazoan phyla: Molecular clocks confirm paleontological estimates. *Proc. Natl. Acad. Sci.* **95**: 606–611.
- Barker, G.C., D.J. Chitwood, and H.H. Rees. 1990. Ecdysteroids in helminths and annelids. *Invert. Reprod. Dev.* **18**: 1–11.
- Barnes, T.M., Y. Kohara, A. Coulson, and S. Hekimi. 1995. Meiotic recombination, noncoding DNA and genomic organization in *Caenorhabditis elegans*. *Genetics* **141**: 159–179.
- Blaxter, M.L., N. Raghavan, I. Ghosh, D. Guiliano, W. Lu, S.A. Williams, B. Slatko, and A.L. Scott. 1996. Genes expressed in *Brugia malayi* infective third stage larvae. *Mol. Biochem. Parasitol.* **77**: 77–93.
- Blumberg, B., H. Kang, J. Bolado Jr., H. Chan, A.G. Craig, T.A. Moreno, K. Umenso, T. Perlmann, E.M. De Robertis, and R.M. Evans. 1998. BXR, an embryonic orphan nuclear receptor activated by a novel class of endogenous benzoate metabolites. *Genes & Dev.* **12**: 1269–1277.
- Blumenthal, T. and K. Steward. 1997. RNA processing and gene structure. In *C. elegans II* (ed. D.L. Riddle, T. Blumenthal, B.J. Meyer, and J.R. Priess), pp. 117–145. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
- Brenner, S. 1974. The genetics of *Caenorhabditis elegans*. *Genetics* **77**: 71–94.
- Carmi, I., J.B. Kocczynski, and B.J. Meyer. 1998. The nuclear hormone receptor SEX-1 is an X-chromosome signal that determines nematode sex. *Nature* **396**: 168–173.
- Cassada, R., E. Isnenghi, M. Culotti, and G. von Ehrenstein. 1981. Genetic analysis of temperature-sensitive embryogenesis mutants in *Caenorhabditis elegans*. *Dev. Biol.* **84**: 193–205.
- Chitwood, D.J. and M.F. Feldlaufer. 1990. Ecdysteroids in axenically propagated *Caenorhabditis elegans* and culture medium. *J. of Nematol.* **22**: 598–607.
- Coulson, A., C. Huynh, Y. Kozono, and R. Shownkeen. 1995. The physical map of the *Caenorhabditis elegans* genome. In *Caenorhabditis elegans: Modern biological analysis of an organism* (ed. H.F. Epstein and D.C. Shakes), Vol. 48, pp. 533–550. Academic Press, San Diego, CA.
- Damiani, R.D., Jr. and S.R. Wessler. 1993. An upstream open reading frame represses expression of *Lc*, a member of the *R/B* family of maize transcriptional activators. *Proc. Natl. Acad. of Sci.* **90**: 8244–8248.
- de Bono, M. and J. Hodgkin. 1996. Evolution of sex determination in *Caenorhabditis*: Unusually high divergence of *tra-1* and its functional consequences. *Genetics* **144**: 587–595.
- Devereux, J., P. Haerberli, and O. Smithies. 1984. A comprehensive set of sequence analysis programs for the VAX. *Nucleic Acids Res.* **12**: 387–395.
- Eeckman, F.H. and R. Durbin. 1995. ACeDB and Macace. In *Caenorhabditis elegans: Modern biological analysis of an organism* (ed. H.F. Epstein and D.C. Shakes), Vol. 48, pp. 583–605. Academic Press, San Diego, CA.
- Escriva, H., R. Safi, C. Hanni, 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. of Sci.* **94**: 6803–6808.
- Evans, R.M. 1988. The steroid and thyroid hormone receptor superfamily. *Science* **240**: 889–895.
- Fields, C. 1990. Information content of *Caenorhabditis elegans* splice site sequences varies with intron length. *Nucleic Acids Res.* **18**: 1509–1512.
- Fisk, G.J. and C.S. Thummel. 1995. Isolation, regulation, and DNA-binding properties of three *Drosophila* nuclear hormone receptor superfamily members. *Proc. Natl. Acad. of Sci.* **92**: 10604–10608.
- Fitch, D.H.A. and W.K. Thomas. 1997. Evolution. In *C. elegans II* (ed. D.L. Riddle, T. Blumenthal, B.J. Meyer and J.R. Priess), pp. 815–850. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
- Forman, B.M. 1997. Orphan nuclear receptors and their ligands. In *Molecular biology of steroid and nuclear hormone receptors* (ed. L.P. Freedman), pp. 281–305. Birkhauser, Boston, MA.
- Forman, B.M., E. Goode, J. Chen, A.E. Oro, D.J. Bradley, T. Perlmann, D.J. Noonan, L.T. Burka, T. McMorris, W.W. Lamph, R.M. Evans, and C. Wienberger. 1995a. Identification of a nuclear receptor that is activated by farnesol metabolites. *Cell* **81**: 687–693.
- Forman, B.M., P. Tontonoz, J. Chen, R.P. Brun, B.M. Spiegelman, and R.M. Evans. 1995b. 15-Deoxy-Delta12,14-prostaglandin J2 is a ligand for the adipocyte determination factor PPARgamma. *Cell* **83**: 803–812.
- Forman, B.M., I. Tzamelis, H.-S. Choi, J. Chen, D. Simha, W. Seol, R.M. Evans, and D.D. Moore. 1998. Androstane metabolites bind to and deactivate the nuclear receptor CAR-beta. *Nature* **395**: 612–615.
- Geballe, A.P. and D.R. Morris. 1994. Initiation codons within 5'-leaders of mRNAs as regulators of translation. *Trends Biochem. Sci.* **19**: 159–164.
- Giguere, V., E.S. Ong, P. Segui, and R.M. Evans. 1987. Identification of a receptor for the morphogen retinoic acid. *Nature* **330**: 624–629.
- Giguere, V., N. Yang, P. Segui, and R.M. Evans. 1988. Identification of a new class of steroid hormone receptors. *Nature* **331**: 91–94.
- Gribskov, M. and R.R. Burgess. 1986. Sigma factors from *E. coli*, *B. subtilis*, and phage *SPO1*, and phage *T4* are homologous proteins. *Nucleic Acids Res.* **14**: 6745–6763.
- Gronemeyer, H. and V. Laudet. 1995. Transcription factors 3: Nuclear receptors. *Protein Profile* **2**: 1173–1308.
- Hinnebusch, A.G. 1997. Translational control of yeast *GCN4*: A window on factors that control initiator-tRNA binding to the ribosome. *J. Biol. Chem.* **272**: 21661–21664.
- Hirsh, D. and R. Vanderslice. 1976. Temperature-sensitive developmental mutants of *Caenorhabditis elegans*. *Dev. Biol.* **49**: 220–235.
- Hodgkin, J. 1997. Appendix 1: Genetics. In *C. elegans II* (ed. D.L.

- Riddle, T. Blumenthal, B.J. Meyer, and J.R. Priess), pp. 881–1047. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
- Howell, A.M. and A.M. Rose. 1990. Essential genes in the *hDf6* region of Chromosome I in *Caenorhabditis elegans*. *Genetics* **126**: 583–592.
- Janowski, B.A., P.J. Willy, T. Rama-Devi, J.R. Falck, and D.J. Mangelsdorf. 1996. An oxysterol signaling pathway mediated by the nuclear receptor LXR $\alpha$ . *Nature* **383**: 728–731.
- Johnsen, R.C. and D.L. Baillie. 1991. Genetic analysis of a major segment [LGV(left)] of the genome of *Caenorhabditis elegans*. *Genetics* **129**: 735–752.
- Kastner, P., M. Mark, and P. Chambon. 1995. Nonsteroid nuclear receptors: What are genetic studies telling us about their role in real life? *Cell* **83**: 859–869.
- Kennedy, B.P., E.J. Aamodt, F.L. Allen, M.A. Chung, M.F.P. Heschl, and J.D. McGhee. 1993. The gut esterase gene (*ges-1*) from the nematodes *Caenorhabditis elegans* and *Caenorhabditis briggsae*. *J. Mol. Biol.* **229**: 890–908.
- Kimura, M. 1990. The present status of neutral theory. In *Population biology of genes and molecules* (ed. N. Takahata and J.F. Corw), pp. 1–16. Baifukan Co., Tokyo, Japan.
- Kliwer, S.A., J.M. Lenhard, T.M. Willson, I. Patel, D.C. Morris, and J.M. Lehmann. 1995. A prostaglandin J2 metabolite binds peroxisome proliferator-activated receptor  $\gamma$  and promotes adipocyte differentiation. *Cell* **83**: 813–819.
- Koelle, M.R., W.S. Talbot, W.A. Segraves, M.T. Bender, P. Cherbas, and D.S. Hogness. 1991. The *Drosophila* EcR gene encodes an ecdysone receptor, a new member of the steroid receptor superfamily. *Cell* **67**: 59–77.
- Koelle, M.R., W.A. Segraves, and D.S. Hogness. 1992. DHR3 — A *Drosophila* steroid receptor homolog. *Proc. Natl. Acad. Sci.* **89**: 6167–6171.
- Kostrouch, Z., M. Kostrouchova, and J.E. Rall. 1995. Steroid/thyroid hormone receptor genes in *Caenorhabditis elegans*. *Proc. Natl. Acad. Sci.* **92**: 156–159.
- Kostrouchova, M., M. Krause, Z. Kostrouch, and J.E. Rall. 1998. CHR3: A *Caenorhabditis elegans* orphan nuclear hormone receptor required for proper epidermal development and molting. *Development* **125**: 1617–1626.
- Kozak, M. 1991. An analysis of vertebrate mRNA sequences: Intimations of translational control. *J. Cell Biol.* **115**: 887–903.
- Kozlova, T., G.V. Pokholkova, G. Tzertzinis, J.D. Sutherland, I.F. Zhimulev, and F.C. Kafatos. 1998. *Drosophila* hormone receptor 38 functions in metamorphosis: A role in adult cuticle formation. *Genetics* **149**: 1465–1475.
- Krause, M. and D. Hirsh. 1987. A trans-spliced leader sequence on actin mRNA in *C. elegans*. *Cell* **49**: 753–761.
- Krezel, W., V. Dupe, M. Mark, A. Dierich, P. Kastner, and P. Chambon. 1996. RXR $\gamma$  null mice are apparently normal and compound RXR $\alpha$ +/-/RXR $\beta$ -/-/RXR $\gamma$ -/- mutant mice are viable. *Proc. Natl. Acad. Sci.* **93**: 9010–9014.
- Kuwabara, P.E. 1996. Interspecies comparison reveals evolution of control regions in the nematode sex-determining gene *tra-2*. *Genetics* **144**: 597–607.
- Kuwabara, P.E. and S. Shah. 1994. Cloning by synteny: Identifying *C. briggsae* homologues of *C. elegans* genes. *Nucleic Acids Res.* **22**: 4414–4418.
- Laudet, V. 1997. Evolution of the nuclear receptor superfamily: Early diversification from an ancestral orphan receptor. *J. Mol. Endocrinol.* **19**: 207–226.
- Laudet, V., C. Hanni, J. Coll, F. Catzeflis, and D. Stehelin. 1992. Evolution of the nuclear receptor gene superfamily. *EMBO J.* **11**: 1003–1013.
- Lavorgna, G., H. Ueda, J. Clos, and C. Wu. 1991. FTZ-F1, a steroid hormone receptor-like protein implicated in the activation of *fushi tarazu*. *Science* **252**: 848–851.
- Lavorgna, G., F.D. Karim, C.S. Thummel, and C. Wu. 1993. Potential role for a FTZ-F1 steroid-receptor superfamily member in the control of *Drosophila* metamorphosis. *Proc. Natl. Acad. Sci.* **90**: 3004–3008.
- Lehmann, J.M., S.A. Kliwer, L.B. Moore, T.A. Smith-Oliver, B.B. Oliver, J.L. Su, S.S. Sundseth, D.A. Winegar, D.E. Blanchard, T.A. Spender, and T.M. Willson. 1997. Activation of the nuclear receptor LXR by oxysterols defines a new hormone response pathway. *J. Biol. Chem.* **272**: 3354–3361.
- Lohmer, S., M. Maddaloni, M. Motto, F. Salamini, and R.D. Thompson. 1993. Translation of the mRNA of the maize transcriptional activator *Opaque-2* is inhibited by upstream open reading frames present in the leader sequence. *Plant Cell* **5**: 65–73.
- Mangelsdorf, D.J. and R.M. Evans. 1995. The RXR heterodimers and orphan receptors. *Cell* **83**: 841–850.
- Mangelsdorf, D.J., C. Thummel, M. Beato, P. Herrlich, G. Schütz, K. Umesono, B. Blumberg, P. Kastner, M. Mark, P. Chambon, and R.M. Evans. 1995. The nuclear receptor superfamily: The second decade. *Cell* **83**: 835–839.
- McCombie, W.R., M.D. Adams, J.M. Kelley, M.G. FitzGerald, T.R. Utterback, M. Khan, M. Dubnick, A.R. Kerlavage, J.C. Venter, and C. Fields. 1992. *Caenorhabditis elegans* expressed sequence tags identify gene families and potential disease gene homologues. *Nature Genet.* **1**: 124–131.
- McKim, K.S., T. Starr, and A.M. Rose. 1992. Genetic and molecular analysis of the *dpy-14* region in *Caenorhabditis elegans*. *Mol. & Gen. Genet.* **233**: 241–251.
- Meagher, R.B. 1995. The impact of historical contingency on gene phylogeny: Plant actin diversity. *Evolution. Biol.* **28**: 195–215.
- Meneely, P.M. and R.K. Herman. 1979. Lethals, steriles and deficiencies in a region of the X chromosome of *Caenorhabditis elegans*. *Genetics* **92**: 99–115.
- Oro, A.E., M. McKeown, and R.M. Evans. 1990. Relationship between the product of the *Drosophila ultraspiracle* locus and the vertebrate retinoid X receptor. *Nature* **347**: 298–301.
- Petkovich, M., N.J. Brand, A. Krust, and P. Chambon. 1987. A human retinoic acid receptor which belongs to the family of nuclear receptors. *Nature* **330**: 444–450.
- Rastinejad, F. 1997. Structure and function of the steroid and nuclear receptor DNA binding domain. In *Molecular biology of steroid and nuclear hormone receptors* (ed. L.P. Freedman), pp. 105–131. Birkhauser, Boston, MA.
- Riddiford, L.M. 1994. Cellular and molecular actions of juvenile hormone I. General considerations and premetamorphic actions. In *Advances in insect physiology*. (ed. P.D. Vans), Vol. 24, pp. 213–273. Academic Press, San Diego, CA.
- Robertson, H.M. 1998. Two large families of chemoreceptor genes in the nematodes *Caenorhabditis elegans* and *Caenorhabditis briggsae* reveal extensive gene duplication, diversification, movement, and intron loss. *Genome Res.* **8**: 449–463.
- Rogalski, T.M., D.G. Moerman, and D.L. Baillie. 1982. Essential genes and deficiencies in the *unc-22 IV* region of *Caenorhabditis elegans*. *Genetics* **102**: 725–736.
- Rosenbluth, R.E., T.M. Rogalski, R.C. Johnsen, L.M. Addison, and D.L. Baillie. 1988. Genomic organization in *Caenorhabditis elegans*: Deficiency mapping on linkage group V(left). *Genetic Res.* **52**: 105–118.
- Sambrook, J., E.F. Fritsch, and T. Maniatis. 1989. *Molecular cloning: A laboratory manual*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
- Segraves, W. and D.S. Hogness. 1990. The E75 ecdysone-inducible gene responsible for the 75B early puff in *Drosophila* encodes two new members of the steroid receptor superfamily. *Genes & Dev.* **4**: 204–219.
- 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.
- Simons, S.S. Jr. 1997. Structure and function of the steroid and nuclear receptor ligand-binding domains. In *Molecular biology of steroid and nuclear hormone receptors* (ed. L.P. Freedman), pp. 35–104. Birkhauser, Boston, MA.
- Sluder, A.E., T. Lindblom, and G. Ruvkun. 1997. The *Caenorhabditis elegans* orphan nuclear hormone receptor gene *nhr-2* functions in early embryonic development. *Dev. Biol.* **184**: 303–319.
- Stone, B.L. and C.S. Thummel. 1993. The *Drosophila* 78C early late

- puff contains *E78*, an ecdysone-inducible gene that encodes a novel member of the nuclear hormone receptor superfamily. *Cell* **75**: 307–320.
- Talbot, W.S., E.A. Swyryd, and D.S. Hogness. 1993. *Drosophila* tissues with different metamorphic responses to ecdysone express different ecdysone receptor isoforms. *Cell* **73**: 1323–1337.
- Thomas, H.E., H.G. Stunnenberg, and A.F. Stewart. 1993. Heterodimerization of the *Drosophila* ecdysone receptor with retinoid-x receptor and ultraspiracle. *Nature* **362**: 471–475.
- Thummel, C. 1995. From embryogenesis to metamorphosis: The regulation and function of *Drosophila* nuclear receptor superfamily members. *Cell* **83**: 871–877.
- Vanfleteren, J.R., Y. Van de Peer, M.L. Blaxter, S.A.R. Tweedie, C. Trotman, L. Lu, M.-L. Van Hauwaert, and L. Moens. 1994. Molecular genealogy of some nematode taxa as based on cytochrome c and globin amino acid sequences. *Mol. Phylogenet. Evol.* **3**: 92–101.
- Walthall, W.W. and J.A. Plunkett. 1995. Genetic transformation of the synaptic pattern of a motoneuron class in *Caenorhabditis elegans*. *J. Neurosci.* **15**: 1035–1043.
- Waterston, R., C. Martin, M. Craxton, C. Huynh, A. Coulson, L. Hillier, R. Durbin, P. Green, R. Shownkeen, N. Halloran et al. 1992. A survey of expressed genes in *Caenorhabditis elegans*. *Nat. Genet.* **1**: 114–123.
- Waterston, R.H., J.E. Sulston, and A.R. Coulson. 1997. The genome. In *C. elegans II* (ed. D.L. Riddle, T. Blumenthal, B.J. Meyer and J.R. Priess), pp. 23–45. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
- White, K.P., P. Hurban, T. Watanabe, and D.S. Hogness. 1997. Coordination of *Drosophila* metamorphosis by two ecdysone-induced nuclear receptors. *Science* **276**: 114–117.
- Wightman, B., I. Ha, and G. Ruvkun. 1993. Posttranscriptional regulation of the heterochronic gene *lin-14* by *lin-4* mediates temporal pattern formation in *C. elegans*. *Cell* **75**: 855–862.
- Wightman, B., R. Baran, and G. Garriga. 1997. Genes that guide growth cones along the *C.elegans* ventral nerve cord. *Development* **124**: 2571–2580.
- Wilson, R., R. Ainscough, K. Anderson, C. Baynes, M. Berks, J. Bonfield, J. Burton, M. Connell, T. Copsey, J. Cooper et al. 1994. 2.2 Mb of contiguous nucleotide sequence from chromosome III of *C. elegans*. *Nature* **368**: 32–38.
- Wilson, T.E., R.E. Paulsen, K.A. Padgett, and J. Milbrandt. 1992. Participation of non-zinc finger residues in DNA-binding by 2 nuclear orphan receptors. *Science* **256**: 107–110.
- Wong, S. 1997. "Orphan nuclear hormone receptors in *Caenorhabditis elegans*: Initial studies of expression and function." M.S.thesis, University of Georgia, Athens, GA.
- Xie, W., J.G. Chipman, D.L. Robertson, R. L. Erikson, and D. L. Simmons. 1991. Expression of a mitogen-responsive gene encoding prostaglandin synthase is regulated by mRNA splicing. *Proc. Natl. Acad. Sci.* **88**: 2692–2696.
- Yamamoto, K.R. 1997. Intracellular receptors: New instruments for a symphony of signals. In *Molecular biology of steroid and nuclear hormone receptors* (ed. L.P. Freedman), pp. vii–x. Birkhauser, Boston, MA.
- Yao, T.P., W.A. Seagraves, A.E. Oro, M. McKeown, and R.M. Evans. 1992. *Drosophila* ultraspiracle modulates ecdysone receptor function via heterodimer formation. *Cell* **71**: 63–72.
- Yao, T.-P., B.M. Forman, Z. Jlang, L. Cherbas, J.-D. Chen, M. McKeown, P. Cherbas, and R.M. Evans. 1993. Functional ecdysone receptor is the product of *EcR* and *Ultraspiracle* genes. *Nature* **366**: 476–479.
- Ye, X., P. Fong, N. Iizuka, D. Choate, and D.R. Cavener. 1997. *Ultrabithorax* and *Antennapedia* 5' untranslated regions promote developmentally regulated internal translation initiation. *Mol. Cell. Biol.* **17**: 1714–1721.
- Yeh, W.-Y. 1991. "Genes acting late in the signalling pathway for *Caenorhabditis elegans* dauer larval development." Ph.D. thesis. University of Missouri, Columbia, MO.
- Zhou, H.-M. 1997. "Characterization of *unc-55*, a gene involved in synaptic differentiation of the motoneurons in *Caenorhabditis elegans*." Ph.D. thesis. Georgia State University, Atlanta, GA.
- Zimmer, A., A.M. Zimmer, and K. Reynolds. 1994. Tissue specific expression of the retinoic acid receptor- $\beta$ 2: Regulation of short open reading frames in the 5'-noncoding region. *J. Cell Biol.* **127**: 1111–1119.
- Zorio, D.A.R., N.N. Cheng, T. Blumenthal, and J. Spieth. 1994. Operons as a common form of chromosomal organization in *C. elegans*. *Nature* **372**: 270–272.

Received August 17, 1998; accepted in revised form December 22, 1998.