Genome inventory and analysis of nuclear hormone receptors in *Tetraodon nigroviridis*

Raghu Prasad Rao Metpally[1], Ramakrishnan Vigneshwar[1, 2] and Ramanathan Sowdhamini[1,*]

[1]*National Centre for Biological Sciences, Tata Institute of Fundamental Research, UAS-GKVK Campus, Bellary Road, Bangalore 560 065, India*

[2]*Present address: Department of Chemical and Biomolecular Engineering, National University of Singapore, Singapore 117576*

**Corresponding author (Email, mini@ncbs.res.in)*

Nuclear hormone receptors (NRs) form a large superfamily of ligand-activated transcription factors, which regulate genes underlying a wide range of (patho) physiological phenomena. Availability of the full genome sequence of *Tetraodon nigroviridis* facilitated a genome wide analysis of the NRs in fish genome. Seventy one NRs were found in *Tetraodon* and were compared with mammalian and fish NR family members. In general, there is a higher representation of NRs in fish genomes compared to mammalian ones. They showed high diversity across classes as observed by phylogenetic analysis. Nucleotide substitution rates show strong negative selection among fish NRs except for pregnane X receptor (PXR), estrogen receptor (ER) and liver X receptor (LXR). This may be attributed to crucial role played by them in metabolism and detoxification of xenobiotic and endobiotic compounds and might have resulted in slight positive selection. Chromosomal mapping and pairwise comparisons of NR distribution in *Tetraodon* and humans led to the identification of nine syntenic NR regions, of which three are common among fully sequenced vertebrate genomes. Gene structure analysis shows strong conservation of exon structures among orthologoues. Whereas paralogous members show different splicing patterns with intron gain or loss and addition or substitution of exons played a major role in evolution of NR superfamily.

[Metpally R P R, Vigneshwar R and Sowdhamini R 2006 Genome inventory and analysis of nuclear hormone receptors in *Tetraodon nigroviridis*; *J. Biosci*. **32** 43–50]

1. Introduction

Nuclear hormone receptors (NRs) are one of the most abundant superfamily of ligand-responsive transcription factors in metazoans (Laudet and Gronemeyer 2002). Upon direct binding to endogenous or exogenous cognate signalling molecules, NRs regulate gene expression by interacting with specific DNA sequences upstream of their target genes influencing a wide range of (patho) physiological phenomena (Gronemeyer and Laudet 1995; Mangelsdorf *et al* 1995). The genetic programs that these receptors establish or modify affect virtually all aspects of the life of metazoan organisms, including growth, development, reproduction, detoxification, apoptosis and metabolic homeostasis (Laudet and Gronemeyer 2002). Peroxisome proliferator-activated receptors (PPARs), farnesoid X receptors (FXRs), liver X receptors (LXRs) (Steinmetz *et al* 2001) etc. mediate the effect of thyroid hormone, steroid hormone, retinoids, vitamin D, glucocorticoids, androgens, estrogen, and progestins, as well as lipids, cholesterol metabolites, and

Keywords. Evolution; genome analysis; nuclear receptors; phylogeny

Abbreviations used: DBA, DNA binding domain; LBD; ligand binding domain; LXR, liver X receptors; NRs, nuclear hormone receptors; PXR, prephone X receptors.

Supplementary Data pertaining to this article is available on the *Journal of Biosciences* Website at *http://www.ias.ac.in/jbiosci/jan2007/pp43-50-suppl.pdf*

bile acids . The remaining are 'orphan' receptors, with no discovered ligands, although it is likely that modulating ligands will be identified for some of these receptors (Giguere 1999; Kliewer *et al* 1999). NRs have been aggressively pursued academically to study the mechanisms of evolution (Escriva Garcia *et al* 2003; Laudet *et al* 1992; Robinson-Rechavi and Laudet 2003), cell signalling and transcription regulation to uncover the molecular rules that define spatial and temporal control of gene expression (Steinmetz *et al* 2001). These are pharmaceutically pursued as the second largest potential drug targets after G-protein coupled receptors as they are associated with major pathologies (Kliewer *et al* 1999; Metpally and Sowdhamini 2005b). Further, in agriculture, NRs can also be pursued as possible new targets for the control of invertebrate pests (Maglich *et al* 2001). The search of new receptors, their ligands and the identification of novel signalling pathways in which they involve have, hence, become a very active and promising research (Escriva *et al* 2000).

NRs are composed of several modular conserved domains associated with different roles: N-terminal transcriptional activating modulatory domain (activation function 1 or AF1 or A/B region) is a ligand-independent variable region, a well conserved central zinc-finger DNA binding domain (DBD or C region), a variable hinge region (D region) that helps to orient the DBD, a conserved ligand binding domain (LBD or E region) and a sparsely populated variable C-terminal region (F region) (Escriva Garcia *et al* 2003; Steinmetz *et al* 2001) are other modules. This domain architecture remains conserved throughout the metazoan lineage, while the receptors evolved to bind a variety of different ligands and DNA binding site topologies (Escriva *et al* 2004).

DBD is responsible for direct interactions with cofactors and specific *cis*-regulatory DNA sequences called hormone response elements (HRE) to subsequently control gene expression (Schwabe *et al* 1993a). DBD, classified as a type-II zinc finger motif, corresponds to a 75–80 amino acid residue long segment which includes two modules, each containing a zinc ion coordinated by four cysteine residues. DBD recognizes and binds a specific sequence of DNA of target gene (Luisi *et al* 1991; Schwabe *et al* 1993a, b). The LBD is a flexible unit made of α-helices consisting of 170 to 210 amino acid residues and acts in response to ligand binding. The ligands are generally buried in the receptor and are thought to interact with its hydrophobic core. The binding of the ligand induces conformational changes that control and influence multiple functions ranging from nuclear translocation, oligomerization, cofactor binding, transcriptional activation and repression (Bledsoe *et al* 2004; Edwards 2000). Some receptors bind DNA as monomers, some as homodimers and some as heterodimers with a common partner, the retinoid X receptor (RXR). Both DBD and LBD contribute to dimerization. In many NRs, transcription is repressed in the absence of ligand, due to a repression function in the LBD. Interaction of the LBD with ligand abolishes repression and activates transcription via a C-terminal AF2 domain, which in some receptors works in tandem with an N-terminal AF1 domain (Shao and Lazar 1999).

The completion of several other vertebrate and invertebrate genome sequencing projects paves the way for "comparative functional genomics". The quest for assigning function to putative gene products exploits the sequence and structural similarities to known genes and further could be elucidated using molecular biology techniques. Such studies have important implications in biology and in understanding the evolutionary and functional relationships within protein superfamilies across distinct organisms (Rubin *et al* 2000). Sequencing of the model organisms can be an important source of information about the function and provide new comparators for NR studies. Teleost fish, *Tetraodon nigroviridis* is one of the smallest known vertebrate genomes. It has all the specialized functions of higher vertebrates and can be a good vertebrate model system to study (Jaillon *et al* 2004; Metpally and Sowdhamini 2005a). To gain insights into the sequence diversity, gene structure and evolution of this interesting superfamily, in this research article, we describe full repertoire of NR superfamily in *Tetraodon* fish their genomic location and gene structure and their phylogenetic relationships with mammals and other fish model systems.

2. Methods

2.1 *Identification of nuclear receptors in* Tetraodon *genome*

The whole genome sequence of the *T. nigroviridis* currently released at Genosope (release v7.1 of *Tetraodon* genome, released on September 2004) and NCBI were used to carry out the analysis (Jaillon *et al* 2004). Protein sequences encoded by nuclear receptor genes in NuReBase (Ruau *et al* 2004) and NucleaRDB (Horn *et al* 2001) were chosen for sequence similarity searches against the *Tetraodon* genome. Nuclear receptors were identified using a comprehensive approach (figure 1) that includes BLAST (Altschul *et al* 1997) homology comparisons and Hmmsearch representative LBD and DBD domains and full length NR sequences, separately from NR subfamilies 0 to 6 were aligned using ClustalX (Thompson *et al* 1997), and the alignments was used to build a profile Hidden Markov Model (HMM) using the HMMER software package(Eddy 1998) against NuReBase and NucleaRDB with E-values as benchmarked from our previous studies (Metpally and Sowdhamini 2005a). Further examined by Hmmpfam of HMMER (Eddy 1998) and RPS-BLAST (Marchler-Bauer *et al* 2003) analysis. Putative nuclear receptor sequences

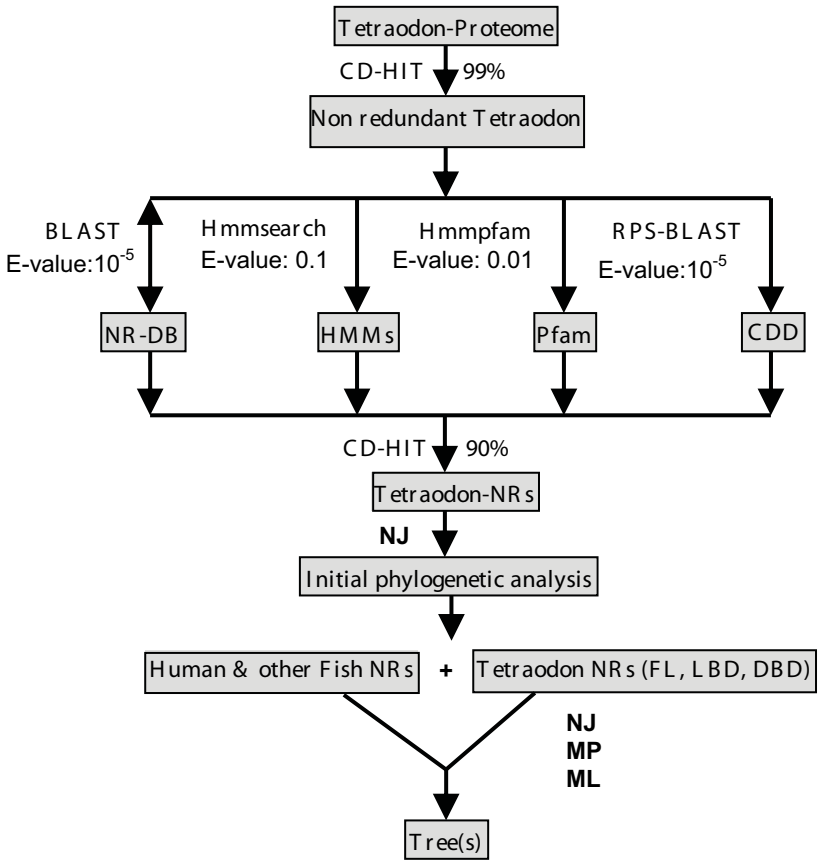


Figure 1. Comprehensive approach for the identification and phylogenetic analysis of *Tetraodon* NRs. All NR sequences from NR Databases (Nurebase and NuclearDB) were compared against *Tetraodon* proteome using BLASTP and hits were searched against NR-DB using reverse BLAST and Hmmsearch using HMMs against *Tetraodon* proteome. As complementary approaches, *Tetraodon* sequences were compared using Hmmpfam against Pfam and RPS-BLAST against CDD respectively. Finally, NR sequences are subjected to phylogenetic analysis using either intact NRs (FL) or LBDs or DBDs along with respective human and fish NRs separately.

were manually checked for specific patterns and presence of DBD and LBD domains. Putative splice variants, polymorphism and duplicates were eliminated by applying 90% sequence identity cut-off using CD-hit (Li *et al* 2001). The corresponding genomic DNA sequences were also searched against the EST database at NCBI using BLASTN with a cutoff E-value of 10e $^{-12}$ (Metpally and Sowdhamini 2005a). Domains of putative NRs of *Tetraodon* were identified by comparing with structural and functional domains specific for NR family which were obtained from Pfam database (Bateman *et al* 2004).

2.2 *Synteny analysis*

2.2a *Ortholog identification*: Two genes, A from genome GA and B from GB, were considered orthologs if B is the best match of gene A in GB and A is the best match of B in GA using BLASTP (Altschul *et al* 1997; Tatusov *et al* 1997) similarity search (Metpally and Sowdhamini 2005a). The genomic locations of *Tetraodon* NRs were identified and mapped onto *Tetraodon* karyogram (supplementary table 1) and were compared with their counterparts in human genome to identify syntenic blocks (table 1).

2.3 *Phylogenetic analysis*

The amino acid sequences of putative NRs of *Tetraodon* were aligned with human nuclear receptor gene family members using CLUSTALX (Thompson *et al* 1997) and edited if required by Jalview (Clamp *et al* 2004). The multiple sequence alignments of full length sequences, LBDs and DBDs separately were used for phylogenetic analysis by implementation of neighbour joining (NJ) and maximum parsimony (MP) in the Phylip 3.6 package (Felsenstein 1989) with a bootstrap of 1000 replicates and of maximum likelihood (ML) method was implemented with 10,000 quartet-puzzling steps in the TREE-PUZZLE (Schmidt *et al* 2002) software with all other parameters as

Table 1. Syntenic blocks containing NR genes in *Tetraodon* and human genomes

Human Synteny block	*Tetraodon* Synteny block	*Tetraodon* chromosome	Size (Mb)	Human chromosome	Size (Mb)
NR1B2: NR1I2	CAG04399: CAG05861	2	5.3	3	95
NR4A3: NR2B1	CAG11936: CAG12025	8	1	9	34
NR1D2: NR1A2: NR1B2	CAG07394: CAG07392	9	0.04	3	1.4
NR1A1: NR1D1: NR1B1	CAG02080: CAG02086	SCAF14676	0.053	17	0.3
NR3A2: NR3B2	CAG03763: CAG10628	10	7.9	14	12
NR1F1:NR2F2	CAF98309: CAG13090	13	4.3	15	153
NR1H4: NR2C1	CAG03422: CAG08700	13	5.2	12	5.4
NR3A1:NR2E1	CAG03596: CAG03617	14	0.25	6	43.5
NR1C3: NR1D2	CAG06739: CAG07050	11	2.8	3	11

published earlier for tree construction of GPCRs (Metpally and Sowdhamini 2005a, b).

2.4 *dN/dS analysis*

The dN/dS ratios for multi-codon regions (i.e. full-length, LBD, DBD) of the nuclear receptor coding sequence were determined with SNAP (Ota and Nei 1994) using Nei and Gojobori method (Nei and Gojobori 1986). Nucleotide sequences of each full length NR and individual domains (DBD and LBD) were codon-aligned in accordance with their corresponding amino acid sequence alignment. For every pair of sequences, we calculated following values: the observed number of synonymous (Sd) and nonsynonymous (Sn) substitutions, the number of potential synonymous (S) and nonsynonymous (N) substitutions, the proportions of Sd/S (ps) and Sn/N (pn), and the corresponding Jukes-Cantor corrected proportions dN and dS (Jukes 1969). In many pairwise comparisons, mutational saturation had been reached (ps or pn > 0.75) and therefore these comparisons were subsequently ignored. To make inferences about selective pressure (positive and negative selection) on individual codons (sites) within the coding sequence of the *Tetraodon* NR genes, the Single Likelihood Ancestor Counting (SLAC) package (*http://www.datamonkey.org*), which implements the Suzuki-Gojobori method (Suzuki and Gojobori 1999), was used.

3. Results and discussion

3.1 *Repertoire of putative* Tetraodon *nuclear receptors*

We identified and analysed the complete *Tetraodon* NR gene repertoire (71 members) by multiple sequence comparison methods (figure 1) and compared it with repertoire of NR genes in human and other fully sequenced fish genomes (fugu and zebra fish). The number of candidate genes identified in *Tetraodon* are about 45% higher than the number of human NR genes, but comparably similar to that of other fish genomes (Maglich *et al* 2003). Moreover, we could only find clear human orthologs for 28 (~39%) of the *Tetraodon* NR genes. About 21 NR genes are domain singletons (8 with only LBD domain and 13 with only DBD domain). Except two of these domain singletons, they did not share sequence similarity with known single domain DAX and SHP (two NRs known to lack DBD). Our results place these with domain singleton NRs close to other respective NRs (with both LBD and DBD) with strong bootstrap support (figure 2). Domain singleton NRs show either partially or completely missing other associated domains. This could be either due to several gaps in the assemblies of the *Tetraodon* genome or they might have arose, by the loss of either DBD or LBD respectively. Either they may have modulating function on other closest associated full NRs as expected with the NR0 family members or absence of any one of these two domains may make these proteins non functional, but exact function is still remain puzzling. Some of these evolved much more rapidly than other family members, as indicated by long branch lengths for these NRs (figure 2).

3.2 *Genomic distribution of* Tetraodon *NRs*

The genomic locations of nuclear receptors were mapped on the *Tetraodon* karyogram (supplementary table 1). We found that NR genes were widely distributed throughout the *Tetraodon* genome except for chromosomes 12, 17, 19, 20 and 21. About 10 NRs are present in scaffolds with undetermined chromosomal assignments (marked by UD in supplementary table 1). We identified nine syntenic regions between *Tetraodon* and human (table 1). Of these, three are observed in other mammals like rat and mouse. These syntenic regions vary from 0.053 Mb to 8 Mb in size. The syntenic blocks in *Tetraodon* are very short

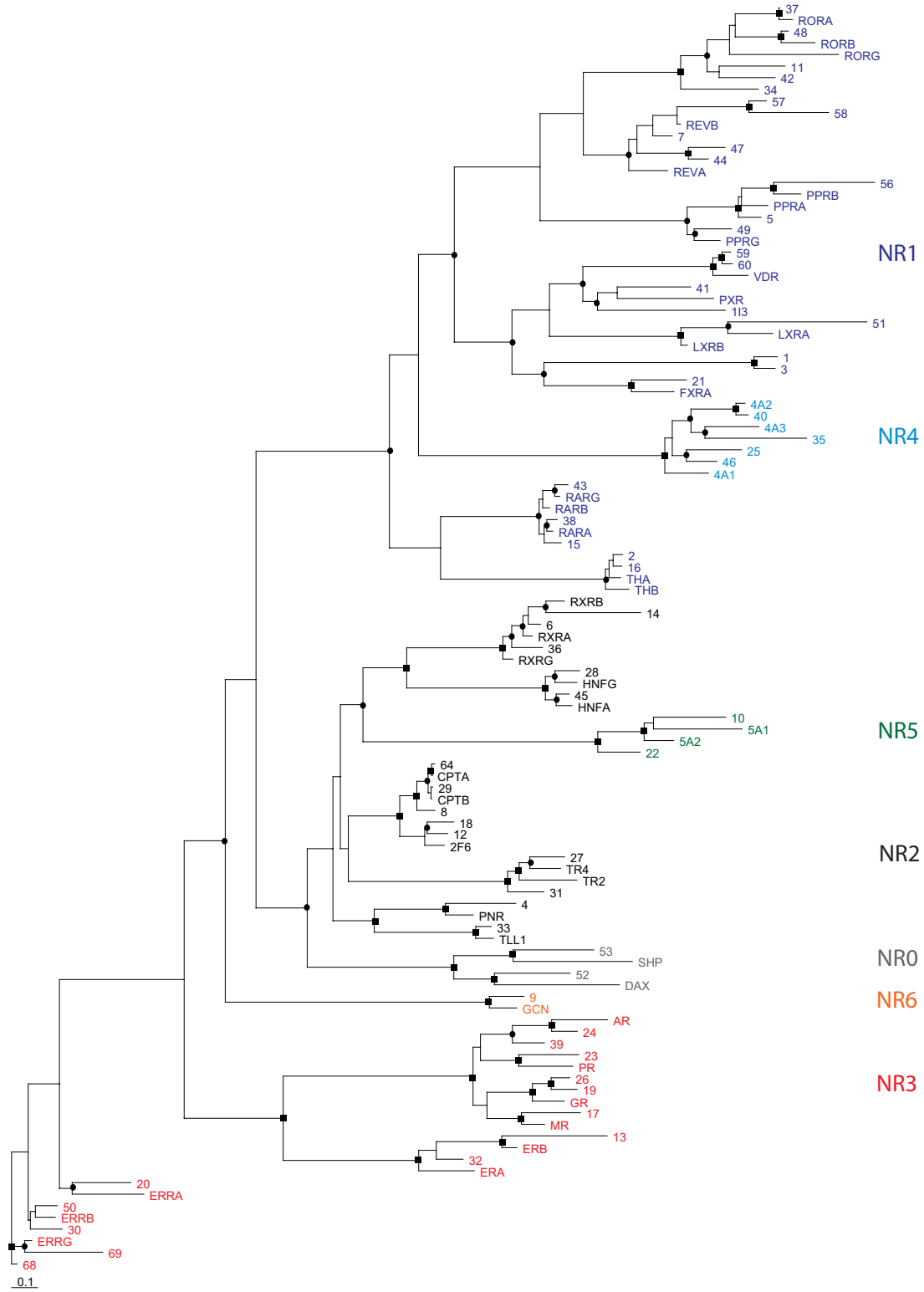


Figure 2. Phylogenetic relationship of the NRs in *Tetraodon* and human genome. The tree was calculated using multiple sequence alignment of LBDs of *Tetraodon* and human NRs by the neighbour joining method on 1000 replicas of the data set. Bootstrap support is indicated with circle on the relevant branch point for 60 to 80% and a square for 80 to 100%.

compared to that of the mammalian counterparts (0.2 Mb to 55 Mb), supporting the reported compactness of the *Tetraodon* genome (Jaillon *et al* 2004). Two syntenic blocks observed between teleosts (*Tetraodon*, fugu, and zebrafish), mammalian (human, rat and mouse) and chicken genomes are highly conserved across vertebrates (one with NR1A1 and NR1B1) and another with NR1D2 and NR1B2) (Koh and Moore 1999); interestingly, only mammals show insertion of one NR receptor (NR1D1 or NR1A2) in each of these blocks with tail to tail orientations of NR1D1 or NR1A2 (table 1). This may be attributed to retrotransposition activity after their divergence from aquatic vertebrates and is supported by the presence of SINE and LINE sequences upstream and/or downstream of the inserted nuclear receptor genes. The insertion and juxtaposition of NR1A2 and NR1D1 genes in mammalian genomes and their absence in avian and fish genomes supports that such an invention might be specific to mammals (Zhang *et al* 2004).

3.3 *Comparative phylogeny of* Tetraodon *NRs*

We constructed separate phylogenetic trees of the NR-DBDs, NR-LBDs and full-length NR genes using multiple phylogenetic methods [NJ, maximum parsimony (MP) and ML) as described in §2 (figure 1) to reconciliate the possibility of disproportionate evolutionary pressure at the individual domains of nuclear receptors (Laudet 1997; Owen and Zelent 2000), and to analyse the topology and relationships of NR subfamilies in the tree(s). *Tetraodon* nuclear receptors could be classified into six distinct families (subfamily one with 28 members; subfamily two: 18; subfamily three: 16; subfamily four: 4; subfamily five: 3; subfamily six :1; and subfamily zero with 2 members), agreeing well with the earlier subfamily classification of NRs (Nuclear Receptors Nomenclature Committee 1999). Interestingly, we found the structures of phylogenetic tree topologies differed in respect to DBDs and LBDs suggesting different rates of evolutionary pressures on DBD and LBD domains. DBDs show high conservation of their sequences with atleast 10 to 25 % higher sequence identity corresponding to LBDs within subfamily (supplementary table 2) and also when compared to LBDs across the genomes (Escriva *et al* 2004) supported by the relatively smaller terminal branch lengths of the DBD members (supplementary figure 1). In general, NRs are highly conserved among fishes and subsequently diverged when compared with their mammalian counterparts. This is apparent in the trees as short terminal branch lengths among fish members and long branch lengths with respect to human members (supplementary figure 2).

3.4 *Negative selection on NR genes*

An important component of any molecular evolutionary analyses is the estimation of synonymous (silent) and nonsynonymous (replacement) nucleotide substitution rates. For our analysis, we used relative frequency of non-synonymous versus synonymous codon substitutions to estimate the selection process acting on these nuclear receptors. dN/Ds analysis of NR receptors across genomes have shown low values (0.01 to 0.2) showing strong negative selection pressure in vertebrate genomes. This is expected since most of them recognize endogenous ligands which have comparatively less chance of variation within organisms. Some of *Tetraodon* NR receptors like pregnane X receptor (PXR), estrogen receptors (ER) and LXR show higher dN/dS ratios (0.3 to 0.45) compared to other NRs and are among the most divergent in the *Tetraodon* NR superfamily. This may be correlated to the crucial role played by these receptors in the metabolism and detoxification of endobiotic and xenobiotic (greater tendency to vary) compounds and may have resulted from slight positive selection (Kliewer *et al* 1998).

3.5 *Gene structure of DBD and LBD domains of* Tetraodon *NR genes*

Protein sequences were compared with genomic DNA sequence, allowing for the identification of exons and introns using Wise2 software. *Tetraodon* DBDs have shown seven out of eight different, conserved patterns of splice junctions observed in mammalian NRs (figure 3) (Zhang *et al* 2004). Specific conserved amino acids (1: glycine; 2: glycine; 3:

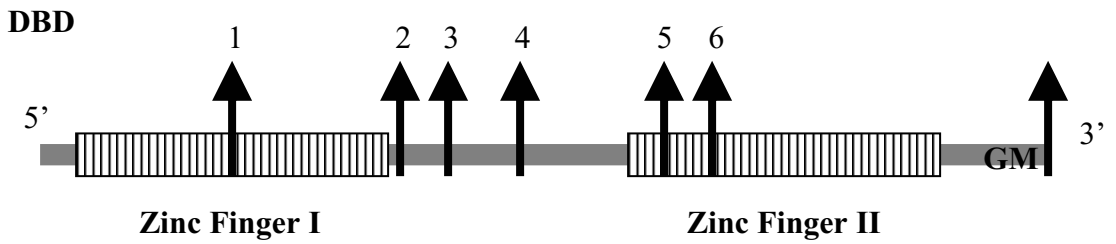


Figure 3. Gene structure of DBD domains of *Tetraodon* NR Genes. Up arrow with numbers one to six represent specific position of splice junctions.

arginine; 4: glycine; 5: arginine) (figure 3) are observed for each pattern of splice junctions of *Tetraodon* DBD. Splice junction is located within the first zinc finger motif in NR2B (1) and 2C (1) subfamily members, it is located between two zinc finger motifs in NR1 (A, I: 3; B, C, F, H: 2), NR3 (4), NR4 (3) and NR5 (2) members and it is located within second zinc finger motif in NR2A (5), 2E (6) and 2F (5) members (figure 3). Where as, splice junctions are lost in DBD domain of NR1H1, NR1D3, NR2F2 and 5, NR5A5 and NR6A1. LBDs show a variety of different exon structures among subfamilies of NRs within *Tetraodon* genome which are differentially conserved among subfamilies. This displays changes including intron gain or loss and exon addition or substitution. This adds up significant diversity to the nuclear receptor superfamily. The exon structures of the DBDs and LBDs are conserved within orthologs across species supported by strong purifying (negative) selection observed by analysing synonymous and nonsynonymous nucleotide substitution rates (dN/dS analysis).

4. Conclusions

We have identified and analysed repertoire of *Tetraodon* nuclear receptors and found high level of orthology with human counterparts. The human and *Tetraodon* NR sequences are analogous in terms of NR subfamilies, but *Tetraodon* display slightly higher number of receptors at the subfamily level. They showed high diversity across subfamilies as observed by phylogenetic analysis. Nucleotide substitution rates show strong negative selection among fish NRs except for PXR, ER and LXR. This may be attributed to crucial role played by them in metabolism and detoxification of xenobiotic and endobiotic compounds and may have resulted in slight positive selection. Chromosomal mapping and pair wise comparisons of NR distribution in *Tetraodon* and humans led to the identification of nine syntenic NR regions, of which three are highly conserved among vertebrate genomes. Gene structure analysis shows strong conservation of exon structures among orthologous members. Whereas paralogous members show different splicing patterns with intron gain or loss and addition or substitution of exons played a major role in evolution of NR superfamily. The *Tetraodon* genome, with its larger set of nuclear receptors, provides an additional and interesting model to study both evolution and function of these receptors.

Acknowledgments

MRPR is a Senior Research Fellow of the Council of Scientific and Industrial Research New Delhi. RS is a Senior Research Fellow of the Wellcome Trust, UK. We also thank NCBS-TIFR for infrastructural support.

References

Altschul S F, Madden T L, Schaffer A A, Zhang J, Zhang Z, Miller W and Lipman D J 1997 Gapped BLAST and PSI-BLAST: a new generation of protein database search programs; *Nucleic Acids Res.* **25** 3389–3402

Bateman A, Coin L, Durbin R, Finn R D, Hollich V, Griffiths-Jones S, Khanna A, Marshall M, Moxon S, Sonnhammer E L, Studholme D J, Yeats C and Eddy S R 2004 The Pfam protein families database; *Nucleic Acids Res.* **32 (**Database issue) D138–D141

Bledsoe R K, Stewart E L and Pearce K H 2004 Structure and function of the glucocorticoid receptor ligand binding domain; *Vitam. Horm.* **68** 49–91

Clamp M, Cuff J, Searle S M and Barton G J 2004 The Jalview Java alignment editor; *Bioinformatics* **20** 426–427

Eddy S R 1998 Profile hidden Markov models; *Bioinformatics* **14** 755–763

Edwards D P 2000 The role of coactivators and corepressors in the biology and mechanism of action of steroid hormone receptors; *J. Mammary Gland Biol. Neoplasia* **5** 307–324

Escriva Garcia H, Laudet V and Robinson-Rechavi M 2003 Nuclear receptors are markers of animal genome evolution; *J. Struct. Funct. Genom.* **3** 177–184

Escriva H, Delaunay F and Laudet V 2000 Ligand binding and nuclear receptor evolution; *Bioessays* **22** 717–727

Escriva H, Bertrand S and Laudet V 2004 The evolution of the nuclear receptor superfamily; *Essays Biochem.* **40** 11–26

Felsenstein J 1989 PHYLIP – Phylogeny Inference Package (Version 3.2); *Cladistics* **5** 164–166

Giguere V 1999 Orphan nuclear receptors: from gene to function; *Endocrinol. Rev.* **20** 689–725

Gronemeyer H and Laudet V 1995 Transcription factors 3: nuclear receptors; *Protein Profile* **2** 1173–1308

Horn F, Vriend G and Cohen F E 2001 Collecting and harvesting biological data: the GPCRDB and NucleaRDB information systems; *Nucleic Acids Res.* **29** 346–349

Jaillon O, Aury J M, Brunet F, Petit J L, Stange-Thomann N, Mauceli E, Bouneau L, Fischer C, *et al* 2004 Genome duplication in the teleost fish *Tetraodon nigroviridis* reveals the early vertebrate proto-karyotype; *Nature* (*London*) **431** 946–957

Jukes T H 1969 Recent advances in studies of evolutionary relationships between proteins and nucleic acids; *Space Life Sci.* **1** 469–490

Kliewer S A, Lehmann J M and Willson T M 1999 Orphan nuclear receptors: shifting endocrinology into reverse; *Science* **284** 757–760

Kliewer S A, Moore J T, Wade L, Staudinger J L, Watson M A, Jones S A, McKee D D, Oliver B B, Willson T M, Zetterstrom R H, Perlmann T and Lehmann J M 1998 An orphan nuclear receptor activated by pregnanes defines a novel steroid signaling pathway; *Cell* **92** 73–82

Koh Y S and Moore D D 1999 Linkage of the nuclear hormone receptor genes NR1D2, THRB, and RARB: evidence for an ancient, large-scale duplication; *Genomics* **57** 289–292

Laudet V 1997 Evolution of the nuclear receptor superfamily: early diversification from an ancestral orphan receptor; *J. Mol. Endocrinol.* **19** 207–226

Laudet V and Gronemeyer H 2002 *The nuclear receptors factsbook* (London: Academic Press)

Laudet V, Hanni C, Coll J, Catzeflis F and Stehelin D 1992 Evolution of the nuclear receptor gene superfamily; *EMBO J.* **11** 1003–1013

Li W, Jaroszewski L and Godzik A 2001 Clustering of highly homologous sequences to reduce the size of large protein databases; *Bioinformatics* **17** 282–283

Luisi B F, Xu W X, Otwinowski Z, Freedman L P, Yamamoto K R and Sigler P B 1991 Crystallographic analysis of the interaction of the glucocorticoid receptor with DNA; *Nature* (*London*) **352** 497–505

Maglich J M, Caravella J A, Lambert M H, Willson T M, Moore J T and Ramamurthy L 2003 The first completed genome sequence from a telost fish (*Fugu rubripes*) adds significant diversity to the nuclear receptor superfamily; *Nucleic Acids Res.* **31** 4051–4058

Maglich J M, Sluder A, Guan X, Shi Y, McKee D D, Carrick K, Kamdar K, Willson T M and Moore J T 2001 Comparison of complete nuclear receptor sets from the human, *Caenorhabditis elegans* and *Drosophila* genomes; *Genome Biol.* **2** RESEARCH-0029

Mangelsdorf D J, Thummel C, Beato M, Herrlich P, Schutz G, Umesano K, Blumberg B, Kastner P, Mark M, Chambon P and Evans R 1995 The nuclear receptor superfamily: the second decade; *Cell* **83** 835–839

Marchler-Bauer A, Anderson J B, DeWeese-Scott C, Fedorova N D, Geer L Y, He S, Hurwitz D I, Jackson J D *et al* 2003 CDD: a curated Entrez database of conserved domain alignments; *Nucleic Acids Res.* **31** 383–387

Metpally R P R and Sowdhamini R 2005a Genome wide survey of G protein-coupled receptors in *Tetraodon nigroviridis*; *BMC Evol. Biol.* **5** 41

Metpally R P R and Sowdhamini R 2005b Cross genome phylogenetic analysis of human and Drosophila G protein-coupled receptors: application to functional annotation of orphan receptors; *BMC Genomics* **6** 106

Nei M and Gojobori T 1986 Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions; *Mol. Biol. Evol.* **3** 418–426

Nuclear Receptors Nomenclature Comittee 1999 A unified nomenclature system for the nuclear receptor superfamily; *Cell* **97** 161–163

Ota T and Nei M 1994 Variance and covariances of the numbers of synonymous and nonsynonymous substitutions per site; *Mol. Biol. Evol.* **11** 613–619

Owen G I and Zelent A 2000 Origins and evolutionary diversification of the nuclear receptor superfamily; *Cell. Mol. Life. Sci.* **57** 809–827

Robinson-Rechavi M and Laudet V 2003 Bioinformatics of nuclear receptors; *Methods Enzymol.* **364** 95–118

Ruau D, Duarte J, Ourjdal T, Perriere G, Laudet V and Robinson-Rechavi M 2004 Update of NUREBASE: nuclear hormone receptor functional genomics; *Nucleic Acids Res.* **32** D165–D167

Rubin G M, Yandell M D, Wortman J R, Gabor Miklos G L, Nelson C R, Hariharan I K, Fortini M E, Li P W, *et al* 2000 Comparative genomics of the eukaryotes; *Science* **287** 2204–2215

Schmidt H A, Strimmer K, Vingron M and von Haeseler A 2002 TREE-PUZZLE: maximum likelihood phylogenetic analysis using quartets and parallel computing; *Bioinformatics* **18** 502–504

Schwabe J W, Chapman L, Finch J T and Rhodes D 1993a The crystal structure of the estrogen receptor DNA-binding domain bound to DNA: how receptors discriminate between their response elements; *Cell* **75** 567–578

Schwabe J W, Chapman L, Finch J T, Rhodes D and Neuhaus D 1993b DNA recognition by the oestrogen receptor: from solution to the crystal; *Structure* **1** 187-204

Shao D and Lazar M A 1999 Modulating nuclear receptor function: may the phos be with you; *J. Clin. Invest.* **103** 1617–1618

Steinmetz A C U, Renaud J P and Moras D 2001 Binding of ligands and activation of transcription by nuclear receptors; *Annu. Rev. Biophys. Biomol. Struct.* **30** 329–359

Suzuki Y and Gojobori T 1999 A method for detecting positive selection at single amino acid sites; *Mol. Biol. Evol.* **16** 1315–1328

Tatusov R L, Koonin E V and Lipman D J 1997 A genomic perspective on protein families; *Science* **278** 631–637

Thompson J D, Gibson T J, Plewniak F, Jeanmougin F and Higgins D G 1997 The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools; *Nucleic Acids Res.* **25** 4876–4882

Zhang Z, Burch P E, Cooney A J, Lanz R B, Pereira F A, Wu J, Gibbs R A, Weinstock G and Wheeler D A 2004 Genomic analysis of the nuclear receptor family: new insights into structure, regulation, and evolution from the rat genome; *Genome Res.* **14** 580–590

ePublication: 20 December 2006

Genome inventory and analysis of nuclear hormone receptors in *Tetraodon nigroviridis*

Raghu Prasad Rao Metpally, Ramakrishnan Vigneshwar and Ramanathan Sowdhamini

J. Biosci. **32(1), January 2007, 43–50**, **© Indian Academy of Sciences**

Supplementary Table 1

Tetraodon nigroviridis nuclear receptors.

NR	Accession No.	No.	Gene	DBD	LBD	Invariable Splice junction (D)	Chromosome
Subfamily 1:							
NR1A1	CAF90676.1	2	TRA	f	f	no	2
NR1A1	CAG02086.1*	16	TRA	f	f	yes	UD
NR1A2	CAG00249.1	54	THB	f	a	no	UD
NR1B1	CAG02080.1	15	RARA	f	f	yes	UD
NR1B1	CAG04399.1*	38	RARA	f	f	yes	2
NR1B3	CAG07392.1	43	RARG	f	f	yes	9
NR1C1	CAF95270.1*	5	PPARA	f	f	yes	13
NR1C1	CAF99979.1	55	PPARA	f	a	no	19
NR1C2	CAG07471.1	56	PPARD	f	p	no	9
NR1C3	CAG07050.1*	49	PPRG	f	f	yes	11
NR1D1	CAG07394.1*	44	REVA	f	f	no	9
NR1D1	CAG06739.1	47	REVA	p	f	yes	11
NR1D2	CAG00250.1	7	REVB	f	f	no	UD
NR1D2	CAG02755.1	57	REVA	a	f	yes	UD
NR1D2	CAG02756.1	58	REVA	f	p	no	UD
NR1F1	CAG11892.1	34	RORA	f	f	yes	8
NR1F1	CAF98309.1*	37	RORA	f	f	yes	13
NR1F2	CAG01439.1	11	RORB	f	f	yes	1
NR1F2	CAG07758.1	42	RORB	f	f	yes	15
NR1F2	CAG06880.1	48	RORB	f	f	yes	11
NR1H3	CAG03422.1*	21	FXRA	f	f	yes	13
NR1H3	CAF99925.1*	51	LXRA	f	f	yes	5
NR1H4	CAF90864.1	1	FXRA	f	f	yes	UD
NR1H4	CAF91991.1	3	FXRB	f	f	yes	UD
NR1I1	CAF94134.1*	59	VDR	a	f	yes	UD
NR1I1	CAF96472.1	60	VDR	a	f	yes	UD
NR1I1	CAF96473.1	61	VDR-beta	f	a	no	UD
NR1I2	CAG05861.1*	41	PXR	f	f	yes	2
Subfamily 2:							
NR2A1	CAG03838.1*	45	HNFA	f	f	yes	9
NR2A2	CAF97945.1	28	HNFG	f	f	yes	6
NR2B1	CAF95413.1	6	RXRA	f	p	yes	4
NR2B1	CAF91378.1	62	RXRA	f	a	no	1
NR2B2	CAG11675.1	14	RXRB	f	f	yes	UD
NR2B2	CAG12025.1*	36	RXRB	f	f	yes	8
NR2B2	CAF88861.1	63	RXRB	f	a	no	UD
NR2C1	CAG08700.1*	31	TR2	f	f	yes	13

Supplementary table 1. (*Continued*)

NR2C2	CAG11327.1*	27	TR4	f	f	yes	9
NR2E1	CAG03617.1*	33	TLL1	f	f	no	14
NR2E2	CAF93476.1	4	PNR	f	f	yes	UD
NR2F1	CAF91926.1	64	CPTA	a	f	yes	12
NR2F1	CAG13569.1	65	CPTA	f	a	no	UD
NR2F2	CAG13090.1*	29	CPTB	f	f	yes	13
NR2F2	CAF94543.1	66	CPTB	f	a	no	UD
NR2F5	CAG00925.1	8	COUPG	f	f	yes	UD
NR2F6	CAG01948.1	12	EAR2	f	f	yes	15
NR2F6	CAG00763.1*	18	EAR2	f	f	yes	1
Subfamily 3:							
NR3A1	CAG03596.1*	32	ERA	f	f	no	14
NR3A2	CAG03763.1	13	ERB	f	f	no	10
NR3A3	CAF90265.1	67	ERG	f	a	no	14
NR3B1	CAG01578.1	20	ERRA	f	f	yes	7
NR3B2	CAG09135.1	30	ERRB	f	f	yes	14
NR3B2	CAG10628.1*	50	ERRB	f	f	yes	10
NR3B3	CAG11044.1	68	ERRG	a	f	yes	5
NR3B3	CAG12248.1	69	ERRG	a	f	no	UD
NR3B3	CAG09068.1	70	ERRG	f	a	no	14
NR3B3	CAG11045.1	71	ERRG	f	a	no	5
NR3C1	CAG11713.1*	19	GR	f	f	no	7
NR3C1	CAF99074.1	26	GR	f	f	no	1
NR3C2	CAG11072.1	17	MR	f	f	no	18
NR3C3	CAG12799.1*	23	PR	f	f	no	16
NR3C4	CAG02975.1*	24	AR	f	f	no	7
NR3C4	CAG08385.1	39	AR	f	f	no	1
Subfamily 4:							
NR4A1	CAF96539.1*	25	NGF1B	f	f	yes	11
NR4A1	CAG03953.1	46	NGFIB	f	f	yes	9
NR4A2	CAG09317.1*	40	NOR1	f	f	yes	2
NR4A3	CAG11936.1	35	NURRI	f	f	yes	8
Subfamily 5:							
NR5A1	CAG01304.1*	10	SF1	f	f	no	UD
NR5A2	CAF92683.1	72	LRH1	f	a	no	1
NR5A5	CAG12178.1	22	FF1C	f	f	yes	3
Subfamily 6:							
NR6A1	CAG01303.1*	9	GCN	f	f	yes	UD
Subfamily 0:							
NR0B1	CAG05777.1*	52	DAX	a	f	yes	2
NR0B2	CAG00032.1*	53	SHP	a	f	yes	8

'f' represent full domain: 'p' represents partial domain; 'a' absence of domain.

* means, it has putative human NR ortholog.

Supplementary Table 2

Sequence conservation of *Tetraodon* NRs between DBDs and LBDs.

Subfamily	Average identity	
	DBD	LBD
Subfamily 1	53%	34%
Subfamily 2	59%	43%
Subfamily 3	65%	41%
Subfamily 4	86%	67%
Subfamily 5	56%	46%

Subfamily 6 has a single member that is why not shown in the table.

Supplementary Table 3

List of nuclear receptors from Human, *Tetraodon*, fugu fish and zebra fish genomes.

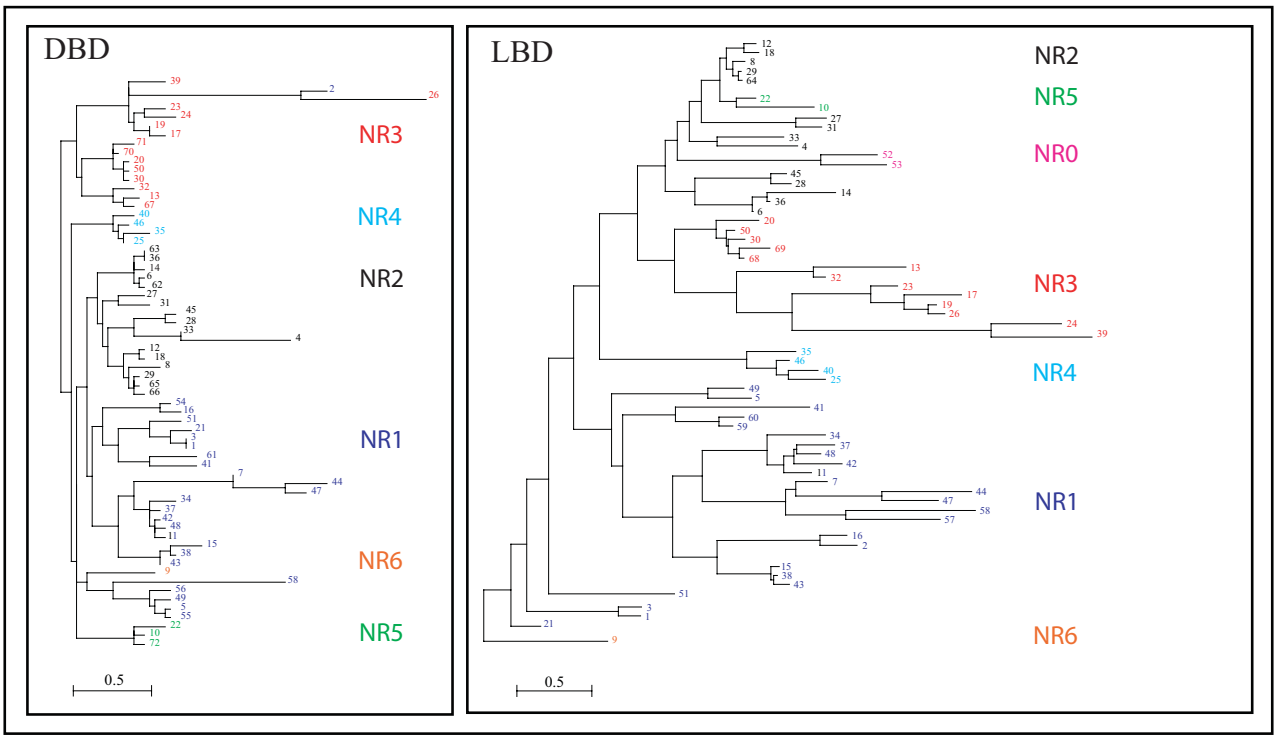
Human Nrs		*Tetraodon* NRs		Fugu fish NRs		Zebra fish NRs	
Gene	Code	Accession No.	No.	Accession No.	No.	Accession No.	No.
1I3_HOMS1	1I3	CAF90864.1	1	SINFRUP00000127989	100	ENSDARP00000000160	200
2F6_HOMS1	2F6	CAF90676.1	2	SINFRUP00000128122	101	ENSDARP00000001911	201
4A1_HOMS1	4A1	CAF91991.1	3	SINFRUP00000130029	102	ENSDARP00000002435	202
4A2_HOMS1	4A2	CAF93476.1	4	SINFRUP00000130211	103	ENSDARP00000002838	203
4A3_HOMS1	4A3	CAF95270.1	5	SINFRUP00000130483	104	ENSDARP00000003080	204
5A1_HOMS1	5A1	CAF95413.1	6	SINFRUP00000130524	105	ENSDARP00000004918	205
5A2_HOMS1	5A2	CAG00250.1	7	SINFRUP00000132228	106	ENSDARP00000005364	206
AR_HOMS1	AR	CAG00925.1	8	SINFRUP00000132438	107	ENSDARP00000007721	207
CPTA_HOMS1	CPTA	CAG01303.1	9	SINFRUP00000132486	108	ENSDARP00000008527	208
CPTB_HOMS1	CPTB	CAG01304.1	10	SINFRUP00000134462	109	ENSDARP00000009236	209
DAX_HOMS1	DAX	CAG01439.1	11	SINFRUP00000134772	110	ENSDARP00000010118	210
ERA_HOMS1	ERA	CAG01948.1	12	SINFRUP00000135072	111	ENSDARP00000010239	211
ERB_HOMS1	ERB	CAG03763.1	13	SINFRUP00000136830	112	ENSDARP00000010876	212
ERRA_HOMS1	ERRA	CAG11675.1	14	SINFRUP00000137435	113	ENSDARP00000011840	213
ERRB_HOMS1	ERRB	CAG02080.1	15	SINFRUP00000138195	114	ENSDARP00000013317	214
ERRG_HOMS1	ERRG	CAG02086.1	16	SINFRUP00000138236	115	ENSDARP00000014027	215
FXRA_HOMS1	FXRA	CAG11072.1	17	SINFRUP00000138841	116	ENSDARP00000015111	216
GCN_HOMS1	GCN	CAG00763.1	18	SINFRUP00000138848	117	ENSDARP00000015784	217
GR_HOMS1	GR	CAG11713.1	19	SINFRUP00000139076	118	ENSDARP00000016299	218
HNFA_HOMS1	HNFA	CAG01578.1	20	SINFRUP00000139078	119	ENSDARP00000017514	219
HNFG_HOMS1	HNFG	CAG03422.1	21	SINFRUP00000139079	120	ENSDARP00000017728	220
LXRA_HOMS1	LXRA	CAG12178.1	22	SINFRUP00000139498	121	ENSDARP00000019436	221
LXRB_HOMS1	LXRB	CAG12799.1	23	SINFRUP00000140232	122	ENSDARP00000019758	222
MR_HOMS1	MR	CAG02975.1	24	SINFRUP00000140234	123	ENSDARP00000021935	223
PNR_HOMS1	PNR	CAF96539.1	25	SINFRUP00000140681	124	ENSDARP00000022973	224
PPRA_HOMS1	PPRA	CAF99074.1	26	SINFRUP00000141263	125	ENSDARP00000023084	225
PPRB_HOMS1	PPRB	CAG11327.1	27	SINFRUP00000143035	126	ENSDARP00000023136	226
PPRG_HOMS1	PPRG	CAF97945.1	28	SINFRUP00000143415	127	ENSDARP00000024987	227
PR_HOMS1	PR	CAG13090.1	29	SINFRUP00000143426	128	ENSDARP00000026313	228

Supplementary table 3 (*Continued*)

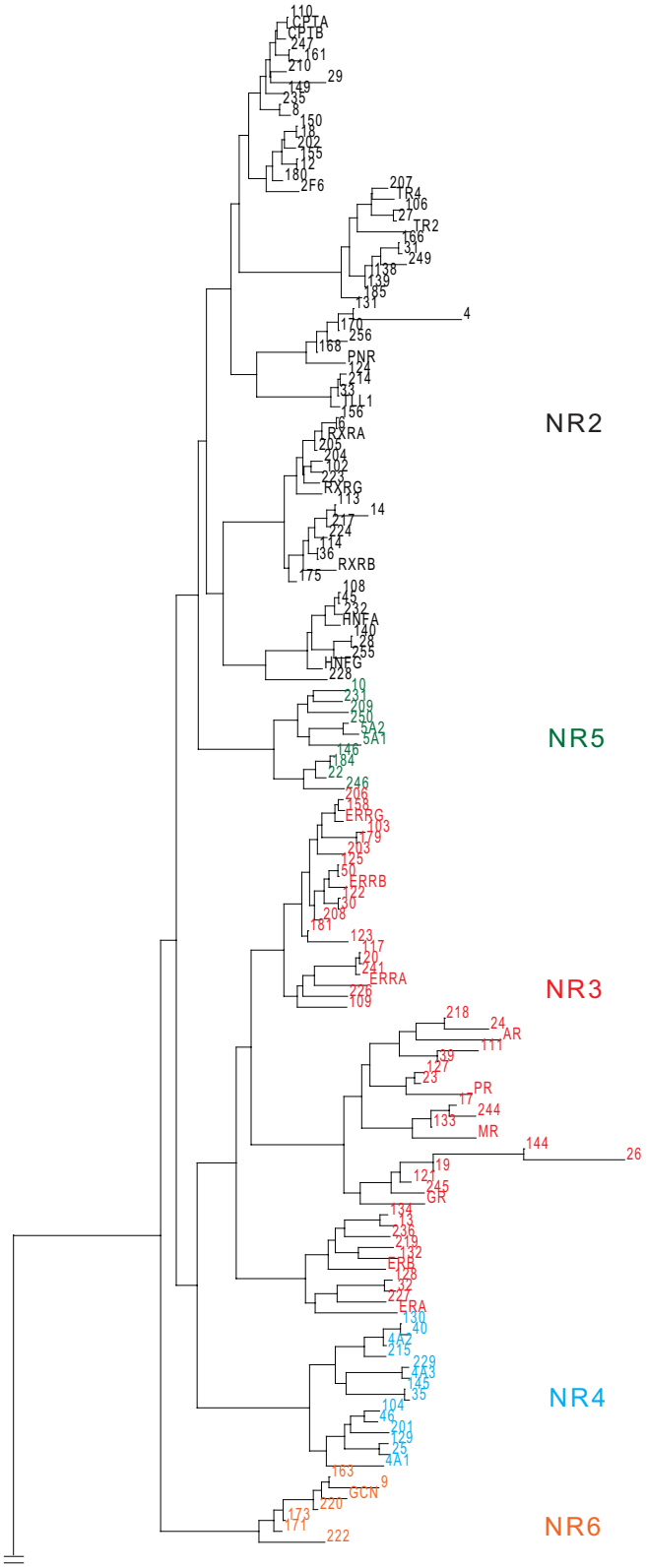
PXR_HOMS1	PXR	CAG09135.1	30	SINFRUP00000144235	129	ENSDARP00000028147	229
RARA_HOMS1	RARA	CAG08700.1	31	SINFRUP00000144980	130	ENSDARP00000028873	230
RARB_HOMS1	RARB	CAG03596.1	32	SINFRUP00000146354	131	ENSDARP00000029204	231
RARG_HOMS1	RARG	CAG03617.1	33	SINFRUP00000146840	132	ENSDARP00000029754	232
REVA_HOMS1	REVA	CAG11892.1	34	SINFRUP00000147278	133	ENSDARP00000036285	233
REVB_HOMS1	REVB	CAG11936.1	35	SINFRUP00000147711	134	ENSDARP00000040467	234
RORA_HOMS1	RORA	CAG12025.1	36	SINFRUP00000148149	135	ENSDARP00000040768	235
RORB_HOMS1	RORB	CAF98309.1	37	SINFRUP00000148150	136	ENSDARP00000041299	236
RORG_HOMS1	RORG	CAG04399.1	38	SINFRUP00000148390	137	ENSDARP00000042466	237
RXRA_HOMS1	RXRA	CAG08385.1	39	SINFRUP00000148638	138	ENSDARP00000042824	238
RXRB_HOMS1	RXRB	CAG09317.1	40	SINFRUP00000148642	139	ENSDARP00000044676	239
RXRG_HOMS1	RXRG	CAG05861.1	41	SINFRUP00000149882	140	ENSDARP00000049550	240
SHP_HOMS1	SHP	CAG07758.1	42	SINFRUP00000150826	141	ENSDARP00000050433	241
THA_HOMS1	THA	CAG07392.1	43	SINFRUP00000151563	142	ENSDARP00000050576	242
THB_HOMS1	THB	CAG07394.1	44	SINFRUP00000151911	143	ENSDARP00000053252	243
TLL1_HOMS1	TLL1	CAG03838.1	45	SINFRUP00000152420	144	ENSDARP00000053819	244
TR2_HOMS1	TR2	CAG03953.1	46	SINFRUP00000152690	145	ENSDARP00000054262	245
TR4_HOMS1	TR4	CAG06739.1	47	SINFRUP00000152795	146	ENSDARP00000057122	246
VDR_HOMS1	VDR	CAG06880.1	48	SINFRUP00000153576	147	ENSDARP00000059975	247
		CAG07050.1	49	SINFRUP00000154296	148	ENSDARP00000061524	248
		CAG10628.1	50	SINFRUP00000156329	149	ENSDARP00000062284	249
		CAF99925.1	51	SINFRUP00000157997	150	ENSDARP00000062419	250
		CAG05777.1	52	SINFRUP00000158010	151	ENSDARP00000063213	251
		CAG00032.1	53	SINFRUP00000158768	152	ENSDARP00000064037	252
		CAG00249.1	54	SINFRUP00000158774	153	ENSDARP00000064044	253
		CAF99979.1	55	SINFRUP00000158776	154	ENSDARP00000065377	254
		CAG07471.1	56	SINFRUP00000161153	155	ENSDARP00000066643	255
		CAG02755.1	57	SINFRUP00000161308	156	ENSDARP00000067473	256
		CAG02756.1	58	SINFRUP00000162567	157		
		CAF94134.1	59	SINFRUP00000162788	158		
		CAF96472.1	60	SINFRUP00000163447	159		
		CAF96473.1	61	SINFRUP00000163457	160		
		CAF91378.1	62	SINFRUP00000163483	161		
		CAF88861.1	63	SINFRUP00000164033	162		
		CAF91926.1	64	SINFRUP00000164938	163		
		CAG13569.1	65	SINFRUP00000165123	164		
		CAF94543.1	66	SINFRUP00000165525	165		
		CAF90265.1	67	SINFRUP00000165728	166		
		CAG11044.1	68	SINFRUP00000165984	167		
		CAG12248.1	69	SINFRUP00000166348	168		
		CAG09068.1	70	SINFRUP00000166643	169		
		CAG11045.1	71	SINFRUP00000166688	170		
		CAF92683.1	72	SINFRUP00000167201	171		
				SINFRUP00000168173	172		
				SINFRUP00000168385	173		

Supplementary table 3 (*Continued*)

SINFRUP00000169072	174
SINFRUP00000169093	175
SINFRUP00000169341	176
SINFRUP00000170179	177
SINFRUP00000171165	178
SINFRUP00000171918	179
SINFRUP00000172227	180
SINFRUP00000172442	181
SINFRUP00000172695	182
SINFRUP00000172796	183
SINFRUP00000173136	184
SINFRUP00000174262	185
SINFRUP00000176119	186
SINFRUP00000176452	187



Supplementary figure 1. Phylogenetic tree of DBD and LBD domains of *Tetraodon* NRs. Tree topologies of DBD and LBDs differed with respect to association of subfamilies. NR1 is closer to NR4 in LBD tree. But NR4 is closer to NR2 and NR3 in DBD tree. Terminal branch lengths of members are shorter in DBD Tree compared to LBD tree. Sequences are indicated by codes as the serial number in supplementary table 1.



Supplementary figure 2. Phylogenetic Tree of the *Tetraodon*, fugu fish, zebra fish and human NRs. As represented in supplementary table 3, *Tetraodon* NRs are numbered below 100, fugu fish NRs from 100 to 199, zebra fish NRs numbered 200 onwards. Whereas human NRs are represent as corresponding gene names.