U.PORTO
FACULDADE DE CIÊNCIAS
UNIVERSIDADE DO PORTO

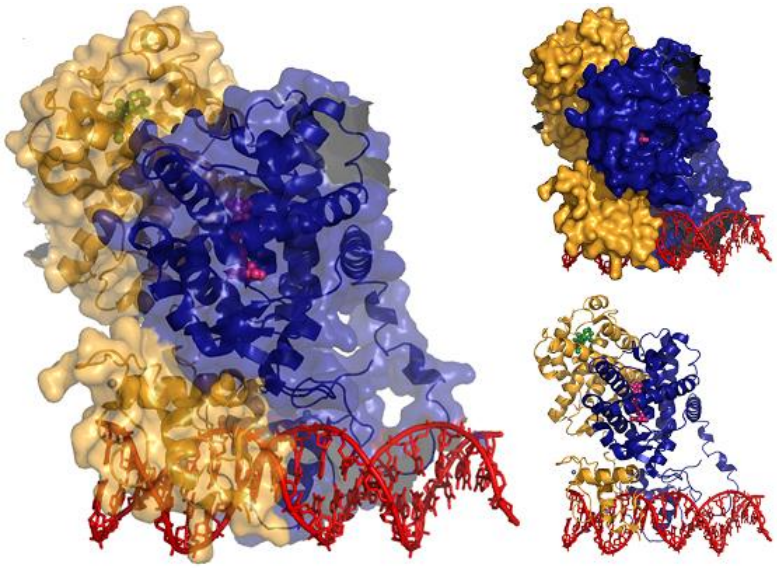
Nuclear Receptors in Metazoan lineages: the cross-talk between Evolution and Endocrine Disruption

Elza Sofia Silva Fonseca

Tese de Doutoramento apresentada à
Faculdade de Ciências da Universidade do Porto
Biologia

2020

PhD



Nuclear Receptors in Metazoan lineages: the cross-talk between Evolution and Endocrine Disruption

Elza Sofia Silva Foseca
Doutoramento em Biologia
Departamento de Biologia
2020

Orientador
Doutor Luís Filipe Costa Castro, Professor Auxiliar,
Faculdade de Ciências da Universidade do Porto,
Centro Interdisciplinar de Investigação Marinha e Ambiental (CIIMAR)

Coorientador
Professor Doutor Miguel Alberto Fernandes Machado e Santos, Professor Auxiliar,
Faculdade de Ciências da Universidade do Porto
Centro Interdisciplinar de Investigação Marinha e Ambiental (CIIMAR)

This thesis was supported by FCT (ref: SFRH/BD/100262/2014), Norte2020 and FEDER (Coral – Sustainable Ocean Exploitation – Norte-01-0145-FEDER-000036 and EvoDis – Norte-01-0145-FEDER-031342).

The present thesis is organized into seven chapters. Chapter 1 consists of a general introduction, providing an overview on Metazoa definition, and a review on the current knowledge of evolution and function of nuclear receptors and their role in endocrine disruption processes. Chapters 2, 3, 4 and 6 correspond to several projects developed during the doctoral programme presented here as independent articles, listed below (three articles published in peer reviewed international journals and one article in final preparation for submission). Chapter 5 was adapted from an article published in a peer reviewed international journal (listed below), in which I executed the methodology regarding the structural and functional analyses of rotifer RXR and I contributed to the writing of the sections referring to these analyses (Material and Methods, Results and Discussion). Finally, Chapter 7 contains a general discussion, the main conclusions and the future challenges.

I declare that I have totally or partially peformed the work included in this thesis, in close cooperation/co-authorships with supervisors and other researchers. I have total responsibiliy for the conception of Chapters 1, 5 and 7, with revisions from the supervising team. In the remaining chapters, I contributed significantly to the development of the research ideas, to produce and analyse the data presented and to the writing of all chapters.

CHAPTER 2

Fonseca E*, Machado AM, Arrondo N, Gomes-dos-Santos A, Veríssimo A, Esteves P, Almeida T, Themudo G, Ruivo R, Román E, da Fonseca RR, Santos MM, Froufe E, Pérez M, Venkatesh B, Castro LFC. Chondrichthyes Offer Unique Insights into the Evolution of the Nuclear Receptor Gene Repertoire in Gnathostomes (In preparation).

CHAPTER 3

Fonseca E*, Ruivo R*****, Lopes-Marques M, Zhang H, Santos MM, Venkatesh B, Castro LF. 2017. LXRα and LXRβ Nuclear Receptors Evolved in the Common Ancestor of Gnathostomes.Genome Biol Evol. 9(1):222-230. doi: 10.1093/gbe/evw305.

CHAPTER 4

Fonseca ESS*, Ruivo R, Machado AM, Conrado F, Tay BH, Venkatesh B, Santos MM, Castro LFC. 2019. Evolutionary Plasticity in Detoxification Gene Modules: The Preservation and Loss of the Pregnane X Receptor in Chondrichthyes Lineages. Int J Mol Sci. 20(9):E2331. doi: 10.3390/ijms20092331.

CHAPTER 5 adapted from

Lee MC*, **Fonseca E**, Park JC, Yoon DS, Choi H, Kim M, Han J, Cho HS, Shin KH, Santos ML, Jung JH, Castro LFC, Lee JS. 2019. Tributyltin Affects Retinoid X Receptor-Mediated Lipid Metabolism in the Marine Rotifer *Brachionus koreanus*. Environ Sci Technol. 53(13):7830-7839. doi: 10.1021/acs.est.9b01359.

CHAPTER 6

Fonseca ESS*, Hiromori Y*, Kaite Y, Ruivo R, Franco JN, Nakanishi T, Santos MM, Castro LFC. 2019. An Orthologue of the Retinoic Acid Receptor (RAR) Is Present in the Ecdysozoa Phylum Priapulida, Genes. 10(12):985. doi: 10.3390/genes10120985.

*First author or joint first authors

Ao astro que me guiou,

Ao anjo que me aconselhou,

Ao lar que me acolheu.

Não são os Lusíadas de Camões

Tão pouco a Odisseia de Homero,

Mas é parte da minha vida que vingou.

Elza Fonseca

"If I have the belief that I can do it, I shall surely acquire the capacity to do it even if I may not have it at the beginning."

Mahatma Gandhi

Agradecimentos

Em primeiro lugar, tenho a agradecer à Fundação para a Ciências e a Tecnologias (FCT), à Faculdade de Ciências da Universidade do Porto (FCUP) e ao Centro Interdisciplinar de Investigação Marinha e Ambiental (CIIMAR) pelo financiamento e disponivbilização de infraestruturas necessários para a realização deste trabalho.

Em segundo lugar, agradeço ao meu orientador Doutor Filipe Castro e ao meu coorientador Doutor Miguel Santos pela receção e integração nas suas equipas, bem como, por todo apoio prestado desde a conceção de todos os trabalhos, ao apoio psicológico e emocional que foram imperativos para a conclusão deste meu percurso. Um especial agradecimento a Filipe Castro, mais do que o orientador da minha tese de doutoramento, foi um conselheiro e amigo que levarei para toda a vida como exemplo a nível profissional e pessoal.

Não poderia deixar de agradecer a todos os meus colegas de laboratório pelo apoio dado tanto a nível da conceção e da execução dos trabalhos, como pelos bons momentos passados nas memoráveis pausas para almoço. Muitos tornaram-se amigos que nunca esquecerei, pois ficarão com um lugar especial no meu coração. Teresa Neuparth, Raquel Ruivo, Ana André, Ana Capitão, Joana Soares, Rita Mendes, Inês Páscoa, André Machado, Susana Barros, Mónica Lopes e Patrícia Ferreira, mais do que colegas de trabalho, são pessoas excecionais com as quais tive o privilégio de trabalhar e com as quais pude partilhar todas as minhas frustrações e desabafos. Um grande obrigada pela vossa, disponibilidade, paciência, amizade e por devolverem à humanidade alguns dos seus valores perdidos.

Não poderia deixar de agradecer a todos os membros integrantes do CIIMAR. Convosco sinto-me em família.

Por último, mas não menos importante, agradeço à minha família, pais, irmã e sobrinho, que estiveram sempre dispostos a ajudar-me a superar todas as dificuldades que foram surgindo.

A todos um enorme obrigada e a vós tudo devo.

Resumo

A multicelularidade é uma caraterística definidora dos metazoários. Neste contexto, o sistema endócrino desempenha um papel essencial na comunicação célula a célula e, consequentemente, na homeostasia destes organismos. Assim, a evolução e diversificação dos sistemas endócrinos dos metazoários representam um tópico fundamental de investigação em biologia. O sistema endócrino opera como uma rede altamente regulada, na qual os recetores nucleares (RNs) constituem componentes críticos. Os RNs são a superfamília de fatores de transcrição intracelulares mais abundante nos genomas dos metazoários. Têm como principal função a regulação da expressão de genes a jusante, normalmente após ativação por ligação específica a um ligando, modulando múltiplos processos biológicos como por exemplo a reprodução e o desenvolvimento embrionário, no seu conjunto definindo o *"anel de fisiologia dos RNs"*. Nos últimos anos, tornou-se evidente que existem diferenças cruciais entre os repertórios genéticos de RNs de vertebrados e invertebrados. Essas diferenças derivam em parte de processos genómicos, tais como, a duplicação em *tandem*, a perda secundária de genes, mutações na região codificante ou reguladora, e duplicações do genoma (DG) que ocorreram em momentos específicos da evolução dos metazoários (ex.: transição invertebrados/vertebrados), contribuindo para a evolução e diversificação dos RNs. Adicionalmente, tem-se verificado uma plasticidade estrutural de RNs ortólogos na acomodação de ligandos. Além disso, tal como as hormonas e os metabolitos resultantes da dieta, os xenobióticos podem também interagir com os RNs, mimetizando ligandos naturais. Desse modo, este grupo alargado de compostos tem a capacidade de bloquear a ação dos RNs, ou promover ora ativação ora repressão da transcrição de genes. Como são conhecidos por causar desequilíbrios fisiológicos nos metazoários, estes compostos são chamados de disruptores endócrinos (DEs). Por conseguinte, a persistência de DEs e de poluentes emergentes no meio ambiente, especialmente em ecossistemas aquáticos, tornou-se num assunto de preocupação global. Assim, de forma a esclarecer os mecanismos de disrupção endócrina por DEs que atuam via RNs em linhagens de metazoários e para evitar efeitos adversos na saúde humana e na fauna selvagem, é necessário investigar a evolução e a função dos RNs em linhagens com importantes posições filogenéticas (ex.: invertebrados vs. vertebrados, lofotrocozoários vs. ecdisozoários).

Neste contexto, os repertórios de genes de RNs, a sua caraterização funcional e exploração por DEs foram investigados em linhagens filogeneticamente informativas para determinar a história evolutiva dos RNs no sistema endócrino dos metazoários.

O estudo do repertório dos genes de RNs em condríctios, uma linhagem precoce dos gnatostomados, contribuiu para elucidar o padrão de diversificação desta superfamília de fatores de transcrição em vertebrados. Foi demonstrada a contribuição das duas rondas de DG na expansão das subfamílias de RNs. Por outro lado, os eventos independentes de perda de genes observados nos gnatostomados demonstraram uma história dinâmica de retenção diferencial de parálogos (ex.: perda de *NR1H3* em anfíbios; perda de *NR1H2* em aves e actinopterígeos; retenção de *NR1I1* em todos os vertebrados; perda de *NR1I2* em aves, répteis e elasmobrânquios; retenção de *NR1I3* em tetrápodes). Além disso, a caraterização funcional de RNs em linhagens de condríctios foi crucial para compreender os mecanismos subjacentes à evolução dos sistemas endócrinos de vertebrados.

Sendo os animais mais abundantes na Terra e, uma vez que na maioria dos casos os seus genomas não passaram por eventos de DG, os invertebrados fornecem uma perspetiva informativa da condição ancestral dos RNs. Uma mudança nas especificidades dos ligandos, bem como, um aumento na afinidade de ligação ao ligando foram observados nos ortólogos de vertebrados (ex.: esteróides para oxisteróis na evolução do recetor X do fígado – anfioxo vs. vertebrados; sensor de baixa afinidade de ácido retinóico (AR) para recetor de alta afinidade de AR na evolução do recetor do ácido retinóico – priapulídeos). Além disso, a suscetibilidade à disrupção endócrina por DEs também foi demonstrada em invertebrados, apesar das diferenças nos repertórios genéticos de RNs em comparação com os dos vertebrados e das diferenças na interação entre o par ligando-recetor quando comparado aos modelos clássicos descritos (ex.: perturbação lipídica mediada por recetores de retinóide X em rotíferos expostos a tributil-estanho).

Em conclusão, a investigação da história evolutiva dos RNs revelou a importância da realização de análises evolutiva e comparativa para obter uma visão holística e realista da evolução e função dos RNs no sistema endócrino dos metazoários e a sua exploração por xenobióticos.

Palavras-chave: Recetores nucleares, disrupção endócrina, disruptores endócrinos, evolução, Metazoa, gnatostomados, Condríctios, Ecdysozoa, Priapulida, rotífero, recetor X do fígado, recetor pregnano X, recetor de retinóide X, recetor do ácido retinóico

Summary

A key trait of the Metazoa is their multicellular nature. In this context, the endocrine system is fundamental to allow cell-to-cell communication and, consequently, the organism homeostasis. Thus, the evolution and diversification of the metazoan endocrine system represents a fundamental research topic in Biology. The endocrine system operates as a highly regulated network, in which nuclear receptors (NRs) are critical components. NRs are the most abundant superfamily of metazoan-specific intracellular transcription factors. They regulate the expression of downstream genes mostly upon binding to ligands, modulating important biological processes such as reproduction or embryonic development, coeherently defining the "*NR ring of physiology*". Over the past years, it has become evident that there are crucial differences in the NR gene repertoires between vertebrates and invertebrates. These differences are mostly due to genomic processes such tandem gene duplication, gene loss, mutations and whole-genome duplications (WGD) that occurred at the invertebrate/vertebrate transition, contributing to evolution and diversification of NRs in Metazoa. Additionally, the ligand pocket of NRs has revealed binding plasticity among metazoan lineages. Moreover, similar to hormones and dietary metabolites, xenobiotics may interact with NRs and mimic natural ligands, blocking the NRs or promoting either activation or repression of gene transcription. As they are known to cause physiological imbalances in metazoans, such compounds are called endocrine-disrupting chemicals (EDCs). In this way, the persistence of EDCs and emerging pollutants in the environment, especially in aquatic ecosystems, has turned into a subject of global concern. Thus, to improve the understanding of the disruption of the endocrine system by EDCs acting *via* NRs across metazoan lineages and to prevent deleterious effects in human health and wildlife, it is mandatory to investigate the evolution and function of NRs focusing on lineages placed at key phylogenetic positions (e.g. invertebrates vs. vertebrates, lophotrochozoans vs. ecdysozoans).

In this way, the NR gene repertoires, the functional characterization of NRs and their exploitation by EDCs were investigated in selected phylogenetic informative lineages to build the evolutionary history of the studied NRs in the endocrine system of metazoans.

The study of the NR gene repertoire in Chondrichthyes, an early branching lineage of the gnathostomes, contributed to elucidate the diversification pattern of this superfamily of transcription factors in vertebrates. The two rounds of WGD were shown to contribute to the expansion of NR subfamilies. On the other hand, the independent

events of gene loss observed in gnathostomes demonstrated a dynamic history of differential paralog retention (e.g. loss of *NR1H3* in amphibians; loss of *NR1H2* in birds and ray-finned fishes; retention of *NR1I1* in all gnathostomes; loss of *NR1I2* in birds, reptiles and elasmobranchs; retention of *NR1I3* in tetrapods). Furthermore, the functional characterization of NRs in chondrichthyan lineages was crucial to comprehend the mechanisms underlying the evolution of endocrine systems of vertebrates.

As the most abundant group of animals on Earth and as their genomes have not experienced WGD in most cases, invertebrates provide an important perspective at the ancestral condition of NRs. A shift in ligand preferences, as well as, an increase in ligand-binding affinity were observed in their vertebrate orthologs (e.g. steroidal compounds to oxysterols in liver X receptor evolution – amphioxus vs. vertebrates; retinoic acid (RA) low-affinity sensor to RA high-affinity receptor in retinoic acid receptor evolution – priapulids). Moreover, the susceptibility to endocrine disruption by EDCs was also exmained in invertebrates, despite the differences in the NR genes repertoire compared to vertebrates and the differences in the interaction of ligand-receptor couple compared to the classical described models (e.g. retinoid X receptor-mediated lipid perturbation in rotifers exposed to tributyltin).

Overall, the investigation of the evolutionary history of NRs revealed the importance of conducting comparative evolutionary analyses to achieve a broader and more realistic vision of the evolution and function of NRs in the endocrine system of metazoans and their exploitation by xenobiotics.

Keywords: Nuclear receptors, endocrine disruption, endocrine-disrupting chemicals, evolution, Metazoa, gnathostomes, Chondrichthyes, Ecdysozoa, Priapulida, rotifer, liver X receptor, pregnane X receptor, retinoid X receptor, retinoic acid receptor

Contents

List of Tables

List of Figures

CHAPTER 5 – The Evolution of the Retinoid X Receptor in Metazoa: Insights into Lipid Metabolism Disruption in a Marine Rotifer

Supplementary Material

CHAPTER 6 – An Ortholog of the Retinoic Acid Receptor (RAR) Is Present in the Ecdysozoa Phylum Priapulida

Supplementary Material

CHAPTER 7 – General Discussion and Conclusions

List of Abbreviations

A

AncNR	Nuclear receptor ancestor
AncSR	Steroid receptor ancestor
ANOVA	Analysis of variance
AF	Activation function
AR	Androgen receptor
ATRA	All-*trans* retinoic acid

B

BLAST	Basic local alignment search tool
bp	baise pairs
BPA	Bisphenol A
BUSCO	Benchmarking Universal Single-Copy Orthologs

C

9cisRA	9-*cis* retinoic acid
CAR	Constitutive androstane receptor
Chr	Chromosome
CNS	Central nervous system
CO1, COI	Cytochrome c oxidase I
Cyp	Cytochromes P450
Cys	Cysteine

D

DBD(s)	DNA-binding domain(s)
DMEM	Dulbecco's modified Eagle's medium
DMSO	Dimethyl sulfoxide
DNA	Deoxyribonucleic acid
DR(s)	Direct repeat(s)

E

E2	17β-estradiol
EC50	Half maximal effective concentration

EcR	Ecdysone receptor
EDC(s)	Endocrine-disrupting chemical(s)
EE2	17α-ethinylestradiol
ER(s)	Estrogen receptor(s)
ER(s)	Everted repeat(s)

F

FXR(s)	Farnesoid X receptor(s)

G

Gb	Gigabytes
Gly	Glycine
GMQE	Global Model Quality Estimation
GR	Glucocorticoid receptor

H

H	Helix
HNFA(s)	Hepatocyte nuclear factor 4 receptor(s)
HRE(s)	Hormone response element(s)

I

I, Ile	Isoleucine
IR(s)	Inverted repeat(s)

J

JGI	Joint Genome Institute

L

L	Leucine
LBD(s)	Ligand-binding domain(s)
LBP(s)	Ligand-binding pocket(s)
LXR(s)	Liver X receptor(s)

M

mg	milligram
mL	milliliter
mM	millimolar
MAFFT	Multiple alignment using fast fourier transform
MR	Mineralocorticoid receptor
MY	Million years
MYA	Million years ago
μm	Micrometer
μM	Micromolar

N

nM	Nanomolar
NR(s)	Nuclear receptor(s)
NSCs	Neurosecretory cells

O

OIST	Okinawa Institute of Science and Technology

P

PE	Paired-end
PCR	Polymerase chain reaction
PNR	Photoreceptor-specific nuclear receptor
POPs	Persistent organic pollutants
PPAR(s)	Peroxisome proliferator-activated receptor(s)
PR	Progesterone receptor
PXR	Pregnane X receptor

Q

QMEAN	Qualitative Model Energy Analysis

R

2R WGD	Two rounds of whole-genome duplication
3R WGD	Three rounds of whole-genome duplication
4R WGD	Four rounds of whole-genome duplication
RA	Retinoic acid

RACE	Rapid amplification of cDNA ends
RAR(s)	Retinoic acid receptor(s)
RAREs	Retinoc acid responsive elements
Rev-Erb(s)	Reverse ErbA receptor(s)
RNA	Ribonucleic acid
ROR(s)	RAR-related orphan receptor(s)
RXR(s)	Retinoid X receptor(s)

S

SE	Standard error
SEM	Standard error of the mean
SRA	Sequence read archive
SR(s)	Steroid receptor(s)
Supnrs	Supplementary nuclear receptors

T

TBT	Tributyltin
THR(s)	Thyroid hormone receptor(s)
TNC	Trans-nonachlor
TPT	Triphenyltin
TR	Thyroid hormone receptor

U

UAS	Upstream activation sequence

V

VDR	Vitamin D receptor

W

WGD(s)	Whole-genome duplication(s)

CHAPTER 1 - General Introduction

1. General Introduction

1.1. The Metazoa

The Swiss botanist brothers Johann and Gaspard Bauhin published *De Plantis a Divis Sanctisve Nomen Habentibus* in 1591 and *Phytopinax* in 1596, respectively, where they described and classified thousands of plants, having partially developed the binomial nomenclature (a 2-part naming method) (Verma 2011; Webster 1970). Almost 200 years later, Carl Linnaeus (*Carolus Linnæus* in Latin), a Swedish botanist, zoologist and physician, created the Linnaean taxonomy, which consists in a binomial nomenclature and a hierarchical classification system, still in use nowadays but with some adaptations. In 1735, he published the first edition of *Systema Naturae*, where he explained his novel system of nature, which was then popularised with the further publications of *Species Plantarum* (1753) and the tenth edition of *Systema Naturae* (1758) (Calisher 2007). Thus, the so-called "father of modern taxonomy" classified nature in three Kingdoms (Animalia, Vegetabilia and Mineralia), the highest hierarchical levels (ranks) within the system. Each kingdom was divided in classes, three for the Mineral kingdom (Petræ, Mineræ and Fossilia), twenty four for the Plant kingdom (Monandria, Diandria, Triandria, Tetrandria, Pentandria, Hexandria, Heptandria, Octandria, Enneandria, Decandria, Dodecandria, Icosandria, Polyandra, Didynamia, Tetradynamia, Monadelphia, Diadelphia, Polyadelphia, Syngenesia, Gynandria, Monoecia, Dioecia, Polygamia and Cryptogamia) and six for the Animal kingdom (Mammalia, Aves, Amphibia, Pisces, Insecta and Vermes) (Linnaeus 1758). Classes were divided into orders and them, in turn, into genera (singular: genus) and then species (Linnaeus 1758; Calisher 2007). Since Linnaean taxonomy, the taxonomic groups and the principles behind them have been significantly modified, as the knowledge and number of described species increased. In 1809, the French biologist Jean-Baptiste Lamarck published the *Philosophie Zoologique*, where he reclassified invertebrates. He maintained the four vertebrates groups Mammalia, Aves, Amphibia and Pisces, but separated the Linnaean class Vermes into molluscs, cirripedes, annelids, crustaceans, worms, radiates, polyps, and infusorians and separated arachnids from Linnaean class Insecta, creating a total of 14 groups (Gould 2011). Later in 1866, the German zoologist Ernst Haeckel introduced the term phyla (singular phylum) as the principal subdivisions of the Animal kingdom and thus at the same level that Linnaeus' classes, having recognized six phyla, Coelenterata, Echinodermata, Articulata, Mollusca, Vertebrata and Spongiae (Valentine 2004). Lamark removed the Mineral kingdom from taxonomy and divided the Animal kingdom into, Metazoa for the multicellular animals, and Protozoa

(Protista) for the single-celled animal and Spongiae (sponges) (Gould 2011; Valentine 2004). Since then, Metazoa became a synonym of Animalia.

In the twentieth century, Herbert Copeland argued for the separation of the organisms without nuclei (simple bacteria) from Protista kingdom, suggesting Monera as the fourth kingdom of life (Copeland 1938) and Robert Whittaker proposed the separation of fungi from Protista kingdom, creating the Fungi kingdom (Whittaker 1969, 1959), dividing the living world into Animalia, Plantae, Fungi, Protista and Monera. However, in 1977 Carl Richard Woese and George Fox distinguished archeabacteria from bacteria (eubacteria) based on ribosomal RNA analysis (Woese & Fox 1977). Then, in 1990, Woese proposed three-domain system, Eucarya, Bacteria and Archea, as the broadest level of life (Woese et al. 1990). Few years later, aiming to establish the new kingdom Chromista (Cavalier-Smith 1981), Thomas Cavalier-Smith proposed to reassume the two-empire system, Procaryota and Eucaryota, firstly suggested by Edouard Chatton (Chatton 1925, 1938). Within each empire or superkingdoms there are the six kingdoms: the kingdom Bacteria (which comprised Archeabacteria as an infrakingdom) in Procaryota and the kingdoms Animalia, Plantae, Fungi, Protozoa and Chromista in Eucaryota (Cavalier-Smith 1998). More recently, the classification of life forms was revised by Cavalier-Smith and his collaborators reintroducing the division of Procaryota into two kingdoms, Bacteria (=Eubacteria) and Archaea (=Archebacteria) (Ruggiero et al. 2015a, 2015b), based on the consensus in the Taxonomic Outline of Bacteria and Archaea and the Catalogue of Life (Roskov et al. 2019) (**Figure 1.1**).

In the frame of this thesis the focus will be on the Animal kingdom or Metazoa. The etymology of the word "animal" comes from the Latin *animalis* and means having breath, having soul or living being (Cresswell 2010). To discriminate if an organism is an animal or not, we should look carefully to some biological features described below and to its evolutionary ancestor. To be part of Metazoa an organism should comprise the following characteristics:

- eukaryotic (cells with a nucleus enclosed within membranes) and multicellular (differentiation of cells into specialized cells);
- heterotrophic nutrition (get the energy to survive from organic material);
- mobility (ability to spontaneously move their bodies);
- sexual reproduction through the generation of sperm and egg cells;
- blastulation (formation of blastula during embryonic development).

All these features are necessary, with few exceptions, to classify an organism as an Animal, but they are not sufficient on their own, since some of them are shared with other kingdoms.

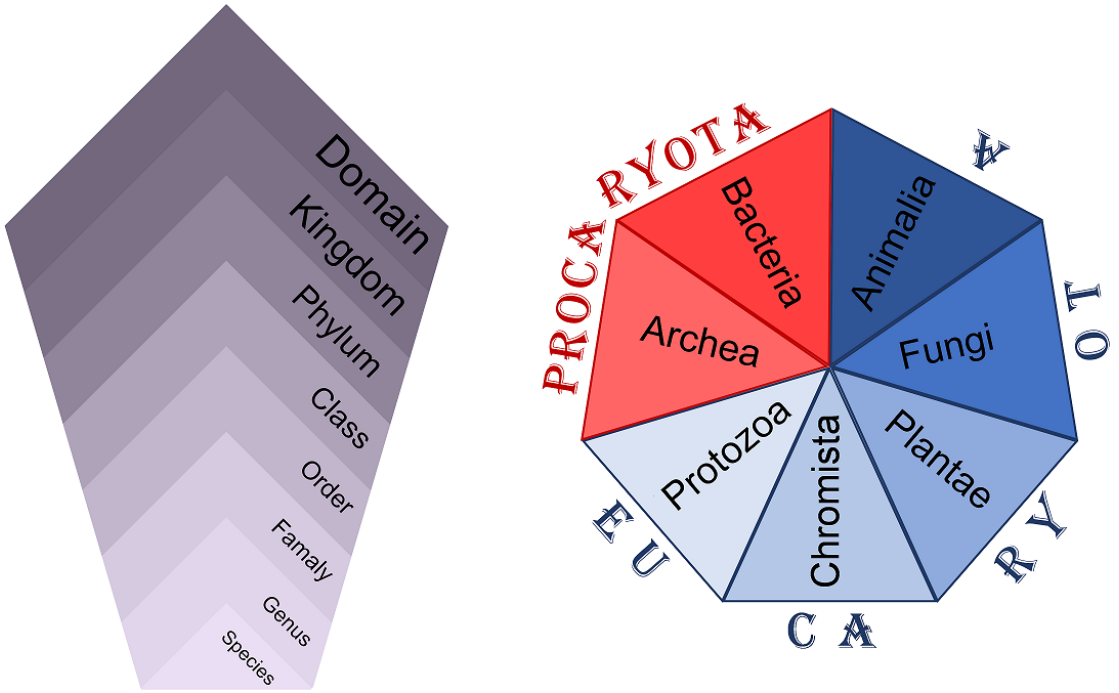


Figure 1.1. Biological classification of organisms in taxonomic ranks and modern classification of life.

Unlike Bacteria and Archaea that are unicellular prokaryotes and protists that are unicellular eukaryotes, plants and fungi are also multicellular eukaryotes (Holland 2011). Nevertheless, plants and fungi do not fulfil other requirements to be catalogued in Metazoa. Unlike animals, plants and fungi have cell walls, plants are autotrophic and fungi are decomposers. Regarding motility, all animals are motile with the exception of sponges (Leys, SP, Degnan 2001), corals (Whalan et al. 2015), most bivalves (Kamino 2008; Takeuchi et al. 2016) and barnacles (the unique sessile crustacean) (Kamino 2008), which are mobile during the larval stage and, after settlement, became sessile in the adult stage. However, we now know that some parts of plants (Volkov et al. 2017; Forterre 2013), slime moulds (protists) (Hoppe & Kutschera 2015) and fungal zoospores (Swafford, AJM & Oakley 2018; Halsall 1976) are also able to move. Sexual reproduction is also observed in plants and fungi, but animals produces gametes (sperm and egg cells) with several sizes which contributed for the evolution of animal behaviour (Holland 2011). Interestingly, most animal phyla are capable of asexual reproduction, mainly agametic reproduction (fragmentation[1] and budding[2]) and parthenogenesis[3] (de Meeûs et al. 2007). Fragmentation is more common in sponges and corals (Padua et al. 2016;

[1] **Fragmentation:** asexual reproduction in multicellular organisms in which an organism is split into fragments and each of them will develop into matured, fully grown individuals that are identical to their parents.
[2] **Budding:** asexual reproduction in which a new organism is formed from small parts of the parent without the division of the parent individual.
[3] **Parthenogenesis:** asexual reproduction in which growth and development of embryos occur in unfertilized eggs

Lirman 2000), but also occurs in annelids and echinoderms (Miyachi et al. 2005; Dolmatov et al. 2018), whereas budding is very common in cnidarians (hydras, jellyfishes) (Shostak & Kankel 1967; Hubot et al. 2017), phoronids (Temereva 2017; Margulis & Chapman 2009) and urochordates (Kawamura et al. 2013). In the case of parthenogenesis, it occurs frequently in Metazoa (de Meeûs et al. 2007; Fields et al. 2015), being universal in rotifers (Welch & Meselson 2001), very common in hexapods (Normark 2003; Harker 1997; Suomalainen et al. 1976; Arakaki et al. 2001) and in some vertebrates, such as, cartilaginous and bony fishes (Beukeboom & Vrijenhoek 1998; Feldheim et al. 2010; Spuway 1953; Dudgeon et al. 2017), amphibians (Beukeboom & Vrijenhoek 1998), reptiles (Hall 1970; Groot et al. 2003) and birds (Ramachandran & McDaniel 2018). Finally, blastulation is a developmental biological process uniquely observed in animals. During embryonic development, a hollow sphere of epithelial cells, the blastula, is formed by successive cell divisions from a single cell, the fertilized egg. This process is imperial in animals to allow cell differentiation to further elaborate tissue and organ specialization (Holland 2011).

Yet, to precisely classify a living being as an animal we should have in mind that Metazoa are a monophyletic group. Thus, the set of organisms encompassed in this group share a common evolutionary origin. So, organisms that possess an animal biologic criterion but do not share a common ancestry with animals cannot be grouped in the Metazoa. In the same way, organisms that share a common ancestor with other animals are categorized in Metazoa even if they had lost some apomorphies (Holland 2011; Gould 2011).

Like animals, plants and fungi are multicellular eukaryotes. Curiously, fungi are more closely related to animals than to plants, and together with choanoflagellates, they form the group Opisthokonta (Holland 2011). Choanoflagellates are unicellular aquatic organisms with a single apical flagellum that can live solitary or as colonies in which each cells can acquire different shapes and functions despite feeding individually (Nielsen 2012; Holland 2011). Interestingly, these organisms are very similar to the feeding cells of sponges, the choanocytes. Molecular and morphological analyses confirmed that choanoflagellates are the sister group of Metazoa (King et al. 2008) and the last unicellular common ancestor shared by choanoflagellates and metazoans was more than 600 million years ago (MYA), in the late Precambrian period. Thus, it was proposed that the most recent ancestor of all metazoans was probably a microscopic ball of single flagellum cells, called choanoblastaea (Holland 2011; Nielsen 2012; King et al. 2008). However, a recent study demonstrated that the choanocyte cells of sponges may not actually be homologous of choanoflagellates and the maintenance or convergence of a

collar-flagellum system during 600 MY was an adaptation to optimize fluid flow (Mah et al. 2014).

The reason why multicellular eukaryotes evolved remains unsolved, but it is now clear that multicellularity provided some advantages to organisms, since it enables a division of labour. As cells may transport nutrients between them through junction molecules, ones can be specialized on feeding and others can be specialized on specific functions, such as, grow, sensation, contraction or secretion (Holland 2011; Nielsen 2012). The evolution of the first multicellular ancestor allowed diversification and radiation of metazoans on Earth. However, the time when the first animals emerged remains poorly understood (Cunningham et al. 2017; Budd & Jensen 2017). Most animal fossil records date from the Cambrian Period at 541 MYA and it was thought that during the early Cambrian almost all animal phyla diverged, an event called Cambrian explosion (Cunningham et al. 2017; Budd & Jensen 2017). Nevertheless, a fossil record of an Ediacaran biota (organisms that lived during the Ediacaran Period) was found in Trezona Formation of South Australia (Cunningham et al. 2017; Budd & Jensen 2017; Maloof et al. 2010). This fossil was from a *Dickinsonia* specimen with approximately 560 MY that was proved to be a metazoan (Hoekzema et al. 2017). Nowadays, there are some evidences that point to emergence of metazoans prior to the Cambrian Period, in the Ediacaran Period (Cunningham et al. 2017; Budd & Jensen 2017; Hoekzema et al. 2017; Maloof et al. 2010).

The evolution of life on Earth led to the appereance of multiple forms, coinciding in many cases with the expansion of genetic toolkits (Carroll 2001). Apart from external factors (environment and ecology), genome novelties were crucial for the emergence of metazoans (Paps & Holland 2018). The metazoan-specific genes were clustered in 25 novel core homology groups, containing genes related to cell adhesion (e.g. fermitin and liprin), cell cycle (*MADD* and *RUN*), signalling (e.g. Wnt and TGF-β pathways, receptors, homeobox genes) and gene regulation (e.g. transcription factors), all classical functions linked to animal multicellularity (Paps & Holland 2018). Interestingly, these new genes were highly retained during evolution of the metazoans, suggesting their critical role in the appearance of multicellularity (Paps 2018).

Today, around 1.296.192 of metazoan species are catalogued belonging to 33 phyla of the Animal kingdom (Ruggiero et al. 2015a; Roskov et al. 2019). Given the high biodiversity of the Animal kingdom, I will briefly introduce the main characteristics of some lineages.

Metazoa can be divided in two groups, the Bilateria and “non-bilaterian” (**Figure 1.2**). Non-bilaterians have no bilateral symmetry and they comprise the four early- diverged

phyla Porifera (sponges), Ctenophora (comb jellies), Placozoa (trichoplax) and Cnidaria (sea anemones, corals, jellyfish) (Lanna 2015; Holland 2011). Sponges lack true epithelia and the gastrulation stage during embryonic development, their cells are differentiated but not organized into distinct organs and they feed by drawing water through chambers containing choanocytes (Lanna 2015; Holland 2011). Only two species were descried in Placozoa, *Trichoplax adhaerens* and *Hoilungia hongkongensis* (Roskov et al. 2019). Like sponges, Placozoa lack anterior-posterior polarity, nerve and muscle cells. They have a flat body formed by only six somatic cell types and they do not have a gut, secreting enzymes to break down food matter into nutrients for absorption through the contact of the lower epithelium with the substratum (Lanna 2015; Holland 2011; Nielsen 2012). Cnidaria and Ctenophora, have two germ layers (ectoderm and endoderm) and are so called diploblastic (Lanna 2015). During the gastrulation stage, Cnidaria developed a single opening gut and Ctenophora developed a gut with an anal pore (primitive anus). Both phyla have radial symmetry, nerve cells arranged in a network around the body and, when adults, they have tentacles to capture preys (Nielsen 2012; Holland 2011; Lanna 2015). For these reasons, it has been accepted that Ctenophora diverged after sponges or diverged as cnidarians relatives (Pisani et al. 2015). However, it is a matter of controversy if this phylum is the earliest branching lineage of the metazoan tree (Moroz et al. 2014; Telford et al. 2016). A recent and more complete study, established Ctenophora as the sister group of all other metazoans and proposed that the most recent common ancestor of all metazoans was a simultaneous hermaphrodite with smooth muscles, bioluminescence, tentacles and a pelagic lifestyle. In this way, as simple and nerveless animals, the sponges evolved from a more complex animal, simplifying secondarily (Whelan et al. 2017).

Bilateria comprise the remaining metazoan phyla (**Figure 1.2**) and, as the name indicates, includes animals with bilateral symmetry. Bilateral symmetry is defined by three axes, front (anterior) to back (posterior), top (dorsal) to bottom (ventral), and left to right, in which just the left and the right plans are symmetric. In other words, a body plan with bilateral symmetry can be divided into mirror images along a central axis by just a and Nephrozoa (**Figure 1.2**). Xenacoelomorpha only have a single opening to a blind gut, a ciliated overgrown epidermis and a nerve net. Despite the similarities with Cnidaria, Xenacoelomorpha display a bilateral symmetry and three germ layers characteristic of Bilateria (Hejnol & Pang 2016; Cannon et al. 2016; Gavilán et al. 2019). The precise phylogenetic position of this clade is still under debate, having been

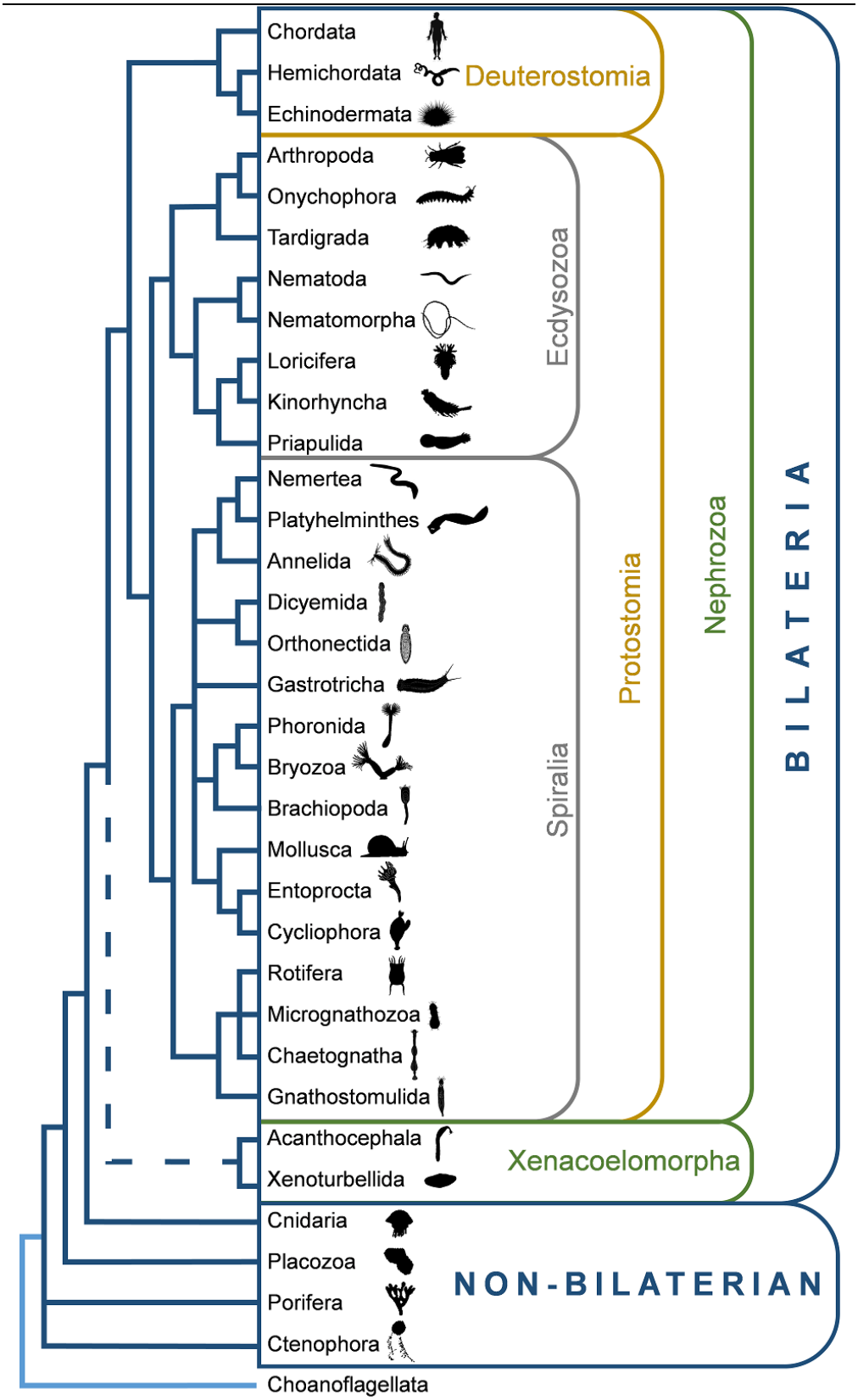


Figure 1.2. Phylogenetic tree of Metazoa and Choanoflagellata. Division of metazoan into major clades. Based on (Cannon et al. 2016; Giribet & Edgecombe 2017; Marlétaz et al. 2019; Telford 2019).

proposed as a member of deuterostomes in Nephrozoa clade (Philippe et al. 2007) or single line, creating left- and right-hand sides (Nielsen 2012; Holland 2011). As more complex animals, Bilateria can be divided in two major clades: Xenacoelomorpha as the earliest bilaterian, being the sister group of Nephrozoa (Hejnol et al. 2009; Hejnol & Pang 2016; Cannon et al. 2016). Thus, the morphological and genomic features of Xenacoelomorpha would provide insights into the bilaterian evolution. Nephrozoa comprises the remaining bilateral animals (Protostomes and Deuterostomes) (**Figure 1.2**) that possess coelom (the body cavity which contains the digestive tract), nephridia (excretory organs) and nerve cords (Jondelius et al. 2002). Protostomia is characterised by several features, such as, spiral cleavage, ventral nerves, schizocoelous development, trochophora larva, protonephridia and a blastopore that becomes the mouth; but not all of these features are observed in all of the animals belonging to this clade (Nielsen 2012). The names Protostomia and Deuterostomia were introduced by Grobben in 1908, when he described that during the embryonic development, the blastopore becomes the mouth in the Protostomia, while it becomes the anus in the Deuterostomia (Grobben 1908). Nevertheless, this feature is not observed in all Protostomes and the presence of ventral nerve cords is the character most observed in Protostomia (Nielsen 2012). Deuterostomia is mostly characterised by radial cleavage, neural tube, enterocoelus development, dipleurula larva, pharyngeal slits and a blastopore that becomes the anus (Nielsen 2012; Holland 2011). Protostomia is divided into Spiralia in which spiral cleavage is observed in most phyla, such as, Annelida, Entoprocta, Gnathostomulida, Mollusca, Nemertea and Platyhelminthes (Hejnol 2010) and into Ecdysozoa that includes the animals that grow by ecdysis (moulting of the cuticle) and that lack spiral cleavage and ciliated epithelium (Nielsen 2012; Ewer 2005). Deuterostomia is divided into Ambulacraria which comprises Echinodermata (animal with pentameric body plan, such as, sea urchins and starfish) and Hemichordata (acorn worms and pterobranchs) both phyla with dipleurula larva (Nielsen 2012) and into Chordata which comprises Cephalochordata, Tunicata and Vertebrata subphyla, possessing all the notochord, the neural tube, the longitudinal muscles along the notochord (used in locomotion) or a tail and the ciliated pharyngeal gill slits (Nielsen 2012).

The Metazoa comprises a very diverse kingdom, with multiple morphological forms and tantalizing adaptations. Importantly, the sampling of taxa and the choice of organism are of critical importance to produce comparative approaches and genome analyses. In this way, a phylogenetic framework is crucical to devise and unfold plausible and testable scenarios about past events sculpting extant phenotypes (Soltis & Soltis 2003).

1.2. Nuclear Receptors

Nuclear receptors (NRs) are the most abundant family of intracellular transcription factors only found in metazoans (Mangelsdorf et al. 1995; Degnan et al. 2009). These transcription factors are termed “receptors” because they are mostly able to bind to ligands and “nuclear” because they translocate to the nucleus bound to the ligand or in the free state to mediate their function as transcription factors (Bunce & Campbell 2010). Most NRs share a common structural architecture organized into five or six modular regions (A to F), sharing variable degrees of homology (Germain et al. 2006; Laudet & Gronemeyer 2002) (**Figure 1.3A**). The poorly conserved N-terminal A/B region contains

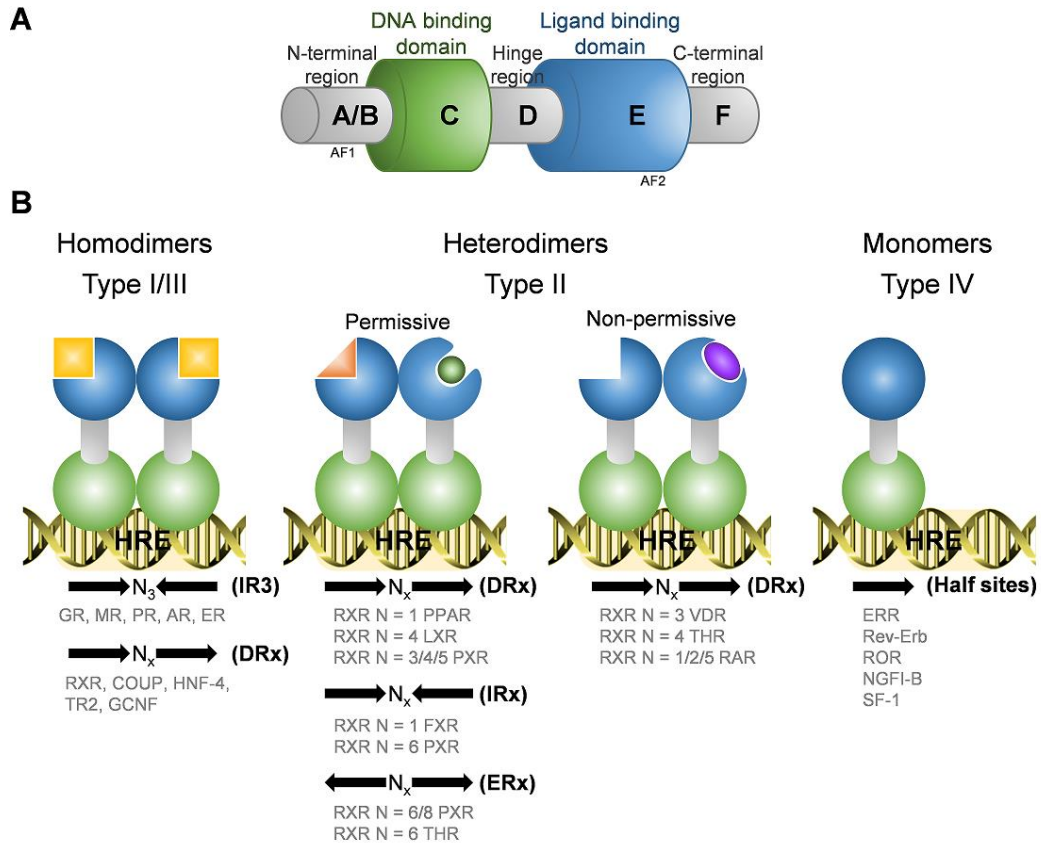


Figure 1.3. Modular structure and DNA binding mechanism of nuclear receptors. **(A)** Typical organization of NRs in six domains: A/B domain in the N-terminal region, containing the AF1 domain; C domain or the highly conserved DNA binding domain; D domain or hinge region; E domain or the conserved ligand binding domain, containing the AF2 domain; F domain not present in all NRs and with unknown function. **(B)** Three types of NRs recognition and binding to hormone response elements (HREs): homodimers, heterodimers and monomers; NRs classification according their DNA binding mechanism: type I, II, III and IV; Examples of NRs and their core recognition motif preference: Direct repeat (DR), inverted repeat (IR) or everted repeat (ER) of the consensus half-site motif (x represents the number of bp separating the two half-sites).

a transcriptional activation function domain (AF1), which the activity is controlled by ligand binding; but can be constitutively active in fusion proteins with heterologous receptors. The A/B region can interact with coactivators or other transcription factors and be subject of alternative splicing and post-translational events (e.g. phosphorylation), being isoform-specific (Germain et al. 2006, 2003). The C region is the highly conserved DNA binding domain (DBD) responsible for recognition and binding to specific DNA sequences called hormone response elements (HREs) in the promotor region of a target gene. The DBD is composed of two tetracysteine zinc-finger motifs (the C-terminal motif [Cys-X5-Cys-X9-Cys-X2-Cys] and the N-terminal motif [Cys-X2-Cys-X13-Cys-X2-Cys], in which the 4 cysteine residues chelate one Zn^{2+} ion) exclusive to NRs and of sequence elements known as P-, D-, T- and A-boxes (Laudet & Gronemeyer 2002; Germain et al. 2003, 2006; Heldin et al. 2016). These structural elements contribute to HRE specificity (P-box in the first zinc-finger motif), to dimerization (D-box in the second zinc-finger motif) and to contact with the DNA backbone (T- and A-boxes) (Germain et al. 2006, 2003). The variable D region function as a flexible hinge between C and E regions. The flexibility of this region allows C and E regions to adopt different conformations and consequently, contributes for the selection of the DNA binding sites (Billas & Moras 2013). The nuclear localization signal is located D region (Germain et al. 2006, 2003). The E region or ligand binding domain (LBD) is less conserved than DBD and is composed of 12 α- helices (H1-12) folded as an antiparallel tri-layered sandwich: the first layer comprises helices H1, H2 and H3 with a varying length loop between H1 and H2 or, when H2 is absent, H1 and H3; the second layer, helices H4, H5, H8, H9 and H11, is sandwiched between the first layer and the third layer containing helices H6, H7 and H10; a β-turn connects H5 to H6 (Laudet & Gronemeyer 2002). The H12 is connected to H11 by a flexible loop which allows H12 to swing upon binding to a ligand, trapping the ligand and stabilizing the conformation of the NR (Hellal-Levy et al. 2000; Laudet & Gronemeyer 2002). In this way, the LBD mediates recognition and interaction of ligands at the ligand binding pocket (LBP), dimerization with LBDs from other NRs, interaction with co-regulators (coactivators [LxxLL motif in H4] or corepressors ([LxxxIxxxI/L motif in H1]) and ligand-dependent transactivation at the ligand-dependant activation function AF2 domain (H12), which enables the recruitment of co-regulators (Germain et al. 2006, 2003). Finally, the F region is not present in all NRs, and its function remains unclear (Germain et al. 2006, 2003).

According to sequence alignment and phylogenetic analyses of the conserved C and E domains, NRs were classified into eight subfamilies (NR1 to NR8) (**Table 1.1**) which share the same modular structure (**Figure 1.3A**), with the exception of NR7

members who have two DBDs (Wu et al. 2007; Kaur et al. 2015). Non-canonical NRs lack one of the two conserved domains (C in vertebrates and E in invertebrates) are grouped into NR0 subfamily (Huang et al. 2015; Germain et al. 2006; Laudet & Gronemeyer 2002; Committee 1999).

Nuclear receptors are typically localized in the cytoplasm or are bound to the HRE in the nucleus as repressive complexes (Heldin et al. 2016). The nuclear translocation and/or recruitment of coactivators and consequent displacement of corepressors is normally promoted by ligand binding (Germain et al. 2003, 2006; Heldin et al. 2016). The recognition and binding to the HRE occur concomitantly to the formation of homodimers or heterodimers, but can also occur as monomers (**Figure 1.3B**) (Glass 1994; Mangelsdorf & Evans 1995; Khorasanizadeh & Rastinejad 2001; Germain et al. 2003, 2006). The particular case of steroid receptors (ER, GR, PR, MR and AR) is rather different, since they appear as homodimers with no bound ligand in the cytoplasm (Germain et al. 2003, 2006). The core recognition motif consists in two hexameric core half-site motif separated by a variable numbers of base pairs (bp). These two half-sites can be arranged as direct repeats (DRs) (consensus nucleotide sequences with head-to-tail orientation), inverted repeats (IRs) (consensus nucleotide sequences with head-to-head orientation), or everted repeats (ERs) (consensus nucleotide sequences with tail-to-tail orientation) (Germain et al. 2003, 2006). The hexameric core motif 5′-PuGGTCA (Pu = A or G) is recognized by all NRs; however, the NR selectivity and the dimerization pattern are conferred by mutations, number of bp separating the two half-sites and their relative orientation (Germain et al. 2003, 2006; Glass 1994; Penvose et al. 2019). Regarding the DNA binding mechanism, NRs can be classified in four types (**Figure 1.3B**). Type I NRs (NR3 subfamily) translocate to the nucleus upon ligand binding and bind to DNA as homodimers, recognizing IR HREs (Heldin et al. 2016; Sever & Glass 2013). Type II NRs (mostly NR1 subfamily) are normally concentrated in the nucleus and bind to DNA as heterodimers with the RXR, recognizing DR, IR or ER HREs. In the absence of ligand, the heterodimer forms a complex with corepressors, repressing target gene transcription. In the presence of a ligand, corepressors dissociate from the complex, allowing the association of coactivators and the consequent target gene transcription activation; in permissive heterodimers, the ligands of both dimer partners are able to activate the complex, whereas in non-permissive heterodimers, the complex is unable to be activated by RXR ligands (Heldin et al. 2016; Sever & Glass 2013; Dawson & Xia 2012). Type III NRs (principally NR2 and NR6) differ from type I in recognizing DR HREs instead of IR HREs. Type IV NRs (NR4, NR5 and some NR1 and NR3) bind DNA as monomers and recognize half-site HREs (Heldin et al. 2016; Sever & Glass 2013).

Table 1.1. List of nuclear receptors.

Nomenclature	Abbreviation	Name
NR1A1	THRα	Thyroid hormone receptor α
NR1A2	THRβ	Thyroid hormone receptor β
NR1B1	RARα	Retinoic acid receptor α
NR1B2	RARβ	Retinoic acid receptor β
NR1B3	RARγ	Retinoic acid receptor γ
NR1B4	RAR	Retinoic acid receptor
NR1C1	PPARα	Peroxisome proliferator-activated receptor α
NR1C2	PPARβ	Peroxisome proliferator-activated receptor β
NR1C3	PPARγ	Peroxisome proliferator-activated receptor γ
NR1D1	Rev-Erbα	Reverse ErbA receptor α
NR1D2	Rev-Erbβ	Reverse ErbA receptorβ
NR1D3	E75	Ecdysone-induced protein 75
NR1D4	Rev-Erbγ	Reverse ErbA receptor γ
NR1E1	E78C	Ecdysone-induced protein 78C
NR1F1	RORα	RAR-related orphan receptor α
NR1F2	RORβ	RAR-related orphan receptor β
NR1F3	RORγ	RAR-related orphan receptor γ
NR1G1	Sex-1	Steroid hormone receptor cnr14
NR1H1	EcR	Ecdysone receptor
NR1H2	LXRβ	Liver X receptor β
NR1H3	LXRα	Liver X receptor α
NR1H4	FXRα	Farnesoid X receptor α
NR1H5	FXRβ	Farnesoid X receptor β
NR1I1	VDR	Vitamin D receptor
NR1I2	PXR	Pregnane X receptor
NR1I3	CAR	Constitutive androstane receptor
NR1J1	DHR96	Nuclear hormone receptor HR96
NR1K1	VDRα-like	VDR/PXR α
NR1K2	VDRβ-like	VDR/PXR β
NR1L	HR97	Hormone receptor-like 97
NR1M1	HR10	Hormone receptor-like 10
NR1N1	HR11	Hormone receptor-like 11
NR1O	-	-
NR1P1-11	-	-

Continued

Table 1.1. List of nuclear receptors (cont.)

Nomenclature	Abbreviation	Name
NR2A1	HNF4α	Hepatocyte nuclear factor 4 receptor α
NR2A2	HNF4γ	Hepatocyte nuclear factor 4 receptor γ
NR2A3	HNF4β	Hepatocyte nuclear factor 4 receptor β
NR2A4	HNF4	Hepatocyte nuclear factor 4 receptor
NR2B1	RXRα	Retinoid X receptor α
NR2B2	RXRβ	Retinoid X receptor β
NR2B3	RXRγ	Retinoid X receptor γ
NR2B4	USP	Ultraspiracle
NR2C1	TR2	Testicular receptor 2
NR2C2	TR4	Testicular receptor 4
NR2D1	DHR78	Nuclear hormone receptor HR78
NR2E1	TLX	Tailless homolog
NR2E2	TLL	Tailless
NR2E3	PNR	Photoreceptor-specific nuclear receptor
NR2E4	DSF	Dissatisfaction nuclear receptor
NR2E5	FAX1	Nuclear hormone receptor FAX-1
NR2F1	COUP-TFI	Chicken ovalbumin upstream promoter transcription factor I
NR2F2	COUP-TFII	Chicken ovalbumin upstream promoter transcription factor II
NR2F3	SVP	Seven-up
NR2F4	COUP-TFIII	Chicken ovalbumin upstream promoter transcription factor III
NR2F5	SVP-46	Seven-up related protein 46
NR2F6	EAR-2	V-erbA-related protein 2
NR3A1	ERα	Estrogen receptor α
NR3A2	ERβ	Estrogen receptor β
NR3B1	ERRα	Estrogen-related receptor α
NR3B2	ERRβ	Estrogen-related receptor β
NR3B3	ERRγ	Estrogen-related receptor γ
NR3C1	GR	Glucocorticoid receptor
NR3C2	MR	Mineralocorticoid receptor
NR3C3	PR	Progesterone receptor
NR3C4	AR	Androgen receptor
NR4A1	NGFIB	Nerve Growth factor IB
NR4A2	NURR1	Nuclear receptor related 1
NR4A3	NOR1	Neuron-derived orphan receptor 1
NR4A4	DHR38	Nuclear hormone receptor HR38

Continued

Table 1.1. List of nuclear receptors (cont.)

Nomenclature	Abbreviation	Name
NR5A1	SF1	Steroidogenic factor 1
NR5A2	LRH1	Liver receptor homolog-1
NR5A3	FTZ-F1α	Nuclear hormone receptor FTZ-F1 α
NR5B1	FTZ-F1β	Nuclear hormone receptor FTZ-F1 β
NR6A1	GCNF	Germ cell nuclear factor
NR6A2	DHR4	Nuclear hormone receptor HR4
NR7A	-	Two DBD receptors
NR8A1	NR8A1	Nuclear receptor 8
NR0A1	KNI	Knirps
NR0A2	KNRL	Knirps related
NR0A3	EG	Eagle
NR0A4	ODR-7	ODR-7
NR0A5	TRX	Trithorax
NR0B1	DAX1	Dosage-sensitive sex reversal-adrenal hypoplasia congenital critical region on the X chromosome, gene 1
NR0B2	SHP	Small heterodimer partner

The nature of ligands can also be used to classify NRs (**Table 1.2**). Typically, NR ligands comprise an array of small lipophilic molecules from endogenous sources (e.g. dietary lipids, vitamins, cholesterol metabolites, hormones) or from exogenous sources (e.g. xenobiotics), that act as (full, partial or inverse) agonists or antagonists (Gronemeyer et al. 2004; Mangelsdorf & Evans 1995; Mangelsdorf et al. 1995; Heldin et al. 2016; Germain et al. 2003, 2006) as they promote a structural rearrangement that allows the bind of the NR with the transcriptional machinery through the interaction of the AF2 domain with co-regulators (Escriva et al. 2004). The endocrine NRs bind hormones (e.g. steroids, thyroid hormone) and vitamins (vitamin D, Vitamin A) with high affinity. The orphan NRs are so-called because their physiological ligands have yet to be identified (Benoit et al. 2006; Mukherjee & Mani 2010). Dietary metabolites (e.g. retinoids, fatty acids, sterols, bile acids) and xenobiotics were identified as ligands of some orphan NRs, which are now categorized as “adopted” orphan or metabolic NRs. These NRs have larger LBPs than endocrine NRs, allowing them to accommodate an array of diverse molecules with low binding affinity (Benoit et al. 2006; Mukherjee & Mani 2010; Kojetin & Burris 2014; Mazaira et al. 2018).

Given the ligand diversity of NRs, they modulate almost all aspects of metazoan life. Embryogenesis, reproduction, development, homeostasis, metabolism, or immunity

Table 1.2. Classification of nuclear receptors based on their ligands.

Endocrine NRs		Metabolic NRs		Orphan NRs	
NR	Ligand	NR	Ligand	NR	Ligand
		PPARs	Fatty acids		Unknown
THRs	Thyroid hormone	LXRs	Oxysterols	DAX1	
RARs	Vitamin A	FXRs	Bile acids	SHP	
VDR	Vitamin D	PXR	Pregnane/xenobiotics	TR2/TR4	
ERs	Estrogen	CAR	Androstane/ xenobiotic	TLX/TLL/PNR	
GR	Glucocorticoids	Rev-Erbs	Haem	COUPs	
MR	Mineralocorticoids	RORs	Sterols	ERRs	
AR	Androgens	RXRs	Retinoids	NURs	
PR	Progesterone	HNF4s	Fatty acids	GCNF	
		SF1, LRH1	Phosphatidylinositols		

are examples of biological processes involving NR-mediated transcription regulation (Bookout et al. 2006; Gronemeyer et al. 2004; Germain et al. 2003). The steroid hormone receptors (GR, MR, AR, PR and ERs) are mostly expressed in reproductive organs and regulate reproduction (Bondesson et al. 2015; Belfiore & LeRoith 2018), whereas SF1 and DAX1 regulate sexual differentiation and steroidogenesis (Beuschlein et al. 2002; Parker & Schimmer 2002). Embryogenesis and development in vertebrates and invertebrates have been associated to RAR (Johnson et al. 2019; Mark et al. 2006) and EcR have been linked to insect development and metamorphosis (Uryu et al. 2015). The NURs receptors are implicated in proliferation, apoptosis and neuronal cell differentiation (Safe et al. 2016). Basal and lipid metabolism, energy homeostasis and inflammation are modulate by receptors such THRs, PPARs, LXRs and FXRs (Weiss et al. 1998; Gullberg et al. 2000; Forrest & Vennstrom 2000; Brent 2000; Marrif et al. 2005; Evans et al. 2004; Ogawa et al. 2005; Kalaany & Mangelsdorf 2006; Zelcer & Tontonoz 2006; Mello 2010; Laurencikiene & Ryden 2012; Ahmadian et al. 2013; Jiao et al. 2015). The xenobiotic and endobiotic metabolism is closely related to PXR and CAR, since they regulate transcription of *CYP* genes which encode cytochrome P450 enzymes involved in the bile acid pathway and drugs metabolism (di Masi et al. 2009). The relationship between the effect of NRs action and their tissue expression profile can be organized according to the typical function of these tissues in the so-called nuclear receptor “ring of physiology” (**Figure 1.4**). So, NRs can be distributed in two major clusters, cluster I (reproduction and central nervous system (CNS) function) which comprise NRs expressed in the CNS, reproductive organs, and adrenals, and cluster II (nutrient metabolism and immunity) which comprise NRs predominantly expressed within the gastro/enterohepatic axis and metabolic tissues, each of them divided in three subclusters (IA – steroidogenesis; IB –

reproduction and development; IC – CNS, circadian and basal metabolism; IIA – bile acids and xenobiotic metabolism; IIB – lipid metabolism; IIC - energy homeostasis and immunity) (Bookout et al. 2006).

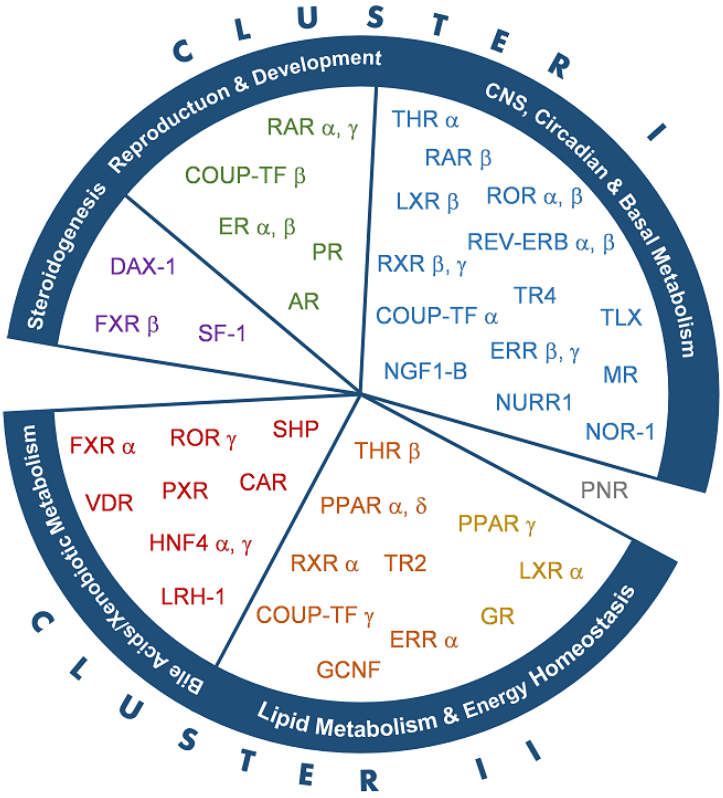


Figure 1.4. The nuclear receptor ring of physiology: relationship between NR expression, function, and physiology. Cluster I: reproduction and CNS function; cluster II: nutrient metabolism and immunity. NRs belonging to subclusters IA, IB, IC, IIA, IIB and IIC are indicated in purple, green, blue, red, orange and yellow, respectively. Note: PNR belongs to cluster II and forms its own subcluster due to its restricted expression in the eye. Adapted from (Bookout et al. 2006).

Beyond the positive and negative regulation of the expression of downstream target genes, NRs participate in the signal transduction pathways crosstalk. Thus, NRs can interact with other transcription factor resulting in a trans-repression of target genes (e.g. GR and either AP1 (c-Fos/c-Jun) or nuclear factor-κB (NF-κB) trans-repression); and they can be target of other signalling pathways that modify them and consequently alter their function either/or even label them for degradation (e.g. phosphorylation, ubiquitylation and acetylation) (Germain et al. 2003; Bunce & Campbell 2010; Gronemeyer et al. 2004). In addition to the genomic action, which can occur over minutes or hours, NRs also mediate non-genomic actions outside the nucleus. These events involve the physical interaction of the holo-NR with another protein, which expeditiously initiate a signalling cascade (e.g. modulation of the activity of ion channels, kinases,

phosphatases) (Unsworth et al. 2018; Ordóñez-Morán & Muñoz 2009; Gronemeyer et al. 2004).

The biological role of NRs in intercellular signalling and their selectivity make them major targets for drug development to treat multiple diseases, such as, inflammation, diabetes, metabolic syndrome, cardio vascular disease and cancer (Germain et al. 2003; Gronemeyer et al. 2004; Shulman & Mangelsdorf 2005; Mazaira et al. 2018; Unsworth et al. 2018).

1.3. Metazoa and Nuclear Receptor Evolution

Nuclear receptors structure is highly conserved, suggesting a common evolutionary origin. In the early 90's it was speculated that NR divergence coincided with the radiation of metazoans, as the role of NRs in cell-cell communication was crucial for metazoan evolution (Laudet et al. 1992). Moreover, it was suggested that NR genes originated by fusion of genes implicated in steroid binding in the cytosol described in yeast with DNA binding sequences (GATA and LIM zinc finger domains) (O'Malley 1989; Moore 1990; Clarke & Berg 1998). The complete genome-sequencing of non-metazoan species (plants, protists and fungi), including the sister group of metazoans, the Choanoflagellata (King et al. 2008), were unsuccessful to retrieve NR sequences. In contrast, NR-coding genes have been found in all metazoan sequenced genomes. Thus, NRs are unique to Metazoa (Laudet et al. 1992; Owen & Zelent 2000; Escriva et al. 2004).

The number of NRs retrieved from some metazoan lineages with sequenced genomes is listed in **Table 1.3**. The different numbers of NRs among several species indicates a highly species-dependent diversification. Good examples are nematodes with a lineage-specific expansion of the *HNF4* gene, leading to more than 250 supplementary nuclear receptors (supnrs) (Bertrand et al. 2004; Robinson-Rechavi et al. 2005), molluscs with the NR1P subfamily (Vogeler et al. 2014) and rotifers with the NR1O subfamily (Kim et al. 2017), apparently unique to these lineages.

Gene duplication is an important mechanism in evolution that allows gene or its duplicate to escape from natural selection, contributing to genome expansion (Ohno 2013). In fact, the variation in genome size among species suggests a tetraploidization event coinciding with the emergence of the first vertebrate at 500 MYA - the "2R hypothesis" (Ohno 2013). This hypothesis describes the two (separate) rounds of WGD (2R WGD) that the vertebrate ancestor underwent, supporting the fact that gene families are generally constituted by up to four members (paralogous genes) originated from duplication in vertebrates, contrasting to gene families (orthologous genes) with normally a unique member in invertebrates (4:1 ratio) (Putnam et al. 2008). The exact timing of each round of duplication is still discussed, but strong evidences support that both occurred before the divergence of gnathostomes (Kuraku et al. 2009; Smith et al. 2013; Mehta et al. 2013). Additionally, in the Actinopterygii lineage, a teleost-specific genome duplication (3R WGD) occurred approximately 450 MYA (Jaillon et al. 2004) after the divergence of Holostei (Amores et al. 2011), with a salmonid-specific genome duplication (4R WGD) occurring in the stem salmonid ancestor (Macqueen & Johnston 2014).

Table 1.3. Number of nuclear receptor in some metazoan lineages.

Phylum	Species	NRs number	References
Porifera	*Amphimedon queenslandica*	2	(Bridgham et al. 2010)
Ctenophora	*Mnemiopsis leidy*	2	(Reitzel et al. 2011)
Placozoa	*Trichoplax adhaerens*	4	(Srivastava et al. 2008)
Cnidaria	*Nematostella vectensis*	17	(Reitzel & Tarrant 2009)
Nematoda	*Caenorhabditis elegans*	284	(Sluder et al. 1999; Robinson-Rechavi et al. 2005)
	Caenorhabditis briggsae	268	(Stein et al. 2003; Bertrand et al. 2004)
Arthropoda	*Drosophila melanogaster*	21	(King-Jones & Thummel 2005)
	Daphnia pulex	25	(Thomson et al. 2009)
	Daphnia magna	26	(Litoff et al. 2014)
Platyhelminthes	*Schistosoma mansoni*	21	(Wu et al. 2006)
Rotifera	*Brachionus koreanus*	32	(Kim et al. 2017)
	Brachionus plicatilis	29	(Kim et al. 2017)
	Brachionus rotundiformis	32	(Kim et al. 2017)
	Brachionus calyciflorus	40	(Kim et al. 2017)
Annelida	*Capitella teleta*	27	(Bodofsky et al. 2017)
Mollusca	*Crassostrea gigas*	43	(Vogeler et al. 2014; Huang et al. 2015)
	Lottia gigantea	33	(Kaur et al. 2015)
	Biomphalaria glabrata	39	(Kaur et al. 2015)
Echinodermata	*Strongylocentrotus purpuratus*	33	(Howard-Ashby et al. 2006)
Chordata	*Ciona intestinalis*	17	(Yagi et al. 2003)
	Branchiostoma floridae	33	(Lecroisey et al. 2012; Schubert et al. 2008)
	Fugu rubripes	70	(Maglich et al. 2003; Bertrand et al. 2004)
	Tetraodon nigroviridis	71	(Metpally et al. 2007)
	Gasterosteus aculeatus	66	(Zhao et al. 2015)
	Oreochromis niloticus	74	(Zhao et al. 2015)
	Oryzias latipes	67	(Zhao et al. 2015)
	Danio rerio	73	(Zhao et al. 2015)
	Xenopus tropicalis	52	(Zhao et al. 2015)
	Pelodiscus sinensis	48	(Zhao et al. 2015)
	Gallus gallus	44	(Zhao et al. 2015)
	Anas platyrhynchos	42	(Zhao et al. 2015)
	Tursiops truncatus	47	(Zhao et al. 2015)
	Mus musculus	49	(Zhao et al. 2015)
	Rattus norvegicus	49	(Zhao et al. 2015)
	Homo sapiens	48	(Robinson-Rechavi et al. 2001; Zhao et al. 2015)

Furthermore, independent specific genome duplications were documented in the mammal red viscacha rat (*Tympanoctomys barrerae*) (Gallardo et al. 1999), in the ray-

finned American paddlefish (*Polyodon spathula*) (Crow et al. 2012) and in amphibian African clawed frog (*Xenopus laevis*) (Session et al. 2016). After duplication, genes underwent rapid evolutionary divergence (Holland et al. 2017), accumulating mutations that can lead to non-functionalization and consequently, gene loss, or to neo-functionalization with conservation of the ancestral function by one duplicate and acquisition of a novel function by the other duplicate. Alternatively, both gene copies can diverge, resulting in sub-functionalization either with an overlap of the ancestral function or with a differential regulation (Louis 2007). In the case of NR superfamily, events of gene duplication or WGD were clearly the driving force for their evolution and diversification, resulting in a (pre)classification into six subfamilies (Laudet et al. 1992; Owen & Zelent 2000; Escriva et al. 2004). Thus, the diversification of NRs was a consequence of two waves of gene duplication. The first wave, before the divergence of cnidarians, not associated with a whole-genome duplication, led to the existence of the six subfamilies and the different groups within each subfamily. The second wave led to the appearance of paralogs within each subfamily and occurred at the divergence of vertebrates, corresponding to WGDs (**Figure 1.5**) (Escriva et al. 1997, 2000, 2004).

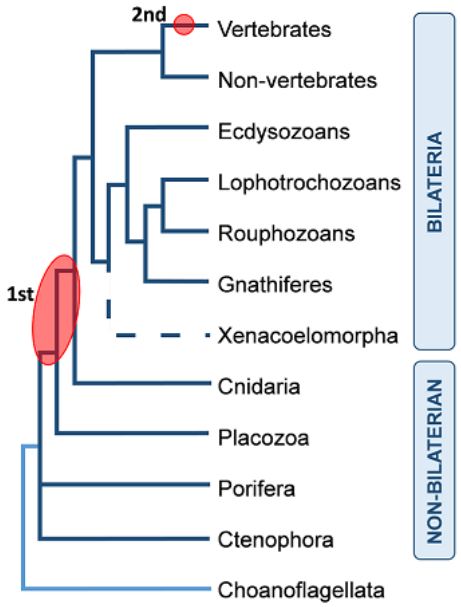


Figure 1.5. Phylogenetic tree of Metazoa and Choanoflagellata. The two waves of genomes duplication (red circles) leading to NR diversification. Phylogeny distribution based on (Laumer et al. 2015).

To attempt to reconstruct the evolutionary history of NRs, several studies proposed that all NRs evolved from a unique NR responsible to activate and/or repress transcription of genes by binding to DNA regulatory regions as a ligand-independent monomer (Escriva et al. 1997, 2000; Owen & Zelent 2000; Escriva et al. 2004). As

metazoans evolved, NRs diversified and acquired the ability to dimerise, whereas the NR modulation upon bind to a ligand was gained multiple and independent times during the NR evolution. The hypothesis of an orphan NR ancestor (AncNR) was accepted, since the phylogenetic analyses did not group NRs with similar ligands (e.g. both RAR and RXR bind to retinoids but are grouped in subfamilies I and II, respectively) and orphan NRs were widely dispersed in the tree (Escriva et al. 1997, 2000; Owen & Zelent 2000; Escriva et al. 2004). However, this conceptual proposal implies that the ligand-dependent regulation of gene transcription evolved independently and many times *de novo* which is not parsimonious (Bridgham et al. 2010). The availability of new complete sequenced genomes, including from early-branching metazoans, helped to better understand the puzzling history of NR evolution and diversification. In fact, the sequenced genome of the demosponge *Amphimedon queenslandica* (phylum Porifera) allowed the identification of two NR genes (NR1 and NR2), which are now widely used as the root of the NR tree (Bridgham et al. 2010; Holzer et al. 2017). In fact, these NRs are members of subfamily II: NR2 is an ortholog of NR2A (HNF4) family and is ubiquitously expressed, whereas NR1 is the unduplicated ortholog of all other NRs and is expressed in specific cells which contact with the external environment (Bridgham et al. 2010). Both NR1 and NR2 were demonstrate to regulate (activate or repress, respectively) transcription upon low-affinity-binding to fatty acids (Bridgham et al. 2010). Thus, it was proposed that the AncNR was able to regulate gene transcription in a ligand-dependent way, acting as a sensor for metabolic-derived compounds, such as, fatty acids (Bridgham et al. 2010; Markov & Laudet 2010). According to this scenario, all NRs evolved from a common ligand-dependent NR through duplications, species-specific gene losses and mutations which contributed for the neo-functionalization (ligand-independent regulation of transcription or high affinity for a particular ligand) of the emerging novel NR genes (Inoue et al. 2010; Markov & Laudet 2010; Gutierrez-Mazariegos et al. 2016) (**Figure 1.6**).

The hypothesis of a ligand-dependent AncNR is a parsimonious reconstruction and is in agreement with the absence of an internal circulatory system in sponges necessary for hormonal signalling (Holzer et al. 2017). However, a precise idea of the AncNR state will only be possible with the characterization of NRs from the other two early-branching metazoan phyla, the Ctenophora and Placozoa (**Table 1.3**). From the four NR found on the placozoan genome (NR2A, NR2B, NR2F and NR3B orthologs) (Srivastava et al. 2008; Baker 2008), only one was characterized, the NR2B ortholog, showing ability to bind retinoids (9-*cis* RA and All-*trans* RA) and to promote gene transcription upon bind to 9cisRA (Novotný et al. 2017; Reitzel et al. 2018). Nevertheless, the two NRs found in Ctenophora are related with NR2A receptors, but lack the DBD. This reinforces the

previous idea of NR expansion from an AnNR NR2A-like with initial radiation of NR2 family, but also resuscitates the debate about the AncNR structure evolution: (1) the loss of DBD in ctenophore NRs was lineage-specific or (2) if Ctenophora is the earliest-branching metazoan lineage, the AncNR could be a non-canonical NR and the fusion of DBD with LBD should have occurred in the ctenophore-sponge split (Reitzel et al. 2011). Thus, more data will be necessary to settle the true AncNR state.

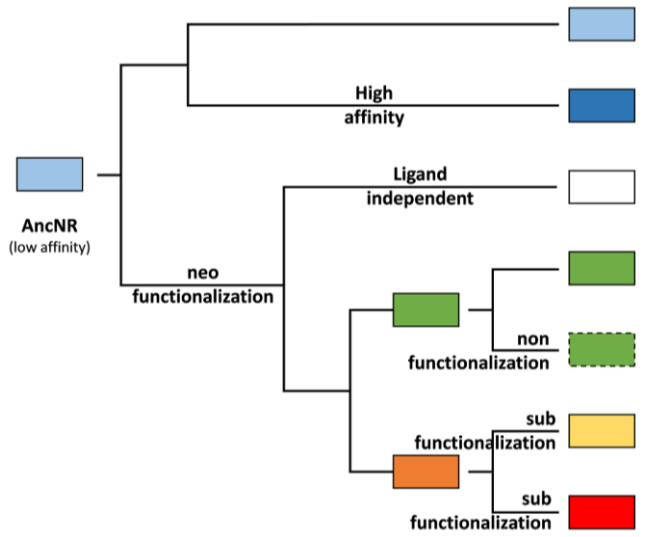


Figure 1.6. Schematic representation of nuclear receptors evolution from a ligand-dependent AncNR. The different colours represent different ligand selectivity; Light colours represent low-affinity binding NRs (sensors); Dark colours represent high-affinity binding NRs (specialized); White represents the loss of ligand-binding ability (orphan receptors); dashed line indicates NR gene loss.

Another important point is the fact that the presence of a NR ortholog in different lineages or species, does not imply the same ligand-specificity or even the same ligand dependence to activate gene transcription, neither the same signalling pathway in cross-species exploitations (Holzer et al. 2017; Katsiadaki 2019). The estrogen and RA signalling pathways represent good examples. Estrogens and other steroids, which are steroid receptors (SRs) *bona fide* ligands in vertebrates, have been identified in non-vertebrates. Nevertheless, the amphioxus ER ortholog has no capacity to bind estrogens (Paris et al. 2008). Similarly, the mollusc ER orthologs are unable to bound estrogens but display a constitutively transcription activation (Thornton & Need, E, Crews 2003; Keay et al. 2006). In this way, it was suggested that either the SR ancestor was ligand-dependent NR with loss of binding ability at least twice, or SR ancestor was an orphan NR with gain of estrogens-binding ability specifically in vertebrates. Later, the annelid ER

ortholog was identified as an estrogen-binding NR and subject of endocrine disruption by environmental contaminants and the AncSR was established as a ligand-dependent NR (Keay & Thornton 2009). A similar panorama is observed in the RA signalling: the mollusc RAR orthologs do not bind retinoids neither environmental contaminants (Urushitani et al. 2013; Gutierrez-Mazariegos et al. 2014; André et al. 2019), while the annelid RAR ortholog binds to retinoids (Handberg-Thorsager et al. 2018).

Finally, is important to understand how specificity and affinity for a particular ligand evolved. The co-evolution of NRs and their ligands is not plausible, as ligands are not encoded by genes but are molecules resulting from complex metabolic pathways mediated by many enzymes. Thus, the ligand exploitation paradigm was proposed as the selection for novel biosynthetic pathways generates a collection of intermediates and prompts new combinations of ligand/receptor pairs (**Figure 1.7**) (Thornton 2001). However, this model fails in not considering the evolution of the pathways that control the synthesis of ligands (Holzer et al. 2017).

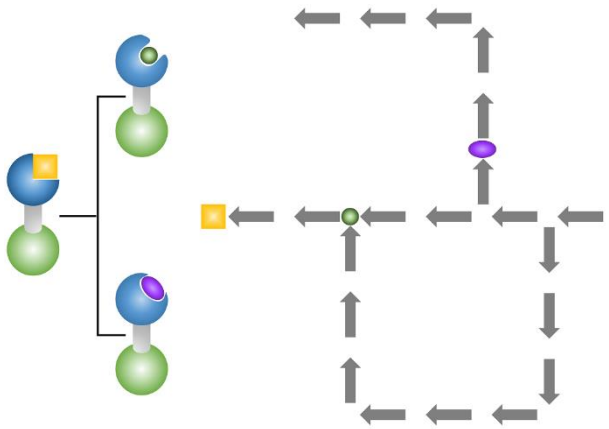


Figure 1.7. Ligand exploitation model. After NR gene duplication, the new receptors exploit intermediary products of the pathway.

1.4. Endocrine Disruption *via* Nuclear Receptors

Metazoans are multicellular organisms, in which cells need to communicate for organism survival. This communication occurs through two distinct signalling mechanisms. One mechanism involves direct contact of specialized cells, the neurons, and the transmission of electrical and chemical signals mediated by neurotransmitters across a synapse (small gap separating neurons). The other mechanism implies the endocrine system that is defined by the secretion and diffusion of chemical messengers (hormones) into the extracellular space over a large distance, affecting cells which express receptors for the released hormones. NRs are critical components of the endocrine system. In vertebrates, the endocrine system is composed of endocrine glands (cluster of specialized cells that produce hormones), hormones and the respective hormone receptors (Johnstone et al. 2014; Hartenstein 2006) (**Figure 1.8A**). Some neurons of the central and peripheral nervous system also produce hormones (neurohormones) and are called neurosecretory cells (NSCs), forming the neuroendocrine system. The neurohormones are released in the extracellular space of NSCs within vesicles and, like the hormones produced by non-neuronal endocrine cells, in the blood stream (Thorndyke & Georges 1988).

The hormones found in Metazoa are principally short peptides (e.g. prolactin, oxytocin, insulin) that are produced inside the cell, in the rough endoplasmic reticulum, and stored in vesicles and their receptors are mostly G-protein-coupled receptors which are embedded in the cell plasma membrane. Peptide hormones can bind to intracellular receptors through intracrine mechanism. Furthermore, some hormones are from lipid origin (e.g. eicosanoids, juvenile hormone), cholesterol derivatives (steroid hormones) or amino acid derivatives (e.g. melatonin, thyroid hormone). Non-peptide hormones can bind to cell membrane receptors (G-protein-coupled receptors) or can bind intracellular receptors, the case of NRs (Thorndyke & Georges 1988; Soberman & Christmas 2003; Nussey & Whitehead 2001; Belfiore & LeRoith 2018) (**Figure 1.8A**).

Neurosecretory cells and neuropeptides were identified in insects, crustaceans, nematodes, annelids, molluscs and even in cnidarians and comb jellies, indicating that the neuroendocrine system already exists in the bilaterian ancestor. Moreover, an array of peptide hormones-like are synthesized by metazoans devoid of nervous system (e.g. placozoans and sponges) and outside the Animal kingdom (e.g. protists, plants, bacteria), suggesting that cell communication *via* chemical signals is not unique to Metazoa (Hartenstein 2006; Roch & Sherwood 2014; Bonett 2016; Kleine & Rossmanith 2016; Wang et al. 2015). As the main player in the regulation of growth, differentiation,

reproduction, and behaviour, the endocrine system is critical for metazoan evolution (Bonett 2016; Duckworth 2015; Hartenstein 2006).

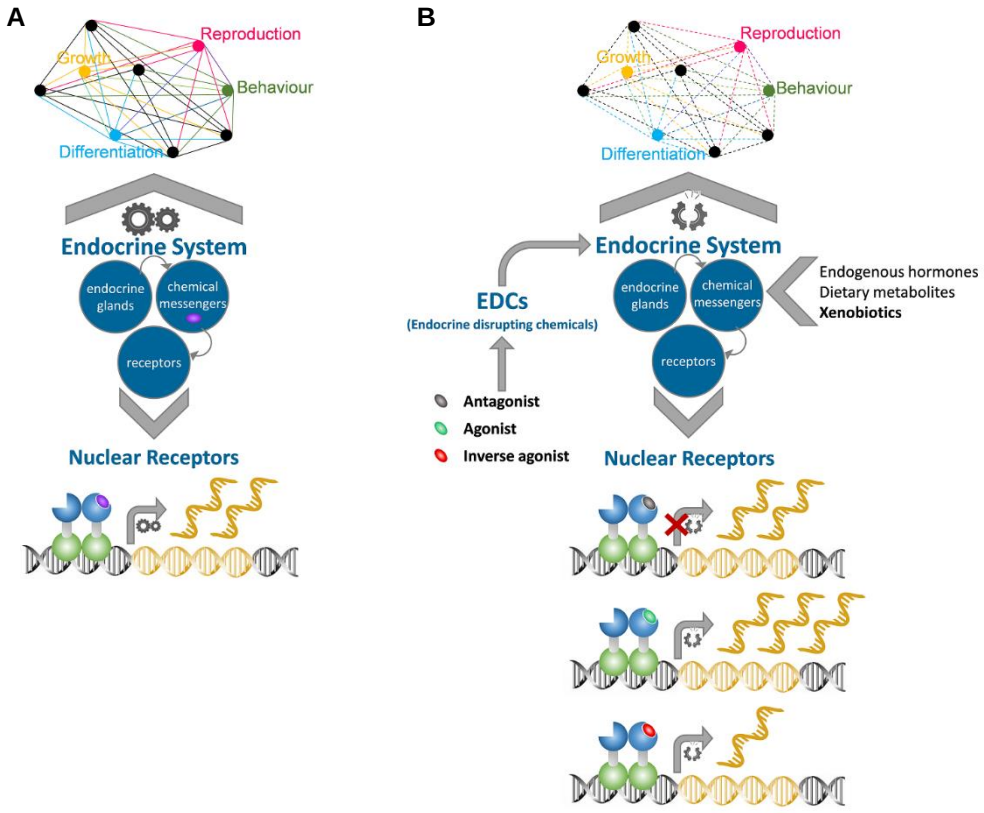


Figure 1.8. Schematic representation of biological processes regulated by the endocrine system through nuclear receptors **(A)** and endocrine disruption by endocrine-disrupting chemicals *via* nuclear receptors **(B)**.

The development of industry and technology in the last decades was accompanied by the increase of chemical synthetic manufacturing. Numerous man-made chemicals from pharmaceutical, agriculture and industrial origins became omnipresent in quotidian life as a result of their benefits (Noguera-Oviedo & Aga 2016). However, they are now persistent in the environment as contaminants, due to the continuous release, bioaccumulation and biomagnification through the food chain (**Figure 1.9**). Today, the environmental contamination by such compounds is a matter of great concern, once they have been reported to impact negatively ecosystems, affecting human health and wildlife (Grün & Blumberg 2006; Diamanti-Kandarakis et al. 2009; Bergman et al. 2013; Kabir et al. 2015; Darbre 2015; Santos et al. 2016; Capitão et al. 2017; Katsiadaki 2019). Some of these chemicals are able to disrupt the endocrine system, compromising the homeostasis of organisms. The referred compounds are the so-called endocrine-disrupting chemicals (EDCs) (Bergman et al. 2013; Diamanti-Kandarakis et al. 2009; Kabir et al. 2015; Heindel & Schug 2014) (**Figure 1.8B**). The aquatic species are more

exposed to EDC contaminants than the terrestrial ones, since all source of chemicals derived from the anthropogenic activity are dumped into aquatic systems (Sumpter 2005; Ahmed et al. 2017; Katsiadaki 2019) (**Figure 1.9**). Many studies have been conducted

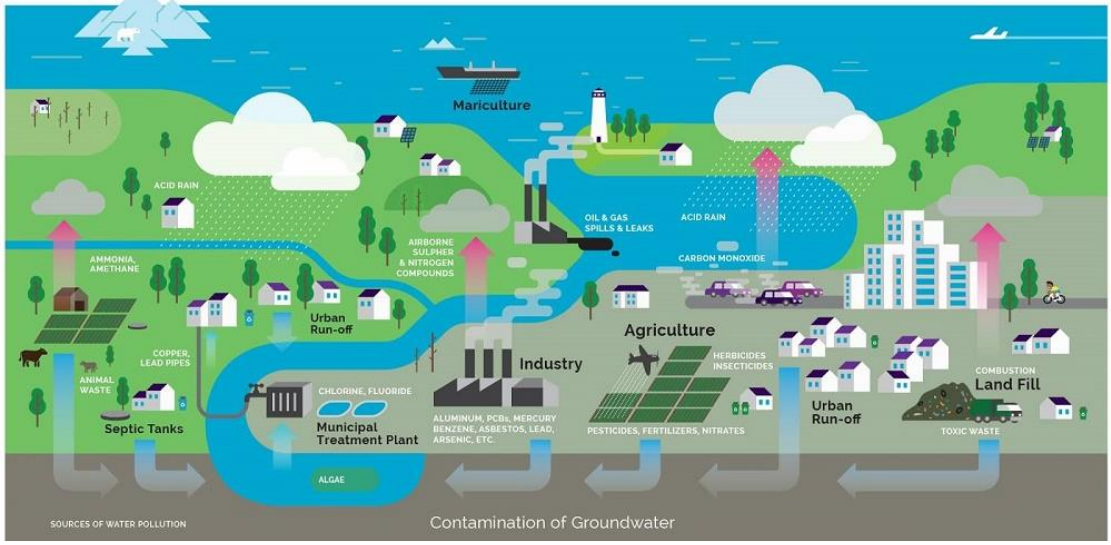


Figure 1.9. Schematic diagram of different sources of aquatic ecosystems contamination. From: https://sazworld.com/effects-of-drinking-polluted-water

to evaluate the effect of EDCs in aquatic organisms. Some well-known examples are imposex (irreversible development of male secondary sexual characteristics in females) in gastropod molluscs (Smith 1971; Bryan et al. 1986; Santos et al. 2000; Abidli et al. 2009; Pascoal et al. 2013; Bettin et al. 1996), adipogenesis in vertebrates (Grün et al. 2006; Iguchi & Katsu 2008) by organotin compounds, lipid homeostasis perturbation by obesogens (Santos et al. 2012; Lyssimachou et al. 2015; Jordão et al. 2015), alteration of fecundity and sex ratio of fish by pharmaceuticals (Soares et al. 2009; Runnalls et al. 2015; Coimbra et al. 2015) and changes in gene and protein expression, physiology and behaviour by exposure to several EDCs (Brander 2013; Sária et al. 2013; Barros et al. 2018).

The EDCs have highly variable chemical structures and properties, as they are synthesized for multiple applications: pesticides, (e.g. organochlorides, organotins), pharmaceutical products (e.g. antidepressants), plastics (e.g. bisphenol A (BPA)), synthetic hormones (e.g. ethinylestradiol) and personal care products (e.g. triclosan) (Schug et al. 2016; Kabir et al. 2015). This variety enables EDCs to interfere with the endocrine system through multiple mechanisms of action. In the non-receptor-mediated mechanism, EDCs can affect the synthesis, transport, metabolism and excretion of hormones and other signalling molecules. In the receptor-mediated mechanism, EDCs can bind to receptors, and, consequently, activate or block them, mimicking or preventing

the action of endogenous molecules (Kiyama & Wada-Kiyama 2015; Lauretta et al. 2019; Bergman et al. 2013). In this mechanism, NRs are major players, as they are prime targets of EDCs (**Figure 1.8B**). Indeed, many endocrine-disrupting mechanisms have been reported as NRs-mediated (Iguchi & Katsu 2008; le Maire et al. 2010; Castro & Santos 2014; Kiyama & Wada-Kiyama 2015). The development of imposex by exposure to organotins has been associated to the RXR signalling pathway (Nishikawa et al. 2004; Castro et al. 2007; Lima et al. 2011; Stange et al. 2012; André et al. 2017); the lipid homeostasis perturbation by obesogens has been linked to the PPARγ signalling pathways (Grün & Blumberg 2006; Riu et al. 2011; Ouadah-Boussouf & Babin 2016; Capitão et al. 2018; Barbosa et al. 2019), as well as, to the modulation of other NRs (FXR, LXR, EcR) by some EDCs (Capitão et al. 2017); and the interaction of xenoestrogens with ERs was associated with abnormal reproductive development (Sumpter & Johnson 2005; Kiyama & Wada-Kiyama 2015).

The endocrine system of marine invertebrates and early branching vertebrate lineages is still poorly described. Moreover, we cannot extrapolate that the function of a given NR and the effect of EDCs on a species or phylum is conserved in the remaining species that express this NR (e. g. unlike annelid and vertebrate RARs, molluscan RAR is not modulated by RA or the EDCs tested so far (Lemaire et al. 2005; Campo-Paysaa et al. 2008; Gutierrez-Mazariegos et al. 2014; Handberg-Thorsager et al. 2018; André et al. 2019). Thus, the screening for different environment contaminants potentially acting as EDCs *via* NRs together with an evolutionary thinking (**Figure 1.10**) are necessary to aboard the global impact of EDCs in the ecosystems during the Anthropocene (Santos et al. 2018; Katsiadaki 2019).

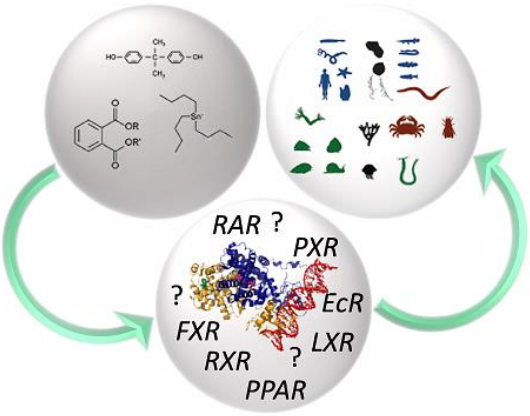


Figure 1.10. An evolutionary framework for endocrine disruption studies: interaction between endocrine-disrupting chemicals, nuclear receptors and the endocrine impact in metazoan biodiversity.

1.5. Aims of the Thesis

Previously, several studies were conducted to retrieve the NR gene repertoire in metazoans (Vogeler et al. 2014; Huang et al. 2015; Kim et al. 2017), denoting the evolutionary path of NRs in the endocrine system, and highlighting endocrine disruption events by EDCs (Grün et al. 2006; Diamanti-Kandarakis et al. 2009; Kabir et al. 2015; Darbre 2015; Vogeler et al. 2017; Capitão et al. 2017; Katsiadaki 2019). Nevertheless, evolution has often been overlooked as most studies focused on model organisms (Castro & Santos 2014) (**Figure 1.11**). Moreover, the function and binding selectivity of orthologous NR genes should not be inferred across lineages (Lemaire et al. 2005; Campo-Paysaa et al. 2008; Gutierrez-Mazariegos et al. 2014; Handberg-Thorsager et al. 2018; André et al. 2019).

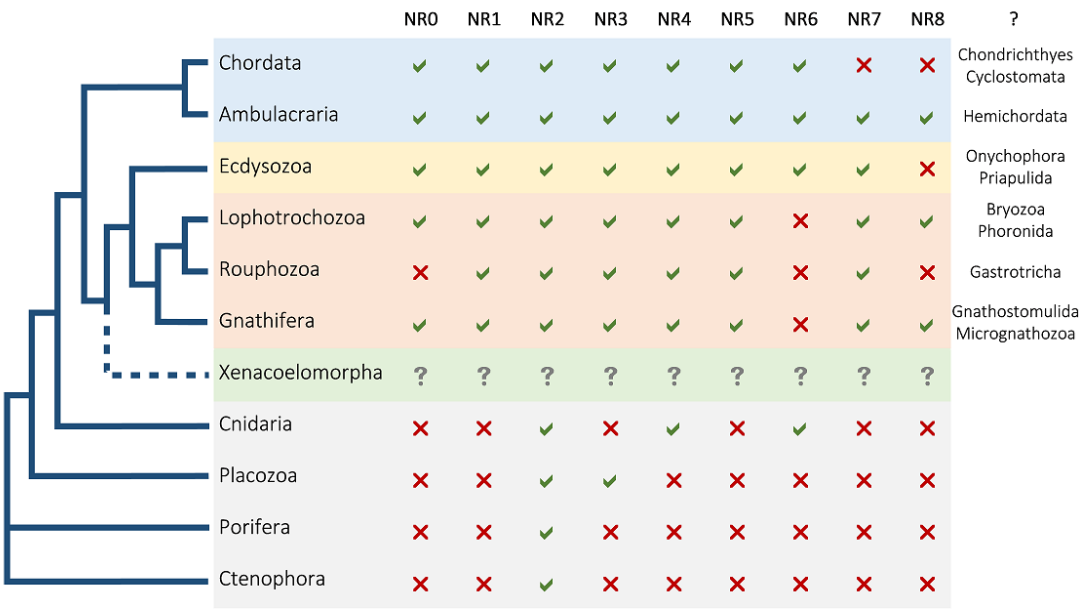


Figure 1.11. General view of nuclear receptor repertoires along the metazoan lineages. Phylogeny distribution based on (Laumer et al. 2015). Repertoire of NRs based on (Wu et al. 2007; Ibarra-Coronado et al. 2011; Weber et al. 2012; Kim et al. 2017; Santos et al. 2018). The last column identifies some lineages whom NR repertoire remains unknown. The symbols ✓, ✕, **?** mean identified, not identified, and unknown NRs, respectivelly.

Therefore, the main aim of this thesis is to contribute for the understanding of the evolution and function of NRs in the endocrine system, considering the significance of taxonomic sampling. Furthermore, this thesis intend to assess how spieces with distinct NR gene repertoires may or not be susceptible to endocrine disruption by emerging contaminants in the ecosystems during the Anthropocene epoch.

To accomplish the proposed goals and taking advantage of the revolution of next generation technology, the work focused on lineages with genomes or transcriptomes publically available and placed in key evolutionary transitions (e.g. Chondrichthyes, the first gnathostome lineage to diverge having fully undergone the vertebrate 2R WGD; Holostei, the ray-finned fishes prior to the teleost-specific 3R WGD; Priapulida, an early branching ecdysozoan; and Rotifera, model organisms for ecotoxicological studies, and on NRs selected according to the five physiological categories defined in the "*NR ring of physiology*" (reproduction and development, dietary-lipid metabolism and energy homeostasis, basal metabolic functions, steroidogenesis, and Xenobiotic metabolism) (Bookout et al. 2006). For this end, specific objectives were set:

1) Collection of NR gene repertoires in the selected species;
2) Functional characterization of the selected NRs, using transactivation assays;
3) Determination of the ability of selected disruptors to activate/antagonize these NRs;
4) Elucidation of the origin and diversification of endocrine systems in Metazoa using non-model invertebrates;
5) Assessment of the ecological impact of selected chemicals acting *via* NRs.

1.6. References

Abidli S, Lahbib Y, Menif N. 2009. Effects of TBT on the imposex development, reproduction and mortality in Hexaplex trunculus (Gastropoda: Muricidae). J Mar Biol Assoc U K. 89:139–146. doi: 10.1017/S0025315408002282.

Ahmadian M et al. 2013. PPAR γ signaling and metabolism : the good, the bad and the future. Nat Med. 19:557–566. doi: 10.1038/nm.3159.

Ahmed M et al. 2017. Progress in the biological and chemical treatment technologies for emerging contaminant removal from wastewater: A critical review. J Hazard Mater. 323:274–298. doi: 10.1016/j.jhazmat.2016.04.045.

Amores A, Catchen J, Ferrara A, Fontenot Q, Postlethwait J. 2011. Genome Evolution and Meiotic Maps by Massively Parallel DNA Sequencing: Spotted Gar, an Outgroup for the Teleost Genome Duplication. Genetics. 188:799–808. doi: 10.1534/genetics.111.127324.

André A et al. 2017. Cloning and functional characterization of a retinoid X receptor orthologue in Platynereis dumerilii: Na evolutionary and toxicological perspective. Chemosphere. 182:753−761. doi: 10.1016/j.chemosphere.2017.05.064.

André A et al. 2019. The retinoic acid receptor (RAR) in molluscs: Function, evolution and endocrine disruption insights. Aquat Toxicol. 208:80–89. doi: 10.1016/j.aquatox.2019.01.002.

Arakaki N, Miyoshi T, Noda H. 2001. Wolbachia-mediated parthenogenesis in the predatory thrips Franklinothrips vespiformis (Thysanoptera: Insecta). Proc Roy Soc L. B. 268:1011–1016. doi: 10.1098/rspb.2001.1628.

Baker M. 2008. Trichoplax, the simplest known animal, contains an estrogen-related receptor but no estrogen receptor: implications for estrogen receptor evolution. Biochem Biophys Res Commun. 375:623–627. doi: 10.1016/j.bbrc.2008.08.047.

Barbosa M et al. 2019. Linking chemical exposure to lipid homeostasis: A municipal waste water treatment plant influent is obesogenic for zebrafish larvae. Ecotoxicol Env. Saf. 182:109406. doi: 10.1016/j.ecoenv.2019.109406.

Barros S et al. 2018. Chronic environmentally relevant levels of simvastatin disrupt embryonic development, biochemical and molecular responses in zebrafish (Danio rerio). Aquat Toxicol. 201:47–57. doi: 10.1016/j.aquatox.2018.05.014.

Belfiore A, LeRoith D. 2018. *Principles of Endocrinology and Hormone Action*. Springer: Cham doi: 10.1007/978-3-319-44675-2.

Benoit G et al. 2006. International Union of Pharmacology. LXVI. Orphan nuclear receptors. Pharmacol Rev. 58:798–836. doi: 10.1124/pr.58.4.10.

Bergman Å et al. 2013. Perspectives | Editorial The Impact of Endocrine Disruption : A Consensus Statement on the State of the Science. Env Heal Perspect. 121:104–106. doi: 10.1289/ehp.1205448.

Bertrand S et al. 2004. Evolutionary genomics of nuclear receptors: from twenty-five ancestral genes to derived endocrine systems. Mol Biol Evol. 21:1923–1937. doi: 10.1093/molbev/msh200.

Bettin C, Oehlmann J, Stroben E. 1996. TBT-induced imposex in marine neogastropods is mediated by an increasing androgen level. Helgol Meeresunters. 50:299–317. doi: 10.1007/BF02367105.

Beukeboom L, Vrijenhoek R. 1998. Evolutionary genetics and ecology of sperm-dependent parthenogenesis. J Evol Biol. 11:755–782. doi: 10.1046/j.1420-9101.1998.11060755.x.

Beuschlein F et al. 2002. SF-1, DAX-1, and acd: molecular determinants of adrenocortical growth and steroidogenesis. Endocr Res. 28:597–607.

Billas I, Moras D. 2013. Allosteric controls of nuclear receptor function in the regulation of transcription. J Mol Biol. 425:2317. doi: 10.1016/j.jmb.2013.03.017.

Bodofsky S, Koitz F, Wightman B. 2017. Conserved and exapted functions of nuclear receptors in animal development. Nucl Recept. Res. 4:101305. doi: 10.11131/2017/101305.

Bondesson M, Hao R, Lin C, Williams C, Gustafsson J. 2015. Estrogen receptor signaling during vertebrate development. Biochim Biophys Acta. 1849:142–151. doi: 10.1016/j.bbagrm.2014.06.005.

Bonett R. 2016. Analyzing endocrine system conservation and evolution. Gen Comp Endocrinol. 234:3–9. doi: 10.1016/j.ygcen.2016.03.011.

Bookout A et al. 2006. Anatomical profiling of nuclear receptor expression reveals a hierarchical transcriptional network. Cell. 126:789–799. doi: 10.1016/j.cell.2006.06.049.

Brander S. 2013. Thinking outside the box: Assessing endocrine disruption in aquatic life. In: Monitoring Water Quality: Pollution Assessment, Analysis, and Remediation. Ajuha, S, editor. Elsevier pp. 103–147. doi: 10.1016/B978-0-444-59395-5.00005-4.

Brent G. 2000. Tissue-specific actions of thyroid hormone: insights from animal models. Rev Endocr Metab Disord. 1:27–33. doi: 10.1023/A:1010056202122.

Bridgham J et al. 2010. Protein evolution by molecular tinkering: diversification of the nuclear receptor superfamily from a ligand-dependent ancestor. PLoSBiol. 8:e1000497. doi: 10.1371/journal.pbio.1000497.

Bryan G, Gibbs P, Hummerstone L, Burt G. 1986. The decline of the Gastropod Nucella lapillus around South-West England: evidence for the effect of tributyltin from antifouling paints. J Mar Biol Assoc U K. 66:611–640. doi: 10.1017/S0025315400042247.

Budd G, Jensen S. 2017. The origin of the animals and a 'Savannah' hypothesis for early bilaterian evolution. Biol Rev Camb Philos Soc. 92:446–473. doi: 10.1111/brv.12239.

Bunce CM, Campbell MJ. 2010. *Nuclear Receptors: Current Concepts and Future Challenges*. 1st ed. Springer Netherlands: Dordrecht doi: 10.1007/978-90-481-3303-1.

Calisher C. 2007. Taxonomy: what's in a name? Doesn't a rose by any other name smell as Sweet? Croat Med J. 48:268–270.

Campo-Paysaa F, Marlétaz F, Laudet V, Schubert M. 2008. Retinoic acid signaling in development: tissue-specific functions and evolutionary origins. Genesis. 46:640–656. doi: 10.1002/dvg.20444.

Cannon J et al. 2016. Xenacoelomorpha is the sister group to Nephrozoa. Nature. 530:89–93. doi: 10.1038/nature16520.

Capitão A et al. 2018. Evolutionary Exploitation of Vertebrate Peroxisome Proliferator-Activated Receptor γ by Organotins. Env. Sci Technol. 52:13951−13959. doi: 10.1021/acs.est.8b04399.

Capitão A, Lyssimachou A, Castro L, Santos M. 2017. Obesogens in the aquatic environment: an evolutionary and toxicological perspective. Env. Int. 106:153−169. doi: 10.1016/j.envint.2017.06.003.

Carroll S. 2001. Chance and necessity: the evolution of morphological complexity and diversity. Nature. 409:1102–1109. doi: 10.1038/35059227.

Castro L et al. 2007. Imposex induction is mediated through the Retinoid X receptor signalling pathway in the neogastropod Nucella lapillus. Aquat Toxicol. 85:57–66. doi: 10.1016/j.aquatox.2007.07.016.

Castro L, Santos M. 2014. To bind or not to bind: the taxonomic scope of nuclear receptor mediated endocrine disruption in invertebrate phyla. Env. Sci Technol. 48:5361−5363. doi: 10.1021/es501697b.

Cavalier-Smith T. 1998. A revised six-kingdom system of life. Biol Rev Camb Philos Soc. 73:203–266. doi: 10.1111/j.1469-185X.

Cavalier-Smith T. 1981. Eukaryote kingdoms, seven or nine? Biosystems. 14:461–481. doi: 10.1016/0303-2647(81)90050-2.

Chatton E. 1925. Pansporella perplexa: amœbien à spores protégées parasite des daphnies: réflexions sur la biologie et la phylogénie des protozoaires. In: Annales des sciences naturelles: Zoologie 10e serie, VII. Masson: Paris pp. 1–84.

Chatton E. 1938. *Titres et travaux scientifiques (1906-1937)*. Impr. E. Sottano: Sette.

Clarke N, Berg J. 1998. Zinc fingers in Caenorhabditis elegans: finding families and probing pathways. Science (80-.). 282:2018–2022. doi: 10.1126/science.282.5396.2018.

Coimbra A et al. 2015. Chronic effects of clofibric acid in zebrafish (Danio rerio): a multigenerational study. Aquat Toxicol. 160:76–86. doi: 10.1016/j.aquatox.2015.01.013.

Committee NRN. 1999. A unified nomenclature system for the nuclear receptor superfamily. Cell. 97:161–163. doi: 10.1016/s0092-8674(00)80726-6.

Copeland H. 1938. The kingdoms of organisms. Q Rev Biol. 13:383–420. doi: 10.1086/394568.

Cresswell J. 2010. *The Oxford Dictionary of Word Origins*. 2nd ed. Oxford University Press: New York.

Crow K, Smith C, Cheng J, Wagner G, Amemiya C. 2012. An Independent Genome Duplication Inferred from Hox Paralogs in the American Paddlefish—A Representative Basal Ray-Finned Fish and Important Comparative Reference. Genome Biol Evol. 4:937–953. doi: 10.1093/gbe/evs067.

Cunningham J, Liu A, Bengtson S, Donoghue P. 2017. The origin of animals: Can molecular clocks and the fossil record be reconciled? Bioessays. 39:1–12. doi: 10.1002/bies.201600120.

Darbre P. 2015. *Endocrine disruption and human health. Overview of EDCs and human health which sets the bigger picture.* :Academic: New York.

Dawson M, Xia Z. 2012. The retinoid X receptors and their ligands. Biochim Biophys Acta. 1821:21–56. doi: 10.1016/j.bbalip.2011.09.014.

Degnan B, Vervoort M, Larroux C, Richards G. 2009. Early evolution of metazoan transcription factors. Curr Opin Genet Dev. 19:591–599. doi: 10.1016/j.gde.2009.09.008.

Diamanti-Kandarakis E et al. 2009. Endocrine-Disrupting Chemicals: An Endocrine Society Scientific Statement. Endocr Rev. 30:293–342. doi: 10.1210/er.2009-0002.

Dolmatov I, Afanasyev S, Boyko A. 2018. Molecular mechanisms of fission in echinoderms: Transcriptome analysis. PLoS One. 13:e0195836. doi: 10.1371/journal.pone.0195836.

Duckworth R. 2015. Neuroendocrine mechanisms underlying behavioral stability: implications for the evolutionary origin of personality. Ann N Y Acad Sci. 1360:54–74. doi: 10.1111/nyas.12797.

Dudgeon C, Coulton L, Bone R, Ovenden J, Thomas S. 2017. Switch from sexual to parthenogenetic reproduction in a zebra shark. Sci Rep. 7:40537. doi: 10.1038/srep40537.

Escriva H et al. 1997. Ligand binding was acquired during evolution of nuclear receptors. Proc Natl Acad Sci U S A. 94:6803–6808. doi: 10.1073/pnas.94.13.6803.

Escriva H, Bertrand S, Laudet V. 2004. The evolution of the nuclear receptor superfamily. Essays Biochem. 40:11–26. doi: 10.1042/bse0400011.

Escriva H, Delaunay F, Laudet V. 2000. Ligand binding and nuclear receptor evolution. Bioessays. 22:717–727. doi: 10.1002/1521-1878(200008)22:8<717::AID-BIES5>3.0.CO;2-I.

Evans R, Barish G, Wang Y. 2004. PPARs and the complex journey to obesity. Nat Med. 10:355–361. doi: 10.1038/nm1025.

Ewer J. 2005. How the ecdysozoan changed its coat. PLoS Biol. 3:e349. doi: 10.1371/journal.pbio.0030349.

Feldheim K et al. 2010. Shark virgin birth produces multiple, viable offspring. J Hered. 101:374–377. doi: 10.1093/jhered/esp129.

Fields A, Feldheim K, Poulakis G, Chapman D. 2015. Facultative parthenogenesis in a critically endangered wild vertebrate. Curr Biol. 25:446–447. doi: 10.1016/j.cub.2015.04.018.

Forrest D, Vennstrom B. 2000. Functions of thyroid hormone receptors in mice. Thyroid. 10:41–52. doi: 10.1089/thy.2000.10.41.

Forterre Y. 2013. Slow, fast and furious: understanding the physics of plant movements. J Exp Bot. 64:4745–4760. doi: 10.1093/jxb/ert230.

Gallardo M, Bickham J, Honeycutt R, Ojeda R, Köhler N. 1999. Discovery of tetraploidy in a mammal. Nature. 401:341. doi: 10.1038/43815.

Gavilán B, Sprecher S, Hartenstein V, Martinez P. 2019. The digestive system of xenacoelomorphs. Cell Tissue Res. 1–14. doi: doi.org/10.1007/s00441-019-03038-2.

Germain P, Altucci L, Bourguet W, Rochette-Egly C, Gronemeyer H. 2003. Nuclear receptor superfamily: principles of signaling. Pure Appl Chem. 75:1619–1664. doi: 10.1351/pac200375111619.

Germain P, Staels B, Dacquet C, Spedding M, Laudet V. 2006. Overview of nomenclature of nuclear receptors. Pharmacol Rev. 58:685–704. doi: 10.1124/pr.58.4.2.

Giribet G, Edgecombe G. 2017. Current Understanding of Ecdysozoa and its Internal Phylogenetic Relationships. Integr Comp Biol. 57:455–466. doi: 10.1093/icb/icx072.

Glass C. 1994. Differential recognition of target genes by nuclear receptor monomers, dimers, and heterodimers. Endocr Rev. 15:391–407. doi: 10.1210/edrv-15-3-391.

Gould SJ. 2011. A Tree Grows in Paris: Lamarck's Division of Worms and Revision of Nature. In: The Lying Stones of Marrakech. Harvard University Press: Cambridge pp. 130–134.

Grobben K. 1908. Die systematische Einteilung des Tierreichs. Verh Zool Bot Ges Wien. 58:491–511.

Gronemeyer H, Gustafsson J, Laudet V. 2004. Principles for modulation of the nuclear receptor superfamily. Nat Rev Drug Discov. 3:950–964. doi: 10.1038/nrd1551.

Groot T, Bruins E, Breeuwer J. 2003. Molecular genetic evidence for parthenogenesis in the Burmese python, Python molurus bivittatus. Heredity (Edinb). 90:130–135. doi: 10.1038/sj.hdy.6800210.

Grün F et al. 2006. Endocrine-disrupting organotin compounds are potent inducers of adipogenesis in vertebrates. Mol Endocrinol. 20:2141−2155. doi: 10.1210/me.2005-0367.

Grün F, Blumberg B. 2006. Environmental obesogens: organotins and endocrine disruption via nuclear receptor signaling. Endocrinology. 147:S50−S55. doi: 10.1210/en.2005-1129.

Gullberg H, Rudling M, Forrest D, Angelin B, Vennstrom B. 2000. Thyroid hormone receptor beta-deficient mice show complete loss of the normal cholesterol 7alpha-hydroxylase (CYP7A) response to thyroid hormone but display enhanced resistance to dietary cholesterol. Mol Endocrinol. 14:1739–1749. doi: 10.1210/mend.14.11.0548.

Gutierrez-Mazariegos J et al. 2014. A mollusk retinoic acid receptor (RAR) ortholog sheds light on the evolution of ligand binding. Endocrinology. 155:4275–4286. doi: 10.1210/en.2014-1181.

Gutierrez-Mazariegos J et al. 2016. Evolutionary diversification of retinoic acid receptor ligand binding pocket structure by molecular tinkering. R Soc Open Sci. 3:150484. doi: 10.1098/rsos.150484.

Hall W. 1970. Three probable cases of parthenogenesis in Lizards (Agamidae, Chamaeleontidae, Gekkonidae). Experientia. 26:1271–1273.

Halsall D. 1976. Zoospore chemotaxis in Australian isolates of Phytophthora species. Can J Microbiol. 22:409–422.

Handberg-Thorsager M et al. 2018. The ancestral retinoic acid receptor was a low-affinity sensor triggering neuronal differentiation. Sci Adv. 4:eaao1261. doi: 10.1126/sciadv.aao1261.

Harker J. 1997. The role of parthenogenesis in the biology of two species of mayfly (Ephemeroptera). Freshw. Biol. 37:287–297. doi: 10.1046/j.1365-2427.1997.00157.x.

Hartenstein V. 2006. The neuroendocrine system of invertebrates: a developmental and evolutionary perspective. J Endocrinol. 190:555–570. doi: 10.1677/joe.1.06964.

Heindel J, Schug T. 2014. The obesogen hypothesis: current status and implications for human health. Curr Env. Heal. Reports. 1:333−340. doi: 10.1016/j.mayocp.2013.04.005.

Hejnol A. 2010. A twist in time-the evolution of spiral cleavage in the light of animal phylogeny. Integr Comp Biol. 50:695–706. doi: 10.1093/icb/icq103.

Hejnol A et al. 2009. Assessing the root of bilaterian animals with scalable phylogenomic methods. Proc Biol Sci. 276:4261–4270. doi: 10.1098/rspb.2009.0896.

Hejnol A, Pang K. 2016. Xenacoelomorpha's significance for understanding bilaterian evolution. Curr Opin Genet Dev. 39:48–54. doi: 10.1016/j.gde.2016.05.019.

Heldin C, Lu B, Evans R, Gutkind J. 2016. Signals and Receptors. Cold Spring Harb Perspect Biol. 8:a005900. doi: 10.1101/cshperspect.a005900.

Hellal-Levy C et al. 2000. Crucial role of the H11-H12 loop in stabilizing the active conformation of the human mineralocorticoid receptor. Mol Endocrinol. 14:1210–1221. doi: 10.1210/mend.14.8.0502.

Hoekzema R, Brasier M, Dunn F, Liu A. 2017. Quantitative study of developmental biology confirms Dickinsonia as a metazoan. Proc Biol Sci. 284:pii: 20171348. doi: 10.1098/rspb.2017.1348.

Holland P. 2011. *The Animal Kingdom: A Very Short Introduction*. 1st ed. Oxford University Press: New York.

Holland P, Marlétaz F, Maeso I, Dunwell T, Paps J. 2017. New genes from old: asymmetric divergence of gene duplicates and the evolution of development. Philos Trans R Soc B Biol Sci. 372:20150480. doi: 10.1101/247361.

Holzer G, Markov G, Laudet V. 2017. Evolution of Nuclear Receptors and Ligand Signaling: Toward a Soft Key-Lock Model? Curr Top Dev Biol. 125:1–38. doi: 10.1016/bs.ctdb.2017.02.003.

Hoppe T, Kutschera U. 2015. Species-specific cell mobility of bacteria-feeding myxamoebae in plasmodial slime molds. Plant Signal Behav. 10:e1074368. doi: 10.1080/15592324.2015.

Howard-Ashby M et al. 2006. Gene families encoding transcription factors expressed in early development of Strongylocentrotus purpuratus. Dev Biol. 300:90–107. doi: 10.1016/j.ydbio.2006.08.033.

Huang W et al. 2015. Evolution of a novel nuclear receptor subfamily with emphasis on the member from the Pacific oyster Crassostrea gigas. Gene. 567:164–172. doi: 10.1016/j.gene.2015.04.082.

Hubot N, Lucas C, Piraino S. 2017. Environmental control of asexual reproduction and somatic growth of Aurelia spp. (Cnidaria, Scyphozoa) polyps from the Adriatic Sea. PLoS One. 12:e0178482. doi: 10.1371/journal.pone.0178482.

Ibarra-Coronado E et al. 2011. A helminth cestode parasite express an estrogen-binding protein resembling a classic nuclear estrogen receptor. Steroids. 76:1149–1159. doi: 10.1016/j.steroids.2011.05.003.

Iguchi T, Katsu Y. 2008. Commonality in signaling of endocrine disruption from snail to human. Bioscience. 58:1061–1067. doi: 10.1641/B581109.

Inoue J et al. 2010. Evolutionary origin and phylogeny of the modern holocephalans (Chondrichthyes: Chimaeriformes): a mitogenomic perspective. Mol Biol Evol. 27:2576–2586. doi: 10.1093/molbev/msq147.

Jaillon O et al. 2004. Genome duplication in the teleost fish Tetraodon nigroviridis reveals the early vertebrate proto-karyotype. Nature. 431:946–957. doi: 10.1038/nature03025.

Jiao Y, Lu Y, Li X. 2015. Farnesoid X receptor: a master regulator of hepatic triglyceride and glucose homeostasis. Acta Pharmacol Sin. 36:44–50. doi: 10.1038/aps.2014.116.

Johnson A, de Hoog E, Tolentino M, Nasser T, Spencer G. 2019. Pharmacological evidence for the role of RAR in axon guidance and embryonic development of a protostome species. Genesis. 30:e23301. doi: 10.1002/dvg.23301.

Johnstone C, Hendry C, Farley A, McLafferty E. 2014. Endocrine system: part 1. Nurs Stand. 28:42–49. doi: 10.7748/ns.28.38.42.e7471.

Jondelius U, Ruiz Trillo I, Baguñà J, Riutort M. 2002. The Nemertodermatida are basal bilaterians and not members of the Platyhelminthes. Zool Scr. 31:201–215. doi: 10.1046/j.1463-6409.2002.00090.x.

Jordão R et al. 2015. Obesogens beyond vertebrates: Lipid perturbation by tributyltin in the crustacean Daphnia magna. Env. Heal. Perspect. 123:813–819. doi: 10.1289/ehp.1409163.

Kabir E, Rahman M, Rahman I. 2015. A review on endocrine disruptors and their possible impacts on human health. Env. Toxicol Pharmacol. 40:241–258. doi: 10.1016/j.etap.2015.06.009.

Kalaany N, Mangelsdorf D. 2006. LXRS and FXR: the yin and yang of cholesterol and fat metabolism. Annu Rev Physiol. 68:159–191. doi: 10.1146/annurev.physiol.68.033104.152158.

Kamino K. 2008. Underwater adhesive of marine organisms as the vital link between biological science and material science. Mar Biotechnol. 10:111–121. doi: 10.1007/s10126-007-9076-3.

Katsiadaki I. 2019. Are marine invertebrates really at risk from endocrine-disrupting chemicals? Curr Opin Env. Sci Heal. 11:37–42. doi: 10.1016/j.coesh.2019.06.005.

Kaur S et al. 2015. The nuclear receptors of Biomphalaria glabrata and Lottia gigantea: implications for developing new model organisms. PLoS One. 10:e0121259. doi: 10.1371/journal.pone.0121259.

Kawamura K, Shiohara M, Kanda M, Fujiwara S. 2013. Retinoid X receptor-mediated transdifferentiation cascade in budding tunicates. Dev Biol. 384:343–355. doi: 10.1016/j.ydbio.2013.10.004.

Keay J, Bridgham J, Thornton J. 2006. The Octopus vulgaris estrogen receptor is a constitutive transcriptional activator: evolutionary and functional implications. Endocrinology. 147:3861–3869. doi: 10.1210/en.2006-0363.

Keay J, Thornton J. 2009. Hormone-activated estrogen receptors in annelid invertebrates: implications for evolution and endocrine disruption. Endocrinology. 150:1731–1738. doi: 10.1210/en.2008-1338.

Khorasanizadeh S, Rastinejad F. 2001. Nuclear-receptor interactions on DNA-response elements. Trends Biochem Sci. 26:384–390. doi: 10.1016/s0968-0004(01)01800-x.

Kim D et al. 2017. Genome-wide identification of nuclear receptor (NR) genes and the evolutionary significance of the NR1O subfamily in the monogonont rotifer Brachionus spp. Gen Comp Endocrinol. 252:219–225. doi: 10.1016/j.ygcen.2017.06.030.

King-Jones K, Thummel C. 2005. Nuclear receptors - a perspective from Drosophila. Nat Rev Genet. 6:311–323. doi: 10.1038/nrg1581.

King N et al. 2008. The genome of the choanoflagellate Monosiga brevicollis and the origin of metazoans. Nature. 451:783–788. doi: 10.1038/nature06617.

Kiyama R, Wada-Kiyama Y. 2015. Estrogenic endocrine disruptors: Molecular mechanisms of action. Env. Int. 83:11–40. doi: 10.1016/j.envint.2015.05.012.

Kleine B, Rossmanith W. 2016. Evolution of the Endocrine System. In: Hormones and the Endocrine System. Springer: Cham pp. 347–353. doi: 10.1007/978-3-319-15060-4.

Kojetin D, Burris T. 2014. REV-ERB and ROR nuclear receptors as drug targets. Nat Rev Drug Discov. 13:197–216. doi: 10.1038/nrd4100.

Kuraku S, Meyer A, Kuratani S. 2009. Timing of Genome Duplications Relative to the Origin of the Vertebrates: Did Cyclostomes Diverge before or after? Mol Biol Evol. 26:47–59. doi: 10.1093/molbev/msn222.

Lanna E. 2015. Evo-devo of non-bilaterian animals. Genet Mol Biol. 38:284–300. doi: 10.1590/S1415-475738320150005.

Laudet V, Gronemeyer H. 2002. *The nuclear receptors factsbook*. Academic Press: London:

Laudet V, Hänni C, Coll J, Catzeflis F, Stéhelin D. 1992. Evolution of the nuclear receptor gene superfamily. EMBO J. 11:1003–1013. doi: 10.1002/j.1460-2075.1992.tb05139.x.

Laumer C et al. 2015. Spiralian phylogeny informs the evolution of microscopic lineages. Curr Biol. 25:2000–2006. doi: 10.1016/j.cub.2015.06.068.

Laurencikiene J, Ryden M. 2012. Liver X receptors and fat cell metabolism. Int J Obes. 36:1494–1502. doi: 10.1038/ijo.2012.21.

Lauretta R, Sansone A, Sansone M, Romanelli F, Appetecchia M. 2019. Endocrine Disrupting Chemicals: Effects on Endocrine Glands. Front Endocrinol. 10:178. doi: 10.3389/fendo.2019.00178.

Lecroisey C, Laudet V, Schubert M. 2012. The cephalochordate amphioxus: a key to reveal the secrets of nuclear receptor evolution. Br. Funct Genomics. 11:156–166. doi: 10.1093/bfgp/els008.

Lemaire G, Balaguer P, Michel S, Rahmani R. 2005. Activation of retinoic acid receptor-dependent transcription by organochlorine pesticides. Toxico Appl Pharmacol. 202:38–49. doi: 10.1016/j.taap.2004.06.004.

Leys, SP, Degnan B. 2001. Cytological basis of photoresponsive behavior in a sponge larva. Biol Bull. 201:323–338. doi: 10.2307/1543611.

Lima D et al. 2011. Tributyltin-induced imposex in marine gastropods involves tissue-specific modulation of the retinoid X receptor. Aquat Toxicol. 101:221–227. doi: 10.1016/j.aquatox.2010.09.022.

Linnaeus C. 1758. *Systema naturae per regna tria naturae: secundum classes, ordines, genera, species, cum characteribus, differentiis, synonymis, locis*. 10th ed. Laurentius Salvius: Stockholm.

Lirman D. 2000. Fragmentation in the branching coral Acropora palmata (Lamarck): growth, survivorship, and reproduction of colonies and fragments. J Exp Mar Bio Ecol. 251:41–57.

Litoff E et al. 2014. Annotation of the Daphnia magna nuclear receptors: comparison to Daphnia pulex. Gene. 552:116–125. doi: 10.1016/j.gene.2014.09.024.

Louis E. 2007. Evolutionary genetics: making the most of redundancy. Nature. 449:673–674. doi: 10.1038/449673a.

Lyssimachou A et al. 2015. The mammalian ″obesogen″ tributyltin targets hepatic triglyceride accumulation and the transcriptional regulation of lipid metabolism in the liver and brain of zebrafish. PLoS One. 10:e0143911. doi: 10.1371/journal.pone.0143911.

Macqueen D, Johnston I. 2014. A well-constrained estimate for the timing of the salmonid whole genome duplication reveals major decoupling from species diversification. Proc Biol Sci. 281:20132881. doi: 10.1098/rspb.2013.2881.

Maglich J et al. 2003. The first completed genome sequence from a teleost fish (Fugu rubripes) adds significant diversity to the nuclear receptor superfamily. Nucleic Acids Res. 31:4051–4058. doi: 10.1093/nar/gkg444.

Mah J, Christensen-Dalsgaard K, Leys S. 2014. Choanoflagellate and choanocyte collar-flagellar systems and the assumption of homology. Evol Dev. 16:25–37. doi: 10.1111/ede.12060.

le Maire A, Bourguet W, Balaguer P. 2010. A structural view of nuclear hormone receptor: endocrine disruptor interactions. Cell Mol Life Sci. 67:1219–1237. doi: 10.1007/s00018-009-0249-2.

Maloof A et al. 2010. Possible animal-body fossils in pre-Marinoan limestones from South Australia. Nat Geosci. 3:653–659. doi: 10.1038/ngeo934.

Mangelsdorf D et al. 1995. The nuclear receptor superfamily: the second decade. Cell. 83:835–839. doi: 10.1016/0092-8674(95)90199-X.

Mangelsdorf D, Evans R. 1995. The RXR heterodimers and orphan receptors. Cell. 83:841–850. doi: 10.1016/0092-8674(95)90200-7.

Margulis L, Chapman MJ. 2009. Animalia. In: Kingdoms and Domains An Illustrated Guide to the Phyla of Life on Earth. Academic press: Massachusetts p. 341.

Mark M, Ghyselinck N, Chambon P. 2006. Function of retinoid nuclear receptors: lessons from genetic and pharmacological dissections of the retinoic acid signaling pathway during mouse embryogenesis. Annu Rev Pharmacol Toxicol. 46:451–480. doi: 10.1146/annurev.pharmtox.46.120604.141156.

Markov G, Laudet V. 2010. Origin and evolution of the ligand-binding ability of nuclear receptors. Mol Cell Endocrinol. 334:21–30. doi: 10.1016/j.mce.2010.10.017.

Marlétaz F, Peijnenburg K, Goto T, Satoh N, Rokhsar D. 2019. A New Spiralian Phylogeny Places the Enigmatic Arrow Worms among Gnathiferans. Curr Biol. 29:312–318. doi: 10.1016/j.cub.2018.11.042.

Marrif H et al. 2005. Temperature homeostasis in transgenic mice lacking thyroid hormone receptor-alpha gene products. Endocrinology. 146:2872–2884. doi: 10.1210/en.2004-1544.

di Masi A, De Marinis E, Ascenzi P, Marino M. 2009. Nuclear receptors CAR and PXR: Molecular, functional, and biomedical aspects. Mol Asp. Med. 30:297–343. doi: 10.1016/j.mam.2009.04.002.

Mazaira G et al. 2018. The Nuclear Receptor Field: A Historical Overview and Future Challenges. Nucl Recept. Res. 5:101320. doi: 10.11131/2018/101320.

de Meeûs T, Prugnolle F, Agnew P. 2007. Asexual reproduction: genetics and evolutionary aspects. Cell Mol Life Sci. 64:1355–1372. doi: 10.1007/s00018-007-6515-2.

Mehta T et al. 2013. Evidence for at least six Hox clusters in the Japanese lamprey (Lethenteron japonicum). Proc Natl Acad Sci U S A. 110:16044–16049. doi: 10.1073/pnas.1315760110.

Mello T. 2010. Nuclear Receptors in the Control of Lipid Metabolism. Curr Cardiovasc Risk Rep. 4:142–149. doi: 10.1007/s12170-010-0080-1.

Metpally R, Vigneshwar R, Sowdhamini R. 2007. Genome inventory and analysis of nuclear hormone receptors in Tetraodon nigroviridis. J Biosci. 32:43–50. doi: 10.1007/s12038-007-0005-4.

Miyachi Y, Kanao T, Okamoto T. 2005. Low dose beta-emitter source induces sexual reproduction instead of fragmentation in an earthworm, Enchytraeus japonensis. J Env. Radioact. 79:1–5. doi: 10.1016/j.jenvrad.2004.04.009.

Moore D. 1990. Diversity and unity in the nuclear hormone receptors: a terpenoid receptor superfamily. New Biol. 2:100–105.

Moroz L et al. 2014. The ctenophore genome and the evolutionary origins of neural systems. Nature. 510:109–114. doi: 10.1038/nature13400.

Mukherjee S, Mani S. 2010. Orphan Nuclear Receptors as Targets for Drug Development. Pharm. Res. 27:1439–1468. doi: 10.1007/s11095-010-0117-7.

Nielsen C. 2012. *Animal Evolution Interrelationships of the Living Phyla*. 3rd ed. Oxford University Press: New York.

Nishikawa J et al. 2004. Involvement of the retinoid X receptor in the development of imposex caused by organotins in gastropods. Env. Sci Technol. 38:6271−6276. doi: 10.1021/es049593u.

Noguera-Oviedo K, Aga D. 2016. Lessons learned from more than two decades of research on emerging contaminants in the environment. J Hazard Mater. 316:242–251. doi: 10.1016/j.jhazmat.2016.04.058.

Normark B. 2003. The evolution of alternative genetic systems in insects. Annu Rev Entomol. 48:397–423. doi: 10.1146/annurev.ento.48.091801.112703.

Novotný J et al. 2017. Trichoplax adhaerens reveals a network of nuclear receptors sensitive to 9-cis-retinoic acid at the base of metazoan evolution. PeerJ. 29:e3789. doi: 10.7717/peerj.3789.

Nussey S, Whitehead S. 2001. Principles of endocrinology. In: Endocrinology: An Integrated Approach. BIOS Scientific Publishers: Oxford.

O'Malley B. 1989. Did eucaryotic steroid receptors evolve from intracrine gene regulators? Endocrinology. 125:1119–1120. doi: 10.1210/endo-125-3-1119.

Ogawa S et al. 2005. Molecular determinants of crosstalk between nuclear receptors and toll-like receptors. Cell. 122:707–721. doi: 10.1016/j.cell.2005.06.029.

Ohno S. 2013. *Evolution by gene duplication*. Springer Science & Business Media: Berlin.

Ordóñez-Morán P, Muñoz A. 2009. Nuclear receptors: genomic and non-genomic effects converge. Cell Cycle. 8:1675–1680. doi: 10.4161/cc.8.11.8579.

Ouadah-Boussouf N, Babin P. 2016. Pharmacological evaluation of the mechanisms involved in increased adiposity in zebrafish triggered by the environmental contaminant tributyltin. Toxicol Appl Pharmacol. 294:32–42. doi: 10.1016/j.taap.2016.01.014.

Owen G, Zelent A. 2000. Origins and evolutionary diversification of the nuclear receptor superfamily. Cell Mol Life Sci. 57:809–827. doi: 10.1007/s000180050043.

Padua A, Leocorny P, Custódio M, Klautau M. 2016. Fragmentation, Fusion, and Genetic Homogeneity in a Calcareous Sponge (Porifera, Calcarea). J Exp Zool A Ecol Genet Physiol. 325:294–303. doi: 10.1002/jez.2017.

Paps J. 2018. What Makes an Animal? The Molecular Quest for the Origin of the Animal Kingdom. Integr Comp Biol. 58:654–665. doi: 10.1093/icb/icy036.

Paps J, Holland P. 2018. Reconstruction of the ancestral metazoan genome reveals an increase in genomic novelty. Nat Commun. 9:1730. doi: 10.1038/s41467-018-04136-5.

Paris M et al. 2008. An amphioxus orthologue of the estrogen receptor that does not bind estradiol: insights into estrogen receptor evolution. BMC Evol Biol. 8:219. doi: 10.1186/1471-2148-8-219.

Parker K, Schimmer B. 2002. Genes essential for early events in gonadal development. Ann Med. 34:171–178.

Pascoal S et al. 2013. Transcriptomics and in vivo tests reveal novel mechanisms underlying endocrine disruption in an ecological sentinel, Nucella lapillus. Mol Ecol. 22:1589–1608. doi: 10.1111/mec.12137.

Penvose A, Keenan J, Bray D, Ramlall V, Siggers T. 2019. Comprehensive study of nuclear receptor DNA binding provides a revised framework for understanding receptor specificity. Nat Commun. 10:2514. doi: 10.1038/s41467-019-10264-3.

Philippe H, Brinkmann H, Martinez P, Riutort M, Baguñà J. 2007. Acoel flatworms are not platyhelminthes: evidence from phylogenomics. PLoS One. 2:e717. doi: 10.1371/journal.pone.0000717.

Pisani D et al. 2015. Genomic data do not support comb jellies as the sister group to all other animals. Proc Natl Acad Sci U S A. 112:15402–15407. doi: 10.1073/pnas.1518127112.

Putnam N et al. 2008. The amphioxus genome and the evolution of the chordate karyotype. Nature. 453:1064–1071. doi: 10.1038/nature06967.

Ramachandran R, McDaniel C. 2018. Parthenogenesis in birds: a review. Reproduction. 155:R245–R257. doi: 10.1530/REP-17-0728.

Reitzel A et al. 2018. Conservation of DNA and Ligand Binding Properties of Retinoid X Receptor from the Placozoan Trichoplax adhaerens to human. J Steroid Biochem Mol Biol. 184:3–10. doi: 10.1016/j.jsbmb.2018.02.010.

Reitzel A et al. 2011. Nuclear receptors from the ctenophore Mnemiopsis leidyi lack a zinc-finger DNA-binding domain: lineage-specific loss or ancestral condition in the emergence of the nuclear receptor superfamily? Evodevo. 2:3. doi: 10.1186/2041-9139-2-3.

Reitzel A, Tarrant A. 2009. Nuclear receptor complement of the cnidarian Nematostella vectensis: phylogenetic relationships and developmental expression patterns. BMC Evol Biol. 9:230. doi: 10.1186/1471-2148-9-230.

Riu A et al. 2011. Characterization of novel ligands of ERα, Er β, and PPARγ: the case of halogenated bisphenol A and their conjugated metabolites. Toxicol Sci. 122:372–382. doi: 10.1093/toxsci/kfr132.

Robinson-Rechavi M, Carpentier A, Duffraisse M, Laudet V. 2001. How many nuclear hormone receptors are there in the human genome? Trends Genet. 17:554–556. doi: 10.1016/S0168-9525(01)02417-9.

Robinson-Rechavi M, Maina C, Gissendanner C, Laudet V, Sluder A. 2005. Explosive lineage-specific expansion of the orphan nuclear receptor HNF4 in nematodes. J Mol Evol. 60:577–586. doi: 10.1007/s00239-004-0175-8.

Roch G, Sherwood N. 2014. Glycoprotein hormones and their receptors emerged at the origin of metazoans. Genome Biol Evol. 6:1466–1479. doi: 10.1093/gbe/evu118.

Roskov Y et al. 2019. *Species 2000 & ITIS Catalogue of Life, 2019 Annual Checklist.* Species 2000, Naturalis: Leiden, Netherlands www.catalogueoflife.org/annual-checklist/2019.

Ruggiero M et al. 2015a. A higher level classification of all living organisms. PLoS One. 10:e0119248. doi: 10.1371/journal.pone.0119248.

Ruggiero M et al. 2015b. Correction: A Higher Level Classification of All Living Organisms. PLoS One. 10:e0130114. doi: 10.1371/journal.pone.0130114.

Runnalls T et al. 2015. From single chemicals to mixtures-Reproductive effects of levonorgestrel and ethinylestradiol on the fathead minnow. Aquat Toxicol. 169:152–167. doi: 10.1016/j.aquatox.2015.10.009.

Safe S et al. 2016. Nuclear Receptor 4A (NR4A) Family – Orphans No More. J Steroid Biochem Mol Biol. 157:48–60. doi: 10.1016/j.jsbmb.2015.04.016.

Santos M et al. 2016. Statins: An undesirable class of aquatic contaminants? Aquat Toxicol. 174.:1–9. doi: 10.1016/j.aquatox.2016.02.001.

Santos M, Reis-Henriques M, Castro L. 2012. Lipid homeostasis perturbation by organotins: biochemical and biological effects of organotins. In: Biochemical and Biological Effects of Organotins. Pagliarani, A, Ventrella, V, & F, T, editors. Bentham pp. 83–96. doi: 10.2174/978160805265311201010083.

Santos M, Ruivo R, Capitão A, Fonseca E, Castro L. 2018. Identifying the gaps: Resources and perspectives on the use of nuclear receptor based-assays to improve hazard assessment of emerging contaminants. J Hazard Mater. 358:508–511. doi: 10.1016/j.jhazmat.2018.04.076.

Santos M, Vieira N, Santos A. 2000. Imposex in the dogwelk Nucella lapillus (L.) along the Portuguese coast. Mar Pollut Bull. 40:643–646. doi: 10.1016/S0025-326X(00)00017-5.

Sárria M, Santos M, Castro L, Vieira N, Monteiro N. 2013. Estrogenic chemical effects are independent from the degree of sex role reversal in pipefish. J Hazard Mater. 263:746–753. doi: 10.1016/j.jhazmat.2013.10.043.

Schubert M et al. 2008. Nuclear hormone receptor signaling in amphioxus. Dev Genes Evol. 218:651–665. doi: 10.1007/s00427-008-0251-y.

Schug T et al. 2016. Minireview: Endocrine Disruptors: Past Lessons and Future Directions. Mol Endocrinol. 30:833–847. doi: 10.1210/me.2016-1096.

Session A et al. 2016. Genome evolution in the allotetraploid frog Xenopus laevis. Nature. 538:336–343. doi: 10.1038/nature19840.

Sever R, Glass C. 2013. Signaling by Nuclear Receptors. Cold Spring Harb Perspect Biol. 5:a016709. doi: 10.1101/cshperspect.a016709.

Shostak S, Kankel D. 1967. Morphogenetic movements during budding in Hydra. Dev Biol. 15:451–463. doi: 10.1016/0012-1606(67)90037-1.

Shulman A, Mangelsdorf D. 2005. Retinoid x receptor heterodimers in the metabolic syndrome. N Engl J Med. 353:604–615. doi: 10.1056/NEJMra043590.

Sluder A, Mathews S, Hough D, Yin, VP, Maina C. 1999. The nuclear receptor superfamily has undergone extensive proliferation and diversification in nematodes. Genome Res. 9:103–120. doi: 10.1101/gr.9.2.103.

Smith B. 1971. Sexuality in the American mud snail, Nassarius obsoletus. Say Proc Malacol Soc L. 39:377–378. doi: 10.1093/oxfordjournals.mollus.a065117.

Smith J et al. 2013. Sequencing of the sea lamprey (Petromyzon marinus) genome provides insights into vertebrate evolution. Nat Genet. 45:415–421. doi: 10.1038/ng.2568.

Soares J et al. 2009. Disruption of zebrafish (Danio rerio) embryonic development after full life-cycle parental exposure to low levels of ethinylestradiol. Aquat Toxicol. 95:330–338. doi: 10.1016/j.aquatox.2009.07.021.

Soberman R, Christmas P. 2003. The organization and consequences of eicosanoid signaling. J Clin Invest. 111:1107–1113. doi: 10.1172/JCI18338.

Soltis D, Soltis P. 2003. The Role of Phylogenetics in Comparative Genetics. Plant Physiol. 132:1790–1800. doi: 10.1104/pp.103.022509.

Spuway H. 1953. Spontaneous parthenogenesis in a fish. Nature. 171:437. doi: 10.1038/171437a0.

Srivastava M et al. 2008. The Trichoplax genome and the nature of placozoans. Nature. 454:955–960. doi: 10.1038/nature07191.

Stange D, Sieratowicz A, Oehlmann J. 2012. Imposex development in Nucella lapillus – evidence for the involvement of retinoid X receptor and androgen signalling pathways in vivo. Aquat Toxicol. 106–107:20–24. doi: 10.1016/j.aquatox.2011.10.010.

Stein L et al. 2003. The genome sequence of Caenorhabditis briggsae: a platform for comparative genomics. PLoS Biol. 1:E45. doi: 10.1371/journal.pbio.0000045.

Sumpter J. 2005. Endocrine disruptors in the Aquatic Enviroment: an overview. Acta Hydrochim Hydrobiol. 33:9–16. doi: 10.1002/aheh.200400555.

Sumpter J, Johnson A. 2005. Lessons from endocrine disruption and their application to other issues concerning trace organics in the aquatic environment. Env. Sci Technol. 39:4321–4332. doi: 10.1021/es048504a.

Suomalainen E, Saura A, Lokki J. 1976. Evolution of parthenogenetic insects. Evol Biol. 9:209–257.

Swafford, AJM, Oakley T. 2018. Multimodal sensorimotor system in unicellular zoospores of a fungus. J Exp Biol. 221:jeb163196. doi: 10.1242/jeb.163196.

Takeuchi T et al. 2016. Bivalve-specific gene expansion in the pearl oyster genome: implications of adaptation to a sessile lifestyle. Zool. Lett. 2:1–13. doi: 10.1186/s40851-016-0039-2.

Telford M. 2019. Evolution: Arrow Worms Find Their Place on the Tree of Life. Curr Biol. 29:R152–R154. doi: 10.1016/j.cub.2018.12.029.

Telford M, Moroz L, Halanych K. 2016. A sisterly dispute. Nature. 529:286–287. doi: 10.1038/529286a.

Temereva E. 2017. Innervation of the lophophore suggests that the phoronid Phoronis ovalis is a link between phoronids and bryozoans. Sci Rep. 7:14440. doi: 10.1038/s41598-017-14590-8.

Thomson S, Baldwin W, Wang Y, Kwon G, Leblanc G. 2009. Annotation, phylogenetics, and expression of the nuclear receptors in Daphnia pulex. BMC Genomics. 10:500. doi: 10.1186/1471-2164-10-500.

Thorndyke M, Georges D. 1988. Functional aspects of peptide neurohormones in protochordates. In: Neurohormones in Invertebrates. Thorndyke, M & Goldsworthy, G, editors. Cambridge University Press: Cambridge. doi: 10.1017/CBO9780511752230.

Thornton J. 2001. Evolution of vertebrate steroid receptors from an ancestral estrogen receptor by ligand exploitation and serial genome expansions. Proc Natl Acad Sci U S A. 98:5671–5676. doi: 10.1073/pnas.091553298.

Thornton J, Need, E, Crews D. 2003. Resurrecting the ancestral steroid receptor: ancient origin of estrogen signaling. Science (80-.). 301:1714–1717. doi: 10.1126/science.1086185.

Unsworth A, Flora G, Gibbins J. 2018. Non-genomic effects of nuclear receptors: insights from the anucleate platelet. Cardiovasc Res. 114:645–655. doi: 10.1093/cvr/cvy044.

Urushitani H et al. 2013. Cloning and characterization of the retinoic acid receptor-like protein in the rock shell, Thais clavigera. Aquat Toxicol. 142–143:403–413. doi: 10.1016/j.aquatox.2013.09.008.

Uryu O, Ameku T, Niwa R. 2015. Recent progress in understanding the role of ecdysteroids in adult insects: Germline development and circadian clock in the fruit fly Drosophila melanogaster. Zool. Lett. 1:32. doi: 10.1186/s40851-015-0031-2.

Valentine JW. 2004. Phyla Are Morphologically Based Branches of the Tree of Life. In: On the Origin of Phyla. University of Chicago Press: Chicago pp. 7–8.

Verma B. 2011. History of Plant Classification. In: Introduction to Taxonomy of Angiosperms. PHI Learning Pvt. Ltd: New Dheli p. 480.

Vogeler S, Galloway T, Isupov M, Bean T. 2017. Cloning retinoid and peroxisome proliferator-activated nuclear receptors of the Pacific oyster and in silico binding to environmental chemicals. PLoS One. 12:e0176024. doi: 10.1371/journal.pone.0176024.

Vogeler S, Galloway T, Lyons B, Bean T. 2014. The nuclear receptor gene family in the Pacific oyster, Crassostrea gigas, contains a novel subfamily group. BMC Genomics. 15:369. doi: 10.1186/1471-2164-15-369.

Volkov A, Xu K, Kolobov V. 2017. Cold plasma interactions with plants: Morphing and movements of Venus flytrap and Mimosa pudica induced by argon plasma jet. Bioelectrochemistry. 118:100–105. doi: 10.1016/j.bioelechem.2017.07.011.

Wang C, Liu Y, Li S, Han G. 2015. Insights into the Origin and Evolution of the Plant Hormone Signaling Machinery. Plant Physiol. 167:872–886. doi: 10.1104/pp.114.247403.

Weber K et al. 2012. Analysis of C. elegans NR2E nuclear receptors defines three conserved clades and ligand-independent functions. BMC Evol Biol. 12:81. doi: 10.1186/1471-2148-12-81.

Webster C. 1970. Bauhin, Jean. Dict. Sci. Biogr. Vol I. 525–527.

Weiss R et al. 1998. Thyroid hormone action on liver, heart, and energy expenditure in thyroid hormone receptor beta-deficient mice. Endocrinology. 139:4945–4952. doi: 10.1210/endo.139.12.6412.

Welch D, Meselson M. 2001. Rates of nucleotide substitution in sexual and anciently asexual rotifers. Proc Natl Acad Sci USA. 98:6720–6724. doi: 10.1073/pnas.111144598.

Whalan S, Wahab M, Sprungala S, Poole A, de Nys R. 2015. Larval settlement: the role of surface topography for sessile coral reef invertebrates. PLoS One. 10:e0117675. doi: 10.1371/journal.pone.0117675.

Whelan N et al. 2017. Ctenophore relationships and their placement as the sister group to all other animals. Nat Ecol Evol. 1:1737–1746. doi: 10.1038/s41559-017-0331-3.

Whittaker R. 1969. New concepts of kingdoms of organisms. Science (80-.). 163:150–160. doi: 10.1126/science.163.3863.150.

Whittaker R. 1959. On the broad classification of organisms. Q Rev Biol. 34:210–226.

Woese C, Fox G. 1977. Phylogenetic structure of the prokaryotic domain: The primary kingdoms. Proc Natl Acad Sci USA. 74:5088–5090. doi: 10.1073/pnas.74.11.5088.

Woese C, Kandler O, Wheelis M. 1990. Towards a natural system of organisms: proposal for the domains Archaea, Bacteria, and Eucarya. PNAS. 87:4576–4579. doi: 10.1073/pnas.87.12.4576.

Wu W, Niles E, El-Sayed N, Berriman M, LoVerde P. 2006. Schistosoma mansoni (Platyhelminthes, Trematoda) nuclear receptors: Sixteen new members and a novel subfamily. Gene. 366:303–315. doi: 10.1016/j.gene.2005.09.013.

Wu W, Niles E, Hirai H, LoVerde P. 2007. Evolution of a novel subfamily of nuclear receptors with members that each contain two DNA binding domains. BMC Evol Biol. 7:27. doi: 10.1186/1471-2148-7-27.

Yagi K et al. 2003. A genomewide survey of developmentally relevant genes in Ciona intestinalis: III. Genes for Fox, ETS, nuclear receptors and NFkappaB. Dev Genes Evol. 213:235–244. doi: 10.1007/s00427-003-0322-z.

Zelcer N, Tontonoz P. 2006. Liver X receptors as integrators of metabolic and inflammatory signaling. J Clin Invest. 116:607–614. doi: 10.1172/JCI27883.

Zhao Y, Zhang K, Giesy J, Hu J. 2015. Families of nuclear receptors in vertebrate models: characteristic and comparative toxicological perspective. Sci Rep. 5:8554. doi: 10.1038/srep08554.

CHAPTER 2 - Chondrichthyes Offer Unique Insights into the Evolution of the Nuclear Receptor Gene Repertoire in Gnathostomes

Elza Fonseca[1,2], André M. Machado[1], Nair Arrondo[3], André Gomes-dos-Santos[1,2], Ana Veríssimo[2,4], Pedro Esteves[2,4], Tereza Almeida[2,4], Gonçalo Themudo[1], Raquel Ruivo[1], Montse Pérez[3], Rute R. da Fonseca[5], Miguel M. Santos[1,2], Elsa Froufe[1], Esther Román[3], Byrappa Venkatesh[6], L. Filipe C. Castro[1,2*]

[1]CIIMAR/CIMAR - Interdisciplinary Centre of Marine and Environmental Research, U.Porto, 4450-208 Matosinho, Portugal

[2]FCUP - Faculty of Sciences, Department of Biology, U.Porto, 4169-007 Porto, Portugal

[3]AQUACOV, Instituto Español de Oceanografía, Centro Oceanográfico de Vigo, 36390 Vigo, Spain

[4]CIBIO – Research Center in Biodiversity and Genetic Resources, InBIO, Associate Laboratory, U.Porto, 4485-661 Vairão, Portugal.

[5]Center for Macroecology, Evolution and Climate, GLOBE Institute, University of Copenhagen, DK-2100 Copenhagen, Denmark.

[6]Comparative Genomics Laboratory, Institute of Molecular and Cell Biology, A*STAR (Agency for Science, Technology and Research), Biopolis Drive, 138673 Singapore, Singapore.

*Corresponding author at: CIIMAR/CIMAR, U.Porto, Avenida General Norton de Matos, S/N, 4450-208 Matosinhos, Portugal Tel.: +351 223 401 800

2. Chondrichthyes Offer Unique Insights into the Evolution of the Nuclear Receptor Gene Repertoire in Gnathostomes

2. Abstract

Nuclear receptors (NRs) are key transcription factors that originated in the ancestor of Metazoans. Most are triggered by binding to either endogenous (e.g. retinoic acid) or exogenous (e.g. xenobiotics) ligands, and their evolution and diversification is tightly linked to the function of endocrine systems. Importantly, they represent classic targets of physiological exploitation by anthropogenic chemicals. The NR gene repertoire in different lineages has been shaped by gene loss, duplication and mutation, denoting a dynamic evolutionary path. As the earliest diverging class of gnathostomes (jawed vertebrates), Chondrichthyes (cartilaginous fishes) offer an exceptional opportunity to address the early diversification of NR gene families and the evolution of the endocrine system in vertebrates. Here we provide novel insights into the NR gene content across chordates as well as an in-depth analysis of the collection of NRs in seven chondrichthyan lineages, including five Elasmobranchii (sharks and skates) and two Holocephali (chimaeras). For this purpose, we generated a low coverage draft genome assembly of the chimaera small-eyed rabbitfish (*Hydrolagus affinis*). We show that Chondrichthyes retain an archetypal NR gene repertoire, similar to that of mammals and in agreement with the two rounds of whole-genome duplication that occurred in gnathostomes ancestry. Furthermore, novel gene members of the non-canonical NR0B receptors were found in the genomes of this lineage. Our findings provide an essential view into the early diversification of NRs in gnathostomes and their endocrine system, paving the way for functional studies.

Keywords: nuclear receptors, genome, chimaera, gene loss, gene duplication

2.1. Introduction

The homeostatic coordination of biological functions such as development or reproduction depends on the action of numerous transcription factors. Among these, Nuclear Receptors (NRs) are the most abundant and peculiar in Metazoan genomes. NR monomers, homodimers or heterodimers typically triggered by ligand binding, selectively modulate transcription upon recognition of specific DNA responsive elements, in the promoter region of target genes (Laudet & Gronemeyer 2002; Germain et al. 2006). Specific ligands comprise a vast array of small lipophilic molecules from endogenous or exogenous sources, such as hormones (e.g. thyroid hormones, steroids),

morphogens (e.g. retinoic acid) or dietary components (e.g. fatty acids and vitamins) (Gronemeyer et al. 2004; Mangelsdorf & Evans 1995; Mangelsdorf et al. 1995). Thus, NR-mediated gene expression is tightly controlled by the combinatorial effect of receptors, ligands and DNA responsive elements. Canonical NRs are divided into eight subfamilies, NR1 to NR8, and are characterized by a structural architecture that includes a DNA-binding domain (DBD) and a ligand-binding domain (LBD). The exception is the subfamily NR7 with two DBD (Wu et al. 2007; Kaur et al. 2015). Non-canonical NRs are grouped in the NR0 subfamily and lack the LBD in invertebrates and the DBD in vertebrates (Laudet & Gronemeyer 2002; Germain et al. 2006; Huang et al. 2015). NRs have distinct binding specificities towards hormonal and non-hormonal compounds which is corroborated by the moderate conservation of their LBD amino acid sequences. Nevertheless, there are some ligand-orphan families of NRs (not exclusively ligand-activated) that exhibit unique structural adaptations to a redundant ligand-binding pocket (Kliewer et al. 1999).

A significant feature of NR biology is their activation or inhibition by the anthropogenic endocrine-disrupting chemicals (EDCs) (e.g. Capitão et al., 2017; Katsiadaki, 2019; le Maire et al., 2010; Lemaire et al., 2005). Some EDCs mimic or block the role of endogenous hormones and other signalling molecules, perturbing normal endogenous endocrine functions and causing physiological imbalances leading to disease (Darbre 2015). Thus, EDCs can represent a major threat to ecosystem health, and understanding their mechanisms of action provides clues to both anticipate deleterious effects in humans and wildlife, and to the evolutionary paths of NRs in Metazoans (Bertrand et al. 2004; Laudet & Gronemeyer 2002).

Currently available Sarcopterygii and Actinopterygii genomes show a clear species-dependent pattern in NR gene content: 48 NRs were identified in humans (Robinson-Rechavi et al. 2001), 49 NRs in mouse and rat, 52 NRs in the Western clawed frog, and 74 and 73 NRs were found in tilapia and zebrafish genomes, respectively (Zhao et al. 2015). The number of NR genes is also extremely variable in invertebrate chordates: 17 NRs in sea squirt (Yagi et al. 2003) and 33 NRs in amphioxus (Lecroisey et al. 2012; Schubert et al. 2008). This variation has probably been shaped by events of gene loss and duplication (tandem or whole-genome), denoting a dynamic evolutionary pattern (Bertrand et al. 2004, 2011; Bridgham et al. 2010).

Still, very few studies have addressed the diversity, function and ligand specificity of NRs in Chondrichthyes (e.g. Carroll et al., 2008; Filowitz et al., 2018; Inoue et al., 2010; Katsu et al., 2019). Chondrichthyes are the oldest extant lineage to have diverged from the last common ancestor of jawed vertebrates, occupying a critical position in the vertebrate phylogenetic tree (Benton et al. 2009). This class of gnathostomes contains

all cartilaginous fishes and is composed of two subclasses: Holocephali (chimaeras) and Elasmobranchii (sharks, skates, and rays), that emerged more than 400 million and approximately 350 million years ago, respectively (Inoue et al. 2010; Ebert et al. 2013). Chondrichthyan species colonise a wide range of ecological habitats: while elasmobranchs are distributed from tropical to polar aquatic ecosystems and tolerate highly contaminated environments (Katsu et al. 2010), holocephalans are deep-water dwellers, living around 500m and deeper, populating all oceans with the exception of the Arctic and Antarctic oceans (Didier et al. 2012). Importantly, cartilaginous fish genomes evolved at a slower rate compared to teleost fishes, sharing many similarities with mammalian genomes (Venkatesh et al. 2006, 2007; Hara et al. 2018; Marra et al. 2019). Thus, Chondrichthyes are fundamental to understand the evolution of vertebrate traits and innovations.

Here we interrogated genomes and transcriptomes of five Elasmobranchii species – whale shark (*Rhincodon typus*), brownbanded bamboo shark (*Chiloscyllium punctatum*), small-spotted catshark (*Scyliorhinus canicula*), cloudy catshark (*S. torazame*) and little skate (*Leucoraja erinacea*) – and two Holocephali species, one with a publicly available genome – elephant shark (*Callorhinchus milii*) – whereas for the small-eyed rabbitfish (*Hydrolagus affinis*) a draft genome assembly was generated for the present study.

2.2. Material and Methods

2.2.1. Sequence Retrieval and Phylogenetic Analysis

Amino acid sequences for putative NRs were retrieved via blast searches against publically available genome and transcriptomes obtained from GenBank, Skatebase (http://skatebase.org/), and Figshare (https://figshare.com/authors/Phyloinformatics_Lab_in_RIKEN_Kobe/4815111) using annotated human (*Homo sapiens*) and zebrafish (*Danio rerio*) NRs sequences as reference. Some Chondrichthyes NR sequences were retrieved via tblastn on Sequence Read Archive (SRA) and assembled with Geneious (v.7.1.7) using human NR sequences as reference. The accession numbers are listed in Supplementary Material **Table S2.1**.

The orthology of the retrieved sequences was inferred from phylogenetic analyses in the context of the NR superfamily. The sequences were aligned with the Multiple Alignment using Fast Fourier Transform (MAFFT) software (v.7) (Katoh & Toh 2010) using the FFT-NS-2 model. A first sequence alignment combined all the retrieved NR amino acid sequences [edited with Geneious (v.7.1.7), columns containing 90% of gaps

were removed; alignment available on Figshare]. The final alignment containing 706 sequences and 894 positions was used in a Bayesian phylogenetic analysis with MrBayes (v.3.2.3) sited in the CIPRES Science Gateway (v.3.3) (Miller et al. 2015). The following parameters were used: generation number=27000000, rate matrix for aa=mixed (Jones), nruns=2, nchains=6, temp=0.15, sampling set to 27000 and burnin to 0.25. A second and more restricted Bayesian phylogenetic analysis with 334 sequences from human, zebrafish, spotted gar (*Lepisosteus oculatus*), whale shark (*Rhincodon typus*), elephant shark (*Callorhinchus milii*) and Florida and European lancelets (*Branchiostoma floridae* and *B. lanceolatum*) (total of 3971 amino acids) was performed using the following parameters: generation number= 7000000, rate matrix for aa=mixed (Jones), nruns=2, nchains=4, temp=0.2, sampling set to 5000 and burnin to 0.25.

2.2.2. Synteny Analysis

To further confirm the orthology inferred from phylogenetics, the genomic location of individual NRs was extracted from human (GRCh38), zebrafish (GRCz10 and GRCz11), spotted gar (LepOcu1), whale shark (ASM164234v2) and elephant shark (Callorhinchus_milii-6.1.3) genomes. The NR-neighbouring genes were collected from GenBank using the human loci as reference to assemble the synteny maps of the remaining species. The orthology of NRb, one of the three novel cephalochordate NRs reported in previous studies (NRa, NRb and NRc) (Schubert et al. 2008; Lecroisey et al. 2012), could not be confirmed with our phylogenetic approach. Thus, the genomic location of this NR was retrieved in both Belcher's (*Branchiostoma belcheri*, Haploidv18h27) and Florida (Version 2 GCA_000003815.1) amphioxus genomes and neighbouring genes were collected and mapped in acorn worm (*Saccoglossus kowalevskii*, Skow_1.1) and purple sea urchin (*Strongylocentrotus purpuratus*, Spur_4.2) genomes to further investigate its orthology status.

2.2.3. Sampling, RNA and DNA Isolation and Genome Sequencing of the Small-eyed Rabbitfish

A small-eyed rabbitfish (*Hydrolagus affinis*) male specimen was collected from the EU Groundfish Survey (Fletán Negro 3L-2018) in the NW Atlantic (NAFO Regulatory Area, Div. 3L from 47.3685, -46.6540 at a depth of 1159 m; to 47.3438, -46.6638 at a depth of 1157 m) (Supplementary Material **Table S2.2**). A small piece of liver and gonad tissue was collected immediately after the capture of the specimen, preserved in RNAlater and stored at -80 ºC, for posterior RNA extraction. Additionally, a portion of muscle was also collected and preserved in ethanol (99%) and stored at -20ºC for DNA

extraction. RNA extractions were performed with illustra RNAspin Mini RNA Isolation Kit (GE Healthcare, UK). The extracted RNA was treated with DNase I on-column. Genomic DNA extraction was done with the DNeasy Blood & Tissue Kit (Qiagen, Hilden, Germany), following the manufacturer's instructions. The integrity of RNA and DNA was assessed on a 1% agarose TAE gel, stained with GelRed™ nucleic acid stain (Biotium, Hayward, CA, USA). Finally, RNA was quantified using a microplate spectrophotometer and a Take3™ Micro-Volume Plate (BioTeK, Winooski, VT, USA) and saved for later RNA-seq analysis. The DNA sample was quantified using a Qubit Fluorometer (Invitrogen, Carlsbad, CA, USA). To sequence the paired-end (PE) short reads we used the Illumina HiSeq X Ten platform (Macrogen, Seoul, South Korea) and generated a total of 76.2 Gbp of raw data.

Prior to genome sequencing, the taxonomic status of the specimen was confirmed by amplification and Sanger sequencing of the mtDNA *COI* gene. A fragment of ~ 599 bp of *CO1* gene was amplified by PCR using the universal primer pair FishF2 and FishR2 (Ward et al. 2005). The amplified DNA template was purified and sequenced (Macrogen, Seoul, South Korea), using the same primers. The novel *COI* gene sequence was deposited in GenBank (Supplementary Material **Table S2.2**).

2.2.4. Cleaning of Raw Dataset, de novo Assembly and Assessment of Small-eyed Rabbitfish Genome

The small-eyed rabbitfish raw sequencing reads were inspected using the FastQC software (v.0.11.8) (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/) (Andrews et al. 2015). Next, the Trimmomatic software (v.0.38) (Bolger et al. 2014) was used to assess quality and trim the raw dataset under the following parameters (LEADING:5 TRAILING:5 SLIDINGWINDOW:4:20 MINLEN:50). In addition, the Lighter tool (v.1.1.1) (Song et al. 2014), with parameters (-k 17 1200000000 0.1), was applied to correct sequencing errors. The previous steps were crucial to increase the accuracy and speed, while decreasing the memory requirements of downstream analyses. Subsequently, the clean raw reads were used to determine general genome features (e.g. genome size estimation, heterozygosity and homozygosity rates, among other), using several software such as GenomeScope2 (k-mers; 19, 21, 23, 25, 27, 29, 31) (v.2.0) (Ranallo-Benavidez et al. 2019), KmerGenie (v.1.7048) (Chikhi & Medvedev 2014), and the Preqc module of SGA software (v.0.10.15) (Simpson 2014) under the default parameters. The KmerGenie and Preqc module as well the Jellyfish software (v.2.2.10) (Marçais & Kingsford 2011) (1st step of GenomeScope2 to build k-mer frequency distributions) were downloaded and run locally in our cluster, while the second

step of GenomeScope2 was calculated in an external webserver (http://qb.cshl.edu/genomescope/genomescope2.0/) (Ranallo-Benavidez et al. 2019).

To perform the *de novo* genome assembly two independent methods were used: the W2RAP pipeline (v.0.1) (Clavijo et al. 2017) and the SOAPdenovo2 software (v.2.04) (Luo et al. 2012). The W2RAP pipeline was run according to the developer's instructions (https://github.com/bioinfologics/w2rap) (Clavijo et al. 2017). Initially, the hist module of KAT software (v.2.4.1) (Mapleson et al. 2017) was used, to determine the k-mer coverage and the best frequency of cut-off (the first down peak of the histogram) to be applied in W2RAP pipeline. Next, the W2RAP was run several times to produce different assemblies, using the general parameters (-t 30 -m 500 --min_freq 5 -d 32 --dump_all 1) and different values of kmer length (-k: 160, 168, 172, 180 and 200). After the production of the first draft genome assemblies, the QUAST (v.5.0.2) (Gurevich et al. 2013) and the Benchmarking Universal Single-Copy Orthologs (BUSCO) (v.3.0.2) (Simão et al. 2015) (using the lineage-specific libraries of Eukaryota, Metazoa and Vertebrata) tools were used, to choose the "best" W2RAP initial assembly. This choice was based on the following criteria: first, the percentage of complete and fragmented BUSCOS found; secondly, the value of N50 and N75; and thirdly, the largest sequence produced. In the end of this process, and although the scaffolding at paired-end scale are included on the W2RAP pipeline (step 8), a re-scaffolding step was performed with SOAPdenovo2 software to improve the contiguity of the "best" initial w2rap assembly. Thus, the SOAPdenovo-fusion module of SOAPdenovo2 was applied to prepare the draft assembly dataset; the mapping and scaffolding modules were run to finalize the re-scaffolding, with different k values (41, 51, 61, 65, 71, 75, and 80). Importantly, before performing the re-scaffolding, the mean insert size (IS) of the raw dataset (392bp) was calculated, a prerequisite to run SOAPdenovo2, using the draft assembly produced by W2RAP (k-200) as reference, Bowtie2 (v.2.3.5) (Langmead & Salzberg 2012) to map the clean raw reads, and finally the CollectInsertSizeMetrics function of Picard tools (v.2.19.2) (McKenna et al. 2010) to determine the IS value. The second method, the SOAPdenovo2 assembler, was applied with the default settings, the IS mentioned above and k-mer values of 23, 41, 51, 61, 65, 71, 75 and 81. In the end of the assembly process, all genome versions were re-assessed with QUAST and BUSCO tools, being the final draft genome assembly selected with the same criteria previously used to determine the best initial W2Rap assembly. Finally, the comp module of KAT was used to quantify the presence or absence of the raw reads k-mers content in the final draft genome assembly. The draft genome assembly, as well as the clean raw reads, were deposited in GenBank (Supplementary Material **Table S2.2**).

2.2.5. NRs in the Genome of the Small-eyed Rabbitfish and Phylogenetic Analysis

To investigate the absence of certain NR in the genome of the elephant shark, we performed blast researches in the small-eyed rabbitfish genome. The retrieved amino acid sequences were aligned with MAFFT software (v.7) (https://mafft.cbrc.jp/alignment/server/) (Katoh & Toh 2010) using the L-INS-i model and the orthologies were assessed by maximum likelihood under the LG substitution model for amino-acid sequences using the PhyML (v.3.0) (Guindon et al. 2010) plugin (v.2.2.4) in Geneious Prime (v.2019.2.3) (https://www.geneious.com/). Branch support was calculated by bootstrapping using 500 replications; the proportion of invariable sites and the gamma distribution parameters were estimated.

2.3. Results and Discussion

2.3.1. Overview of NRs in Chordates

The results from previous studies are summarized in **Tables 2.1** and **2.2**. Briefly, Zhao and collaborators (Zhao et al. 2015) investigated the NR gene repertoire in 12 vertebrate species representing different lineages: mammals (human; mouse, *Mus musculus*; rat, *Rattus norvegicus*; dolphin, *Tursiops truncatus*), birds (chicken, *Gallus gallus*; wild duck, *Anas platyrhynchos*), a reptile (Chinese softshell turtle, *Pelodiscus sinensis*), an amphibian (Western clawed frog, *Xenopus tropicalis*) and teleost fishes (zebrafish; Japanese medaka, *Oryzias latipes*; Nile tilapia, *Oreochromis niloticus*; stickleback, *Gasterosteus aculeatus*) (Tables 1 and 2). Additionally, NRs had been identified in pufferfishes (fugu, *Takifugu rubripes*; tetraodon, *Tetraodon nigroviridis*) (Metpally et al. 2007; Maglich et al. 2003; Bertrand et al. 2004), and in invertebrate chordates: Tunicata (sea squirt, *Ciona intestinalis*) (Yagi et al. 2003) and Cephalochordata (Florida amphioxus, *Branchiostoma floridae*) (Lecroisey et al. 2012; Schubert et al. 2008). Together, this string of studies indicates that the NR gene superfamily composition is rather variable among chordate lineages (e.g. Bertrand et al., 2004; Lecroisey et al., 2012). Yet, a detailed view of the NR repertoire in cartilaginous fishes is still missing. As the earliest diverging class of gnathostomes (jawed vertebrates), together with their slowly evolving genome (Venkatesh et al. 2007), Chondrichthyes offer an exceptional opportunity to address the early diversification of NR gene families and the evolution of the endocrine system (e.g. Fonseca et al., 2017). In this study, we interrogated genomes and transcriptomes of five Elasmobranchii

Table 2.1. Number of nuclear receptor genes identified in several species by previous studies.

	Mammals				Birds		Reptiles	Amphibians	Fishes	Urochordates	Cephalochordates
	Homo sapiens[a, b]	*Mus musculus*[b]	*Rattus norvegicus*[b]	*Tursiops truncatus*[b]	*Gallus gallus*[b]	*Anas platyrhynchos*[b]	*Pelodiscus sinensis*[b]	*Xenopus tropicalis*[b]		*Ciona intestinalis*[c]	*Banchiostoma floridae*[d, e]
NR0	2	2	2	2	2	2	2	2	See Table 2	0	1
NR1	19	20	20	19	17	16	18	20		9	15
NR2	12	12	12	11	11	10	13	14		4	7
NR3	9	9	9	9	8	8	9	10		1	3
NR4	3	3	3	3	3	3	3	3		1	1
NR5	2	2	2	2	2	2	2	2		1	2
NR6	1	1	1	1	1	1	1	1		1	1
Add. NR	0	0	0	0	0	0	0	0		0	3
TOTAL	**48**	**49**	**49**	**47**	**44**	**42**	**48**	**52**	**66-74**	**17**	**33**

[a]Robinson-Rechavi et al. 2001; [b]Zhao et al. 2015; [c]Yagi et al. 2003; [d]Lecroisey et al. 2012; [e]Schubert et al. 2008

Table 2.2. Number of nuclear receptor genes identified in fishes by previous studies.

	Teleosts						Holosteans	Chondrichthyes	
	Danio rerio[a]	*Oryzias latipes*[a]	*Oreochromis niloticus*[a]	*Gasterosteus aculeatus*[a]	*Fugu rubripes*[b,c]	*Tetraodon nigroviridis*[d]	*Lepisosteus oculatus*	Elasmobranchii	Holocephali
NR0	3	2	3	2	4	4			
NR1	29	25	30	24	27	27			
NR2	19	16	18	16	16	17			
NR3	12	14	14	15	15	15			
NR4	4	5	5	4	4	4			
NR5	4	4	3	4	3	3			
NR6	2	1	1	1	1	1			
Add. NR	0	0	0	0	0	0			
TOTAL	**73**	**67**	**74**	**66**	**70**	**71**	**?**	**?**	**?**

[a]Zhao et al. 2015; [b]Maglich et al. 2003; [c]Bertrand et al.2004; [d]Metpally et al. 2007

species: whale shark, brownbanded bamboo shark (*Chiloscyllium punctatum*), small-spotted catshark (*Scyliorhinus canicula*), cloudy catshark (*S. torazame*) and little skate (*Leucoraja erinacea*); and one Holocephali species: elephant shark. Given the paucity of genomic data from Holocephali, we additionally generated a draft genome assembly from the small-eyed rabbitfish (see below). A thorough examination is provided in the following sections of this manuscript. Moreover, we searched for NRs in the reptile green anole (*Anolis carolinensis*) and in the non-teleost ray-finned fish spotted gar. Our research retrieved 52 NRs in the spotted gar, contrasting with the 73 NRs previously found in zebrafish (**Figures 2.1** and **2.2**). As expected, we did not find evidence of 3R genome duplicates characteristic of teleost fishes, since holosteans, such as spotted gar, are pre-3R actinopterygians (Braasch et al. 2016). In the green anole, we found a NR repertoire very similar to that of the Chinese softshell turtle (Zhao et al. 2015), but with some previously unreported novelties (**Figure 2.2**; Supplementary Material **Figure S2.1**). First, regarding the subfamily NR1, we identified single gene copies of *NR1B3*, *NR1D4* and *NR1H2*, not described in turtle. While two gene copies of *NR1F3* were previously identified in turtle, similarly to birds (Zhao et al. 2015), our analyses demonstrate that in anole and frog only one copy of *NR1F3* exists, while the remaining sequences are actually *NR1F2-like* copies (Supplementary Material **Figure S2.1**). We

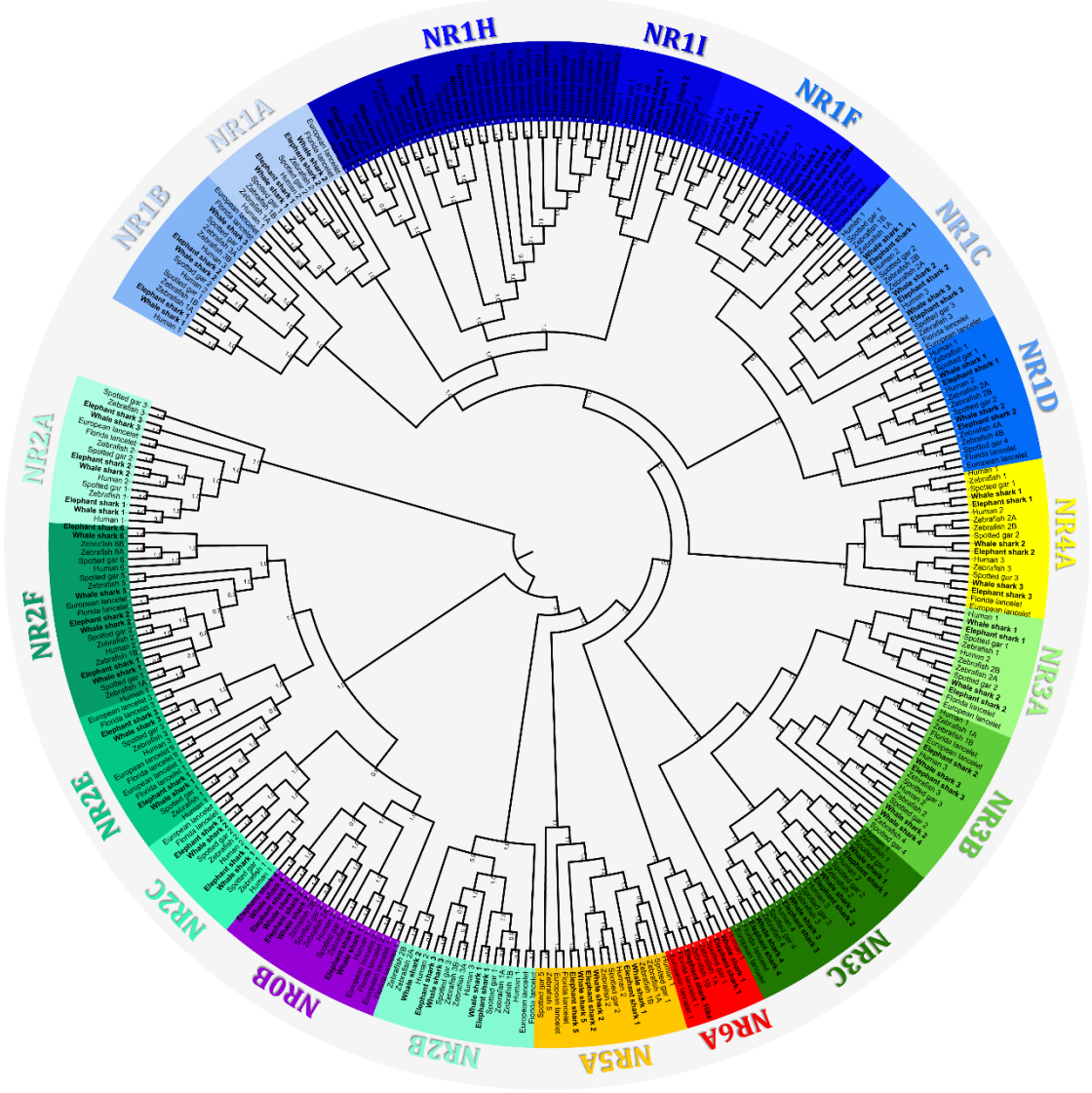


Figure 2.1. Bayesian phylogenetic analysis of NR amino acid sequences of human, zebrafish, spotted gar, whale shark, elephant shark and Florida and European lancelets; numbers at nodes indicate posterior probabilities.

next investigated the subfamily NR2. Two *NR2F1* and one *NR2F6* genes were identified in turtle, whereas in anole, single copies of *NR2F1* and *NR2F5* were retrieved, yet no *NR2F6* was found (Supplementary Material **Figure S2.1**). Moreover, instead of the *NR2B3* identified in turtle, we found a *NR2B2* ortholog in anole (Supplementary Material **Figure S2.1**). Nonetheless, all *NR2F5*, *NR2F6*, *NR2B2* and *NR2B3* genes were found in gecko (*Gekko japonicus*) (Supplementary Material **Table S2.3**). Thus, besides lineage-specific losses, the differential absence of *NR2B2* and *NR2F5* genes in turtle and, *NR2B3* and *NR2F6* genes in anole could also be explained by genome sequencing or annotation gaps.

To further broaden the revision of NRs in chordates, we re-investigated the NR gene collection in chicken, sea squirt and amphioxus (*B. floridae* and *B. lanceolatum*) (Putnam et al. 2008; Marlétaz et al. 2018). We identified four new NRs in chicken

(*NR1B3*, *NR1D1*, *NR2F5* and *NR4A1*) and a second *NR0B* ortholog in amphioxus (**Figures 2.1** and **2.2**; Supplementary Material **Figure S2.1**). No *NR2B2* gene ortholog was retrieved in chicken and duck genomes by previous and present studies (Zhao et al. 2015); however, we found a copy of this gene in other birds (Okarito kiwi, Bengalese finch, bald eagle and golden eagle) (Supplementary Material **Table S2.3**).

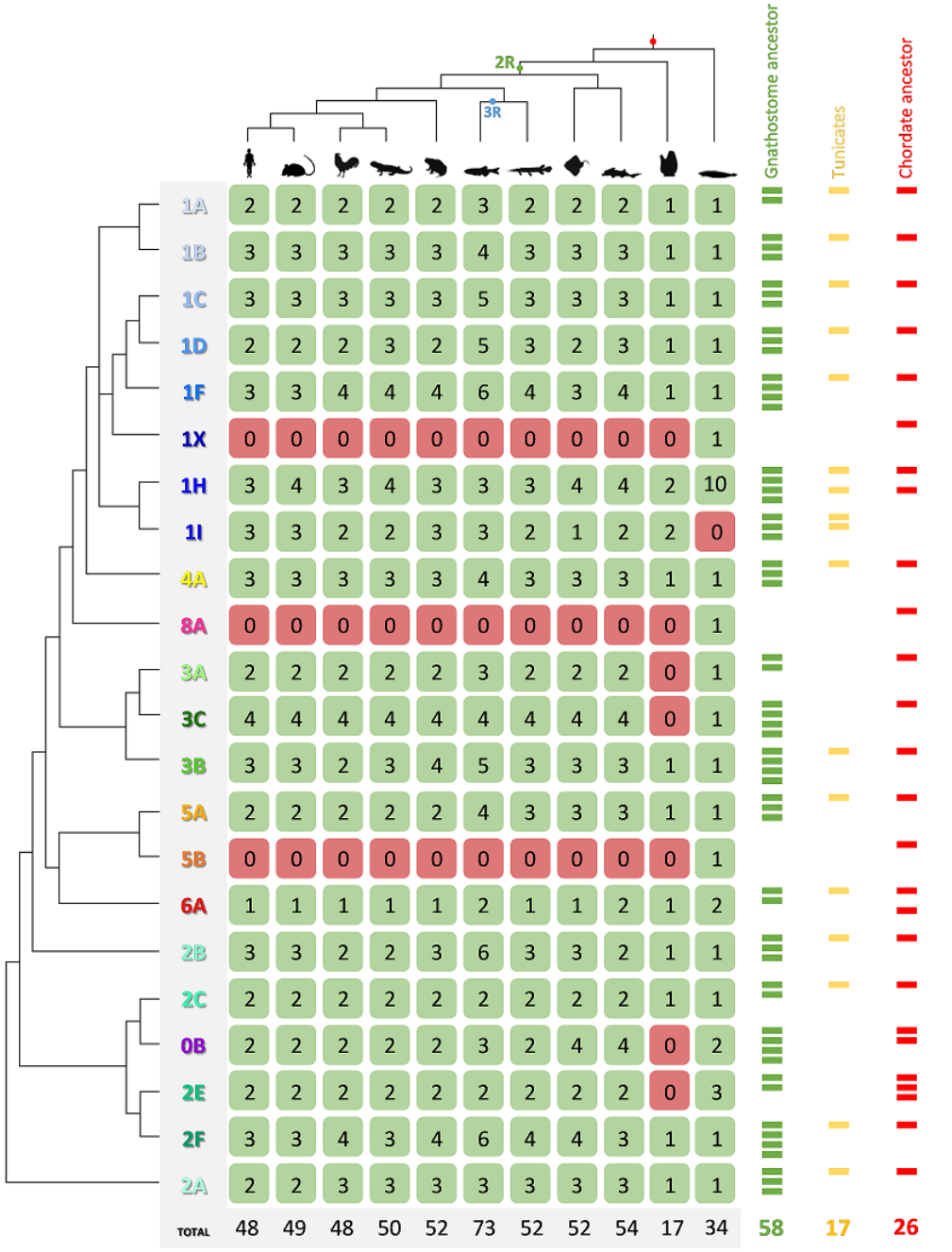


Figure 2.2. Phylogenetic representation of NRs distribution across Chordata. The numbered green boxes refer to the number of NR orthologs, while red boxes indicate the absence of NR orthologs in the genome of the represented species. On the right, green, yellow and red boxes represent the estimated number of NR copies in gnathostome ancestor, tunicates and chordate ancestor, respectively.

Considering the NR gene collection in cephalochordates, previous investigations (Schubert et al. 2008; Lecroisey et al. 2012) reported three NRs (*NRa*, *NRb* and *NRc*), with uncertain orthology. In these studies, *NRa* and *NRb* were suggested to be rapidly evolving duplicates of amphioxus NRs (Lecroisey et al. 2012; Schubert et al. 2008); BLAST analyses associated *NRa* to NR2 or NR3 subfamilies and *NRb* to NR5 or NR6 subfamilies (Schubert et al. 2008; Lecroisey et al. 2012). Alternatively, these NRs could be orthologs of an ancient NR subfamily that was secondarily lost in multiple metazoan lineages. In our phylogenetic analyses, we found out that *NRa* robustly clustered with NR8A, a new NR subfamily identified in ambulacrarians, molluscs, annelids and cnidarians (Lecroisey et al. 2012; Huang et al. 2015; Howard-Ashby et al. 2006). This new NR subfamily was characterised for the first time in the Pacific oyster (*Crassostrea gigas*), and the studies performed suggested that *NR8A1* could be involved in embryogenesis (Huang et al. 2015). The orthology of amphioxus *NRb* remains unclear (Supplementary Material **Figure S2.1**); however, we suggest that *NRb* could be categorised as a *NR6A2-like*. To corroborate this hypothesis, we analysed the sea urchin and acorn worm *NR6A2 locus* to address the neighbouring genes (synteny) and to compare with the amphioxus *NRb locus*. In the Florida amphioxus, *NRb* and the selected ambulacrarian neighbouring genes are placed on different scaffolds, but in Belcher's amphioxus we identified *NRb* and two acorn worm *NR6A2* neighbouring genes (*SMG7* and *NR5B1*) in the same scaffold (Supplementary Material **Figure S2.2**). Thus, we provide tentative support for linking *NRb* to the *NR6A2* receptor. Finally, the amphioxus NRc gene was associated to NR1 subfamily, more precisely to NR1X (2DBD) receptors (Wu et al. 2007, 2006) (Supplementary Material **Figure S2.1**). These NRs display an extra DBD and a single LBD, previously identified in various invertebrate genomes, such as molluscs (Wu et al. 2007; Vogeler et al. 2014), platyhelminths (Wu et al. 2007) and echinoderms (Howard-Ashby et al. 2006). This finding is supported by the NRc amino acid sequence analysis of Florida and European amphioxus, both displaying two DBDs and a single LBD (Supplementary Material **Figure S2.3**).

2.3.2. Detailed Examination of NRs in Chondrichthyes

Our searches for NR DBDs and LBDs sequences in the genomes and transcriptomes of whale shark, brownbanded bamboo shark, small-spotted catshark, cloudy catshark, little skate and elephant shark retrieved full or partial open reading frames (ORF) sequences of a total of 52 NRs in both Elasmobranchii and Holocephali (**Figures 2.1** and **2.2**). All the NRs were classified into seven NR subfamilies (NR0-NR6), showing orthology with human and actinopterygian species (**Figure 2.1**). For most NRs, we observed a conservative evolutionary relationship, consistent with traditional

morphological and molecular systematics (**Figure 2.1**; Supplementary Material **Figure S2.1**) (Zhao et al. 2015). To support the phylogenetic analyses and to distinguish between gene loss or absence of sequencing data, we analysed the genomic location (synteny) of NR *loci* in human, zebrafish, spotted gar and elephant shark (Supplementary Material **Figure S2.4**).

2.3.2.1. NR1 Subfamily

In this study, we found 18 and 21 NRs in Elasmobranchii and Holocephali species, respectively, belonging to the NR1 subfamily. In detail, two NR1A and three NR1C paralogs were found in all the examined Chondrichthyes; two NR1D paralogs were found in most Chondrichthyes; three NR1B paralogs were recognized in the whale shark and little skate, whereas only two were found in the elephant shark (**Figure 2.1**; Supplementary Material **Figure S2.1**). The genomic locations of the human NR1A, NR1B, NR1C and NR1D orthologs were next inspected. In humans, *NR1A1*, *NR1B1* and *NR1D1* genes are located on the same chromosome (Chr 17), and a cluster with *NR1A2*, *NR1B2* and *NR1D2* genes is located in a second chromosome (Chr 3). The same pattern was observed for zebrafish and spotted gar (Supplementary Material **Figure S2.4**). Our analysis of the elephant shark genome located the cluster *NR1A2-NR1B2-NR1D2* in the same scaffold (Sca NW_006890093.1), but *NR1A1*, *NR1B1* and *NR1D1* genes were distributed in three different scaffolds (Sca NW_006891573.1, Sca NW_006890316.1 and Sca NW_006891334.1, respectively). The NR1C paralogs were located in different chromosomes, with the exception of *NR1C3* which is also located in human Chr 3. In the elephant shark, the three NR1C paralogs were also recovered at three different scaffolds neighboured by the same human gene orthologs. Regarding *NR1B3*, we identified orthologs in whale shark and little skate but no sequence was retrieved for the elephant shark (**Figure 2.1** and Supplementary Material **Figures S2.1** and **S2.4**). Yet, orthologs of neighbouring genes of human *NR1B3* were found in three distinct scaffolds of elephant shark genome (Supplementary Material **Figure S2.4**). Furthermore, in zebrafish, spotted gar and green anole, an additional NR1D paralog (*NR1D4*s), previously identified in teleosts (Zhao et al. 2015), was also retrieved (Supplementary Material **Figure S2.1**). No *NR1D4* paralog was found in Chondrichthyes genomes. In zebrafish and spotted gar, *NR1B3* paralogs are followed by *NR1D4* paralogs (Supplementary Material **Figure S2.4**). The *NR1B3* paralog is encoded by Elasmobranchii genomes (**Figure 2.1**) and we were able to identify neighbouring genes of *NR1B3* in elephant shark genome (Supplementary Material **Figure S2.4**).

In the NR1F group, we detected four gene paralogs in the chimaera and three in sharks (**Figures 2.1** and **2.2**). Our blast searches failed to retrieve *NR1F3* paralogs in

sharks and we were unable to further confirm the putative loss by synteny due to the poor genome assembly of the whale shark currently available. Moreover, we found a *NR1F2-like* receptor, previously identified in frog, turtle, birds and teleosts (Zhao et al. 2015), and now in green anole, spotted gar and Chondrichthyes. Thus, it is reasonable to deduce that *NR1F2-like* was lost in the ancestor of mammals (Supplementary Material **Figures S2.1** and **S2.4**).

In a previous study, we addressed the evolution of NR1H and NR1I groups both characterized by independent gene loss events (Fonseca et al. 2017, 2019). Briefly, we showed that *NR1H2* and *NR1H3* are gnathostome-specific paralogs (Fonseca et al. 2017). Here, we were able to retrieve the full or near full sequence of these paralogs in Chondrichthyes with the exception of the very short sequence of the elephant shark *NR1H3* paralog (**Figure 2.1**). Similarly, *NR1H4* (this study) and *NR1H5* (this study and Cai et al. (2007)) paralogs were identified in Chondrichthyes species, indicating that NR1H paralogs are gnathostome-specific. Contrasting with the ubiquitous occurrence of *NR1I1* in all studied vertebrate genomes, *NR1I3* displays a tetrapod-specific occurrence and even more interesting, *NR1I2* was demonstrated to be probably lost in Elasmobranchii (Fonseca et al. 2019), paralleling the scenario described for the *NR1F3* gene.

2.3.2.2. NR2 Subfamily

In the NR2 subfamily a total of 14 and 12 NRs were identified in Elasmobranchii and Holocephali, respectively. Chondrichthyes display three NR2A paralogs (*NR2A1*, *NR2A2* and *NR2A3*) similarly to birds, reptiles, amphibians and teleosts (**Figures 2.1** and **2.2**). Also, their genomes encode two paralogs in both NR2C (*NR2C1* and *NR2C2*) and NR2E (*NR2E1* and *NR2E3*) groups (**Figures 2.1** and **2.2**). Concerning the NR2B group, we found three paralogs (*NR2B1*, *NR2B2* and *NR2B3*) in Elasmobranchii and two in elephant shark (*NR2B1* and *NR2B3*) (**Figures 2.1** and **2.2**). Similarly, we recovered four NR2F paralogs (*NR2F1*, *NR2F2*, *NR2F5* and *NR2F6*) but only three paralogs (*NR2F1*, *NR2F2* and *NR2F6*) in Elasmobranchii and Holocephali genomes, respectively (**Figures 2.1** and **2.2**). We next inspected the genomic regions of *NR2B2* and *NR2F5* in human, zebrafish and spotted gar to recover the *NR2B2* and *NR2F5* locations in elephant shark genome. However, the complete reconstruction of elephant shark *NR2B2* and *NR2F5 loci* was not possible due to fragmentation of this regions in the present assembly (Supplementary Material **Figure S2.4**). Thus, the presence or absence of *NR2B2* and *NR2F5* in Holocephali remains inconclusive. Yet, given that the full set of gene paralogs was retrieved in Elasmobranchii genomes, we put forward that the all

three NR2B and four NR2F genes were encoded in the gnathostome ancestor genome (**Figure 2.2**).

2.3.2.3. NR3 Subfamily

Regarding the NR3 subfamily, we recognized 9 NRs in both Elasmobranchii and Holocephali. The two NR3A paralogs (*NR3A1* and *NR3A2*) and the four NR3C paralogs (*NR3C1*, *NR3C2*, *NR3C3* and *NR3C4*) found in the studied gnathostomes were also retrieved in our analysis of chondrichthyan genomes and transcriptomes (**Figures 2.1** and **2.2**). Conversely, not all the NR3B gene paralogs were retrieved in our analysis. Chondrichthyan genomes display *NR3B2*, *NR3B3* and *NR3B4* paralogs. The later was secondarily lost in reptiles, birds and mammals (**Figure 2.1**, Supplementary Material **Figure S2.1** and (Zhao et al. 2015)). Equal to chicken and spotted gar, no *NR3B1* was found in the investigated chondrichthyan genomes. The genomic location analysis for this paralog suggests the independent loss of *NR3B1* in Chondrichthyes and holosteans. However, the elephant shark *NR3B1* neighbouring gene orthologs were found in different scaffolds and the genomic location of spotted gar *NR3B1* is poorly assembled (Supplementary Material **Figure S2.4**). Therefore, we suggest that the gnathostome ancestor displayed two NR3A paralogs, four NR3B paralogs and four NR3C paralogs (**Figure 2.2**).

2.3.2.4. NR4, NR5 and NR6 Subfamilies

The three NR4A paralogs (*NR4A1*, *NR4A2* and *NR4A3*), two NR5A paralogs (*NR5A1* and *NR5A2*) and the *NR6A1* ortholog were recovered in all examined chondrichthyan genomes (**Figures 2.1** and **2.2**). Additionally, orthologs of zebrafish *NR5A5* were also identified in the spotted gar and chondrichthyan genomes, having been secondarily lost in tetrapods (**Figure 2.1** and Supplementary Material **Figure S2.1**). However, neither phylogenetics nor the synteny analyses were sufficiently robust to support the accurate orthology of the chondrichthyan *NR5A5* (**Figure 2.1**; Supplementary Material **Figures S2.1** and **S2.4**). Importantly, we deduced the existence of a *NR6A1-like* gene ortholog located next to *NR5A2* in the elephant shark genome (**Figure 2.1**; Supplementary Material **Figures S2.1** and **S2.4**). The analyses of the human, zebrafish and spotted gar *NR5A2 locus* allowed us to corroborate the loss of *NR6A1-like* gene in these species. The Elasmobranchii *NR5A2 locus* is dispersed into several scaffolds (data not shown), impeding a stronger support to the conclusion about *NR6A1-like* gene loss in these species.

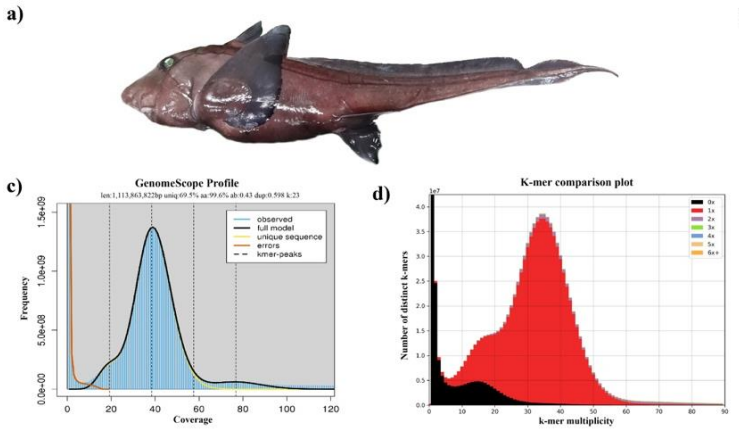
2.3.2.5. NR0 Subfamily

NR0B genes are non-canonical receptors which lack a DBD (Laudet & Gronemeyer 2002). Until the present study only two gene paralogs (*NR0B1* and *NR0B2*) had been described in vertebrates (Zhao et al. 2015). Our searches into chondrichthyan genomes and transcriptomes retrieved two novel NR0B paralogs, which we name *NR0B3* and *NR0B4* (**Figures 2.1** and **2.2**). Furthermore, by examining the synteny location of *NR0B3* and *NR0B4 loci* in human, zebrafish and spotted gar (Supplementary Material **Figure S2.4**), we were able to deduce the secondary loss of these genes in the Euteleostomi lineage. Interestingly, we were able to show that the previously named *NR0B2* ortholog in frog is a *NR0B4* gene (Supplementary Material **Figure S2.4**). In effect, the analyses of NR0B gene *loci* in this lineage (data not shown) supports that *NR0B2* and *NR0B3* paralogs were lost, while *NR0B1* and *NR0B4* paralogs were retained. Therefore, we propose that gnathostome ancestor had four NR0B paralogs (**Figure 2.2**).

2.3.3 *De novo* Genome Assembly of the Small-eyed Rabbitfish Genome

Our analysis of chondrichthyan genomes allowed the identification of 52 NR. Yet, in 5 cases we unable to confirm the presence of a given NR in the available genome of the elephant shark. To further clarify if these cases of NR gene absence in Holocephali (or Chondrichthyes) are due to lineage-specific losses or result from missing sequencing data or assembly gaps, we generated a draft genome assembly small-eyed rabbitfish (**Figure 2.3a**). The genomic sequencing produced approximately 205M of paired end raw reads that, after trimming and quality control resulted in about 184M reads to use for further analyses (**Figure 2.3b**). To estimate the genome size of small-eyed rabbitfish species were used three in silico methods: GenomeScope2, KmerGenie and the module PreQC of SGA software that allowed to find a range of values to the probable genome size of the species. This range of methods was applied due the low coverage of dataset – 36x, and in order to avoid an erroneous single genome size estimation value. Thus, we estimated a genome size between 1.08 and 1.27 Gb for the small-eyed rabbitfish (**Figure 2.3b**). In addition, the GenomeScope2 analyses still allowed to determine the rate of heterozygosity between 0.37-0.48 % (**Figure 2.3c**; Supplementary Material **Table S2.4**). These values are smaller than genome size estimations for Chimaeriformes present in Animal Genome Size Database (AGSD - http://www.genomesize.com), ranging between 1.51 (spotted ratfish, *Hydrolagus colliei*) and 2.01 Gb stipulated for chimaera species. Interestingly, the unique Chimaeriformes species with an available

genome assembly, the elephant shark (NCBI - GCA_000165045.2) (Venkatesh et al. 2014), has a genome size of 974.50 Mb, slightly smaller than the estimated for the small-eyed rabbitfish.



Raw data	
Raw sequencing reads (PE)	205663228
Clean reads	183667702
Assembly v. Haf_172k_RS_71 statistics (>500bp)	
Genome Size estimated, gb	0,941-1,268
Number of contigs [1,2,3]	229072
Contig N50 size, kb	20,16
Number of Scaffolds	213581
Scaffold N50 size, kb	22,29
Longest scaffold, kb	397,232
Total length Scaffolds, gb	1,137
Heterozygosity, %,	0,41-0,38
BUSCO Completeness (%)	
BUSCO's library	**Vertebrata**
Complete	63.5
Complete and single copy	62.7
Complete and duplicated	0.8
Fragmented	23.9
Missing	12.6
Total BUSCO found	87,4

1) Kmergenie (1,268 Mb); **2)** Genomescope k19,21,23,25,27,29,31 (941-1,065 Mb); **3)** SgaPreQC (1,228 Mb)

Figure 2.3. A draft genome assembly of *Hydrolagus affinis*. **a)** Photograph of the sequenced specimen collected in Northwest Atlantic (photograph credits to Nair Arrondo); **b)** General statistics of the genome assembly; **c)** GenomeScope2 k-mer frequency distribution profile, showing genome size estimation, heterozygosity, unique content and repeat content (i.e. len, ab, uniq, dup respectively) for 23 k-mer; **d)** Stacked histogram showing read k-mer frequency spectrum versus *H. affinis* genome assembly. Colours represent the frequency of read content present in the final assembly.

The draft genome assembly of small-eyed rabbitfish was performed using two approaches, with the W2RAP pipeline and SOAPdenovo2. In total, twenty versions of the small-eyed rabbitfish draft genome were made. Briefly, we applied the W2RAP pipeline to perform five assemblies and SOAPdenovo2 to perform eight assemblies with different k-mer values (see methods for details). Across the thirteen-initial assemblies, the N50 contig and scaffolds values varied between 0.542 – 15.857 kb and 10.200 – 19.681 kb, respectively, the largest sequence ragging between 183.560 – 405.902 kb and the rate of total BUSCOS (complete + fragmented) found in Vertebrata database varied between 56.3 and 86.1 % (Supplementary Material **Table S2.5**). Comparing the above-mentioned metrics in the initial assemblies, we selected the Haf_172k version of the draft genome assembly of the small-eyed rabbitfish to perform the re-scaffolding, with different k-mer values. This process allowed to obtain a final draft genome assembly, (Haf_172k_RS_71), with an N50 scaffold value of 22.290 kb, total length of 1.136 Gb and a percentage of 77.9, 84.0, 87.6% BUSCOS found in Eukaryote, Metazoa and Vertebrata databases, respectively (**Figure 2.3b**). Notwithstanding, we applied KAT software to verify the k-mer content of the final draft assembly in relation to the clean raw read dataset. In **Figure 2.3d**, we can observe that the major part of read k-mer content is present in the assembly (red colour of the histogram), mainly in the homozygotic peak

(±34x; k-mer multiplicity), while the k-mer content missing (black colour of the histogram) is mainly present in the heterozygotic peak (±18x; k-mer multiplicity). It is clear on this histogram some of the side effects of the low coverage approaches, such as unclear separation of the heterozygotic and homozygotic peak (essential to improve the quality and contiguity of genomic resources).

We next investigated the NR gene catalogue in this draft genome assembly, focusing on the elephant shark missing genes. In the elephant shark genome, no *NR1D4* and *NR1B3* genes, as well as, no *NR2B2*, *NR2F5* and *NR3B1* genes were found (**Figure 2.1**; Supplementary Material **Figure S2.4**). Thus, we conducted blast searches on the small-eyed rabbitfish genome, and assessed the orthology of the retrieved sequences with phylogenetic analyses (**Figure 2.4**). Our analyses retrieved the three paralogous

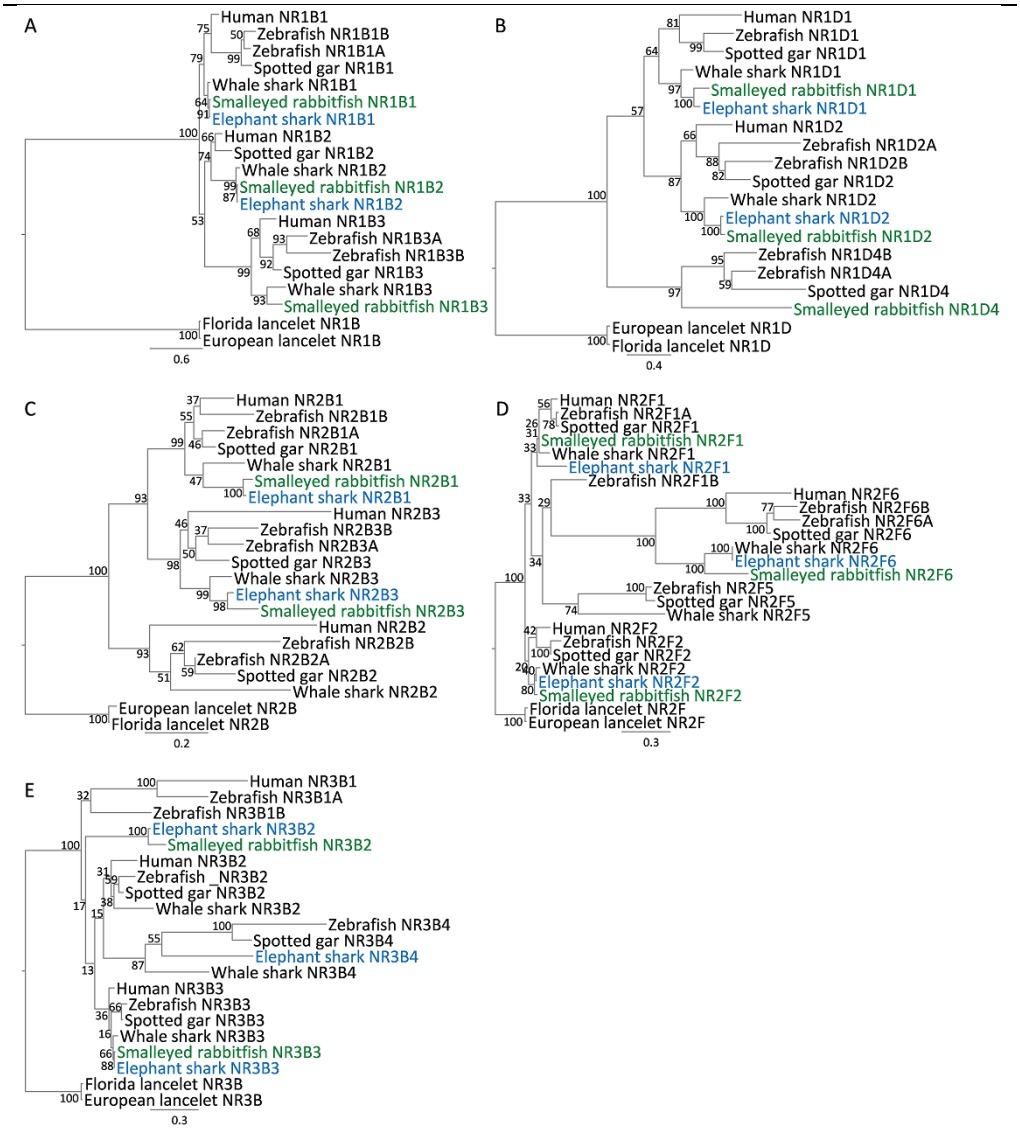


Figure 2.4. Phylogenetic trees of NRs recovered from small-eyed rabbitfish genome. **a)** NR1B; **b)** NR1D; **c)** NR2B; **d)** NR2F; **e)** NR3B.

genes of both NR1B and NR1D subfamilies (**Figure 2.4a, b**; Supplementary Material **Table S2.6**), allowing to propose that the gnathostome ancestor genome encoded three NR1B and three NR1D paralogs (**Figure 2.2**). Nevertheless, we failed to recovered *NR2B2*, *NR2F5* and *NR3B1* paralogs (**Figure 4c, d, e**; Supplementary Material **Table S2.6**). Overall, we found 54 NRs in Holocephali genomes and our results suggest that the genome of the gnathostome ancestor encoded a total of 58 NRs (**Figure 2.2**).

2.3.4. NR Gene Repertoire in Chondrichthyes: Functional Considerations

The evolution of the Metazoa coincided with the appearance of numerous novelties at the genome level, including NRs (e.g. Paps and Holland, 2018). In effect, NRs are one of the largest family of transcription factors found in extant Metazoa genomes (Mangelsdorf et al. 1995). As transcription factors, NRs regulate the expression of genes involved in biological process, such as, reproduction, development, metabolism and immunity, being prime targets for hormones of the endocrine system (Evans & Mangelsdorf 2014). Over the past decade, some studies have been conducted in cartilaginous fishes to elucidate the evolution of signalling pathways involving NRs. With the present work we took advantage of novel and existing genomic data to gain insight into the emergence and evolution of vertebrate signalling pathways involving NRs. Regarding Chondrichthyes, we were able to determine that the NR repertoire is relatively stable and comparable to the NR repertoire of mammals. Yet, repertoire stability does not necessarily imply functional conservation. In fact, NR-mediated gene transcription operates within complex interaction networks including NRs and partner NRs, ligands, and DNA-binding sites, a network which is further entangled by the transcriptional modulation of NRs, their expression patterns, as well as downstream target genes and physiological processes: establishing the so called "*NR ring of physiology*" (Bookout et al. 2006; Siddiq et al. 2017). Although not exhaustive, some studies have provided clues on the functional evolution of chondrichthyan NRs: highlighting cases of conserved responsiveness (estrogen receptors, (Katsu et al. 2010) (Filowitz et al., 2018); liver X receptors, (Fonseca et al. 2017)) or weak activity towards typical mammalian ligands (farnesoid X receptor β (Cai et al. 2007)); and also, lineage-specific specializations including interactions with elasmobranch-specific hormones (mineralocorticoid receptor and glucocorticoid receptor, (Carroll et al. 2008, 2011)) or distinct activation and tissue expression profiles (pregnane X receptor, (Fonseca et al. 2019). Still, the current knowledge on NR function and evolution in Chondrichthyes is still sparse and further

studies are required to illuminate NR-dependent networks within this group. Importantly, and given their phylogenetic placement, the assessment of NR functions in Chondrichthyes is crucial to decipher when NR functions emerged and how mechanisms of receptor promiscuity, hormone exploitation or co-evolution of metabolic signalling pathways have shaped the evolutionary history of vertebrate NRs (Thornton 2001; Carroll et al. 2008).

2.4. Conclusions

Overall, the investigation of Chondrichthyes genomes allowed the identification of the collection of NRs in Chondrichthyes offering a valuable tool to comprehend the evolution of endocrine systems and physiology of vertebrates.

Acknowledgments: This research was supported by Project No. 031342, co-financed by COMPETE 2020, Portugal 2020 and the European Union through the ERDF, and by Fundação para a Ciência e a Tecnologia (FCT) through national funds. PhD grant awarded to Elza Fonseca (SFRH/BD/100262/2014) by FCT. R.R.F. thanks the Danish National Research Foundation for its support of the Center for Macroecology, Evolution, and Climate (grant DNRF96). This research is part of the CIIMAR lead initiative *PortugalFishomics*.

2.5. References

Andrews S, Krueger F, Seconds-Pichon A, Biggins F, Wingett S. 2015. FastQC. A quality control tool for high throughput sequence data. Babraham Bioinformatics. Babraham Inst. https://www.bioinformatics.babraham.ac.uk/projects/fastqc/.

Benton M, Donoghue P, Asher R. 2009. Calibrating and constraining molecular clocks. In: The timetree of Life. Hedges, S & Kumar, S, editors. Oxford University Press pp. 35–86.

Bertrand S et al. 2004. Evolutionary genomics of nuclear receptors: from twenty-five ancestral genes to derived endocrine systems. Mol Biol Evol. 21:1923–1937. doi: 10.1093/molbev/msh200.

Bertrand S, Belgacem M, Escriva H. 2011. Nuclear hormone receptors in chordates. Mol Cell Endocrinol. 334:67–75. doi: 10.1016/j.mce.2010.06.017.

Bolger A, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics. 30:2114–2120. doi: 10.1093/bioinformatics/btu170.

Bookout A et al. 2006. Anatomical profiling of nuclear receptor expression reveals a hierarchical transcriptional network. Cell. 126:789–799. doi: 10.1016/j.cell.2006.06.049.

Braasch I et al. 2016. The spotted gar genome illuminates vertebrate evolution and facilitates human-teleost comparisons. Nat Genet. 48:427–437. doi: 10.1038/ng.3526.

Bridgham J et al. 2010. Protein evolution by molecular tinkering: diversification of the nuclear receptor superfamily from a ligand-dependent ancestor. PLoSBiol. 8:e1000497. doi: 10.1371/journal.pbio.1000497.

Cai S, Xiong L, Wray C, Ballatori N, Boyer J. 2007. The farnesoid X receptor FXRalpha/NR1H4 acquired ligand specificity for bile salts late in vertebrate evolution. Am J Physiol Regul Integr Comp Physiol. 293:1400–1409. doi: 10.1152/ajpregu.00781.2006.

Capitão A, Lyssimachou A, Castro L, Santos M. 2017. Obesogens in the aquatic environment: an evolutionary and toxicological perspective. Env. Int. 106:153−169. doi: 10.1016/j.envint.2017.06.003.

Carroll S, Bridgham J, Thornton J. 2008. Evolution of hormone signaling in elasmobranchs by exploitation of promiscuous receptors. Mol Biol Evol. 25:2643–2652. doi: 10.1093/molbev/msn204.

Carroll S, Ortlund E, Thornton J. 2011. Mechanisms for the evolution of a derived function in the ancestral glucocorticoid receptor. PLoS Genet. 7:e1002117. doi: 10.1371/journal.pgen.1002117.

Chikhi R, Medvedev P. 2014. Informed and automated k-mer size selection for genome assembly. Bioinformatics. 30:31–37. doi: 10.1093/bioinformatics/btt310.

Clavijo B et al. 2017. W2RAP: a pipeline for high quality, robust assemblies of large complex genomes from short read data. bioRxiv. doi: 10.1101/110999.

Darbre P. 2015. Endocrine disruption and human health. Overview of EDCs and human health which sets the bigger picture. :Academic: New York.

Didier D, Kemper J, Ebert D. 2012. Phylogeny, biology, and classification of extant holocephalans. In: Biology of Sharks and their Relatives. Carrier, J, Musick, J, & Heithaus, M, editors. CRC Press: Boca Raton, FL pp. 97–122.

Ebert D, Ho H, White W, Carvalho M. 2013. Introduction to the systematics and biodiversity of sharks, rays, and chimaeras (Chondrichthyes) of Taiwan. Zootaxa. 3752:5–19.

Evans R, Mangelsdorf D. 2014. Nuclear Receptors, RXR, and the Big Bang. Cell. 157:255–266. doi: 10.1016/j.cell.2014.03.012.

Filowitz G, Rajakumar R, O'Shaughnessy K, Cohn M. 2018. Cartilaginous Fishes Provide Insights into the Origin, Diversification, and Sexually Dimorphic Expression of Vertebrate Estrogen Receptor Genes. Mol Biol Evol. 35:2695–2701. doi: 10.1093/molbev/msy165.

Fonseca E et al. 2019. Evolutionary Plasticity in Detoxification Gene Modules: The Preservation and Loss of the Pregnane X Receptor in Chondrichthyes Lineages. Int J Mol Sci. 10:E2331. doi: 10.3390/ijms20092331.

Fonseca E et al. 2017. LXRα and LXRβ nuclear receptors evolved in the common ancestor of gnathostomes. Genome Biol Evol. 9:222–230. doi: 10.1093/gbe/evw305.

Germain P, Staels B, Dacquet C, Spedding M, Laudet V. 2006. Overview of nomenclature of nuclear receptors. Pharmacol Rev. 58:685–704. doi: 10.1124/pr.58.4.2.

Gronemeyer H, Gustafsson J, Laudet V. 2004. Principles for modulation of the nuclear receptor superfamily. Nat Rev Drug Discov. 3:950–964. doi: 10.1038/nrd1551.

Guindon S et al. 2010. New Algorithms and Methods to Estimate Maximum-Likelihood Phylogenies: Assessing the Performance of PhyML 3.0. Syst Biol. 59:307–321. doi: 10.1093/sysbio/syq010.

Gurevich A, Saveliev V, Vyahhi N, Tesler G. 2013. QUAST: quality assessment tool for genome assemblies. Bioinformatics. 29:1072–1075. doi: 10.1093/bioinformatics/btt086.

Hara Y et al. 2018. Shark genomes provide insights into elasmobranch evolution and the origin of vertebrates. Nat Ecol Evol. 2:1761–1771. doi: 10.1038/s41559-018-0673-5.

Howard-Ashby M et al. 2006. Gene families encoding transcription factors expressed in early development of Strongylocentrotus purpuratus. Dev Biol. 300:90–107. doi: 10.1016/j.ydbio.2006.08.033.

Huang W et al. 2015. Evolution of a novel nuclear receptor subfamily with emphasis on the member from the Pacific oyster Crassostrea gigas. Gene. 567:164–172. doi: 10.1016/j.gene.2015.04.082.

Inoue J et al. 2010. Evolutionary origin and phylogeny of the modern holocephalans (Chondrichthyes: Chimaeriformes): a mitogenomic perspective. Mol Biol Evol. 27:2576–2586. doi: 10.1093/molbev/msq147.

Katoh K, Toh H. 2010. Parallelization of the MAFFT multiple sequence alignment program. Bioinformatics. 26:1899–1900. doi: 10.1093/bioinformatics/bty121.

Katsiadaki I. 2019. Are marine invertebrates really at risk from endocrine-disrupting chemicals? Curr Opin Env. Sci Heal. 11:37–42. doi: 10.1016/j.coesh.2019.06.005.

Katsu Y et al. 2010. Cloning and functional characterization of Chondrichthyes, cloudy catshark, Scyliorhinus torazame and whale shark, Rhincodon typus estrogen receptors. Gen Comp Endocrinol. 168:496–504. doi: 10.1016/j.ygcen.2010.06.010.

Katsu Y et al. 2019. Transcriptional activation of elephant shark mineralocorticoid receptor by corticosteroids, progesterone, and spironolactone. Sci Signal. 12:eaar2668. doi: 10.1126/scisignal.aar2668.

Kaur S et al. 2015. The nuclear receptors of Biomphalaria glabrata and Lottia gigantea: implications for developing new model organisms. PLoS One. 10:e0121259. doi: 10.1371/journal.pone.0121259.

Kliewer S, Lehmann J, Willson T. 1999. Orphan nuclear receptors: shifting endocrinology into reverse. Science (80-.). 284:757–760. doi: 10.1126/science.284.5415.757.

Langmead B, Salzberg S. 2012. Fast gapped-read alignment with Bowtie 2. Nat Methods. 9:357–359. doi: 10.1038/nmeth.1923.

Laudet V, Gronemeyer H. 2002. The nuclear receptors factsbook. Academic Press: London:

Lecroisey C, Laudet V, Schubert M. 2012. The cephalochordate amphioxus: a key to reveal the secrets of nuclear receptor evolution. Br. Funct Genomics. 11:156–166. doi: 10.1093/bfgp/els008.

Lemaire G, Balaguer P, Michel S, Rahmani R. 2005. Activation of retinoic acid receptor-dependent transcription by organochlorine pesticides. Toxico Appl Pharmacol. 202:38–49. doi: 10.1016/j.taap.2004.06.004.

Luo R et al. 2012. SOAPdenovo2: an empirically improved memory-efficient short-read de novo assembler. Gigascience. 1:18. doi: 10.1186/2047-217X-1-18.

Maglich J et al. 2003. The first completed genome sequence from a teleost fish (Fugu rubripes) adds significant diversity to the nuclear receptor superfamily. Nucleic Acids Res. 31:4051–4058. doi: 10.1093/nar/gkg444.

le Maire A, Bourguet W, Balaguer P. 2010. A structural view of nuclear hormone receptor: endocrine disruptor interactions. Cell Mol Life Sci. 67:1219–1237. doi: 10.1007/s00018-009-0249-2.

Mangelsdorf D et al. 1995. The nuclear receptor superfamily: the second decade. Cell. 83:835–839. doi: 10.1016/0092-8674(95)90199-X.

Mangelsdorf D, Evans R. 1995. The RXR heterodimers and orphan receptors. Cell. 83:841–850. doi: 10.1016/0092-8674(95)90200-7.

Mapleson D, Accinelli G, Kettleborough G, Wright J, Clavijo B. 2017. KAT: a K-mer analysis toolkit to quality control NGS datasets and genome assemblies. Bioinformatics. 33:574–576. doi: 10.1093/bioinformatics/btw663.

Marçais G, Kingsford C. 2011. A fast, lock-free approach for efficient parallel counting of occurrences of k-mers. Bioinformatics. 27:764–770. doi: 10.1093/bioinformatics/btr011.

Marlétaz F et al. 2018. Amphioxus functional genomics and the origins of vertebrate gene regulation. Nature. 564:64–70. doi: 10.1038/s41586-018-0734-6.

Marra N et al. 2019. White shark genome reveals ancient elasmobranch adaptations associated with wound healing and the maintenance of genome stability. Proc Natl Acad Sci U S A. 116:4446–4455. doi: 10.1073/pnas.1819778116.

McKenna A et al. 2010. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res. 20:1297–1303. doi: 10.1101/gr.107524.110.

Metpally R, Vigneshwar R, Sowdhamini R. 2007. Genome inventory and analysis of nuclear hormone receptors in Tetraodon nigroviridis. J Biosci. 32:43–50. doi: 10.1007/s12038-007-0005-4.

Miller M et al. 2015. A RESTful API for Access to Phylogenetic Tools via the CIPRES Science Gateway. Evol. Bioinform Online. 11:43–48. doi: 10.4137/EBO.S21501.

Paps J, Holland P. 2018. Reconstruction of the ancestral metazoan genome reveals an increase in genomic novelty. Nat Commun. 9:1730. doi: 10.1038/s41467-018-04136-5.

Putnam N et al. 2008. The amphioxus genome and the evolution of the chordate karyotype. Nature. 453:1064–1071.

Ranallo-Benavidez T, Jaron K, Schatz M. 2019. GenomeScope 2.0 and Smudgeplots: Reference-free profiling of polyploid genomes. bioRxiv. doi: 10.1101/747568.

Robinson-Rechavi M, Carpentier A, Duffraisse M, Laudet V. 2001. How many nuclear hormone receptors are there in the human genome? Trends Genet. 17:554–556. doi: 10.1016/S0168-9525(01)02417-9.

Schubert M et al. 2008. Nuclear hormone receptor signaling in amphioxus. Dev Genes Evol. 218:651–665. doi: 10.1007/s00427-008-0251-y.

Siddiq M, Hochberg G, Thornton J. 2017. Evolution of protein specificity: Insights from ancestral protein reconstruction. Curr Opin Struct Biol. 47:113–122. doi: 10.1016/j.sbi.2017.07.003.

Simão F, Waterhouse R, Ioannidis P, Kriventseva E, Zdobnov E. 2015. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformatics. 31:3210–3212. doi: 10.1093/bioinformatics/btv351.

Simpson J. 2014. Exploring genome characteristics and sequence quality without a reference. Bioinformatics. 30:1228–1235. doi: 10.1093/bioinformatics/btu023.

Song L, Florea L, Langmead B. 2014. Lighter: fast and memory-efficient sequencing error correction without counting. Genome Biol. 15:509. doi: 10.1186/s13059-014-0509-9.

Thornton J. 2001. Evolution of vertebrate steroid receptors from an ancestral estrogen receptor by ligand exploitation and serial genome expansions. Proc Natl Acad Sci U S A. 98:5671–5676. doi: 10.1073/pnas.091553298.

Venkatesh B et al. 2006. Ancient noncoding elements conserved in the human genome. Science (80-.). 314:1892. doi: 10.1126/science.1130708.

Venkatesh B et al. 2014. Elephant shark genome provides unique insights into gnathostome evolution. Nature. 505:174–179. doi: 10.1038/nature12826.

Venkatesh B et al. 2007. Survey Sequencing and Comparative Analysis of the Elephant Shark (Callorhinchus milii) Genome. PLoS Biol. 5:e101. doi: 10.1371/journal.pbio.0050101.

Vogeler S, Galloway T, Lyons B, Bean T. 2014. The nuclear receptor gene family in the Pacific oyster, Crassostrea gigas, contains a novel subfamily group. BMC Genomics. 15:369. doi: 10.1186/1471-2164-15-369.

Ward R, Zemlak T, Innes B, Last P, Hebert P. 2005. DNA barcoding Australia's fish species. Philos Trans R Soc L. B Biol Sci. 360:1847–1857. doi: 10.1098/rstb.2005.1716.

Wu W, Niles E, El-Sayed N, Berriman M, LoVerde P. 2006. Schistosoma mansoni (Platyhelminthes, Trematoda) nuclear receptors: Sixteen new members and a novel subfamily. Gene. 366:303–315. doi: 10.1016/j.gene.2005.09.013.

Wu W, Niles E, Hirai H, LoVerde P. 2007. Evolution of a novel subfamily of nuclear receptors with members that each contain two DNA binding domains. BMC Evol Biol. 7:27. doi: 10.1186/1471-2148-7-27.

Yagi K et al. 2003. A genomewide survey of developmentally relevant genes in Ciona intestinalis: III. Genes for Fox, ETS, nuclear receptors and NFkappaB. Dev Genes Evol. 213:235–244. doi: 10.1007/s00427-003-0322-z.

Zhao Y, Zhang K, Giesy J, Hu J. 2015. Families of nuclear receptors in vertebrate models: characteristic and comparative toxicological perspective. Sci Rep. 5:8554. doi: 10.1038/srep08554.

2.6. Supplementary Material

Table S2.1. List of accession numbers of nuclear receptors.

NR	Spesies	Protein ACC number	DB source
NR0B1	*Homo sapiens*	NP_000466.2	NM_000475.4
	Mus musculus	NP_031456.1	NM_007430.5
	Gallus gallus	NP_989924.1	NM_204593.1
	Anolis carolinensis	XP_008105579.1	XM_008107372.2
	Xenopus tropicalis	XP_002933661.1	XM_002933615.4
	Danio rerio	NP_001076416.1	NM_001082947.1
	Lepisosteus oculatus	XP_006639240.1	XM_006639177.2
	Rhincodon typus	XP_020386254.1	XM_020530665.1
	Chiloscyllium punctatum		Chipu0000450.t1
	Callorhinchus milii	XP_007903610.1	XM_007905419.1
NR0B2	*Homo sapiens*	NP_068804.1	NM_021969.2
	Mus musculus	NP_035980.1	NM_011850.3
	Gallus gallus	NP_001026064.2	NM_001030893.2
	Anolis carolinensis	XP_003227509.1	XM_003227461.3
	Danio rerio A	NP_001243120.1	NM_001256191.1
	Danio rerio B	XP_001338278.1	XM_001338242.4
	Lepisosteus oculatus	XP_006631461.1	XM_006631398.2
	Rhincodon typus	XP_020369760.1	XM_020514171.1
	Scyliorhinus torazame		Scyto0016001.t1
	Chiloscyllium punctatum		Chipu0005253.t1
	Leucoraja erinacea		SRX2488463/4/5/6
	Scyliorhinus canicula		SRX651774
	Callorhinchus milii	XP_007893128.1	XM_007894937.1
NR0B3	*Rhincodon typus*	XP_020390993.1	XM_020535404.1
	Scyliorhinus torazame		Scyto0005422.t1
	Chiloscyllium punctatum		Chipu0018193.t1
	Callorhinchus milii	XP_007892781.1	XM_007894590.1_SRX154859
NR0B4	*Xenopus tropicalis*	XP_012818749.1	XM_012963295.2
	Rhincodon typus	XP_020387135.1	XM_020531546.1
	Scyliorhinus torazame		Scyto0002851.t1
	Chiloscyllium punctatum		Chipu0005589.t1
	Leucoraja erinacea		SRX2488463/4/5/6
	Callorhinchus milii	XP_007903240.1	XM_007905049.1
NR0B-1	*Branchiostoma floridae*	XP_002609613.1	XM_002609567.1
	Branciostoma lanceolatum	BL16095	
NR0B-2	*Branchiostoma floridae*	XP_002599038.1	XM_002598992.1
	Branciostoma lanceolatum	BL00088	
NR1A1	*Homo sapiens*	P10827.1	P10827
	Mus musculus	NP_001300912.1	NM_001313983.1
	Gallus gallus	P04625.1	P04625
	Anolis carolinensis	XP_008111555.1	XM_008113348.2
	Xenopus tropicalis	XP_012807993.1	XM_012952539.1
	Danio rerio A	NP_571471.1	NM_131396.1
	Danio rerio B	XP_001921013.4	XM_001920978.6
	Lepisosteus oculatus	XP_015217671.1	XM_015362185.1
	Rhincodon typus	XP_020368506.1	XM_020512917.1
	Scyliorhinus torazame		Scyto0001273.t1_Scyto0001274.t1
	Chiloscyllium punctatum		Chipu0011015.t1
	Leucoraja erinacea		SRX2488464
	Scyliorhinus canicula		EF672346.1_SRX651774
	Callorhinchus milii	XP_007883023.1	XM_007884832.1
NR1A2	*Homo sapiens*	P10828.2	P10828
	Mus musculus	AAI19553.1	BC119552.1
	Gallus gallus	NP_001239150	NM_001252221.2
	Anolis carolinensis	XP_003226266.1	XM_003226218.3
	Xenopus tropicalis	XP_012820319	XM_012964865.2
	Danio rerio	BAN67994.1	AB759513.1
	Lepisosteus oculatus	XP_015212984	XM_015357498.1
	Rhincodon typus		Rhity0003347.t1
	Leucoraja erinacea		SRX2488464
	Scyliorhinus torazame		Scyto0006427.t1
	Chiloscyllium punctatum		Chipu0002468.t1
	Scyliorhinus canicula		SRX651774
	Callorhinchus milii	XP_007893682	XM_007895491.1
NR1A	*Ciona intestinalis*	NP_001027658.1	NM_001032486.1
	Branchiostoma floridae	ABS11249.1	EF672344.1

	Branciostoma lanceolatum	BL95391	
NR1B1	*Homo sapiens*	NP_000955.1	NM_000964.3
	Mus musculus	P11416.1	P11416
	Gallus gallus	Q90966.1	Q90966
	Anolis carolinensis	XP_003222504.1	XM_003222456.3
	Xenopus tropicalis	NP_001164665.1	NM_001171194.1
	Danio rerio A	NP_571481.2	NM_131406.2
	Danio rerio B	NP_571474.1	NM_131399.1
	Lepisosteus oculatus	XP_006638289.1	XM_006638226.2
	Leucoraja erinacea		SRX2488464
	Scyliorhinus canicula		SRX651774
	Rhincodon typus	XP_020377259.1	XM_020521670.1
	Scyliorhinus torazame		Scyto0002336.t1
	Chiloscyllium punctatum		Chipu0011018.t1
	Callorhinchus milii	XP_007883927.1	XM_007885736.1
NR1B2	*Homo sapiens*	P10826.2	P10826
	Mus musculus	P22605.1	P22605
	Gallus gallus	NP_990657.1	NM_205326.1
	Anolis carolinensis	XP_008117474.1	XM_008119267.2
	Xenopus tropicalis	XP_002932450.2	XM_002932404.4
	Oryzias latipes	XP_020562893.1	XM_020707234.1
	Scleropages formosus	XP_018605782.1	XM_018750266.1
	Lepisosteus oculatus	XP_015212912.1	XM_015357426.1
	Leucoraja erinacea		SRX2488464
	Scyliorhinus canicula		SRX651774
	Rhincodon typus	XP_020376747.1_XP_020381359.1	SRX657786
	Scyliorhinus torazame		Scyto0003262.t1
	Chiloscyllium punctatum		Chipu0002469.t1
	Callorhinchus milii	XP_007893684.1	XM_007895493.1
NR1B3	*Homo sapiens*	AAA52692.1	M24857.1
	Mus musculus	AAA40035.1	M34476.1
	Gallus gallus	CAA52153.2	X73973.2
	Colinus virginianus	OXB61961.1	
	Coturnix japonica	XP_015705888.1	
	Anolis carolinensis	XP_008101794.1	XM_008103587.2
	Xenopus tropicalis	XP_002936679.1	XM_002936633.4
	Danio rerio A	NP_571414.1	NM_131339.1
	Danio rerio B	NP_001076779.1	NM_001083310.1
	Lepisosteus oculatus	XP_015199750.1	XM_015344264.1
	Leucoraja erinacea		SRX2488464
	Rhincodon typus		SRX657786
NR1B	*Ciona intestinalis*	NP_001071806.1	NM_001078338.1
	Branchiostoma floridae	XP_002598475.1	XM_002598429.1
	Branciostoma lanceolatum	BL10401	
NR1C1	*Homo sapiens*	NP_001001928.1	NM_001001928.2
	Mus musculus	NP_001106889.1	NM_001113418.1
	Gallus gallus	NP_001001464.1	NM_001001464.1
	Anolis carolinensis	XP_003221452.1	XM_003221404.3
	Xenopus tropicalis	XP_002940784.2	XM_002940738.4
	Danio rerio A	NP_001154805.1	NM_001161333.1
	Danio rerio B	NP_001096037.1	NM_001102567.1
	Lepisosteus oculatus	XP_015208771.1_SRX661019	XM_015353285.1
	Leucoraja erinacea		SRX2488464
	Scyliorhinus canicula		SRX036537_SRX651773/4/5
	Rhincodon typus	XP_020370894.1	XM_020515305.1
	Scyliorhinus torazame		Scyto0002575.t1
	Chiloscyllium punctatum		Chipu0007271.t1
	Callorhinchus milii	XP_007892874.1	XM_007894683.1
NR1C2	*Homo sapiens*	NP_001165289.1	NM_001171818.1
	Mus musculus	NP_035275.1	NM_011145.3
	Gallus gallus	NP_990059.1	NM_204728.1
	Anolis carolinensis	XP_003220387.1	XM_003220339.3
	Xenopus tropicalis	XP_004910760.1	XM_004910703.3
	Danio rerio A	XP_699900.6	XM_694808.8
	Danio rerio B	NP_571543.1	NM_131468.2
	Lepisosteus oculatus	XP_015198262.1	XM_015342776.1
	Leucoraja erinacea		SRX2488464
	Scyliorhinus canicula		SRX036537_SRX651773/4/5
	Rhincodon typus	XP_020370048.1	XM_020514459.1
	Scyliorhinus torazame		Scyto0013770.t1
	Chiloscyllium punctatum		Chipu0002705.t1

	Callorhinchus milii	XP_007892312.1	XM_007894121.1
NR1C3	*Homo sapiens*	NP_005028.4	NM_005037.5
	Mus musculus	NP_001120802.1	NM_001127330.2
	Gallus gallus	NP_001001460.1	NM_001001460.1
	Anolis carolinensis	XP_016847098.1	XM_016991609.1
	Xenopus tropicalis	XP_012810576.1	XM_012955122.2_SRX485825
	Danio rerio	NP_571542.1	NM_131467.1
	Lepisosteus oculatus	XP_006631094.2	XM_006631031.2
	Leucoraja erinacea		SRX2488464
	Scyliorhinus canicula		SRX036537_SRX651773/4/5
	Rhincodon typus	XP_020366499.1	XM_020510910.1
	Scyliorhinus torazame		Scyto0007316.t1
	Chiloscyllium punctatum		Chipu0000207.t1
	Callorhinchus milii	XP_007901520.1	XM_007903329.1
NR1C	*Ciona intestinalis*	NP_001071801.1	NM_001078333.1
	Branchiostoma floridae	XP_002598634.1	XM_002598588.1
	Branciostoma lanceolatum	BL01590_BL27633	
NR1D1	*Homo sapiens*	NP_068370.1	NM_021724.4
	Mus musculus	NP_663409.2	NM_145434.4
	Gallus gallus	XP_015155042.1	XM_015299556.1
	Anolis carolinensis	XP_003222497.1	XM_003222449.3
	Xenopus tropicalis	NP_001093675.1	NM_001100205.1
	Danio rerio	NP_991292.2	NM_205729.2
	Lepisosteus oculatus	XP_015217762.1	XM_015362276.1_SRX661021
	Rhincodon typus	XP_020368505.1	XM_020512916.1
	Scyliorhinus torazame		Scyto0001275.t1
	Chiloscyllium punctatum		Chipu0011014.t1
	Callorhinchus milii	XP_007882839.1	XM_007884648.1
NR1D2	*Homo sapiens*	NP_005117.3	NM_005126.4
	Mus musculus	NP_035714.3	NM_011584.4
	Gallus gallus	NP_990536.2	NM_205205.2
	Anolis carolinensis	XP_008117168.2	XM_008118961.2
	Xenopus tropicalis	OCA31744.1	CM004448.1
	*Danio rerio*A	NP_001124064.2	NM_001130592.2
	*Danio rerio*B	NP_571140.1	NM_131065.1
	Lepisosteus oculatus	XP_006635830.1	XM_006635767.2
	Leucoraja erinacea	ctg64377	
	Rhincodon typus	XP_020373766.1	XM_020518177.1
	Scyliorhinus torazame		Scyto0006428.t1
	Chiloscyllium punctatum		Chipu0002467.t1
	Callorhinchus milii	XP_007893681.1	XM_007895490.1
NR1D4	*Anolis carolinensis*	XP_008101793.1	XM_008103586.1
	Danio rerio A	NP_001272465.1	NM_001285536.1
	Danio rerio B	NP_001272462.1	NM_001285533.1
	Lepisosteus oculatus	XP_015199579.1	XM_015344093.1
NR1D	*Ciona intestinalis*	NP_001071962.1	NM_001078494.1
	Branchiostoma floridae	XP_002598635.1	XM_002598589.1
	Branciostoma lanceolatum	BL08608	
NR1F1	*Homo sapiens*	NP_599022.1	NM_134260.2
	Mus musculus	NP_038674.1	NM_013646.2
	Gallus gallus	NP_001276816.1	NM_001289887.1
	Anolis carolinensis	XP_008118164.1	XM_008119957.1
	Xenopus tropicalis	XP_012822203.1	XM_012966749.2
	Danio rerio A	NP_001103637.1	NM_001110167.1
	Danio rerio B	NP_957361.1	NM_201067.1
	Lepisosteus oculatus	XP_006628975.1	XM_006628912.2
	Scyliorhinus canicula		SRX651774
	Rhincodon typus	XP_020377941.1_SRX657786	XM_020522352.1
	Scyliorhinus torazame		Scyto0016606.t1_Scyto0003854.t1
	Chiloscyllium punctatum		Chipu0019601.t1_Chipu0025440.t1
	Callorhinchus milii	XP_007905765.1	XM_007907574.1
NR1F2	*Homo sapiens*	NP_008845.2	NM_006914.3
	Mus musculus	NP_001036819.1	NM_001043354.2
	Gallus gallus	NP_990424.1	NM_205093.1
	Anolis carolinensis	XP_016846264.1	XM_016990775.1
	Xenopus tropicalis	XP_002940077.1	XM_002940031.4
	Danio rerio	NP_001076325.1	NM_001082856.1
	Lepisosteus oculatus	XP_006627046.2	XM_006626983.2
	Scyliorhinus canicula		SRX651774
	Rhincodon typus	XP_020385586.1_XP_020375824.1	SRX657786
	Scyliorhinus torazame		Scyto0003224.t1

	Chiloscyllium punctatum		Chipu0003514.t1
	Callorhinchus milii	XP_007897563.1	XM_007899372.1
NR1F2 like	*Gallus gallus*	XP_003642912.2	XM_003642864.3
	Anolis carolinensis		SRX111454
	Xenopus tropicalis	XP_002938868.1	XM_002938822.4
	Danio rerio A	NP_001076288	NM_001082819.1
	Danio rerio B	NP_001264023.1	NM_001277094.2
	Lepisosteus oculatus	XP_015221273.1	XM_015365787.1
	Scyliorhinus canicula		SRX651774
	Rhincodon typus		Rhity0000094.t1
	Scyliorhinus torazame		Scyto0021577.t1
	Chiloscyllium punctatum		Chipu0003660.t1
	Callorhinchus milii	XP_007908356.1	XM_007910165.1
NR1F3	*Homo sapiens*	NP_005051.2	NM_005060.3
	Mus musculus	NP_035411.2	NM_011281.3
	Gallus gallus	XP_015135499.1	XM_015280013.1
	Aquila chrysaetos canadensis	XP_011597873.1	XM_011599571.1
	Numida meleagris	XP_021232065.1	XM_021376390.1
	Anolis carolinensis	XP_016854407.1	XM_016998918.1
	Xenopus tropicalis	XP_012825302.2	XM_012969848.2
	Danio rerio	XP_001344049.3	XM_001344013.5
	Lepisosteus oculatus	XP_015224330.1	XM_015368844.1
	Callorhinchus milii	XP_007883515.1	XM_007885324.1
NR1F	*Ciona intestinalis*	NP_001072021.1	NM_001078553.1
	Branchiostoma floridae	XP_002597918.1	XM_002597872.1
	Branciostoma lanceolatum	BL08657	
NR1H2	*Homo sapiens*	NP_009052.3	NM_007121.5
	Mus musculus	NP_001272446.1	NM_001285517.1
	Anolis carolinensis	XP_003222765.1	XM_003222717.3
	Xenopus tropicalis	NP_001072853.1	NM_001079385.1
	Leucoraja erinacea	AQR58541.1_SRX036536	KY094508.1
	Scyliorhinus canicula		ctg29778_ctg35662_ctg1931
	Rhincodon typus	XP_020390950.1	XM_020535361.1
	Scyliorhinus torazame		Scyto0014090.t1
	Chiloscyllium punctatum		Chipu0009762.t1
	Callorhinchus milii	XP_007883658.1	XM_007885467.1
NR1H3	*Homo sapiens*	NP_001238864.1	NM_001251935.1
	Mus musculus	NP_038867.2	NM_013839.4
	Gallus gallus	NP_989873.1	NM_204542.2
	Anolis carolinensis	XP_003214647.1	XM_003214599.3
	Danio rerio	NP_001017545.2	NM_001017545.3
	Lepisosteus oculatus	AQR58545.1	KY094512.1_SRX661023
	Leucoraja erinacea	AQR58540.1	KY094507.1
	Scyliorhinus canicula	AQR58544.1_SRX651773	KY094511.1
	Rhincodon typus	XP_020372680.1	XM_020517091.1
	Scyliorhinus torazame		Scyto0019021.t
	Chiloscyllium punctatum		Chipu0018244.t1
	Callorhinchus milii		SRX699034 partial
NR1H4	*Homo sapiens*	NP_005114.1	NM_005123.3
	Mus musculus	NP_001157172.1	NM_001163700.1
	Gallus gallus	NP_989444.1	NM_204113.2
	Anolis carolinensis	XP_003221144.1	XM_003221096.2
	Xenopus tropicalis	XP_002936891.2	XM_002936845.4
	Danio rerio	AAH92785.1	BC092785.1
	Lepisosteus oculatus	XP_015207318.1	XM_015351832.1
	Scyliorhinus canicula		SRX651773
	Rhincodon typus		SRX657786
	Scyliorhinus torazame		Scyto0000963.t1
	Chiloscyllium punctatum		Chipu0017012.t1_Chipu0012034.t1
	Callorhinchus milii	XP_007893586.1	XM_007895395.1
NR1H5	*Mus musculus*	NP_941060.2	NM_198658.2
	Gallus gallus	XP_004935064.1	XM_004935007.2
	Anolis carolinensis	XP_016848846.1	XM_016993357.1
	Xenopus tropicalis	NP_001107159.1	NM_001113687.1
	Danio rerio	NP_001116713.1	NM_001123241.1
	Lepisosteus oculatus	XP_015197747.1	XM_015342261.1
	Leucoraja erinacea	ABP98947.1	EF520727.1
	Scyliorhinus canicula		SRX651773
	Rhincodon typus	XP_020386533.1	XM_020530944.1
	Scyliorhinus torazame		Scyto0021828.t1_Scyto0017960.t1
	Chiloscyllium punctatum		Chipu0015608.t1

	Callorhinchus milii	XP_007899810.1	XM_007901619.1
NR1H-1	*Ciona intestinalis*	BAE06541.1	AB210536.1
	Branchiostoma floridae 8	EEN45473.1 (222287)	GG666659.1
	Branciostoma lanceolatum 8	BL08609_evm6	
NR1H-2	*Ciona intestinalis*	FAA00147.1	BR000116.1
	Branchiostoma floridae 1	86645	
	Branchiostoma floridae 2	124680	
	Branchiostoma floridae 3	124679	
	Branchiostoma floridae 4	124681	
	Branchiostoma floridae 5	128090	
	Branchiostoma floridae 6	124948	
	Branchiostoma floridae 7	156544	
	Branchiostoma floridae 9	222341	
	Branchiostoma floridae 10	253062	
	Branciostoma lanceolatum 1	BL97043_cuf5	
	Branciostoma lanceolatum 2	BL97043_cuf9	
	Branciostoma lanceolatum 3	BL16685_cuf1	
	Branciostoma lanceolatum 4	BL97044_cuf6	
	Branciostoma lanceolatum 5	BL09785_cuf0	
	Branciostoma lanceolatum 6	BL20884_cuf7	
	Branciostoma lanceolatum 7	BL05413_cuf1	
	Branciostoma lanceolatum 9	BL23676_cuf1	
	Branciostoma lanceolatum 10	BL97043_cuf15	
NR1I1	*Homo sapiens*	NP_000367.1	NM_000376.2
	Mus musculus	NP_033530.2	NM_009504.4
	Gallus gallus	NP_990429.1	NM_205098.1
	Anolis carolinensis	XP_008101719.1	XM_008103512.2
	Xenopus tropicalis	XP_002935703.1	XM_002935657.4
	Danio rerio A	NP_570994.1	NM_130919.1
	Danio rerio B	NP_001153457.1	NM_001159985.1
	Lepisosteus oculatus	XP_015199783.1	XM_015344297.1
	Leucoraja erinacea	AIM62165.1	KJ925051.1
	Rhincodon typus		SRX657786_Rhity0019804.t1
	Scyliorhinus torazame		Scyto0013739.t1_DN103660_c0_g1_i3
	Chiloscyllium punctatum		Chipu0017931.t1
	Callorhinchus milii	XP_007908698.1	XM_007910507.1
NR1I2	*Homo sapiens*	NP_003880.3	NM_003889.3
	Mus musculus	NP_035066.1	NM_010936.3
	Xenopus tropicalis	NP_001091887.1	NM_001098417.1
	Danio rerio	XP_005167477.1	XM_005167420.4
	Lepisosteus oculatus	XP_006639043.1	XM_006638980.2
	Callorhinchus milii	XP_007894039.1	XM_007895848.1
NR1I3	*Homo sapiens*	NP_001070950.1	NM_001077482.2
	Mus musculus	NP_033933.2	NM_009803.5
	Gallus gallus	NP_990033.1	NM_204702.1
	Anolis carolinensis	XP_003230590.2	XM_003230542.3
	Xenopus tropicalis	ADW81978.1	HM117646.1
NR1I	*Ciona intestinalis* A	AHB39788.1	KC561370.1
	Ciona intestinalis B	NP_001037831.1	NM_001044366.1
NR2A1	*Homo sapiens*	NP_000448.3	NM_000457.4
	Mus musculus	NP_032287.2	NM_008261.3
	Gallus gallus	NP_001026026.1	NM_001030855.1
	Anolis carolinensis	XP_003220655.2	XM_003220607.3
	Xenopus tropicalis	XP_002933229.2	XM_002933183.4
	Danio rerio	NP_919349.1	NM_194368.1
	Lepisosteus oculatus	XP_015220204.1	XM_015364718.1
	Rhincodon typus	XP_020383050.1	XM_020527461.1
	Scyliorhinus torazame		Scyto0012432.t1
	Chiloscyllium punctatum		Chipu0001016.t1
	Callorhinchus milii	XP_007909687.1	XM_007911496.1
NR2A2	*Homo sapiens*	NP_004124.4	NM_004133.4
	Mus musculus	NP_038948.1	NM_013920.2
	Gallus gallus	XP_425924.4	XM_425924.5
	Anolis carolinensis	XP_003219633.1	XM_003219585.3
	Xenopus tropicalis	XP_002939634.1	XM_002939588.3
	Danio rerio	NP_001068579.2	NM_001075111.2
	Lepisosteus oculatus	XP_006634664.1	XM_006634601.2
	Rhincodon typus		Rhity0000857.t1
	Scyliorhinus torazame		Scyto0006727.t1
	Chiloscyllium punctatum		Chipu0004779.t1
	Callorhinchus milii	XP_007885199.1	XM_007887008.1

NR2A3	*Gallus gallus*	NP_001025747.2	NM_001030576.3
	Anolis carolinensis	XP_008120106.1	XM_008121899.2
	Xenopus tropicalis	XP_002936047.1	XM_002936001.4
	Danio rerio	NP_991109.1	NM_205546.1
	Lepisosteus oculatus	XP_006641400.1	XM_006641337.2
	Scyliorhinus torazame		Scyto0006379.t1
	Chiloscyllium punctatum		Chipu0018237.t1
	Rhincodon typus	XP_020367523.1	XM_020511934.1
	Callorhinchus milii	XP_007887435.1	XM_007889244.1
NR2A	*Ciona intestinalis*	NP_001071735.1	NM_001078267.1
	Branchiostoma floridae	XP_002612502.1	XM_002612456.1
	Branciostoma lanceolatum	BL12447	
NR2B1	*Homo sapiens*	NP_002948.1	NM_002957.5
	Mus musculus	NP_035435.1	NM_011305.3
	Gallus gallus	XP_003642339.2	XM_003642291.3
	Anolis carolinensis	XP_003222989.3	XM_003222941.3
	Xenopus tropicalis	XP_012824678.1	XM_012969224.2
	Danio rerio A	NP_001155023.1	NM_001161551.1
	Danio rerio B	NP_571228.1	NM_131153.1
	Lepisosteus oculatus	XP_015222432.1	XM_015366946.1
	Scyliorhinus canicula		SRX651773
	Rhincodon typus	XP_020375413.1	XM_020519824.1
	Scyliorhinus torazame		Scyto0022845.t1
	Chiloscyllium punctatum		Chipu0020531.t1
	Callorhinchus milii	XP_007901074.1	XM_007902883.1
NR2B2	*Homo sapiens*	NP_001257330.1	NM_001270401.1
	Mus musculus	NP_001192143.1	NM_001205214.1
	Anolis carolinensis	XP_008122539.2	XM_008124332.2
	Xenopus tropicalis	NP_001015937.1	NM_001015937.2
	Danio rerio A	NP_571350.1	NM_131275.1
	Danio rerio B	NP_571313.1	NM_131238.1
	Lepisosteus oculatus	XP_015195289.1	XM_015339803.1
	Scyliorhinus canicula		SRX651773
	Rhincodon typus	XP_020377920.1	XM_020522331.1_SRX657786
	Scyliorhinus torazame		Scyto0012071.t1
NR2B3	*Homo sapiens*	NP_008848.1	NM_006917.4
	Mus musculus	NP_033133.1	NM_009107.3
	Gallus gallus	NP_990625.1	NM_205294.1
	Xenopus tropicalis	XP_004913829.2	XM_004913772.3
	*Danio rerio*A	NP_571292.3	NM_131217.3
	*Danio rerio*B	NP_001002345.1	NM_001002345.1
	Lepisosteus oculatus	XP_006634677.2	XM_006634614.2
	Scyliorhinus canicula		SRX651774
	Rhincodon typus	XP_020381037.1	XM_020525448.1_SRX657786
	Scyliorhinus torazame		Scyto0019558.t1_Scyto0013776.t1
	Chiloscyllium punctatum		Chipu0019949.t1 _Chipu0024820.t1_Chipu0020786.t1
	Callorhinchus milii	XP_007901150.1	XM_007902959.1
NR2B	*Ciona intestinalis*	NP_001071809.1	NM_001078341.1
	Branchiostoma floridae	AAM46151.1	AF378829.1
	Branchiostoma lanceolatum	ANP24206.1	KX118110.1
NR2C1	*Homo sapiens*	NP_003288.2	NM_003297.3
	Mus musculus	NP_035759.3	NM_011629.3
	Gallus gallus	NP_989455.1	NM_204124.1
	Anolis carolinensis	XP_016849232.1	XM_016993743.1
	Xenopus tropicalis	NP_001016207.1	NM_001016207.2
	Danio rerio	XP_005164874.1	XM_005164817.3
	Lepisosteus oculatus	XP_006633373.1	XM_006633310.2
	Rhincodon typus	XP_020382857.1	XM_020527268.1
	Scyliorhinus torazame		Scyto0006343.t1_Scyto0024715.t1_Scyto0013014.t1
	Chiloscyllium punctatum		Chipu0016339.t1
	Callorhinchus milii	XP_007907268.1	XM_007909077.1
NR2C2	*Homo sapiens*	NP_003289.2	NM_003298.4
	Mus musculus	NP_001334271.1	NM_001347342.1
	Gallus gallus	XP_414462.3	XM_414462.5
	Anolis carolinensis	XP_008103523.1	XM_008105316.2
	Xenopus tropicalis	XP_012817160.1	XM_012961706.2
	Danio rerio	NP_001116766.1	NM_001123294.1
	Lepisosteus oculatus	XP_006630783.1	XM_006630720.2
	Rhincodon typus	XP_020379844.1_XP_020378982.1	SRX657786
	Scyliorhinus torazame		Scyto0018489.t1_Scyto0020489.t1_Scyto0020684.t1_Scyto0020737.t1
	Chiloscyllium punctatum		Chipu0018349.t1

	Callorhinchus milii	XP_007888647.1	XM_007890456.1
NR2C	*Ciona intestinalis*	NP_001071910.1	NM_001078442.1
	Branchiostoma floridae	XP_002599699.1	XM_002599653.1
	Branchiostoma lanceolatum	BL23000	
NR2E1	*Homo sapiens*	NP_001273031.1	NM_001286102.1
	Mus musculus	NP_689415.1	NM_152229.2
	Gallus gallus	NP_990501.1	NM_205170.1
	Anolis carolinensis	XP_008119032.1	XM_008120825.2
	Xenopus tropicalis	XP_004914612.1	XM_004914555.3
	Danio rerio	NP_001003608.1	NM_001003608.1
	Lepisosteus oculatus	XP_006626027.1	XM_006625964.2
	Rhincodon typus	XP_020383667.1	XM_020528078.1
	Scyliorhinus torazame		Scyto0017942.t1
	Chiloscyllium punctatum		Chipu0006639.t1
	Callorhinchus milii	XP_007891086.1	XM_007892895.1
NR2E2	*Homo sapiens*	CAB82769.1	AJ276674.1
	Mus musculus	NP_038736.1	NM_013708.4
	Gallus gallus	NP_989925.1	NM_204594.1
	Anolis carolinensis	XP_003227397.1	XM_003227349.3
	Xenopus tropicalis	NP_001090633.1	NM_001097164.1
	Danio rerio	NP_001007369.1	NM_001007368.1
	Lepisosteus oculatus	XP_006628992.1	XM_006628929.2
	Rhincodon typus	XP_020392512.1	XM_020536923.1
	Scyliorhinus torazame		Scyto0005111.t1
	Chiloscyllium punctatum		Chipu0012309.t1
	Callorhinchus milii	XP_007908162.1	XM_007909971.1
NR2E	*Branchiostoma floridae* 1	XP_002612243.1	XM_002612197.1
	Branchiostoma floridae 1-like	XP_002609373.1	XM_002609327.1
	Branchiostoma floridae 3	XP_002595944.1	XM_002595898.1
	Branchiostoma lanceolatum 1	BL05890	
	Branchiostoma lanceolatum 1-like	BL22964	
	Branchiostoma lanceolatum 3	BL11879	
NR2F1	*Homo sapiens*	NP_005645.1	NM_005654.5
	Mus musculus	NP_034281.2	NM_010151.3
	Gallus gallus	XP_003643114.1	XM_003643066.3
	Anolis carolinensis	XP_008101253.1	XM_008103046.2
	Xenopus tropicalis	XP_012815199.1	XM_012959745.2
	Danio rerio A	NP_571255.1	NM_131180.1
	Danio rerio B	NP_956886.1	NM_200592.1
	Lepisosteus oculatus	XP_006626565.1	XM_006626502.2
	Scyliorhinus canicula	AAS49607.1	AY393842.1_ SRX651774
	Rhincodon typus	XP_020392814.1	XM_020537225.1
	Scyliorhinus torazame		Scyto0006932.t1
	Chiloscyllium punctatum		Chipu0006013.t1
	Callorhinchus milii	XP_007891252.1	XM_007893061.1
NR2F2	*Homo sapiens*	NP_066285.1	NM_021005.3
	Mus musculus	NP_033827.2	NM_009697.3
	Gallus gallus	NP_989752.1	NM_204421.1
	Anolis carolinensis	XP_008117243.2	XM_008119036.2
	Xenopus tropicalis	NP_001107703.1	NM_001114231.1
	Danio rerio	NP_571258.1	NM_131183.1
	Lepisosteus oculatus	XP_006628930.1	XM_006628867.2
	Rhincodon typus	XP_020382960.1	XM_020527371.1
	Scyliorhinus torazame		Scyto0000699.t1
	Chiloscyllium punctatum		Chipu0008489.t1
	Callorhinchus milii	XP_007905008.1	XM_007906817.1
NR2F5	*Anolis carolinensis*	XP_003228600.2	XM_003228552.3
	Xenopus tropicalis	XP_002938509.1	XM_002938463.4
	Danio rerio	NP_571261.1	NM_131186.1
	Lepisosteus oculatus	XP_015192161.1	XM_015336675.1
	Rhincodon typus	XP_020380563.1	XM_020524974.1
	Chiloscyllium punctatum		Chipu0020572.t1
NR2F6	*Homo sapiens*	NP_005225.2	NM_005234.3
	Mus musculus	NP_034280.2	NM_010150.2
	Gallus gallus	XP_015155582.1	XM_015300096.1
	Xenopus tropicalis	NP_001004841.1	NM_001004841.1
	Danio rerio A	NP_991120.1	NM_205557.1
	Danio rerio B	NP_998404.1	NM_213239.1
	Lepisosteus oculatus	XP_015221106.1	XM_015365620.1
	Rhincodon typus		Rhity0016697.t1
	Scyliorhinus torazame		Scyto0025131.t1

	Chiloscyllium punctatum		Chipu0026370.t1_Chipu0015782.t1
	Callorhinchus milii	XP_007896989.1	XM_007898798.1
NR2F	*Ciona intestinalis*	NP_001071673.1	NM_001078205.1
	Branchiostoma floridae	AAO61416.1	AY211769.1
	Branchiostoma lanceolatum	BL07175	
NR3A1	*Homo sapiens*	AAA52399.1	M12674.1
	Mus musculus	NP_031982.1	NM_007956.5
	Gallus gallus	XP_015139536.1	XM_015284050.1
	Anolis carolinensis	NP_001277446.1	NM_001290517.1
	Xenopus tropicalis	BAE81788.1	AB244211.1
	Danio rerio	XP_009297713.1	XM_009299438.2
	Lepisosteus oculatus	XP_006625908.1	XM_006625845.2
	Leucoraja erinacea		SRX036634
	Scyliorhinus canicula		CLONE
	Rhincodon typus	XP_020388534.1_XP_020368227.1	
	Scyliorhinus torazame		Scyto0008126.t1
	Chiloscyllium punctatum		Chipu0011139.t1_Chipu0011138.t1_Chipu0011137.t1_Chipu0011136.t1
	Callorhinchus milii	XP_007892594.1	XM_007894403.1
NR3A2	*Homo sapiens*	BAA24953.1	AB006590.1
	Mus musculus	NP_997590.1	NM_207707.1
	Gallus gallus	NP_990125.1	NM_204794.2
	Anolis carolinensis	XP_016846689.1	XM_016991200.1
	Xenopus tropicalis	NP_001035101.1	NM_001040012.2
	Danio rerio A	NP_851297.1	NM_180966.2
	Danio rerio B	NP_777287.2	NM_174862.3
	Lepisosteus oculatus	XP_006632252.1	XM_006632189.2
	Scyliorhinus canicula		clone
	Rhincodon typus	XP_020380159.1	XM_020524570.1
	Chiloscyllium punctatum		Chipu0021385.t1
	Callorhinchus milii	BAX07664.1	LC068848.1
NR3A	*Branchiostoma floridae*	ACF16007	EU714009.1
	Branchiostoma lanceolatum	BL17851	
NR3B1	*Homo sapiens*	NP_004442.3	NM_004451.4
	Mus musculus	NP_031979.2	NM_007953.2
	Anolis carolinensis	XP_003230133.2	XM_003230085.2
	Xenopus tropicalis	NP_001072756.1	NM_001079288.1
	Danio rerio A	NP_998120.1	NM_212955.1
	Danio rerio B	XP_005160912.1	XM_005160855.3
NR3B2	*Homo sapiens*	NP_004443.3	NM_004452.3
	Mus musculus	NP_036064.3	NM_011934.4
	Gallus gallus	XP_015143195.1	XM_015287709.1
	Anolis carolinensis	XP_008119299.1	XM_008121092.2
	Xenopus tropicalis	OCA25838.1	CM004450.1
	Danio rerio	NP_001311468.1	NM_001324539.1
	Lepisosteus oculatus	XP_015206123.1	XM_015350637.1
	Rhincodon typus		SRX657786_Rhity0023221.t1
	Chiloscyllium punctatum		DRX104910_Chipu0023578.t1_Chipu0027822.t1_Chipu0028318.t1
	Callorhinchus milii	XP_007884349.1 _XP_007909977.1	XM_007886158.1_XM_007911786.1
NR3B3	*Homo sapiens*	NP_001429.2	NM_001438.3
	Mus musculus	NP_036065.1	NM_011935.3
	Gallus gallus	NP_001007082.1	NM_001007081.1
	Anolis carolinensis	XP_008124062.1	XM_008125855.2
	Xenopus tropicalis	XP_017949231.1	XM_018093742.1
	Danio rerio	XP_005158828.1	XM_005158771.4
	Lepisosteus oculatus	XP_015201956.1	XM_015346470.1
	Leucoraja erinacea	ctg10651	
	Rhincodon typus		Rhity0001164.t1
	Scyliorhinus torazame		Scyto0006980.t1
	Chiloscyllium punctatum		Chipu0001857.t1
	Callorhinchus milii	XP_007902690.1	XM_007904499.1
NR3B4	*Xenopus tropicalis*	XP_002938860.2	XM_002938814.4
	Danio rerio	XP_001921093.3	XM_001921058.7
	Lepisosteus oculatus	XP_006627644.1	XM_006627581.2
	Rhincodon typus		Rhity0014964.t1_XM_020528994.1
	Scyliorhinus torazame		Scyto0023574.t1
	Chiloscyllium punctatum		Chipu0014347.t1
	Callorhinchus milii	XP_007884860.1	XM_007886669.1
NR3B	*Ciona intestinalis*	NP_001071700.1	NM_001078232.1
	Branchiostoma floridae	AAU88062	AY738654.1
	Branchiostoma lanceolatum	BL08547	
NR3C1	*Homo sapiens*	NP_000167.1	NM_000176.2

	Mus musculus	NP_032199.3	NM_008173.3
	Gallus gallus	NP_001032915.1	NM_001037826.1
	Anolis carolinensis	XP_003217421.1	XM_003217373.3
	Xenopus tropicalis	NP_001016967.1	NM_001016967.3
	Danio rerio	NP_001018547.2	NM_001020711.3
	Lepisosteus oculatus	XP_015204941.1	XM_015349455.1
	Leucoraja erinacea	ABD46744.1	DQ382338.1_SRX2488466
	Scyliorhinus canicula	AEF12277.1	JF896319.1_SRX651773
	Rhincodon typus	XP_020370084.1	XM_020514495.1
	Scyliorhinus torazame		Scyto0001101.t1
	Chiloscyllium punctatum		Chipu0015189.t1
	Callorhinchus milii	XP_007899521.1	XM_007901330.1
NR3C2	*Homo sapiens*	NP_000892.2	NM_000901.4
	Mus musculus	NP_001077375.1	NM_001083906.1
	Gallus gallus	NP_001152817.1	NM_001159345.1
	Anolis carolinensis	XP_008110199.1	XM_008111992.2
	Xenopus tropicalis	XP_002933547.2	XM_002933501.4
	Danio rerio	NP_001093873.1	NM_001100403.1
	Lepisosteus oculatus	XP_006629556.1	XM_006629493.2
	Leucoraja erinacea	ABD46745.1	DQ382339.1
	Rhincodon typus	XP_020379699.1_XP_020389419.1	XM_020524110.1_XM_020533830.1
	Scyliorhinus torazame		Scyto0001313.t1
	Chiloscyllium punctatum		Chipu0013822.t1
	Callorhinchus milii	XP_007902220.1	XM_007904029.1
NR3C3	*Homo sapiens*	NP_000917.3	NM_000926.4
	Mus musculus	NP_032855.2	NM_008829.2
	Gallus gallus	NP_990593.1	NM_205262.1
	Anolis carolinensis	XP_016848163.1	XM_016992674.1
	Xenopus tropicalis	XP_002935617.1	XM_002935571.4
	Danio rerio	NP_001159807.1	NM_001166335.1
	Lepisosteus oculatus	XP_006628126.2	XM_006628063.2
	Leucoraja erinacea	ABD46747.1	DQ382341.1
	Scyliorhinus canicula		
	Rhincodon typus	XP_020366030.1	XM_020510441.1
	Scyliorhinus torazame		Scyto0018491.t1
	Chiloscyllium punctatum		Chipu0014491.t1
	Callorhinchus milii	XP_007900068.1	XM_007901877.1
NR3C4	*Homo sapiens*	NP_000035.2	NM_000044.4
	Mus musculus	NP_038504.1	NM_013476.4
	Gallus gallus	NP_001035179.1	NM_001040090.1
	Anolis carolinensis	XP_008118585.1	XM_008120378.1
	Xenopus tropicalis	XP_002941888.2	XM_002941842.4
	Danio rerio	NP_001076592.1	NM_001083123.1
	Lepisosteus oculatus	XP_006632826.1	XM_006632763.2
	Leucoraja erinacea	ABD46746.1	DQ382340.1
	Rhincodon typus		XM_020515388.1_Rhity0010156.t1
	Scyliorhinus torazame		Scyto0007431.t1_Scyto0010180.t1
	Chiloscyllium punctatum		Chipu0000531.t1
	Callorhinchus milii	XP_007892478.1	XM_007894287.1
NR3C	*Branchiostoma floridae*	ACB10649.1	EU371729.1
	Branchiostoma lanceolatum	BL05603	
NR4A1	*Homo sapiens*	NP_002126.2	NM_002135.4
	Mus musculus	NP_034574.1	NM_010444.2
	Gallus gallus	XP_015129004.1	XM_015273518.1
	Anolis carolinensis	XP_008102184.1	XM_008103977.2
	Xenopus tropicalis	NP_001072303.1	NM_001078835.1
	Danio rerio	NP_001002173.1	NM_001002173.1
	Lepisosteus oculatus	XP_015199566.1	XM_015344080.1
	Rhincodon typus		Rhity0017553.t1
	Scyliorhinus torazame		Scyto0021457.t1_Scyto0022272.t1
	Chiloscyllium punctatum		Chipu0020028.t1
	Callorhinchus milii	XP_007908140.1	XM_007909949.1
NR4A2	*Homo sapiens*	NP_006177.1	NM_006186.3
	Mus musculus	NP_038641.1	NM_013613.2
	Gallus gallus	XP_015145687.1	XM_015290201.1
	Anolis carolinensis	XP_016851341.1	XM_016995852.1
	Xenopus tropicalis	NP_001093678.1	NM_001100208.1
	Danio rerio A	NP_001106956.1	NM_001113484.1
	Danio rerio B	NP_001002406.1	NM_001002406.1
	Lepisosteus oculatus	XP_006636492.1	XM_006636429.2
	Rhincodon typus	XP_020383762.1	XM_020528173.1

	Scyliorhinus torazame		Scyto0003452.t1
	Chiloscyllium punctatum		Chipu0000439.t1
	Callorhinchus milii	XP_007888111.1	XM_007889920.1
NR4A3	*Homo sapiens*	NP_775292.1	NM_173200.2
	Mus musculus	NP_056558.1	NM_015743.3
	Gallus gallus	XP_015137891.1	XM_015282405.1
	Anolis carolinensis	XP_008111173.1	XM_008112966.2
	Xenopus tropicalis	XP_017950372.1	XM_018094883.1
	Danio rerio	NP_001166100.1	NM_001172629.1
	Lepisosteus oculatus	XP_015212798.1	XM_015357312.1
	Rhincodon typus	XP_020387536.1	XM_020531947.1
	Scyliorhinus torazame		Scyto0010033.t1
	Chiloscyllium punctatum		Chipu0003458.t1
	Callorhinchus milii	XP_007895627.1	XM_007897436.1
NR4A	*Ciona intestinalis*	NP_001071779.1	NM_001078311.1
	Branchiostoma floridae	g20862.t1	
	Branchiostoma lanceolatum	BL13378	
NR5A1	*Homo sapiens*	NP_004950.2	NM_004959.4
	Mus musculus	NP_001303616.1	NM_001316687.1
	Gallus gallus	NP_990408.1	NM_205077.1
	Anolis carolinensis	XP_008122440.1	XM_008124233.2
	Xenopus tropicalis	NP_001139213.1	NM_001145741.1
	Danio rerio A	NP_571869.1	NM_131794.1
	Danio rerio B	NP_997999.1	NM_212834.1
	Lepisosteus oculatus	XP_006640764.2	XM_006640701.2
	Leucoraja erinacea		SRX2488466
	Rhincodon typus	XP_020371669.1	Rhity0006719.t1
	Scyliorhinus torazame		Scyto0006891.t1
	Chiloscyllium punctatum		Chipu0006102.t1
	Callorhinchus milii	XP_007882936.1	XM_007884745.1_SRX154861
NR5A2	*Homo sapiens*	NP_995582.1	NM_205860.2
	Mus musculus	NP_109601.1	NM_030676.3
	Gallus gallus	NP_990409.1	NM_205078.1
	Anolis carolinensis	XP_008113229.1	XM_008115022.2
	Xenopus tropicalis	NP_001011100.1	NM_001011100.1
	Danio rerio	NP_001300658.1	NM_001313729.1
	Lepisosteus oculatus	XP_006634936.2	XM_006634873.2
	Leucoraja erinacea		SRX2488466
	Rhincodon typus	XP_020386768.1	XM_020531179.1_ SRX657786
	Scyliorhinus torazame		Scyto0010544.t1
	Chiloscyllium punctatum		Chipu0016931.t1_Chipu0010474.t1
	Callorhinchus milii	XP_007885508.1	XM_007887317.1
NR5A5	*Danio rerio*	NP_999944.2	NM_214779.2
	Lepisosteus oculatus	XP_006631632.1	XM_006631569.2
	Leucoraja erinacea		SRX2488466
	Rhincodon typus	XP_020372121.1	XM_020516532.1_SRX657786_Rhity0026103.t1
	Scyliorhinus torazame		Scyto0016411.t1
	Chiloscyllium punctatum		Chipu0007707.t1
	Callorhinchus milii	XP_007884535.1	XM_007886344.1_SRX154858
NR5A	*Ciona intestinalis*	NP_001071721.1	NM_001078253.1
	Branchiostoma floridae	XP_002596353.1	XM_002596307.1
	Branchiostoma lanceolatum	BL14032	
NR5B	*Branchiostoma floridae*	g46332.t1	
	Branchiostoma lanceolatum	BL01871	
NR6A1	*Homo sapiens*	NP_201591.2	NM_033334.3
	Mus musculus	NP_034394.1	NM_010264.4
	Gallus gallus	XP_015135068.1	XM_015279582.1
	Anolis carolinensis	XP_016854129.1	XM_016998640.1
	Xenopus tropicalis	NP_001008005.1	NM_001008004.1
	Danio rerio A	NP_571331.2	NM_131256.2
	Danio rerio B	NP_001028892.1	NM_001033720.1
	Lepisosteus oculatus	XP_015222619.1	XM_015367133.1
	Leucoraja erinacea		SRX2488466
	Scyliorhinus canicula		SRX036537
	Rhincodon typus	XP_020372124.1	XM_020516535.1_SRX657786
	Chiloscyllium punctatum		Chipu0006101.t1
	Callorhinchus milii	XP_007905518.1	XM_007907327.1
	*Callorhinchus milii*like	XP_007885551.1	XM_007887360.1
NR6A2	*Branchiostoma floridae*	XP_002608117.1	XM_002608071.1
	Branchiostoma lanceolatum	BL05739	
	Saccoglossus kowalevskii	XP_006814372.1	XM_006814309.1

	Strongylocentrotus purpuratus	AAY41406.1	DQ018372.1
NR8A1	*Branchiostoma floridae*	XP_002590068.1	XM_002590022.1_DRX029482
	Branchiostoma lanceolatum	BL24556	
	Saccoglossus kowalevskii	XP_002737359.1	XM_002737313.1
	Strongylocentrotus purpuratus	XP_011665772.1	XM_011667470.1
2DBD (NR1X)	*Branchiostoma floridae*	g30470.t1	
	Branchiostoma lanceolatum	BL15669_BL15670	
	Saccoglossus kowalevskii b	XP_006815259.1	XM_006815196.1
	Saccoglossus kowalevskii g	XP_006817041.1	XM_006816978.1
	Saccoglossus kowalevskii d	XP_006818146.1	XM_006818083.1
	Strongylocentrotus purpuratus b	XP_011681136.1	XM_011682834.1
	Strongylocentrotus purpuratus d	XP_011661943.1	XM_011663641.1

Table S2.2. MixS descriptors and accession numbers of tissue sample, raw data and assemblies of *Hydrolagus affinis*.

Item	**Description**
investigation_type	Eukaryote
project_name	*Hydrolagus affinis* genome
lat_lon	
geo_loc_name	
tissues	
collection_date	
biome	Sea water (ENVO:00002149)
feature	Ocean (ENVO:00000015)
material	Sea water (ENVO:00002149)
env_package	Water
seq_meth	Illumina HiSeq X Ten
Collector	
Maturity	
Datasets Generated	**Acession Numbers of NCBI**
Genome raw data	in process..
Genome Assembly	in process..
COI sequence	MN701085

Table S2.3. List of accession numbers of *NR2B* and *NR2F* genes in reptiles and birds.

Species			**NR**	**Accession number**	
Reptiles	*Gekko japonicus*	Schlegel's Japanese gecko	NR2B1	XP_015272804.1	XM_015417318.1
	Gekko japonicus	Schlegel's Japanese gecko	NR2B2	XP_015265750.1	XM_015410264.1
	Gekko japonicus	Schlegel's Japanese gecko	NR2B3	XP_015261691.1	XM_015406205.1
	Gekko japonicus	Schlegel's Japanese gecko	NR2F1	XP_015282619.1	XM_015427133.1
	Python bivittatus	Burmese python	NR2F2	XP_015746728.1	XM_015891242.1
	Gekko japonicus	Schlegel's Japanese gecko	NR2F5	XP_015270008.1	XM_015414522.1
	Gekko japonicus	Schlegel's Japanese gecko	NR2F6	XP_015262308.1	XM_015406822.1
Birds	*Haliaeetus leucocephalus*	Bald eagle	NR2B2	XP_010568205.1	XM_010569903.1
	Aquila chrysaetos canadensis	Golden eagle	NR2B2	XP_011599174.1	XM_011600872.1
	Lonchura striata domestica	Bengalese finch	NR2B2	XP_021403355.1	XM_021547680.1
	Apteryx rowi	Okarito kiwi	NR2B2	XP_025914020.1	XM_026058235.1

Table S2.4. GenomeScope2 (k-mer 19, 21, 23, 25, 27, 29 and 31), Kmergenie and Sga Preqc statistics of *Hydrolagus affinis* WGS reads.

GenomeScope2	**K = 19**		**K = 21**		**K = 23**	
Property	*min*	*max*	*min*	*max*	*min*	*max*
Homozygous (%)	99.528	99.5823	99.5413	99.5809	99.5519	99.5886
Heterozygosity (%)	0.417664	0.47198	0.419083	0.458696	0.411448	0.448061
Genome Haploid Length (bp)	1,078,650,590	1,080,878,402	1,098,032,621	1,100,040,460	1,111,900,584	1,113,863,822
Genome Repeat Length (bp)	394,767,261	395,582,601	357,672,455	358,326,487	339,275,389	339,874,434
Genome Unique Length (bp)	683,883,329	685,295,801	740,360,166	741,713,972	772,625,195	773,989,388
Model Fit (%)	68.4603	98.5673	70.9466	98.6549	72.7316	98.6566
Read Error Rate (%)	0.325752	0.325752	0.350625	0.350625	0.361024	0.361024

GenomeScope2	**k = 25**		**K = 27**		**K = 29**	
Property	*min*	*min*	*max*	*min*	*max*	*max*
Homozygous (%)	99.5635	99.5983	99.5829	99.6106	99.5929	99.6198
Heterozygosity (%)	0.401737	0.436541	0.389377	0.41714	0.380195	0.407128
Genome Haploid Length (bp)	1,122,661,665	1,124,610,356	1,131,077,312	1,132,967,035	1,137,826,984	1,139,717,115
Genome Repeat Length (bp)	324,714,705	325,278,337	312,409,048	312,930,998	300,832,894	301,332,631
Genome Unique Length (bp)	797,946,960	799,332,019	818,668,264	820,036,037	836,994,090	838,384,484
Model Fit (%)	74.2192	98.8781	75.5402	98.9148	76.6726	98.8856
Read Error Rate (%)	0.363953	0.363953	0.364027	0.364027	0.361135	0.361135

GenomeScope2	**K =31**	
Property	*min*	*max*
Homozygous (%)	99.602	99.6106
Heterozygosity (%)	0.371528	0.41714
Genome Haploid Length (bp)	1,143,135,686	1,132,967,035
Genome Repeat Length (bp)	290,319,184	312,930,998
Genome Unique Length (bp)	852,816,502	820,036,037
Model Fit (%)	77.6871	98.9148
Read Error Rate (%)	0.357174	0.364027

Kmergenie Property	
Genome Haploid Length (bp)	1,268,245,615
Predicted Best Kmer	101
SGA Preqc Property	
Genome Haploid Length (bp)	1,227,700,000
Pcr Duplication Proportion	0.079
Mean quality score	34,25-36
Fragment Size	400

GenomeScope2 Reports:

kmer 19 report: http://qb.cshl.edu/genomescope/genomescope2.0/analysis.php?code=BEPE81d3LajeTCElQXCh

Kmer 21 report: http://qb.cshl.edu/genomescope/genomescope2.0/analysis.php?code=vmjCBTUZK7u1k9BZF7Rp

Kmer 23 report: http://qb.cshl.edu/genomescope/genomescope2.0/analysis.php?code=CG01OZEF0ARBvoCXQArk

Kmer 25 report: http://qb.cshl.edu/genomescope/genomescope2.0/analysis.php?code=XqyWNMYNYjfeWGUzhPaj

Kmer 27 report: http://qb.cshl.edu/genomescope/genomescope2.0/analysis.php?code=0eeUgOLWb7aLLAyv5czw

Kmer 29 report: http://qb.cshl.edu/genomescope/genomescope2.0/analysis.php?code=TzBtKxSFld2RP3UawZEL

Kmer 31 report: http://qb.cshl.edu/genomescope/genomescope2.0/analysis.php?code=cfzjbnH9fDWdWRRIf3br

Table S2.5. Genome assembly versions of *Hydrolagus affinis*.

Assembly_name	Version	Status	Method	K-mer	Sequences (>=0 bp)	Sequences* (>=500 bp)	Sequences (>=1000 bp)	Sequences (>=5000 bp)	Sequences (>=10000 bp)	Sequences (>=25000 bp)	Sequences (>=50000 bp)	Total length (>= 0bp)
Haf_160k	1st	Contig	w2rap	160	259849	259849	143986	49313	29198	9152	1982	1119238096
Haf_168k	1st	Contig	w2rap	168	255567	255567	144552	49418	29123	9114	2007	1116603524
Haf_172k	1st	Contig	w2rap	172	252783	252783	144535	49432	29087	9124	1961	1113541132
Haf_180k	1st	Contig	w2rap	180	244979	244979	143770	49462	28930	8980	1907	1102264526
Haf_200k	1st	Contig	w2rap	200	223532	223532	138397	49054	28167	8385	1723	1054317262
Haf_Soap_23k	1st	Contig	SOAPdenovo2	23	2567985	376860	109564	319	7	0	0	659181621
Haf_Soap_41k	1st	Contig	SOAPdenovo2	41	1604546	474104	243229	9489	571	1	0	864037833
Haf_Soap_51k	1st	Contig	SOAPdenovo2	51	1344093	442944	249243	20168	2208	19	0	920594278
Haf_Soap_61k	1st	Contig	SOAPdenovo2	61	1144031	414990	239535	29458	4970	86	0	947829957
Haf_Soap_65k	1st	Contig	SOAPdenovo2	65	1073253	413308	238907	30779	5541	108	1	949727678
Haf_Soap_71k	1st	Contig	SOAPdenovo2	71	1029176	435013	248597	26267	3893	60	0	933533863
Haf_Soap_75k	1st	Contig	SOAPdenovo2	75	1062461	489467	262632	15096	1277	11	0	900064179
Haf_Soap_81k	1st	Contig	SOAPdenovo2	81	1351907	534577	182681	439	0	0	0	764521376
Haf_160k	1st	Scaffold	w2rap	160	1502462	252172	136309	41643	26074	10202	3017	1409199803
Haf_168k	1st	Scaffold	w2rap	168	1270677	248112	137097	41968	26113	10140	2983	1365942086
Haf_172k	1st	Scaffold	w2rap	172	1164641	245464	137216	42118	26172	10125	2935	1343448764
Haf_180k	1st	Scaffold	w2rap	180	1005627	237927	136718	42412	26215	10075	2809	1303642952
Haf_200k	1st	Scaffold	w2rap	200	820316	217019	131884	42542	25848	9389	2522	1228654866
Haf_Soap_23k	1st	Scaffold	SOAPdenovo2	23	657372	183136	143347	57255	25772	4340	572	1024367589
Haf_Soap_41k	1st	Scaffold	SOAPdenovo2	41	921334	186020	121873	50148	28578	8323	1725	1156675419
Haf_Soap_51k	1st	Scaffold	SOAPdenovo2	51	1698662	189473	121280	48443	28584	9218	2188	1272804801
Haf_Soap_61k	1st	Scaffold	SOAPdenovo2	61	1891015	197182	123271	47403	27964	9353	2355	1315680082
Haf_Soap_65k	1st	Scaffold	SOAPdenovo2	65	1843129	203531	125707	46965	27701	9138	2271	1308307371
Haf_Soap_71k	1st	Scaffold	SOAPdenovo2	71	1643084	187131	118356	45568	27362	9248	2319	1260148876
Haf_Soap_75k	1st	Scaffold	SOAPdenovo2	75	1448035	189175	120565	45841	26814	8398	1986	1197723998
Haf_Soap_81k	1st	Scaffold	SOAPdenovo2	81	1029369	208998	142375	47541	22426	4492	641	1021896317
Haf_172k_RS_41	2nd	Re-Scaffold	SOAPdenovo2	41	1096396	225160	131553	42666	26913	10625	3268	1342611353
Haf_172k_RS_51	2nd	Re-Scaffold	SOAPdenovo2	51	1081597	221435	131499	42722	26949	10709	3317	1341509332
Haf_172k_RS_61	2nd	Re-Scaffold	SOAPdenovo2	61	1068479	217679	131161	42738	26982	10769	3358	1340312334
Haf_172k_RS_65	2nd	Re-Scaffold	SOAPdenovo2	65	1063073	216096	130825	42747	26959	10802	3377	1339740971
Haf_172k_RS_71	2nd	Re-Scaffold	SOAPdenovo2	71	1055100	213581	130447	42742	26979	10853	3414	1338731327
Haf_172k_RS_75	2nd	Re-Scaffold	SOAPdenovo2	75	1050112	212081	130270	42756	26946	10869	3438	1338011799
Haf_172k_RS_80	2nd	Re-Scaffold	SOAPdenovo2	80	1044307	210175	130030	42852	26940	10895	3455	1337140803

* All statistics (GC (%) ; N50 ; N75 ; L50 ; L75 ;N's per 100 kbp) are based on contigs/scaffolds of size ≥500 bp.
** (C : Complete Buscos [S: Complete and single copy, %;D:Complete and duplicated, %], F: Fragmented, %, M: Missing, %, n: Number of sequences in Database)

Continued

Continuation

Assembly_name	Total length* (>=500 bp)	Total length (>=1000 bp)	Total length (>=5000 bp)	Total length (>=10000 bp)	Total length (>=50000 bp)	Largest sequence	GC (%)*	N50*	N75*	L50*	L75*	# N's per 100 kbp*
Haf_160k	1119238096	1039091495	844281625	700434615	140983318	231031	42.39	15789	5145	18003	48357	0.00
Haf_168k	1116603524	1039317385	843978932	699034572	142333634	222527	42.35	15824	5192	17930	48138	0.00
Haf_172k	1113541132	1037869609	842564677	697395503	139121198	222549	42.33	15857	5211	17918	47981	0.00
Haf_180k	1102264526	1031274959	836264510	689738889	135422177	230894	42.27	15664	5279	17867	47598	0.00
Haf_200k	1054317262	994841628	805301900	656047709	120253086	216970	42.11	15245	5440	17729	46260	0.00
Haf_Soap_23k	351313887	167415627	1948603	87831	0	17860	40.57	542	233	334128	793109	117.80
Haf_Soap_41k	673578505	510167778	63678250	7070896	0	30324	41.68	1300	568	174663	426128	45.65
Haf_Soap_51k	754940472	618097046	146292159	28344148	0	37207	41.87	1788	713	131070	334925	34.07
Haf_Soap_61k	801493323	678039781	230803487	66598273	0	47046	41.88	2267	846	103388	275231	27.51
Haf_Soap_65k	812000668	689323979	244740684	75112443	67561	67561	41.84	2369	890	98994	263273	25.00
Haf_Soap_71k	799230211	667674711	200625554	51625824	0	42907	41.71	2134	859	110436	283627	21.09
Haf_Soap_75k	757888271	597539645	105357075	15844310	0	37113	41.54	1622	737	146526	352954	20.39
Haf_Soap_81k	5317330 14	286362288	2555304	0	0	9795	41.01	768	423	292481	625227	21.01
Haf_160k	1120006661	1039860060	845078653	734498659	232141449	278637	42.39	19681	5166	13779	40645	68.58
Haf_168k	1117349324	1040063185	844748887	732321086	229537467	318369	42.35	19639	5224	13830	40650	66.75
Haf_172k	1114273232	1038601709	843320906	730306829	225441116	397232	42.33	19535	5253	13894	40630	65.70
Haf_180k	1102969926	1031980359	836979465	722160596	214323276	397186	42.27	19279	5335	14003	40520	63.95
Haf_200k	1054968962	995493328	805958163	687338799	188954037	272171	42.11	18367	5523	14139	39734	61.77
Haf_Soap_23k	947809106	919004458	704267940	481459391	38149189	243617	40.66	10200	4870	25025	58591	28406.15
Haf_Soap_41k	1024141092	978834571	816259513	661980597	123261316	405902	41.79	15924	6485	16694	41710	13493.18
Haf_Soap_51k	1052791255	1004904691	844530966	702602706	159598069	461863	41.98	17590	6835	15311	39074	10269.34
Haf_Soap_61k	1060347970	1008878700	844012538	705122151	171544787	253442	41.99	17989	6659	14965	38979	8189.88
Haf_Soap_65k	1055299778	1001229481	830614456	693288718	165079161	260641	41.95	17521	6327	15218	40015	7519.34
Haf_Soap_71k	1030970297	983290314	824347248	694373649	170321009	252424	41.82	18512	6860	14244	36866	6976.74
Haf_Soap_75k	995885512	948325913	784622794	648897148	141239925	280270	41.65	16765	6278	15169	39130	6947.57
Haf_Soap_81k	885114182	838309955	620383038	442987618	42612844	183560	41.15	10013	4036	22383	57204	9443.64
Haf_172k_RS_41	1130479926	1064336959	883596718	771495032	255788860	397232	42.39	21429	6448	12963	36369	173.32
Haf_172k_RS_51	1133300233	1069645946	888686307	776503458	260072531	397232	42.39	21714	6569	12840	35967	196.72
Haf_172k_RS_61	1135425695	1074116782	893095601	780996321	264903502	397232	42.39	21985	6703	12717	35570	220.49
Haf_172k_RS_65	1136180663	1075726885	895215513	782837094	267146079	397232	42.39	22102	6775	12639	35349	233.52
Haf_172k_RS_71	1137079911	1078110237	898002668	785826922	270324035	397232	42.39	22290	6871	12553	35036	253.40
Haf_172k_RS_75	1137591126	1079531949	899604833	787061219	272306092	397232	42.39	22399	6922	12496	34869	267.47
Haf_172k_RS_80	1138110018	1081222756	901916880	788716903	273960898	397232	42.39	22541	6988	12433	34676	282.32

Continued

* All statistics (GC (%) ; N50 ; N75 ; L50 ; L75 ;N's per 100 kbp) are based on contigs/scaffolds of size ≥500 bp.
** (C : Complete Buscos [S: Complete and single copy, %;D:Complete and duplicated, %], F: Fragmented, %, M: Missing, %, n: Number of sequences in Database)

Continuation

Assembly_name	Buscos – Eukariota**	Buscos – Metazoa**	Buscos – Vetebrata**
Haf_160k	-	-	-
Haf_168k	-	-	-
Haf_172k	-	-	-
Haf_180k	-	-	-
Haf_200k	-	-	-
Haf_Soap_23k	-	-	-
Haf_Soap_41k	-	-	-
Haf_Soap_51k	-	-	-
Haf_Soap_61k	-	-	-
Haf_Soap_65k	-	-	-
Haf_Soap_71k	-	-	-
Haf_Soap_75k	-	-	-
Haf_Soap_81k	-	-	-
Haf_160k	C:55.8%[S:55.1%,D:0.7%],F:19.8%,M:24.4%,n:303	C:65.1%[S:64.2%,D:0.9%],F:17.5%,M:17.4%,n:978	C:59.7%[S:58.9%,D:0.8%],F:25.6%,M:14.7%,n:2586
Haf_168k	C:55.8%[S:55.1%,D:0.7%],F:20.1%,M:24.1%,n:303	C:64.2%[S:63.3%,D:0.9%],F:17.3%,M:17.5%,n:978	C:59.8%[S:59.0%,D:0.8%],F:25.7%,M:14.5%,n:2586
Haf_172k	C:55.5%[S:54.5%,D:1.0%],F:20.8%,M:23.7%,n:303	C:64.1%[S:63.1%,D:1.0%],F:19.0%,M:16.9%,n:978	C:60.0%[S:59.2%,D:0.8%],F:26.1%,M:13.9%,n:2586
Haf_180k	C:52.8%[S:52.1%,D:0.7%],F:21.8%,M:25.4%,n:303	C:63.3%[S:62.7%,D:0.6%],F:18.9%,M:17.8%,n:978	C:59.0%[S:58.3%,D:0.7%],F:26.8%,M:14.2%,n:2586
Haf_200k	C:51.8%[S:50.8%,D:1.0%],F:21.8%,M:26.4%,n:303	C:59.2%[S:58.5%,D:0.7%],F:22.7%,M:18.1%,n:978	C:56.3%[S:55.5%,D:0.8%],F:28.2%,M:15.5%,n:2586
Haf_Soap_23k	C:36.9%[S:36.6%,D:0.3%],F:28.1%,M:35.0%,n:303	C:45.6%[S:45.2%,D:0.4%],F:28.5%,M:25.9%,n:978	C:36.3%[S:36.0%,D:0.3%],F:36.8%,M:26.9%,n:2586
Haf_Soap_41k	C:48.9%[S:47.9%,D:1.0%],F:25.1%,M:26.0%,n:303	C:59.2%[S:58.8%,D:0.4%],F:22.0%,M:18.8%,n:978	C:50.8%[S:50.3%,D:0.5%],F:30.3%,M:18.9%,n:2586
Haf_Soap_51k	C:52.5%[S:51.8%,D:0.7%],F:22.1%,M:25.4%,n:303	C:59.8%[S:59.2%,D:0.6%],F:22.7%,M:17.5%,n:978	C:52.2%[S:51.7%,D:0.5%],F:30.5%,M:17.3%,n:2586
Haf_Soap_61k	C:50.1%[S:48.8%,D:1.3%],F:25.4%,M:24.5%,n:303	C:60.0%[S:59.4%,D:0.6%],F:21.8%,M:18.2%,n:978	C:53.3%[S:52.8%,D:0.5%],F:29.4%,M:17.3%,n:2586
Haf_Soap_65k	C:48.5%[S:47.5%,D:1.0%],F:26.4%,M:25.1%,n:303	C:57.7%[S:57.4%,D:0.3%],F:22.8%,M:19.5%,n:978	C:51.8%[S:51.3%,D:0.5%],F:30.9%,M:17.3%,n:2586
Haf_Soap_71k	C:50.5%[S:49.5%,D:1.0%],F:23.8%,M:25.7%,n:303	C:56.5%[S:56.1%,D:0.4%],F:23.8%,M:19.7%,n:978	C:50.8%[S:50.2%,D:0.6%],F:30.9%,M:18.3%,n:2586
Haf_Soap_75k	C:44.6%[S:43.9%,D:0.7%],F:25.4%,M:30.0%,n:303	C:50.8%[S:50.5%,D:0.3%],F:25.3%,M:23.9%,n:978	C:44.9%[S:44.6%,D:0.3%],F:32.6%,M:22.5%,n:2586
Haf_Soap_81k	C:24.4%[S:24.4%,D:0.0%],F:28.7%,M:46.9%,n:303	C:31.0%[S:31.0%,D:0.0%],F:28.7%,M:40.3%,n:978	C:19.9%[S:19.7%,D:0.2%],F:36.4%,M:43.7%,n:2586
Haf_172k_RS_41	C:56.8%[S:55.8%,D:1.0%],F:19.8%,M:23.4%,n:303	C:67.2%[S:66.3%,D:0.9%],F:16.6%,M:16.2%,n:978	C:62.9%[S:62.1%,D:0.8%],F:24.2%,M:12.9%,n:2586
Haf_172k_RS_51	C:56.4%[S:55.4%,D:1.0%],F:20.8%,M:22.8%,n:303	C:66.9%[S:66.0%,D:0.9%],F:17.1%,M:16.0%,n:978	C:62.8%[S:61.9%,D:0.9%],F:24.3%,M:12.9%,n:2586
Haf_172k_RS_61	C:57.4%[S:56.4%,D:1.0%],F:18.8%,M:23.8%,n:303	C:66.2%[S:65.3%,D:0.9%],F:17.1%,M:16.7%,n:978	C:63.0%[S:62.2%,D:0.8%],F:24.3%,M:12.7%,n:2586
Haf_172k_RS_65	C:57.1%[S:55.8%,D:1.3%],F:18.8%,M:24.1%,n:303	C:66.1%[S:65.2%,D:0.9%],F:17.1%,M:16.8%,n:978	C:63.4%[S:62.6%,D:0.8%],F:24.0%,M:12.6%,n:2586
Haf_172k_RS_71	C:58.4%[S:57.1%,D:1.3%],F:19.5%,M:22.1%,n:303	C:67.2%[S:66.3%,D:0.9%],F:16.8%,M:16.0%,n:978	C:63.7%[S:62.9%,D:0.8%],F:23.9%,M:12.4%,n:2586
Haf_172k_RS_75	C:57.8%[S:56.8%,D:1.0%],F:19.5%,M:22.7%,n:303	C:67.1%[S:66.1%,D:1.0%],F:16.4%,M:16.5%,n:978	C:63.0%[S:62.3%,D:0.7%],F:24.4%,M:12.6%,n:2586
Haf_172k_RS_80	C:57.1%[S:55.8%,D:1.3%],F:19.5%,M:23.4%,n:303	C:67.1%[S:66.1%,D:1.0%],F:16.7%,M:16.2%,n:978	C:63.1%[S:62.3%,D:0.8%],F:24.6%,M:12.3%,n:2586

* All statistics (GC (%) ; N50 ; N75 ; L50 ; L75 ;N's per 100 kbp) are based on contigs/scaffolds of size ≥500 bp.
** (C : Complete Buscos [S: Complete and single copy, %;D:Complete and duplicated, %], F: Fragmented, %, M: Missing, %, n: Number of sequences in Database)

Table S2.6. List of nuclear receptors researched in *Hydrolagus affinis* genome.

Look for	Found	Scaffolds
NR1B3	NR1B1	scaffold57996_scaffold10182_C2295255_scaffold70118
	NR1B2	C1780265_scaffold14376_scaffold66502
	NR1B3	C2031625_C2225317
NR1D4	NR1D1	C2305947
	NR1D2	C2315167
	NR1D4	scaffold49229
NR2B2	NR2B1	C2253421_C2205335_ C2046755_scaffold63182
	NR2B2	not found
	NR2B3	C2328249
NR2F5	NR2F1	scaffold68454
	NR2F2	scaffold68454
	NR2F5	not found
	NR2F6	scaffold33262_scaffold63091
NR3B1	NR3B1	not found
	NR3B2	C2194775_C2182649_C2204871
	NR3B3	C2322617_C2300843
	NR3B4	not found

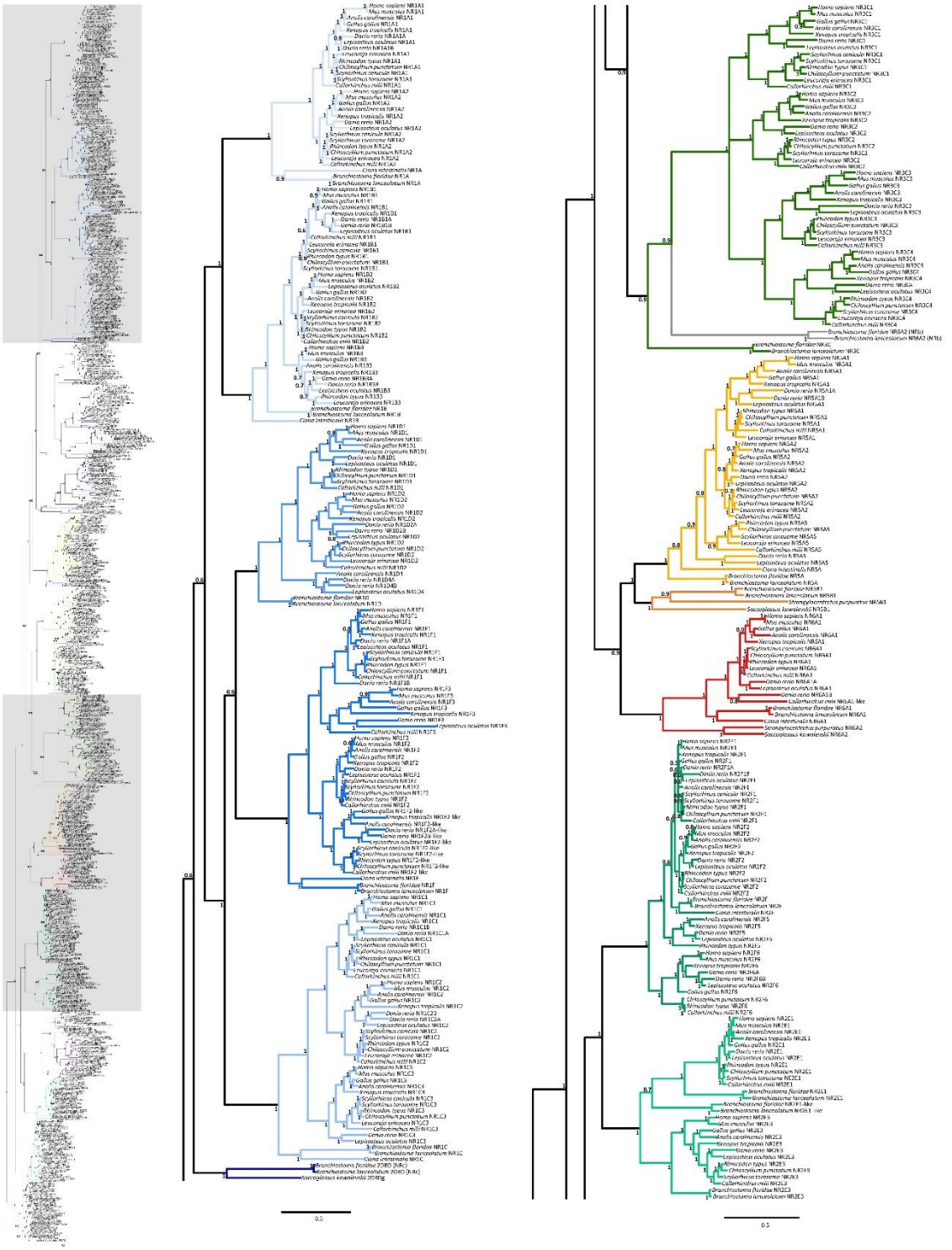


Figure S2.1 part 1 of 2

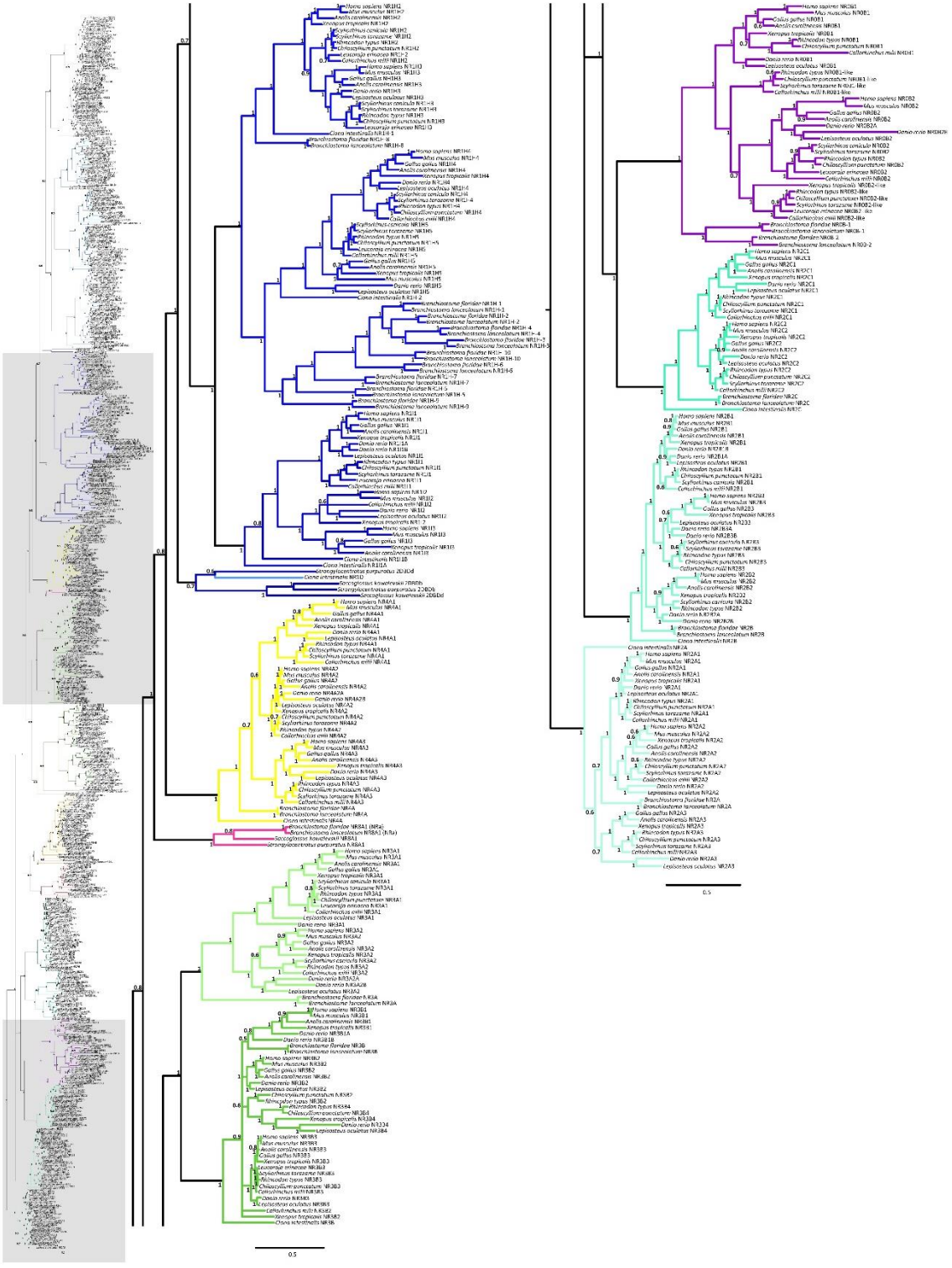


Figure S2.1 part 2 of 2

Figure S2.1. Bayesian phylogenetic analysis of NR amino acid sequences of human, mouse, chicken, green anole, Western clawed frog, zebrafish, spotted gar, whale shark, brownbanded bamboo shark, small-spotted catshark, cloudy catshark, little skate, elephant shark, sea squirt, and Florida and European lancelets; numbers at nodes indicate posterior probabilities.

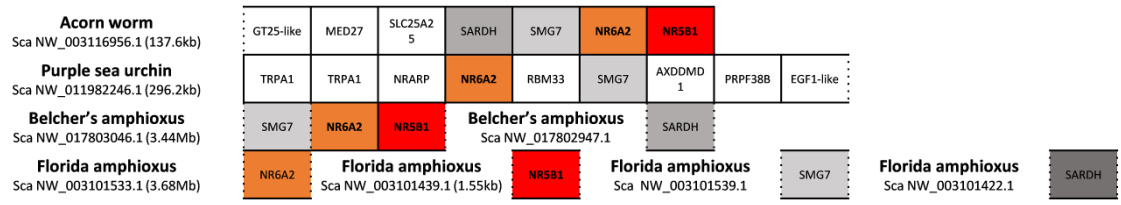


Figure S2.2. Synteny analysis of *NR6A2* in amphioxus species, acorn worm and purple sea urchin.

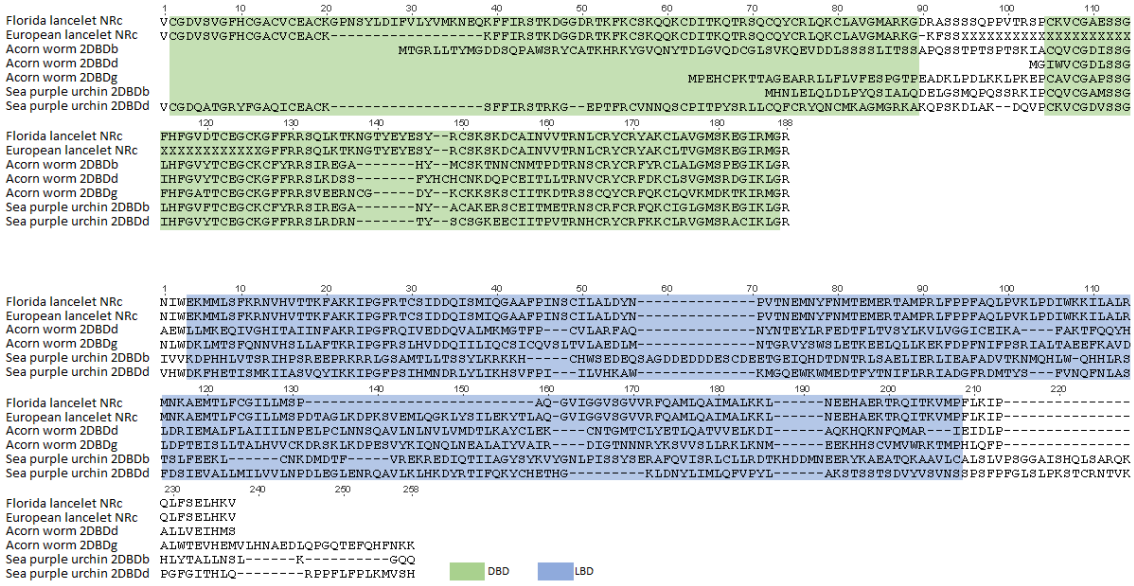


Figure S2.3. 2DBD receptor amino acid sequences alignment of Florida and European lancelets, acorn worm and purple sea urchin. MAFFT alignment using E-INS-i algorithm.

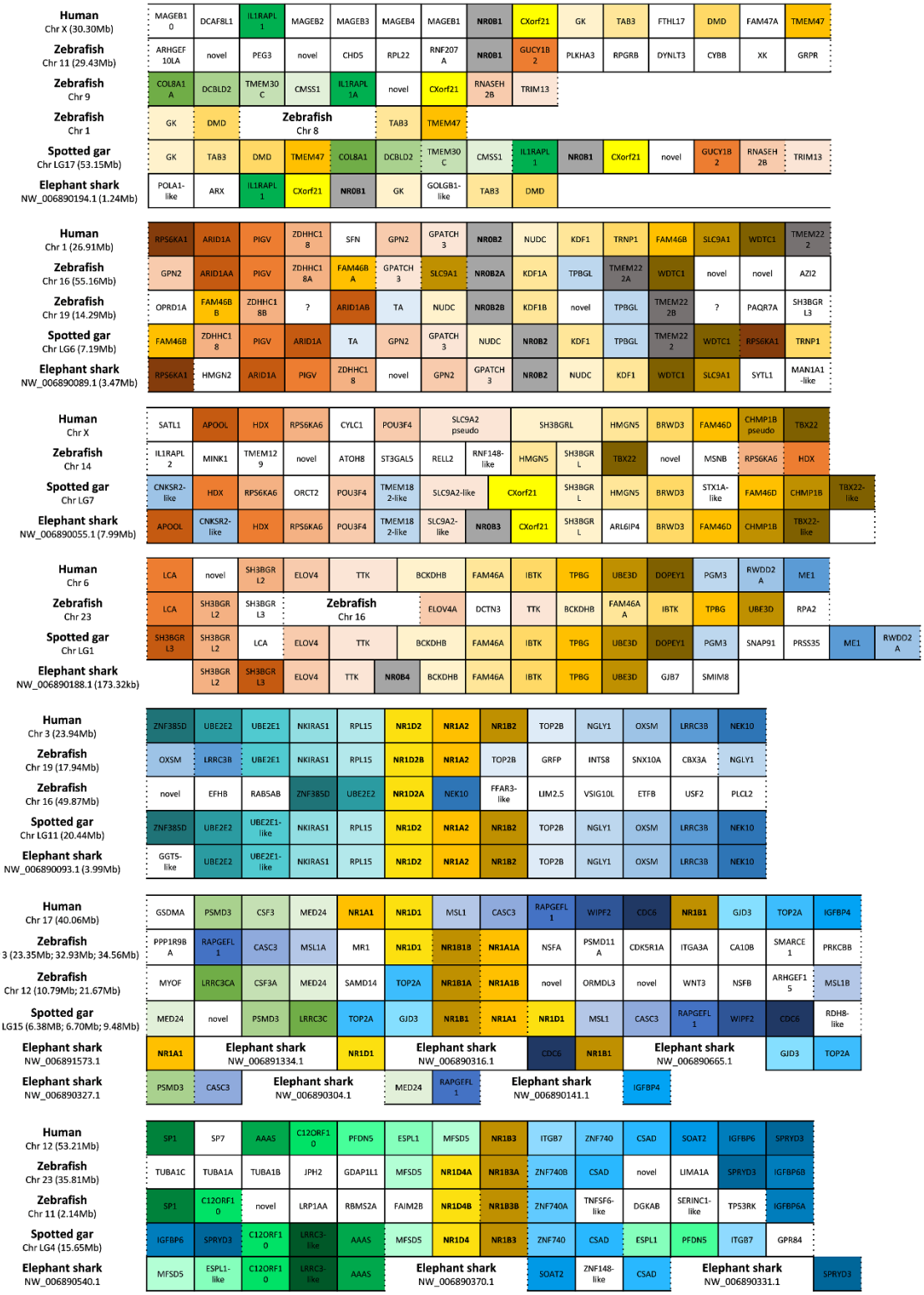


Figure S2.4 part 1 of 8

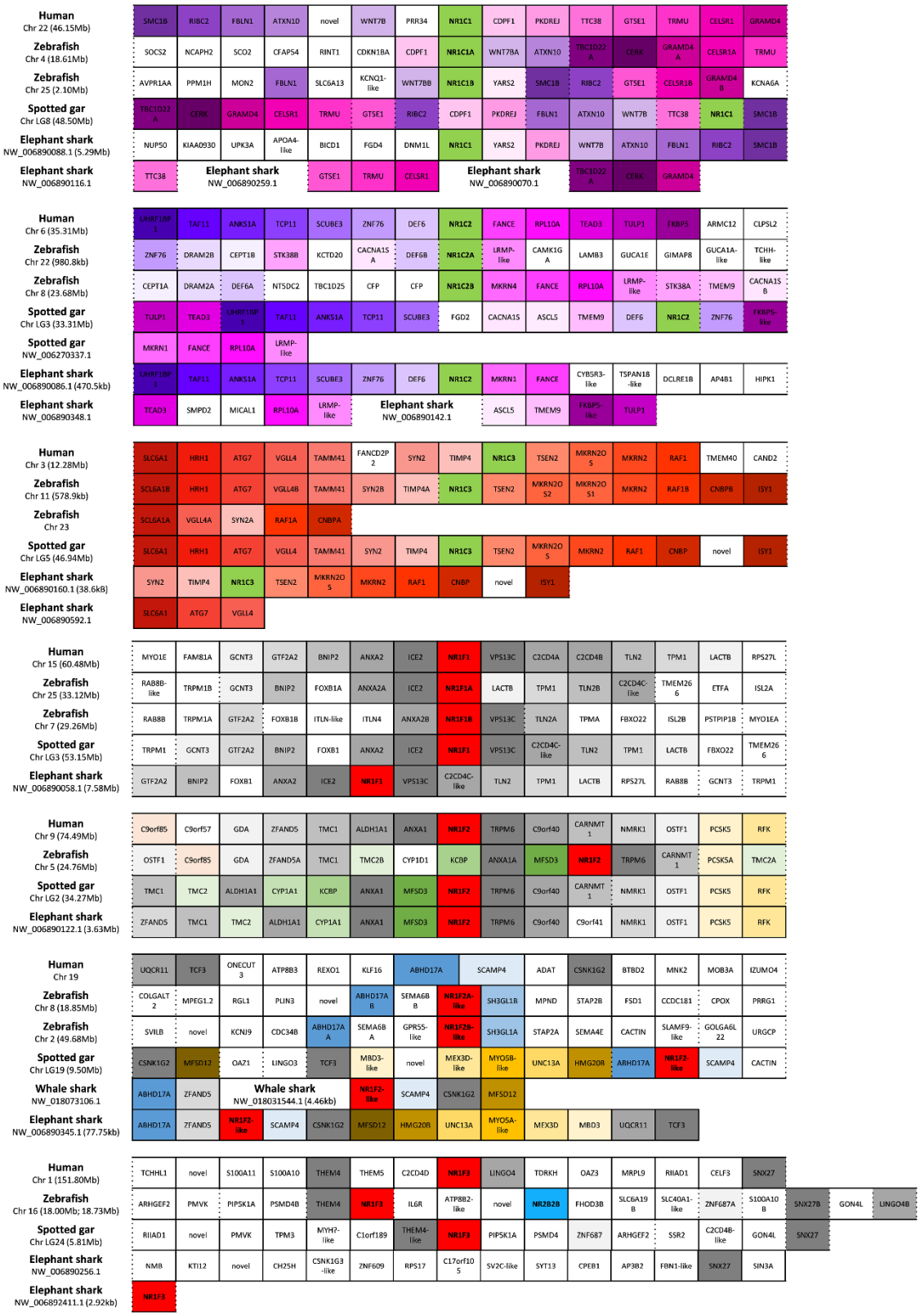


Figure S2.4 part 2 of 8

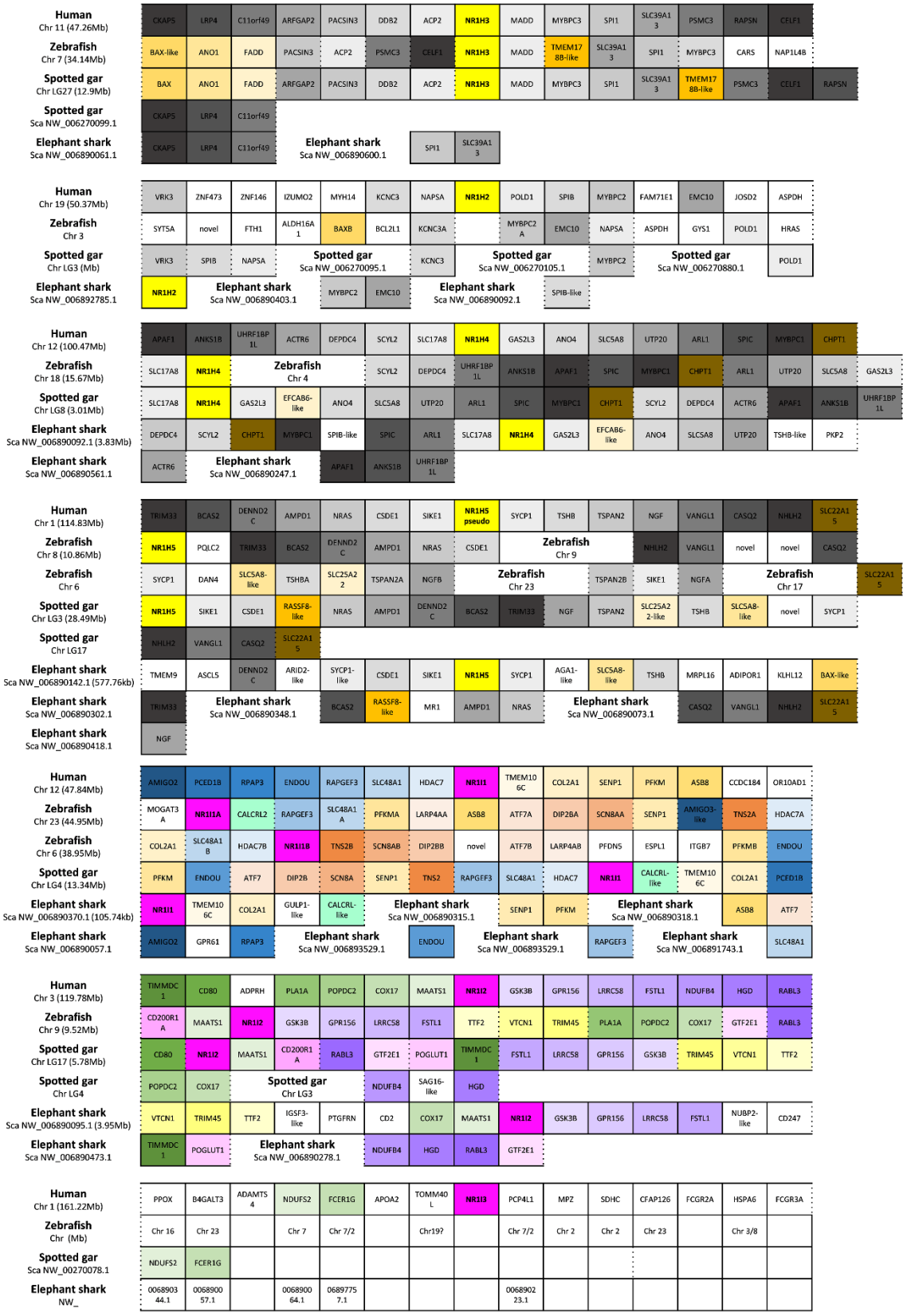


Figure S2.4 part 3 of 8

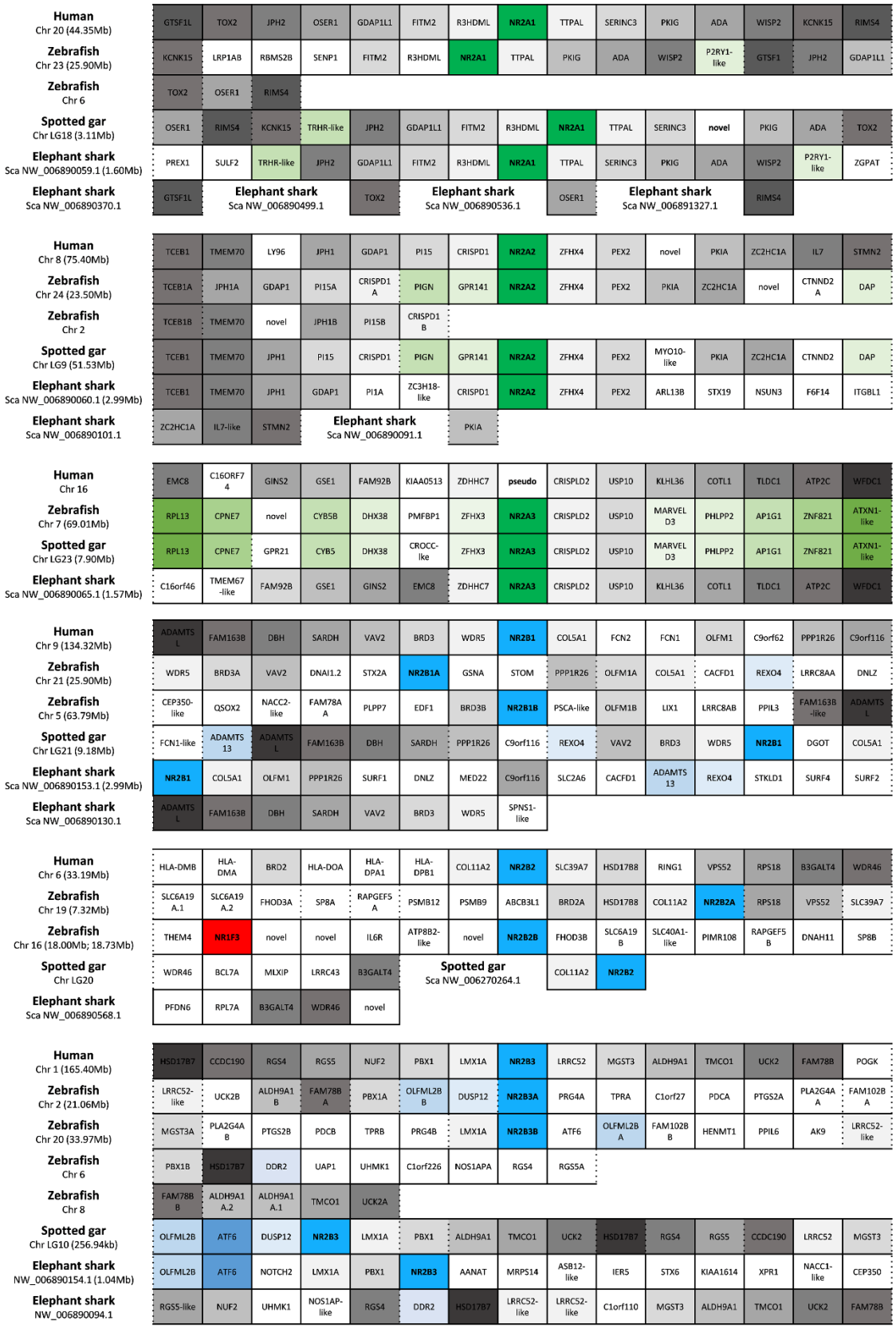


Figure S2.4 part 4 of 8

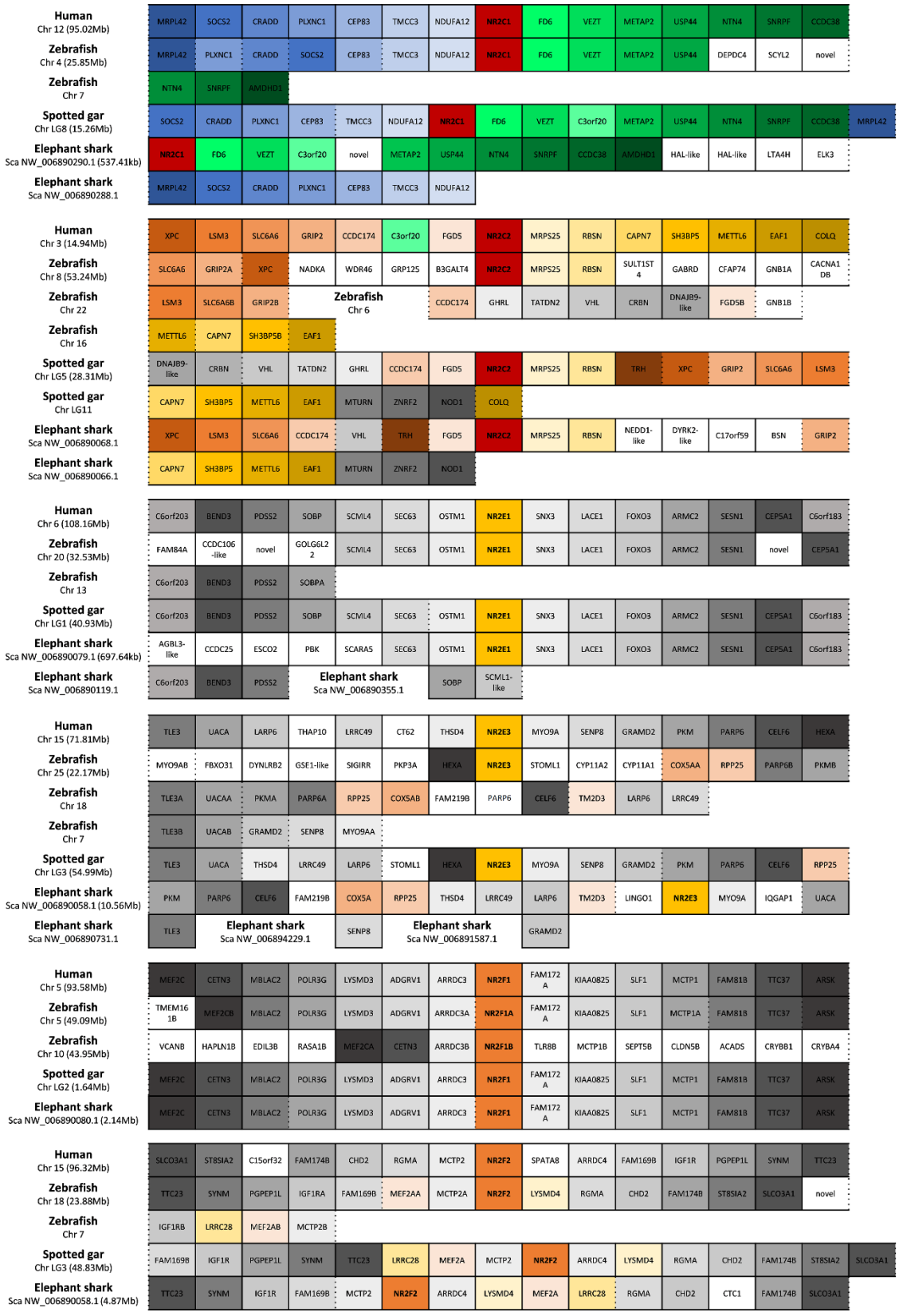


Figure S2.4 part 5 of 8

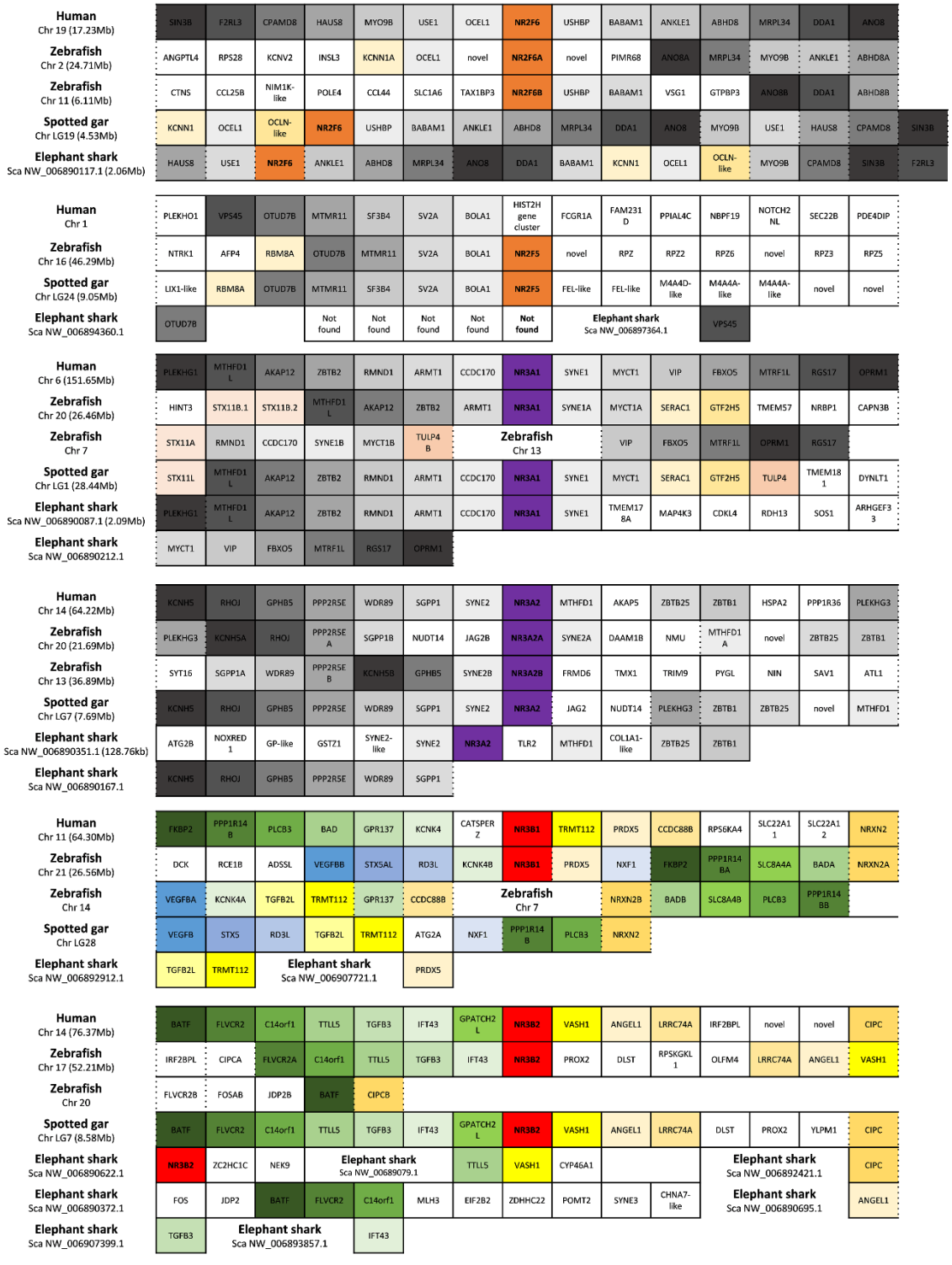


Figure S2.4 part 6 of 8

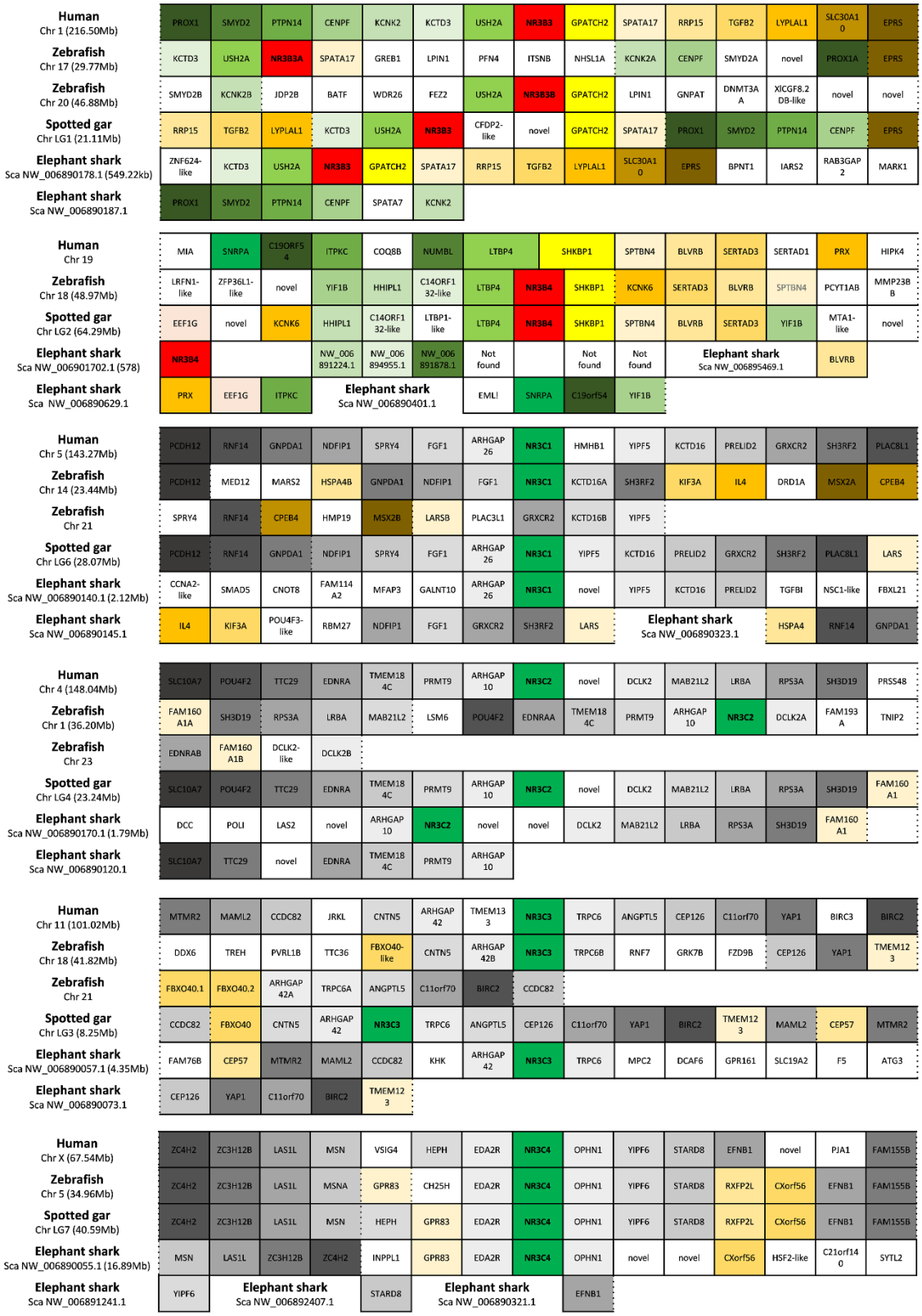


Figure S2.4 part 7 of 8

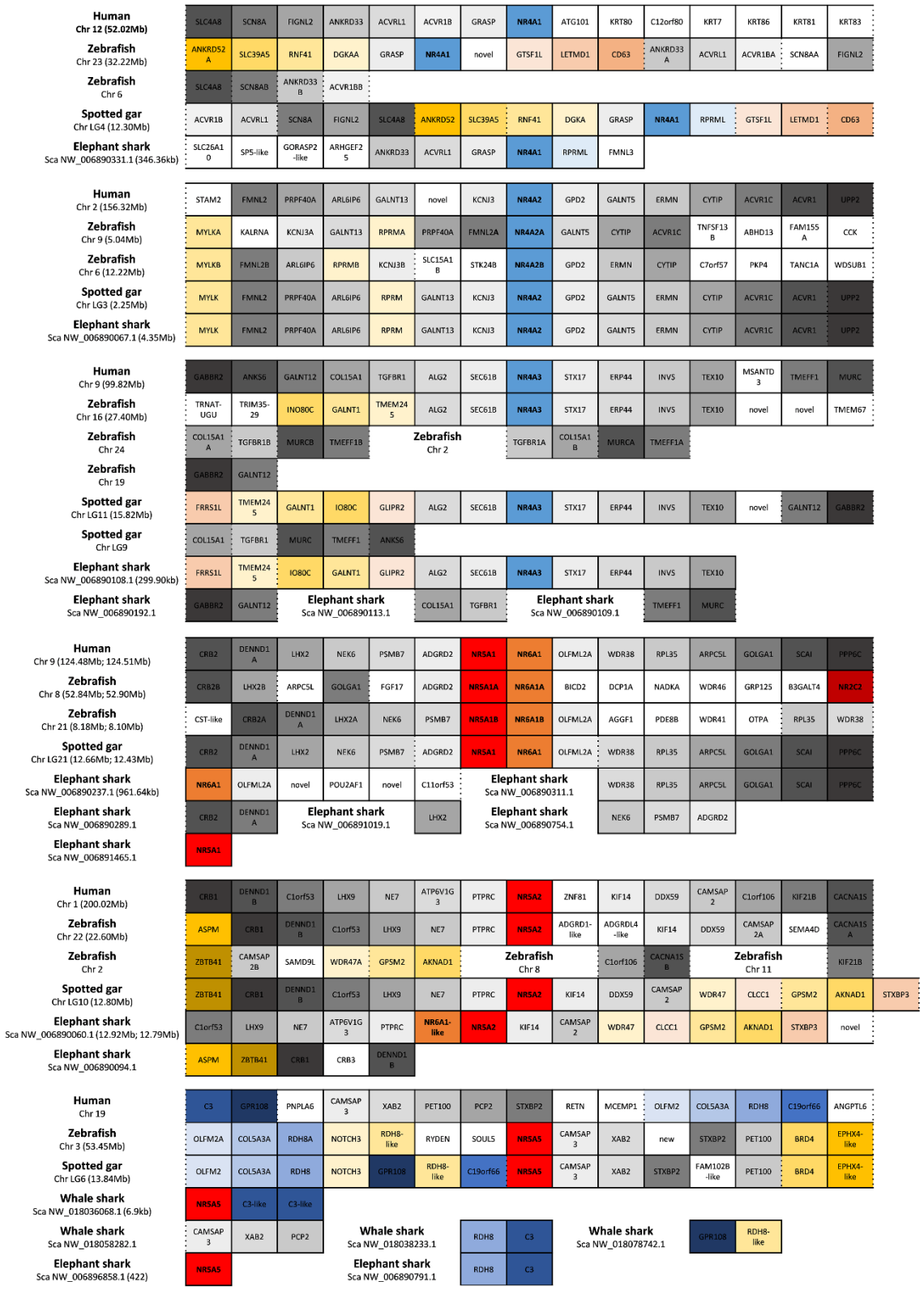


Figure S2.4 part 8 of 8

Figure S2.4. Synteny analyses of NRs *loci* in human, zebrafish, spotted gar and elephant shark.

CHAPTER 3 - *LXRα* and *LXRβ* Nuclear Receptors Evolved in the Common Ancestor of Gnathostomes

GBE

LXRα and *LXRβ* Nuclear Receptors Evolved in the Common Ancestor of Gnathostomes

Elza Fonseca[1,2,†], Raquel Ruivo[1,†], Mónica Lopes-Marques[1,3], Huixian Zhang[4], Miguel M. Santos[1,2], Byrappa Venkatesh[4,5], and L. Filipe C. Castro[1,2,*]

[1]CIIMAR/CIMAR – Interdisciplinary Centre of Marine and Environmental Research, U. Porto, Portugal

[2]Department of Biology, FCUP – Faculty of Sciences, U. Porto, Portugal

[3]ICBAS – Institute of Biomedical Sciences Abel Salazar – U. Porto, Portugal

[4]Institute of Molecular and Cell Biology, Agency for Science, Technology and Research, Biopolis, Singapore

[5]Department of Paediatrics, Yong Loo Lin School of Medicine, National University of Singapore, Singapore

[†]These authors contributed equally to this work.
*Corresponding author: E-mail: filipe.castro@ciimar.up.pt.

Accepted: December 29, 2016

3. *LXRα* and *LXRβ* Nuclear Receptors Evolved in the Common Ancestor of Gnathostomes

3. Abstract

Nuclear receptors (NRs) regulate numerous aspects of the endocrine system. They mediate endogenous and exogenous cues, ensuring a homeostatic control of development and metabolism. Gene duplication, loss and mutation have shaped the repertoire and function of NRs in metazoans. Here we examine the evolution of a pivotal orchestrator of cholesterol metabolism in vertebrates, the liver X receptors (LXRs). Previous studies suggested that *LXRα* and *LXRβ* genes emerged in the mammalian ancestor. However, we show through genome analysis and functional assay that *bona fide LXRα* and *LXRβ* orthologs are present in reptiles, coelacanth and chondrichthyans but not in cyclostomes. These findings show that LXR duplicated before gnathostome radiation, followed by asymmetric paralog loss in some lineages. We suggest that a tighter control of cholesterol levels in vertebrates was achieved through the exploitation of a wider range of oxysterols, an ability contingent on ligand binding pocket remodelling.

Keywords: Nuclear receptors; cholesterol; Liver X receptor; chordates

3.1. Introduction

The appearance of complex endocrine systems, coordinating distinct biological functions such as development, metabolism or reproduction, represents a hallmark of bilaterian evolution (Bertrand et al. 2004). This increased complexity required a homeostatic coalescence of tissue-specific metabolic pathways and signalling cascades, with nuclear receptors (NRs) as major mediators of endocrine processes. These metazoan-specific transcription factors are mostly triggered by ligands of diverse origin (hormonal, nutritional, environmental), selectively modulating transcription upon recognition of specific DNA responsive elements, in the promoter region of target genes (Laudet and Gronemeyer 2002).

Evolutionary events like gene duplication, loss or mutation significantly contributed to functionally diversify NRs with likely impacts on organism physiology (e.g. Bertrand et al. 2004; Escriva et al. 2006; Bridgham et al. 2008; Carroll et al. 2008; Bridgham et al. 2010; Ogino et al. 2016). Gene duplication was particularly relevant in the case of vertebrates (e.g. Gutierrez-Mazariegos et al. 2016). As with many other gene families, the vertebrate NR repertoire augmented as a consequence of whole-genome duplications (WGD) in early vertebrate evolution (Thornton 2001; Bertrand et al., 2004;

Escriva et al. 2006; Lecroisey et al. 2012). This is thought to underscore the complexity and elaboration of the vertebrate endocrine system.

Here we scrutinize a particularly intriguing case involving a NR group, the liver X receptor (*LXR*, NR1H). Liver X Receptor plays a critical role in cholesterol homeostasis, regulating the expression of genes involved in the efflux, transport, and excretion (Kalaany and Mangelsdorf 2006; Laurencikiene and Ryden 2012). Originally classified as "ligand orphan" (Willy et al. 1995), it was later discovered that oxysterols (cholesterol oxidized derivatives) such as 24(S)-hydroxycholesterol, 22(R)-hydroxycholesterol and 24(S),25-epoxycholesterol are their *bona fide* ligands at physiological concentrations (Janowski et al. 1996; Lehmann et al. 1997). The repertoire of *LXRs* genes is surprisingly unequal in the investigated taxa. In mammals, two genes *LXRα* and *LXRβ* that share similar binding properties have been identified. Despite these similarities they control specific as well as overlapping physiological processes (Peet et al. 1998; Alberti et al. 2001; Juvet et al. 2003; Steffensen et al. 2003; Gerin et al. 2005; Korach-Andre et al. 2010). In contrast to mammals, a single gene was identified in birds (e.g. *Gallus gallus*), teleosts (e.g. *Danio rerio*) and amphibians (e.g. *Xenopus tropicalis* and *X. laevis*) and tunicates (Maglich et al. 2003; Reschly et al. 2008; Krasowski et al. 2011). This phylogenetic distribution was interpreted as a result from a duplication of a single *LXR* gene in mammalian ancestry (Maglich et al. 2003; Reschly et al. 2008; Krasowski et al. 2011). Upon duplication in the mammalian ancestor one paralog retained a more ubiquitous expression, while the second evolved specific roles in cholesterol metabolism (Reschly et al. 2008). However, an alternative hypothesis involving secondary loss of one *LXR* independently in multiple lineages would also account for the observed evolutionary pattern.

To discriminate between these evolutionary scenarios we investigated a broad range of chordate clades, including the chondrichthyans, cyclostomes and cephalochordates.

3.2 Material and Methods

3.2.1. Sequence Retrieval and Phylogenetic Analysis

Amino acid sequences were retrieved through BLAST searches in the publically available genome databases (Ensembl, GenBank, Skatebase; http://skatebase.org/), using as reference annotated human LXRα and LXRβ sequences. Sequence sampling included representatives of major vertebrate lineages: mammals (*Homo sapiens*, *Pan troglodytes*, *Mus musculus*, *Cavia porcellus*, *Dasypus novemcinctus*, *Oryctolagus cuniculus*), reptiles (*Thamnophis sirtalis*, *Python bivittatus*, *Anolis carolinensis*, *Alligator*

mississippiensis, *Alligator sinensis*) birds (*G. gallus*, *Meleagris gallopavo*, *Ficedula albicollis*, *Taeniopygia guttata*, *Anas platyrhynchos*), amphibians (*X.laevis*, *X. tropicalis*), sarcopterygii (*Latimeria chalumnae*), euteleostei (*Takifugu rubripes*, *Siganus canaliculatus*, *Oryzias latipes*, *Oreochromis niloticus*, *D. rerio*) osteoglossomorpha (*Scleropages formosus*), holostei (*Lepisosteus oculatus*), chondrichthyans (*L. erinacea*, *S. canicula*, *Callorhinchus milii*), cyclostomes (*L. japonicum* and *P. marinus*) and four invertebrate deuterostomes (*Ciona intestinalis*, *Branchiostoma lanceolatum*, *B. floridae* and *Saccoglossus kowalevskii*). Retrieved sequences and corresponding accession numbers are listed in the Supplementary Material online. All sequences were aligned with MAFFT alignment software (Katoh and Toh 2010) using the E-INS-i model. Sequence alignment was visualized and edited in Geneious® v7.1.7 (available upon request). The columns containing 90% of gaps were stripped. The final sequence alignment contained 86 sequences and 547 positions and was used to perform phylogenetic analysis. Bayesian phylogenetic calculation was performed with MrBayes v 3.2.3 sited in the CIPRES Science Gateway V3.3 (Miller et al. 2015). Calculation parameters were as follows, generation number = 1000000, rate matrix for aa = mixed (Jones), nruns = 2, nchains = 4, temp = 0.2, sampling set to 500 and burnin to 0.25.

3.2.2. Synteny Analysis

LXRα and *LXR*β genes were localized onto the human chromosomes Chr11 and Chr19, respectively, corresponding *LXR* gene and the neighbouring genes were collected from Ensembl and GenBank databases. Human *loci* (GRCh38) were further used as a reference to assemble the synteny maps of the remaining species: *G. gallus* (Galgal4), *A. carolinensis* (AnoCar2.0), *X. tropicalis* (GCF_000004195.3), *X. laevis* (GCF_001663975.1), *L. chalumnae* (LatCha1), *D. rerio* (GRCz10), *L. oculatus* (LepOcu1), *B. floridae* (GCF_000003815.1) and *B. lanceolatum* (BraLan2). Synteny statistics was calculated using CHSminer v1.1 (Wang et al. 2009) search parameters were maintained as default: maximal gap c 30 and size c 2, with the exception of the *D. rerio* Ch7 versus *H. sapiens* Chr11, were maximal gap was set to 80 to accommodate the highly rearranged locus in *D. rerio*.

3.2.3. Animal Sampling and *LXR* Isolation

Leucoraja erinacea were collected from the coast of Woods Hole, MA. All tissues were collected and preserved in RNAlater and stored at -20 ◦C. Total RNA was isolated using an Illustra RNAspin Mini RNA Isolation Kit (GE Healthcare, UK) according to the manufacturer's recommendations, including the oncolumn treatment of isolated RNA with RNase-free DNase I. Using 500ng of liver RNA as input cDNA was synthetized with

the iScript cDNA Synthesis Kit (Bio-Rad), according to the manufacturers' recommendations. Using two partial *LXR*α-like segments retrieved through BLAST searches in Skatebase, a set of primers were designed to isolate the partial open reading frame (ORF) of *LerLXRs* with Phusion Flash master mix (Thermo Fisher Scientific). The isolated partial ORF was further extended through rapid amplification of cDNA ends (RACE) technique. For this 5´RACE ready cDNA was prepared from previously isolated RNA using the SMARTER RACE cDNA amplification kit (Clontech) according to manufactures recommendations. The full ORF of *LerLXR*α was amplified using Phusion Flash master mix (Thermo Fisher Scientific). For LXR isolation in amphioxus (KY094511), *B. lanceolatum*, adult specimens were collected from Ria Formosa, Portugal. Total RNA and cDNA synthesis were performed as described earlier. A combination of PCR strategies (e.g., degenerate primers, RACE PCR and genome database search) was employed to isolate the full ORF of the *LXR* ortholog.

3.2.4. Construction of Plasmid Vectors

The ligand binding domain (LDB) including the hinge region of *H. sapiens LXR*β (HsaLXRβ; U07132.1), *L. erinacea LXR*α (LerLXRα) and *L. erinacea LXR*β (LerLXRβ) (transcriptome Contig89816 from Skatebase), *L. japonicum LXR* (LjaLXR) and *B. lanceolatum LXR* (BlaLXR) were isolated by PCR with Phusion Flash master mix (Thermo Fisher Scientific) using the specific primers (HsaLXRβ PF-ACTGGGATCCTAGATCCGGAAGAAGAAGATTCGG and PR-ATATCTAGATCACTCG TGGACGTCCCAGAT; LerLXRα PF-ACTGGGATCCGGAAGAAAATGAAGAAGCTGG AG and PR-ATATCTAGAAGTCATTCCTGCATGTCCCAG; LerLXRβ PF-ACTGGGATC CAGAAGAAGCAGAGGAAGCGGGAG and PR-ATATCTAGACCCTCCGTCACTCATG CAC; LjaLXR PF-CCCTCTAGACGTCGGAAAAACGACGAACC and PR-AAAGGTACC TCACTCGTGAACGTCCCAGA; BlaLXR PF-GCATCTAGACTCCGCGACAGAGCACC and PR-CCGGGTACCCTACTGTGGAACGTCCCATAT). PCR reaction comprised an initial denaturation step at 98 °C for 10 s followed by 40 cycles of denaturation at 98 °C for 1 s annealing at 62 °C (*LerLXRs*) or 60 °C (*LjaLXR* and *BlaLXR*) for 5 s and extension at 72 °C for 15 s, with an final extension step for 60 s. The resulting PCR products and pBIND (AF264722; Promega) were digested with *Bam*HI and *Xba*I (*LerLXRs*) or *Xba*I and *Kpn*I (*LjaLXR* and *BlaLXR*) restriction enzymes (Promega) and ligated withT4 ligase (Promega) to produce GAL4-LBD "chimeric" receptor.

3.2.5. Chemical and Solutions

The synthetic LXR agonist T0901317 and 24(S)-hydroxycholesterol (24-HC) were obtained from Enzo, 25-hydroxycholesterol (25-HC) and 24(S),25-epoxycholesterol (24,25-EC) were obtained from Santa Cruz Biotechnology. All test compounds were diluted in DMSO in order to obtain the desired concentrations.

3.2.6. Transfection and Transactivation Assays

COS-1 cells were seeded into a 24-well plate at a concentration of $2x10^5$ cells/ml 24 h prior to transfection. Cells were transfected with lipofectamine 2000 reagent (Invitrogen) and the transfection medium OptiMEM (Life Technologies) according manufacturer's indications, using 500 ng of pBIND constructions and 1,000 ng of pGL4.31[luc2P/GAL4UAS/Hygro] vector (DQ487213; Promega). After 5 h of incubation, transfection media was replaced with phenol red-free DMEM (PAN-Biotech) supplemented with 10% of charcoal stripped fetal bovine serum (PAA Laboratories) and cells were treated with varying concentration of oxysterols (ranging from 10^1 to $10^{5.5}$ nM) in DMSO. Cells were lysed 24 h after transfection and assayed for luciferase activity with Dual-Luciferase Reporter Assay System (Promega), according to the manufacturer's instructions. All transfections were performed in triplicate. Data was presented as means ± standard error (SE) from three separate experiments.

3.2.7. Statistical Analysis

SigmaPlot 11.0 software was used to calculate the EC50 values from the sigmoidal dose-response curves and the differences between groups variation were analysed with one-way ANOVA. Holm–Sidak was used to identify significant differences in the normalized-fold activation of the LXR receptors with the several compounds tested. The level of significance (*P* value) was set to 0.05.

3.3. Results

3.3.1. Asymmetric *LXR* Gene Collection in Vertebrate Lineages

To examine the *LXR* gene repertoire in vertebrate species, we searched the genome and transcriptome sequences of selected species from mammals, birds, reptiles, amphibians, coelacanth, teleosts, lepisosteiformes, chondrichthyans and cyclostomes. Confirming previous findings, *LXR*α and *LXR*β were identified in mammals (Reschly et al. 2008; Krasowski et al. 2011). In contrast, single-copy *LXR* genes were retrieved from spotted gar, most of teleosts and birds, consistent with previous observations (Reschly et al. 2008; Krasowski et al. 2011). However, our extensive

searches uncovered some gnathostome lineages with two *LXR* sequences: the Asian arowana (osteoglossomorpha), the coelacanth and the anole lizard. Further scrutiny of the available transcriptomes of the elephant shark, revealed a complete sequence and two nonoverlapping *LXR* sequence fragments (Supplementary Material **Figure S3.1**). Searches of additional genome and transcriptome sequences of cartilaginous fishes, the little skate *Leucoraja erinacea* and the small spotted catshark *Scyliorhinus canicula*, yielded several overlapping partial sequences with similarity to either *LXR*α or *LXR*β from bony vertebrates. Through a combination of PCR strategies we were able to recover the full or near-full coding sequence of two *LXR* genes in the little skate and the small spotted catshark. In contrast, a single *LXR*-like sequence was identified in the genome and transcriptome datasets of the Japanese lamprey, *Lethenteron japonicum*. Searches to the genome assembly and transcriptomes of the sea lamprey, *Petromyzon marinus*, allowed also the identification of a single *LXR*-like gene although too short for phylogenetic analysis (not shown). The investigation of the genome sequences from two cephalochordate species allowed the recovery of 10 *LXR*/*FXR*-like sequences (NR1H1-10), similar to those found in previous studies (Bertrand et al. 2011; Lecroisey et al. 2012). However, a clear identity was only verified in seven, as three *B. floridae* sequences have been discontinued, namely NR1H2 (XP_002224320.1) NR1H4 (XP_002224321.1) and NR1H10 (XP_002246474.1). Additionally, another sequence (NR1H9) shows a truncated DBD (not shown). Thus, a total of six cephalochordate sequences were considered for the main phylogenetic analysis.

To assign orthology/paralogy of the recovered sequences we next carried out phylogenetic analysis (**Figure 3.1** and Supplementary Material **Figure S3.2**). Two monophyletic clades containing, respectively, *LXR*α and *LXR*β are observed, out-grouped by single-copy *LXR* sequences from the Japanese lamprey, cephalochordates, sea squirt and hemichordate (**Figure 3.1** and Supplementary Material **Figure S3.2**). Thus, data derived from genome, transcriptome and phylogenetics indicate that *LXR*α orthologs are present in all the examined gnathostome species (except amphibians), while *LXR*β is found in mammals, reptiles, amphibians, coelacanth and cartilaginous fish. The little skate and the small spotted catshark *LXRs* genes robustly groups with the *LXR*α and *LXR*β clades, respectively (**Figure 3.1**), providing unequivocal support for their orthology. The Asian arowana sequences are both of the *LXR*α type, a probable consequence of the teleost- specific genome duplication (3R) or a lineage-specific duplication. In summary, phylogenetic analysis suggests a much earlier origin of *LXR*α and *LXR*β than the timing of mammalian radiation (Reschly et al. 2008), predating

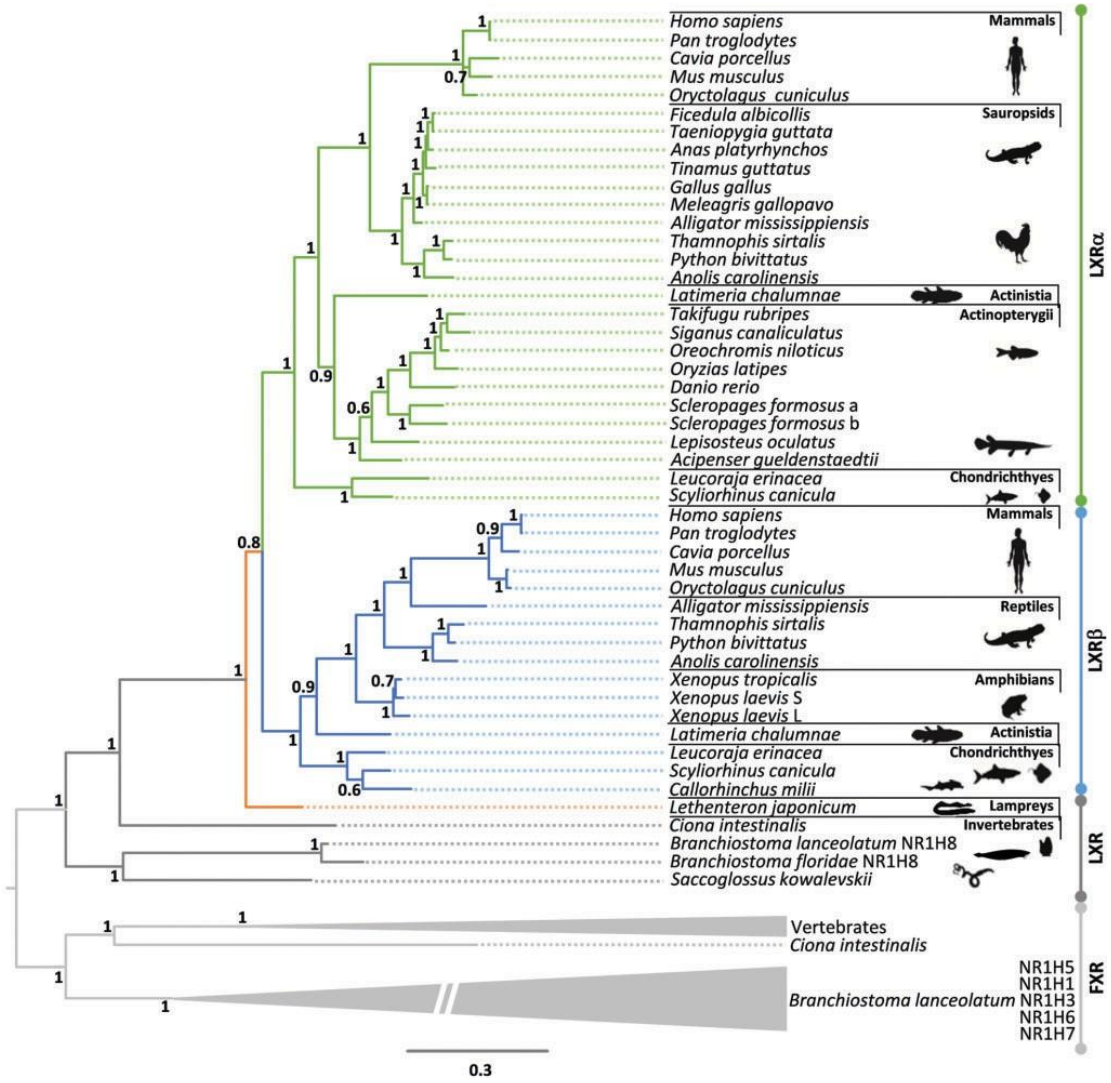
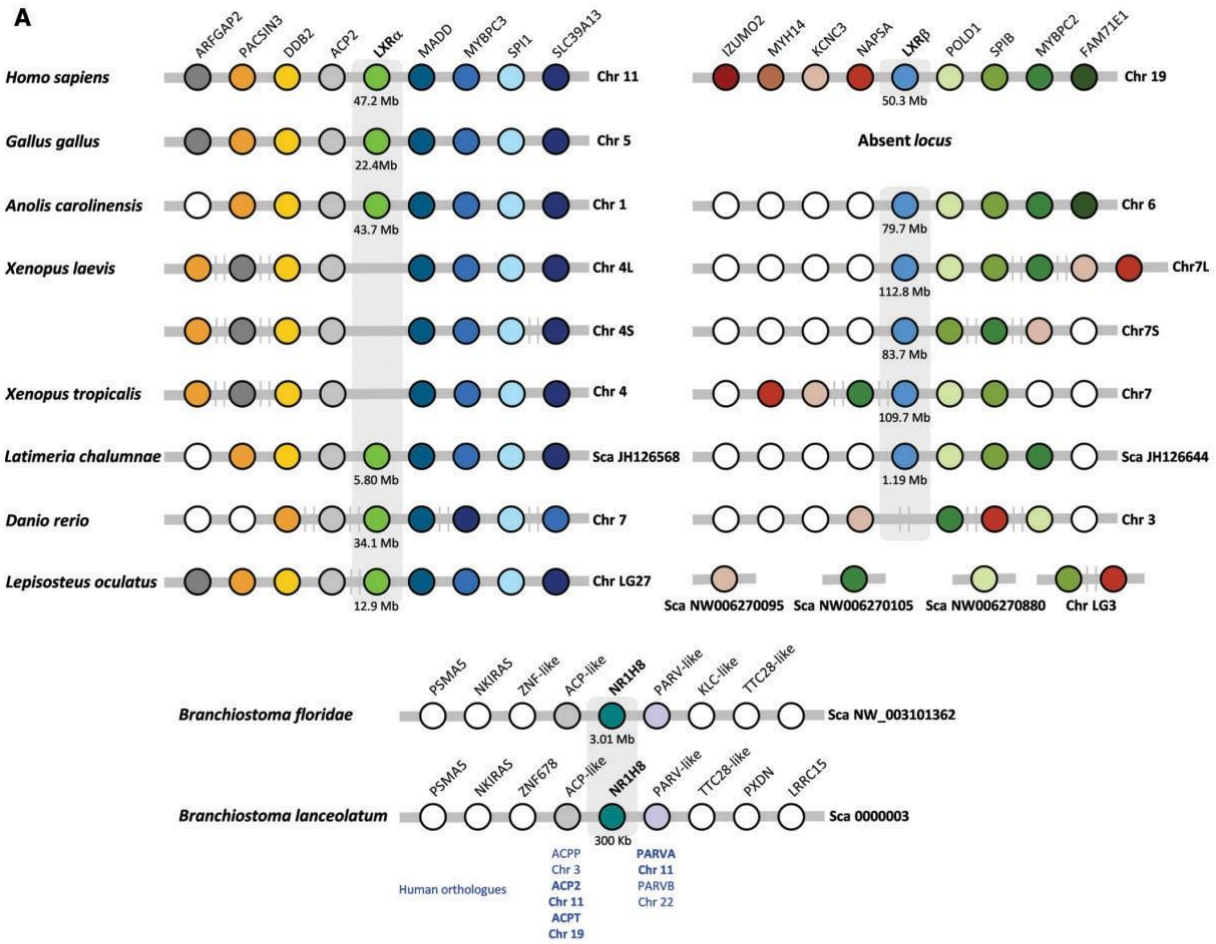


Figure 3.1. Phylogenetic analysis of NR1H nuclear receptors (LXRs/FXRs). Bayesian phylogenetic tree of LXR and FXR amino acid sequences; numbers at nodes indicate posterior probabilities.

gnathostome divergence but after splitting from cyclostomes (**Figure 3.1**). Our analysis also confirms that the unusual NR1H gene number in cephalochordates is the result of a lineage specific expansion of the *FXR* clade as previously suggested (**Figure 3.1** and Supplementary Material **Figure S3.2**) (Bertrand et al. 2011; Lecroisey et al. 2012).

3.3.2. Synteny Analysis Supports Lineage Specific Events of Gene Loss

To discriminate between true gene loss and absence of sequencing data we next investigated the synteny of *LXR* genes in selected species with available genome data (**Figure 3.2**). We find strong synteny conservation in the examined *LXRα loci.* In both *Xenopus* species the loss of *LXRα* is confirmed since the *locus* is conserved with no



B

	Species	Homologous region 1 coordinates	Species	Homologous region 2 coordinates	S	P value
LXRα	*G. gallus*	**Chr5** 20371044 : 26157686	*H. sapiens*	**Chr11** 40135753 : 47606114	45	1.468 E-128
	A. carolinensis	**Chr1** 42299177 : 44314211	*H. sapiens*	**Chr11** 46299212 : 47616211	27	2.676 E-85
	D. rerio	**Chr7** 33172830 : 41003835	*H. sapiens*	**Chr11** 45261852 : 48192393	11*	1.283 E-55
LXRβ	*A. carolinensis*	**Chr6** 78845194 : 80632227	*H. sapiens*	**Chr19** 49497156 : 51228979	29	7.626 E-86
	X. tropicalis	**GL172742.1** 32955 : 1074361	*H. sapiens*	**Chr19** 48373723 : 51298481	12	5.894 E-39

Figure 3.2. Synteny maps of *LXRα* and *LXRβ loci*. **(A)** Detail of the *LXRα locus* and *LXRβ locus* in the selected vertebrate and cephalochordate species; Chr and Sca indicate chromosome and scaffold, respectively. **(B)** Statistical support of synteny analysis; *P* values indicate the probability of identifying non-homologous chromosomal segments, and S indicates the size of the chromosomal segment identified.

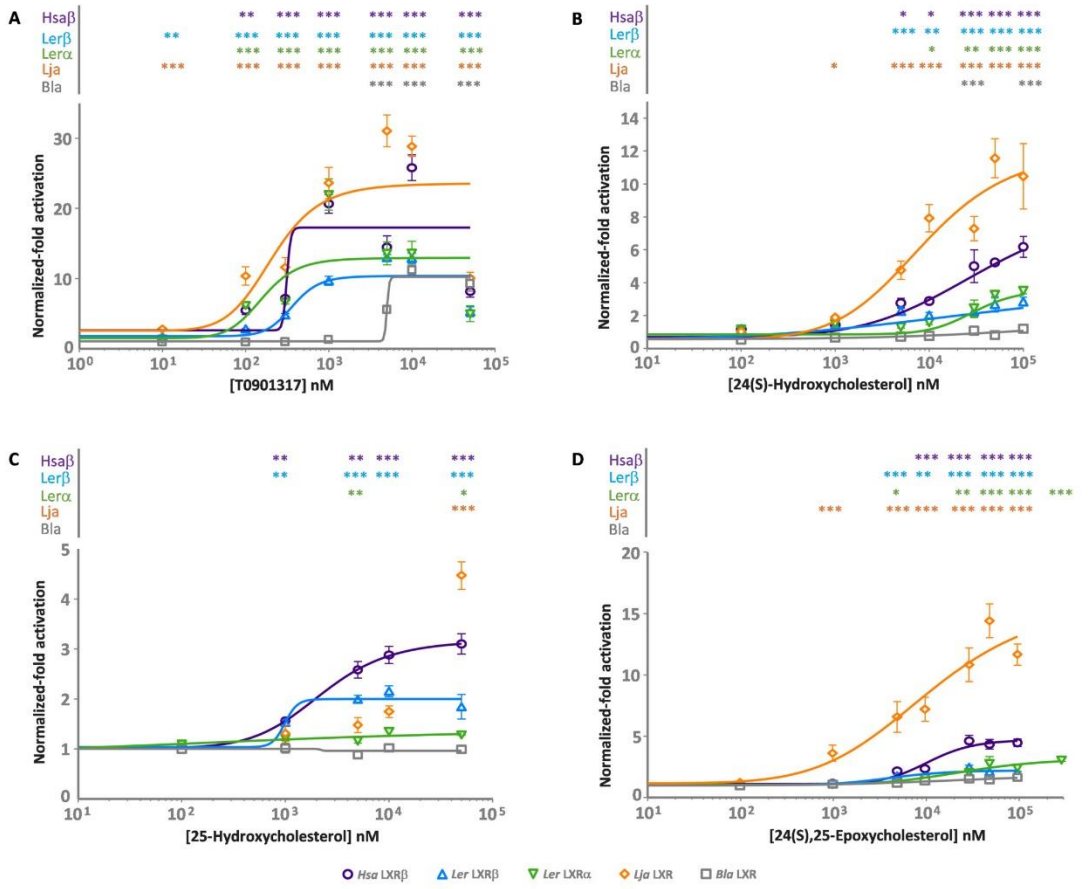
LXR-like intervening sequence. Orthologous flanking genes in the *LXRβ loci* are not as evident but still statistically supported, in reptiles and amphibians (**Figure 3.2**). This *locus* is entirely absent in birds as previously noted (Lovell et al. 2014). The *LXRβ* of the elephant shark maps to a single gene scaffold impeding synteny comparisons (not shown). The cephalochordate *LXR locus* shows some degree of conservation when compared with vertebrates (**Figure 3.2**). One flanking gene, *ACP-like*, presents the

corresponding human orthologs, *ACP2* and *ACPT*, located in close proximity to *LXR*α (Chr11 p11.2 47.20 Mb) and *LXR*β (Chr19 q13.33 50.8 Mb), respectively. A second flanking gene in the cephalochordate *LXR locus*, *PARV-like*, has the human orthologs locating to the *LXR*α *locus*, *PARVA* (Chr11 p15.3 12.40 Mb) and *PARVB* in human chromosome 22 (Chr22 q13.31 44 Mb) (**Figure 3.2**). In contrast, the cephalochordate *FXR*-like genes localize to separate and nonsyntenic genomic locations with respect to the *LXR loci* (Supplementary Material **Figure S3.3**).

3.3.3. Conserved and Derived Ligand Specificity of Chordate *LXRs*

To determine the ligand binding properties of the little skate LXRα and LXRβ, and the single proteins of Japanese lamprey and European amphioxus, we investigated their capacity to bind to physiological and synthetic LXR ligands. The synthetic compound T0901317 is known to be a LXR agonist in several species. The agonistic response was observed for the two little skate receptors and for the lamprey receptor with significant activations ($P < 0.001$), in all of the tested concentrations (**Figure 3.3A**), suggesting that this synthetic compound is also a potent agonist for these three receptors (*L. erinacea* LXRα EC_{50}=0.2 μM, *L. erinacea* LXRβ EC_{50}=0.4 μM, *L. japonicum* LXR EC_{50}= 0.2 μM). Interestingly, amphioxus LXR is less sensitive to this agonist, displaying statistical significant activation ($P < 0.001$) only at higher concentrations (EC_{50}= 5.0 μM). We next assayed three oxysterols: 24(S)-hydroxycholesterol (24-HC), 25-hydroxycholesterol (25-HC) and 24(S),25-epoxycholesterol (24,25-EC) (**Figure 3.3B–D**). Previous studies including human and mouse LXRs reported that 24-HC and 24,25-EC robustly activated both LXRα and LXRβ; 25-HC, on the other hand, induced lower transcriptional responses at the tested concentrations, with a more prominent decrease observed for LXRα (Reschly et al. 2008). A similar pattern was observed for the little skate α and β isoforms. The lamprey LXR was also strongly induced by 24-HC and 24,25-EC, producing the highest maximal activations. Upon exposure to 25-HC, lamprey LXR transcriptional response was also prominent and statistically significant ($P < 0.001$), yet only at higher concentrations. Contrary to the vertebrate isoforms, the amphioxus LXR was residually, but significantly, activated by 24-HC alone. Overall, the obtained EC_{50} values for the tested LXR/oxysterol pairs are much higher, in the micromolar range than those of T0901317 (**Figure 3.3E**). Likewise, the respective maximal activations were lower than for T0901317, similar to mammalian LXRs (Reschly et al. 2008). Regarding the Japanese lamprey LXR exposed to 25-HC, due to the absence of a dose–response

plateau, within the tested concentration range, the estimation of EC_{50} and maximal activation values was not performed.



E

	Hsa LXRβ		*Ler* LXRβ		*Ler* LXRα		*Lja* LXR		*Bla* LXR	
	EC_{50} (μM)	$Max_{fold\ A}$	EC_{50} (μM)	$Max_{fold\ A}$	EC_{50} (μM)	$Max_{fold\ A}$	EC_{50} (μM)	$Max_{fold\ A}$	EC_{50} (μM)	$Max_{fold\ A}$
T0901317	0.3	17.3	0.4	10.4	0.2	13.0	0.2	23.6	5.0	10.3
24-HC	36.6	8.8	153.9	4.8	29.3	3.8	8.6	12.1	14.7	1.5
25-HC	2.2	3.2	1.0	2.0	2.7	1.4	-	-	2.2	1.0
EC	11.0	4.7	4.4	2.2	26.7	3.2	11.4	16.2	53.9	2.1

Figure 3.3. Functional analysis of *L. erinacea*, *L. japonicum* and *B. lanceolatum* LXRs LBD. Dose-response curves for LXRs activation by T0901317 **(A)**, 24(S)-hydroxycholesterol **(B)**, 25-hydroxycholesterol **(C)** and 24(S),25-epoxycholesterol **(D)** for *H. sapiens* LXRβ (○), *L. erinacea* LXRβ (Δ) and LXRα (∇), *L. japonicum* (◊) and *B. lanceolatum* (□); EC_{50} and maximum normalized-fold activation ($Max_{fold\ A}$) values for HsaLXRβ, LerLXRβ and LerLXRα **(E)**. The activation of LXR was normalized to the control condition (DMSO without ligand) represented by 10^{-2} M. Hsa stands for *H. sapiens*, Ler stands for *L. erinacea*, Lja stands for *L. japonicum* and Bla stands for *B. lanceolatum*. The values represented are the means with ±SE from three separate experiments. * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$.

3.4. Discussion

By performing a comprehensive search of the *LXR* gene repertoire in chordate genomes, we unveil the accurate evolutionary functional diversification of this NR group (**Figure 3.4**). A single *LXR* member was previously found in basal chordates such as

tunicates (Reschly et al. 2008), similar to what we describe here in cephalochordates and cyclostomes. In contrast, we also report that *LXR*α and *LXR*β are gnathostome-specific paralogs found in a wide array of lineages including the chondrichthyans, with independent gene loss of either paralog in several lineages (Supplementary Material **Figure S3.4**). The overall timing of *LXR* gene duplication as determined from our phylogenetic analysis is coincident with that of whole-genome duplications in vertebrate ancestry (1R/2R) (Putnam et al. 2008; Smith and Keinath 2015), although the absence of synteny data from lamprey and chondrichthyans impedes a detailed discussion.

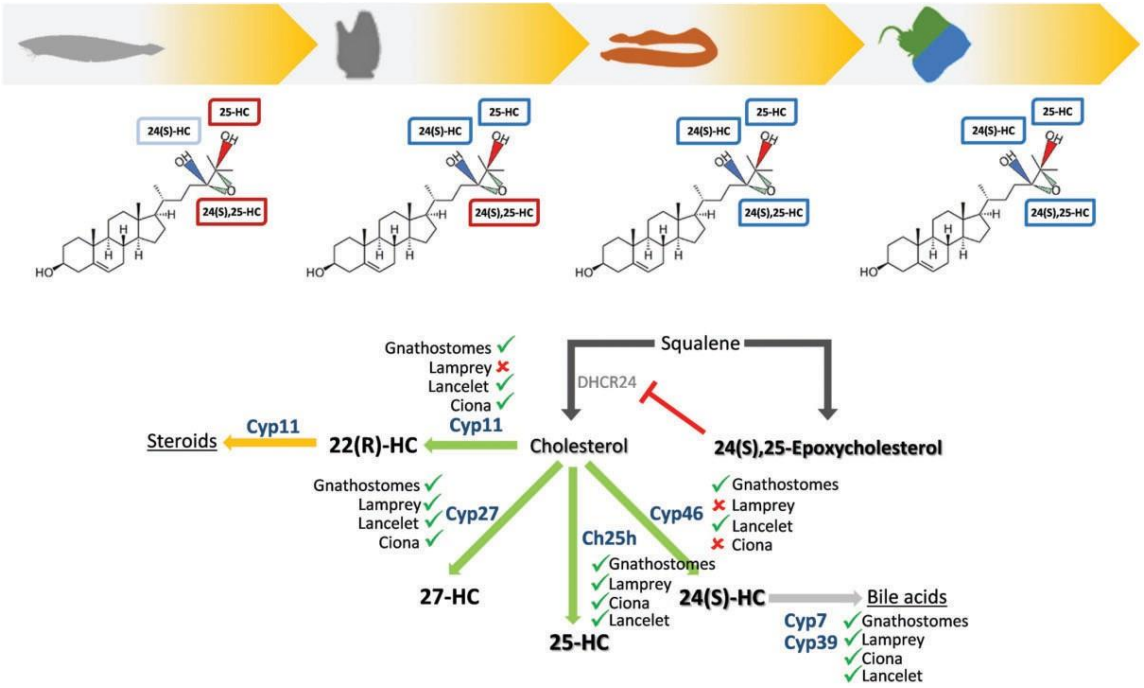


Figure 3.4. Elaboration of the metabolic and signalling oxysterols cascades in chordate evolution. Top: Binding specificities of LXRs to oxysterols in cephalochordates, tunicates, cyclostomes and chondrichthyans: red and blue boxes represent no activation and activation of LXR-dependent transcription, respectively, and light blue boxes represent residual LXR activation. Bottom: Schematic representation of cholesterol synthesis and oxidation pathways: green ✓ and red × stand for presence or absence of the corresponding cytochrome P450s oxygenase (CYP) gene, respectively (Nelson et al. 2013).

Unlike vertebrates, the amphioxus ortholog displays low to null capacity to activate transcription upon exposure to any of the tested oxysterols. The tunicate LXR, we should recall, exhibits an intermediary pharmacology: not activated by synthetic LXR agonists, yet induced by some oxysterols and other steroidal compounds, such as androstenol and androstanol. Besides steroids, the tunicate LXR was strongly activated by 6-Formylindolo(3,2-b)carbazole, a tryptophan photoproduct and proposed endogenous aryl hydrocarbon receptor ligand (Reschly et al. 2008). Certainly, if steroids, or other polycyclic compounds, are the tunicate LXR physiological ligands, they differ from those

of vertebrates (Reschly et al. 2008). Yet, vertebrate LXR ligand capacity seems contingent on an emerging ability to accommodate oxysteroid backbones: originally limited to a narrower set of oxysterols in the ancestor of tunicates and vertebrates. The broader variety of oxysterol ligands, capable of specifically activating LXRs, first appeared in the ancestor of vertebrates, as depicted in our results with the cyclostome and chondrichthyan LXR receptors. In agreement, the comparison of the LXR ligand binding pockets (LBP) from invertebrate and vertebrate species indicates that a significant remodelling occurred, after the separation of tunicates from vertebrates, with multiple substitutions entrenched in the vertebrate lineage (Supplementary Material **Figure S3.5**). Curiously, this parallels the assembly of bile acid synthesis pathways, of which oxysterols serve as intermediates. Classic and alternate synthesis pathways are triggered by cytochrome P450s (CYPs) oxygenase-dependent cholesterol oxidation (**Figure 3.4**). Similarly to the NR repertoire, the original, tandemly repeated, CYP clans were expanded by duplication (Nelson et al. 2013). According to the ligand exploitation paradigm, selection for novel biosynthetic pathways generates a collection of intermediates and prompts new combinations of ligand/receptor pairs (Thornton 2001). Thus, it is conceivable that both metabolic (oxysterols) and signalling (LXRs) pathways evolved, in parallel and opportunistically, towards the integrated network observed in gnathostomes.

3.5. Conclusions

The coexistence of different complements of *LXR* genes in gnathostome lineages is enigmatic. In effect, the retention of *LXR*α and *LXR*β in many lineages indicates that both NRs evolved separate roles in the aftermath of gene duplication. Paradoxical though, the vertebrate *LXR* paralogs examined to date exhibit some redundancy in ligand specificity. Interestingly, the zebrafish and the amphibian *LXR*α and *LXR*β, respectively, also show conserved ligand specificity compared with both mammalian *LXRs* (Reschly et al. 2008).Thus, the coexistence of one (birds, amphibians and teleosts) or two (chondrichthyans, coelacanths, reptiles and mammals) *LXR* paralogs in diverse gnathostome lineages does not imply distinguishable differences in *LXR* ligand preference. Given the apparent functional redundancy of both receptors at ligand preference, it is possible that there could be differences in the transcriptional regulation of the genes coding for the receptors per se. Data from mouse supports this hypothesis. In fact, Peroxisome Proliferator-Activated Receptors (*PPARs*), major regulators of lipid metabolism, were suggested to induce the transcription of *LXR*α, but not *LXR*β, indirectly regulating cholesterol metabolism (Tobin et al. 2000; Chawla et al. 2001). Nevertheless,

differential activation mechanics, such as nuclear targeting, corepressor or coactivator binding, phosphorylation, could favour specific, isoform-specific, metabolic pathways.

Supplementary Material: Supplementary data are available at Genome Biology and Evolution online.

Acknowledgments: We thank Neelakanteswar Aluru and Pedro Morais for the samples of little skate and amphioxus, respectively. This work was supported by Norte2020 and FEDER (Coral—Sustainable Ocean Exploitation—Norte-01-0145-FEDER-000036), EXPL/MAR-EST/1540/2012. SFRH/BD/79305/2011 to E.F., SFRH/BD/84238/2012 to M.L.-M. and SFRH/BPD/72519/2010 to R.R. L.F.C.C. and M.M.S. are supported by the European Regional Development Fund (ERDF) through the COMPETE—Operational Competitiveness Programme and POPH—Operational Human Potential Programme and national funds through FCT—Foundation for Science and Technology (UID/Multi/04423/2013). This work was supported by the *Branchiostoma lanceolatum* genome consortium that provided access to the *Branchiostoma lanceolatum* genome sequence. We thank three anonymous reviewers for their comments and suggestions.

3.6. References

Alberti S, et al. 2001. Hepatic cholesterol metabolism and resistance to dietary cholesterol in *LXRbeta*-deficient mice. J Clin Invest. 107:565– 573. doi: 10.1172/JCI9794.

Bertrand S, et al. 2004. Evolutionary genomics of nuclear receptors: from twenty-five ancestral genes to derived endocrine systems. Mol Biol Evol. 21:1923–1937. doi: 10.1093/molbev/msh200.

Bertrand S, Belgacem MR, Escriva H. 2011. Nuclear hormone receptors in chordates. Mol Cell Endocrinol. 334:67–75. doi: 10.1016/j.mce.2010.06.017.

Bridgham JT, Brown JE, Rodriguez-Mari A, Catchen JM, Thornton JW. 2008. Evolution of a new function by degenerative mutation in cephalochordate steroid receptors. PLoS Genet. 4:e1000191. doi: 10.1371/journal.pgen.1000191.

Bridgham JT, et al. 2010. Protein evolution by molecular tinkering: diversification of the nuclear receptor superfamily from a ligand-dependent ancestor. PLoS Biol. 8:e1000497. doi: 10.1371/journal.pbio.1000497.

Carroll SM, Bridgham JT, Thornton JW. 2008. Evolution of hormone signaling in elasmobranchs by exploitation of promiscuous receptors. Mol Biol Evol. 25:2643–2652. doi: 10.1093/molbev/msn204.

Chawla A, et al. 2001. A PPAR gamma-LXR-ABCA1 pathway in macrophages is involved in cholesterol efflux and atherogenesis. Mol Cell 7:161–171. doi: 10.1016/s1097-2765(01)00164-2.

Escriva H, et al. 2006. Neofunctionalization in vertebrates: the example of retinoic acid receptors. PLoS Genet. 2(7):e102. doi: 10.1371/journal.pgen.0020102.

Gerin I, et al. 2005. *LXRbeta* is required for adipocyte growth, glucose homeostasis, and beta cell function. J Biol Chem. 280:23024– 23031. doi: 10.1074/jbc.M412564200.

Gutierrez-Mazariegos J, et al. 2016. Evolutionary diversification of retinoic acid receptor ligand-binding pocket structure by molecular tinkering. R Soc Open Sci. 3:150484. doi: 10.1098/rsos.150484.

Janowski BA, Willy PJ, Devi TR, Falck JR, Mangelsdorf DJ. 1996. An oxysterol signalling pathway mediated by the nuclear receptor *LXR alpha*. Nature 383:728–731. doi: 10.1038/383728a0.

Juvet LK, et al. 2003. On the role of liver X receptors in lipid accumulation in adipocytes. Mol Endocrinol. 17:172–182. doi: 10.1210/me.2001-0210.

Kalaany NY, Mangelsdorf DJ. 2006. *LXRS* and *FXR*: the yin and yang of cholesterol and fat metabolism. Annu Rev Physiol. 68:159–191. doi: 10.1146/annurev.physiol.68.033104.152158.

Katoh K, Toh H. 2010. Parallelization of the MAFFT multiple sequence alignment program. Bioinformatics 26:1899–1900. doi: 10.1093/bioinformatics/btq224.

Korach-Andre M, et al. 2010. Separate and overlapping metabolic functions of *LXRalpha* and *LXRbeta* in C57Bl/6 female mice. Am J Physiol Endocrinol Metab. 298:E167–E178. doi: 10.1152/ajpendo.00184.2009.

Krasowski MD, Ni A, Hagey LR, Ekins S. 2011. Evolution of promiscuous nuclear hormone receptors: *LXR*, *FXR*, *VDR*, *PXR*, and *CAR*. Mol Cell Endocrinol. 334:39–48. doi: 10.1016/j.mce.2010.06.016.

Laudet V, Gronemeyer H. 2002. *The nuclear receptors factsbook*. London: Academic press.

Laurencikiene J, Ryden M. 2012. Liver X receptors and fat cell metabolism. Int J Obes (Lond). 36:1494–1502. doi: 10.1038/ijo.2012.21.

Lecroisey C, Laudet V, Schubert M. 2012. The cephalochordate amphioxus: a key to reveal the secrets of nuclear receptor evolution. Brief Funct Genomics 11:156–166. doi: 10.1093/bfgp/els008.

Lehmann JM, et al. 1997. Activation of the nuclear receptor LXR by oxysterols defines a new hormone response pathway. J Biol Chem. 272:3137–3140. doi: 10.1074/jbc.272.6.3137.

Lovell PV, et al. 2014. Conserved syntenic clusters of protein coding genes are missing in birds. Genome Biol. 15:565. doi: 10.1186/s13059-014-0565-1.

Maglich JM, et al. 2003. The first completed genome sequence from a teleost fish (*Fugu rubripes*) adds significant diversity to the nuclear receptor superfamily. Nucleic Acids Res. 31:4051–4058. doi: 10.1093/nar/gkg444.

Miller MA, et al. 2015. A RESTful API for access to phylogenetic tools via the CIPRES science gateway. Evol Bioinform Online 11:43–48. doi: 10.4137/EBO.S21501.

Nelson DR, Goldstone JV, Stegeman JJ. 2013. The cytochrome P450 genesis *locus*: the origin and evolution of animal cytochrome P450s. Philos Trans R Soc Lond B Biol Sci. 368:20120474. doi: 10.1098/rstb.2012.0474.

Ogino Y, et al. 2016. Neofunctionalization of androgen receptor by gain-of-function mutations in teleost fish lineage. Mol Biol Evol. 33:228–244. doi: 10.1098/rstb.2012.0474.

Peet DJ, et al. 1998. Cholesterol and bile acid metabolism are impaired in mice lacking the nuclear oxysterol receptor *LXR alpha*. Cell 93:693– 704. doi: 10.1016/s0092-8674(00)81432-4.

Putnam NH, et al. 2008. The amphioxus genome and the evolution of the chordate karyotype. Nature 453:1064–1071. doi: 10.1038/nature06967.

Reschly EJ, et al. 2008. Ligand specificity and evolution of liver X receptors. J Steroid Biochem Mol Biol. 110:83–94. doi: 10.1016/j.jsbmb.2008.02.007.

Smith JJ, Keinath MC. 2015. The sea lamprey meiotic map improves resolution of ancient vertebrate genome duplications. Genome Res. 25:1081–1090. doi: 10.1101/gr.184135.114.

Steffensen KR, et al. 2003. Gene expression profiling in adipose tissue indicates different transcriptional mechanisms of liver X receptors alpha and beta, respectively. Biochem Biophys Res Commun. 310:589–593. doi: 10.1016/j.bbrc.2003.08.139.

Thornton JW. 2001. Evolution of vertebrate steroid receptors from an ancestral estrogen receptor by ligand exploitation and serial genome expansions. Proc Natl Acad Sci U S A. 98:5671–5676. doi: 10.1073/pnas.091553298.

Tobin KA, et al. 2000. Cross-talk between fatty acid and cholesterol metabolism mediated by liver X receptor-alpha. Mol Endocrinol. 14:741–752. doi: 10.1210/mend.14.5.0459.

Wang Z, Ding G, Yu Z, Liu L, Li Y. 2009. CHSMiner: a GUI tool to identify chromosomal homologous segments. Algorithms Mol Biol. 4:1–7. doi: 10.1186/1748-7188-4-2.

Willy PJ, et al. 1995. *LXR*, a nuclear receptor that defines a distinct retinoid response pathway. Genes Dev. 9:1033–1045. doi: 10.1101/gad.9.9.1033.

3.7. Supplementary Material

Retrieved sequences for phylogenetic reconstruction and corresponding accession numbers are listed as follows: *H. sapiens* LXRα (NP_001238863.1), *H. sapiens* LXRβ (P55055.2), *P. troglodytes* LXRα (XP_009458632.1), *P. troglodytes* LXRβ (XP_009434358.1), *M. musculus* LXRα (NP_038867.2), *C. porcellus* LXRα (XP_003465264.1), *C. porcellus* LXRβ (XP_013003335.1), *D. novemcinctus* LXRα (XP_004457991.1), *D. novemcinctus* LXRβ (XP_012379317.1), *O. cuniculus* LXRα (NP_001171885.1), *O. cuniculus* LXRβ (NP_001171965.1), *T. sirtalis* LXRα (XP_013926197.1), *T. sirtalis* LXRβ (XP_013920824.1), *P. bivittatus* LXRα (XP_007431628.1), *P. bivittatus* LXRβ (XP_007423157.1), *A. carolinensis* LXRα (XP_003214647.1), *A. carolinensis* LXRβ (XP_003222765.1), *A. mississippiensis* LXRα (XP_006261463.1), *A. mississippiensis* LXRβ (XP_006274461.1), *A. sinensis* LXRβ (XP_006014951.1), *G. gallus* LXRα (NP_989873.1), *M. gallopavo* LXRα(ENSMGAT00000012056), *F. albicollis* LXRα (XP_005047160.1), *T. guttata* LXRα (XP_002200060.1), *A. platyrhynchos* LXRα (NP_001297352.1), *X. laevis* LXRβ-L (NP_001086083.1), *X. laevis* LXRβ-S (XP_018083760.1), *X. tropicalis* LXRβ (NP_001072853.1), *L. chalumnae* LXRα (XP_005987002.1), *L. chalumnae* LXRβ (XP_005990991.1), *T. rubripes* LXRα (XP_003969673.1), *S. canaliculatus* LXRα (AFH35110.1), *O. latipes* LXRα (XP_004066988.1), *O. niloticus* LXRα (XP_005455771.1), *D. rerio* LXRα (XP_009301489.1), *L. oculatus* LXRα (assembly of SRA file SRX661023 and the isolated KY094512), *S. formosus* LXRαb (KPP69102.1), *S. formosus* LXRαa (KPP71249.1), *L. erinacea* LXRα (KY094507), *L. erinacea* LXRβ (KY094508), *S. canicula* LXRα (KY094510), *S. canicula* LXRβ (assembly of ctg29778, ctg35662, ctg1931), *L. japonicum* LXR (KY094509), *C. milii* LXRβ (XP_007883658.1), *C. intestinalis* LXR (BAE06541.1), *S. kowalevskii* LXR (NP_001161579.1), *H. sapiens* FXRα (NP_005114.1), *P. troglodytes* FXRα (XP_003952320.1), *M. musculus* FXRα (NP_001157172.1), *M. musculus* FXRβ (NP_941060.2), *O. cuniculus* FXRα (ENSOCUP00000027131), *O. cuniculus* FXRβ (ENSOCUP00000021724), *G. gallus* FXRα (ENSGALP00000018906), *G. gallus* FXRβ (ENSGALP00000003395), *T. guttata* FXRα (ENSTGUP00000009262), *T. guttata* FXRβ (ENSTGUG00000000964), *A. mississippiensis* FXRα (XP_014450696.1), *A. mississippiensis* FXRβ (XP_006260806.1), *T. sirtalis* FXRα (XP_013920207.1), *T. sirtalis* FXRβ (XP_013909551.1), *P. bivittatus* FXRα (XP_007438686.1), *P. bivittatus* FXRβ (XP_007434736.1) *A. carolinensis* FXRα (ENSACAP00000012155), *A. carolinensis*

FXRβ (ENSACAP00000003240), *X. tropicalis* FXRα (XP_002936891.2), *X. tropicalis* FXRβ (NP_001107159.1), *P. formosa* FXRα (XP_007545600.1), *P. formosa* FXRβ (XP_016536194.1), *X. maculatus* FXRα (XP_005802311.1), *X. maculatus* FXRβ (XP_005805852.1), *O. niloticus* FXRα (XP_013124133.1), *O. niloticus* FXRβ (ENSONIP00000011635), *T. rubripes* FXRα (ENSTRUP00000018054), *T. rubripes* FXRβ (ENSTRUP00000020496), *D. rerio* FXRα (ENSDARP00000061793), *D. rerio* FXRβ (ENSDARP00000105819), *L. oculatus* FXRα (XP_006633125.1), *L. oculatus* FXRβ (XP_015197747.1), *C. intestinalis* FXR (FAA00147.1), *B. floridae* NR1H8 (XP_002589462.1), *B. lanceolatum* NR1H1 (BL97043), *B. lanceolatum* NR1H3 (BL16685), *B. lanceolatum* NR1H5 (BL09785), *B. lanceolatum* NR1H6 (BL20884), *B. lanceolatum* NR1H7 (BL05413), *B. lanceolatum* NR1H8 (BL08609).

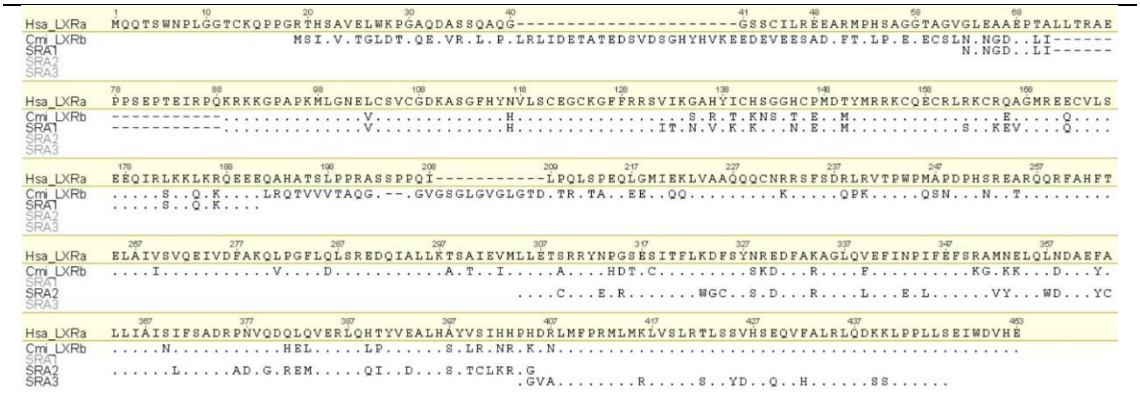


Figure S3.1. MAFFT alignment of *Callorhinchus milii* partial LXRα sequences. The human LXRα (NP_001238863.1) was set as reference and the elephant shark LXRβ (XP_007883658.1) was used as a control for the alignment of the elephant shark partial sequences LXRα-like retrieved through blast searches in SRA file (SRX154858) available in NCBI; Hsa stands for *H. sapiens* and Cmi stands for *C. milii*.

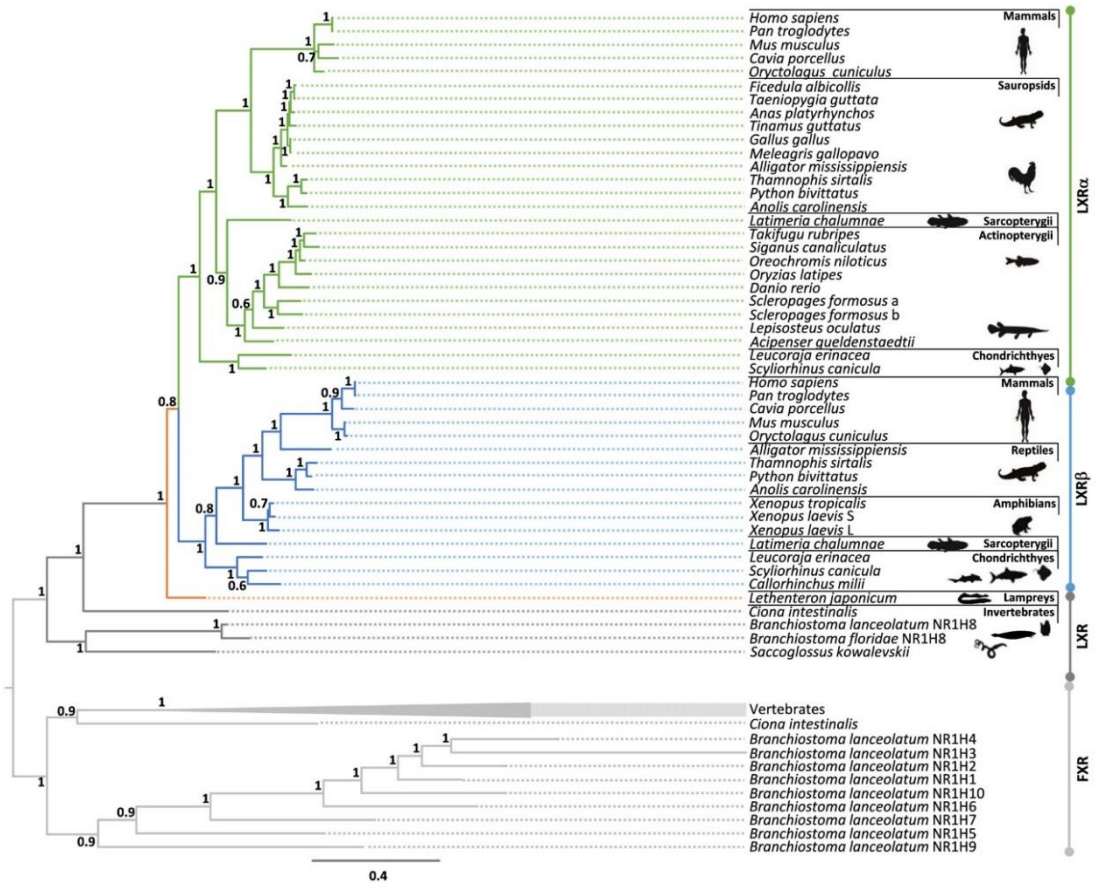


Figure S3.2. Phylogenetic analysis of NR1H nuclear receptors (LXRs/FXRs) with the *B. lanceolatum* NR1H full gene set. The approach used in the main tree was applied with the final sequence alignment containing 90 sequences and 554 positions; numbers at nodes indicate posterior probabilities.

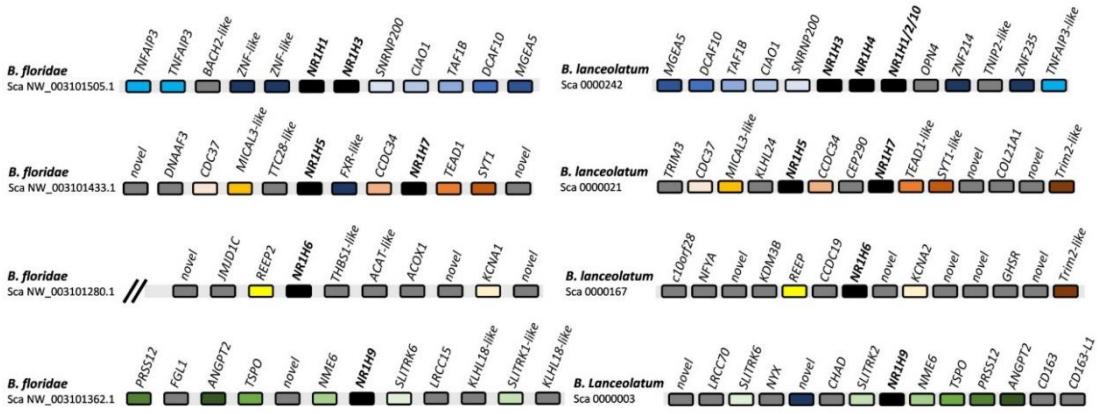


Figure S3.3. Synteny maps of *B. floridae* and *B. lanceolatum* FXR-like *loci*. Sca indicates scaffold; Colour code indicates orthologs genes between *B. floridae* and *B. lanceolatum*.

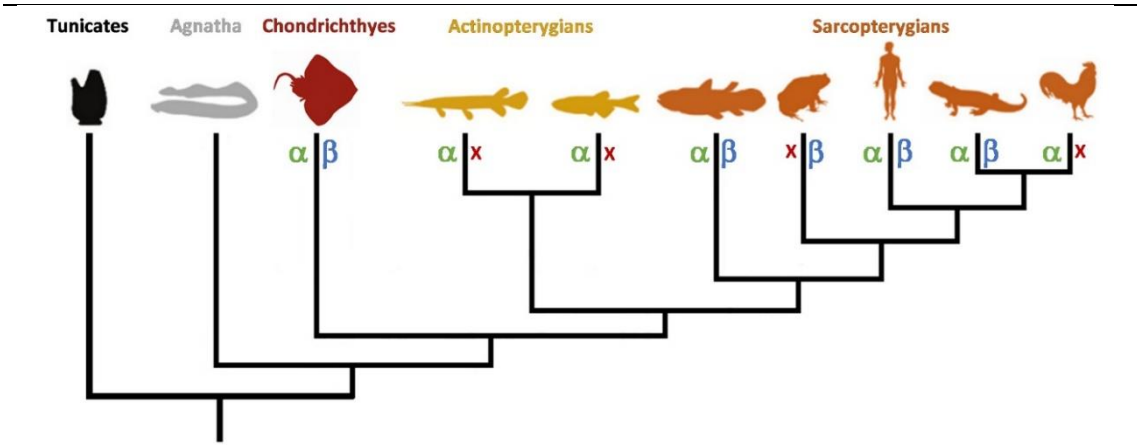


Figure S3.4. Simplified phylogeny of *LXR* gene evolution in chordates. The **α** indicates the presence of *LXRα* paralog, the **β** indicates the presence of *LXRβ* paralog and **X** indicates LXR paralog loss.

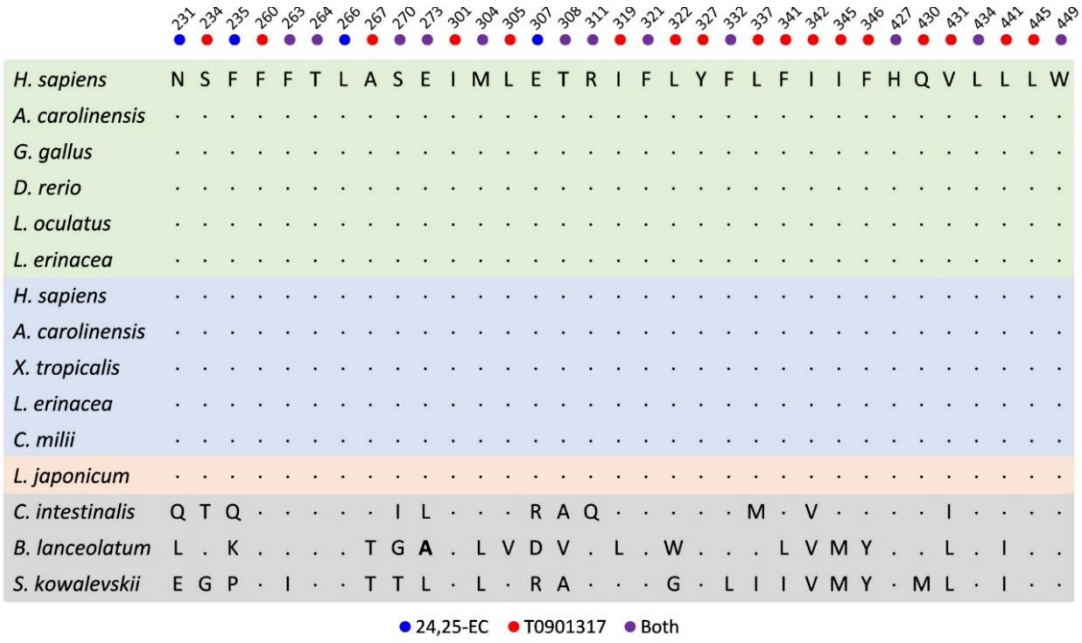


Figure S3.5. Conservation analysis of ligand binding residues of LXRs which interact with T091317 and 24(S),25-epoxycholesterol using human LXRα as reference (Farnegardh et al. 2003; Williams et al. 2003).

References

Farnegardh M, Bonn T, Sun S, Ljunggren J, Ahola H, Wilhelmsson A, Gustafsson JA, Carlquist M. 2003. The three-dimensional structure of the liver X receptor beta reveals a flexible ligand-binding pocket that can accommodate fundamentally different ligands. J Biol Chem. 278:38821-38828. doi: 10.1074/jbc.M304842200.

Williams S, Bledsoe RK, Collins JL, Boggs S, Lambert MH, Miller AB, Moore J, McKee DD, Moore L, Nichols J, et al. 2003. X-ray crystal structure of the liver X receptor beta ligand binding domain: regulation by a histidine-tryptophan switch. J Biol Chem. 278:27138-27143. doi: 10.1074/jbc.M302260200.

CHAPTER 4 - Evolutionary Plasticity in Detoxification Gene Modules: The Preservation and Loss of the Pregnane X Receptor in Chondrichthyes Lineages

International Journal of *Molecular Sciences*

MDPI

Article

Evolutionary Plasticity in Detoxification Gene Modules: The Preservation and Loss of the Pregnane X Receptor in Chondrichthyes Lineages

Elza S. S Fonseca[1,2], Raquel Ruivo[1], André M. Machado[1], Francisca Conrado[1], Boon-Hui Tay[3], Byrappa Venkatesh[3], Miguel M. Santos[1,2] and L. Filipe C. Castro[1,2*]

1 CIIMAR/CIMAR—Interdisciplinary Centre of Marine and Environmental Research, 4450-208 Matosinhos, Portugal; fonseca.ess@gmail.com (E.S.S.F.); ruivo.raquel@gmail.com (R.R.); andre.machado@ciimar.up.pt (A.M.M.); fi.silva@campus.fct.unl.pt (F.C.); santosmafm@gmail.com (M.M.S.)
2 FCUP—Faculty of Sciences, Department of Biology, University of Porto, 4150-177 Porto, Portugal
3 Comparative Genomics Laboratory, Institute of Molecular and Cell Biology, A*STAR (Agency for Science, Technology and Research), Biopolis, Singapore 138673, Singapore; mcblab46@imcb.a-star.edu.sg (B.-H.T.); mcbbv@imcb.a-star.edu.sg (B.V.)
* Correspondence: filipe.castro@ciimar.up.pt; Tel.: +351-22-3401800

Received: 25 February 2019; Accepted: 6 May 2019; Published: 10 May 2019

4. Evolutionary Plasticity in Detoxification Gene Modules: The Preservation and Loss of the Pregnane X Receptor in Chondrichthyes Lineages

4. Abstract

To appraise how evolutionary processes, such as gene duplication and loss, influence an organism's xenobiotic sensitivity is a critical question in toxicology. Of particular importance are gene families involved in the mediation of detoxification responses, such as members of the nuclear receptor subfamily 1 group I (NR1I), the pregnane X receptor (*PXR*), and the constitutive androstane receptor (*CAR*). While documented in multiple vertebrate genomes, *PXR* and *CAR* display an intriguing gene distribution. *PXR* is absent in birds and reptiles, while *CAR* shows a tetrapod-specific occurrence. More elusive is the presence of *PXR* and *CAR* gene orthologs in early branching and ecologically-important Chondrichthyes (chimaeras, sharks and rays). Therefore, we investigated various genome projects and use them to provide the first identification and functional characterization of a Chondrichthyan PXR from the chimaera elephant shark (*Callorhinchus milii*, Holocephali). Additionally, we substantiate the targeted *PXR* gene loss in Elasmobranchii (sharks and rays). Compared to other vertebrate groups, the chimaera *PXR* ortholog displays a diverse expression pattern (skin and gills) and a unique activation profile by classical xenobiotic ligands. Our findings provide insights into the molecular landscape of detoxification mechanisms and suggest lineage-specific adaptations in response to xenobiotics in gnathostome evolution.

Keywords: nuclear receptors; gene loss; detoxification; endocrine disruption

4.1. Introduction

Nuclear receptors (NRs) are central constituents of animal endocrine systems. These ligand-dependent sensors act as transcription factors, regulating key physiological processes including metabolism, development, reproduction and nutrient utilization (Bookout et al. 2006). Importantly, NRs are also directly exploited by xenobiotics, causing numerous examples of physiological impairment (e.g. (Capitão et al. 2018; Santos et al. 2018)). Two critical components of the vertebrate "chemical defensome" are the pregnane X receptor (*PXR*) and the constitutive androstane receptor (*CAR*) (Goldstone et al. 2006). These are part of NR1I subfamily which also includes the vitamin D receptor (*VDR*). *PXR* and *CAR* were originally identified as xenobiotic sensors, since they regulate genes involved in drug metabolism such as phase I cytochrome P450

(e.g. *CYP3A4*, *CYP2B6* and *CYP2C*), phase II transferases (e.g. UDP glucuronosyl transferase and glutathione-S -transferase) and drug transporters. Moreover, *PXR* is notoriously involved in other metabolic processes including energy homeostasis, inflammatory responses, cell proliferation, apoptosis and tumour development (di Masi et al. 2009; Zhuo et al. 2014; Tolson & Wang 2010). The taxonomic distribution of *VDR*/*PXR*/*CAR* gene orthologs is remarkably mutable (Cruzeiro et al. 2016; Kim et al. 2017; Mathäs et al. 2012; Eide et al. 2018). In vertebrate species, *VDR* is found in both cyclostomes (lampreys) and gnathostomes (Kollitz et al. 2015); *CAR* occurs in tetrapods (Mathäs et al. 2012; Zhao et al. 2015); while *PXR* genes have been described and characterized in amphibians (Mathäs et al. 2012) and mammals (Mathäs et al. 2012). On the other hand, teleost genomes such as zebrafish (Danio rerio), also retain *PXR*, but this is not an universal condition found throughout teleost lineages (Eide et al. 2018; Mathäs et al. 2012), consistent with the highly derived nature of their genomes (Ravi & Venkatesh 2018). Importantly, synteny supports the hypothesis that the absence of *PXR* in birds, reptiles and some teleosts, as well as *CAR* in ray-finned fish is due to secondary gene loss (Mathäs et al. 2012). Genome comparisons between human and teleost *PXR*, *CAR*, and *VDR* orthologous genomic regions further implicates whole-genome duplications (2R) as the underlying cause of the *NR1I* gene expansion (Mathäs et al. 2012; Bertrand et al. 2004). These observations suggest that *VDR*, *PXR* and *CAR* first appeared in the ancestors of vertebrates and should be present in early lineage genomes such as Chondrichthyes (cartilaginous fish). Consistently, *VDR* has been described and functionally characterized in cartilaginous fishes (Kollitz et al. 2015). *PXR* and *CAR* orthologs have not been described in Chondrichthyes, although the presence of the former has been suggested (Mathäs et al. 2012). Here we thoroughly investigate the gene repertoire of the NR1I subfamily, central components of the “chemical defensome”, in Chondrichthyes. Cartilaginous fishes are divided into two branches, the Holocephali (Chimaeras) and Elasmobranchii (sharks, rays and skate). Together, they are a highly diversified group of early branching vertebrates, representing important components of aquatic ecosystems and food webs, and thus are key ecological indicators (Didier 2002; Walker et al. 2008).

4.2. Material and Methods

4.2.1. Sequence Retrieval and Phylogenetic Analysis

Amino acid sequences were retrieved through blast searches in the publicly available genome databases, using as reference annotated human VDR, PXR and CAR sequences. Sequence sampling included major vertebrate lineages: mammals (*Homo*

sapiens, Mus musculus, *Rattus norvegicus*, *Oryctolagus cuniculus*, *Sus scrofa*, *Bos taurus*), birds (*Gallus gallus, Anas platyrhynchos*), reptiles (*Anolis carolinensis*), amphibians (*Xenopus tropicalis*), euteleostei (*Danio rerio*, *Cyprinus carpio*, *Oryzias latipes*, *Oreochromis niloticus*) osteoglossomorpha (*Scleropages formosus*), holostei (*Lepisosteus oculatus*), chondrichthyans (Elasmobranchii: *Leucoraja. erinacea*, *Chiloscyllium punctatum*, *Scyliorhinus torazame*, *Rhyncodon typus*; Chimaera: *Callorhinchus milii*), cyclostomes (*Petromyzon marinus*) and invertebrates (*Ciona intestinalis*). Retrieved sequences and corresponding protein accession numbers are listed in the **Table S4.1**. A multiple alignment of the retrieved sequences was obtained with MAFFT alignment software (Katoh & Toh 2010) using default parameters. The final sequence alignment contained 52 sequences and 659 positions was used to construct a phylogenetic tree with MrBayes v 3.2.3 (CIPRES, San Diego, CA, USA; http://www.phylo.org/index.php/) sited in the CIPRES Science Gateway V3.3 (Miller et al. 2015). The Bayesian analyses was performed under a mixed substitution model with two independent runs of four chains (one cooled and 3 heated) for 1×10^6 generations and trees were sampled every 500 generations with a burnin set to 0.25 until the average standard deviation of the split frequencies remained <0.01. The statistical support for each branch is indicated at the nodes and expressed as Bayesian posterior probabilities (Nascimento et al. 2017). FigTree v1.3.1 was used to visualize the tree. Geneious® v7.1.7 (Biomatters Ltd., Auckland, New Zealand) was used to calculate the amino acid identity between human, mouse, zebrafish and elephant shark PXRs.

4.2.2. Synteny Analysis

The genomic region containing the human *PXR*, *CAR* and *VDR* genes were localized at chromosomes 3 (119.78Mb), 1 (161.22Mb) and 12 (47.84Mb), respectively. The human neighbouring genes as well the respective *loci* (GRCh38.p7) were collected from the GenBank database and used as reference to assemble the synteny maps of zebrafish, elephant shark, cloudy catshark, brownbanded bamboo shark and whale shark. To find the orthologs genes in the genomes of zebrafish (GRCz10), elephant shark (Callorhinchus_milii-6.1.3) and whale shark (ASM164234v2), we perform a BLAST of the human neighbour genes; four flanking genes from both sides of each target gene were considered, against the above-mentioned genomes. In the case of the cloudy catshark (Storazame_v1.0), brownbanded bamboo shark (Cpunctatum_v1.0), and the new version of whale shark (Rtypus_kobe_v1.0), the flanking genes found in elephant shark, as well the previous reference genes of human, were blasted (blast-n: -word_size 10, -outfmt 6, -num_threads 50) against the three recently built elasmobranchs genomes (https://figshare.com/authors/Phyloinformatics_Lab_in_RIKEN_Kobe/4815111).

Importantly, we used the new version of the whale shark genome (Rtypus_kobe_v1.0) to complete the previous syntenic map of the whale shark (ASM164234v2). Next, we manually inspected the blast-n results and using the qstart, qend, sstart, send and bit score options of outfmt6 format of blast software, reconstructed the structure for each gene. To confirm the neighbors homology in non-annotated genomes (*C. punctatum*, *S. torazame* and *R. typus* new version), we used the following strategy: (1) .fasta and .gff files of each genome were used to extract the predicted coding region of each homolog candidate gene (if the blast approach detected the gene fragmentated in different scaffolds, we considered the biggest); (2) after, we performed reciprocal blast-n (with megablast and dc-megablast algorithm) searches of all candidates genes in Nucleotide *database of NCBI (NT-NCBI);* (3) if each candidate gene matched against the expected genes (references above mentioned, or orthologs genes in other species), it was kept and used to build the synteny maps.

4.2.3. Construction of Plasmid Vectors

The *PXR* hinge region and ligand binding domain (LBD) was isolated from zebrafish using a polymerase chain reaction (PCR) approach with the specific primers F: 5'-atttCTAGAATGAAGAGAGAGCTGATCATGTC-3' and R: 5'- aattGGTACCCTTTG TGAGGACTTAGGTGTC-3' and the Phusion Flash master mix (Thermo Fisher Scientific, Waltham, MA, USA), according to the protocol from the supplier. The hinge and LBD of the elephant shark *PXR* was synthesized by IDT - Integrated DNA Technologies, Inc. (https://eu.idtdna.com/pages/home). Both *PXR*s were digested with *Xba*I and *Kpn*I restriction enzymes (NZYtech) and ligated to pBIND (AF264722; Promega) with T4 ligase (Promega, Madison, WI, USA) to produce a GAL4-LBD "chimeric" receptor. The chimeric receptor produces a hybrid protein that contains the Gal4 DNA binding domain (DBD) and acts on an upstream activation sequence (UAS) response element. Plasmid sequences were confirmed using Sanger sequencing (Eurofins GATC, Constance, Germany).

4.2.4. Gene Expression

Total RNA was extracted from the following tissues of elephant shark using Trizol reagent (Gibco BRL): brain, gill, heart, intestine, kidney, liver, muscle, ovary, skin and testis. One microgram of total RNA from each tissue was reverse-transcribed into cDNA with Superscript II (Invitrogen, Carlsbad, CA, USA) and used as a template for RT-PCR. The following primer pair which spans an intron was used to amplify elephant shark PXR: PXR_F, 5'-TGGAAGATCTCCTGGAAGCACATC-3' and PXR_R, 5'-GAAGTTACGCTG GAGCTTGTAGTC-3'. Actin was amplified as an internal control to verify the integrity of

cDNA using the primers: Actin_F, 5'-GGTATTGTCACCAACTGGGAC-3' and Actin_R, 5'-AGATGGG CACAGTGTGGGTG-3. The PCR cycles comprised an initial denature step of 95 °C for 30 s followed by 35 cycles of 95 °C for 10 s, 60 °C for 30 s and 72 °C for 30 s and a final elongation step of 72 °C for 5 min.

4.2.5. Transfection and Transactivation Assays

Cell culture and transactivation assays were performed as described in Fonseca et al. 2017. All ligands used (E2, EE2, TNC and BPA) were purchased from Sigma Aldrich (Sintra, Portugal). All compounds were resuspended in dimethyl sulfoxide (DMSO) at a final concentration of 0.1, 1 and 10 µM, and 5, 50 and 100 µM for BPA. Briefly, Cos-1 cells (Sigma, Sintra Portugal) were maintained in DMEM (PAN-Biotech, Aidenbach, Bayern, Germany) supplemented with 10% fetal bovine serum (PAN-Biotech Aidenbach, Bayern, Germany) and 1% penicillin/streptomycin (PAN-Biotech) at 37 °C with a humidified atmosphere and 5% CO_2. Cells were seeded in 24-well culture plates and after 24 hours cells were transfected with 0.5 µg of pBIND constructs (pBIND-CmiPXRLBD or pBIND-DrePXRLBD) and 1µg of pGL4.31[luc2P/GAL4UAS/Hygro] luciferase reporter vector (DQ487213; Promega, Madison, WI, USA), containing five UAS elements upstream the firefly luciferase reporter gene, using lipofectamine 2000 reagent (Invitrogen, Carlsbad, CA, USA), in Opti-MEM (Gibco, Carlsbad, CA, USA), according manufacturer's indications. After 5 h of incubation, transfection media was replaced with medium containing the test compounds (E2, EE2 and TNC - 1, 1 and 10µM, and BPA - 5, 50 and 100 µM) dissolved in DMSO (0.1%). Cells were lysed 24 h after transfection and assayed for Firefly luciferase (reporter pGL4.31) and *Renilla* luciferase (pBIND) activities with Dual-Luciferase Reporter Assay System (Promega, Madison, WI, USA), according to the manufacturer's instructions. All transfections were performed with two technical replicates per condition in three independent assays. The results were expressed as fold-induction resulting from the ratio between luciferase (reporter pGL4.31) and *Renilla* (internal control for transfection efficiency luminescent activity), and then normalized by the DMSO control. Transactivation data was presented as means of the normalized values (n = 3) and the bars with standard error of the mean (SEM) from the three separate experiments.

4.2.6. Statistical Analysis

The means of the technical replicates were used for statistical analysis with one-way analysis of variance (ANOVA), followed by the Holm-Sidak method in SigmaPlot 11.0 software (Systat Software Inc., San Jose, CA, USA). The level of significance was set to 0.05.

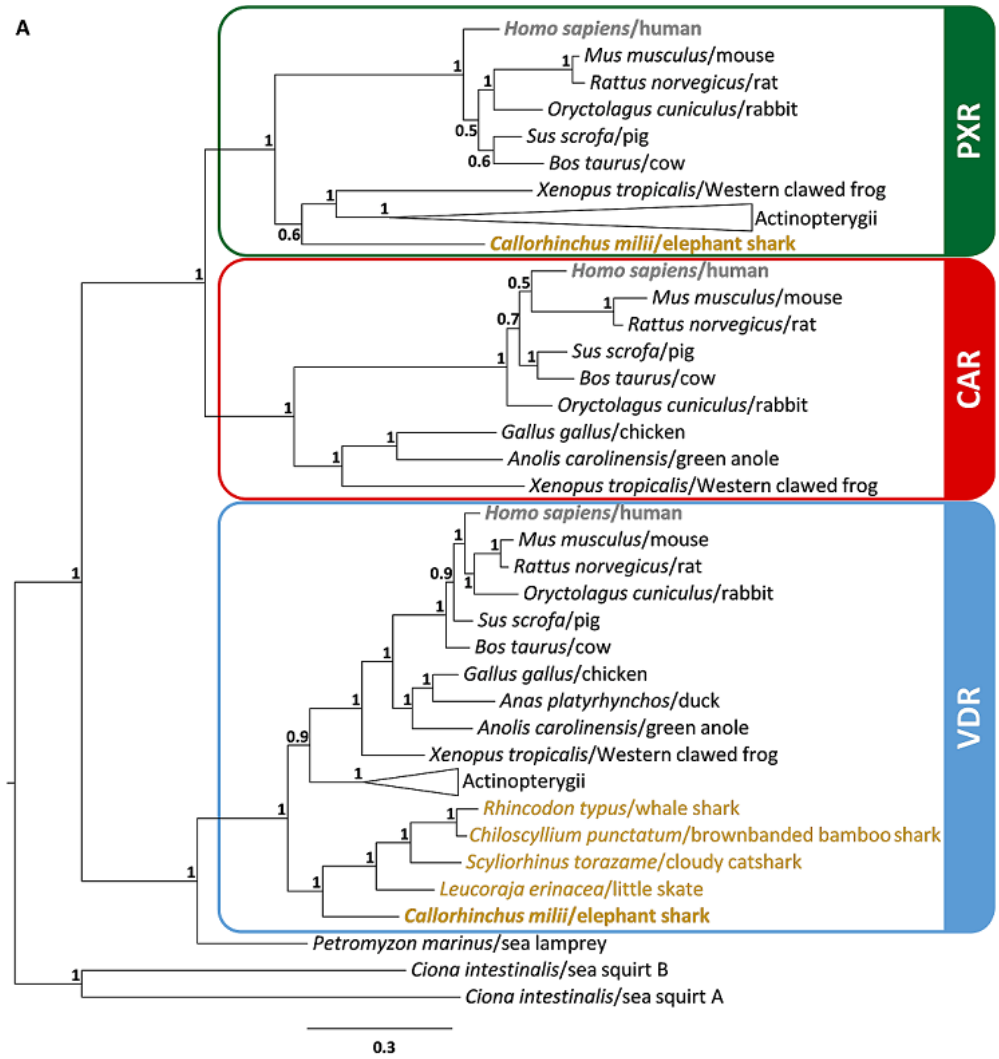
4.3. Results

4.3.1. Identification of Nuclear Receptor Subfamily 1 Group I (NR1I) Ortholog Genes in Chondrichthyes

To determine the gene complement of *VDR/PXR/CAR*-like genes in Chondrichthyes species, we examined five genome datasets from two subclasses: Holocephali (chimaeras) and Elasmobranchii (sharks and rays) (King et al. 2011; Wyffels et al. 2014; Read et al. 2017; Venkatesh et al. 2014; Hara et al. 2018). Our search identified one or two genes with similarity to *VDR/PXR/CAR* genes in the elephant shark (*Callorrhyncus milii*), little skate (*Leucoraja erinacea*), cloudy catshark (*Scyliorhinus torazame*), brownbanded bamboo shark (*Chiloscyllium punctatum*) and whale shark (*Rhyncodon typus*). To establish the orthology of the retrieved sequences, we performed phylogenetic analysis (**Figure 4.1A**), including described *VDR/PXR/CAR* gene sequences from mammals, birds, reptiles, amphibians, teleosts, lepisosteiformes, cyclostomes and tunicates (**Table S4.1**). Three statistically supported sequence clades were retrieved, consistent with the separation into *VDR*, *PXR* and *CAR* genes (**Figure 4.1A**). We next compared the two critical functional domains of the *CmiPXR*, the DNA binding domain (DBD) and the ligand binding domain (LBD). *CmiPXR* shares around 70% of sequence identity to other *PXR*-DBDs (**Figure 4.1B**), with the LBD identity values displaying significantly lower values (**Figure 4.1B**).

4.3.2. Synteny Analysis of NR1I Ortholog Genes

To further verify the orthology of these novel gene sequences and to discriminate between true gene loss or absence of sequencing data, we next verified the genomic location of *VDR*, *PXR* and *CAR* at the syntenic locations in the genomes of the elephant shark, cloudy catshark, brownbanded bamboo shark and whale shark, using the human and zebrafish gene *loci* composition as reference as shown in **Figure 4.2**. The scattered assembly and small contiguous size of the current little skate genome (LER_WGS_1-GCA_000238235.1) impeded a consistent comparative analysis at this stage. The *PXR* gene from elephant shark is flanked by *MAATS1* and *GSK3B* genes (scaffold NW_006890095.1 3.95Mb), and the overall *locus* composition is similar to that of other vertebrate species (**Figure 4.2**). In the Elasmobranchii species analysed here and despite the global synteny conservation, no intervening *PXR*-like sequence is found between *MAATS1* and *GSK3B* (brownbanded bamboo shark – scaffold scf_chipu00000056; whale shark – scaffold scf_rhity00002454; cloudy catshark –



B

DNA-binding domain

Ligand-binding domain	Hsa	Mmu	Dre	Cmi
Hsa		94.3	73.9	76.1
Mmu	78.3		69.3	73.9
Dre	43.9	45.9		67.8
Cmi	46.5	44.4	44.9	

Figure 4.1. The NR1I gene repertoire in Chondrichthyes. **(A)** Bayesian phylogenetic tree of vitamin D receptor (*VDR*), pregnane X receptor (*PXR*) and constitutive androstane receptor (*CAR*) genes. The numbers at the nodes represent the statistical support expressed in Bayesian posterior probabilities. Actinopterygii are represented by *Danio rerio* (zebrafish), *Cyprinus carpio* (European carp), *Oryzias latipes* (medaka), *Oreochromis niloticus* (Nile tilapia), *Scleropages formosus* (Asian arowana) and *Lepisosteus oculatus* (spotted gar); Chondrichthyes are highlighted in yellow: Elasmobranchii (little skate, brownbanded bambooshark, whale shark and cloudy catshark) and Holocephali (elephant shark) in bold. **(B)** Percentage of amino acids identity of DNA and ligand binding domains between human (Hsa), mouse (Mmu), zebrafish (Dre) and elephant shark (Cmi) PXRs.

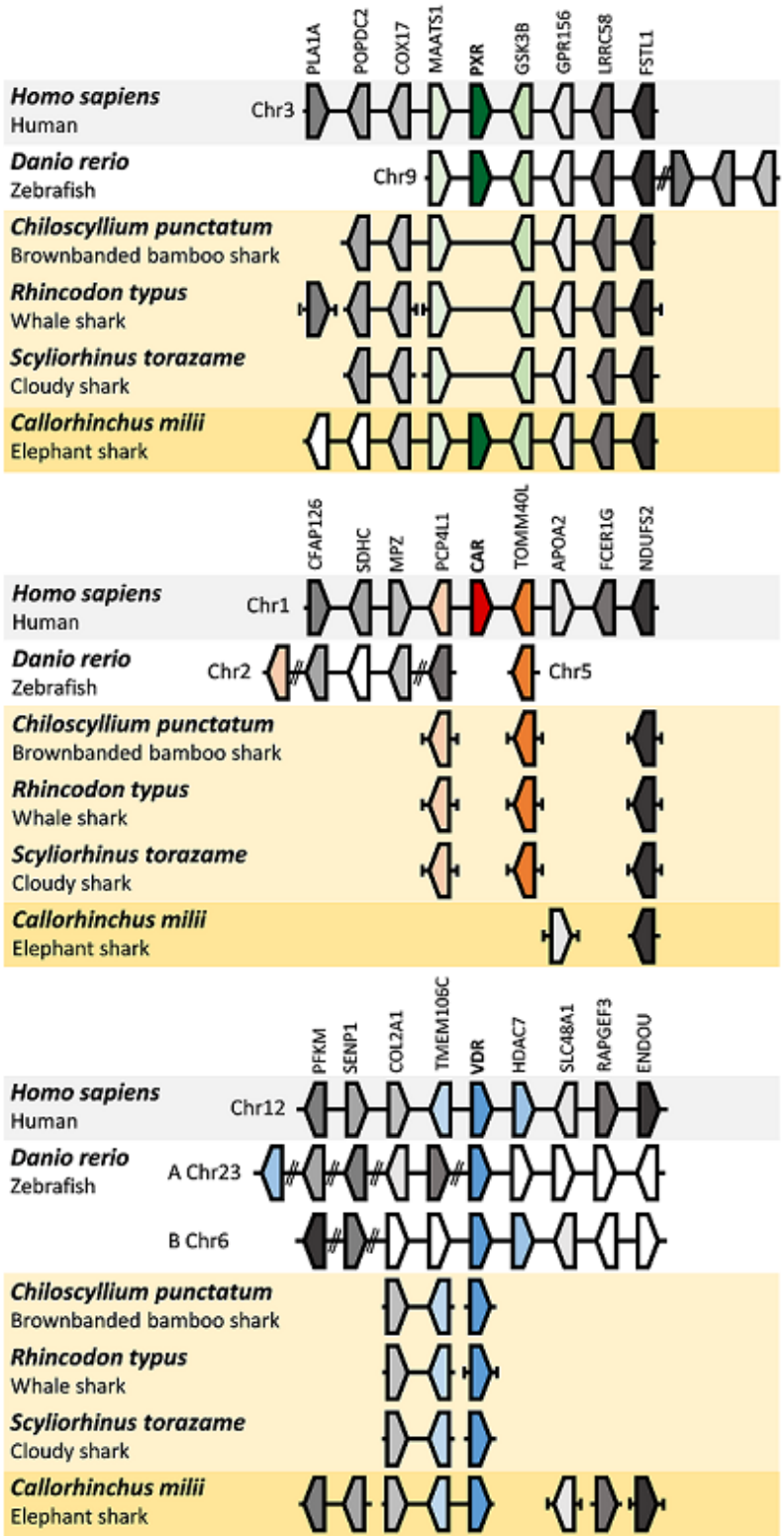


Figure 4.2. Schematic representation of syntenic pregnane X receptor (*PXR*), constitutive androstane receptor (*CAR*), and vitamin D receptor (*VDR*) regions. Human, zebrafish, brownbanded bamboo shark, whale shark, cloudy catshark and elephant shark genomic locations of *VDR*, *PXR* and *CAR* genes. The genomic locations of human *PXR*, *CAR* and *VDR* were used as reference and highlighted in grey. The double slashes in zebrafish chromosomes symbolise discontinuity in the chromosome representation. The scaffolds with chondrichthyan orthologs were highlighted in yellow (dark yellow for elephant shark).

scaffold scf_scyto00010339) (**Figure 4.2**, **Table S4.2- S4.4**). In the case of the *VDR locus*, we searched the scaffolds containing the human *VDR* flanking genes *TMEM106C* and *HDAC7*, but no *HDAC7* ortholog was found in any of the Chondrichthyes genomes. In the elephant shark, *VDR* was found in the same scaffold than *TMEM106C* (scaffold NW_006890370.1 105.7kb), contrary to Elasmobranchii *VDRs* (brownbanded bamboo shark – scaffold scf_chipu00001415 28.2kb; whale shark – scaffold NW_018047310.1 2.9kb; cloudy catshark – scaffolds scf_scyto00007144 and scf_scyto00012969), which were found on different scaffolds than the *TMEM106C* orthologs (brownbanded bamboo shark – scaffold scf_chipu00001599; whale shark – scaffold NW_018032445.1; cloudy catshark – scaffold scf_scyto00007676), probably due to missing sequencing data for the intervening genomic region (**Figure 4.2**, **Table S4.2-S4.4**). In Chondrichthyes, the *CAR locus* is dispersed in comparison to humans (**Figure 4.2**, **Table S4.2- S4.4**).

4.3.3. Gene Expression Analysis of the Elephant Shark Pregnane X Receptor (*PXR*)

Next, we investigated the gene expression profile of the *CmiPXR* in a tissue panel. Our analysis indicates that elephant shark gene ortholog displays a unique pattern, with restricted expression in the skin and gills (**Figure 4.3**).

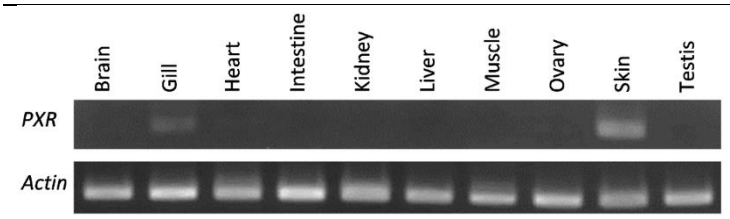


Figure 4.3. *PXR* expression pattern on an elephant shark tissue panel.

4.3.4. Transactivation Assays of *CmiPXR*

Given the differences on the LBD sequence of *CmiPXR*, we explored the capacity to transactivate gene expression in the presence of classical *PXR* ligands from different chemical categories: the natural and the synthetic steroid hormone 17β-estradiol (E2) and 17α-ethinylestradiol (EE2) respectively, and the environmental contaminants trans-nonachlor (TNC) and bisphenol A (BPA), using a mammalian cell-based activation assay and the zebrafish *PXR* (*DrePXR*) as control. Both E2 and EE2 significantly activated ($P<0.05$) *DrePXR* and *CmiPXR* at high concentrations (**Figure 4.4**). Regarding the effect

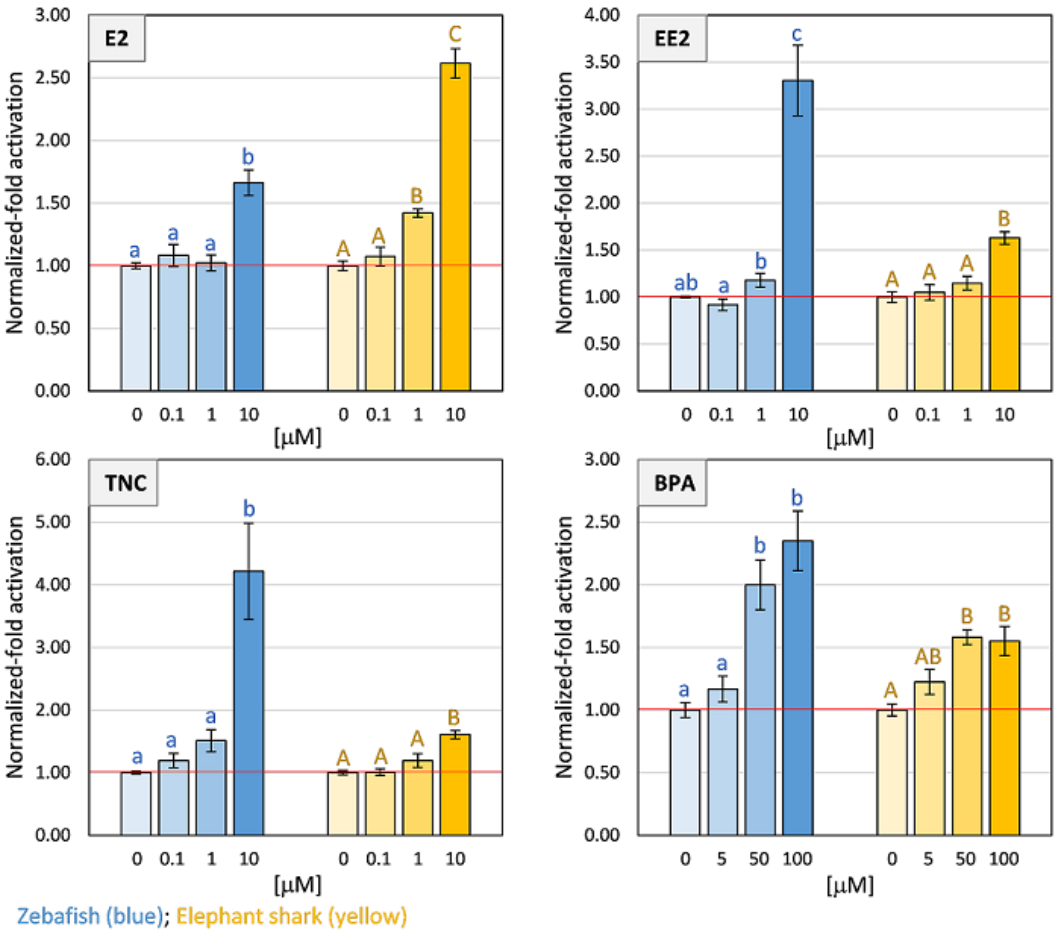


Figure 4.4. Transactivation activity of luciferase reporter gene performed in COS-1 cells mediated by the PXR ligand binding domain (LBD) pBIND constructs in the presence of 17β-estradiol (E2), 17α-ethinylestradiol (EE2), trans-nonachlor (TNC) and bisphenol A (BPA). Data represent means ± S.E.M. from three separate experiments (n = 3). The results were normalized to the control condition (DMSO without ligand). The red horizontal line represents the level of the control condition (no fold activation). Significant differences between the tested concentrations and the solvent control were inferred using one-way ANOVA. The lowercase letters (zebrafish) and the uppercase letters (elephant shark) were used to mark significant differences ($p < 0.05$).

of the two environmental pollutants tested, both *DrePXR* and *CmiPXR* were significantly activated (*P*<0.05) when exposed to the highest tested concentration of TNC and the two highest concentrations of BPA (**Figure 4.4**).

4.4. Discussion

By performing a comprehensive search of the *NR1I* gene repertoire in early branching gnathostome genomes, we unfold the evolutionary history of this fundamental component of detoxification response. Notably, we were able to deduce that *PXR* while present in the elephant shark, a Holocephali, has been most likely lost in the investigated Elasmobranchii species. The two newly identified genes in the elephant shark fall into the *PXR* and *VDR* clades, while the single gene identified in little skate, cloudy catshark,

brownbanded bamboo shark and whale shark are *bona fide VDR* orthologs. The orthology of the new gene sequences found in this study was further confirmed with the syntenic analysis of *VDR*, *PXR* and *CAR* locations in the genomes of the elephant shark, cloudy catshark, brownbanded bamboo shark and whale shark. Regarding the *CAR locus*, its dispersed composition in Chondrichthyes compared to human, impedes a formal conclusion on the loss of *CAR* in these species although this is the likeliest scenario. Additionally, gene orthologs of both *PXR* and *CAR* were also not found in the two currently available genomes of cyclostomes (sea lamprey and Japanese lamprey, not shown). Furthermore, our analysis of amino acid identity between DBD and LBD of human, mouse, zebrafish and elephant shark PXRs is consistent with previous studies for other species (Bainy et al. 2013; Krasowski et al. 2005b). As expected, despite the substantial variation in sequence identity among vertebrates, we observe that the PXR-DBD is more conserved than LBD between species. This suggests that different PXR should recognize similar response elements in the promotor of target genes, but their activation might be triggered upon binding to different ligands. The large and flexible ligand binding pocket of PXR, allows this receptor to accommodate a huge and diverse ligand range such as endogenous ligands (5β-pregnane, progesterone, testosterone, lithocholic acids, 17β-estradiol), antibiotics, drugs, carcinogens and an array of environmental pollutants (Goodwin et al. 2003; Kliewer et al. 1998; Krasowski et al. 2005a; Lehmann et al. 1998; Moore et al. 2000). The occurrence of different gene complements of the *NR1I* subfamily in Chondrichthyes parallels similar findings in other vertebrate lineages (Eide et al. 2018; Handschin et al. 2000; Moore, LB et al. 2002). Recently, an extensive investigation into 76 fish genomes put forward that approximately half of these species have lost *PXR (Eide et al. 2018),* in line with the description made here in cartilaginous fishes. Moreover, xenobiotic exposure experiments with classic xenobiotic PXR ligands in cod (*PXR*-absent) did not show a clear transcription activation of genes coding for P450 cytochrome enzymes (*cyp3a*), as observed in mice or zebrafish (Eide et al. 2018). In addition, promotor analysis raised the interesting possibility the *PXR*-absent species might have their *cyp3a* and *cyp1a* genes regulated by an unrelated xenobiotic-sensing transcription factor, the Aryl Hydrocarbon Receptor (*AHR*) (Eide et al. 2018). Whether this is the case in Chondrichthyes that have lost *PXR* remains to be investigated. Together, these results suggest that the transcriptional regulation of detoxification gene modules is exceptionally plastic and has been rewired during vertebrate evolution.

The expression of *PXR* gene in elephant shark is confined to skin and gills what is in clear contrast with what is observed in other species. For example, *PXR* gene

transcripts in zebrafish are found in the liver, eye, intestine, brain, heart, and kidney, while in mammals *PXR* is expressed in the liver, gastrointestinal tract, brain and retina (Bainy et al. 2013).

The ligand binding profile of the chimaera *PXR* ortholog is also puzzling, with *CmiPXR* exhibiting a smaller activation for the selected environmental pollutants, in contrast to *PXR* from zebrafish (Milnes et al. 2008). Yet, distinct sensitivities towards xenobiotics have been reported across species (Lille-Langøy et al. 2015; Milnes et al. 2008; Scheer et al. 2010; Sui et al. 2010). Thus, we cannot fully discard the existence of a distinct set of potential *PXR*-interacting xenobiotics, or other unidentified endogenous ligands, for chimaeras. Nonetheless, our transactivation results, together with the unique expression profile, raise the interesting possibility that *CmiPXR* could act as a specialized steroid-like sensor in the skin. In fact, previous studies suggested putative effects of skin expressed *PXR* in humans and rodents: for example, induction of keratinocyte proliferation, immune hyper-responsiveness, modulation of DNA repair mechanisms and overall skin barrier functions (Elentner et al. 2018; Schmuth et al. 2014). On the other hand, it is known that estrogens participate in skin homeostasis, by modulating collagen deposition, wound healing and scarring, and maintaining skin hydration and elasticity (Shah & Maibach 2001). Furthermore, both *PXR* and estrogen have been directly or indirectly linked to fibrous connective tissue equilibrium (Frazier-Jessen et al. 1996; Schmuth et al. 2014). Thus, we hypothesize that *PXR* could play a role in estrogen-dependent skin maintenance in chimaeras, contributing to the peculiar appearance of their smooth, rubbery and scale-less skin.

4.5. Conclusions

In this study, we decipher the early evolution of central components of the vertebrate "*chemical defensome*". Our findings indicate that *PXR* gene orthologs are present in the Holocephali but have been probably lost in Elasmobranchii. Moreover, the chimaera *PXR* gene displays a unique pattern of gene expression. Future studies will be required to dissect the molecular wiring of detoxification gene modules in Chondrichthyes.

Supplementary Materials: Supplementary materials can be found at www.mdpi.com/1422-0067/20/x/xx/s1.

Funding: This research was supported by Project No. 031342 co-financed by COMPETE 2020, Portugal 2020 and the European Union through the ERDF, and by Fundação para a Ciência e a Tecnologia (FCT) through national funds; PhD grant

awarded to Elza Fonseca (SFRH/BD/100262/2014) by FCT. The work in BV lab was supported by the Biomedical Research Council of A*STAR, Singapore.

Author Contributions: Conceived and designed the experiments: EF, RR, AM, MMS, LFCC; Performed the experiments: EF, AM, FC, BT, BV; Analysed the data: EF, RR, AM, FC, BV, MMS, LFCC; Wrote the paper: EF, RR, MMS, LFCC.

Conflicts of Interest: The authors declare no conflict of interest.

4.6. References

Bainy A et al. 2013. Functional characterization of a full length pregnane X receptor, expression in vivo, and identification of PXR alleles, in zebrafish (*Danio rerio*). Aquat Toxicol. 142–143:447–457. doi: 10.1016/j.aquatox.2013.09.014.

Bertrand S et al. 2004. Evolutionary genomics of nuclear receptors: from twenty-five ancestral genes to derived endocrine systems. Mol Biol Evol. 21:1923–1937. doi: 10.1093/molbev/msh200.

Bookout A et al. 2006. Anatomical profiling of nuclear receptor expression reveals a hierarchical transcriptional network. Cell. 126:789–799. doi: 10.1016/j.cell.2006.06.049.

Capitão A et al. 2018. Evolutionary Exploitation of Vertebrate Peroxisome Proliferator-Activated Receptor γ by Organotins. Env. Sci Technol. doi: 10.1021/acs.est.8b04399.

Cruzeiro C et al. 2016. A mollusk VDR/PXR/CAR-like (NR1J) nuclear receptor provides insight into ancient detoxification mechanisms. Aquat Toxicol. 174:61–69. doi: 10.1016/j.aquatox.2016.02.007.

Didier D. 2002. Chimaeras. In: The living marine resources of the Western Central Atlantic. Volume 1: Introduction, mollusks, crustaceans, hagfishes, sharks, batoid fishes, and chimaeras. Carpenter, KE, editor. FAO Species Identification Guide for Fishery Purposes and American Society of Ichthyologists and Herpetologists Special Publication No. 5: Rome, Italy pp. 592–599. http://www.fao.org/tempref/docrep/fao/009/y4160e/y4160e41.pdf.

Eide M et al. 2018. Independent losses of a xenobiotics receptor across teleost evolution. Sci Rep. 8:10404. doi: 10.1038/s41598-018-28498-4.

Elentner A et al. 2018. Epidermal Overexpression of Xenobiotic Receptor PXR Impairs the Epidermal Barrier and Triggers Th2 Immune Response. J Invest Dermatol. 138:109–120. doi: 10.1016/j.jid.2017.07.846.

Fonseca E et al. 2017. LXRalpha and LXRbeta Nuclear Receptors Evolved in the Common Ancestor of Gnathostomes. Genome Biol Evol. 9:222–230. doi: 10.1093/gbe/evw305.

Frazier-Jessen M, Mott F, Witte P, Kovacs E. 1996. Estrogen suppression of connective tissue deposition in a murine model of peritoneal adhesion formation. J Immunol. 156:3036–3042.

Goldstone J et al. 2006. The chemical defensome: environmental sensing and response genes in the *Strongylocentrotus purpuratus* genome. Dev Biol. 300:366–384. doi: 10.1016/j.ydbio.2006.08.066.

Goodwin B et al. 2003. Identification of bile acid precursors as endogenous ligands for the nuclear xenobiotic pregnane X receptor. Proc Natl Acad Sci USA. 100:223–228. doi: 10.1073/pnas.0237082100.

Handschin C, Podvinec M, Meyer U. 2000. CXR, a chicken xenobiotic-sensing orphan nuclear receptor, is related to both mammalian pregnane X receptor (PXR) and constitutive androstane receptor (CAR). Proc Natl Acad Sci U S A. 97:10769–10774. doi: 10.1073/pnas.97.20.10769.

Hara Y et al. 2018. Shark genomes provide insights into elasmobranch evolution and the origin of vertebrates. Nat Ecol Evol. 2:1761–1771. doi: 10.1038/s41559-018-0673-5.

Katoh K, Toh H. 2010. Parallelization of the MAFFT multiple sequence alignment program. Bioinformatics. 26:1899–1900. doi: 10.1093/bioinformatics/bty121.

Kim D et al. 2017. Genome-wide identification of nuclear receptor (NR) genes and the evolutionary significance of the NR1O subfamily in the monogonont rotifer Brachionus spp. Gen Comp Endocrinol. 252:219–225. doi: 10.1016/j.ygcen.2017.06.030.

King B, Gillis J, Carlisle H, Dahn R. 2011. A natural deletion of the HoxC cluster in elasmobranch fishes. Science. 334:1517. doi: 10.1126/science.1210912.

Kliewer S et al. 1998. An orphan nuclear receptor activated by pregnanes defines a novel steroid signaling pathway. Cell. 92:73–82. doi: 10.1016/S0092-8674(00)80900-9.

Kollitz E et al. 2015. Molecular cloning, functional characterization, and evolutionary analysis of vitamin D receptors isolated from basal vertebrates. PLoS One. 10:e0122853. doi: 10.1371/journal.pone.0122853.

Krasowski M, Yasuda K, Hagey L, Schuetz E. 2005a. Evolution of the pregnane X receptor: adaptation to cross-species differences in biliary bile salts. Mol Endocrinol. 19:1720–1739. doi: 10.1210/me.2004-0427.

Krasowski M, Yasuda K, Hagey L, Schuetz E. 2005b. Evolutionary selection across the nuclear hormone receptor superfamily with a focus on the NR1I subfamily (vitamin

D, pregnane X, and constitutive androstane receptors). Nucl Recept. 3:2. doi: 10.1186/1478-1336-3-2.

Lehmann J et al. 1998. The human orphan nuclear receptor PXR is activated by compounds that regulate CYP3A4 gene expression and cause drug interactions. J Clin Invest. 102:1016–1023. doi: 10.1172/JCI3703.

Lille-Langøy R et al. 2015. Environmental contaminants activate human and polar bear (Ursus maritimus) pregnane X receptors (PXR, NR1I2) differently. Toxicol Appl Pharmacol. 284:54–64. doi: 10.1016/j.taap.2015.02.001.

di Masi A, De Marinis E, Ascenzi P, Marino M. 2009. Nuclear receptors CAR and PXR: Molecular, functional, and biomedical aspects. Mol Asp. Med. 30:297–343. doi: 10.1016/j.mam.2009.04.002.

Mathäs M et al. 2012. Evolutionary history and functional characterization of the amphibian xenosensor CAR. Mol Endocrinol. 26:14–26. doi: 10.1210/me.2011-1235.

Miller M et al. 2015. A RESTful API for Access to Phylogenetic Tools via the CIPRES Science Gateway. Evol. Bioinform Online. 11:43–48. doi: 10.4137/EBO.S21501.

Milnes M et al. 2008. Activation of steroid and xenobiotic receptor (SXR, NR1I2) and its orthologs in laboratory, toxicologic, and genome model species. Env. Heal. Perspect. 116:880–885. doi: 10.1289/ehp.10853.

Moore, LB et al. 2002. Pregnane X receptor (PXR), constitutive androstane receptor (CAR), and benzoate X receptor (BXR) define three pharmacologically distinct classes of nuclear receptors. Mol Endocrinol. 16:977–986. doi: 10.1210/mend.16.5.0828.

Moore L et al. 2000. Orphan nuclear receptors constitutive androstane receptor and pregnane X receptor share xenobiotic and steroid ligands. J Biol Chem. 275:15122–15127. doi: 10.1074/jbc.M001215200.

Nascimento F, Reis M, Yang Z. 2017. A biologist's guide to Bayesian phylogenetic analysis. Nat Ecol Evol. 1:1446–1454. doi: 10.1038/s41559-017-0280-x.

Ravi V, Venkatesh B. 2018. The divergent genomes of teleosts. Annu Rev Anim Biosci. 6:47–68. doi: 10.1146/annurev-animal-030117-014821.

Read T et al. 2017. Draft sequencing and assembly of the genome of the world's largest fish, the whale shark: Rhincodon typus Smith 1828. BMC Genomics. 18:532. doi: 10.1186/s12864-017-3926-9.

Santos M, Ruivo R, Capitão A, Fonseca E, Castro L. 2018. Identifying the gaps: Resources and perspectives on the use of nuclear receptor based-assays to improve hazard assessment of emerging contaminants. J Hazard Mater. 358:508–511. doi: 10.1016/j.jhazmat.2018.04.076.

Scheer N, Ross J, Kapelyukh Y, Rode A, Wolf C. 2010. In vivo responses of the human and murine pregnane X receptor to dexamethasone in mice. Drug Metab Dispos. 38:1046–1053. doi: 10.1124/dmd.109.031872.

Schmuth M, Moosbrugger-Martinz V, Blunder S, Dubrac S. 2014. Role of PPAR, LXR, and PXR in epidermal homeostasis and inflammation. Biochim Biophys Acta. 1841:463–473. doi: 10.1016/j.bbalip.2013.11.012.

Shah M, Maibach H. 2001. Estrogen and skin. An overview. Am J Clin Dermatol. 2:143–150. doi: 10.2165/00128071-200102030-00003.

Sui Y et al. 2010. Bisphenol A and its analogues activate human pregnane X receptor. Env. Heal. Perspect. 120:399–405. doi: 10.1289/ehp.1104426.

Tolson A, Wang H. 2010. Regulation of drug-metabolizing enzymes by xenobiotic receptors: PXR and CAR. Adv Drug Deliv Rev. 62:1238–1249. doi: 10.1016/j.addr.2010.08.006.

Venkatesh B et al. 2014. Elephant shark genome provides unique insights into gnathostome evolution. Nature. 505:174–179. doi: 10.1038/nature12826.

Walker T et al. 2008. Rapid assessment of sustainability for ecological risk of shark and other chondrichthyan bycatch species taken in the Southern and Eastern Scalefish and Shark Fishery. Queenscliff, Australia http://frdc.com.au/Archived-Reports/FRDC Projects/2002-033-DLD.pdf.

Wyffels J et al. 2014. SkateBase, an elasmobranch genome project and collection of molecular resources for chondrichthyan fishes. F1000Res. 3:191. doi: 10.12688/f1000research.4996.1.

Zhao Y, Zhang K, Giesy J, Hu J. 2015. Families of nuclear receptors in vertebrate models: characteristic and comparative toxicological perspective. Sci Rep. 5:8554. doi: 10.1038/srep08554.

Zhuo W et al. 2014. Role of pregnane X receptor in chemotherapeutic treatment. Cancer Chemother Pharmacol. 74:217–227. doi: 10.1007/s00280-014-2494-9.

4.7. Supplementary Material

Table S4.1. List of sequences used for phylogenetic reconstruction of *VDR*, *PXR* and *CAR* genes and corresponding accession numbers.

Species	NR1I1/VDR	NR1I2/PXR	NR1I3/CAR
Homo sapiens	NP_000367.1	NP_003880.3	NP_001070950.1
Mus musculus	NP_033530.2	NP_035066.1	NP_033933.2
Rattus norvegicus	NP_058754.1	NP_443212.1	NP_075230.1
Sus scrofa	NP_001090883.1	NP_001033094.1	NP_001033085.1
Bos taurus	NP_001161404.2	NP_001096696.1	NP_001073236.1
Oryctolagus cuniculus	XP_017194999.1	NP_001075536.1	XP_017201220.1
Gallus gallus	NP_990429.1	-	NP_990033.1
Anas platyrhynchos	XP_021122847.1	-	-
Anolis carolinensis	XP_008101719.1	-	XP_003230590.2
Xenopus tropicalis	XP_002935703.1	NP_001091887.1	ADW81978.1
Danio rerio A	NP_570994.1	XP_005167477.1	-
Danio rerio B	NP_001153457.1		-
Cyprinus carpio A	XP_018939074.1	APX55174.1	-
Cyprinus carpio B	XP_018951347.1		-
Oryzias latipes A	NP_001121988.1	ABV29345.1	-
Oryzias latipes B	NP_001121989.1		-
Oreochromis niloticus A	XP_005454147.1	NP_001269825.1	-
Oreochromis niloticus B	XP_003441588.1		-
Scleropages formosus A	XP_018615991.1	XP_018599579.1	-
Scleropages formosus B	XP_018581562.1		-
Lepisosteus oculatus	XP_015199783.1	XP_006639043.1	-
Leucoraja erinacea	AIM62165.1	-	-
Chiloscyllium punctatum	VDR_Cpunctatum[1]	-	-
Rhyncodon typus	VDR_Rtypus[2]	-	-
Scyliorhinus torazame	VDR_Storazame[3]	-	-
Callorhinchus milii	XP_007908698.1	XP_007894039.1	-
Petromyzon marinus	AAP05810.1	-	-
Ciona intestinalis A	AHB39788.1	-	-
Ciona intestinalis B	NP_001037831.1	-	-

[1,3]These sequences were built bioinformatically, using the .fasta and gff files of *C. punctatum* and *S. torazame* species.

[2]This sequence were built bioinformatically, using the .fasta and gff files of the new version of *R. typus* species (Rtypus_kobe_v1.0), the previous genome (ASM164234v2) and the raw data of the PRJNA255419 bioproject.

Table S4.2. Blast-n output of human, elephant shark and whale shark (first version) genome sequences against the new version of the whale shark genome. The blast-n valeus were used to reconstruct the synteny of the *locus* in three target *PXR*, *CAR* and *VDR.*

Bait Species	Accession Number	Gene Symbol	Scaffolds of *R. typus* Genome	ID%	Length	Gaps	qstart	qend	sstart	send	E-valeu	Bitscore	Strand
R. typus	XM_020529942.1	*COX17*	scf_rhity00047988	100.000	125	0	1	125	871	995	7.24e-59	231	-
R. typus	XM_020529942.1	*COX17*	scf_rhity00001107	100.000	86	0	125	210	4203	4288	3.47e-37	159	-
R. typus	XM_020529932.1	*POPDC2*	scf_rhity00001107	100.000	494	0	1	494	28051	28544	0.0	913	-
R. typus	XM_020529932.1	*POPDC2*	scf_rhity00001107	99.420	345	1	597	940	38885	39229	1.54e-176	625	-
R. typus	XM_020529932.1	*POPDC2*	scf_rhity00001107	100.000	75	0	940	1014	46552	46626	2.44e-30	139	-
R. typus	XM_020521822.1	*MAATS1*	scf_rhity00035281	100.000	158	0	1	158	753	910	3.74e-76	292	+
R. typus	XM_020521822.1	*MAATS1*	scf_rhity00002454	98.864	88	1	155	242	1522	1608	5.13e-35	156	+
R. typus	XM_020521822.1	*MAATS1*	scf_rhity00002454	97.015	67	2	234	300	14663	14727	1.13e-21	111	+
R. typus	XM_020521822.1	*MAATS1*	scf_rhity00002454	100.000	177	0	300	476	19096	19272	1.03e-86	327	+
R. typus	XM_020521822.1	*MAATS1*	scf_rhity00002454	100.000	180	0	475	654	21188	21367	2.21e-88	333	+
R. typus	XM_020521822.1	*MAATS1*	scf_rhity00002454	99.379	161	0	651	811	28853	29013	3.74e-76	292	+
R. typus	XM_020521822.1	*MAATS1*	scf_rhity00002454	100.000	188	0	810	997	42870	43057	7.88e-93	348	+
R. typus	XM_020521822.1	*MAATS1*	scf_rhity00002454	100.000	142	0	994	1135	50626	50767	2.94e-67	263	+
R. typus	XM_020521822.1	*MAATS1*	scf_rhity00002454	100.000	120	0	1133	1252	58392	58511	4.98e-55	222	+
R. typus	XM_020521822.1	*MAATS1*	scf_rhity00002454	100.000	74	0	1250	1323	66112	66185	1.86e-29	137	+
R. typus	XM_020521822.1	*MAATS1*	scf_rhity00002454	100.000	148	0	1323	1470	67129	67276	1.36e-70	274	+
R. typus	XM_020521822.1	*MAATS1*	scf_rhity00002454	100.000	226	0	1469	1694	69016	69241	5.92e-114	418	+
R. typus	XM_020521822.1	*MAATS1*	scf_rhity00002454	100.000	152	0	1690	1841	70020	70171	8.11e-73	281	+
R. typus	XM_020521822.1	*MAATS1*	scf_rhity00002454	98.958	96	1	1842	1936	71091	71186	1.83e-39	171	+
R. typus	XM_020521822.1	*MAATS1*	scf_rhity00002454	100.000	181	0	1932	2112	73561	73741	6.13e-89	335	+
H. sapiens / C . milii	NM_003889.3 / XM_007910507.1	***PXR***	Not Found	-	-	-	-	-	-	-	-	-	-
R. typus	XM_020510094.1	*GSK3B*	scf_rhity00002454	100.000	89	0	1	89	267358	267270	5.06e-38	165	-
R. typus	XM_020510094.1	*GSK3B*	scf_rhity00002454	100.000	194	0	89	282	201200	201007	2.16e-96	359	-
R. typus	XM_020510094.1	*GSK3B*	scf_rhity00002454	100.000	84	0	283	366	181541	181458	3.04e-35	156	-
R. typus	XM_020510094.1	*GSK3B*	scf_rhity00002454	100.000	113	0	365	477	158782	158670	2.30e-51	209	-
R. typus	XM_020510094.1	*GSK3B*	scf_rhity00002454	100.000	131	0	478	608	150157	150027	2.27e-61	243	-
R. typus	XM_020510094.1	*GSK3B*	scf_rhity00002454	99.107	112	0	607	718	134673	134562	3.86e-49	202	-
R. typus	XM_020510094.1	*GSK3B*	scf_rhity00002454	100.000	103	0	713	815	131647	131545	8.35e-46	191	-
R. typus	XM_020510094.1	*GSK3B*	scf_rhity00002454	100.000	100	0	812	911	130814	130715	3.88e-44	185	-
R. typus	XM_020510094.1	*GSK3B*	scf_rhity00002454	100.000	189	0	908	1096	121245	121057	1.30e-93	350	-
R. typus	XM_020510094.1	*GSK3B*	scf_rhity00002454	99.038	104	0	1095	1198	118094	117991	1.08e-44	187	-
R. typus	XM_020510094.1	*GSK3B*	scf_rhity00002454	100.000	74	0	1193	1266	105787	105714	1.10e-29	137	-
R. typus	XM_020510092.1	*GPR156*	scf_rhity00000674	98.780	82	1	1	81	47016	46935	1.24e-31	145	-
R. typus	XM_020510092.1	*GPR156*	scf_rhity00000674	100.000	141	0	76	216	43310	43170	1.18e-66	261	-
R. typus	XM_020510092.1	*GPR156*	scf_rhity00000674	100.000	114	0	214	327	35559	35446	1.21e-51	211	-
R. typus	XM_020510092.1	*GPR156*	scf_rhity00000674	100.000	128	0	327	454	25807	25680	1.99e-59	237	-
R. typus	XM_020510092.1	*GPR156*	scf_rhity00000674	99.010	101	0	453	553	25552	25452	9.47e-43	182	-
R. typus	XM_020510092.1	*GPR156*	scf_rhity00000674	99.592	245	0	548	792	21279	21035	8.46e-123	448	-
R. typus	XM_020510092.1	*GPR156*	scf_rhity00000674	100.000	165	0	784	948	16350	16186	5.39e-80	305	-
R. typus	XM_020510092.1	*GPR156*	scf_rhity00000674	100.000	1415	0	947	2361	14923	13509	0.0	2614	-
R. typus	XM_020510093.1	*LRRC58*	scf_rhity00000674	100.000	68	0	1	68	159847	159780	1.37e-26	126	-
R. typus	XM_020510093.1	*LRRC58*	scf_rhity00000674	100.000	72	0	66	137	141792	141721	8.17e-29	134	-
R. typus	XM_020510093.1	*LRRC58*	scf_rhity00000674	100.000	131	0	137	267	134871	134741	1.30e-61	243	-
R. typus	XM_020510093.1	*LRRC58*	scf_rhity00000674	100.000	279	0	267	545	134278	134000	6.94e-144	516	-
R. typus	XM_020510093.1	*LRRC58*	scf_rhity00000674	100.000	196	0	543	738	133472	133277	9.57e-98	363	-
C. milii	XM_007895843.1	*FST1L*	scf_rhity00000674	81.928	83	0	463	545	360877	360795	2.09e-09	71.3	-
C. milii	XM_007895843.1	*FST1L*	scf_rhity00000674	83.962	106	0	563	668	358851	358746	7.41e-19	102	-
C. milii	XM_007895843.1	*FST1L*	scf_rhity00000674	84.286	70	0	669	738	357979	357910	7.51e-09	69.4	-
C. milii	XM_007895843.1	*FST1L*	scf_rhity00000674	78.947	76	0	928	1003	352263	352188	7.57e-04	52.8	-
C. milii	XM_007895843.1	*FST1L*	scf_rhity00000674	89.011	91	0	1420	1510	242446	242356	3.42e-22	113	-
C. milii	XM_007895843.1	*FST1L*	scf_rhity00000674	86.916	107	0	1525	1631	237430	237324	2.05e-24	121	-
C. milii	XM_007895843.1	*FST1L*	scf_rhity00000674	80.916	131	0	1677	1807	231151	231021	2.06e-19	104	-
C. milii	XM_007895843.1	*FST1L*	scf_rhity00000674	79.167	120	3	1808	1925	227198	227081	3.47e-12	80.5	-
C. milii	XM_007895843.1	*FST1L*	scf_rhity00000674	92.174	115	1	1930	2044	225312	225199	1.20e-36	161	-
C. milii	XM_007895843.1	*FST1L*	scf_rhity00000674	85.938	128	1	2039	2166	220985	220859	7.30e-29	135	-
C. milii	XM_007895843.1	*FST1L*	scf_rhity00000674	85.246	61	0	2203	2263	218613	218553	3.50e-07	63.9	-
C. milii	XM_007895843.1	*FST1L*	scf_rhity00000674	87.234	47	0	2261	2307	208555	208509	2.10e-04	54.7	-
H. sapiens	NM_001102566.1	*PCP4L1*	scf_rhity00056002	80.000	85	0	97	181	1353	1437	2.75e-08	63.9	-
H. sapiens	NM_001077482.2	***CAR***	Not Found	-	-	-	-	-	-	-	-	-	-
R. typus	XM_020523825.1	*TOMM40L*	scf_rhity00005237	100.000	97	0	292	388	95212	95308	1.47e-42	180	-
R. typus	XM_020523825.1	*TOMM40L*	scf_rhity00005237	100.000	104	0	386	489	96063	96166	1.89e-46	193	-
R. typus	XM_020523825.1	*TOMM40L*	scf_rhity00005237	100.000	110	0	487	596	96544	96653	8.74e-50	204	-
R. typus	XM_020523825.1	*TOMM40L*	scf_rhity00005237	100.000	127	0	593	719	97034	97160	3.10e-59	235	-
R. typus	XM_020523825.1	*TOMM40L*	scf_rhity00027030	98.810	84	0	715	798	2599	2516	1.16e-33	150	-
R. typus	XM_020523825.1	*TOMM40L*	scf_rhity00027030	98.291	117	2	789	905	2068	1954	8.74e-50	204	-
R. typus	XM_020523825.1	*TOMM40L*	scf_rhity00155028	100.000	142	0	897	1038	270	129	1.42e-67	263	-
R. typus	XM_020523967.1	*NDUFS2*	scf_rhity00028733	100.000	223	0	1	223	2002	1780	1.42e-112	412	-
R. typus	XM_020523967.1	*NDUFS2*	scf_rhity00010013	98.361	122	1	218	339	7337	7217	1.54e-52	213	-
R. typus	XM_020523967.1	*NDUFS2*	scf_rhity00010013	100.000	78	0	335	412	6090	6013	5.71e-32	145	-
R. typus	XM_020523967.1	*NDUFS2*	scf_rhity00010013	98.824	85	0	408	492	4874	4790	3.41e-34	152	-
R. typus	XM_020523967.1	*NDUFS2*	scf_rhity00010013	100.000	88	0	488	575	4453	4366	1.58e-37	163	-
R. typus	XM_020523967.1	*NDUFS2*	scf_rhity00010013	100.000	125	0	573	697	4240	4116	4.26e-58	231	-
R. typus	XM_020523967.1	*NDUFS2*	scf_rhity00010013	100.000	133	0	693	825	3956	3824	1.52e-62	246	-
R. typus	XM_020523967.1	*NDUFS2*	scf_rhity00096406	100.000	99	0	824	922	1297	1199	1.21e-43	183	-

R. typus	XM_020523967.1	*NDUFS2*	scf_rhity00096406	100.000	87	0	920	1006	591	505	5.67e-37	161	-
R. typus	XM_020523967.1	*NDUFS2*	scf_rhity00067765	100.000	61	0	1005	1065	1858	1798	1.61e-22	113	-
R. typus	XM_020523967.1	*NDUFS2*	scf_rhity00111163	100.000	39	0	1063	1101	945	907	2.73e-10	73.1	-
C. milii	XM_007910528.1	*COL2A1*	scf_rhity00026090	83.041	171	2	105	274	2000	1831	3.75e-34	154	-
C. milii	XM_007910528.1	*COL2A1*	scf_rhity00004791	91.139	79	0	535	613	9914	9992	2.96e-20	108	-
C. milii	XM_007910528.1	*COL2A1*	scf_rhity00004791	91.489	47	0	612	658	13353	13399	1.81e-07	65.8	-
C. milii	XM_007910528.1	*COL2A1*	scf_rhity00004791	87.273	55	0	658	712	13569	13623	6.50e-07	63.9	-
C. milii	XM_007910528.1	*COL2A1*	scf_rhity00004791	91.071	56	1	766	820	15147	15202	3.00e-10	75.0	-
C. milii	XM_007910528.1	*COL2A1*	scf_rhity00004791	83.929	56	0	918	973	17943	17998	3.91e-04	54.7	-
C. milii	XM_007910528.1	*COL2A1*	scf_rhity00004791	93.182	44	0	973	1016	18765	18808	1.81e-07	65.8	-
C. milii	XM_007910528.1	*COL2A1*	scf_rhity00004791	81.731	104	1	1067	1169	25812	25915	1.39e-13	86.1	-
C. milii	XM_007910528.1	*COL2A1*	scf_rhity00004791	89.130	46	0	1171	1216	26398	26443	3.02e-05	58.4	-
C. milii	XM_007910528.1	*COL2A1*	scf_rhity00004791	79.787	94	0	1216	1309	27715	27808	1.40e-08	69.4	-
C. milii	XM_007910528.1	*COL2A1*	scf_rhity00004791	88.679	53	0	1312	1364	29421	29473	1.81e-07	65.8	-
C. milii	XM_007910528.1	*COL2A1*	scf_rhity00004791	76.768	99	2	1531	1628	35687	35784	3.91e-04	54.7	-
C. milii	XM_007910528.1	*COL2A1*	scf_rhity00004791	80.000	105	0	1886	1990	43049	43153	2.32e-11	78.7	-
C. milii	XM_007910528.1	*COL2A1*	scf_rhity00004791	80.282	71	0	1988	2058	43243	43313	3.91e-04	54.7	-
C. milii	XM_007910528.1	*COL2A1*	scf_rhity00004791	88.000	50	0	2097	2146	44621	44670	8.41e-06	60.2	-
C. milii	XM_007910528.1	*COL2A1*	scf_rhity00004791	76.074	163	0	2533	2695	53858	54020	1.39e-13	86.1	-
C. milii	XM_007910528.1	*COL2A1*	scf_rhity00004791	83.673	98	0	2695	2792	54312	54409	8.29e-16	93.5	-
C. milii	XM_007910528.1	*COL2A1*	scf_rhity00004791	78.761	113	0	2799	2911	62253	62365	8.35e-11	76.8	-
C. milii	XM_007910528.1	*COL2A1*	scf_rhity00004791	84.211	57	0	2909	2965	62456	62512	1.09e-04	56.5	-
C. milii	XM_007910528.1	*COL2A1*	scf_rhity00004791	80.531	113	4	2963	3073	64620	64730	4.99e-13	84.2	-
C. milii	XM_007910528.1	*COL2A1*	scf_rhity00004791	85.455	55	0	3073	3127	64961	65015	3.02e-05	58.4	-
C. milii	XM_007910528.1	*COL2A1*	scf_rhity00004791	85.321	109	0	3127	3235	65724	65832	6.36e-22	113	-
C. milii	XM_007910528.1	*COL2A1*	scf_rhity00004791	80.000	105	1	3285	3389	70146	70249	8.35e-11	76.8	-
C. milii	XM_007910528.1	*COL2A1*	scf_rhity00004791	81.879	298	1	3391	3687	70405	70702	4.65e-63	250	-
C. milii	XM_007910528.1	*COL2A1*	scf_rhity00004791	86.702	188	0	3685	3872	71094	71281	7.89e-51	209	-
C. milii	XM_007910528.1	*COL2A1*	scf_rhity00004791	86.853	251	2	3867	4116	72281	72530	5.93e-72	279	-
C. milii	XM_007910528.1	*COL2A1*	scf_rhity00004791	93.919	148	0	4116	4263	73104	73251	2.82e-55	224	-
R. typus	XM_020524727.1	*TMEM106C*	scf_rhity00004791	99.200	125	0	1	125	99575	99451	1.01e-56	226	+
R. typus	XM_020524727.1	*TMEM106C*	scf_rhity00004791	98.551	69	0	120	188	99519	99451	1.37e-25	122	+
R. typus	XM_020524727.1	*TMEM106C*	scf_rhity00004791	98.592	71	1	162	232	99519	99450	3.80e-26	124	+
R. typus	XM_020524727.1	*TMEM106C*	scf_rhity00004791	97.826	46	0	183	228	99519	99474	8.34e-13	80.5	+
R. typus	XM_020524727.1	*TMEM106C*	scf_rhity00004791	99.333	150	1	223	371	99108	98959	4.61e-70	270	+
R. typus	XM_020524727.1	*TMEM106C*	scf_rhity00004791	100.000	53	0	367	419	98612	98560	2.30e-18	99.0	+
R. typus	XM_020524727.1	*TMEM106C*	scf_rhity00004791	100.000	55	0	420	474	93438	93384	1.78e-19	102	+
R. typus	XM_020524727.1	*TMEM106C*	scf_rhity00004791	97.368	114	3	464	576	86269	86158	3.69e-46	191	+
C. milii	XM_007910507.1	*VDR*	scf_rhity00008426	79.739	153	0	1	153	19353	19505	6.88e-22	111	-
C. milii	XM_007910507.1	*VDR*	scf_rhity00008426	87.879	132	0	150	281	31282	31413	3.13e-35	156	-
C. milii	XM_007910507.1	*VDR*	scf_rhity00011943	82.927	123	0	286	408	1588	1710	6.88e-22	111	-
C. milii	XM_007910507.1	*VDR*	scf_rhity00011943	84.091	88	0	684	771	18784	18871	4.17e-14	86.1	-
C. milii	XM_007910507.1	*VDR*	scf_rhity00013198	83.212	137	4	765	899	4730	4864	3.18e-25	122	-
C. milii	XM_007910507.1	*VDR*	scf_rhity00010247	82.075	106	0	921	1026	6305	6410	8.96e-16	91.6	-
H. sapiens	NM_001098416.3	*HDAC7*	Not Found	-	-	-	-	-	-	-	-	-	-

Table S4.3. Blast-n output of human, elephant shark and whale shark (first version) genome sequences against the brownbanded bamboo shark genome. The blast-n valeus were used to reconstruct the synteny of the *locus* in three target *PXR*, *CAR* and *VDR*.

Bait Species	Accession Number	Gene Symbol	Scaffolds of *C. punctatums* Genome	ID%	Length	Gaps	qstart	qend	sstart	send	E-valeu	Bitscore	Strand
R. typus	XM_020529932.1	*POPDC2*	scf_chipu00000056	95.951	494	0	1	494	4992671	4993164	0.0	802	-
R. typus	XM_020529932.1	*POPDC2*	scf_chipu00000056	95.614	114	0	492	605	5000060	5000173	1.41e-43	183	-
R. typus	XM_020529932.1	*POPDC2*	scf_chipu00000056	95.942	345	1	597	940	5003176	5003520	2.01e-156	558	-
R. typus	XM_020529932.1	*POPDC2*	scf_chipu00000056	94.667	75	0	940	1014	5015810	5015884	1.45e-23	117	-
R. typus	XM_020529942.1	*COX17*	scf_chipu00000056	93.496	123	0	3	125	4956135	4956257	2.61e-44	183	-
R. typus	XM_020529942.1	*COX17*	scf_chipu00000056	97.619	84	0	125	208	4964003	4964086	1.23e-32	145	-
R. typus	XM_020521822.1	*MAATS1*	scf_chipu00000056	91.667	156	0	1	156	4775262	4775107	2.94e-53	217	+
R. typus	XM_020521822.1	*MAATS1*	scf_chipu00000056	92.771	83	0	160	242	4771175	4771093	2.37e-24	121	+
R. typus	XM_020521822.1	*MAATS1*	scf_chipu00000056	93.333	60	0	241	300	4759671	4759612	6.69e-15	89.8	+
R. typus	XM_020521822.1	*MAATS1*	scf_chipu00015270	92.222	180	0	297	476	369	548	6.23e-65	255	+
R. typus	XM_020521822.1	*MAATS1*	scf_chipu00000056	98.333	180	0	475	654	4750844	4750665	2.82e-83	316	+
R. typus	XM_020521822.1	*MAATS1*	scf_chipu00000056	95.570	158	0	654	811	4746570	4746413	2.24e-64	254	+
R. typus	XM_020521822.1	*MAATS1*	scf_chipu00000056	92.021	188	0	810	997	4731836	4731649	1.04e-67	265	+
R. typus	XM_020521822.1	*MAATS1*	scf_chipu00000056	95.775	142	0	994	1135	4726590	4726449	3.78e-57	230	+
R. typus	XM_020521822.1	*MAATS1*	scf_chipu00075376	90.840	131	3	1133	1263	642	769	6.45e-40	172	+
R. typus	XM_020521822.1	*MAATS1*	scf_chipu00000056	93.243	74	0	1250	1323	4708525	4708452	5.13e-21	110	+
R. typus	XM_020521822.1	*MAATS1*	scf_chipu00000056	96.622	148	0	1323	1470	4707512	4707365	3.75e-62	246	+
R. typus	XM_020521822.1	*MAATS1*	scf_chipu00000056	96.460	226	0	1469	1694	4706591	4706366	1.65e-100	374	+
R. typus	XM_020521822.1	*MAATS1*	scf_chipu00000056	94.737	152	0	1690	1841	4705776	4705625	2.26e-59	237	+
R. typus	XM_020521822.1	*MAATS1*	scf_chipu00000056	92.708	96	1	1842	1936	4704759	4704664	2.35e-29	137	+
R. typus	XM_020521822.1	*MAATS1*	scf_chipu00000056	86.957	184	1	1932	2112	4702337	4702154	2.29e-49	204	+
H. sapiens / C . milii	NM_003889.3 / XM_007910507.1	***PXR***	Not Found	-	-	-	-	-	-	-	-	-	-
R. typus	XM_020510094.1	*GSK3B*	scf_chipu00000056	100.000	89	0	1	89	4481147	4481235	6.41e-38	165	-
R. typus	XM_020510094.1	*GSK3B*	scf_chipu00000056	95.337	193	0	89	281	4550303	4550495	1.01e-80	307	-
R. typus	XM_020510094.1	*GSK3B*	scf_chipu00000056	100.000	84	0	283	366	4580702	4580785	3.86e-35	156	-
R. typus	XM_020510094.1	*GSK3B*	scf_chipu00000056	96.460	113	0	365	477	4607809	4607921	1.37e-44	187	-

R. typus	XM_020510094.1	*GSK3B*	scf_chipu00000056	99.237	131	0	478	608	4615195	4615325	1.34e-59	237	-
R. typus	XM_020510094.1	*GSK3B*	scf_chipu00000056	98.165	109	0	607	715	4627538	4627646	1.06e-45	191	-
R. typus	XM_020510094.1	*GSK3B*	scf_chipu00000056	100.000	103	0	713	815	4630583	4630685	1.06e-45	191	-
R. typus	XM_020510094.1	*GSK3B*	scf_chipu00000056	98.000	100	0	812	911	4631396	4631495	1.07e-40	174	-
R. typus	XM_020510094.1	*GSK3B*	scf_chipu00000056	98.942	189	0	908	1096	4644749	4644937	3.57e-90	339	-
R. typus	XM_020510094.1	*GSK3B*	scf_chipu00000056	95.050	101	0	1095	1195	4649239	4649339	2.98e-36	159	-
R. typus	XM_020510094.1	*GSK3B*	scf_chipu00000056	95.946	74	0	1193	1266	4656936	4657009	1.41e-24	121	-
R. typus	XM_020510092.1	*GPR156*	scf_chipu00000056	94.872	78	0	1	78	4385393	4385470	7.38e-25	122	-
R. typus	XM_020510092.1	*GPR156*	scf_chipu00000056	92.414	145	0	72	216	4387707	4387851	1.98e-50	207	-
R. typus	XM_020510092.1	*GPR156*	scf_chipu00000056	95.726	117	0	211	327	4397052	4397168	7.18e-45	189	-
R. typus	XM_020510092.1	*GPR156*	scf_chipu00000056	91.406	128	0	327	454	4409602	4409729	5.59e-41	176	-
R. typus	XM_020510092.1	*GPR156*	scf_chipu00000056	95.098	102	0	453	554	4409864	4409965	1.56e-36	161	-
R. typus	XM_020510092.1	*GPR156*	scf_chipu00000056	94.606	241	0	548	788	4412776	4413016	1.85e-100	374	-
R. typus	XM_020510092.1	*GPR156*	scf_chipu00000056	96.364	165	0	784	948	4417282	4417446	6.93e-70	272	-
R. typus	XM_020510092.1	*GPR156*	scf_chipu00000056	91.081	1424	4	940	2361	4418947	4420368	0.0	1923	-
R. typus	XM_020510093.1	*LRRC58*	scf_chipu00000056	84.615	65	0	73	137	4283016	4283080	3.83e-08	65.8	-
R. typus	XM_020510093.1	*LRRC58*	scf_chipu00000056	96.947	131	0	137	267	4286041	4286171	7.73e-55	220	-
R. typus	XM_020510093.1	*LRRC58*	scf_chipu00000056	94.585	277	0	269	545	4287447	4287723	1.18e-117	429	-
R. typus	XM_020510093.1	*LRRC58*	scf_chipu00000056	96.429	196	0	543	738	4289865	4290060	5.73e-86	324	-
C. milii	XM_007895843.1	*FST1L*	scf_chipu00000056	80.556	108	0	460	567	4020679	4020786	3.40e-13	84.2	-
C. milii	XM_007895843.1	*FST1L*	scf_chipu00000056	81.731	104	0	563	666	4022328	4022431	2.63e-14	87.9	-
C. milii	XM_007895843.1	*FST1L*	scf_chipu00000056	88.571	70	0	669	738	4023195	4023264	9.46e-14	86.1	-
C. milii	XM_007895843.1	*FST1L*	scf_chipu00000056	79.070	86	0	781	866	4026553	4026638	5.73e-06	60.2	-
C. milii	XM_007895843.1	*FST1L*	scf_chipu00000056	78.947	76	0	928	1003	4029827	4029902	0.001	52.8	-
C. milii	XM_007895843.1	*FST1L*	scf_chipu00000056	87.129	101	2	1411	1510	4153418	4153517	4.34e-22	113	-
C. milii	XM_007895843.1	*FST1L*	scf_chipu00000056	86.916	107	0	1525	1631	4161165	4161271	2.59e-24	121	-
C. milii	XM_007895843.1	*FST1L*	scf_chipu00000056	83.206	131	0	1677	1807	4165635	4165765	2.59e-24	121	-
C. milii	XM_007895843.1	*FST1L*	scf_chipu00000056	80.952	105	3	1808	1910	4172054	4172156	4.40e-12	80.5	-
C. milii	XM_007895843.1	*FST1L*	scf_chipu00000056	92.373	118	1	1927	2044	4173639	4173755	3.28e-38	167	-
C. milii	XM_007895843.1	*FST1L*	scf_chipu00000056	88.793	116	1	2039	2154	4179859	4179973	1.99e-30	141	-
C. milii	XM_007895843.1	*FST1L*	scf_chipu00000056	85.246	61	0	2203	2263	4181469	4181529	4.43e-07	63.9	-
C. milii	XM_007895843.1	*FST1L*	scf_chipu00000056	89.362	47	0	2261	2307	4194117	4194163	5.73e-06	60.2	-
H. sapiens	NM_001102566.1	*PCP4L1*	scf_chipu00019604	79.268	82	0	100	181	1318	1399	1.62e-06	58.4	-
H. sapiens	NM_001077482.2	***CAR***	Not Found	-	-	-	-	-	-	-	-	-	
R. typus	XM_020523825.1	*TOMM40L*	scf_chipu00000039	91.304	46	1	96	140	1210516	1210471	7.06e-07	62.1	-
R. typus	XM_020523825.1	*TOMM40L*	scf_chipu00055375	95.876	97	0	292	388	900	996	8.76e-36	158	-
R. typus	XM_020523825.1	*TOMM40L*	scf_chipu00055375	91.346	104	0	386	489	1658	1761	2.45e-31	143	-
R. typus	XM_020523825.1	*TOMM40L*	scf_chipu00169312	89.189	111	0	486	596	405	515	3.17e-30	139	-
R. typus	XM_020523825.1	*TOMM40L*	scf_chipu00021705	93.966	116	0	604	719	2789	2674	2.42e-41	176	-
R. typus	XM_020523825.1	*TOMM40L*	scf_chipu00122405	85.000	80	0	716	795	732	811	5.42e-13	82.4	-
R. typus	XM_020523825.1	*TOMM40L*	scf_chipu00006027	90.351	114	1	793	906	10566	10454	5.27e-33	148	-
R. typus	XM_020523825.1	*TOMM40L*	scf_chipu00006027	89.437	142	0	897	1038	7081	6940	1.87e-42	180	-
R. typus	XM_020523967.1	*NDUFS2*	scf_chipu00004839	83.550	231	6	2	223	18979	18751	3.28e-50	206	-
R. typus	XM_020523967.1	*NDUFS2*	scf_chipu00050391	94.118	119	1	218	336	735	852	1.99e-42	180	-
R. typus	XM_020523967.1	*NDUFS2*	scf_chipu00062202	97.436	78	0	335	412	640	717	1.57e-28	134	-
R. typus	XM_020523967.1	*NDUFS2*	scf_chipu00062202	96.552	87	1	403	489	1478	1563	2.60e-31	143	-
R. typus	XM_020523967.1	*NDUFS2*	scf_chipu00004839	90.909	88	0	488	575	8527	8440	4.39e-24	119	-
R. typus	XM_020523967.1	*NDUFS2*	scf_chipu00004839	91.200	125	0	573	697	8336	8212	1.19e-39	171	-
R. typus	XM_020523967.1	*NDUFS2*	scf_chipu00004839	90.977	133	0	693	825	7794	7662	1.99e-42	180	-
R. typus	XM_020523967.1	*NDUFS2*	scf_chipu00004839	92.929	99	0	824	922	4501	4403	7.24e-32	145	-
R. typus	XM_020523967.1	*NDUFS2*	scf_chipu00004839	90.526	95	2	914	1006	4314	4220	9.43e-26	124	-
R. typus	XM_020523967.1	*NDUFS2*	scf_chipu00037270	88.525	61	0	1005	1065	2241	2181	9.64e-11	75.0	-
C. milii	XM_007910528.1	*COL2A1*	scf_chipu00001599	82.486	177	2	99	274	154649	154824	4.75e-34	154	-
C. milii	XM_007910528.1	*COL2A1*	scf_chipu00001599	86.000	50	0	440	489	164583	164632	4.96e-04	54.7	-
C. milii	XM_007910528.1	*COL2A1*	scf_chipu00001599	89.873	79	0	535	613	167343	167421	1.75e-18	102	-
C. milii	XM_007910528.1	*COL2A1*	scf_chipu00001599	91.489	47	0	612	658	171023	171069	2.29e-07	65.8	-
C. milii	XM_007910528.1	*COL2A1*	scf_chipu00001599	89.091	55	0	658	712	171240	171294	1.77e-08	69.4	-
C. milii	XM_007910528.1	*COL2A1*	scf_chipu00001599	91.071	56	1	766	820	172856	172911	3.81e-10	75.0	-
C. milii	XM_007910528.1	*COL2A1*	scf_chipu00001599	90.909	44	0	973	1016	176005	176048	1.07e-05	60.2	-
C. milii	XM_007910528.1	*COL2A1*	scf_chipu00001599	82.075	106	1	1070	1175	183936	184040	1.36e-14	89.8	-
C. milii	XM_007910528.1	*COL2A1*	scf_chipu00001599	89.130	46	0	1171	1216	184516	184561	3.83e-05	58.4	-
C. milii	XM_007910528.1	*COL2A1*	scf_chipu00001599	79.787	94	0	1216	1309	185910	186003	1.77e-08	69.4	-
C. milii	XM_007910528.1	*COL2A1*	scf_chipu00001599	85.938	64	4	1307	1368	187431	187492	2.29e-07	65.8	-
C. milii	XM_007910528.1	*COL2A1*	scf_chipu00001599	83.607	61	0	1786	1846	198458	198518	3.83e-05	58.4	-
C. milii	XM_007910528.1	*COL2A1*	scf_chipu00001599	82.243	107	0	1884	1990	202542	202648	1.05e-15	93.5	-
C. milii	XM_007910528.1	*COL2A1*	scf_chipu00001599	90.244	41	0	2106	2146	203736	203776	4.96e-04	54.7	-
C. milii	XM_007910528.1	*COL2A1*	scf_chipu00001599	77.914	163	0	2533	2695	212510	212672	1.75e-18	102	-
C. milii	XM_007910528.1	*COL2A1*	scf_chipu00001599	82.653	98	0	2695	2792	212983	213080	4.89e-14	87.9	-
C. milii	XM_007910528.1	*COL2A1*	scf_chipu00001599	78.632	117	1	2797	2911	220830	220946	1.06e-10	76.8	-
C. milii	XM_007910528.1	*COL2A1*	scf_chipu00001599	86.207	58	2	2909	2965	221036	221092	2.96e-06	62.1	-
C. milii	XM_007910528.1	*COL2A1*	scf_chipu00001599	82.524	103	0	2963	3065	223076	223178	3.78e-15	91.6	-
C. milii	XM_007910528.1	*COL2A1*	scf_chipu00001599	85.455	55	0	3073	3127	223418	223472	3.83e-05	58.4	-
C. milii	XM_007910528.1	*COL2A1*	scf_chipu00001599	88.991	109	0	3127	3235	224091	224199	1.72e-28	135	-
C. milii	XM_007910528.1	*COL2A1*	scf_chipu00001599	80.952	105	1	3285	3389	227792	227895	2.27e-12	82.4	-
C. milii	XM_007910528.1	*COL2A1*	scf_chipu00001599	81.544	298	1	3391	3687	228051	228348	2.74e-61	244	-
C. milii	XM_007910528.1	*COL2A1*	scf_chipu00001599	87.831	189	0	3685	3873	228756	228944	1.28e-54	222	-
C. milii	XM_007910528.1	*COL2A1*	scf_chipu00001599	91.628	215	0	3902	4116	229975	230189	2.07e-77	298	-
C. milii	XM_007910528.1	*COL2A1*	scf_chipu00001599	93.243	148	0	4116	4263	230763	230910	1.66e-53	219	-
R. typus	XM_020524727.1	*TMEM106C*	scf_chipu00001599	94.340	106	1	472	576	241387	241282	3.67e-37	161	+

R. typus	XM_020524727.1	*TMEM106C*	scf_chipu00001599	96.296	54	0	421	474	252505	252452	1.75e-15	89.8	+
R. typus	XM_020524727.1	*TMEM106C*	scf_chipu00001599	92.453	53	0	367	419	257715	257663	1.37e-11	76.8	+
R. typus	XM_020524727.1	*TMEM106C*	scf_chipu00001599	96.000	150	1	223	371	258189	258040	1.27e-61	243	+
R. typus	XM_020524727.1	*TMEM106C*	scf_chipu00000025	74.843	159	7	63	220	4017375	4017223	2.96e-08	65.8	+
C. milii	XM_007910507.1	*VDR*	scf_chipu00001415	89.216	102	0	52	153	196819	196718	8.65e-27	128	-
C. milii	XM_007910507.1	*VDR*	scf_chipu00001415	87.023	131	0	151	281	177111	176981	6.64e-33	148	-
C. milii	XM_007910507.1	*VDR*	scf_chipu00001415	78.431	153	0	286	438	143743	143591	1.89e-18	100	-
C. milii	XM_007910507.1	*VDR*	scf_chipu00001415	82.796	93	1	687	778	123346	123254	6.83e-13	82.4	-
C. milii	XM_007910507.1	*VDR*	scf_chipu00001415	83.582	134	4	768	899	97142	97011	4.03e-25	122	-
C. milii	XM_007910507.1	*VDR*	scf_chipu00001415	82.353	119	0	921	1039	85238	85120	1.46e-19	104	-
C. milii	XM_007910507.1	*VDR*	scf_chipu00001415	81.250	64	0	1159	1222	72435	72372	5.36e-04	52.8	-
H. sapiens	NM_001098416.3	*HDAC7*	Not Found	-	-	-	-	-	-	-	-	-	-

Table S4.4. Blast-n output of human, elephant shark and whale shark (first version) genome sequences against the cloudy catshark genome. The blast-n valeus were used to reconstruct the synteny of the *locus* in three target *PXR*, *CAR* and *VDR*.

Bait Species	Accession Number	Gene Symbol	Scaffolds of S. torazame Genome	ID%	Length	Gaps	qstart	qend	sstart	send	E-valeu	Bitscore	Strand
R. typus	XM_020529932.1	*POPDC2*	scf_scyto00004451	86.842	494	0	1	494	110476	110969	1.23E-154	553	-
R. typus	XM_020529932.1	*POPDC2*	scf_scyto00004451	93.023	344	0	597	940	121940	122283	1.26E-139	503	-
R. typus	XM_020529932.1	*POPDC2*	scf_scyto00004451	92.920	113	0	493	605	114005	114117	6.75E-38	165	-
R. typus	XM_020529932.1	*POPDC2*	scf_scyto00004451	85.333	75	0	940	1014	128612	128686	9.05E-12	78.7	-
R. typus	XM_020529942.1	*COX17*	scf_scyto00004451	88.618	123	0	3	125	25599	25721	3.48E-34	150	-
R. typus	XM_020529942.1	*COX17*	scf_scyto00004451	90.698	86	0	125	210	36134	36219	1.27E-23	115	-
R. typus	XM_020521822.1	*MAATS1*	scf_scyto00010339	87.898	157	0	1	157	102924	103080	1.10E-43	185	+
R. typus	XM_020521822.1	*MAATS1*	scf_scyto00010954	85.882	85	1	155	239	5469	5552	8.85E-15	89.8	+
R. typus	XM_020521822.1	*MAATS1*	scf_scyto00010954	85.075	67	2	234	300	11980	12044	4.14E-08	67.6	+
R. typus	XM_020521822.1	*MAATS1*	scf_scyto00010954	90.714	140	0	337	476	17821	17960	3.05E-44	187	+
R. typus	XM_020521822.1	*MAATS1*	scf_scyto00010954	93.333	180	0	475	654	20228	20407	3.81E-68	267	+
R. typus	XM_020521822.1	*MAATS1*	scf_scyto00010954	85.185	162	2	651	811	42066	42226	1.43E-37	165	+
R. typus	XM_020521822.1	*MAATS1*	scf_scyto00010954	85.638	188	0	810	997	83401	83588	1.41E-47	198	+
R. typus	XM_020521822.1	*MAATS1*	scf_scyto00010954	87.413	143	2	994	1135	84704	84845	5.14E-37	163	+
R. typus	XM_020521822.1	*MAATS1*	scf_scyto00010954	83.916	143	5	1119	1260	92673	92810	1.45E-27	132	+
R. typus	XM_020521822.1	*MAATS1*	scf_scyto00019089	85.366	82	1	1250	1329	13565	13646	4.12E-13	84.2	+
R. typus	XM_020521822.1	*MAATS1*	scf_scyto00019089	92.763	152	1	1320	1470	14575	14726	1.08E-53	219	+
R. typus	XM_020521822.1	*MAATS1*	scf_scyto00019089	90.708	226	0	1469	1694	16466	16691	1.04E-78	302	+
R. typus	XM_020521822.1	*MAATS1*	scf_scyto00019089	90.132	152	0	1690	1841	21232	21383	1.41E-47	198	+
R. typus	XM_020521822.1	*MAATS1*	scf_scyto00012507	90.698	86	1	1852	1936	8615	8700	5.25E-22	113	+
R. typus	XM_020521822.1	*MAATS1*	scf_scyto00012507	82.209	163	0	1932	2094	22807	22969	2.41E-30	141	+
H. sapiens / C . milii	NM_003889.3 / XM_007910507.1	***PXR***	Not Found	-	-	-	-	-	-	-	-	-	
R. typus	XM_020510094.1	*GSK3B*	scf_scyto00012507	93.151	73	0	1194	1266	70445	70373	1.45E-20	108	-
R. typus	XM_020510094.1	*GSK3B*	scf_scyto00087619	90.385	104	0	1095	1198	696	593	1.85E-29	137	-
R. typus	XM_020510094.1	*GSK3B*	scf_scyto00001496	95.238	189	0	908	1096	523941	524129	2.23E-78	300	-
R. typus	XM_020510094.1	*GSK3B*	scf_scyto00001496	96.000	100	0	812	911	507821	507920	3.05E-37	163	-
R. typus	XM_020510094.1	*GSK3B*	scf_scyto00001496	97.087	103	0	713	815	506971	507073	1.41E-40	174	-
R. typus	XM_020510094.1	*GSK3B*	scf_scyto00001496	96.330	109	0	607	715	504105	504213	3.03E-42	180	-
R. typus	XM_020510094.1	*GSK3B*	scf_scyto00001496	96.947	131	0	478	608	488029	488159	1.79E-54	220	-
R. typus	XM_020510094.1	*GSK3B*	scf_scyto00001496	95.575	113	0	365	477	476725	476837	8.42E-43	182	-
R. typus	XM_020510094.1	*GSK3B*	scf_scyto00001496	98.810	84	0	283	366	438710	438793	2.38E-33	150	-
R. typus	XM_020510094.1	*GSK3B*	scf_scyto00001496	92.228	193	0	89	281	413329	413521	1.35E-70	274	-
R. typus	XM_020510094.1	*GSK3B*	scf_scyto00001496	100.000	89	0	1	89	334378	334466	8.48E-38	165	-
R. typus	XM_020510092.1	*GPR156*	scf_scyto00001496	80.336	1429	18	947	2361	238954	240374	0	1062	-
R. typus	XM_020510092.1	*GPR156*	scf_scyto00001496	94.545	165	0	784	948	237552	237716	9.23E-65	255	-
R. typus	XM_020510092.1	*GPR156*	scf_scyto00001496	88.163	245	0	548	792	213963	214207	7.03E-76	292	-
R. typus	XM_020510092.1	*GPR156*	scf_scyto00001496	90.196	102	0	453	554	206120	206221	4.51E-28	134	-
R. typus	XM_020510092.1	*GPR156*	scf_scyto00001496	86.719	128	0	327	454	205611	205738	7.50E-31	143	-
R. typus	XM_020510092.1	*GPR156*	scf_scyto00001496	88.889	117	0	211	327	186015	186131	2.08E-31	145	-
R. typus	XM_020510092.1	*GPR156*	scf_scyto00001496	86.525	141	0	76	216	167043	167183	9.63E-35	156	-
R. typus	XM_020510092.1	*GPR156*	scf_scyto00001496	92.683	82	1	1	81	157588	157669	4.54E-23	117	-
R. typus	XM_020510093.1	*LRRC58*	scf_scyto00000214	91.603	131	0	137	267	1067413	1067543	4.83E-43	182	-
R. typus	XM_020510093.1	*LRRC58*	scf_scyto00000214	90.614	277	0	269	545	1067963	1068239	3.45E-99	368	-
R. typus	XM_020510093.1	*LRRC58*	scf_scyto00000214	92.857	196	0	543	738	1074112	1074307	3.58E-74	285	-
C. milii	XM_007895843.1	*FST1L*	scf_scyto00000214	74.419	172	4	87	257	715659	715826	1.26E-08	69.4	-
C. milii	XM_007895843.1	*FST1L*	scf_scyto00000214	81.481	108	0	460	567	728379	728486	9.68E-15	89.8	-
C. milii	XM_007895843.1	*FST1L*	scf_scyto00000214	84.466	103	0	564	666	744011	744113	1.24E-18	102	-
C. milii	XM_007895843.1	*FST1L*	scf_scyto00000214	90.141	71	0	668	738	744837	744907	7.48E-16	93.5	-
C. milii	XM_007895843.1	*FST1L*	scf_scyto00000214	80.198	101	0	778	878	748572	748672	7.53E-11	76.8	-
C. milii	XM_007895843.1	*FST1L*	scf_scyto00000214	91.209	91	0	1420	1510	892369	892459	2.65E-25	124	-
C. milii	XM_007895843.1	*FST1L*	scf_scyto00000214	85.185	135	0	1508	1642	903903	904037	9.47E-30	139	-
C. milii	XM_007895843.1	*FST1L*	scf_scyto00000214	85.714	133	0	1677	1809	912086	912218	2.63E-30	141	-
C. milii	XM_007895843.1	*FST1L*	scf_scyto00000214	80.342	117	6	1813	1925	925742	925854	1.62E-12	82.4	-
C. milii	XM_007895843.1	*FST1L*	scf_scyto00000214	89.916	119	1	1927	2045	927808	927925	1.22E-33	152	-
C. milii	XM_007895843.1	*FST1L*	scf_scyto00000214	87.200	125	1	2039	2163	939494	939617	2.63E-30	141	-
C. milii	XM_007895843.1	*FST1L*	scf_scyto00000214	83.607	61	0	2203	2263	941514	941574	2.73E-05	58.4	-
C. milii	XM_007895843.1	*FST1L*	scf_scyto00000214	95.745	47	0	2261	2307	944073	944119	7.53E-11	76.8	-
H. sapiens	NM_001102566.1	*PCP4L1*	scf_scyto00161034	83.529	85	0	97	181	1030	946	4.56E-13	80.5	-
H. sapiens	NM_001077482.2	***CAR***	Not Found	-	-	-	-	-	-	-	-	-	
R. typus	XM_020523825.1	*TOMM40L*	scf_scyto00132904	91.525	59	0	168	226	98	40	7.17E-13	82.4	-

R. typus	XM_020523825.1	*TOMM40L*	scf_scyto00013824	91.892	74	0	225	298	52160	52233	1.53E-19	104	-
R. typus	XM_020523825.1	*TOMM40L*	scf_scyto00013824	86.458	96	0	293	388	54607	54702	4.25E-20	106	-
R. typus	XM_020523825.1	*TOMM40L*	scf_scyto00013824	88.462	104	0	386	489	60099	60202	3.26E-26	126	-
R. typus	XM_020523825.1	*TOMM40L*	scf_scyto00013824	88.785	107	0	487	593	60788	60894	7.02E-28	132	-
R. typus	XM_020523825.1	*TOMM40L*	scf_scyto00010387	88.235	119	2	602	719	103337	103220	1.17E-30	141	-
R. typus	XM_020523825.1	*TOMM40L*	scf_scyto00010387	85.333	75	0	724	798	101406	101332	9.27E-12	78.7	-
R. typus	XM_020523825.1	*TOMM40L*	scf_scyto00010387	88.235	119	2	602	719	103337	103220	1.17E-30	141	-
R. typus	XM_020523825.1	*TOMM40L*	scf_scyto00010387	85.333	75	0	724	798	101406	101332	9.27E-12	78.7	-
R. typus	XM_020523825.1	*TOMM40L*	scf_scyto00010387	90.845	142	0	897	1038	95986	95845	1.14E-45	191	-
R. typus	XM_020523967.1	*NDUFS2*	scf_scyto00049214	86.179	123	0	101	223	153	31	2.07E-28	134	-
R. typus	XM_020523967.1	*NDUFS2*	scf_scyto00391764	91.026	78	0	335	412	584	507	4.52E-20	106	-
R. typus	XM_020523967.1	*NDUFS2*	scf_scyto00391764	86.364	66	0	424	489	322	257	4.58E-10	73.1	-
R. typus	XM_020523967.1	*NDUFS2*	scf_scyto00432542	90.909	88	0	488	575	118	205	5.80E-24	119	-
R. typus	XM_020523967.1	*NDUFS2*	scf_scyto00432542	87.200	125	0	573	697	303	427	3.44E-31	143	-
R. typus	XM_020523967.1	*NDUFS2*	scf_scyto00254062	82.090	134	0	692	825	950	817	7.51E-23	115	-
R. typus	XM_020523967.1	*NDUFS2*	scf_scyto00034789	87.879	99	0	824	922	3516	3418	2.09E-23	117	-
C. milii	XM_007910528.1	*COL2A1*	scf_scyto00020566	82.738	168	2	132	298	2862	2696	2.93E-32	148	-
C. milii	XM_007910528.1	*COL2A1*	scf_scyto00081496	94.595	37	0	311	347	1412	1376	5.07E-05	58.4	-
C. milii	XM_007910528.1	*COL2A1*	scf_scyto00007676	89.873	79	0	535	613	149294	149216	2.31E-18	102	-
C. milii	XM_007910528.1	*COL2A1*	scf_scyto00007676	91.489	47	0	612	658	144968	144922	3.03E-07	65.8	-
C. milii	XM_007910528.1	*COL2A1*	scf_scyto00007676	87.755	49	0	664	712	144750	144702	5.07E-05	58.4	-
C. milii	XM_007910528.1	*COL2A1*	scf_scyto00007676	91.071	56	1	766	820	143056	143001	5.04E-10	75.0	-
C. milii	XM_007910528.1	*COL2A1*	scf_scyto00007676	95.000	40	0	979	1018	140679	140640	1.09E-06	63.9	-
C. milii	XM_007910528.1	*COL2A1*	scf_scyto00007676	81.188	101	1	1070	1169	135788	135688	1.08E-11	80.5	-
C. milii	XM_007910528.1	*COL2A1*	scf_scyto00007676	80.682	88	0	1222	1309	134426	134339	2.34E-08	69.4	-
C. milii	XM_007910528.1	*COL2A1*	scf_scyto00007676	78.218	101	0	1377	1477	131776	131676	3.03E-07	65.8	-
C. milii	XM_007910528.1	*COL2A1*	scf_scyto00007676	84.483	58	0	1789	1846	124466	124409	5.07E-05	58.4	-
C. milii	XM_007910528.1	*COL2A1*	scf_scyto00007676	81.982	111	0	1883	1993	115073	114963	3.87E-16	95.3	-
C. milii	XM_007910528.1	*COL2A1*	scf_scyto00007676	77.570	107	0	1988	2094	114874	114768	3.03E-07	65.8	-
C. milii	XM_007910528.1	*COL2A1*	scf_scyto00007676	78.981	157	0	2539	2695	102368	102212	4.97E-20	108	-
C. milii	XM_007910528.1	*COL2A1*	scf_scyto00007676	85.047	107	0	2695	2801	102062	101956	1.38E-20	110	-
C. milii	XM_007910528.1	*COL2A1*	scf_scyto00007676	79.091	110	0	2802	2911	93356	93247	1.40E-10	76.8	-
C. milii	XM_007910528.1	*COL2A1*	scf_scyto00007676	76.577	111	0	2963	3073	89829	89719	3.92E-06	62.1	-
C. milii	XM_007910528.1	*COL2A1*	scf_scyto00007676	86.538	52	0	3076	3127	89465	89414	5.07E-05	58.4	-
C. milii	XM_007910528.1	*COL2A1*	scf_scyto00007676	88.073	109	0	3127	3235	88127	88019	1.06E-26	130	-
C. milii	XM_007910528.1	*COL2A1*	scf_scyto00007676	81.416	113	1	3285	3397	83843	83732	5.00E-15	91.6	-
C. milii	XM_007910528.1	*COL2A1*	scf_scyto00007676	84.247	292	1	3397	3687	83428	83137	7.69E-73	283	-
C. milii	XM_007910528.1	*COL2A1*	scf_scyto00007676	89.947	189	0	3685	3873	81999	81811	3.63E-61	244	-
C. milii	XM_007910528.1	*COL2A1*	scf_scyto00007676	87.251	251	2	3867	4116	80534	80285	2.14E-73	285	-
C. milii	XM_007910528.1	*COL2A1*	scf_scyto00007676	93.919	148	0	4116	4263	79787	79640	4.73E-55	224	-
R. typus	XM_020524727.1	*TMEM106C*	scf_scyto00007676	87.333	150	1	223	371	55335	55484	8.06E-40	171	+
R. typus	XM_020524727.1	*TMEM106C*	scf_scyto00007676	86.792	53	0	367	419	55787	55839	1.82E-06	60.2	+
R. typus	XM_020524727.1	*TMEM106C*	scf_scyto00007676	94.545	55	0	420	474	60578	60632	3.00E-14	86.1	+
R. typus	XM_020524727.1	*TMEM106C*	scf_scyto00007676	93.204	103	0	474	576	66169	66271	2.92E-34	152	+
C. milii	XM_007910507.1	*VDR*	scf_scyto00007144	87.500	112	0	42	153	53238	53127	3.19E-27	130	-
C. milii	XM_007910507.1	*VDR*	scf_scyto00007144	88.550	131	0	151	281	27042	26912	4.06E-36	159	-
C. milii	XM_007910507.1	*VDR*	scf_scyto00007144	81.098	164	0	284	447	645	482	8.86E-28	132	-
C. milii	XM_007910507.1	*VDR*	scf_scyto00012969	87.952	83	0	689	771	7071	7153	8.98E-18	99.0	-
C. milii	XM_007910507.1	*VDR*	scf_scyto00012969	85.926	135	0	765	899	20563	20697	1.14E-31	145	-
C. milii	XM_007910507.1	*VDR*	scf_scyto00012969	85.088	114	0	921	1034	29548	29661	2.48E-23	117	-
C. milii	XM_007910507.1	*VDR*	scf_scyto00012969	72.556	266	4	1039	1302	38914	39177	2.52E-13	84.2	-
H. sapiens	NM_001098416.3	*HDAC7*	Not Found	-	-	-	-	-	-	-	-	-	-

CHAPTER 5 - The Evolution of the Retinoid X Receptor in Metazoa: Insights into Lipid Metabolism Disruption in a Marine Rotifer

Adapted from:

Environmental Science & Technology

Article

Cite This: *Environ. Sci. Technol.* 2019, 53, 7830−7839

pubs.acs.org/est

Tributyltin Affects Retinoid X Receptor-Mediated Lipid Metabolism in the Marine Rotifer *Brachionus koreanus*

Min-Chul Lee,[†,#] Elza Fonseca,[‡,§,#] Jun Chul Park,[†] Deok-Seo Yoon,[†] Hyuntae Choi,[∥] Moonkoo Kim,[⊥,#] Jeonghoon Han,[†] Hyeon-Seo Cho,[¶] Kyung-Hoon Shin,[∥] Miguel L. Santos,[‡,§] Jee-Hyun Jung,[⊥,#] L. Filipe C. Castro,[‡,§] and Jae-Seong Lee[*,†]

[†]Department of Biological Science, College of Science, Sungkyunkwan University, Suwon 16419, South Korea

[‡]Department of Biology, Faculty of Sciences, University of Porto, 4169-007 Porto, Portugal

[§]Interdisciplinary Centre of Marine and Environmental Research, University of Porto, 4450-208 Matosinhos, Portugal

[∥]Department of Marine and Convergence Sciences, College of Science and Technology, Hanyang University, Ansan 15588, South Korea

[⊥]Risk Assessment Research Center, Korea Institute of Ocean Science & Technology, Geoje 53201, South Korea

[#]Department of Marine Environmental Science, Korea University of Science and Technology, Daejeon 34113, South Korea

[¶]College of Fisheries and Ocean Sciences, Chonnam National University, Yeosu 550-749, South Korea

Supporting Information

5. The Evolution of the Retinoid X Receptor in Metazoa: Insights into Lipid Metabolism Disruption in a Marine Rotifer

5. Abstract

Nuclear receptors (NRs) are transcription factors executing an essential function in cellular metabolism. Among its numerous members, the retinoid X receptor (RXR) is a central player in the action of the endocrine system. Importantly, RXR is able to operate as a homodimer and as a heterodimer with other NRs. Additionally, RXR has been found to be a critical actor in various processes of endocrine disruption resulting from the exposure to organotins. Negative physiological effects such imposex in gastropod molluscs and lipid perturbation across different lineages have been clearly linked to the abnormal molecular exploitation of NRs, including RXR. Thus, given its prominent role in the overall organism homeostasis, RXR is present in the genomes of most extant metazoan species examined to date. Here, we expand on the phylogenetic distribution of RXR across the metazoan tree of life by exploring multiple next generation sequencing projects. Furthermore, we employed sequence comparison and *in silico* modelling to address amino acid residue conservation and the mode of action of a known class of xenobiotics, organotins. As a proof of concept, we show that the *RXR* ortholog from a rotifer, *Brachionus koreanus*, is activated (transactivation assay) by tributyltin (TBT), a model obesogen, despite the absence of a critical residue normally required for activation. These results demonstrate the critical importance of considering comparative pipelines and functional tests to decipher the effect of xenobiotics across evolutionary scales. More globally, our work supports a wider taxonomic scope of lipid perturbation due to xenobiotic exposure that occurs *via* RXR in aquatic animals.

5.1. Introduction

The release of man-made contaminants into aquatic environments and the interplay with global changes is a hall-mark of the Anthropocene, posing a serious threat to the long-term preservation of ecosystems. Examples include plasticizers, pesticides, detergents, pharmaceuticals and other emergent compounds, which are central to human industrialized societies. Endocrine disrupting chemicals (EDCs) are substances, with the potential to alter endocrine function causing physiological imbalance, reproductive impairment and metabolic defects (Sumpter & Johnson 2005; Tyler et al. 1998; Grün & Blumberg 2009; Baker et al. 2012; Söffker & Tyler 2012). Understanding the mechanistic features of EDCs action represents a scientific challenge, but it is also important to anticipate deleterious effects on the ecosystem.

The ability of numerous EDCs to mimic or block the function of signalling molecules represents a major threat to physiological homeostasis. Fundamental players in this context are the nuclear receptors (NRs). These transcription factors are abundant in metazoan genomes, being mostly triggered by ligand binding, thus regulating the expression of downstream genes. NRs and their evolution are key to understand the evolution of endocrine systems and contaminant exploitation (Bertrand et al. 2004; Holzer et al. 2017; Castro & Santos 2014; Tohyama et al. 2016). The emergence of full genome sequences brought, in recent years, a radical change to our understanding of NR genes diversification. One outstanding question is whether we have currently the full scenario of NR evolution, as many other metazoan lineages remain poorly investigated. Despite the significant wealth of NR evolutionary research, key phyla such as Brachiopoda, Rotifera, Bryozoa, Phoronida, Priapulida, Kinorhyncha and many others remain uncharacterized. The presence of a receptor *per se* does not warrant a similar molecular/physiological role nor truly comparable structural ligand. In agreement, various observations have hinted that ligand-receptor couples may not be stable in an evolutionary time scales and consequently, exhibiting dramatic changes in binding specificities. Estrogen receptors (ERs) from protostome species have been shown to be unresponsive (molluscs) or responsive (annelids and rotifers) to estradiol (Thornton & Need, E, Crews 2003; Keay & Thornton 2009; Jones et al. 2017).

Given that the NR mode of action implies tight physical binding to specific ligands, it is not surprising that they are prime targets of EDCs. Thus, understanding the mechanisms of action of EDCs provides clues to anticipate and comprehend deleterious effects in humans and wildlife. In fact, several studies have linked NRs to endocrine disruption. In the later, a significant example comes from organotins, tributyltin (TBT) and triphenyltin (TPT), the prime cause of imposex development in gastropod molluscs. We and others showed that the TBT-dependent activation of the retinoid X receptor (RXR) is key to imposex development (Nishikawa et al. 2004; Castro et al. 2007). However, in vertebrates TBT also binds the peroxisome proliferator-activated receptor gamma (PPARγ) eliciting adipogenesis (Capitão et al. 2018). Interestingly, lipid metabolism is also clearly affected by organotins in *Daphnia magna*, while through the activation of different NR signalling pathways (Jordão et al. 2015). Overall, the lack of systematic approaches and the poor taxonomic sampling are extremely problematic. Here, we expand this comparative approach by examining RXR. This NR is fundamental since due the exclusive capacity to operate both as homo and heterodimer with other NRs including PPAR. Importantly, organotins, such as TBT, activate the module RXR-PPAR heterodimer and induce obese phenotypes (Kanayama et al. 2005; Grün & Blumberg 2006; Grün et al. 2006). Therefore, it is of critical importance to investigate also RXR,

particularly in PPAR-absent lineages and consider the capacity of this receptor to mediate disturbances in lipid metabolism upon TBT exposure.

5.2. Material and Methods

5.2.1. Sequence Alignment and Phylogenetics

The *Homo sapiens* RAR amino acid sequences and the RXR amino acid sequences of *H. sapiens*, *Danio rerio*, *Branchiostoma floridae*, *B. lanceolatum*, *Crassostrea gigas*, *Lottia gigantean*, *Patella vulgata*, *Nucella lapillus*, *Reishia clavigera*, *Acanthochitona crinita*, *Platynereis dumerilii*, *Capitella teleta*, *Lingula anatina*, *Brachionus koreanus*, *B. rotundiformis*, *B. plicatilis*, *B. calyciflorus*, *Schistosoma japonicum*, *S. haematobium*, S. *mansoni, Priapulus caudatus*, *Uca pugilator*, *Daphnia magna*, *Aurelia aurita* and *Trichoplax adhaerens* were retrieved from the GenBank database (accession numbers on Supplementary Material **Table S5.1**). The full length *RXR* genes of *Membranipora membranacea* and *Halicryptus spinulosus*, partial gene sequence of *Bonellia viridis RXR* and *Bugula neritina RXR1* and *RXR2*, and two partial *RXR*-like sequences from *Xenoturbella bocki* were retrieved throught a BLAST approach using Sequence Read Archive (SRA) files. The open reading frame (ORF) of *RXR* genes from *B. viridis, Phoronopsis californica*, *Megathiris detruncata*, *B. neritina* and *X. bocki* were isolated using a polymerase chain reaction (PCR) approach briefly described in the Supplementary Material. The 44 amino acid sequences were aligned with the MAFFT server (v.7) (Katoh & Toh 2010), generating an alignment with 1223 positions used in a Bayesian phylogenetic analysis with MrBayes (v.3.2.3) sited in the CIPRES Science Gateway (v.3.3) (Miller et al. 2015). The following parameters were used: generation number=10000000, rate matrix for aa=mixed (Jones), nruns=2, nchains=4, temp=0.20, sampling set to 1000 and burnin to 0.25. The statistical support for each branch is expressed as Bayesian posterior probabilities (Nascimento et al. 2017) and indicated at the nodes. FigTree (v.1.3.1) was used to visualize the tree.

5.2.2. Construction of Plasmid Vectors

The *RXR* hinge region and ligand binding domain (LBD) were isolated using PCR with specific primers (Forward primer: aaaTCTAGACAAAAAAACAAAGACAAGCCAG; Reverse primer: aaaGGTACCTTAATAAACTTCTTGTTTTTTGACAAA) and Phusion Flash High-Fidelity PCR Master Mix (Thermo Fisher Scientific Co., Waltham, MA, USA), according to the protocol from the supplier. The PCR product was then digested with *Xba*I and *Kpn*I restriction enzymes (NZYTech, Lisbon, Portugal) and ligated to the pBIND

plasmid (AF264722; Promega, Madison, WI, USA) with T4 ligase (Promega, Madison, WI, USA) to produce a GAL4- LBD "chimeric" receptor. The chimeric receptor produces a hybrid protein, containing the Gal4 DNA binding domain (DBD) that acts on an upstream activation sequence (UAS) response element. Plasmid sequences were confirmed by sequencing (Eurofins GATC, Constance, Germany). The *RXR* mutant *B. koreanus* (Cys335Ile) was produced by NZYTech (Lisbon, Portugal).

5.2.3. Transactivation Assay

Cell culture and transactivation assays were performed as previously described in (Fonseca et al. 2017). The ligands TBT and 9-*cis* retinoic acid (9cisRA) were purchased from Sigma-Aldrich (Sintra, Portugal) and were resuspended in dimethyl sulfoxide (DMSO) at a final concentration of 1, 10, and 100 nM for TBT and 1 µM for 9cisRA. Briefly, Cos-1 cells (Sigma-Aldrich) were maintained in Dulbecco's modified Eagle's medium (DMEM) (PAN-Biotech, Aidenbach, Bayern, Germany) supplemented with 10% fetal bovine serum (PAN- Biotech, Aidenbach, Bayern, Germany) and 1% penicillin/ streptomycin (PAN-Biotech, Aidenbach, Bayern, Germany) at 37 °C with 5% CO_2 (humidified atmosphere). Cells were seeded in 24-well culture plates. After 24 h, cells were transfected with 0.5 µg of pBIND-*BkoRXR*-LBD or the pBIND-*BkoRXR*-LBD mutant and 1 µg of pGL4.31 [luc2P/ GAL4UAS/Hygro] luciferase reporter vector (Promega, Madison, WI, USA), which contains five UAS elements upstream of the firefly luciferase reporter gene, using lipofectamine 2000 reagent (Invitrogen, Carlsbad, CA, USA) in Opti-MEM (Gibco, Carlsbad, CA, USA) according to manufacturer's indications. After a 5 h incubation, the transfection media was replaced with medium containing TBT (1, 10, and 100 nM) or 9cisRA (1 µM) dissolved in DMSO (0.1%). Cells were lysed 24 h after transfection. The Dual-Luciferase Reporter Assay System (Promega, Madison, WI, USA) was used to assay firefly luciferase (reporter pGL4.31) and *Renilla* luciferase (pBIND) activities following the manufacturer's instructions. All transfections were performed with two technical replicates per condition in three independent assays. The results are expressed as the fold induction calculated as the ratio between firefly luciferase (reporter pGL4.31) and *Renilla* luciferase (internal control for transfection efficiency) and normalized by dividing by the control (DMSO), and the mean of the technical replicates was used in the statistical analysis.

5.2.4. Comparative Homology Modelling

The SWISS- MODEL homology modelling workspace was used to calculate the homology models, with the alignment mode. The crystal structure of the human RXRα ligand binding domain in complex with tributyltin (3E94) was used as the template. The

RXR sequence from *B. koreanus* was aligned to that of human RXRα using MAFFT L-INS-I, and the target-template alignment was submitted to SWISS-MODEL for homology modelling predictions. The Global Model Quality Estimation (GMQE) and Qualitative Model Energy Analysis (QMEAN) parameters (Benkert et al. 2011, 2009) were used to evaluate the resulting model. Finally, visualization and analysis of the *B. koreanus* RXR model were performed using PyMOL software, version 1.3 (Schrodinger 2010).

5.2.5. Statistical Analysis

For the statistical analyses, SigmaPlot software, version 11.0 was used to conduct one-way ANOVA, followed by the Holm-Sidak test. Student's *t*-test was used to compare wild-type and mutant RXR under the same treatment conditions. Overall, *P*-values <0.05 were considered significant.

5.3. Results

5.3.1. Phylogenetic Analysis of RXR in Metazoan Lineages

A set of currently available genomes and transcriptomes from multiple metazoan phyla were investigated to retrieve RXR amino acid sequences combined with *RXR*-like genes isolation from several species to determine the occurrence of *RXR* in Metazoa (Supplementary Material **Table S5.1**). In all of the analysed species a single sequence was found, with the exception of the platyhelminth *Schistosoma mansoni* in which two *RXR* genes were previously identified (Freebern et al. 1999a, 1999b) and, unexpectedly, bryozoans (**Figure 5.1**). Thus, *RXR* occurs in the vast majority of metazoans examined to date, with the additional novelty of two genes that most likely emerged in the Bryozoa ancestor (**Figure 5.1**).

5.3.2. *In vitro* Interaction of Rotifer RXR with TBT and 9cisRA

As part of a larger study, aiming to decipher the structural interactions of organotins such as TBT and *RXR* orthologs from different metazoan lineages, we proceed with a detailed sequence comparative analysis (Supplementary Material **Figure S5.1**). Of the species considered in the phylogenetic analysis we were able to deduce that the *RXR* ortholog from rotifers does not contain a Cys residue previously found to be critical for organotin binding. Thus, to analyse the *in vitro* interaction of RXR with 9cisRA (natural ligand) and TBT in *B. koreanus*, transfection and transactivation assays were conducted. A significant activation (TBT 1 nM $P = 0.049$; TBT 10 and 100 nM and 9cisRA $P < 0.001$) of firefly luciferase was observed under Bk-RXR exposure at all tested concentrations of TBT and 9cisRA compared to the control (**Figure 5.2A**). Analysis of the amino acidic

composition of the Bk-RXR LBD (**Figure 5.2B**) revealed poor conservation compared to human RXRs (31.9 to 33.6%, data not shown). To determine how differences in the amino acid contents affected the structure of the Bk-RXR LBD, a comparative homology model was employed using the crystal structure of human RXRα LBD complexed with TBT (3E94) as the template. The resulting model revealed that the rotifer RXR pocket was differently modulated compared to the human version (**Figure 5.2C**). To discriminate the residue responsible for binding to TBT, mutant RXR (Cys335Ile) was tested (Supplementary Material **Figure S5.1**) and showed a significant decrease (P = 0.050) of transactivation induction at 1 nM TBT compared to the wild type.

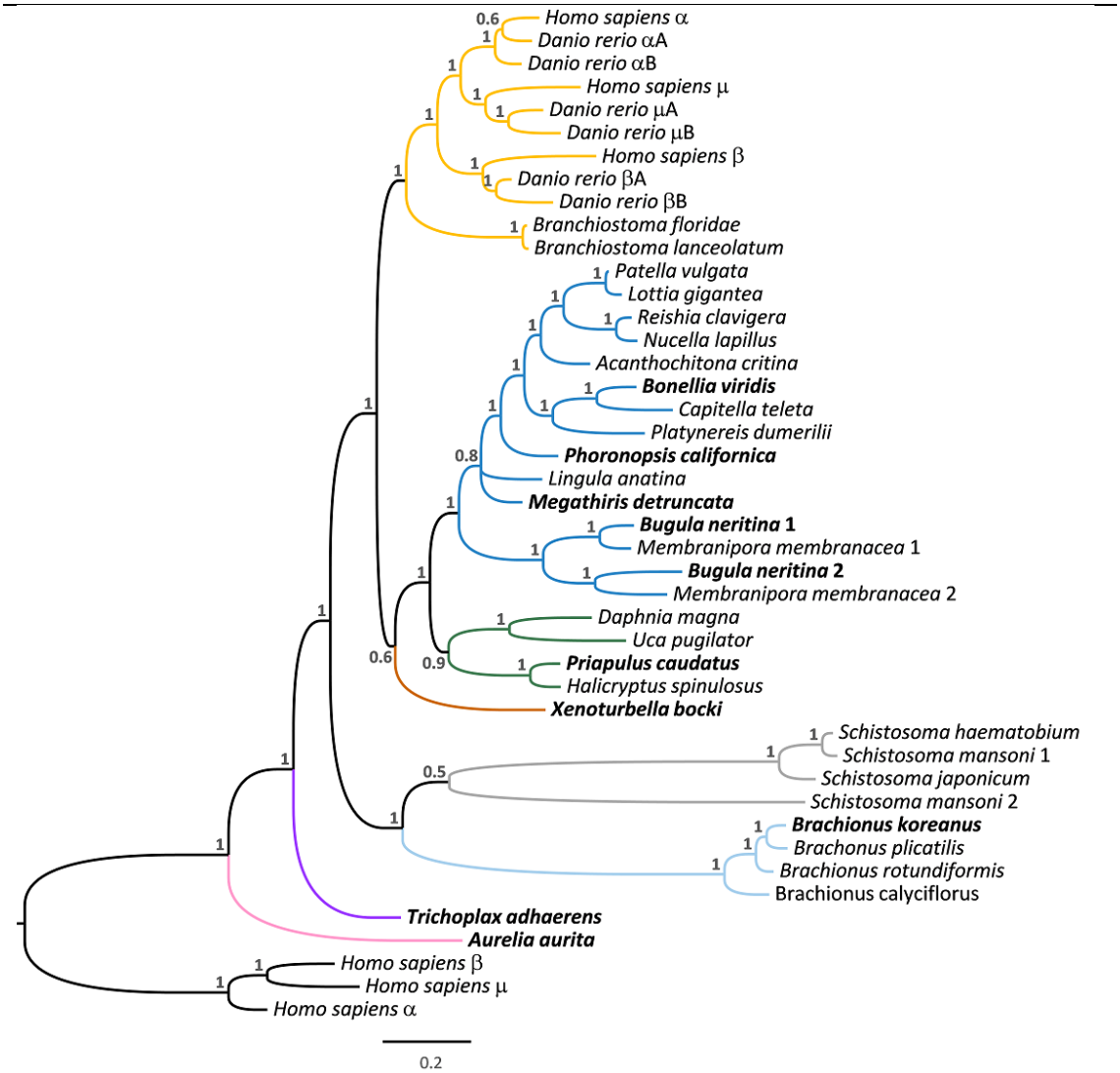


Figure 5.1. Bayesian analysis of the phylogenetic distribution of RXR among metazoan lineages. Chordata (yellow), Lophotrochozoa: Mollusca, Anelida, Phoronida, Brachiopoda and Bryozoa (blue), Platyhelminthes (grey), Rotifera (light blue), Ecdysozoa: Arthropoda and Priaplida (green), Xenoturbellida (orange), Cnidaria (pink), and Placozoa (purple); numbers at nodes indicate Bayesian posterior probabilities. Human RAR amino acid sequences were used as outgroup to root the tree.

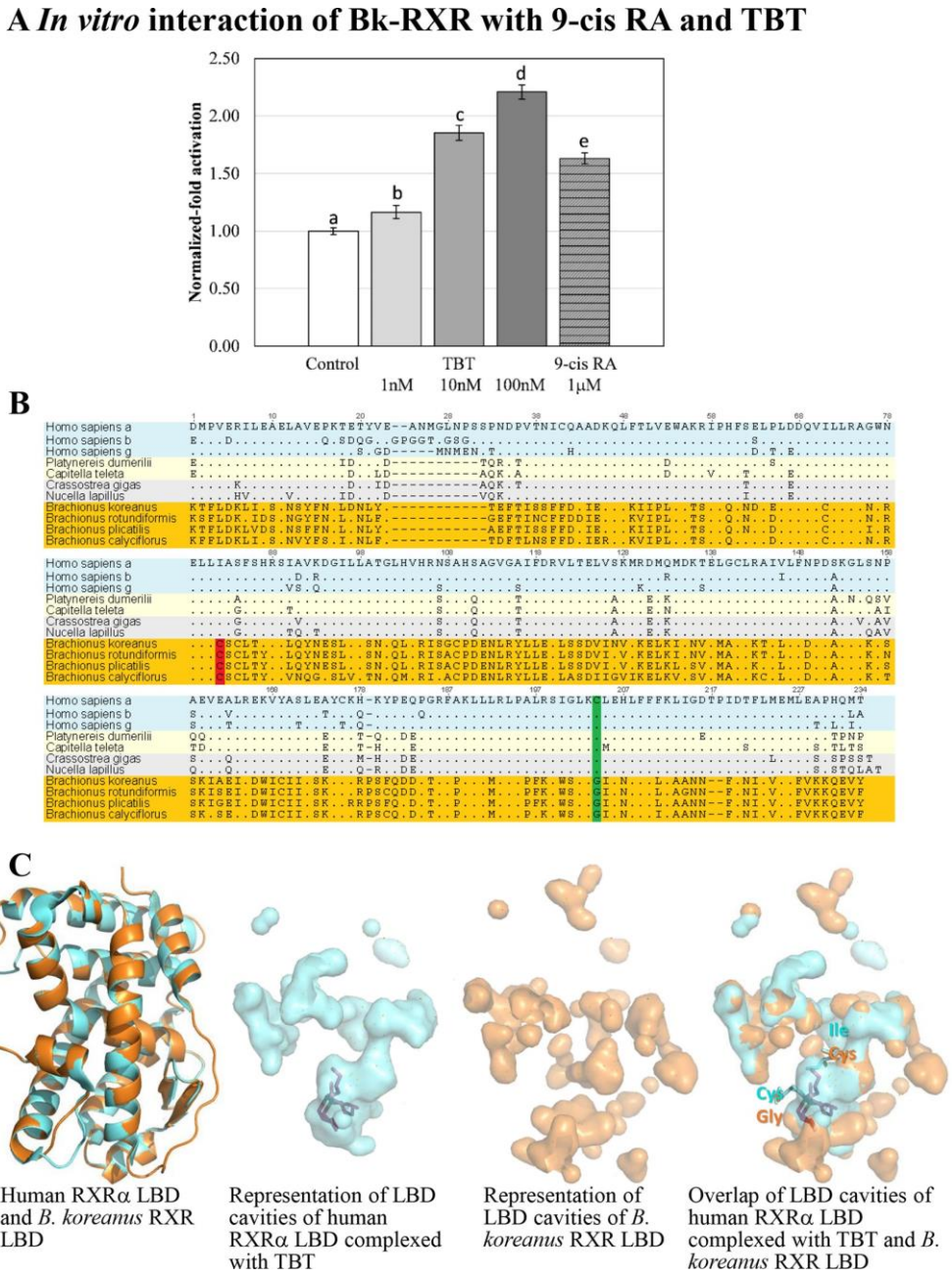


Figure 5.2. (A) Reporter gene activity in chimeric RXR *B. koreanus* upon binding to TBT (1, 10, or 100 nM) or 9cisRA (1 μM), using DMSO as the control. Data are expressed as fold induction over the control (DMSO) and are plotted as the mean ± SEM of three separate experiments. Significant differences are indicated by the letters above the bars ($P < 0.05$). **(B)** Alignment of RXR ligand-binding domains from *Homo sapiens*, *Platynereis dumerilii, Capitella teleta*, *Crassostrea gigas* (XP_011434492.1), *Nucella lapillus*, *Brachionus koreanus*, *B. rotundiformis*, *B. plicatilis*, and *B. calyciflorus*. The cysteine residue responsible for RXR activation upon TBT binding in vertebrates is highlighted in green. The cysteine residue hypothetically responsible for TBT binding in rotifer RXR is highlighted in red. **(C)** Comparative homology model of *Brachionus koreanus* RXR ligand binding domain (LBD); from left to right: homology model of Bk-RXR LBD (orange) and crystal structure of human RXRα LBD (cyan) (3E94); representation of human RXRα LBD cavities complexed with TBT (pink, from crystal structure 3E94); representation of Bk-RXR LBD cavities; overlap of human and *Brachionus koreanus* RXR LBD cavities. Cys, cysteine; Ile, isoleucine; Gly, glycine.

5.4. Discussion

In this work, we successfully identified full length *RXR* gene orthologs in all of the examined metazoan species. These covered an ample set of phyla, configuring the conclusion that *RXR* represents an essential gene for the homeostasis and development of animals (Paps & Holland 2018). Curiously, we were also able to deduce the occurrence of two *RXR* genes in bryozoan species. The physiological consequences of this novelty remain unknown and should be investigated in the future.

In recent decades, multiple studies have described the critical role of NRs in the mediation of endocrine disruptors. Classic examples include imposex in marine snails or obesogenic responses in mammals (Grün et al. 2006). In effect, the later phenotype – lipid impairment – is phylogenetically much wider than previously anticipated. In several vertebrate taxa, exposure to TBT disrupts lipid metabolism (Capitão et al. 2017), due to PPARγ direct activation. Importantly, organotins have also been found to activate at nanomolar levels RXR (Nishikawa et al. 2004; Castro et al. 2007). Since RXR is a heterodimer partner with PPAR, it is unknown whether direct activation of RXR can elicit similar obesogenic outcomes. Moreover, in several metazoan lineages, *PPAR* gene orthologs are absent. Yet, *RXR* as shown here is largely conserved throughout metazoan species. In this study, we focused on rotifers as part of larger study addressing the exploitation of RXR by xenobiotics such as organotins. This group of metazoans is a very popular ecotoxicological model, due to the small size (between 100 and 200 μm), very easy to maintain in controlled laboratory conditions and finally display a very short reproductive cycle (approximately 2 hours) (Dahms et al. 2011; Lubzens 1987). Moreover, genomic resources are available for some species including *B. koreanus* (http://rotifer.skku.edu:8080/Bk). Additionally, a close examination of the rotifer RXR sequence showed the absence of a critical Cys residue, which has been considered critical for TBT binding. Thus, to investigate the interaction of RXR with TBT in *B. koreanus*, transactivation assays were conducted. Firefly luciferase transactivation was significantly induced by rotifer RXR at the tested concentrations. Moreover, sequence comparisons and mutant analysis suggested that the rotifer RXR ligand pocket is modulated differently compared to other species (le Maire et al. 2009; Kanayama et al. 2005; Castro et al. 2007; Grün et al. 2006; Nishikawa et al. 2004). The cysteine residue (Cys432) suggested to be critical for TBT binding (le Maire et al. 2009) is substituted with a glycine (Gly) in rotifer RXR, suggesting that Bk-RXR could be nonresponsive to TBT. However, despite the altered composition of the rotifer RXR LBD, including the substitution of residues key to the binding of TBT to RXR, its ability to transactivate upon binding to TBT was demonstrated. These findings suggest a new binding orientation of

TBT to RXR in rotifers, although the exact residue has not been identified, TBT has been shown to bind to RXR and transactivate gene expression in various species (Capitão et al. 2017; André et al. 2017). Together with the analysis presented here, the *in vivo* effects of TBT exposure on various lipid metabolic variables were also demonstrated (Lee et al. 2019). Overall, the data indicated that TBT induces changes in the fatty acid profiles of the rotifer *B. koreanus*, particularly decreasing the polyunsaturated fatty acids content, and that these changes are likely associated with rotifer-specific signalling or feedback mechanisms of lipid homeostasis, resulting in lipid accumulation.

5.5. Conclusions

In this study, we screened multiple metazoan genomes for the occurrence of RXR. This NR is present in all of the examined species, although with some important differences. We found that RXR from rotifers does not contain a previously considered critical residue. Yet, TBT is able to induce transcription via RXR probably exploring a novel active structural configuration. This finding correlates with the lipid disruption phenotypes observed in this group of animals when exposed to organotins.

5.6. References

André A et al. 2017. Cloning and functional characterization of a retinoid X receptor orthologue in *Platynereis dumerilii*: Na evolutionary and toxicological perspective. Chemosphere. 182:753−761. doi: 10.1016/j.chemosphere.2017.05.064.

Baker M, Sasik R, Gerwick L, Hardiman G. 2012. The Praeger Handbook of Environmental Health. Greenwood Publishing Group Inc: Westport, CT, USA.

Bertrand S et al. 2004. Evolutionary genomics of nuclear receptors: from twenty-five ancestral genes to derived endocrine systems. Mol Biol Evol. 21:1923–1937. doi: 10.1093/molbev/msh200.

Capitão A et al. 2018. Evolutionary Exploitation of Vertebrate Peroxisome Proliferator-Activated Receptor γ by Organotins. Env. Sci Technol. 52:13951–13959. doi: 10.1021/acs.est.8b04399.

Capitão A, Lyssimachou A, Castro L, Santos M. 2017. Obesogens in the aquatic environment: an evolutionary and toxicological perspective. Env. Int. 106:153−169. doi: 10.1016/j.envint.2017.06.003.

Castro L et al. 2007. Imposex induction is mediated through the Retinoid X receptor signalling pathway in the neogastropod *Nucella lapillus*. Aquat Toxicol. 85:57–66. doi: 10.1016/j.aquatox.2007.07.016.

Castro L, Santos M. 2014. To bind or not to bind: the taxonomic scope of nuclear receptor mediated endocrine disruption in invertebrate phyla. Env. Sci Technol. 48:5361−5363. doi: 10.1021/es501697b.

Dahms H, Hagiwara A, Lee J. 2011. Ecotoxicology, ecophysiology, and mechanistic studies with rotifers. Aquat Toxicol. 101:1−12. doi: 10.1016/j.aquatox.2010.09.006.

Fonseca E et al. 2017. LXRα and LXRβ nuclear receptors evolved in the common ancestor of gnathostomes. Genome Biol Evol. 9:222–230. doi: 10.1093/gbe/evw305.

Freebern W, Niles E, LoVerde P. 1999a. RXR-2, a member of the retinoid x receptor family in *Schistosoma mansoni*. Gene. 233:33–38. doi: 10.1016/s0378-1119(99)00161-4.

Freebern W, Osman A, Niles E, Christen L, LoVerde P. 1999b. Identification of a cDNA encoding a retinoid X receptor homologue from *Schistosoma mansoni*. Evidence for a role in female-specific gene expression. J Biol Chem. 274:4577–4585. doi: 10.1074/jbc.274.8.4577.

Grün F et al. 2006. Endocrine-disrupting organotin compounds are potent inducers of adipogenesis in vertebrates. Mol Endocrinol. 20:2141−2155. doi: 10.1210/me.2005-0367.

Grün F, Blumberg B. 2009. Endocrine disrupters as obesogens. Mol Cell Endocrinol. 304:19–29. doi: 10.1016/j.mce.2009.02.018.

Grün F, Blumberg B. 2006. Environmental obesogens: organotins and endocrine disruption via nuclear receptor signaling. Endocrinology. 147:S50−S55. doi: 10.1210/en.2005-1129.

Holzer G, Markov G, Laudet V. 2017. Evolution of Nuclear Receptors and Ligand Signaling: Toward a Soft Key-Lock Model? Curr Top Dev Biol. 125:1–38. doi: 10.1016/bs.ctdb.2017.02.003.

Jones B et al. 2017. Conservation of estrogen receptor function in invertebrate reproduction. BMC Evol Biol. 17:65. doi: 10.1186/s12862-017-0909-z.

Jordão R et al. 2015. Obesogens beyond vertebrates: Lipid perturbation by tributyltin in the crustacean *Daphnia magna*. Env. Heal. Perspect. 123:813–819. doi: 10.1289/ehp.1409163.

Kanayama T, Kobayashi N, Mamiya S, Nakanishi T, Nishikawa J. 2005. Organotin compounds promote adipocyte differentiation as agonists of the peroxisome proliferator-activated receptor/retinoid X receptor pathway. Mol Pharmacol. 67:766−774. doi: 10.1124/mol.104.008409.

Katoh K, Toh H. 2010. Parallelization of the MAFFT multiple sequence alignment program. Bioinformatics. 26:1899–1900. doi: 10.1093/bioinformatics/bty121.

Keay J, Thornton J. 2009. Hormone-activated estrogen receptors in annelid invertebrates: implications for evolution and endocrine disruption. Endocrinology. 150:1731–1738. doi: 10.1210/en.2008-1338.

Lee M et al. 2019. Tributyltin Affects Retinoid X Receptor-Mediated Lipid Metabolism in the Marine Rotifer *Brachionus koreanus*. Env. Sci Technol. 53:7830–7839. doi: 10.1021/acs.est.9b01359.

Lubzens E. 1987. Raising rotifers for use in aquaculture. Hydrobiologia. 147:245−255. doi: 10.1007/BF00025750.

le Maire A et al. 2009. Activation of RXR-PPAR heterodimers by organotin environmental endocrine disruptors. EMBO Rep. 10:367−373. doi: 10.1038/embor.2009.8.

Miller M et al. 2015. A RESTful API for Access to Phylogenetic Tools via the CIPRES Science Gateway. Evol. Bioinform Online. 11:43–48. doi: 10.4137/EBO.S21501.

Nascimento F, Reis M, Yang Z. 2017. A biologist's guide to Bayesian phylogenetic analysis. Nat Ecol Evol. 1:1446–1454. doi: 10.1038/s41559-017-0280-x.

Nishikawa J et al. 2004. Involvement of the retinoid X receptor in the development of imposex caused by organotins in gastropods. Env. Sci Technol. 38:6271−6276. doi: 10.1021/es049593u.

Paps J, Holland P. 2018. Reconstruction of the ancestral metazoan genome reveals an increase in genomic novelty. Nat Commun. 9:1730. doi: 10.1038/s41467-018-04136-5.

Schrodinger L. 2010. The PyMOL Molecular Graphics System. http://www.pymol.org.

Söffker M, Tyler C. 2012. Endocrine disrupting chemicals and sexual behaviors in fish - a critical review on effects and possible consequences. Crit Rev Toxicol. 42:653–668. doi: 10.3109/10408444.2012.692114.

Sumpter J, Johnson A. 2005. Lessons from endocrine disruption and their application to other issues concerning trace organics in the aquatic environment. Env. Sci Technol. 39:4321–4332. doi: 10.1021/es048504a.

Thornton J, Need, E, Crews D. 2003. Resurrecting the ancestral steroid receptor: ancient origin of estrogen signaling. Science. 301:1714–1717. doi: 10.1126/science.1086185.

Tohyama S et al. 2016. Evolution of estrogen receptors in ray-finned fish and their comparative responses to estrogenic substances. J Steroid Biochem Mol Biol. 158:189–197. doi: 10.1016/j.jsbmb.2015.12.009.

Tyler C, Jobling S, Sumpter J. 1998. Endocrine disruption in wildlife: a critical review of the evidence. Crit Rev Toxicol. 28:319–361. doi: 10.1080/10408449891344236.

5.7. Supplementary Material

Sampling and RNA Extraction

Each *Bonellia viridis, Phoronopsis californica*, *Megathiris detruncata*, *Bugula neritina* and *Xenoturbella bocki* specimens were preserved on RNA later (Invitrogen, Carlsbad, CA, USA), immediately after their collection for further total RNA extraction. A small portion of each specimen was homogenized with PureZOL RNA Isolation Reagent® (Bio-Rad, Hercules, CA, USA) and the nucleic acids were extracted with chloroform according to the manufacturer's instructions. The illustra RNAspin Mini RNA Isolation (GE Healthcare, Chicago, IL, USA) kit was used to isolate the total RNA from the aqueous phase obtained in the first step and genomic contamination was prevented by an on-column DNAse I digestion step. RNA was eluted in 30 μL of RNase-free water. The synthesis of first-strand cDNA (1 μg) and of 5´and 3´ cDNA for Rapid amplification of cDNA ends (RACE) were performed with the iScript™cDNA Synthesis Kit (Bio-Rad, Hercules, CA, USA) and the SMARTer™ RACE cDNA Amplification kit (Clontech, Mountain View, CA, USA), respectively, following to the manufacturer's guidelines.

Full ORF *RXR* Genes Isolation

The full ORF RXR genes were obtained by permorming a combination of RACE, nested, hemi-nested and/or degenerate polymerase chain reactions (PCRs) with Phusion Flash High-Fidelity PCR Master Mix (Thermo Fisher Scientific Co., Waltham, MA, USA), according to the manufacturer's protocol. Sequences obtained with degenerated PCR primers and BLAST searches were used to design gene specific primers. Degenerated, specific RACE and full ORF primers are listed on Supplementary Material **Table S5.2**. The obtained fragments were purified using NZYGelpure (Nzytech, Lisbon, Portugal), cloned into Nzy5α competent cells (Nzytech, Lisbon, Portugal), using the pGEM-T Easy Vector System (Promega, Madison, WI, USA). All the obtained sequences were confirmed by sequencing (Eurofins GATC, Constance, Germany).

Table S5.1. List of sequences used for phylogenetic analysis of RXR in metazoan lineages and the corresponding accession numbers.

Phylum	Species	Nuclear receptor	Accession number
Chordata	*Homo sapiens*	RARα	NP_000955.1
		RARβ	NP_000956.2
		RARγ	NP_000957.1
		RXRα	NP_002948.1
		RXRβ	NP_068811.1
		RXRγ	NP_008848.1
	Danio rerio	RXRα	a NP_001155023.1; b NP_571228.1
		RXRβ	a NP_571350.1; b NP_57313.1
		RXRγ	a NP_571292.3; b XP_005160723.1
	Branchiostoma floridae	RXR	AAM46151.1
	Branchiostoma lanceolatum	RXR	ANP24206.1
Mollusca	*Lottia gigantea*	RXR	ESO92876.1
	Patella vulgata	RXR	ALQ43971.1
	Nucella lapillus	RXR	ABS70715.1
	Reishia clavigera	RXR	AAU12572.1
	Acanthochitona crinita	RXR	QAX24918.1
Annelida	*Bonellia viridis*	RXR	Bvi_RXR[a]
	Platynereis dumerilii	RXR	AVR59237.1
	Capitella teleta	RXR	ELT93409.1
Phoronida	*Phoronopsis californica*	RXR	Pca_RXR[a]
Brachiopoda	*Megathiris detruncata*	RXR	Mde_RXR[a]
	Lingula anatina	RXR	XP_013412668.1
Bryozoa	*Bugula neritina*	RXR1	Bne_RXR1[a]
		RXR2	Bne_RXR2[a]
	Membranipora membranacea	RXR1	SRX1121923
		RXR2	SRX1121923
Plathyhelminthes	*Schistosoma japonicum*	RXR	AFP95235.1
	Schistosoma haematobium	RXR	XP_012793373.1
	Schistosoma mansoni	RXR1	XP_018645908.1
		RXR2	AAD45325.1
Rotifera	*Brachionus koreanus*	RXR	ASL70628.1
	Brachionus plicatilis	RXR	ASL70592.1
	Brachionus rotundiformis	RXR	ASL70517.1
	Brachionus calyciflorus	RXR	ASL70559.1
Priapulida	*Priapulus caudatus*	RXR	QFQ33540.1
	Halicryptus spinulosus	RXR	SRX1343820
Arthropoda	*Uca pugilator*	RXR	AAC32789.3
	Daphnia magna	RXR	ABF74729.1
Xenoturbellida	*Xenoturbella bocki*	RXR	Xbo_RXR[a]
Cnidaria	*Aurelia aurita*	RXR	AGT42223.1
Placozoa	*Trichoplax adhaerens*	RXR	XP_002109459.1

[a]These sequences were isolated in this stuby by PCR

Table S5.2. List of primer sets used to isolate *RXR* genes. F, forward; R, reverse; Tm, primer melting temperature.

PCRs		Primer sequences (5′→ 3′)	Tm (ºC)	Cycles
	Bonellia viridis			
5' RACE	R1	CGATCAGCAACTCATTCCAACCTGC	67	45
	R2	CCAGGCACTTCATGTAGCGGCAGTA	67	45
	R3	TTGCACCCTTCGCAGCTGTAAACAC	67	45
Full ORF	F	ATGCACGGTAACAGACAGC	57	35
	R	TCATGTGTTGGTCTGGAGAG		
	Phoronopsis californica			
Degenerated	F	GCAGCTGGTCACCCTGGTNGARTGGGC	50	45
	R	GGCAGCCCAGCTCGGTYTTRTCCAT		
5' RACE	R1	TACCAACGCCTGCTTGATGAGCACT	67	45
	R2	TGAAGTGAGGCACCCTTTTAGCCCA	67	45
3' RACE	F1	TGGGCTAAAAGGGTGCCTCACTTCA	67	45
	F2	GCAGGCGTTGGTACCATATTTGACAG	67	45
Full ORF	F	GTGAAATGAAAAATGGATAGATC	55	35
	R	CTAACTGACAGGACTCGG		
	Megathiris detruncata			
Degenerated	F	GCAGCTGGTCACCCTGGTNGARTGGGC	50	40
	R	GGCAGCCCAGCTCGGTYTTRTCCAT		
5' RACE	R1	GAACATGTAAACCTGTAGCAAGCAGG	65	45
	R2	GCAATTAGGAGTTCATTCCAGCCAGC	65	35
	R3	TTCTTTCATCTCGACAGGCGTAGG	65	40
3' RACE	F1	CCTGCTTGCTACAGGTTTACATGTTC	65	45
	F2	CTGAACTGGTAGCTAAGATGAGAGAA	65	35
Full ORF	F	ATGGGTACTATAACATGGGTA	55	35
	R	TCATGTTGTTGGGCTGGGA		
	***Bugula neritina* 1**			
Partial ORF	F	CGTGTCGAGATGACAAATGC		
	R1	GTCGAAGATTGTCCCTACGC	58	40
	R2	CTATACTGCGCAAGGCAGGT		
5' RACE	R1	GTCGACGTACTGAGAGTTGCTTGGCTC	65	40
	R2	TCTGTCTGTGCTGCTCAGTACAGCGTC	65	40
	R3	CACGCAGCATTTGTCATCTCGACAC	65	45
	R4	TGAGGTGAACCGAGGCTTGAGTGAG	68	40
3' RACE	F1	GAGCCAAGCAACTCTCAGTACGTCGAC	65	40
	F2	CAGATCCATGGCTGTTCAAGATGGAATAC	65	40
Full ORF	F	CTAACCGTGAAATGAATGGTC	57	35
	R	TTATTGCAGAGGGGCTTCTAA		
	***Bugula neritina* 2**			
5' RACE	R1	TGTTGGTGAGAGGGTCGGAGGCTAT	65	45
	R2	CCGTCTCCGTGACTGTTGGACTTTG	65	45
3' RACE	F1	AGCTCACCTGTCTCAAGGCCATCGT	69	45
	F2	CCGGCAGGTTTGCTAAGCTCCTTCT	69	45
Full ORF	F	ATGGGGATGAATGGAGCG	56	35
	R	CTAAGAGAGGTGTGTGTGG		
	Xenoturbella bocki			
Partial ORF	F	CAATCTGTGCTATTTGTGGTG	57.5	40
	R	CATCTCCATCAAGAACGAATC		
5' RACE	R1	TCCACTCCCAGCTCAGCCTCTAGGA	65	45
	R2	CAGTGCTGGCAGCGATTTCTCTGTT	65	45
	R3	CCGGAGATGTTGTGAGTAAGAGTGG	65	45
3' RACE	F1	GAGTCCTCACAGAGCTCGTCGCAAA	65	45
	F2	ACTAACGCAGACGACCCCAGCAGAT	65	45
Full ORF	F	CAGTACAATGAATCCCATAGG	57	35
	R	GTTAAGCCATCGTTGTTGTTG		

Figure S5.1. Alignment of RXR ligand-binding domains from *Homo sapiens*, *Branchiotoma lanceolatum, Bonellia viridis*, *Phoronopsis californica*, *Megathiris detruncata*, *Bugula neritina*, *Priapulus caudatus*, *Xenoturbella bocki, Aurelia aurita*, and *Trichoplax adhaerens*. The cysteine residue responsible for RXR activation upon TBT binding in vertebrates is highlighted in green.

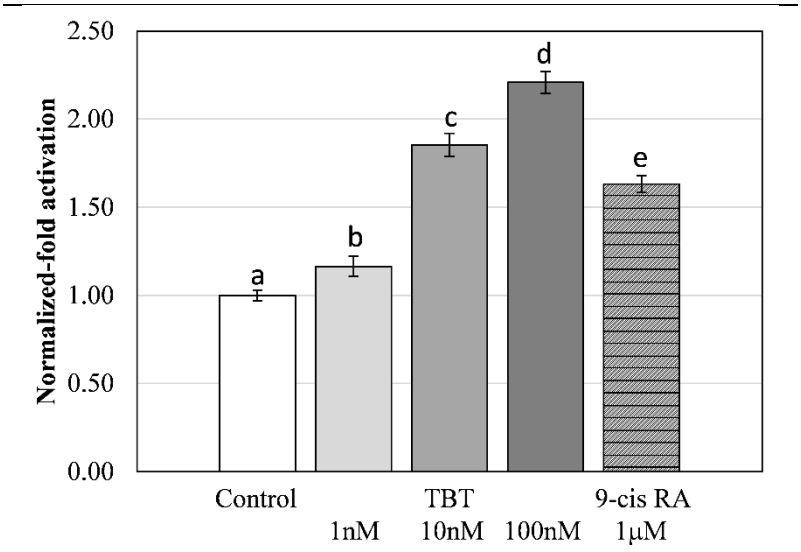


Figure S5.2. Reporter gene activity induced by the chimeric *B. koreanus* RXR mutant upon binding to TBT (1, 10, or 100 nM) or 9cisRA (1 μM). The data are expressed as fold induction over the control (DMSO) and are plotted as the mean±SEM of three separate experiments. Significant differences between the treatments and between the mutant and the wild type for the same treatment are indicated as the letters and * above each bar, respectively ($P < 0.05$).

CHAPTER 6 - An Ortholog of the Retinoic Acid Receptor (RAR) Is Present in the Ecdysozoa Phylum Priapulida

genes

MDPI

Article

An orthologue of the Retinoic Acid Receptor (RAR) Is Present in the Ecdysozoa Phylum Priapulida

Elza S. S. Fonseca [1,2,†], Youhei Hiromori [3,4,†], Yoshifumi Kaite[3], Raquel Ruivo[1], João N. Franco[1], Tsuyoshi Nakanishi[3,*], Miguel M. Santos[1,2,*]and L. Filipe C. Castro [1,2,*]

[1] CIIMAR/CIMAR - Interdisciplinary Centre of Marine and Environmental Research, U.Porto, 4450-208 Matosinhos, Portugal; fonseca.ess@gmail.com (E.F.); ruivo.raquel@gmail.com (R.R.); joaonunofranco@gmail.com (J.N.F.)
[2] FCUP - Faculty of Sciences, Department of Biology, U.Porto, 4169-007 Porto, Portugal
[3] Laboratory of Hygienic Chemistry and Molecular Toxicology, Gifu Pharmaceutical University, Gifu 501-1196 Japan; hayake5@hotmail.com (Y.H.); 135018@gifu-pu.ac.jp (Y.K.)
[4] Faculty of Pharmaceutical Sciences, Suzuka University of Medical Science, Suzuka 513-8670 Japan
* Correspondence: filipe.castro@ciimar.up.pt (L.F.C.C.); nakanishi@gifu-pu.ac.jp (T.N.); santos@ciimar.up.pt (M.M.S.)
† These authors contributed equally to this work

Received: 18 October 2019; Accepted: 27 November 2019; Published: 29 Novenber 2019

6. An Ortholog of the Retinoic Acid Receptor (RAR) Is Present in the Ecdysozoa Phylum Priapulida

6. Abstract

Signalling molecules and their cognate receptors are central components of the Metazoa endocrine system. Defining their presence or absence in extant animal lineages is critical to accurately devise evolutionary patterns, physiological shifts and the impact of endocrine-disrupting chemicals. Here, we address the evolution of retinoic acid (RA) signalling in the Priapulida worm, *Priapulus caudatus* Lamarck, 1816, an Ecdysozoa. RA signalling has been shown to be central to chordate endocrine homeostasis, participating in multiple developmental and physiological processes. Priapulids, with their slow rate of molecular evolution and phylogenetic position, represent a key taxon to investigate the early phases of Ecdysozoa evolution. By exploring a draft genome assembly, we show by means of phylogenetics and functional assays, that an ortholog of the nuclear receptor retinoic acid receptor (RAR) subfamily, a central mediator of RA signalling, is present in Ecdysozoa, contrary to previous perception. We further demonstrate that the Priapulida RAR displays low-affinity for retinoids (similar to annelids), and is not responsive to common endocrine disruptors acting *via* RAR. Our findings provide a timeline for RA signalling evolution in the Bilateria and give support to the hypothesis that the increase in RA affinity towards RAR is a late acquisition in the evolution of the Metazoa.

Keywords: Bilateria; nuclear receptors; xenobiotics; endocrine system

6.1. Introduction

Retinoic acid (RA) is a critical regulator of multiple biological processes in vertebrates including cell differentiation and embryonic development (Ross et al. 2000; Samarut & Rochette-Egly 2012), central nervous system development (Niederreither & Dollé 2008; Escrivà et al. 2002; Schubert et al. 2004), organ formation and tissue maintenance (Cunningham & Duester 2015; Maden & Hind 2003; Bertrand, S et al. 2007) and vision (Cvekl & Wang 2009). Retinoids, such all-*trans* retinoic acid (ATRA), 9-*cis* retinoic acid (9cisRA) and 13-*cis* retinoic acid (13cisRA) are active metabolites of vitamin A (retinol), known to bind and modulate the retinoic acid receptor (RAR) and retinoid X receptor (RXR), the central mediators of RA signalling (Albalat 2009). RAR and RXR belong to the nuclear receptor (NR) superfamily and are ligand-dependent transcription factors that regulate the expression of specific genes (Rochette-Egly & Germain 2009;

Germain et al. 2006). RAR heterodimerizes with RXR and recognizes specific RA responsive elements (RAREs) in the regulatory region of the target gene (Germain et al. 2003; Chatagnon et al. 2015). Upon binding to ligands, the position of the helix 12 on the ligand-binding domain (LBD) is modified, allowing the recruitment of coactivators and consequently, the activation of gene transcription (Rochette-Egly & Germain 2009; Chatagnon et al. 2015; Germain et al. 2003). The emergence of various non-chordate genome sequences established that RA signalling is not chordate-specific, since signalling components such as *RAR* and *RXR* gene orthologs have been found in species from Ambulacraria (echinoderms and hemichordates) (Albalat & Cañestro 2009; Cañestro et al. 2006; Howard-Ashby et al. 2006; Marlétaz et al. 2006; Ollikainen et al. 2006; Simões-Costa et al. 2008). Recently, RAR was also functionally characterized in various mollusc species and in a second Lophotrochozoa clade, the annelid worm *Platynereis dumerilii*. Yet, functional studies in mollusc species demonstrated the loss of RA binding affinity towards RAR (Campo-Paysaa et al. 2008; Gutierrez-Mazariegos et al. 2014; Urushitani et al. 2013; André et al. 2019). In contrast, the annelid *Platynereis RAR* ortholog showed a conserved capacity to bind and respond to retinoids, but with lower affinity compared to vertebrate *RAR* paralogs (Handberg-Thorsager et al. 2018). Further, it was demonstrated that RAs trigger neuronal differentiation, a role previously described only in chordates (Janesick et al. 2015; Campo-Paysaa et al. 2008; Escrivà et al. 2002). Surprisingly, in a second annelid species, Helobdella robusta (leech), *RAR* and other RA signalling components are absent (Albalat & Cañestro 2009).The search for *RAR* gene orthologs in Ecdysozoa (e.g. arthropods and nematodes) genomes was previously unsuccessful, implying that *RAR* was probably lost in this superphylum (Albalat & Cañestro 2009). Overall, these studies suggest that 1) *RAR* evolution was shaped by events of secondary gene loss during Bilateria evolution, notably in the whole Ecdysozoa lineage and Appendicularia (Tunicata) (Albalat & Cañestro 2009; Marlétaz et al. 2006; Cañestro et al. 2006) and 2) the bilaterian *RAR* ancestor was a RA low-affinity sensor, with the ability to bind retinoids and activate transcription of target genes – annelids (*Platynereis*), or a receptor without capacity to bind ligands as seen in molluscs (Handberg-Thorsager et al. 2018; Urushitani et al. 2013; Gutierrez-Mazariegos et al. 2014; André et al. 2019). Additionally, NRs are prime targets of endocrine-disrupting chemicals (EDCs) (e.g (Castro & Santos 2014)), with various examples denoting the impact of EDCs acting *via* NRs (Capitão et al. 2017; Balaguer et al. 2017). Yet, variations in NR gene complement and sequence variation as well as the molecular architecture of endocrine systems are of paramount importance to recognize the mechanisms of action of EDCs (Keay & Thornton 2009; Vogeler et al. 2017), particularly in invertebrate lineages [e.g. (Katsiadaki 2019; Scott 2018)]. In the specific case of *RAR*, environmental

contaminants such as pesticides have been shown to exploit mammalian RARs (Lemaire et al. 2005), but not the molluscan *RAR* (André et al. 2019). To further scrutinize the evolution RA signalling, specifically if *RAR* is absent or present in other extant Ecdysozoa lineages, we investigated the genome of the penis worm, *Priapulus caudatus* Lamarck, 1816 (Scalidophora, Priapulida). Priapulids are mud-dwelling, carnivorous marine worms with a tubular body shape and an eversible proboscis (**Figure 6.1A**; see **Video S6.1**) that altered little since the arthropod/priapulid common ancestor (over 520 million years ago) (Sansom 2016; Webster et al. 2006; Wills et al. 2012). Their morphological and developmental characteristics together with their slow rate of molecular evolution suggest Priapulida as a key phylum to understand the evolution of Ecdysozoa.

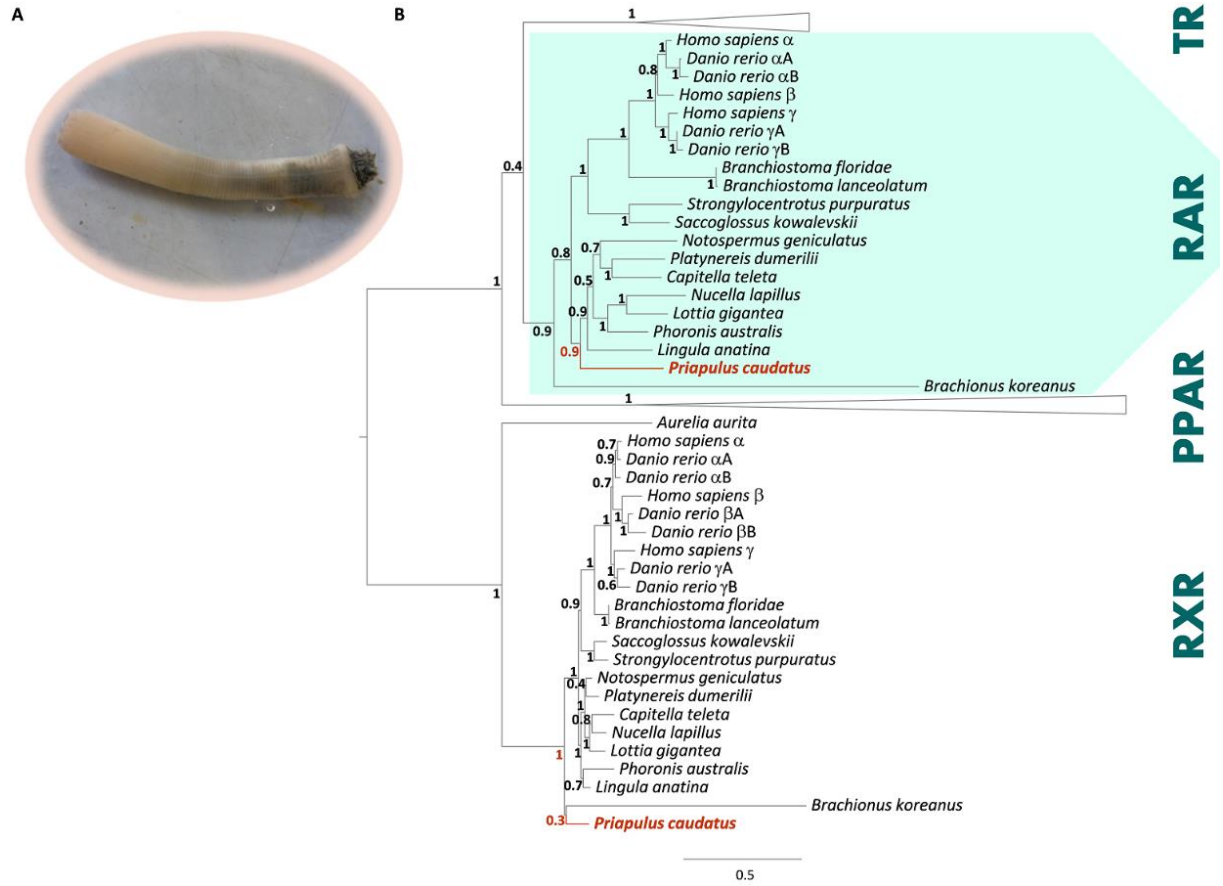


Figure 6.1. The retinoic acid receptor (RAR) and retinoid X receptor (RXR) nuclear receptors of *P. caudatus*. **(A)** A specimen photograph of *P. caudatus*. **(B)** Maximum likelihood phylogenetic tree of RAR, RXR, thyroid hormone receptor (TR) and peroxisome proliferator-activated receptor (PPAR) nuclear receptors; numbers at nodes indicate posterior probabilities calculated using aBayes. Photograph by João N. Franco.

6.2. Material and Methods

6.2.1. Sampling

One adult and two juvenile specimens of *P. caudatus* were collected at the Gullmarn fjord, Sweden, and preserved on RNA later (Invitrogen, Carlsbad, CA, USA) for further total RNA extraction.

6.2.2. RNA Extraction

The adult priapulid was dissected into small portions and homogenized with PureZOL RNA Isolation Reagent® (Bio-Rad, Hercules, CA, USA). The extraction of nucleic acids was performed with chloroform according to the manufacturer's instructions and the resulting aqueous phase was used to isolate the total RNA with the illustra RNAspin Mini RNA Isolation (GE Healthcare, Chicago, IL, USA) kit. A step of on-column DNAse I digestion was included to exclude genomic contamination and the RNA was eluted with RNase-free water, starting from the ethanol step. The iScript™cDNA Synthesis Kit (Bio-Rad, Hercules, CA, USA) was used for cDNA synthesis and performed according to the manufacturer's recommendations, using 1000 ng of the RNA previously isolated.

6.2.3. *RXR* and *RAR* Gene Isolation

A BLAST approach conducted on the publicly available *P. caudatus* draft genome (GCA_000485595.2, Priapulus_caudatus-5.0.1) to investigate the presence of *RAR* and *RXR*-like sequences. The open reading frame (ORF) of *P. caudatus RAR* and *RXR* were deduced from the genome assembly and isolated using a polymerase chain reaction (PCR) with specific primers (Supplementary Material **Table S6.1**). In the case of *RXR*, two partial nucleotide sequences were used to design specific primers (Supplementary Material **Table S6.1**) and a partial *P. caudatus RXR* containing the termination codon was isolated by PCR. To obtain the remaining sequence, the partial RXR isolated sequence was extended using the SMARTer™ RACE cDNA Amplification Kit (Clontech, Mountain View, CA, USA) following the manufacturer instructions, using specific RACE PCR primers (Supplementary Material **Table S6.1**). The Phusion Flash High-Fidelity PCR Master Mix (ThermoFisher, Waltham, MA, USA) was used in all PCR reactions and the obtained products were purified with NZYGelpure (Nzytech, Lisbon, Portugal), cloned into pGEM-T Easy Vector System (Promega, Madison, WI, USA). The sequences were confirmed by automated Sanger sequencing (Eurofins GATC). *RAR* and *RXR P. caudatus* sequences have been deposited in GenBank (Accession numbers MK780070 and MK780071, respectively).

6.2.4. Sequence and Phylogenetic Analysis

Amino acid sequences of RAR, RXR, thyroid hormone receptor (TR) and peroxisome proliferator-activated receptor (PPAR) from various Metazoa taxonomic groups were recovered through BLASTp searches in GenBank, Joint Genome Institute (JGI) Genome Portal and Okinawa Institute of Science and Technology (OIST) Marine Genomics Unit Genome Browser. Retrieved sequences and corresponding protein accession numbers are listed in the Supplementary Material **Table S6.2**. The collected sequences were aligned with Multiple Alignment using Fast Fourier Transform (MAFFT) programme v7 (L-INS-i method) (Katoh & Toh 2010), visualized and edited in Geneious ®v7.1.7. Based on previous studies (Renaud et al. 1995; Hisata et al. 1998; Gesto et al. 2016; Handberg-Thorsager et al. 2018; Escrivà et al. 2006), the amino acids residues that interact with ATRA were identified. The columns containing gaps were stripped, resulting in a final alignment contained 71 sequences and 277 positions. A Maximum Likelihood phylogenetic analysis was performed using the PhyML 3.0 server with the amino acid substitution model LG + G + I and the evolutionary model automatically selected (Guindon et al. 2010). The branch support for phylogenetic trees was calculated using aBayes. FigTree v1.3.1 was used to visualize the tree.

6.2.5. Construction of Plasmid Vectors

The hinge region and LBD of *RAR* and *RXR* were isolated from human and penis worm by PCR with specific primers (Supplementary Material **Table S6.3**) and cloned into pBIND and/or pACT vectors (Promega, accession numbers AF264722 and AF264723.1), to produce “chimeric” receptors with the yeast transcriptional activator GAL4 (RAR-LBD-GAL4) or the viral enhancer, VP16 (RXR-LBD-VP16), which acts on proximal downstream promoters, respectively (Duffy 2002; Hagmann et al. 1997). The priapulid RAR LBD was amplified by PCR from pGEM-pCauRAR with specific primer (FP: 5’- ACTGGATCCTCGATTATGTCTATGCAACAGCGA -3’, RP: 5’- GATTCTAGAAC TAGTGATTTCACGGTATGCAG -3’) and the product was digested with BamHI and XbaI. The digested fragment was subcloned into BamHI–XbaI site of pCold-TF vector (TAKARA bio, accession number AB213654), for the priapulid RAR LBD−His6-tagged trigger factor hybrid protein. Plasmid sequences were confirmed using automated Sanger sequencing (Eurofins GATC). The human RARα LBD was previously cloned into pGEX-4T-1 vector (GE Healthcare Life Sciences, accession number U13853), for the human RARα LBD-Glutathione S-transferase hybrid protein (Gutierrez-Mazariegos et al. 2014)

6.2.6. Chemicals and Solutions

ATRA, 9cisRA, 13cisRA, endrin, dieldrin, and sterile Dimethyl sulfoxide (DMSO) were purchased from Sigma-Aldrich (St. Louis, MO, USA). The stock solutions were prepared in DMSO: ATRA, 9cisRA and 13cisRA at 0.1, 1 and 10 mM, endrin and dieldrin at 10 mM.

6.2.7. Cell Culture and Transactivation Assays

Cos-1 cells were maintained in Dulbecco's Modified Eagle Medium (DMEM) (PAN-Biotech, Aidenbach, Bayern, Germany). A supplementation with 10% fetal bovine serum (PAN-Biotech, Aidenbach, Bayern, Germany) and 1% penicillin/streptomycin (PAN-Biotech, Aidenbach, Bayern, Germany) was used. Cells were maintained at 37 °C with a humidified atmosphere and 5% CO_2. Cells were seeded in 24-well culture plates at a density of 2×10^5 live cells/well and after 24 hours, cells were transfected with 0.5 µg of pBIND constructs (pBIND-PcauRAR-LBD or pBIND-HsaRARγ-LBD (positive control)) and 1µg of pGL4.31 luciferase reporter vector (DQ487213; Promega), containing five upstream activation sequence (UAS) elements upstream the firefly luciferase reporter gene or, in the case of heterodimer transfection assays, with 0.5 µg of pBIND constructs (pBIND-PcauRAR-LBD or pBIND-HsaRARγ-LBD), 0.5 µg pACT constructs (pACT-PcauRXR-LBD or pACT-HsaRXRα-LBD (positive control)) and 0.5 µg of pGL4.31, using lipofectamine 2000 reagent (Invitrogen, Carlsbad, CA, USA) in Opti-MEM (Gibco, Carlsbad, CA, USA), to a final volume of 350 µL. After 5 h of incubation, transfection media was replaced by DMEM phenol-free supplemented with 10% dextran-coated charcoal-treated serum (Invitrogen, Carlsbad, CA, USA) and 1% penicillin/streptomycin (Invitrogen, Carlsbad, CA, USA) containing the test compounds. Final concentrations were the follows: 0.1, 1 and 10 µM ATRA, 9cisRA or 13cisRA and 10 µM organochlorine pesticides (endrin and dieldrin) dissolved in DMSO (0.1%). Cells were lysed 24 h after transfection. Firefly luciferase (reporter pGL4.31) and *Renilla* luciferase (pBIND) activities were assayed with Dual-Luciferase Reporter Assay System (Promega, Madison, WI, USA), according to the manufacturer's instructions. Two technical replicates per condition in three independent assays were performed for all transfections. The results were expressed as fold-induction resulting from the ratio between luciferase (reporter pGL4.31) and *Renilla* (internal control for transfection efficiency luminescent activity), and then normalized by the DMSO control. Transactivation data was calculated as means of the normalized values (n = 3) and the bars with standard error of the mean (SEM) from the three separate experiments. The means of the technical replicates were used for statistical analysis with one-way ANOVA, followed by Holm-Sidak method in

SigmaPlot software v11.0. Data were transformed whenever the normality failed. The level of significance was set to 0.05.

6.2.8. Ligand Binding Assays

The PcauRAR LBD−His6-tagged trigger factor hybrid protein was expressed in Escherichia coli BL21 (DE3) cells containing the chaperon plasmid pG-Tf2 (Takara Bio, Kusatsu, Shiga, Japan) and purified by using His-select nickel affinity gel (Sigma−Aldrich, St. Louis, MO, USA). The HsaRARα LBD-Glutathione S-transferase hybrid protein was expressed in Escherichia coli BL21 (DE3) cells and purified by Glutathione Sepharose 4B (GE Healthcare, Chicago, IL, USA) (positive control). Ligand binding assay was assessed as previously described (Nakanishi et al. 2005; Hiromori et al. 2009, 2016; Gutierrez-Mazariegos et al. 2014). In brief, the purified protein (12.5 µg/mL) was incubated with 10 nM of all-*trans* retinoic acid [11, 12-^{3}H] ([^{3}H]ATRA; 1.665 TBq/mmol; Amersham Biosciences) or 10 nM of 9-*cis* retinoic acid, [11, 12-^{3}H] ([^{3}H]9cisRA; 1.95 TBq/mmol; PerkinElmer). Unlabelled ATRA and 9cisRA were used to compete for [3H]ATRA or [3H]9cisRA in this assay to determine the binding preferences of PcauRAR LBD or HsaRARα LBD Hydroxyapatite was added to precipitate the receptor protein and bound radioactive compounds After after an incubation step at 4 °C for 1 h, hydroxyapatite was added to precipitate the receptor protein and bound radioactive compounds. The hydroxyapatite pellet was washed., and thenThe radioactivity in the pellet was determined by liquid scintillation counting.

6.3. Results

6.3.1. Phylogenetic and Sequence Analysis of Priapulid *RAR*

By thoroughly examining a genome draft of *P. caudatus*, we established the presence of sequences with high similarity to *RAR* and *RXR* respectively. Since *RAR* sequences are absent from previously analysed Ecdysozoa genomes, we went on to validate this initial screening. We experimentally isolated the full-length sequences of both *RAR* and *RXR* from *P. caudatus* cDNA. These encode two protein sequences with 491 (RAR) and 404 (RXR) amino acids. To determine the orthology of the isolated sequence a phylogenetic analysis was conducted (**Figure 6.1B**), including RAR, RXR, TR and PPAR amino acid sequences of vertebrates (human and zebrafish), cephalochordates (Florida and European lancelet), ambulacrarians (acorn worm and purple sea urchin), molluscs (Atlantic dogwhelk and owl limpet), annelids (Dumeril's clam worm and polychaete worm), a nemertean (ribbon worm), a phoronid (phoronid worm),

a brachiopod (common Oriental lamp shell), a rotifer (Korean monogonont rotifer), and a cnidarian (moon jelly).

The predicted RAR (PcauRAR) and RXR (PcauRXR) sequences of the priapulid worm robustly clustered in the respective clade (**Figure 6.1B**). Next, we examined the amino acid sequence alignment with PcauRAR (**Figure 6.2**), which revealed that, similarly to other RARs, PcauRAR has a conserved modular structure typical of NRs,

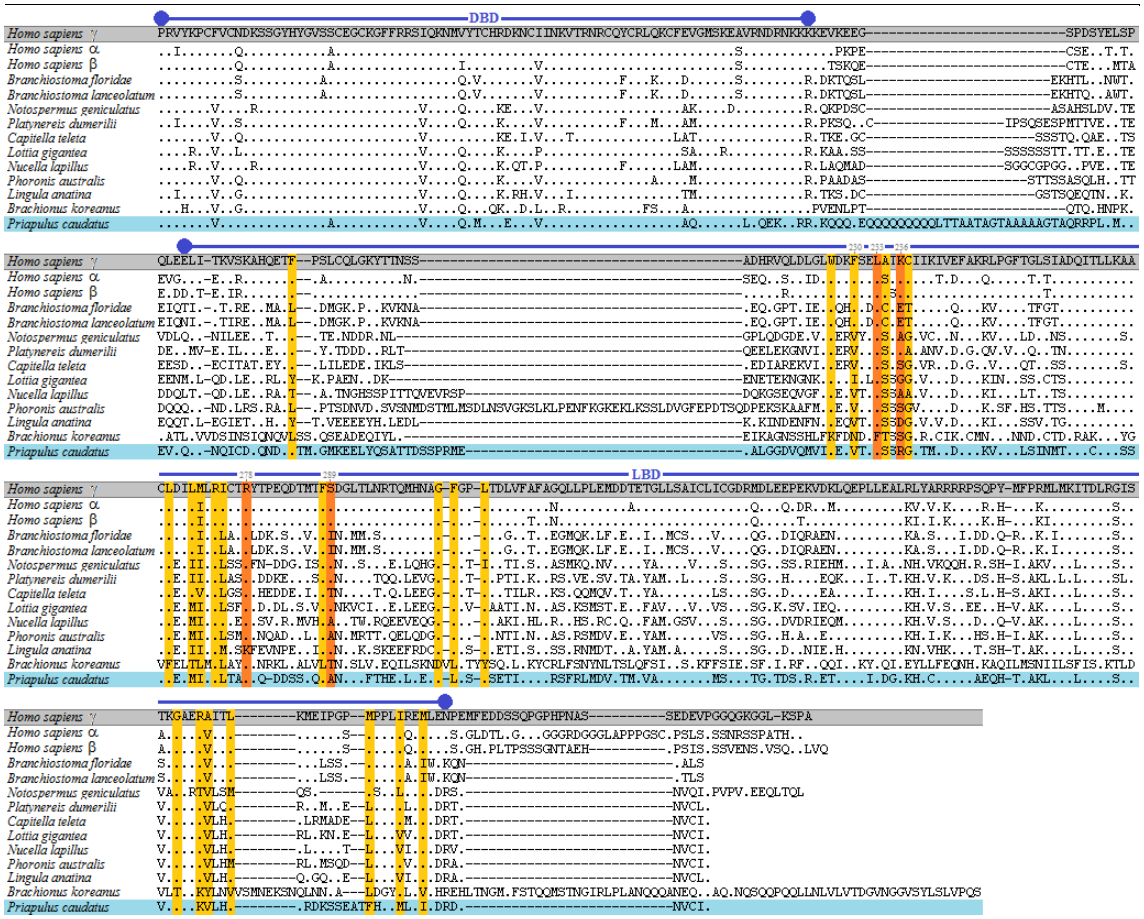


Figure 6.2. Amino acid sequence alignment of the RAR DNA- and ligand- binding domains from human, lancelets, molluscs, annelids, nemertean, brachiopod, phoronid, rotifer and priapulid RAR protein sequences. Key amino acid residues that interact with ATRA in the human RARγ ligand-binding pocket (LBP) are highlighted: orange - direct or indirect hydrogen bonds, yellow - hydrophobic and Van der Waals interactions (Renaud et al. 1995; Hisata et al. 1998; Gesto et al. 2016; Handberg-Thorsager et al. 2018; Escrivà et al. 2006). The DBD and LBD are delimited by the upper blue lines.

with a conserved DNA- binding domain (DBD) and a moderately conserved LBD (Aranda & Pascual 2001). The PcauRAR-DBD shares approximately 80-82% of sequence identity with human and molluscs RARs, 84% with annelids RARs, 81-87% with nemertean, brachiopod and phoronid RARs, and 78% with rotifer RAR. Regarding the PcauRAR-LBD, the identity with human RARs decreases to 39-41%, with molluscs

RARs to 47-49%, with annelids, nemertean and brachiopod RARs to 49-54%, and with phoronid and rotifer RARs to only 33% and 20%, respectively (**Figure 6.2**).

Analysis of the key amino acid residues known to interact with ATRA in human RARγ (Hsa RARγ) ligand binding pocket (LBP) (Renaud et al. 1995; Escrivà et al. 2006), showed that 14 out of 25 are different (**Figure 6.2**), a feature also observed in previous studies: 9 to 11 amino acids in mollusc (André et al. 2019) and 8 in annelids (Handberg-Thorsager et al. 2018). In nemertean 12 amino acids are not conserved, and, similarly to molluscs, 9 amino acids are not conserved in brachiopod and phoronid. In rotifer only 3 of the key 25 amino acids are conserved. Regarding these 25 amino acids residues, 4 (Leu233, Lys236, Arg278, and Ser289) participate in stable hydrogen bonds with the carboxyl group of ATRA in HsaRARγ. In PcauRAR, as well as, in phoronid and brachiopod RARs, two of these residues are not conserved (Lys236>Arg, Ser, Asp in priapulid, phoronid, brachiopod; Arg278>Lys in brachiopod; and Ser289>Ala in priapulid and phoronid), and only one in nemertean RAR (Lys236>Arg), whereas in the annelid RAR they are all conserved (**Figure 6.2**). Furthermore, the Phe230 residue which was demonstrated to play a crucial role for RA binding, enabling transactivation properties (Renaud et al. 1995), is replaced by a Val residue among lophotrochozoans and priapulid. The mutation to a Phe in *Platynereis* results on a decreased ability of the annelid RAR to activate transcription upon binding to ATRA (Handberg-Thorsager et al. 2018), suggesting a similar consequence for priapulid.

6.3.2. Functional Characterization of the Priapulid *RAR* Ortholog

To unfold the binding properties of PcauRAR and compare with both mollusc and annelid RARs, we next investigated the binding profile of the priapulid ortholog to transactivate target gene transcription, performing transactivation assays with GAL4-LBD chimeric receptors. Thus, we tested the ability of PcauRAR-LBD-GAL4 to bind retinoids (ATRA, 9cisRA and 13cisRA) and to activate a luciferase reporter gene (**Figure 6.3**). Our results show that PcauRAR is able to significantly ($P<0.05$) activate transcription upon binding to retinoids at concentrations of 1 and 10 μM (**Figure 6.3B**), but at a lesser degree than HsaRARγ (**Figure 6.3A**). The results obtained with transactivation assays were next confirmed by a competitive ligand binding assay. The ability of PcauRAR to bind to ATRA and 9cisRA was clearly demonstrated (**Figure 6.4**).

In vertebrates, RAR dimerizes with RXR (Chambon 2005; Germain et al. 2003; Chatagnon et al. 2015). Thus, we next assayed the capacity of RAR to transactivate luciferase transcription as a heterodimer with RXR, using a two-hybrid protein-protein interaction strategy (pBind/pACT system). The interaction between the chimeric proteins

(RAR-LBD-GAL4 and RXR-LBD-VP16) was first verified (Supplementary Material **Figure S6.1**) and then, the activation of the heterodimer was tested with ATRA, 9cisRA and 13cisRA at 10 μM (**Figure 6.3C**). As predicted, the RAR/RXR heterodimer activates luciferase transcription upon binding to the tested retinoids (P<0.05) in both human and priapulid (**Figure 6.3C**).

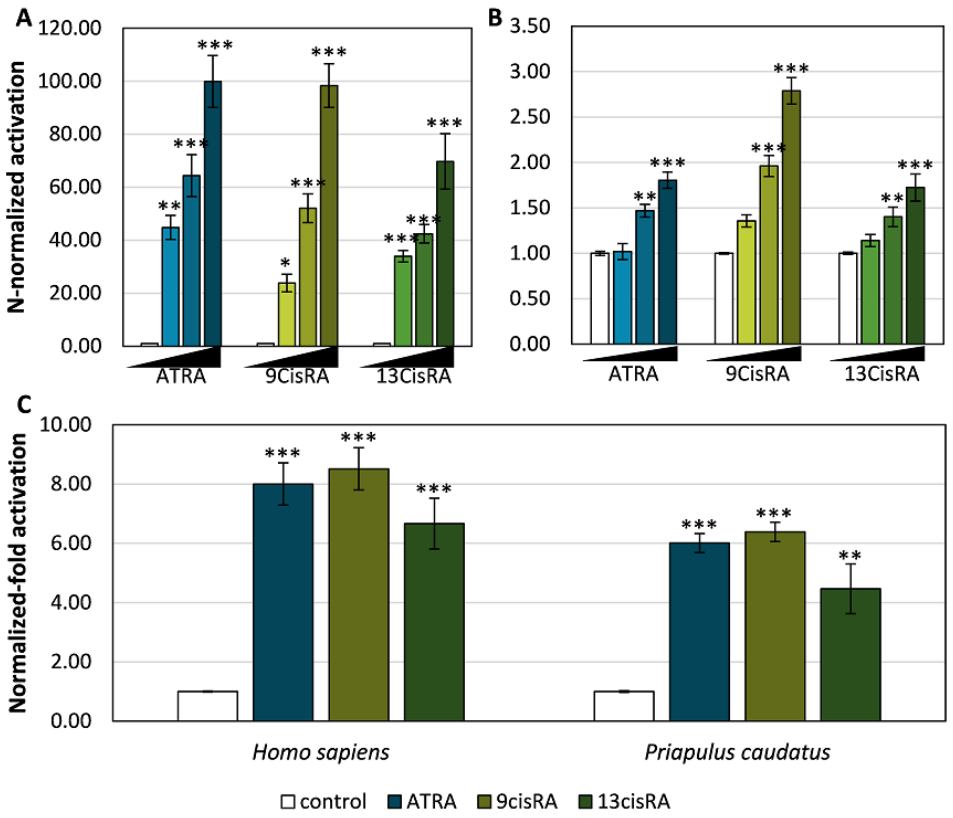


Figure 6.3. Luciferase transcription transactivation mediated by chimeric receptors in the presence of ATRA, 9cisRA or 13cisRA at a final concentration of 0.1, 1 and 10 μM). **(A)** Human RAR-LBD-GAL4; **(B)** Priapulid RAR-LBD-GAL4; **(C)** Human and priapulid RAR/RXR heterodimers. Data represent means ± SEM from three separate experiments (n = 3). The results were normalized to the control condition (DMSO without ligand). Significant differences between the tested concentrations and the solvent control were inferred using one-way ANOVA. Asterisks denote significant differences (* P<0.05, ** P<0.01, *** P<0.001) between the tested compound and the control.

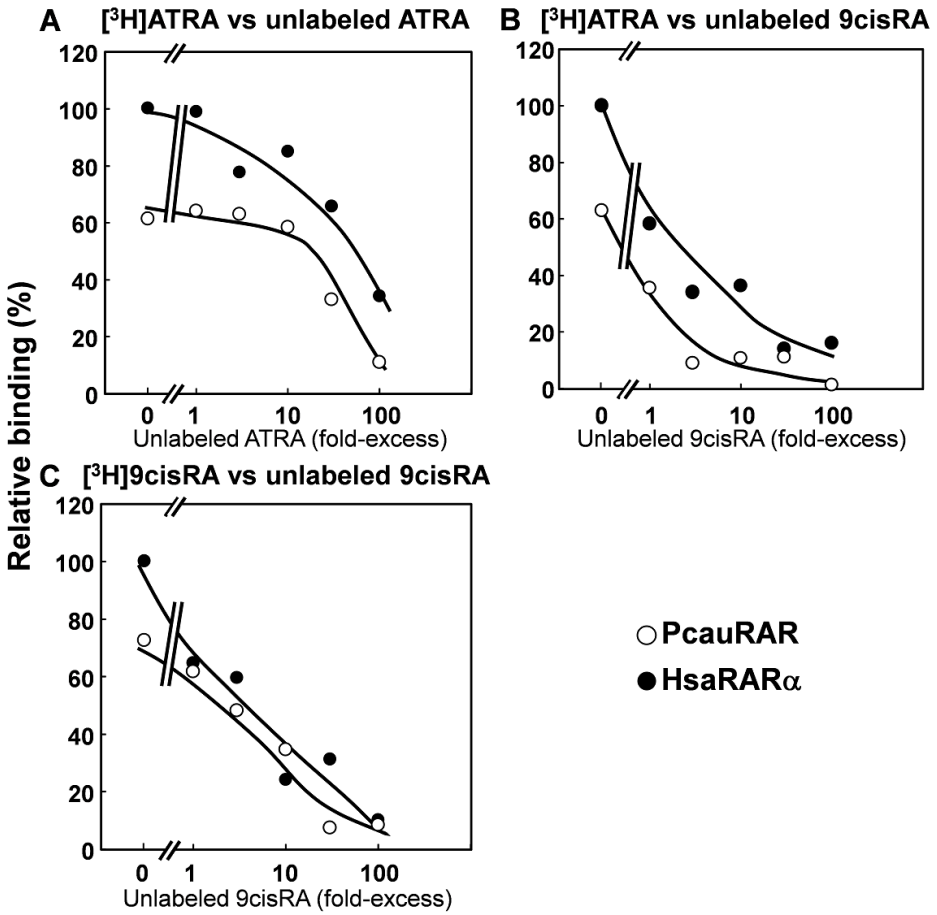


Figure 6.4. Competition by ATRA and 9cisRA with [^{3}H] ATRA and [^{3}H] 9cisRA for binding to the LBD of PcauRAR and HsaRARα. The LBD of PcauRAR (○) and HsaRARα protein (●) was incubated with increasing concentrations of unlabelled ATRA **(A)** or 9cisRA **(B, C)** as competitors in the presence of [^{3}H]ATRA **(A, B)** or [^{3}H]9cisRA **(C)** as ligand. Specific binding of the radio ligands was defined as total binding minus that occurring in the presence of 1000-fold molar excess of unlabelled ATRA **(A, B)** or 9cisRA **(C)**. Results were expressed as percentage of specific binding of the radio ligands. The binding of each radio ligand to HsaRARα in the absence of unlabelled competitors was set at 100%. Each experiment was performed at least twice, and representative curves are shown.

6.3.3. Pesticides Do Not Activate Transcription *Via* the Priapulid *RAR*

NRs are classical targets of EDCs (Bertrand et al. 2004; Darbre 2015; Laudet & Gronemeyer 2002). RARs in particular have been shown to bind and activate transcription in the presence of specific toxicants (Kamata et al. 2008; Lemaire et al. 2005). To address whether two organochlorine pesticides (endrin and dieldrin) known as endocrine disruptive chemicals (EDCs) acting *via* human RARs are also binding to PcauRAR, we performed transactivation assays. Importantly, these pesticides are persistent in fishes and sediments from the Baltic Sea, the geographic range of *P.*

caudatus (Strandberg, Strandberg, et al. 1998; Strandberg, Bandh, et al. 1998; Falandysz & Strandberg 2004; Schubert et al. 2016). As previously shown with mollusc RARs (André et al. 2019), these EDCs were unable to promote luciferase transcription through PcauRAR activation (P>0.05) (**Figure 6.5**).

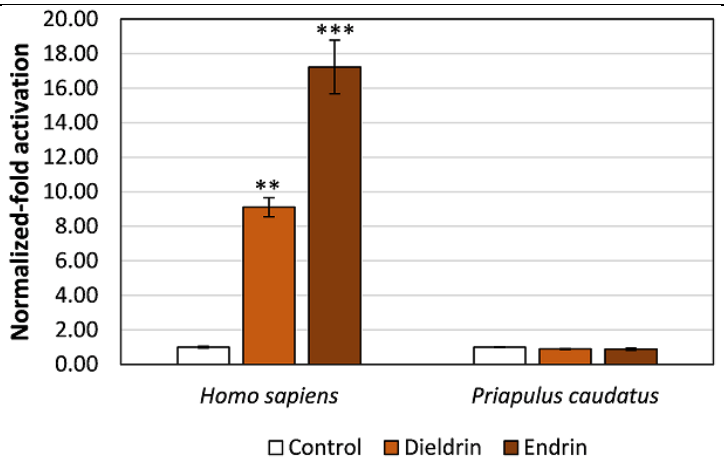


Figure 6.5. Luciferase transcription transactivation mediated by human and priapulid chimeric receptors (RAR-LBD-GAL4) in the presence of 10 µM endrin or dieldrin (organochlorine pesticides). Data represent means ± SEM from three separate experiments (n = 3). The results were normalized to the control condition (DMSO without ligand). Significant differences between the tested concentrations and the solvent control were inferred using one-way ANOVA. Asterisks denote significant differences (** P<0.01, *** P<0.001) between the tested compound and the control.

6.4. Discussion

The emergence of non-chordate sequenced genomes has significantly modified the evolutionary consensus of RA signalling as a chordate-specific feature. In effect, *RAR* and other RA signalling components were described in non-chordate metazoans, such as ambulacrarians (echinoderms and hemichordates) (Cañestro et al. 2006) and lophotrochozoans (molluscs and annelids) (Campo-Paysaa et al. 2008; Gutierrez-Mazariegos et al. 2014; Urushitani et al. 2013; Handberg-Thorsager et al. 2018; André et al. 2019). Strikingly, we establish that a retinoid-activated RAR was retained in the Ecdysozoa *P. caudatus*. Our findings strongly support earlier studies that *RAR* originated in the Bilateria ancestor (Gutierrez-Mazariegos et al. 2014; André et al. 2019; Handberg-Thorsager et al. 2018), and substantiate the likely loss of this transcription factor in most lineages leading to extant Ecdysozoa species examined so far (**Figure 6.6**). Together, these results emphasize the importance of Priapulida to decipher Ecdysozoa evolution, in particular that of NR biology (Telford et al. 2008; Sansom 2016).

By inspecting a RAR protein sequence alignment, we show that the penis worm ortholog exhibits the characteristic modular structure of NRs and displays a higher sequence homology with annelid RARs than with mollusc and vertebrate RARs. The retinoid binding profile of PcauRAR was corroborated with both transactivation assays and a competitive ligand binding assay that clearly established the ability of PcauRAR to

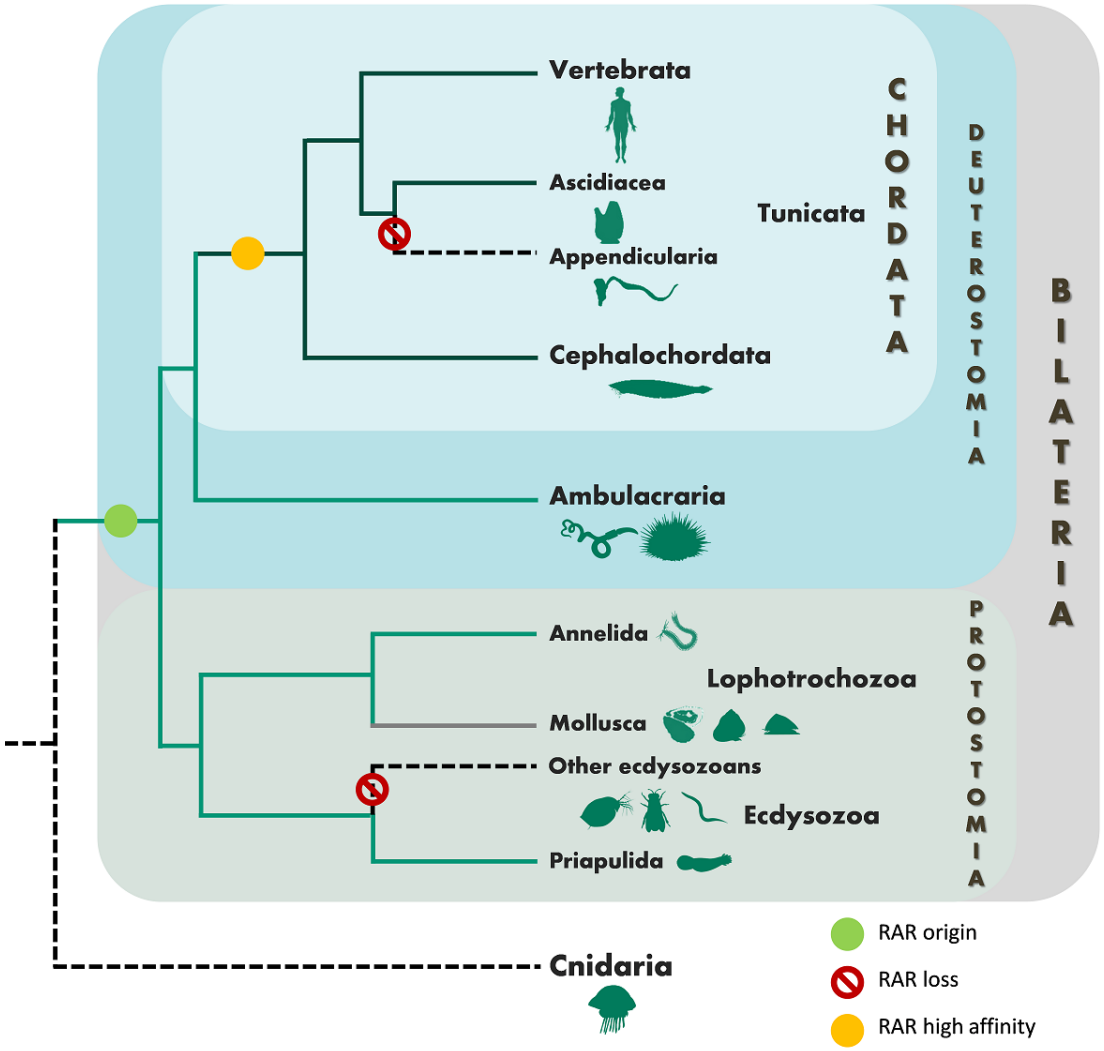


Figure 6.6. The evolution of RAR in Metazoa lineages. Black dashed lines represent no RAR known, light green full lines represent RA low-affinity sensors, dark green full lines represent RA high-affinity receptors, grey full line represent RA unresponsive sensors.

bind to ATRA and 9cisRA, as it had been previously demonstrated with the annelid RAR (Handberg-Thorsager et al. 2018), but not in molluscs (Urushitani et al. 2013; Gutierrez-Mazariegos et al. 2014; André et al. 2019). Yet, our findings consistently show that PcauRAR exhibits a low affinity for the tested ligands (retinoids) - in the micromolar range, similar to previous findings for the *Platynereis* RAR ortholog (Handberg-

Thorsager et al. 2018). This is in stark contrast with the affinity shown by chordate orthologs (nanomolar scale). In effect, the operating mode of PcauRAR in the presence of retinoids (ATRA, 9cisRA and 13cisRA) significantly induced luciferase transcription *via* PcauRAR activation, but at lower levels than HsaRARγ, as suggested by the LBP composition. Previous studies with crystallographic analysis of human RARγ (Renaud et al. 1995) and *Platynereis* RAR (Handberg-Thorsager et al. 2018) in complex with ATRA revealed a strong divergence in the structural interaction on how ATRA binds the RAR LBP residues in these species. In human RARs, 25 amino acid residues are crucial for the interaction with ATRA, with 4 of these residues forming direct or indirect hydrogen bonds with the carboxyl group of retinoids (Renaud et al. 1995; Escrivà et al. 2006). Despite the conservation of these 4 residues, the interaction of retinoids with annelid RAR-LBP is dominated by loose van der Waals forces and no hydrogen bond with retinoid carboxyl group have been described (Handberg-Thorsager et al. 2018). Thus, given the similarities of annelid and priapulid RAR sequences and ligand affinities, we anticipate an annelid-RAR-like structural interaction between ATRA and the priapulid RAR-LBP. A similar outcome is also expected for the RAR orthologs from nemertean, brachiopod and phoronid given the sequence similarities. Moreover, while in *Platynereis* ATRA displays a higher capacity to activate transcription *via* RAR, we find a similar pattern but for 9cisRA. Interestingly, 9cisRA was not detected in *Platynereis* tissues (Handberg-Thorsager et al. 2018) and, presently, RA levels are unknown in priapulids. Furthermore, we did not explore the possibility of other endogenous and uncharacterized ligands to bind and activate transcription *via* RAR in priapulids, although this possibility should deserve future investigation. Overall, it remains a tantalizing question of the exact *in vivo* functions of PcauRAR and whether these are conserved between annelids and priapulids (and other protostomes). Additionally, the finding of RAR in Priapulida raises the interesting possibility that other RA signalling and metabolic components might be present in other protostome phyla such as Loricifera and Bryozoa. Future studies should be undertaken to firmly explore these hypotheses.

Finally, we examined whether PcauRAR can be exploited by EDCs by testing two organochloride pesticides, which have low water solubility, but are extremely persistent and particularly stable in soil (Garnaga-Budrė 2012). Dieldrin was found in zooplankton and fishes from the Baltic Sea at concentrations between 15 and 170 ng/g lipid (Strandberg, Bandh, et al. 1998; Strandberg, Strandberg, et al. 1998; Schubert et al. 2016);, and <0.2-9.9 g and <0.15-0.8 g of dieldrin and endrin, respectively, were found per g of sludge and sediments (Falandysz & Strandberg 2004). Moreover, these compounds are known to disrupt the endocrine system in humans through the

modulation of RA signalling pathways (Kamata et al. 2008; Lemaire et al. 2005). In agreement, with the study conducted in molluscs (André et al. 2019), we show that the tested pesticides were not able to activate PcauRAR and consequently induce gene transcription. To understand the mechanisms of action of EDCs in invertebrate lineages is problematic given the paucity of appropriate comparative approaches. For instance, of the various Ecdysozoa groups, only three (Insecta, Crustacea and Nematoda) have been thoroughly examined from an endocrinology standpoint (Katsiadaki 2019). Moreover, several aquatic pollutants have been reported to retard growth and moulting, and influence mortality in crustaceans (Hosamani et al. 2017; Lagadic et al. 2018; Vogt et al. 2018) and to affect growth, reproduction, life span and gene expression in nematodes (Höss & Weltje 2007; Chen et al. 2019). Importantly, one of the clearest examples of endocrine disruption - imposex development in gastropods upon exposure to organotins, was shown to result from a specific interaction with the highly conserved NR RXR (Castro et al. 2007; Lima et al. 2011; Stange et al. 2012). Moreover, the inclusion of comparative approaches and evolutionary thinking has highlighted the conserved and divergent biological responses to xenobiotics mediated by NRs (e.g. PPAR (Capitão et al. 2018), PXR (Fonseca et al. 2019), ER (Paris et al. 2008)). Thus, defining the gene complement of NRs, their set of "natural" ligands and *in vivo* functions across the diversity of protostome phyla is fundamental to comprehend the impacts of EDCs in the Anthropocene epoch (Santos et al. 2018).

6.5. Conclusions

We provide here the first characterization of an Ecdysozoa *RAR*. Our findings, contribute to further clarify the early evolution of the RA gene module in Metazoa, supporting the hypothesis that *RAR* emerged as RA low-affinity sensor in the Bilateria, with the high-affinity RA binding profile acquired in chordates.

Supplementary Materials: The following are available online at www.mdpi.com/xxx/s1, **Video S6.1**: Live specimens of *P. caudatus*; **Figure S6.1**: Analysis of the interaction between priapulid and human RAR-LBD-GAL4 with RXR-LBD-VP16 partner through a mammalian two-hybrid assay in COS-1 cells with no ligands; **Table S6.1:** List of primers used to isolate *P. caudatus RAR* and *RXR* genes; **Table S6.2:** List of sequences used for phylogenetic reconstruction of *TR*, *RAR*, *PPAR* and *RXR* genes and corresponding accession numbers; **Table S6.3:** List of primers used to amplify hinge region to LBD of RAR and RXR to be cloned into pBIND or pACT expression vectors.

Funding: This research was funded by COMPETE 2020, Portugal 2020 and the European Union through the ERDF, grant number 031342, and a Grant-in-Aid for Scientific Research (C) from the Ministry of Education, Culture, Sports, Science and Technology (Japan), grant number 18K06737 and by FCT through national funds. A PhD grant awarded to Elza Fonseca (SFRH/BD/100262/2014) was funded by the FCT.

Acknowledgments: We acknowledge Richard K. Wilson, Director, and the McDonnell Genome Institute, Washington University School of Medicine for the *P. caudatus* genome assembly 5.0.

Conflicts of Interest: The authors declare no conflict of interest.

6.6. References

Albalat R. 2009. The retinoic acid machinery in invertebrates: ancestral elements and vertebrate innovations. Mol Cell Endocrinol. 313:23–35. doi: 10.1016/j.mce.2009.08.029.

Albalat R, Cañestro C. 2009. Identification of Aldh1a, Cyp26 and RAR orthologs in protostomes pushes back the retinoic acid genetic machinery in evolutionary time to the bilaterian ancestor. Chem Biol Interact. 178:188–196. doi: 10.1016/j.cbi.2008.09.017.

André A et al. 2019. The retinoic acid receptor (RAR) in molluscs: Function, evolution and endocrine disruption insights. Aquat Toxicol. 208:80–89. doi: 10.1016/j.aquatox.2019.01.002.

Aranda A, Pascual A. 2001. Nuclear hormone receptors and gene expression. Physiol Rev. 81:1269–1304. doi: 10.1152/physrev.2001.81.3.1269.

Balaguer P, Delfosse V, Grimaldi M, Bourguet W. 2017. Structural and functional evidences for the interactions between nuclear hormone receptors and endocrine disruptors at low doses. C R Biol. 340:414–420. doi: 10.1016/j.crvi.2017.08.002.

Bertrand, S et al. 2007. Unexpected novel relational links uncovered by extensive developmental profiling of nuclear receptor expression. PLoS Genet. 3:e188. doi: 10.1371/journal.pgen.0030188.

Bertrand S et al. 2004. Evolutionary genomics of nuclear receptors: from twenty-five ancestral genes to derived endocrine systems. Mol Biol Evol. 21:1923–1937. doi: 10.1093/molbev/msh200.

Campo-Paysaa F, Marlétaz F, Laudet V, Schubert M. 2008. Retinoic acid signaling in development: tissue-specific functions and evolutionary origins. Genesis. 46:640–656. doi: 10.1002/dvg.20444.

Cañestro C, Postlethwait J, Gonzalez-Duarte R, Albalat R. 2006. Is retinoic acid genetic machinery a chordate innovation? Evol Dev. 8:394–406. doi: 10.1111/j.1525-142X.2006.00113.x.

Capitão A et al. 2018. Evolutionary Exploitation of Vertebrate Peroxisome Proliferator-Activated Receptor γ by Organotins. Environ. Sci. Technol. 52. doi: 10.1021/acs.est.8b04399.

Capitão A, Lyssimachou A, Castro L, Santos M. 2017. Obesogens in the aquatic environment: an evolutionary and toxicological perspective. Env. Int. 106:153−169. doi: 10.1016/j.envint.2017.06.003.

Castro L et al. 2007. Imposex induction is mediated through the Retinoid X receptor signalling pathway in the neogastropod *Nucella lapillus*. Aquat Toxicol. 85:57–66. doi: 10.1016/j.aquatox.2007.07.016.

Castro L, Santos M. 2014. To bind or not to bind: the taxonomic scope of nuclear receptor mediated endocrine disruption in invertebrate phyla. Env. Sci Technol. 48:5361−5363. doi: 10.1021/es501697b.

Chambon P. 2005. The nuclear receptor superfamily: a personal retrospect on the first two decades. Mol Endocrinol. 19:1418–1428. doi: 10.1210/me.2005-0125.

Chatagnon A et al. 2015. RAR/RXR binding dynamics distinguish pluripotency from differentiation associated cis-regulatory elements. Nucleic Acids Res. 43:4833–4854. doi: 10.1093/nar/gkv370.

Chen H et al. 2019. A review of toxicity induced by persistent organic pollutants (POPs) and endocrine-disrupting chemicals (EDCs) in the nematode *Caenorhabditis elegans*. J Env. Manag. 237:519–525. doi: 10.1016/j.jenvman.2019.02.102.

Cunningham T, Duester G. 2015. Mechanisms of retinoic acid signalling and its roles in organ and limb development. Nat Rev Mol Cell Biol. 16:110–123. doi: 10.1038/nrm3932.

Cvekl A, Wang W. 2009. Retinoic acid signaling in mammalian eye development. Exp Eye Res. 89:280–291. doi: 10.1016/j.exer.2009.04.012.

Darbre P. 2015. Endocrine disruption and human health. Overview of EDCs and human health which sets the bigger picture. :Academic: New York.

Duffy J. 2002. GAL4 system in Drosophila: a fly geneticist's Swiss army knife. Genesis. 34:1–15. doi: 10.1002/gene.10150.

Escrivà H et al. 2006. Neofunctionalization in vertebrates: The example of retinoic acid receptors. PloS Genet. 2:e102. doi: 10.1371/journal.pgen.0020102.

Escrivà H, Holland N, Gronemeyer H, Laudet V, Holland L. 2002. The retinoic acid signaling pathway regulates anterior/posterior patterning in the nerve cord and

pharynx of amphioxus, a chordate lacking neural crest. Development. 129:2905–2916.

Falandysz J, Strandberg B. 2004. Persistent Organochlorine Compounds in Sludge and Sediments from the Gdańsk Region, Baltic Sea. Pol J Env. Stud. 13:133–138.

Fonseca E et al. 2019. Evolutionary Plasticity in Detoxification Gene Modules: The Preservation and Loss of the Pregnane X Receptor in Chondrichthyes Lineages. Int J Mol Sci. 10:pii: E2331. doi: 10.3390/ijms20092331.

Garnaga-Budrė G. 2012. Integrated assessment of pollution in the Baltic Sea. Ekologija. 58:331–355. doi: 10.6001/ekologija.v58i3.2531.

Germain P, Altucci L, Bourguet W, Rochette-Egly C, Gronemeyer H. 2003. Nuclear receptor superfamily: principles of signaling. Pure Appl Chem. 75:1619–1664. doi: 10.1351/pac200375111619.

Germain P, Staels B, Dacquet C, Spedding M, Laudet V. 2006. Overview of nomenclature of nuclear receptors. Pharmacol Rev. 58:685–704. doi: 10.1124/pr.58.4.2.

Gesto M et al. 2016. Retinoid level dynamics during gonad recycling in the limpet Patella vulgata. Gen Comp Endocrinol. 225:142–148. doi: 10.1016/j.ygcen.2015.

Guindon S et al. 2010. New Algorithms and Methods to Estimate Maximum-Likelihood Phylogenies: Assessing the Performance of PhyML 3.0. Syst Biol. 59:307–321. doi: 10.1093/sysbio/syq010.

Gutierrez-Mazariegos J et al. 2014. A mollusk retinoic acid receptor (RAR) ortholog sheds light on the evolution of ligand binding. Endocrinology. 155:4275–4286. doi: 10.1210/en.2014-1181.

Hagmann M, Georgiev O, Schaffner W. 1997. The VP16 paradox: herpes simplex virus VP16 contains a long-range activation domain but within the natural multiprotein complex activates only from promoter-proximal positions. J Virol. 71:5952–5962.

Handberg-Thorsager M et al. 2018. The ancestral retinoic acid receptor was a low-affinity sensor triggering neuronal differentiation. Sci Adv. 4:eaao1261. doi: 10.1126/sciadv.aao1261.

Hiromori Y et al. 2016. Ligand activity of group 15 compounds possessing triphenyl substituent for the RXR and PPARγ nuclear receptors. Biol Pharm Bull. 39:1596–1603. doi: 10.1248/bpb.b16-00186.

Hiromori Y, Nishikawa J, Yoshida I, Nagase H, Nakanishi T. 2009. Structure-dependent activation of peroxisome proliferator-activated receptor (PPAR) γ by organotin compounds. Chem Biol Interact. 180:238–244. doi: 10.1016/j.cbi.2009.03.006.

Hisata K, Fujiwara S, Tsuchida Y, Ohashi M, Kawamura K. 1998. Expression and function of a retinoic acid receptor in budding ascidians. Dev Genes Evol. 208:537–546. doi: 10.1007/s004270050213.

Hosamani N, Reddy S, Reddy R. 2017. Crustacean Molting: Regulation and Effects of Environmental Toxicants. J Mar. Sci Res Dev. 7:236. doi: 10.4172/2155-9910.1000236.

Höss S, Weltje L. 2007. Endocrine disruption in nematodes: effects and mechanisms. Ecotoxicology. 16:15–28. doi: 10.1007/s10646-006-0108-y.

Howard-Ashby M et al. 2006. Gene families encoding transcription factors expressed in early development of Strongylocentrotus purpuratus. Dev Biol. 300:90–107. doi: 10.1016/j.ydbio.2006.08.033.

Janesick A, Wu S, Blumberg B. 2015. Retinoic acid signaling and neuronal differentiation. Cell Mol Life Sci. 72:1559–1576. doi: 10.1007/s00018-014-1815-9.

Kamata R, Shiraishi F, Nishikawa J, Yonemoto J, Shiraishi H. 2008. Screening and detection of the in vitro agonistic activity of xenobiotics on the retinoic acid receptor. Toxicol Vitr. 22:1050–1061. doi: 10.1016/j.tiv.2008.01.002.

Katoh K, Toh H. 2010. Parallelization of the MAFFT multiple sequence alignment program. Bioinformatics. 26:1899–1900. doi: 10.1093/bioinformatics/bty121.

Katsiadaki I. 2019. Are marine invertebrates really at risk from endocrine-disrupting chemicals? Curr Opin Env. Sci Heal. 11:37–42. doi: 10.1016/j.coesh.2019.06.005.

Keay J, Thornton J. 2009. Hormone-activated estrogen receptors in annelid invertebrates: implications for evolution and endocrine disruption. Endocrinology. 150:1731–1738. doi: 10.1210/en.2008-1338.

Lagadic L et al. 2018. Tributyltin: Advancing the Science on Assessing Endocrine Disruption with an Unconventional Endocrine-Disrupting Compound. Rev Env. Contam Toxicol. 245:65–127. doi: 10.1007/398_2017_8.

Laudet V, Gronemeyer H. 2002. The nuclear receptors factsbook. Academic Press: London:

Lemaire G, Balaguer P, Michel S, Rahmani R. 2005. Activation of retinoic acid receptor-dependent transcription by organochlorine pesticides. Toxico Appl Pharmacol. 202:38–49. doi: 10.1016/j.taap.2004.06.004.

Lima D et al. 2011. Tributyltin-induced imposex in marine gastropods involves tissue-specific modulation of the retinoid X receptor. Aquat Toxicol. 101:221–227. doi: 10.1016/j.aquatox.2010.09.022.

Maden M, Hind M. 2003. Retinoic acid, a regeneration-inducing molecule. Dev Dyn. 226:237–244. doi: 10.1002/dvdy.10222.

Marlétaz F, Holland L, Laudet V, Schubert M. 2006. Retinoic acid signaling and the evolution of chordates. Int J Biol Sci. 2:38–47. doi: 10.7150/ijbs.2.38.

Nakanishi T et al. 2005. Trialkyltin compounds bind retinoid X receptor to alter human placental endocrine functions. Mol Endocrinol. 19:2502–2516. doi: 10.1210/me.2004-0397.

Niederreither K, Dollé P. 2008. Retinoic acid in development: towards an integrated view. Nat Rev Genet. 9:541–553. doi: 10.1038/nrg2340.

Ollikainen N, Chansawangbhuwana C, Baker M. 2006. Evolution of the thyroid hormone, retinoic acid, ecdysone and liver X receptors. Integr Comp Biol. 46:815–826. doi: 10.1093/icb/icl035.

Paris M et al. 2008. An amphioxus orthologue of the estrogen receptor that does not bind estradiol: insights into estrogen receptor evolution. BMC Evol Biol. 8:219. doi: 10.1186/1471-2148-8-219.

Renaud J et al. 1995. Crystal structure of the RAR-γ ligand-binding domain bound to all-trans retinoic acid. Nature. 378:681–689. doi: 10.1038/378681a0.

Rochette-Egly C, Germain P. 2009. Dynamic and combinatorial control of gene expression by nuclear retinoic acid receptors (RARs). Nucl Recept Signal. 7:e005. doi: 10.1621/nrs.07005.

Ross S, McCaffery P, Drager U, De Luca L. 2000. Retinoids in embryonal development. Physiol Rev. 80:1021–1054. doi: 10.1152/physrev.2000.80.3.1021.

Samarut E, Rochette-Egly C. 2012. Nuclear retinoic acid receptors: conductors of the retinoic acid symphony during development. Mol Cell Endocrinol. 348:348–360. doi: 10.1016/j.mce.2011.03.025.

Sansom R. 2016. Preservation and phylogeny of Cambrian ecdysozoans tested by experimental decay of Priapulus. Sci Rep. 6:32817. doi: 10.1038/srep32817.

Santos M, Ruivo R, Capitão A, Fonseca E, Castro L. 2018. Identifying the gaps: Resources and perspectives on the use of nuclear receptor based-assays to improve hazard assessment of emerging contaminants. J Hazard Mater. 358:508–511. doi: 10.1016/j.jhazmat.2018.04.076.

Schubert M, Holland N, Escrivà H, Holland L, Laudet V. 2004. Retinoic acid influences anteroposterior positioning of epidermal sensory neurons and their gene expression in a developing chordate (amphioxus). Proc Natl Acad Sci USA. 101:10320–10325. doi: 10.1073/pnas.0403216101.

Schubert S et al. 2016. Persistent organic pollutants in Baltic herring (*Clupea harengus*)-an aspect of gender. Env. Monit Assess. 188:368. doi: 10.1007/s10661-016-5363-7.

Scott A. 2018. Is there any value in measuring vertebrate steroids in invertebrates? Gen Comp Endocrinol. 265:77–82. doi: 10.1016/j.ygcen.2018.04.005.

Simões-Costa M, Azambuja A, Xavier-Neto J. 2008. The search for non-chordate retinoic acid signaling: Lessons from chordates. J Exp Zool B Mol Dev Evol. 310:54–72. doi: 10.1002/jez.b.21139.

Stange D, Sieratowicz A, Oehlmann J. 2012. Imposex development in *Nucella lapillus* – evidence for the involvement of retinoid X receptor and androgen signalling pathways in vivo. Aquat Toxicol. 106–107:20–24. doi: 10.1016/j.aquatox.2011.10.010.

Strandberg B, Bandh C, et al. 1998. Concentrations, biomagnification and spatial variation of organochlorine compounds in a pelagic food web in the northern part of the Baltic Sea. Sci Total Env. 217:143–154. doi: 10.1016/s0048-9697(98)00173-9.

Strandberg B, Strandberg L, et al. 1998. Concentrations and spatial variations of cyclodienes and other organochlorines in herring and perch from the Baltic Sea. Sci Total Env. 215:69–83. doi: 10.1016/s0048-9697(98)00114-4.

Telford M, Bourlat S, Economou A, Papillon D, Rota-Stabelli O. 2008. The evolution of the Ecdysozoa. Philos Trans R Soc L. B Biol Sci. 363:1529–1537. doi: 10.1098/rstb.2007.2243.

Urushitani H et al. 2013. Cloning and characterization of the retinoic acid receptor-like protein in the rock shell, *Thais clavigera*. Aquat Toxicol. 142–143:403–413. doi: 10.1016/j.aquatox.2013.09.008.

Vogeler S, Galloway T, Isupov M, Bean T. 2017. Cloning retinoid and peroxisome proliferator-activated nuclear receptors of the Pacific oyster and in silico binding to environmental chemicals. PLoS One. 12:e0176024. doi: 10.1371/journal.pone.0176024.

Vogt É, Model J, Vinagre A. 2018. Effects of Organotins on Crustaceans: Update and Perspectives. Front Endocrinol. 9:65. doi: 10.3389/fendo.2018.00065.

Webster B et al. 2006. Mitogenomics and phylogenomics reveal priapulid worms as extant models of the ancestral Ecdysozoan. Evol Dev. 8:502–510. doi: 10.1111/j.1525-142X.2006.00123.x.

Wills M, Gerber S, Ruta M, Hughes M. 2012. The disparity of priapulid, archaeopriapulid and palaeoscolecid worms in the light of new data. J Evol Biol. 25:2056–2076. doi: 10.1111/j.1420-9101.2012.02586.x.

6.7. Supplementary Material

Table S6.1. List of primers used to isolate *Priapulus caudatus RAR* and *RXR* genes.

Nuclear receptor	Used for	Oligonuleotide sequence 5'→ 3'	Tm (°C)	Extension time (s)
RAR	full ORF PCR	F: ATGTCTATGCAACAGCGAATATTCGCT R: TCACGGTATGCAGACGTTCTCGT	65	22
RXR	partial ORF PCR	F: CTCGCAAACTCGAAGCACATC R: CTACTCGCTGTTGCTCTCGT	60	15
	RACE PCR	R: GCCCATCGCGAGACACTTCTGGTAT	69	8
	nested RACE PCR	R: GAAGAAGCCCTTGCATCCCTCACAG	69	8
	full ORF PCR	F: ATGAGAATTTACGAGGCGTG R: CTACTCGCTGTTGCTCTC	61	20

Table S6.2. List of sequences used for phylogenetic reconstruction of RAR and RXR and corresponding accession numbers.

Species	Nuclear receptor	Accession number
Homo sapiens	RARα	NP_000955.1
	RXRα	NP_002948.1
	RARβ	NP_000956.2
	RXRβ	NP_068811.1
	RARγ	NP_000957.1
	RXRγ	NP_008848.1
Danio rerio	RARα	a NP_571481.2; b NP_571474.1
	RXRα	a NP_001155023.1; b NP_571228.1
	RXRβ	a NP_571350.1; b NP_57313.1
	RARγ	a NP_571414.1; b NP_001076779.1
	RXRγ	a NP_571292.3; b XP_005160723.1
Banchiostoma floridae	RAR	XP_002598475.1
	RXR	AAM46151.1
Branchiostoma lanceolatum	RAR	ANP24205.1
	RXR	ANP24206.1
Strongylocentrotus purpuratus	RAR	XP_011680540.1
	RXR	XP_011661795.1
Saccoglossus kowalevskii	RAR	XP_006825846.1
	RXR	ADB22634.1
Lottia gigantea	RAR	ESO98546.1
	RXR	ESO92876.1
Nucella lapillus	RAR	AIB06349.1
	RXR	ABS70715.1
Platynereis dumerilii	RAR	AVR59236.1
	RXR	AVR59237.1
Capitella teleta	RAR	ELU07684.1
	RXR	ELT93409.1
Priapulus caudatus	RAR	QFQ33539.1[a]
	RXR	QFQ33540.1[a]
Aurelia aurita	RXR	AGT42223.1

[a]These sequences were isolated in this stuby by PCR

Table S6.3. List of primers used to amplify hinge region to LBD of RAR and RXR to be cloned into pBIND or pACT expression vectors.

Species	Nuclear receptor	Oligonuleotide sequence 5'→ 3'	Tm (°C)	Restriction enzymes
Priapulus caudatus	RAR	F: aaaGGATCCGCAAGAAGCAGCAGCAGGA	58	BamHI
		R: aaaTCTAGATCACGGTATGCAGACGTTCT		XbaI
	RXR	F: aaaTCTAGACAGCGCGTGAAGGAGAAG	56	XbaI
		R: aaaGGTACCCTACTCGCTGTTGCTCTC		KpnI
Homo sapiens	RARg	F: catgGGATCCTAGAGGTGAAGGAAGAAGGG	58	BamHI
		R: gcatgTCTAGATCAGGCTGGGGACTTCAG		XbaI
	RXRa	F: aattTCTAGAGCCGTGCAGGAGGAGCGGCA	62	XbaI
		R: aattGGTACCAGTCATTTGGTGCGGCGCCT		KpnI

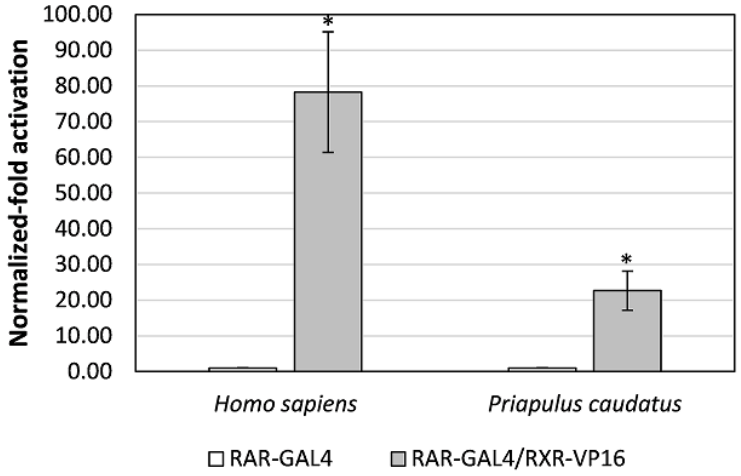


Figure S6.1. Analysis of the interaction between priapulid and human RAR-LBD-GAL4 with RXR-LBD-VP16 partner through a mammalian two-hybrid assay in COS-1 cells with no ligands. Data represent normalized means ± SEM to the control from three separate experiments (n = 3). Cells transfected with no RXR-LBD-VP16 partner were used as control. Significant differences (* $P<0.05$) were inferred using one-way ANOVA.

CHAPTER 7 - General Discussion and Conclusions

7. General Discussion and Conclusions

7.1. General Discussion

The evolutionary appearance of Metazoa on Earth was featured by multiple novelties at the genomic level, including the emergence of the NR superfamily (Paps & Holland 2018; Escriva et al. 2003). As the most abundant family of intracellular transcription factors (Mangelsdorf et al. 1995; Degnan et al. 2009), NRs have been identified and largely studied on vertebrate models (Chung & Cooney 2003; Zhao et al. 2015; Capitão et al. 2018) and in some invertebrate species (e.g. King-Jones & Thummel 2005; Reitzel & Tarrant 2009; Vogeler et al. 2014; Cruzeiro et al. 2016; Kim et al. 2017; A André et al. 2019). Importantly, NRs are key components of the endocrine system, allowing and transducing cell-to-cell communication in multicellular organisms (Hartenstein 2006; Bonett 2016). The endocrine NRs, such as, endocrine steroid hormone receptors (ERs, GR, MR, AR and PR), THR, RAR and VDR, have endocrine steroids (estrogen, glucocorticoids, mineralocorticoids, androgens, progesterone), thyroid hormone, vitamin A and vitamin D as natural ligands, respectively (Gronemeyer et al. 2004). These NRs participate in the modulation of important physiologic processes, such as reproduction, basal metabolism, embryonic development and dietary calcium uptake and metabolism, among others (Bookout et al. 2006).

Another significant feature of NR biology is related with their exploitation by xenobiotics. The widespread occurrence of environmental contaminants in world ecosystems is a matter of great concern, particularly in aquatic environments. In numerous cases, these compounds have been identified as EDCs, affecting human health and wildlife (Diamanti-Kandarakis et al. 2009; Bergman et al. 2013; Kabir et al. 2015; Noguera-Oviedo & Aga 2016). Moreover, several studies reported endocrine disruption mediated by the interaction of EDCs with NRs in aquatic organisms (le Maire et al. 2010; Castro & Santos 2014). Imposex in gastropods (Nishikawa et al. 2004; Castro et al. 2007; Lima et al. 2011; Stange et al. 2012; André et al. 2017), fish feminization by estrogen-like chemicals (Soares et al., 2009) and lipid homeostasis perturbation by obesogens (Grün & Blumberg 2006; Riu et al. 2011; Ouadah-Boussouf & Babin 2016; Capitão et al. 2017; Capitão et al. 2018; Barbosa et al. 2019) are some of the best documented examples. The studies conducted so far have been focused on animal models. Thus, the endocrine disruption of some marine invertebrates and basal vertebrates remains unresolved. The present work intend to explore new insights into the evolution and function of NRs in the endocrine system of metazoans, exploring the taxonomic significance of the chosen lineages, in the context of the Anthropocene epoch.

For this, the number of NRs in metazoan lineages was investigated, considering informative phylogenetic nodes at key evolutionary transitions (e.g. Chondrichthyes, Ecdysozoa vs Lophotrochozoa). Additionally, multiple NRs were functionally characterised, testing their exploitation by EDCs and binding affinities towards presumed natural ligands. To solve gene repertoires in target phyla, the gene orthology of all the retrieved sequences was firstly established through rigorous phylogenetic analyses and, whenever possible, combined with synteny analyses. Synteny was also used to discriminate between true gene loss and absence of sequencing data. The functional characterization of NRs was performed trough *in vitro* transactivation assays to explore the capacity of NRs to activate gene transcription upon binding to classical ligands and environmental contaminants.

To decipher the exact number of NRs in each metazoan lineage is a fundamental approach to tackle some the questions raised in this thesis. Yet, until recently the difficulty to assess genomic and transcriptomic information in a wide range of lineages has hampered a detailed analysis (Santos et al. 2018). The revolution of next generation technology has largely resolved this bottleneck, allowing multi-genome comparisons to be carried out. Moreover, the function and the interaction of natural ligands or EDCs with a specific NR should not be extrapolated between lineages. The genome occurrence of a NR gene ortholog in a given species should not *per se* be associated with a conserved function or role. The molluscan *RAR* gene ortholog is an insightful example. Unlike vertebrates and the annelid RARs, this receptor it is not modulated by RA neither by the EDCs tested (Lemaire et al. 2005; Campo-Paysaa et al. 2008; Gutierrez-Mazariegos et al. 2014; Handberg-Thorsager et al. 2018; André et al. 2019). Thus, the screening of new environmental contaminants (and natural ligands) as potential EDCs acting via NRs and the expansion of genome datasets with an evolutionary framework is crucial in near future to determine the impact of such compounds in the ecosystems (**Figure 7.1**).

The variation of NR gene among lineages (Bertrand et al. 2004; Lecroisey et al. 2012; Zhao et al. 2015) is most likely the result of genomic processes such as gene loss and duplication (tandem or whole-genome), denoting a dynamic evolutionary path (Bertrand et al. 2004, 2011; Bridgham et al. 2010). To have a broader view on the NRs collection in chordates, the NR gene repertoires from chicken (bird), sea squirt (tunicate) and amphioxus (cephalochordate) were revised. Additionally, a search was also conducted for NRs in the green anole (reptile) and the spotted gar (non-teleost ray-finned fish). This approach allowed the identification of paralogs or NR subfamilies not previously found in the genomes of some of these lineages and to resolve some orthology misassignments. Previous studies identified 33 NRs in amphioxus, three of

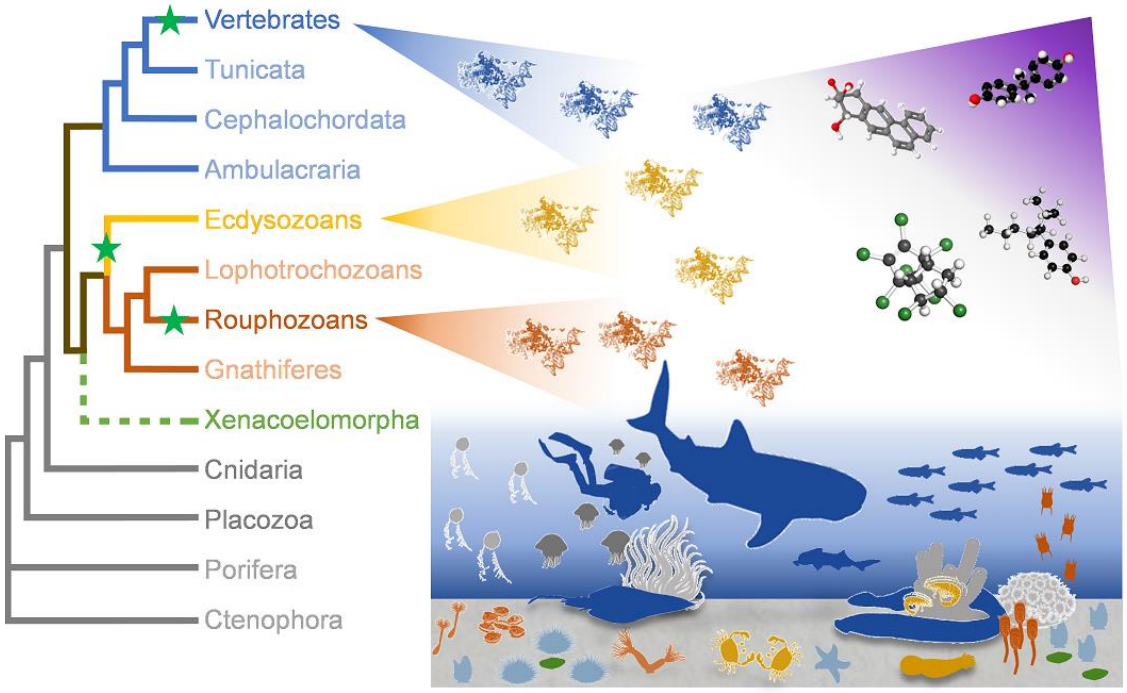


Figure 7.1. Evolution of NRs in the endocrine system of the metazoans and the impact in the ecosystem.

them with uncertain orthology (Schubert et al. 2008; Lecroisey et al. 2012). This analysis of lancelet's NRs established their orthology (see **Chapter 2** for details). Moreover, a second gene copy from the NR0B subfamily was identified, updating to 34 the number of NRs in cephalochordate genomes. The revision of the NRs from chicken added four new genes to the previous repertoire of 44 NRs. Despite the identification of paralogs in the green anole genome not described in turtle (Zhao et al. 2015), both repertoires are very similar (see **Chapter 2** for details). In fishes, we noticed that the collection of NRs has mostly been focused on teleosts (Maglich et al. 2003; Bertrand et al. 2004; Metpally et al. 2007; Zhao et al. 2015), a lineage of ray-finned fishes which underwent a third event of WGD (3R WGD) approximately at 450 MYA after the divergence of Holostei (Jaillon et al. 2004; Amores et al. 2011). Here, the NR repertoire of the holostean spotted gar was investigated, as a non-teleost ray-finned fish prior to the 3R duplication, in order to identify tetrapod ancestor orthologous NR genes which may not be found in teleost genomes due to the extensive rearrangements (Amores et al. 2011; Braasch et al. 2016). As expected, the composition was comparable to teleosts but the number of paralogs was much more similar to the remaining vertebrates (see **Chapter 2** for details). Despite the vast number of studies on the evolution of NRs in vertebrates, very few have been directed towards Chondrichthyes (e.g. Inoue et al. 2010) and a description of their NRs is still missing. Chondrichthyes are a class of cartilaginous fishes with pelvic claspers divided in two subclasses: the Holocephali which contains around 40 species of

chimaeras and the Elasmobranchii which comprises approximately 1100 species of sharks, rays and skates (Didier et al. 2012). They diverged more than 450 MYA, being the oldest living lineage divergent from the last common ancestor of jawed vertebrates (gnathostomes) (Benton et al. 2009). The position of cartilaginous fishes in the vertebrate phylogenetic tree, their lower genome evolutionary rate and the similarities shared with mammalian genomes make them a very attractive group to study the evolution of molecular mechanisms in vertebrates (Venkatesh et al. 2006, 2007, 2014; Hara et al. 2018). Thus, to address the early diversification of NR genes and the evolution of endocrine system in vertebrates, the NR gene repertoire of Chondrichthyes was extensively investigated, using publically available chondrichthyan genomes and transcriptomes (Venkatesh et al. 2007; Wang et al. 2012; Read et al. 2017; Hara et al. 2018) and additionally generating a novel draft genome assembly from a second Chimaeriformes species, the small-eyed rabbitfish. The later provided a valuable resource to infer about true holocephalan-specific gene loss. Generically, new members of the non-canonical NR0B subfamily were found and it was estimated a total of 52 NRs in Elasmobranchii and 54 NRs in Holocephali (see **Chapter 2** for details), numbers consistent with the 2R WGD that occurred approximately at 500 MYA in the invertebrate/vertebrate transition (e.g. Putnam et al. 2008), denoting a relatively stable repertoire compared to mammalians ones. These findings provide an overall view into the early diversification of NRs in gnathostomes. However, to assess the true lineage-specific loss of some NR genes, new genomes from other species are needed to overcome the missing data or annotation gaps in the sequenced genomes currently available.

The investigation into the NR gene collection in Chondrichthyes provided solid evidence for the presence of NR1H and NR1I subfamilies for which no functional studies were available (**Chapters 3** and **4**). The members of these subfamilies are activated by cholesterol derivatives and modulate the cholesterol homeostasis and energy metabolism (Krasowski et al. 2011). NRs of NR1H subfamily, specially LXRs, regulate the transcription of genes involved in the efflux, transport and excretion of cholesterol, modulating cholesterol and fat metabolism upon binding to oxysterols (Kalaany & Mangelsdorf 2006; Bookout et al. 2006). In mammals, two paralogs (*LXRα/NR1H3* and *LXRβ/NR1H2*) were previously described, with a sole paralog identified in the remaining non-mammalian vertebrate species (Reschly et al. 2008b; Krasowski et al. 2011; Zhao et al. 2015). Based on the sequence data, it was suggest that the duplication of a single *LXR* gene occurred concomitantly with the evolution of mammals (Krasowski et al. 2011). Yet, data from chondrichthyan genomes suggests an alternative scenario, placing the

LXR gene duplication event at the split of the gnathostomes with independent losses of either paralog in several lineages (**Figure 7.2**). To gain further insight into the evolution of LXR function, the binding profile of amphioxus, lamprey and chondrichthyan *LXR* orthologs were characterized and it was demonstrated that the ability to induce gene transcription upon binding to oxysterols emerged with the capacity of LXR to accommodate oxysteroid backbones, originally limited to a slighter set of oxysterols in the tunicate ancestor (Reschly et al. 2008b). These findings suggest the binding capacity expansion to a wider diversity of oxysterols in the vertebrate ancestor (see **Chapter 3** for details). Considering the nature of the vertebrate LXR ligands, the evolutionary binding profile of LXR was reconstructed by exploitation of intermediaries of the bile acid synthesis pathways, proposing a parallel and opportunistic evolution of both metabolic and signalling pathways in gnathostomes (see **Chapter 3** for details). Moreover, given the apparent redundancy in LXR paralog function (Korach-André et al. 2010) and their ligand specificity (Reschly et al. 2008b; **Chapter 3**), the retention of both *LXR* genes in Chondrichthyes and mammals during evolution suggests that differences in their transcriptional regulation or post-transductional modifications could favour one of the paralogs for a specific metabolic pathways (see **Chapter 3** for details).

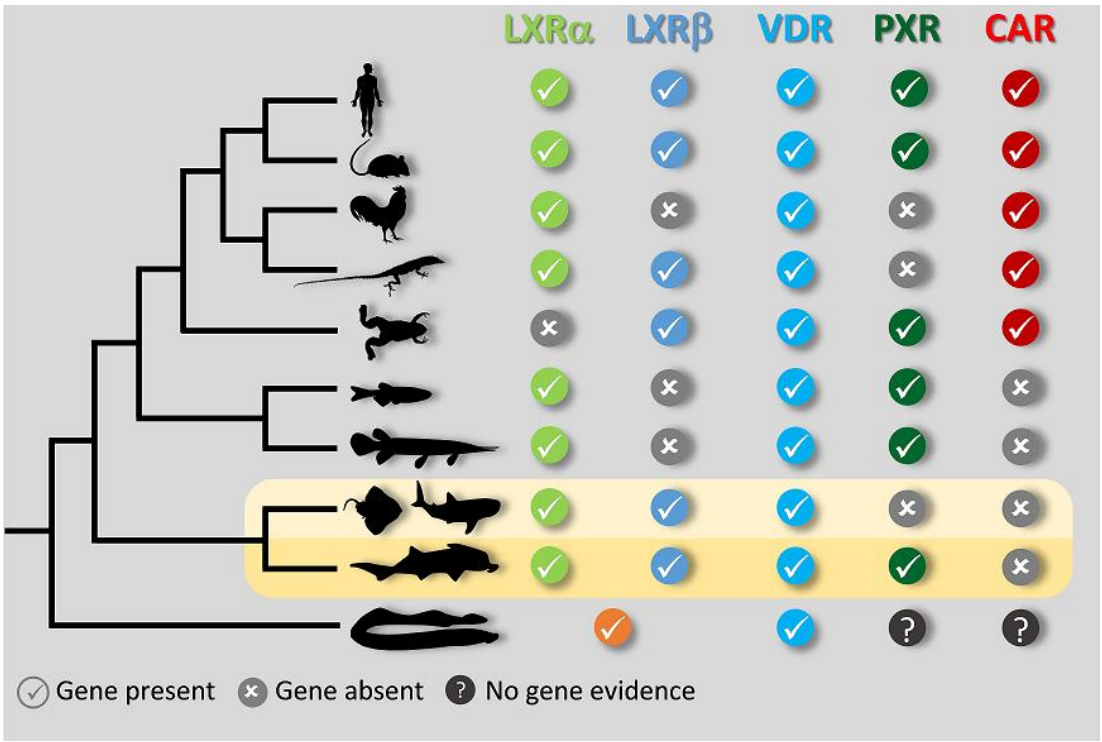


Figure 7.2. Distribution of *LXR*, *VDR*, *PXR* and *CAR* genes in vertebrates.

I next examined the NR1I subfamily, which is composed by *VDR* (*NR1I1*), *PXR* (*NR1I2*) and *CAR* (*NR1I3*). While *VDR* regulates hundreds of vitamin D-responsive genes involved in in calcium and phosphate homeostasis, *PXR* and *CAR* regulate the expression of genes involved in detoxification responses, such as, CYP enzymes and drug transporters (Krasowski et al. 2011), being part of the gene network responsible for protection against endogenous or exogenous toxic compounds, the so-called vertebrate "*chemical defensome*" (Goldstone et al. 2006). Unlike *VDR*, which has been identified in cyclostomes and gnathostomes (Kollitz et al. 2015), *CAR* is apparently a tetrapod novelty (Mathäs et al. 2012; Zhao et al. 2015). The distribution of *PXR* is much more variable, having been shaped by gene loss events in birds, reptiles, and in some teleosts (Mathäs et al. 2012; Eide et al. 2018). The investigation in chondrichthyan genomes identified a holocephalan *PXR* ortholog, suggesting an early emergence of *PXR* in the gnathostome ancestor and proposed another event of *PXR* gene loss in elasmobranchs (**Figure 7.2**, see **Chapter 4** for more details). Consistent with the low identity of the LBD sequence across species (Krasowski et al. 2005; Bainy et al. 2013), orthologous *PXR* genes are activated by a huge array of structurally distinct molecules with variable sensitivities (Handschin et al. 2000; Moore, LB et al. 2002; Milnes et al. 2008; Scheer et al. 2010; Sui et al. 2010; Lille-Langøy et al. 2015). In this study, we verified that elephant shark PXR is activated by estrogen, being less sensitive to selected environmental pollutants than zebrafish PXR. However, a set of non-tested xenobiotics with potential to interact with the holocephalan PXR cannot be discarded. Additionally, we found a unique expression profile confined to skin and gills, suggesting that PXR could act as a specialized steroid-like sensor in skin. In fact, estrogens are known to participate in skin homeostasis (Shah & Maibach 2001), and further studies assumed the role of PXR in human and rodent skin (Schmuth et al. 2014; Elentner et al. 2018), supporting our hypothesis for the role of an estrogen-dependent PXR in the maintenance of the smooth, rubbery and scale-less skin of chimaeras (see **Chapter 4** for details). These results revealed the plasticity of the transcriptional regulation of detoxification genes during vertebrate evolution.

Invertebrate lineages represent the vast majority of metazoans, comprising about 95% of the species (GIGA Community of Scientists 2014; Roskov et al. 2019). Less than 1% of invertebrates have their genomes sequenced (GIGA Community of Scientists 2014), and only a minority have had their NR gene repertoires identified and characterized (Santos et al. 2018). Thus, the disruption of the endocrine system by EDCs having NRs as pivotal targets is still poorly explored in invertebrates (Katsiadaki 2019). However, in a few cases xenobiotic compounds have been shown to target NRs in invertebrate lineages. For example, the organotin TBT has been classified as an obesogen, and known to target the vertebrate RXR-PPAR heterodimers (Kanayama et

al. 2005; Grün & Blumberg 2006; le Maire et al. 2009; Capitão et al. 2018; Barbosa et al. 2019), which is responsible to modulate lipid metabolism (Mello 2010). In contrast, in invertebrate gastropod molluscs the same chemical, TBT, has been described to promote RXR-mediated imposex, a reproductive endocrine impairment (Nishikawa et al. 2004; Castro et al. 2007; Lima et al. 2011; Stange et al. 2012; André et al. 2017). Moreover, disturbance of lipid homeostasis in molluscs, arthropods and tunicates has also been observed (Capitão et al. 2017). However, no studies have been conducted in these lineages to establish if NR-mediated lipid homeostasis perturbation by exposure to TBT is implicated. Moreover, outside deuterostomes, *PPAR* orthologous genes have only been reported in molluscs (Vogeler et al. 2017; Kim et al. 2017), suggesting that the disruption of lipid metabolism by TBT in such *PPAR*-absent species can involve other NRs, as *RXR* which has been identified in most metazoans (Sladek 2011; Philip et al. 2012; **Chapter 5**). The marine rotifer *Brachionus koreanus*, model organism for ecotoxicological (Dahms et al. 2011), has no *PPAR* ortholog gene in its recently sequenced genome (Kim et al. 2017) and the exposure to TBT has been associated to a decrease of polyunsaturated fatty acids and an increase of saturated fatty acids, suggesting the lipid accumulation as either a rotifer-specific signalling or feedback mechanisms of lipid homeostasis or as a detoxification mechanism (Lee et al. 2019). Furthermore, rotifer RXR was demonstrated to induce gene transcription upon binding to TBT despite the altered composition of the rotifer RXR LBD (see **Chapter 5** for details). These results suggest that *RXR* might participate in the modulation of the lipid metabolism disruption of *B. koreanus* exposed to TBT. However, future studies should be conducted to assess the engagement of other NRs (e.g. NR1J) in this process.

In addition to lipid metabolism, other biological processes are modulated by NRs and subject to endocrine disruption. In vertebrates, cell proliferation, differentiation and apoptosis, organogenesis, reproduction and embryonic development are modulated by the RA-signalling pathways through RXR-RAR heterodimers (Ross et al. 2000; Maden & Hind 2003; Theodosiou et al. 2010; Cunningham & Duester 2015; Ghyselinck & Duester 2019). Firstly assigned as a vertebrate novelty, RA-signalling modules were identified and studied in invertebrate chordates, such ascidians and amphioxus (Hisata et al. 1998; Kamimura et al. 2000; Escrivà et al. 2002, 2006). Further genome sequencing of numerous non-chordate invertebrate species suggested an earlier origin of metabolic and signalling modules, with loss of *RAR* orthologs in appendicularian and ecdysozoan species (arthropods and nematodes) (Cañestro et al. 2006; Howard-Ashby et al. 2006; Marlétaz et al. 2006; Campo-Paysaa et al. 2008; Albalat & Cañestro 2009; Albalat 2009). However, lophotrochozoans *RAR* orthologs were identified and

characterized in several molluscs (Campo-Paysaa et al. 2008; Gutierrez-Mazariegos et al. 2014; Urushitani et al. 2013; André et al. 2019) and in one annelid (Handberg-Thorsager et al. 2018). These findings suggested that *RAR* evolution was defined by events of gene loss in several bilaterian lineages and its ability to bind RA and activate gene transcription was lost in molluscs (Handberg-Thorsager et al. 2018; André et al. 2019). The early branching phylogenetic position of the Priapulida phylum in the Ecdysozoa tree and their slow evolving molecular characters (Webster et al. 2006), prompted our attention to investigate the publically available genome of the priapulid worm *P. caudatus* and to scrutinize the evolutionary history of *RAR* in Metazoa. Our searches retrieved a priapulid *RAR* ortholog that we demonstrated to bind RA and activate gene transcription (see **Chapter 6** for details), similarly to previous findings with the annelid *RAR* ortholog (Handberg-Thorsager et al. 2018). Furthermore, we verified that the priapulid RAR is unresponsive to the tested RAR endocrine disruptor pesticides which are persistent in *P. caudatus* habitat (see **Chapter 6** for details). Further studies will be necessary to explore other endogenous ligands and xenobiotics to bind and modulate gene transcription *via* RAR in priapulids. Overall, these data contributed to elucidate the earlier origin of *RAR* in Bilateria ancestor, formerly as a RA low-affinity sensor and then, as RA high-affinity receptor in chordates. Therefore, we highlighted Priapulida as an important model to elucidate Ecdysozoa evolution, and we emphasized the need to go through all metazoan phyla for a broader and assertive view of the evolutionary history of Metazoa.

7.2. Conclusions

Gene duplication is an evolutionary process contributing to genome expansion and allowing one duplicate to evolve a new function or splitting ancestral functions between the duplicates (Chapal et al. 2019). The results presented in this thesis emphasize the importance of the 2R WGD in the expansion and diversification of the NR superfamily genes and the pliable paralog retention prompted by independent events of gene loss in vertebrates. Additionally, the functional characterization of NRs suggested an exploitation of metabolic intermediates by the NR duplicates, resulting in sub-functionalization or neo-functionalization of these paralogs by accumulation of mutations.

On the other hand, the study of NRs in invertebrates provides a closer perspective at the ancestral state of each NR and, often a shift in ligand preferences is observed in their vertebrate orthologs (Reschly et al. 2008a, 2008b). The results showed the conservation of some endocrine disruption mechanisms among invertebrate and vertebrate lineages, despite the differences observed in their NR gene repertoires. Finally, the global analysis highlight the importance of non-classical model organisms in

order to bring new insights into evolution and function of NRs, raising the taxonomic significance on the evolutionary thinking.

Overall, this work offers a broader picture of the NR superfamily gene evolution in Metazoa (**Figure 7.3**). The collection of NR genes here presented clearly shows the dynamism of the evolutionary history of NRs, resulting in the plasticity of the endocrine system across metazoans and their exploitation by EDCs.

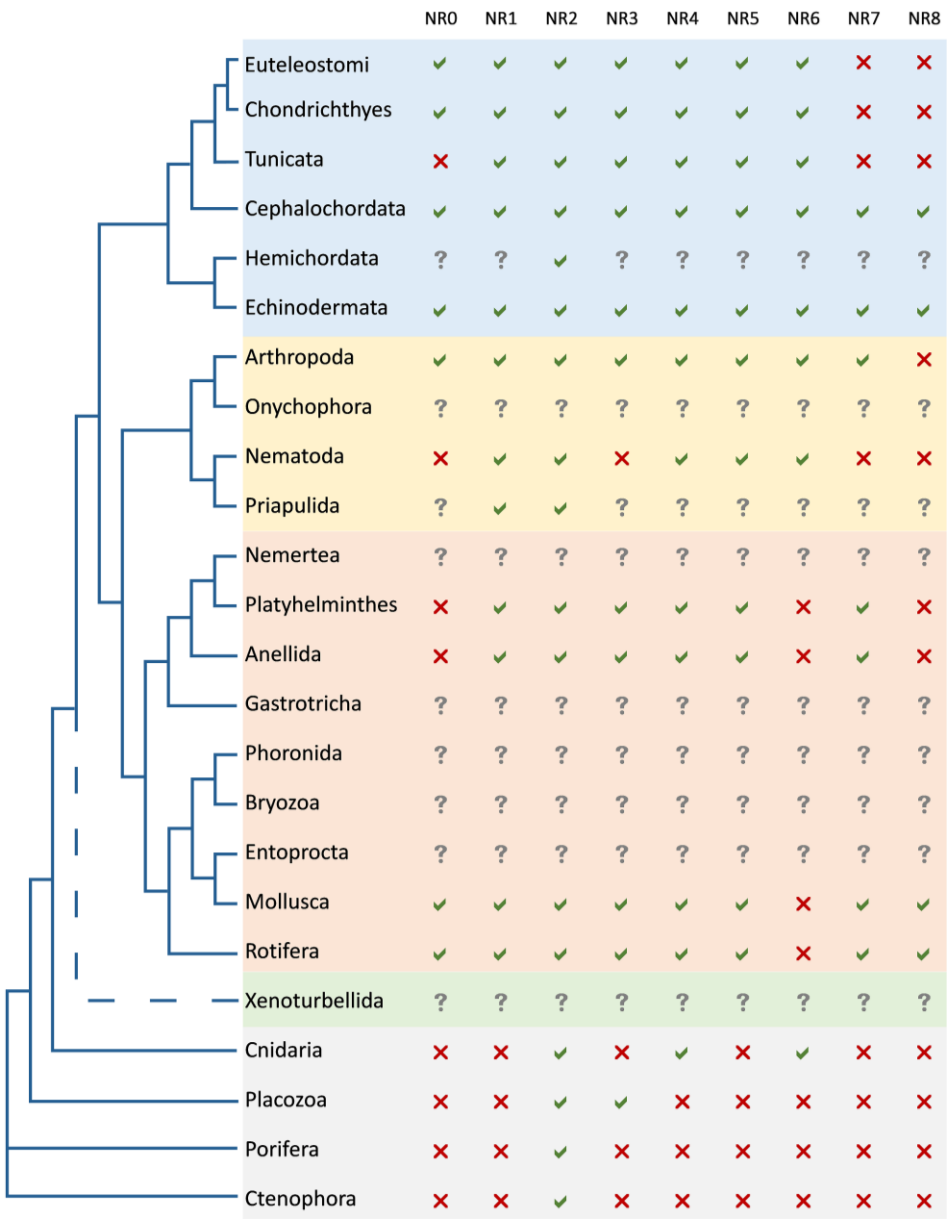


Figure 7.3. Distribution of the NR subfamilies among metazoans (current knowledge).

7.3. Future Challenges

Based on the findings of the present work, to better understand the evolution of the endocrine system in vertebrates it would be important to complete the functional characterization of NRs in Chondrichthyes in combination with the full identification and functional characterization of NRs in Cyclostomes which are early-branching within the vertebrates. In addition, the results obtained with the rotifer model signal the importance of exploring the possibility of other non-classical interactions between TBT and less conserved *RXR* among metazoan phyla, to assess how these lineages can be susceptible or not to NR-mediated endocrine disruption upon exposure to organotins, when no interactions were predictable. Taken the evolutionary thinking and having in consideration that the studies of evolution and function of NRs in the endocrine system has been mostly confined to Chordata, Mollusca, Annelid, Nematoda and Arthropoda (**Figure 7.3**), it would be pertinent in the future to complete the collection of NRs in Priapulida and to expand this approach to other Lophotrochozoa and Ecdysozoa phyla (Phoronida, Bryozoa, Gastrotricha, Loricifera, Onychophora, Kinorhyncha). Additionally, an inspection on *Xenoturbella bocki* (from Xenacoelomorpha) will be interesting given all the controversy around its phylogenetic classification.

Given the vast quantity of genomic data from next generation sequencing, this is the most stimulating Era to study the evolution of Metazoa putting in context the role of NRs in the endocrine disruption phenomenon.

7.4. References

Albalat R. 2009. The retinoic acid machinery in invertebrates: ancestral elements and vertebrate innovations. Mol Cell Endocrinol. 313:23–35. doi: 10.1016/j.mce.2009.08.029.

Albalat R, Cañestro C. 2009. Identification of Aldh1a, Cyp26 and RAR orthologs in protostomes pushes back the retinoic acid genetic machinery in evolutionary time to the bilaterian ancestor. Chem Biol Interact. 178:188–196. doi: 10.1016/j.cbi.2008.09.017.

Amores A, Catchen J, Ferrara A, Fontenot Q, Postlethwait J. 2011. Genome Evolution and Meiotic Maps by Massively Parallel DNA Sequencing: Spotted Gar, an Outgroup for the Teleost Genome Duplication. Genetics. 188:799–808. doi: 10.1534/genetics.111.127324.

André A et al. 2017. Cloning and functional characterization of a retinoid X receptor orthologue in *Platynereis dumerilii*: Na evolutionary and toxicological perspective. Chemosphere. 182:753−761. doi: 10.1016/j.chemosphere.2017.05.064.

André A et al. 2019. The retinoic acid receptor (RAR) in molluscs: Function, evolution and endocrine disruption insights. Aquat Toxicol. 208:80–89. doi: 10.1016/j.aquatox.2019.01.002.

Bainy A et al. 2013. Functional characterization of a full length pregnane X receptor, expression in vivo, and identification of PXR alleles, in zebrafish (*Danio rerio*). Aquat Toxicol. 142–143:447–457. doi: 10.1016/j.aquatox.2013.09.014.

Barbosa M et al. 2019. Linking chemical exposure to lipid homeostasis: A municipal waste water treatment plant influent is obesogenic for zebrafish larvae. Ecotoxicol Env. Saf. 182:109406. doi: 10.1016/j.ecoenv.2019.109406.

Benton M, Donoghue P, Asher R. 2009. Calibrating and constraining molecular clocks. In: The timetree of Life. Hedges, S & Kumar, S, editors. Oxford University Press pp. 35–86.

Bergman Å et al. 2013. Perspectives | Editorial The Impact of Endocrine Disruption : A Consensus Statement on the State of the Science. Env Heal Perspect. 121:104–106. doi: 10.1289/ehp.1205448.

Bertrand S et al. 2004. Evolutionary genomics of nuclear receptors: from twenty-five ancestral genes to derived endocrine systems. Mol Biol Evol. 21:1923–1937. doi: 10.1093/molbev/msh200.

Bertrand S, Belgacem M, Escriva H. 2011. Nuclear hormone receptors in chordates. Mol Cell Endocrinol. 334:67–75. doi: 10.1016/j.mce.2010.06.017.

Bonett R. 2016. Analyzing endocrine system conservation and evolution. Gen Comp Endocrinol. 234:3–9. doi: 10.1016/j.ygcen.2016.03.011.

Bookout A et al. 2006. Anatomical profiling of nuclear receptor expression reveals a hierarchical transcriptional network. Cell. 126:789–799. doi: 10.1016/j.cell.2006.06.049.

Braasch I et al. 2016. The spotted gar genome illuminates vertebrate evolution and facilitates human-teleost comparisons. Nat Genet. 48:427–437. doi: 10.1038/ng.3526.

Bridgham J et al. 2010. Protein evolution by molecular tinkering: diversification of the nuclear receptor superfamily from a ligand-dependent ancestor. PLoSBiol. 8:e1000497. doi: 10.1371/journal.pbio.1000497.

Campo-Paysaa F, Marlétaz F, Laudet V, Schubert M. 2008. Retinoic acid signaling in development: tissue-specific functions and evolutionary origins. Genesis. 46:640–656. doi: 10.1002/dvg.20444.

Cañestro C, Postlethwait J, Gonzalez-Duarte R, Albalat R. 2006. Is retinoic acid genetic machinery a chordate innovation? Evol Dev. 8:394–406. doi: 10.1111/j.1525-142X.2006.00113.x.

Capitão A et al. 2018. Evolutionary Exploitation of Vertebrate Peroxisome Proliferator-Activated Receptor γ by Organotins. Env. Sci Technol. 52:13951–13959. doi: 10.1021/acs.est.8b04399.

Capitão A, Lyssimachou A, Castro L, Santos M. 2017. Obesogens in the aquatic environment: an evolutionary and toxicological perspective. Env. Int. 106:153−169. doi: 10.1016/j.envint.2017.06.003.

Castro L et al. 2007. Imposex induction is mediated through the Retinoid X receptor signalling pathway in the neogastropod *Nucella lapillus*. Aquat Toxicol. 85:57–66. doi: 10.1016/j.aquatox.2007.07.016.

Castro L, Santos M. 2014. To bind or not to bind: the taxonomic scope of nuclear receptor mediated endocrine disruption in invertebrate phyla. Env. Sci Technol. 48:5361−5363. doi: 10.1021/es501697b.

Chapal M, Mintzer S, Brodsky S, Carmi M, Barkai N. 2019. Resolving noise-control conflict by gene duplication. PLoS Biol. 17:e3000289. doi: 10.1371/journal.pbio.3000289.

Chung A, Cooney A. 2003. The varied roles of nuclear receptors during vertebrate embryonic development. Nucl Recept Signal. 1:e007. doi: 10.1621/nrs.01007.

Cruzeiro C et al. 2016. A mollusk VDR/PXR/CAR-like (NR1J) nuclear receptor provides insight into ancient detoxification mechanisms. Aquat Toxicol. 174:61–69. doi: 10.1016/j.aquatox.2016.02.007.

Cunningham T, Duester G. 2015. Mechanisms of retinoic acid signalling and its roles in organ and limb development. Nat Rev Mol Cell Biol. 16:110–123. doi: 10.1038/nrm3932.

Dahms H, Hagiwara A, Lee J. 2011. Ecotoxicology, ecophysiology, and mechanistic studies with rotifers. Aquat Toxicol. 101:1–12. doi: 10.1016/j.aquatox.2010.09.006.

Degnan B, Vervoort M, Larroux C, Richards G. 2009. Early evolution of metazoan transcription factors. Curr Opin Genet Dev. 19:591–599. doi: 10.1016/j.gde.2009.09.008.

Diamanti-Kandarakis E et al. 2009. Endocrine-Disrupting Chemicals: An Endocrine Society Scientific Statement. Endocr Rev. 30:293–342. doi: 10.1210/er.2009-0002.

Didier D, Kemper J, Ebert D. 2012. Phylogeny, biology, and classification of extant holocephalans. In: Biology of Sharks and their Relatives. Carrier, J, Musick, J, & Heithaus, M, editors. CRC Press: Boca Raton, FL pp. 97–122.

Eide M et al. 2018. Independent losses of a xenobiotics receptor across teleost evolution. Sci Rep. 8:10404. doi: 10.1038/s41598-018-28498-4.

Elentner A et al. 2018. Epidermal Overexpression of Xenobiotic Receptor PXR Impairs the Epidermal Barrier and Triggers Th2 Immune Response. J Invest Dermatol. 138:109–120. doi: 10.1016/j.jid.2017.07.846.

Escrivá GH, Laudet V, Robinson-Rechavi M. 2003. Nuclear receptors are markers of animal genome evolution. J Struct Funct Genomics. 3:177–184. doi: 10.1023/A:1022638706822.

Escrivà H et al. 2006. Neofunctionalization in vertebrates: The example of retinoic acid receptors. PloS Genet. 2:e102. doi: 10.1371/journal.pgen.0020102.

Escrivà H, Holland N, Gronemeyer H, Laudet V, Holland L. 2002. The retinoic acid signaling pathway regulates anterior/posterior patterning in the nerve cord and pharynx of amphioxus, a chordate lacking neural crest. Development. 129:2905–2916.

Ghyselinck N, Duester G. 2019. Retinoic acid signaling pathways. Development. 146:dev167502. doi: 10.1242/dev.167502.

GIGA Community of Scientists. 2014. The Global Invertebrate Genomics Alliance (GIGA): developing community resources to study diverse invertebrate genomes. J Hered. 105:1–18. doi: 10.1093/jhered/est084.

Goldstone J et al. 2006. The chemical defensome: environmental sensing and response genes in the *Strongylocentrotus purpuratus* genome. Dev Biol. 300:366–384. doi: 10.1016/j.ydbio.2006.08.066.

Gronemeyer H, Gustafsson J, Laudet V. 2004. Principles for modulation of the nuclear receptor superfamily. Nat Rev Drug Discov. 3:950–964. doi: 10.1038/nrd1551.

Grün F, Blumberg B. 2006. Environmental obesogens: organotins and endocrine disruption via nuclear receptor signaling. Endocrinology. 147:S50−S55. doi: 10.1210/en.2005-1129.

Gutierrez-Mazariegos J et al. 2014. A mollusk retinoic acid receptor (RAR) ortholog sheds light on the evolution of ligand binding. Endocrinology. 155:4275–4286. doi: 10.1210/en.2014-1181.

Handberg-Thorsager M et al. 2018. The ancestral retinoic acid receptor was a low-affinity sensor triggering neuronal differentiation. Sci Adv. 4:eaao1261. doi: 10.1126/sciadv.aao1261.

Handschin C, Podvinec M, Meyer U. 2000. CXR, a chicken xenobiotic-sensing orphan nuclear receptor, is related to both mammalian pregnane X receptor (PXR) and constitutive androstane receptor (CAR). Proc Natl Acad Sci U S A. 97:10769–10774. doi: 10.1073/pnas.97.20.10769.

Hara Y et al. 2018. Shark genomes provide insights into elasmobranch evolution and the origin of vertebrates. Nat Ecol Evol. 2:1761–1771. doi: 10.1038/s41559-018-0673-5.

Hartenstein V. 2006. The neuroendocrine system of invertebrates: a developmental and evolutionary perspective. J Endocrinol. 190:555–570. doi: 10.1677/joe.1.06964.

Hisata K, Fujiwara S, Tsuchida Y, Ohashi M, Kawamura K. 1998. Expression and function of a retinoic acid receptor in budding ascidians. Dev Genes Evol. 208:537–546. doi: 10.1007/s004270050213.

Howard-Ashby M et al. 2006. Gene families encoding transcription factors expressed in early development of *Strongylocentrotus purpuratus*. Dev Biol. 300:90–107. doi: 10.1016/j.ydbio.2006.08.033.

Inoue J et al. 2010. Evolutionary origin and phylogeny of the modern holocephalans (Chondrichthyes: Chimaeriformes): a mitogenomic perspective. Mol Biol Evol. 27:2576–2586. doi: 10.1093/molbev/msq147.

Jaillon O et al. 2004. Genome duplication in the teleost fish *Tetraodon nigroviridis* reveals the early vertebrate proto-karyotype. Nature. 431:946–957. doi: 10.1038/nature03025.

Kabir E, Rahman M, Rahman I. 2015. A review on endocrine disruptors and their possible impacts on human health. Env. Toxicol Pharmacol. 40:241–258. doi: 10.1016/j.etap.2015.06.009.

Kalaany N, Mangelsdorf D. 2006. LXRs and FXR: the yin and yang of cholesterol and fat metabolism. Annu Rev Physiol. 68:159–191. doi: 10.1146/annurev.physiol.68.033104.152158.

Kamimura M, Fujiwara S, Kawamura K, Yubisui T. 2000. Functional retinoid receptors in budding ascidians. Dev Growth Differ. 42:1–8. doi: 10.1046/j.1440-169x.2000.00478.x.

Kanayama T, Kobayashi N, Mamiya S, Nakanishi T, Nishikawa J. 2005. Organotin compounds promote adipocyte differentiation as agonists of the peroxisome proliferator-activated receptor/retinoid X receptor pathway. Mol Pharmacol. 67:766−774. doi: 10.1124/mol.104.008409.

Katsiadaki I. 2019. Are marine invertebrates really at risk from endocrine-disrupting chemicals? Curr Opin Env. Sci Heal. 11:37–42. doi: 10.1016/j.coesh.2019.06.005.

Kim D et al. 2017. Genome-wide identification of nuclear receptor (NR) genes and the evolutionary significance of the NR1O subfamily in the monogonont rotifer *Brachionus* spp. Gen Comp Endocrinol. 252:219–225. doi: 10.1016/j.ygcen.2017.06.030.

King-Jones K, Thummel C. 2005. Nuclear receptors - a perspective from *Drosophila*. Nat Rev Genet. 6:311–323. doi: 10.1038/nrg1581.

Kollitz E et al. 2015. Molecular cloning, functional characterization, and evolutionary analysis of vitamin D receptors isolated from basal vertebrates. PLoS One. 10:e0122853. doi: 10.1371/journal.pone.0122853.

Korach-André M et al. 2010. Separate and overlapping metabolic functions of LXRalpha and LXRbeta in C57Bl/6 female mice. Am J Physiol Endocrinol Metab. 298:E167–E178. doi: 10.1152/ajpendo.00184.2009.

Krasowski M, Ni A, Hagey L, Ekins S. 2011. Evolution of promiscuous nuclear hormone receptors: LXR, FXR, VDR, PXR, and CAR. Mol Cell Endocrinol. 334:39–48. doi: 10.1016/j.mce.2010.06.016.

Krasowski M, Yasuda K, Hagey L, Schuetz E. 2005. Evolution of the pregnane X receptor: adaptation to cross-species differences in biliary bile salts. Mol Endocrinol. 19:1720–1739. doi: 10.1210/me.2004-0427.

Lecroisey C, Laudet V, Schubert M. 2012. The cephalochordate amphioxus: a key to reveal the secrets of nuclear receptor evolution. Br. Funct Genomics. 11:156–166. doi: 10.1093/bfgp/els008.

Lee M et al. 2019. Tributyltin Affects Retinoid X Receptor-Mediated Lipid Metabolism in the Marine Rotifer Brachionus koreanus. Env. Sci Technol. 53:7830–7839. doi: 10.1021/acs.est.9b01359.

Lemaire G, Balaguer P, Michel S, Rahmani R. 2005. Activation of retinoic acid receptor-dependent transcription by organochlorine pesticides. Toxico Appl Pharmacol. 202:38–49. doi: 10.1016/j.taap.2004.06.004.

Lille-Langøy R et al. 2015. Environmental contaminants activate human and polar bear (*Ursus maritimus*) pregnane X receptors (PXR, NR1I2) differently. Toxicol Appl Pharmacol. 284:54–64. doi: 10.1016/j.taap.2015.02.001.

Lima D et al. 2011. Tributyltin-induced imposex in marine gastropods involves tissue-specific modulation of the retinoid X receptor. Aquat Toxicol. 101:221–227. doi: 10.1016/j.aquatox.2010.09.022.

Maden M, Hind M. 2003. Retinoic acid, a regeneration-inducing molecule. Dev Dyn. 226:237–244. doi: 10.1002/dvdy.10222.

Maglich J et al. 2003. The first completed genome sequence from a teleost fish (Fugu rubripes) adds significant diversity to the nuclear receptor superfamily. Nucleic Acids Res. 31:4051–4058. doi: 10.1093/nar/gkg444.

le Maire A et al. 2009. Activation of RXR-PPAR heterodimers by organotin environmental endocrine disruptors. EMBO Rep. 10:367−373. doi: 10.1038/embor.2009.8.

le Maire A, Bourguet W, Balaguer P. 2010. A structural view of nuclear hormone receptor: endocrine disruptor interactions. Cell Mol Life Sci. 67:1219–1237. doi: 10.1007/s00018-009-0249-2.

Mangelsdorf D et al. 1995. The nuclear receptor superfamily: the second decade. Cell. 83:835–839. doi: 10.1016/0092-8674(95)90199-X.

Marlétaz F, Holland L, Laudet V, Schubert M. 2006. Retinoic acid signaling and the evolution of chordates. Int J Biol Sci. 2:38–47. doi: 10.7150/ijbs.2.38.

Mathäs M et al. 2012. Evolutionary history and functional characterization of the amphibian xenosensor CAR. Mol Endocrinol. 26:14–26. doi: 10.1210/me.2011-1235.

Mello T. 2010. Nuclear Receptors in the Control of Lipid Metabolism. Curr Cardiovasc Risk Rep. 4:142–149. doi: 10.1007/s12170-010-0080-1.

Metpally R, Vigneshwar R, Sowdhamini R. 2007. Genome inventory and analysis of nuclear hormone receptors in *Tetraodon nigroviridis*. J Biosci. 32:43–50. doi: 10.1007/s12038-007-0005-4.

Milnes M et al. 2008. Activation of steroid and xenobiotic receptor (SXR, NR1I2) and its orthologs in laboratory, toxicologic, and genome model species. Env. Heal. Perspect. 116:880–885. doi: 10.1289/ehp.10853.

Moore, LB et al. 2002. Pregnane X receptor (PXR), constitutive androstane receptor (CAR), and benzoate X receptor (BXR) define three pharmacologically distinct

classes of nuclear receptors. Mol Endocrinol. 16:977–986. doi: 10.1210/mend.16.5.0828.

Nishikawa J et al. 2004. Involvement of the retinoid X receptor in the development of imposex caused by organotins in gastropods. Env. Sci Technol. 38:6271−6276. doi: 10.1021/es049593u.

Noguera-Oviedo K, Aga D. 2016. Lessons learned from more than two decades of research on emerging contaminants in the environment. J Hazard Mater. 316:242–251. doi: 10.1016/j.jhazmat.2016.04.058.

Ouadah-Boussouf N, Babin P. 2016. Pharmacological evaluation of the mechanisms involved in increased adiposity in zebrafish triggered by the environmental contaminant tributyltin. Toxicol Appl Pharmacol. 294:32–42. doi: 10. 1016/j.taap.2016.01.014.

Paps J, Holland P. 2018. Reconstruction of the ancestral metazoan genome reveals an increase in genomic novelty. Nat Commun. 9:1730. doi: 10.1038/s41467-018-04136-5.

Philip S et al. 2012. Adaptive evolution of the Retinoid X receptor in vertebrates. Genomics. 99:81−89. doi: 10.1016/j.ygeno.2011.12.001.

Putnam N et al. 2008. The amphioxus genome and the evolution of the chordate karyotype. Nature. 453:1064–1071. doi: 10.1038/nature06967.

Read T et al. 2017. Draft sequencing and assembly of the genome of the world's largest fish, the whale shark: *Rhincodon typus* Smith 1828. BMC Genomics. 18:532. doi: 10.1186/s12864-017-3926-9.

Reitzel A, Tarrant A. 2009. Nuclear receptor complement of the cnidarian *Nematostella vectensis*: phylogenetic relationships and developmental expression patterns. BMC Evol Biol. 9:230. doi: 10.1186/1471-2148-9-230.

Reschly E et al. 2008a. Evolution of the bile salt nuclear receptor FXR in vertebrates. J Lipid Res. 49:1577–1587. doi: 10.1194/jlr.M800138-JLR200.

Reschly E et al. 2008b. Ligand specificity and evolution of liver X receptors. J Steroid Biochem Mol Biol. 110:83–94. doi: 10.1016/j.jsbmb.2008.02.007.

Riu A et al. 2011. Characterization of novel ligands of ERα, Er β, and PPARγ: the case of halogenated bisphenol A and their conjugated metabolites. Toxicol Sci. 122:372–382. doi: 10.1093/toxsci/kfr132.

Roskov Y et al. 2019. Species 2000 & ITIS Catalogue of Life, 2019 Annual Checklist. Species 2000, Naturalis: Leiden, Netherlands www.catalogueoflife.org/annual-checklist/2019.

Ross S, McCaffery P, Drager U, De Luca L. 2000. Retinoids in embryonal development. Physiol Rev. 80:1021–1054. doi: 10.1152/physrev.2000.80.3.1021.

Santos M, Ruivo R, Capitão A, Fonseca E, Castro L. 2018. Identifying the gaps: Resources and perspectives on the use of nuclear receptor based-assays to improve hazard assessment of emerging contaminants. J Hazard Mater. 358:508–511. doi: 10.1016/j.jhazmat.2018.04.076.

Scheer N, Ross J, Kapelyukh Y, Rode A, Wolf C. 2010. *In vivo* responses of the human and murine pregnane X receptor to dexamethasone in mice. Drug Metab Dispos. 38:1046–1053. doi: 10.1124/dmd.109.031872.

Schmuth M, Moosbrugger-Martinz V, Blunder S, Dubrac S. 2014. Role of PPAR, LXR, and PXR in epidermal homeostasis and inflammation. Biochim Biophys Acta. 1841:463–473. doi: 10.1016/j.bbalip.2013.11.012.

Schubert M et al. 2008. Nuclear hormone receptor signaling in amphioxus. Dev Genes Evol. 218:651–665. doi: 10.1007/s00427-008-0251-y.

Shah M, Maibach H. 2001. Estrogen and skin. An overview. Am J Clin Dermatol. 2:143–150. doi: 10.2165/00128071-200102030-00003.

Sladek F. 2011. What are nuclear receptor ligands? Mol Cell Endocrinol. 334:3–13. doi: 10.1016/j.mce.2010.06.018.

Soares J et al. 2009. Disruption of zebrafish (*Danio rerio*) embryonic development after full life-cycle parental exposure to low levels of ethinylestradiol. Aquat Toxicol. 95:330–338. doi: 10.1016/j.aquatox.2009.07.021.

Stange D, Sieratowicz A, Oehlmann J. 2012. Imposex development in *Nucella lapillus* – evidence for the involvement of retinoid X receptor and androgen signalling pathways in vivo. Aquat Toxicol. 106–107:20–24. doi: 10.1016/j.aquatox.2011.10.010.

Sui Y et al. 2010. Bisphenol A and its analogues activate human pregnane X receptor. Env. Heal. Perspect. 120:399–405. doi: 10.1289/ehp.1104426.

Theodosiou M, Laudet V, Schubert M. 2010. From carrot to clinic: an overview of the retinoic acid signaling pathway. Cell Mol Life Sci. 67:1423–1445. doi: 10.1007/s00018-010-0268-z.

Urushitani H et al. 2013. Cloning and characterization of the retinoic acid receptor-like protein in the rock shell, Thais clavigera. Aquat Toxicol. 142–143:403–413. doi: 10.1016/j.aquatox.2013.09.008.

Venkatesh B et al. 2006. Ancient noncoding elements conserved in the human genome. Science (80-.). 314:1892. doi: 10.1126/science.1130708.

Venkatesh B et al. 2014. Elephant shark genome provides unique insights into gnathostome evolution. Nature. 505:174–179. doi: 10.1038/nature12826.

Venkatesh B et al. 2007. Survey Sequencing and Comparative Analysis of the Elephant Shark (*Callorhinchus milii*) Genome. PLoS Biol. 5:e101. doi: 10.1371/journal.pbio.0050101.

Vogeler S, Galloway T, Isupov M, Bean T. 2017. Cloning retinoid and peroxisome proliferator-activated nuclear receptors of the Pacific oyster and *in silico* binding to environmental chemicals. PLoS One. 12:e0176024. doi: 10.1371/journal.pone.0176024.

Vogeler S, Galloway T, Lyons B, Bean T. 2014. The nuclear receptor gene family in the Pacific oyster, *Crassostrea gigas*, contains a novel subfamily group. BMC Genomics. 15:369. doi: 10.1186/1471-2164-15-369.

Wang Q et al. 2012. Community annotation and bioinformatics workforce development in concert-Little Skate Genome Annotation Workshops and Jamborees. Database (Oxford). bar064. doi: 10.1093/database/bar064.

Webster B et al. 2006. Mitogenomics and phylogenomics reveal priapulid worms as extant models of the ancestral Ecdysozoan. Evol Dev. 8:502–510. doi: 10.1111/j.1525-142X.2006.00123.x.

Zhao Y, Zhang K, Giesy J, Hu J. 2015. Families of nuclear receptors in vertebrate models: characteristic and comparative toxicological perspective. Sci Rep. 5:8554. doi: 10.1038/srep08554.