



**Programa de Pós-Graduação em Biodiversidade e Evolução
Universidade Federal da Bahia
Instituto de Biologia**

A evolução molecular dos receptores nucleares em cordados

Joisilene de Jesus dos Santos

**Salvador
2023**

Joisilene de Jesus dos Santos

A evolução molecular dos receptores nucleares em cordados

Dissertação apresentada ao Instituto de Biologia da Universidade Federal da Bahia para obtenção do título de mestre em Biodiversidade e Evolução pelo Programa de Pós-Graduação em Biodiversidade e Evolução.

Orientador (a): Vanessa Rodrigues Paixão-Côrtes

Salvador
2023

A ficha catalográfica vai ser confeccionada conforme as normas de Biblioteconomia.

Comissão julgadora

Vanessa Rodrigues Paixão-Côrtes
(Orientadora e presidente da banca)

Carlos Eduardo Guerra Amorim

Janaina Lima de Oliveira

Salvador
2023

À família e amigos.

“Ensinar não é transferir conhecimento, mas criar as possibilidades para a sua própria produção ou a sua construção”.

AGRADECIMENTOS

- ↪ Dra. Vanessa Rodrigues Paixão-Côrtes por toda orientação, confiança, companheirismo e amizade ao longo desses anos de parceria, sem a senhora não seria possível a realização desse projeto.
- ↪ A todos os professores do PPPGBioEvo que durante as disciplinas do programa e no Seminário de avaliação de projeto (SAP) contribuíram com o aperfeiçoamento da minha pesquisa.
- ↪ Marcus Javan pedrosa de Jesus pela parceria na elaboração de fórmulas no Excel, além do auxílio com as coletas nos bancos de dados genômicos.
- ↪ Kalena Ellen por me emprestar seu computador para agilizar a obtenção de dados da minha pesquisa.
- ↪ Colegas de Laboratório: Hallana e Gabriela por tirar algumas dúvidas, que foram essenciais para o andamento do trabalho.
- ↪ Cássia Cristina Sacramento por sempre me incentivar a continuar, quando tudo parecia difícil.
- ↪ Bárbara pelo apoio e amizade durante a construção desse projeto.
- ↪ À minha família e amigos por me apoiar em todos os momentos da minha vida.

Sumário

RESUMO	9
ABSTRACT	10
INTRODUÇÃO GERAL	11
ORIGEM E EVOLUÇÃO DO GENOMA DOS CORDADOS	13
ORIGEM E EVOLUÇÃO DOS RECEPTORES NUCLEARES	18
CLASSIFICAÇÃO DOS RECEPTORES NUCLEARES	21
OBJETIVO GERAL.....	32
OBJETIVOS ESPECÍFICOS	32
METODOLOGIA	32
RESULTADOS E DISCUSSÃO	34
AGNATOS/VERTEBRADOS SEM MANDÍBULA	39
CONCLUSÃO.....	54
REFERÊNCIAS	55

RESUMO

Receptores Nucleares (NRs) pertencem a uma superfamília de fatores de transcrição, que são ativados por ligantes, atuam na regulação da expressão de vários genes, e estão envolvidos em importantes funções celulares, tais como: desenvolvimento, homeostase, reprodução, crescimento, apoptose e metabolismo. Os NRs possivelmente surgiram a partir de uma recombinação quimérica de genes que estavam presentes há 800 milhões de anos atrás. O objetivo deste trabalho é descrever a possível origem e os padrões evolutivos dos NRs em cordados. A metodologia consiste em coletas em bancos de dados genômicos disponíveis e no uso de um conjunto de software para análise das sequências de DNA e proteínas. Para o alinhamento das sequências utilizamos o servidor Guidance 2, árvore dos genes da família foi construída no Mega 10, análise do padrão evolutivo dos genes foi aferida por meio do Pacote PAML 4.9 e a conservação da sintenia foi recuperada utilizando o browser Genômicus. Caracterizamos os NRs em 289 cordados, encontramos 22 NRs nos cefalocordados, 17 NRs Urocordados, 39 NRs no agnatas, e de 46 NRs a 76 genes em vertebrados, aves e peixes, respectivamente. Os dados revelaram que as sequências dos DBDs se mantiveram conservadas ao longo do processo evolutivo, para 45 NRs é o melhor modelo é o neutro (M1), e dois genes, NR4A1 e RORC, o melhor modelo que admite seleção (M8), mais somente 10% dos sítios destes genes estão sobre relaxamento da seleção negativa. A sintenia é conservada na maioria dos ortólogos nos tetrápodes. As análises filogenéticas, agrupam os genes de acordo com os grupos funcionais. Nossos dados de sintenia e de análise filogenética apoiam a hipótese de que os primeiros NRs terem sido sensores lipídicos, de baixa afinidade e através de duplicações, mutações e perdas de genes linhagem ou espécie-específicos ocorreram eventos de neofuncionalização nos NRs emergentes.

ABSTRACT

Nuclear Receptors (NRs) belong to a superfamily of transcription factors, activated by ligands, and act regulating the expression of several genes and involved on cellular functions, such as: development, homeostasis, reproduction, growth, apoptosis and metabolism. NRs possibly arose from a chimerical recombination of genes that were present 800 million years ago. The aim of this work is to describe the possible origin and evolutionary patterns of NRs in chordates. We search the available genomic databases and the use of a set of software for analysis of DNA and protein sequences. We use PAML 4.9 package for standard evolutionary analysis. We verified the syntenic conservation using the Genomicus browser (93V). We characterized the NRs in 289 chordates, found 22 NRs in cephalochordates, 17 NRs in Urochordates, 39 NRs in agnates and 46 NRs to 76 genes in vertebrates, birds and fish, respectively. The data revealed that the sequences of DBDs remained conserved throughout the evolutionary process. For 45 NRs the best model is M1 (neutral), and for two genes, NR4A1 and RORC, the best model is M8 (selection), but only 10% of the sites of these genes are under relaxed constrain. We found a conserved syntenic in most tetrapod orthologs. Our phylogenetic analyzes cluster genes according to their functional groups. Our analysis supports the hypothesis that the first NRs were lipid sensors with low affinity, and through duplications, gene losses, and neofunctionalization events originated new NRs.

INTRODUÇÃO GERAL

O número de espécies que habitam nosso planeta variou, consideravelmente, no decorrer de 3,5 bilhões de anos de evolução. Cada uma das espécies, atuais e extintas, possui um genoma particular, mas que compartilham inúmeros genes. Existem genes encontrados em todos os vertebrados, e em invertebrados como a *Drosophila* (*Drosophila melanogaster* - Meigen, 1830) e o verme (*Caenorhabditis elegans* - Maupas, 1900). Entre eles, estão os genes que participam do processo imunológico (como os codificadores de anticorpos e receptores de células T para o antígeno), moléculas de sinalização celular (como os genes ARHGAPs), e ainda fatores de transcrição (como os receptores nucleares (NRs) os genes HOX) que possuem um papel importante na morfogênese dos organismos multicelulares (Brunet et al., 2016; Carroll, 2008).

A história evolutiva dos genomas dos eucariotos revela diversas duplicações nos genomas de várias linhagens, essas cópias de material genético que servem como substratos para a evolução molecular e representam um importante acontecimento do processo evolutivo. A organização do genoma dos cordados sofreu influência das sucessivas duplicações do genoma inteiro (Whole Genome Duplication; WGD), chamadas de 1R e 2R (Round, R; 2RWGD). Em seguida, houve um retorno para o estado diplóide, que ocorreu de forma progressiva. Porém, alguns genes foram mantidos duplicados, formando e/ou expandindo famílias de genes. Posteriormente, ocorreram duplicações em pequena escala (small-scale duplicates, SSD) e outros eventos de WGD, como o 3R que ocorreu na linhagem dos peixes ósseos (Teleósteos), modificando de uma forma específica, o genoma da linhagem dos vertebrados (Pasquier et al., 2017). Esses eventos por serem encontrados em diversas famílias gênicas podem gerar duplicações, que dificultam a determinação de homologia entre os genes (ortólogos x parálogos). Os genes parálogos que surgiram da retenção de cópias da duplicação genômica também pode ser denominados ortólogos.

Nas análises filogenéticas de famílias gênicas é comum o fenômeno da atração de ramos longos, onde linhagens que evoluíram rapidamente agrupam conjuntamente, independentemente da verdadeira relação evolutiva (Markov & Laudet, 2011). Existem métodos alternativos para determinar a origem de certos genes. Vários estudos de

comparação de genomas têm demonstrado que existe uma conservação na sintenia (ordem) ou na vizinhança de genes (proximidade) em um cromossomo, em diferentes espécies de eucariotos e que organismos que se divergiram a milhões de anos apresentam segmentos de DNA homólogos contendo genes em posições relativamente idênticas (Paixão-Côrtes et al., 2013). Essa vizinhança compartilhada demonstra que esses genes podem ter importantes relações funcionais ou mecanismos regulatórios compartilhados, que foram mantidos juntos durante a evolução. A partir desse pressuposto, é possível desvendar padrões da evolução de famílias de genes e como estes influenciaram a organização do genoma dos cordados (Paixão-Côrtes et al., 2013)

Os Receptores Nucleares (NRs) são uma família de proteínas que atuam como fatores de transcrição e receptores de pequenas moléculas lipofílicas. A família dos NRs estão presentes em todos os metazoários (1.2 a 0.9 bilhões de anos atrás, período pré-cambriano), o primeiro ramo animal onde surgiram NRs funcionais foram as esponjas (Porífera). Os NRs expandiram em número e complexidade durante a diversificação dos metazoários, durante o Cambriano (0.5 bilhões de anos atrás). Os NRs estão envolvidos com o controle de diversos processos celulares, imprescindíveis ao funcionamento biológico, como o desenvolvimento, a homeostase, o metabolismo e a reprodução (Novac & Heinzl, 2004) (Badofsky et al., 2017).

A maioria dos NRs conhecidos, tem a sua função de fator de transcrição, ativada por ligantes (pequenas moléculas lipofílicas), embora alguns NRs funcionem de maneira independente do ligante. Alguns exemplos de ligantes específicos da família dos NRs incluem uma gama de compostos endógenos, como ácidos biliares, retinóicos, hormônios esteroides, hormônio tireoideano e vitamina D (Weikum et al., 2018). A ligação entre o ligante e um receptor nuclear (NR) resulta em uma mudança conformacional no receptor, que, por sua vez, liga-se a sequências específicas de DNA em todo o genoma. Uma vez ligadas ao DNA, proteínas co-reguladoras, remodeladores de cromatina e a maquinaria transcricional geral são recrutados para as regiões do DNA (cis regulatórias). Após a ativação, os NRs ativam ou inibem a transcrição de genes através da ligação a cis- elementos de ação específicos nos promotores dos genes alvo. As ações dos NRs são influenciadas por diversos fatores, como disponibilidade dos ligantes, a estrutura do sítio alvo, influências de outras vias de sinalização e interações com outros fatores celulares (Lonard & O'Malley, 2007).

Além dos sítios de ligação ao DNA canônicos (motivos 5'-AGAACA-3'-like ou 5'-AGGTCA-3'-like), alguns NRs podem se ligar a outros sítios genômicos, como observado em grampos de RNA (hairpins), com afinidade semelhante à observada em eventos de ligação canônica ao DNA, esses achados sugerem novas rotas potenciais para a regulação da atividade dos NRs via RNA.

Esse papel duplo, receptor e fator de transcrição torna os NRs peças-chaves para a evolução. NRs conjuntamente com bHLH-PAS genes são as únicas famílias de fatores de transcrição que possuem a habilidade de ser uma ligação direta entre a regulação gênica e o ambiente metabólico, confirmando o papel central dos NRs na diversificação animal (Markov & Laudet, 2011).

Embora os NRs sejam bem estudados por apresentar alta relevância para diversos aspectos da fisiologia humana, e por serem alvos farmacológicos promissores para diversas doenças, algumas questões ainda permanecem em aberto, no qual, este trabalho busca contribuir com a compreensão do conjunto básico de NRs que contribuíram para a diversificação deste grupo após os vários eventos de duplicação genômica (WGD), minimizando lacunas no conhecimento da dinâmica de ganhos e perdas gênicas dessa superfamília.

ORIGEM E EVOLUÇÃO DO GENOMA DOS CORDADOS

Os cordados são divididos em três subfilos: cefalocordados (Leptocardii), Urocordados (Ascidiacea; tunicados) e vertebrados. Os Cefalocordados, comumente conhecidos como lancelet ou anfioxo, constituem o grupo irmão de Vertebrata e Tunicata (Fig.1) (Bertrand & Escriva, 2011; Carvalho et al., 2021).

Anfioxos (lancelet) são organismos pequenos, bentônicos e filtradores marinhos (Bertrand & Escriva, 2011). Possuem três gêneros: Branchiostoma, Epigonichthys e Asymmetron (Figura.1). Os anfioxos são organismos modelos no estudo da história evolutiva dos cordados, em virtude do seu genoma, morfologia e localização filogenética (Bertrand & Escriva, 2011; Kon et al., 2007).

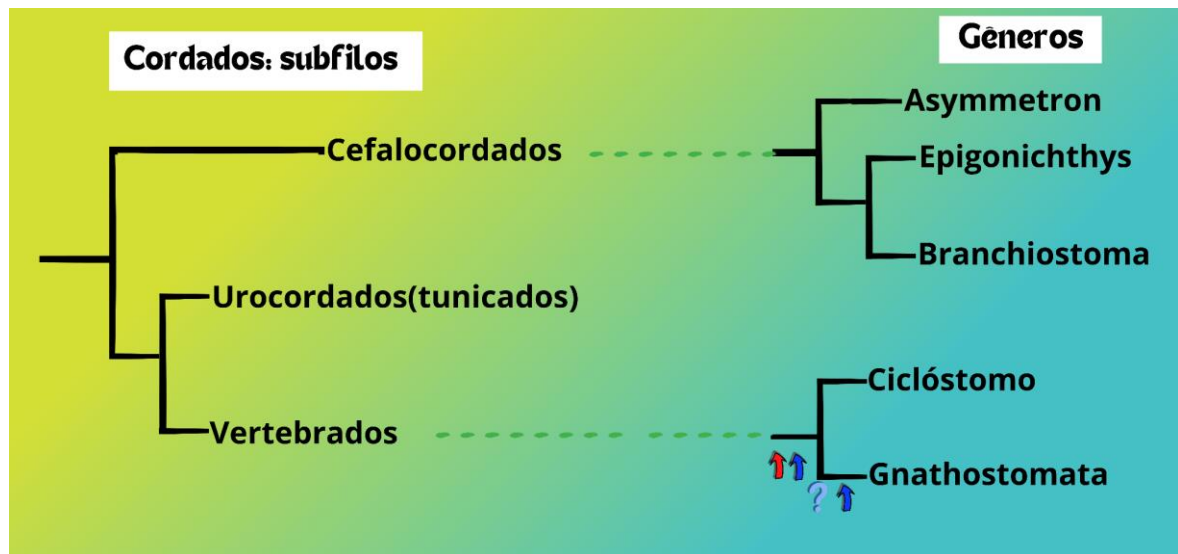


Figura 1. (Adaptado: Bertrand & Escriva, 2011) **Relações filogenéticas entre subfilos de cordados.** (A) O filo dos cordados pode ser dividido em três subfilos: cefalocordados (anfioxo), que possuem posição basal; vertebrados; e urocordados. (B) O subfilo cefalocordado é dividido em três gêneros: Branchiostoma, Epigonichthys e Asymmetron. (C) genomas de vertebrados confirmam que duas rodadas de duplicação do genoma completo ocorreram especificamente na linhagem de vertebrados. As setas indicam os pontos evolutivos em que ocorreram as duas duplicações completas do genoma. Acredita-se que o primeiro evento de duplicação (seta vermelha) tenha ocorrido na base da linhagem dos vertebrados, antes da divisão ciclóstomo-gnatostomado, enquanto o momento exato da segunda duplicação do genoma (setas azuis) ainda é motivo de debate (Kon et al., 2007).

Os urocordados(tunicados) é subfilo de organismos marinhos, dividido em três classes: Ascidiacea (classe mais abundante), Thaliacea e Larvacea. Esse grupo caracteriza-se por serem organismos sésseis (fixos a substratos) e filtradores (Lecroisey et al., 2012), além de possuir celulose em várias regiões do corpo, geralmente largas e altamente cristalinas (Kimura & Itoh, 2007).

Os tunicados têm sido utilizados intensamente em investigações biológicas. O que favorece o estabelecimento de conceitos chaves para a determinação de hipóteses evolutivas, o que engloba, conceitos básicos sobre desenvolvimento e mecanismos transcricionais (Lemaire, 2009).

Os vertebrados são um grupo de animais que possuem características exclusivas, tais como, tecido ósseo, glóbulos vermelhos, glândula tireoide e espinha dorsal feita de numerosos ossos ou segmentos cartilaginosos (as vértebras). Fazem parte desse amplo grupo, todos os mamíferos, aves, répteis, anfíbios, peixes ósseos, tubarões e lampreias. Análises de sequências moleculares confirmaram que todos os deuterostômios (equinodermos, hemicordados e cordados) estão intimamente relacionados entre si (Figura 2). (Prothero & Tamura, 2022).

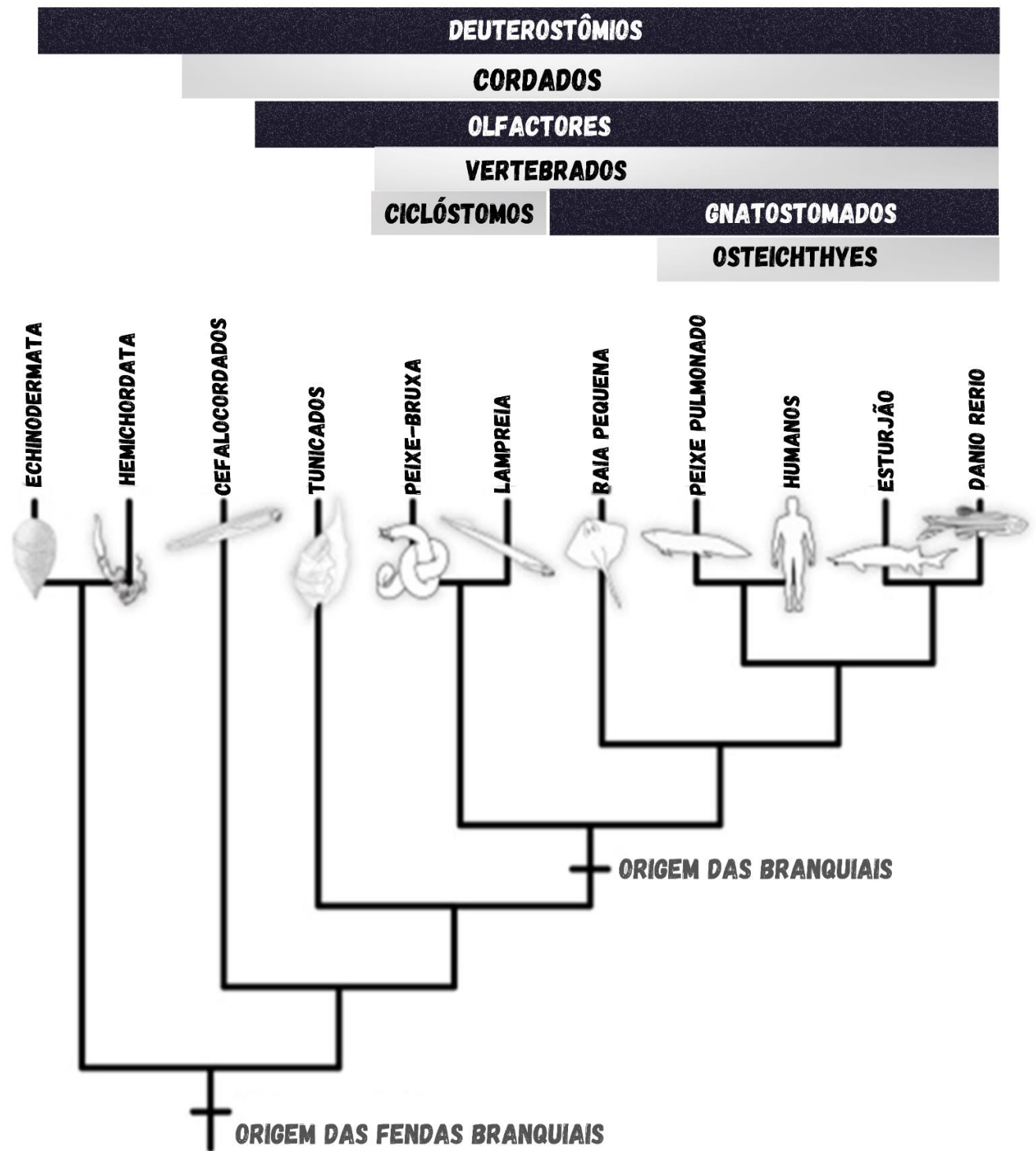


Figura 2. Adaptado. Compreensão atual das inter-relações de deuterostômios, incluindo cordados, olfactores, vertebrados, ciclóstomos e gnatostomados (P. Donoghue, 2017).

Os vertebrados possuem notocorda, sistema circulatório com um coração de duas câmaras. Durante sua embriologia, eles têm uma região que se desenvolve ao longo da coluna vertebral chamada crista neural (Prothero & Tamura, 2022.)

A origem dos vertebrados é caracterizada por mudanças genômicas, embriológicas e fenotípicas. Dados de estudos anteriores sugerem que duplicações no genoma inteiro impulsionaram eventos de macroevolução dos organismos. Favorecendo, consecutivamente, o surgimento de possíveis novidades adaptativas (Donoghue & Purnell, 2009; P. C. J. Donoghue & Keating, 2014).

As duplicações de genes e dos genomas têm sido muito debatidas, desde que o geneticista e biólogo evolutivo Susumu Ohno, propôs que duas rodadas de duplicação do genoma completo (2R) ocorreram na base da linhagem dos vertebrados (Ohno, 1970). Essa poliploidização foi observada em análises de sequências completas do genoma de espécies ao longo da linhagem de cordados (Dehal & Boore, 2005), corroborado quando a primeira sequência do genoma do anfioxo publicado por Putnam et al. (2008).

Os genes duplicados surgem da duplicação do genoma inteiro ou da duplicação de genes individuais. Após um evento de duplicação do genoma as duplicatas geralmente são perdidas (pseudogenização), a menos que suas funções e as dos genes originais diverjam, seja por subfuncionalização dos dois genes, ou por ganho de função da duplicata. Houve, pelo menos, três grandes eventos de duplicação do genoma inteiro (WGD) na evolução animal. Dois ocorreram perto da base dos vertebrados (aproximadamente a 500 e a 350 milhões de anos atrás) e o terceiro na base dos peixes teleósteos cerca de 230 milhões de anos atrás (Pasquier et al., 2017).

O primeiro evento de duplicação provavelmente ocorreu antes da separação da linhagem dos cefalocordados (anfioxos [Branchiostoma]) dos demais cordados cerca de 580 a 500 milhões de anos atrás (Kumar et al., 2017). Os urocordados (tunicados ou ascídias), parecem também ter, pelo menos, um único evento de duplicação, entretanto o genoma destes animais parece estar evoluindo rapidamente, com uma redução de tamanho (70-173 Mb em comparação com 520 Mb dos anfioxos e 3 Gb dos humanos), com perda de alguns genes-chave do desenvolvimento (por exemplo, vários genes Hox) e duplicação

independente de outros (por exemplo, Pax2/5/8 no apendicular *Oikopleura dioica*) (Holland, 2013).

As análises dos genomas dos agnatos inicialmente não resolveram definitivamente se o 2RWGD aconteceu antes ou depois da separação das linhagens de vertebrados sem e com mandíbula. Existem pelo menos duas hipóteses: uma análise propõe apenas uma rodada de WGD na base dos vertebrados, juntamente com duplicações em escala cromossômica em lampreias; a segunda hipótese propõe que o segundo evento de duplicação (2RWGD) precedeu a separação dos agnatos (lampreias [Petromyzontiformes] e peixes bruxas [Myxiniformes]) dos demais vertebrados (Gnatostomados) (Holland, 2018). Recentemente alguns estudos reforçam a hipótese que o 2RWGD aconteceu antes da separação dos vertebrados em duas linhagens (Cardoso et al., 2020; Ermakova et al., 2020).

Os agnatas, vertebrados basais, têm caracteres vertebrados chave que faltam nos cordados invertebrados - cefalocordados (anfioxos) e urocordados (tunicados ou Ascídias). Esses caracteres incluem a crista neural, telencéfalo, organizador ístmico no limite mesencéfalo/romencefálico e olhos pareados. A evolução de todas essas características subsequentes ao 2RWGD dá suporte à hipótese de que duplicatas gênicas deram aos vertebrados as ferramentas para aperfeiçoar estruturas antigas, além de criar novas morfologias. Os vertebrados retiveram preferencialmente réplicas de genes de desenvolvimento, incluindo aqueles que codificam fatores de transcrição e proteínas nas vias de sinalização (Holland, 2013), como os NRs.

ORIGEM E EVOLUÇÃO DOS RECEPTORES NUCLEARES

Os receptores nucleares (NRs) são uma superfamília de fatores de transcrição, filogeneticamente, relacionados, funcionalmente envolvidos em diversas funções moleculares (homeostase, reprodução, desenvolvimento, metabolismo). Esses podem atuar como ativadores ou repressores da transcrição gênica quando associados a um ligante. (Robinson-Rechavi, Garcia and Laudet, 2003; Miglioli *et al.*, 2021). Ou ainda, diversos NRs, conhecidos como receptores órfãos, que não possuem ligantes específicos, atuam formando dímeros com outros NRs, formando complexos e se ligando a motivos de DNA, o nos leva a hipótese de que NRs tinham inicialmente a função de fator de transcrição (DBD), e

posteriormente, adquiriu a capacidade de receptor através do acoplamento do domínio LDB (Badofsky et al., 2017; Markov & Laudet, 2011)

Os NRs, possivelmente, surgiram muito cedo na linhagem de metazoários e se diversificaram através de inúmeras duplicações e perdas gênicas (Schubert et al., 2008), os receptores adrenais e os receptores esteroides sexuais evoluíram muito mais tarde, em linhagens específicas, precedendo e durante a evolução dos vertebrados, começando com o aparecimento dos primeiros receptores de esteroides nos cefalocordados (Baker, 2019).

Nos humanos são encontrados 49 NRs no genoma completo, sendo um pseudogene (Robinson-Rechavi *et al.*, 2001). Porém, estudos anteriores demonstram que em animais invertebrados, como o anfioxo, possuem um número reduzido de NRs, em contraste, peixes teleósteos um complemento maior com cerca de 68 genes (Krasowski et al., 2011; J. Zhang et al., 2004). De certo, o número de NRs divergente entre as espécies, resultante das duplicações de genes e/ou perdas de genes em linhagens específicas durante processo evolutivo (Hwang et al., 2014; Schubert et al., 2008), indubitavelmente, contribuiu para diversificar funcionalmente os NRs e incluí-los na caixa de ferramentas genética básica que os animais adaptaram e refinaram no curso da evolução (Carroll, 2008; Fonseca et al., 2020).

ORGANIZAÇÃO ESTRUTURAL DOS RECEPTORES NUCLEARES

Durante a evolução, os genes da família NRs foram bastante conservados nos animais terrestres (E. Reschly & Krasowski, 2006). Os membros da superfamília dos receptores nucleares apresentam uma organização estrutural semelhante, caracterizada pela presença de cinco domínios funcionais. O domínio A/B ou região N-terminal, domínio C ou de ligação ao DNA (DBD/ *ligand binding domain* o que o torna um fator de transcrição), Domínio D (região dobradiça), Domínio E ou de ligação ao ligante (LBD/ *ligand binding domain*, o que o torna um receptor) e o F terminal (região C-terminal) (Fig.3).



Figura.3 – Representação esquemática da organização estrutural e funcional dos receptores nucleares. Adaptado: <https://pfam.xfam.org/family/PF00104>

A região N-terminal (domínio A/B) contém uma região ativa (AF-1). A região A/B possui comprimento variável entre 50-500 aminoácidos. A região mais conservada é o domínio de ligação ao DNA (DBD, domínio C), caracterizada por possuir um motivo, que garante a especificidade da ligação molecular ao DNA, favorecendo a dimerização intermolecular (Robinson-Rechavi et al., 2003). O domínio D atua como uma dobradiça entre a região C (DBD) e a região E, que corresponde ao domínio de ligação ao ligante (LBD) no qual encontra-se o segundo domínio de ativação (AF-2), este, por sua vez, crucial para interações de NRs com coativadores e correpresores. Por fim, a região F, que não está presente em todas as NRs, não é muito bem caracterizada e não possui função clara conhecida. (Königshofer et al., 2021).

Os NRS se ligam aos elementos de respostas ao DNA, que são sequências motivo em regiões regulatórias próximas as regiões promotoras dos genes alvos (cis; repetições diretas ou invertidas espaçadas 5'-AGAACA-3' ou 5'-AGGTCA-3'). Os NRs se ligam ao DNA, como monômeros ou dímeros (homodímeros ou heterodímeros) (Helsen and Claessens, 2014).

O DBD é o domínio mais conservado entre os receptores nucleares. Consiste em dois módulos de dedo de zinco que formam a estrutura central. A α -hélice no primeiro módulo de dedo de zinco permite as interações específicas da sequência com o DNA. O segundo módulo de dedo de zinco permite que os DBDs-receptores hetero ou homodímero. O DBD também contém uma extensão C-terminal variável que estabiliza a ligação ao DNA fazendo contato no sulco menor do DNA ou participa dos contatos proteína-proteína do dímero DBD (Helsen and Claessens, 2014).

CLASSIFICAÇÃO DOS RECEPTORES NUCLEARES

Os NRs caracterizam-se por ser o maior grupo de reguladores transcricionais com 49 RNs conhecidos, divididos em 19 grupos (Figura 4 - Tabela suplementar 1). A classificação foi desenvolvida segundo um sistema unificado de nomenclatura (Committee, 1999). Os NRs também podem ser classificados como três subtipos de NR (Endócrinos, Órfãos adotados, e Órfãos) com base em seus ligantes fisiológicos e funções potenciais. Além disso, eles podem ser separados em cinco classes, de acordo com algumas especificidades funcionais (Receptores Esteroides, Receptores Heterodímeros; Sensores Lipídicos, Órfãos enigmáticos e Receptores Órfãos)(Sonoda, Pei and Evans, 2008).

Os receptores endócrinos que atuam todos como receptores de alta afinidade para hormônios sexuais, lipossolúveis e vitaminas. Esta classe inclui receptores para os hormônios esteroides (AR, ER, GR, PR e MR); Hormônios que formam heterodímeros (receptor do hormônio tireoidiano; TR) e vitaminas A (receptores do ácido retinóico; RAR) e D (receptor da vitamina D; VDR). Os receptores de esteroides funcionam como homodímeros e TR, VDR e RAR formam heterodímeros com o Receptor Retinóide X (RXR) (Sonoda, Pei and Evans, 2008).

Os receptores “desorfanizados”, inicialmente não possuíam ligantes conhecidos, e aos poucos esses ligantes foram identificados, como receptores lipídicos, como o ligante derivado da vitamina A, o ácido 9-cis retinóico, com alta afinidade para RXR, além dos ligantes para PPAR α , LXR, FXR e PXR. Incluídos entre os órfãos adotados estão os adotados “enigmáticos”, para os quais um ligante foi identificado, pelo menos para um dos subtipos, mas a natureza da regulação dependente de ligante na fisiologia não foi estabelecida. Este grupo inclui receptores cuja atividade e/ou interação com cofatores pode ser modulado em condições não fisiológicas por estrogênios (receptores relacionados ao estrogênio; ERR β e c) e/ou por moléculas endógenas como o colesterol (retinóides receptor órfão; ROR α), retinóides (ROR β), androstano (receptor de androstano constitutivo; CAR) ou fosfolipídios (fator esteroideogênico-1; SF-1 e homólogo do receptor de fígado humano-1; LRH-1) através da ligação ao bolso de ligação do ligante. Também incluído nesta classe está um receptor que é ligado constitutivamente por uma molécula de ácido graxo

provavelmente como um componente estrutural (fator nuclear hepatócito 4a; HNF4a) (Sonoda, Pei and Evans, 2008).

A terceira classe é composta de verdadeiros órfãos cujos ligantes (natural ou sintético) não foram identificados. Esta classe inclui receptores que provavelmente não são regulados por ligantes baseados no tamanho do bolso de ligação do ligante e na posição da alfa-hélice. Esses receptores são provavelmente regulados por disponibilidade do coativador, expressão do próprio receptor, covalente modificação ou uma combinação (Sonoda, Pei and Evans, 2008).

RECEPTORES ORFÃOS

1D

0B

2C

2E

2F

4A

5A

6A

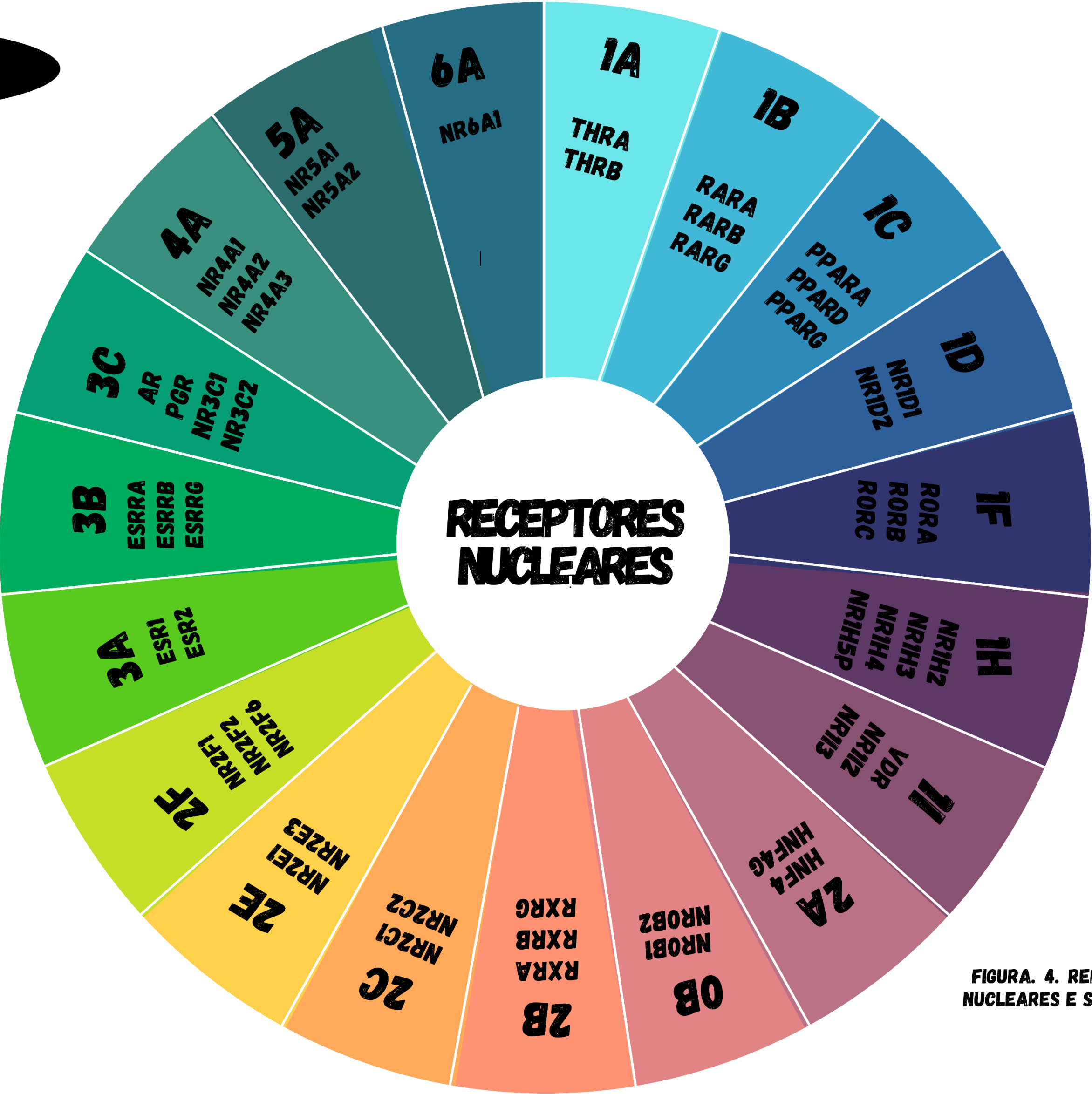


FIGURA. 4. REPRESENTAÇÃO DE TODOS OS RECEPTORES NUCLEARES E SEUS RESPECTIVOS GRUPOS.

(COMMITTEE , 1999)

0B. Receptor Nuclear da subfamília 0 grupo B (NR0B)

A família NR0B, representada em humanos por NR0B1 e NR0B2, não possuem a estrutura de domínio modular de ligação ao DNA (DBD), sendo denominados, “domínio singletons”. (Krasowski et al., 2005; Z. Zhang et al., 2004). O NR0B1, também conhecido como DAX-1, presente no cromossomo X, determina durante o processo de desenvolvimento a formação dos testículos e da glândula adrenal adulta (Fernandes-freitas et al., 2020). Some-se a isto, a reversão do sexo masculino para feminino, em caso, de possíveis duplicações, e hipoplasia adrenal congênita se ocorrer mutações e deleções (Ludbrook & Harley, 2004). Já o NR0B2 também denominado SHP, presente no cromossomo 1, parece estar envolvido em uma diversidade funções associadas a regulação da homeostase metabólica, regula a transcrição de uma variedade de genes alvo e de funções fisiológicas, em síntese, age diretamente com os receptores nucleares e outros fatores de transcrição. (Lee et al., 2007).

1A. Receptor do hormônio tireoide (THR)

A subfamília THR é composta por THRA e THRB, caracteriza-se por serem ativados por hormônios tireoidianos, e funcionalmente, estarem envolvidos na homeostase e desenvolvimento em vertebrados. (Miglioli et al., 2021). Estudos sugerem a presença de endóstilo, órgão faríngeo considerado homólogo ao folículo da glândula tireóide em Urocordados (Tunicata), Vertebrata (Lampreia) e Cefalocordados, em hipótese, especula-se que esse hormônio tireoidiano é uma característica antiga nos cordados e, possivelmente, eventos de duplicação influenciou a diversificação tecidual mais extensa e especializada (Michio Ogasawara, 2000; Paris et al., 2008).

1B. Receptor de ácido retinóico (RAR)

Os receptores de ácido retinóico (RARs) são importantes mediadores da sinalização celular, envolvidos no processo de morfogênese, desenvolvimento e diferenciação (Cheng et al., 2015). Existem três isotipos de RARs (RARA, RARB, RARG), caracterizados por compartilharem estruturas e funções semelhantes. As diversas propriedades transcricionais contribuem para as funções biológicas distintas, e a possível conservação durante a radiação evolutiva dos vertebrados (Hauksdóttir et al., 2003).

1C. Receptor Alfa Ativado por Proliferador de Peroxissoma (PPAR)

Os receptores Alfa Ativado por Proliferador de Peroxissoma, podem ser encontrados em três subtipos (PPARA, PPPARD, PPARG). Ácidos graxos e eicosanoides foram identificados como ligantes naturais para os PPARs. Caracteriza-se por ser homólogos, porém localiza-se em tecidos divergentes e por distintos genes (Berger & Moller, 2001). Dentre seus relevantes papéis biológicos, destacam-se, metabolismo, regulação da homeostase, crescimento celular, desenvolvimento de câncer. Em consequência, os PPARs vem sendo bastante estudados por sua capacidade de melhorar as síndromes metabólicas e nos tratamentos de obesidade e aterosclerose (Hong *et al.*, 2019).

1D. Receptor Nuclear da subfamília 1 grupo D membro 1 (NR1D)

Essa subfamília é representada por NR1D1 e NR1D2, pertencem ao grupo dos receptores órfãos, NRs que não possuem ligantes específicos ou se ligam com baixa afinidade a hormônios. Esse grupo regula o metabolismo de carboidratos e lipídios, além de ritmos circadianos que influenciam na determinação dos padrões de sono e alimentação. O sistema circadiano, como muitas outras famílias multigênicas, sofreu influências de transformações gênicas que acarretou a alteração no número de cópias presentes nos animais. Nos insetos é encontrada em uma única cópia, enquanto, em vertebrados aparecem duplicados. A análise filogenética de trabalhos anteriores, indica que membros de cada grupo de genes circadianos foram gerados por um evento de duplicação de genes nos estágios iniciais da evolução dos vertebrados (Layeghifard *et al.*, 2008).

1F. Receptor órfão alfa relacionado a RAR

Essa subfamília possui, três isoformas (RORA, RORB e RORC), codificada por gene distintos, desempenha um papel regulador na manutenção de uma variedade de processos fisiológicos e patológicos, atuando como um repressor osteogênico na regulação da formação óssea e assim como os NR1D, atua na regulação do ritmo circadiano (Shujiong Feng, 2015).

1H. Receptor Nuclear da subfamília 1 grupo H (NR1H)

A subfamília NR1H é formada por NR1H2 (LXRa), NR1H3(LXRb), NR1H4 E NR1H5P. Sua funcionalidade está associada à homeostase do colesterol, o qual regula a

expressão de genes envolvidos no efluxo, transporte e excreção. No passado, esse grupo era classificado como receptores órfãos, pois não se conhecia seus ligantes específicos (Zhang *et al.*, 2017). Mas atualmente, oxiesteróis como 24(S)-hidroxicolesterol e 24(S),25-epoxicolesterol, ácidos biliares, foram identificados como seus ligantes específicos (Reschly *et al.*, 2008; Fonseca *et al.*, 2017).

Estudos apontam que em cordados os LXRs em invertebrados, á exemplo *Ciona intestinalis* tem ligantes específicos diferentes dos LXRs encontrados em vertebrados. Uma hipótese associada para este dado é a possibilidade de um único gene ancestral LXR presente em invertebrados ter divergido no início da evolução dos vertebrados em resposta a diferenças no metabolismo lipídico entre invertebrados e vertebrados (Reschly *et al.*, 2008).

Os NR1H4, também conhecido como FXR, possuem duas isoformas FXRa e FXRb, também denominado NR1H5P. Em humanos e primatas, FXRb é um pseudogene. No entanto, em outros mamíferos (cães, coelhos, roedores) codifica um receptor hormonal nuclear funcional que parece estar envolvido na biossíntese do colesterol, sendo relevantes para o estudo de doenças hepatobiliares e gastrointestinais (Gadaleta *et al.*, 2010)

Os pseudogenes são segmentos de DNA não gênicos que exibem um alto grau de similaridade de sequência com genes funcionais, porém contêm defeitos disruptivos, tais como códons de parada prematuros, mutações no sítio de splicing e mutações de deslocamento de quadro, que os impedem, consequentemente, de serem expressos adequadamente (Zhang, Nielsen and Yang, 2004).

1I. Receptor Nuclear da subfamília 1 grupo I (NR1I)

A subfamília NR1I é composta por NR1I2, NR1I3 e VDR. Desempenham funções associadas a regulação do metabolismo, transporte e excreção de compostos exógenos, hormônios esteroides, vitaminas, sais biliares e xenobióticos (Ekins *et al.*, 2008).

Logo, os membros dessa subfamília possuem a capacidade de se ligar compostos endógenos potencialmente tóxicos com baixa afinidade e iniciar mudanças na expressão gênica induzindo o metabolismo a excreção de compostos, como por exemplo, a indução da expressão do citocromo P450 3A4 em humanos, funcionalmente importante em mamíferos, por oxidar medicamentos, toxinas, alimentos, geralmente, inativando-as e tornando-as mais

fáceis de serem eliminadas. Estes também atuam na síntese de lipídios, como o colesterol (E. Reschly & Krasowski, 2006).

O papel de desintoxicação do VDR parece limitado a detectar altas concentrações de certos sais biliares tóxicos, como o ácido litocólico, enquanto o NR1I2 e NR1I3 têm a capacidade de reconhecer compostos estruturalmente diversos (E. Reschly & Krasowski, 2006). Estudos científicos sugerem que essa subfamília surgiu da duplicação de um gene ancestral, possivelmente, no invertebrado *Ciona intestinalis* (E. Reschly & Krasowski, 2006). Imagina-se que a diversidade das ações biológicas dessa subfamília está associada com a capacidade dos organismos de adaptar-se às pressões seletivas (Bikle, 2011).

2A. Hepatócito Nuclear família do fator 4 (HNF4)

A subfamília HNF4 é formada por HNF4A e HNF4G, é um fator de transcrição envolvido na expressão específica de genes no fígado, rim e intestino (Chartier *et al.*, 1994). Em humanos, o HNF4 se liga a ácidos graxos endógenos como ligantes e regula a diferenciação de hepatócitos, metabolismo energético, desintoxicação de xenobióticos e manutenção de células-tronco na linhagem germinativa. Além disso, O HNF4 também participa do desenvolvimento primário do endoderma em sapos, regula a expressão de fatores de transcrição necessários para a especificação do endoderma em camundongos, podendo ser expressos em alguns estágios no desenvolvimento da *Ciona intestinalis* (Miglioli *et al.*, 2021). A presença desses NRs em organismos basais, sugere que esses receptores estavam presentes no início do processo evolutivo dos cordados.

2B. Receptor X Retinóico (RXR)

Os receptores retinóides X (RXRs) são membros altamente conservados da família de receptores de hormônios nucleares, envolvidos em vários processos do sistema endócrino (Fuchs *et al.*, 2014). Influenciados pelos múltiplos eventos de duplicação, esse grupo de NRs órfãos, possuem três subtipos (RXRA, RXRB, RXRG) em vertebrados, codificados por genes distintos, sendo encontrada várias isoformas para cada RXR. Em peixes, devido a uma duplicação específica do genoma inteiro (3R), há um gene RXRB adicional, mas apenas em zebrafish (entre os genomas de teleósteos sequenciados), são encontradas cópias adicionais dos genes RXRA e RXRG. Em vertebrados, RXRB é o subtipo amplamente expresso, RXRA é encontrado, principalmente, no fígado, rim, epiderme, intestino e domina a expressão de

RXR na pele, enquanto RXRG mostra uma expressão restrita em músculos, glândula pituitária e certas regiões do o cérebro.(Philip *et al.*, 2012).

Segundo os dados do trabalho de Evans & Mangelsdorf (2014), o primeiro ligante endógeno (ácido 9- *cis* retinóico, um metabólito da vitamina A) para um receptor nuclear órfão foi encontrado no RXR. Essa descoberta revelou a existência de uma via de sinalização anteriormente desconhecida, fornecendo caminhos para norteadores para compreender outros receptores órfãos e possíveis ligantes específicos.

Estudos demonstram que o RXR está presente nos genomas da maioria das espécies de metazoários (Fonseca *et al.*, 2020.), e que possivelmente, surgiu após a divergência de linhagens de cordados vertebrados e não vertebrados(urocordados) (Maeng *et al.*, 2012). Nesse sentido, indica-se que duplicação do genoma inteiro de vertebrados favoreceu seleção positiva nos parálogos, além de divergência funcional entre todos os pares de RXRs de vertebrados (Philip *et al.*, 2012).

2C. Receptor Nuclear da subfamília 2 grupo C (NR2C)

Essa subfamília composta por três membros em mamíferos, NR2C1 e NR2C2, participam de processos biológicos, como desenvolvimento, diferenciação celular e homeostase. Caracteriza-se por um grupo de receptor nuclear órfão que pode atuar como repressor ou ativador da transcrição. Estes parece desempenhar uma função fundamental na repressão da γ -globina durante a eritropoiese (Cui *et al.*, 2011), enquanto o NRC1 atua na proliferação e diferenciação de células-tronco. O parálogo NR2C2 atua na fertilidade masculina por meio do controle da espermatogênese e também no envelhecimento precoce (O'Geen *et al.*, 2010).

Segundo dados J. L. Baker *et al.*, (2016), o NR2C1, possivelmente, em nível populacional, possui diferenças específicas nas células tronco relacionadas ao desenvolvimento da pluripotencialidade, o que hipoteticamente sugere, aparecimento de novidades adaptativas ao longo do processo evolutivo.

2E. Receptor Nuclear da subfamília 2 grupo E(NR2E)

A subfamília NR2E em mamíferos é subdividida em NR2E1 E NR2E3. São considerados receptores nucleares órfãos, por não apresentarem ligantes específicos,

expressos em estruturas derivadas do olho e do prosencéfalo. Dados de trabalho do Kitambi & Hauptmann, (2007), mostraram que esses NRs são expressos durante o desenvolvimento embrionário e pós-embrionário.

Para compreender o desenvolvimento e diferenciação dos fotorreceptores, muitos estudos foram realizados, utilizando o peixe-zebra como organismo modelo. A escolha experimental deve-se à conservação da arquitetura das células da retina, ao formato da retina, aos hábitos diurnos, além da estrutura genômica e a sequência de aminoácidos serem semelhantes aos dos humanos. Resultados mostram que o perfil de expressão de desenvolvimento do peixe-zebra é consistente com as funções anatômicas conservadas evolutivamente nas estruturas do olho e do cérebro dos ortólogos humanos (Kitambi and Hauptmann, 2007; Connor *et al.*, 2022).

2F. Receptor Nuclear da subfamília 2 grupo F (NR2F)

Essa subfamília composta por três membros em mamíferos, NR2F1 e NR2F2, NR2F6, são altamente expressos no sistema nervoso central. Em geral, são ativados por ligantes (esteroides, hormônios tireoidianos e retinóides), mas podem não possuir nenhum ligante identificado, que se denomina receptores órfãos (Philip *et al.*, 2012).

Dados sugerem que NR2F formam grupos parálogos que surgiram da série estabelecida de duplicação do genoma completo (WGDs). Os NR2F6 são mais divergentes em gnatostomados, em contraste com os outros NR2F que foram perdidos independentemente em várias linhagens de vertebrados (Coppola and Waxman, 2021).

3A. Receptor Nuclear da subfamília 3 grupo A (NR3A)

A subfamília NR3A é classificada em ESR1 e ESR2. Esses NRs parecem mediar a ação do estrogênio no desenvolvimento reprodutivo, câncer de mama, ossos. Ainda segundo dados, o Polimorfismos de nucleotídeo único (SNPs) nos genes ESR1/ ESR2 tem ação direta na perda progressiva de massa óssea, tornando os ossos enfraquecidos e predispostos a fraturas, doença denominada osteoporose (Hongyue Ren,2022; Hu et al., 2020; de Mattos et al., 2014; Zhou et al., 2021).

3B. Receptor Nuclear da subfamília 3 grupo B (NR3B)

A subfamília NR3B, possui três subtipos: ESRRRA, ESRRB e ESRRG. Embora os NR3B compartilhem semelhanças de sequência com os receptores de estrogênio, eles não possuem ligantes específicos. Os NR3B parecem estar envolvidos na regulação da homeostase hepática. No entanto, a expressão do ESRRRA é ampla, mais abundante do que os outros dois membros do grupo, encontrada em tecidos metabolicamente ativos, como coração, rins, intestino, músculo esquelético e fígado. Geralmente (Xia, Dufour and Giguère, 2019).

3C. Receptor Nuclear da subfamília 3 grupo C (NR3C)

A subfamília é composta por 4 membros: NR3C1, NR3C2, PGR(NR3C3) e AR (NR3C4). Caracteriza-se por serem agonistas endógenos (molécula que pode se ligar e ativar um receptor para induzir uma reação biológica), divididos em 3-hidroxiesteróides (estrone e 17 β - estradiol) e 3-cetosteróides (diidrotestosterona, aldosterona, cortisol, corticosterona, progesterona e testosterona) (Cain, et al., 2019). Evolutivamente, esses receptores possivelmente surgiram de um ancestral de lampreia e hagfish, o que sugere que esses receptores estavam presentes no táxon que evoluiu na base da linhagem dos vertebrados (Baker and Katsu, 2017; Katsu, Oka and Baker, 2017).

O gene NR3C1 codifica o receptor de glucocorticoide humano fator de transcrição dependente do ligante e ativa a transcrição de genes responsivos a glicocorticóides. Esse NR está relacionado com a homeostase basal e o estresse (Palma-Gudiel *et al.*, 2015). O NR3C2 é o receptor de mineralocorticóides, está associado ao estresse, sintomas depressivos elevados, e redução da indução de cortisol (Cui, Xu and Jiang, 2019). O receptor de progesterona (PGR) está envolvido com ovulação, desenvolvimento da mama, estabelecimento da gravidez em mamíferos (Zhu et al., 2015). O receptor de andrógeno (testosterona e diidrotestosterona (DHT)). é membro da família de receptores nucleares do hormônio esteroide e um fator de transcrição nuclear dependente de ligante. Dada a sua ampla expressão em muitas células e tecidos, está relacionada com as ações biológicas, o que inclui, desenvolvimento, manutenção dos sistemas reprodutivos, musculoesqueléticos, cardiovasculares, imunológicos, neurais e hematopoiéticos (Davey and Grossmann, 2016).

4A. Receptor Nuclear da subfamília 4 grupo A (NR4A)

A família do receptor nuclear NR4A possui três membros: NR4A1, NR4A2 e NR4A3. Estão envolvidos inúmeros processos biológicos, tais como: apoptose e metabolismo, diferenciação neuronal e com doenças neurodegenerativas (Kon *et al.*, 2007).

Esses NRs órfãos constituem um grupo evolucionário antigo e altamente conservado de fatores de transcrição. Em contraste com outros NRs dessa superfamília, os receptores NR4A funcionam como fatores de transcrição independentes de ligantes e genes de resposta imediata ou precoce, que são rapidamente induzidos por uma pleiotropia de estímulos ambientais (Zhao and Bruemmer, 2010).

A. Receptor Nuclear da subfamília 5 grupo A (NR5A)

A subfamília NR5A, em mamíferos, contém dois membros (NR5A1 E NR5A2), estes estão envolvidos em diversas funções de desenvolvimento cartilagem e digestório). Estudos sugerem, ainda, que o NR5A2 está associado a cânceres de mama, cólon, ovário e pâncreas (Yang *et al.*, 2020; Hale *et al.*, 2014). O ligante específico desse grupo é desconhecido, comumente conhecido como receptor órfão. Ademais, os parálogos, possivelmente, surgiram durante eventos de duplicação do genoma inteiro, durante o desenvolvimento dos vertebrados (Kuo *et al.*, 2005).

6A. Receptor Nuclear da subfamília 6 grupo A (NR6A)

O gene NR6A1 nucleares inclui um grupo de fatores de transcrição dependentes de ligantes que se ligam a esteroides e outras moléculas lipofílicas. vários membros desta superfamília são receptores órfãos para os quais ainda não foram identificados ligantes. Seu padrão de expressão pode estar envolvido na neurogênese e no desenvolvimento de células germinativas. O splicing alternativo resulta em várias variantes de transcrição. Doenças associadas ao NR6A1 incluem Carcinoma Embrionário e Teratocarcinoma (Zechel, 2005; Lan *et al.*, 2003).

O NR6A1 é importante para a sobrevivência embrionária de anfíbios e mamíferos. Em *Xenopus* a expressão de uma forma dominante-negativa desse NR, resulta em malformações da cabeça, olhos, orelhas, glândula de cimento e somitos, resultando em letalidade embrionária (Chung *et al.*, 2001).

OBJETIVO GERAL

Descrever o cenário evolutivo da família gênica dos Receptores Nucleares (NRs), analisando o padrão de perdas e ganhos de membros dessa família nos genomas de 289 cordados, além do padrão de evolução molecular.

OBJETIVOS ESPECÍFICOS

- Quantificar a dinâmica de ganhos e perdas genéticas da Família NRs nos 289 cordados; que possuem os genomas completos;
- Determinar se os NRs dos cordados compartilham genes das mesmas famílias como vizinhos (sintenia), para auxiliar o estabelecimento das origens dos NRs;
- Comparar as taxas evolutivas do domínio DBD dos genes NRs;
- Construir árvores gênicas dos DBD dos NRs, para contribuir com a determinação quais genes seriam mais antigos e os que surgiram mais recentemente por duplicação

METODOLOGIA

COLETA DE DADOS

As sequências dos genes da família NRs em Cordados, foram consultadas nos bancos de dados públicos disponíveis: (<http://www.ensembl.org/>); UCSC (<http://genome.ucsc.edu>) e NCBI Genbank (<http://www.ncbi.nlm.nih.gov/genbank/>), Uniprot (<https://www.uniprot.org/>). Adicionamos a ferramenta Blast/Blat para recuperar as sequências adicionais que não estão anotadas e nem foram recuperadas pelo serviço de busca de dados BioMart, utilizando a sequência humana como referência.

Para verificar presença e ausência dos genes da NRs em cordados, utilizamos a ferramenta de mineração de dados BioMart (<http://www.ensembl.org/biomart/martview/>), com parâmetros restritos a 05(cinco) domínios proteicos (IPR035500/NHR-like_dom_sf, IPR000536/Nucl_hrmn_rcpt_lig-bd, IPR001723/Nuclear_hrmn_rcpt, IPR001628/Znf_hrmn_rcpt, IPR013088/Znf_NHR/GATA), estas são sequências proteicas conservadas são essenciais no estudo de uma investigação evolutiva. Complementar a essa atividade, adicionamos a busca por serviço da web de modelos ocultos de Markov, o HMMER (<http://hmmer.org/>), recurso que detecta homólogos e auxilia nas análises das estruturas proteínas.

Buscamos a sintenia e vizinhança compartilhada dos genes da família NRs nos cromossomos dos cordados (inicialmente os 49 humanos, mas genes de linhagens específicas podem ser comparados), com uma distância de 15 genes franqueadores dos dois lados, utilizando o browser Genomicus (93V.). Este browser, além de mostrar visualmente a ordem e a presença de genes compartilhados entre espécies distintas (ortólogos), também pode ser utilizado para comparar a sintenia dentro de uma mesma espécie (parálogos). Adicionalmente, pesquisas manuais nos navegadores do genoma Ensembl e UCSC com o mesmo propósito. Resgatamos dados quanto a arquitetura das proteínas, quanto a presença dos domínios de interação, regiões desordenadas e de baixa complexidade nos browsers UniProt (<http://www.uniprot.org/>) e Pfam 30 (<http://pfam.xfam.org/>).

ANÁLISE DE DADOS

Avaliamos a presença em comum dos genes NRs, em cada uma das espécies dos cordados, nos grandes clados dos cordados, para estabelecer um padrão de perdas e ganhos (e.g. quais aparecem somente em mamíferos).

A conservação da sintenia nos ortólogos e entre os parálogos de uma espécie foram analisadas de forma descritiva (presença/ausência) salientando famílias gênicas que são encontradas conjuntamente com os NRs para auxiliar a construção do cenário evolutivo (a conservação da sintenia/vizinhança é uma evidência de homologia, mesmo quando as sequências não são conservadas)

As sequências nucleotídicas e protéicas para as espécies de vertebrados obtidas através dos bancos de dados para os genes NRs serão alinhadas no server GUIDANCE2 (<http://guidance.tau.ac.il/ver2/>), utilizando o algoritmo MAFFT.

O Pacote PAML 4.9a será utilizado para verificar o padrão evolutivo dos genes da família NRs (somente serão analisadas sequências completas, no caso de vários transcritos para o mesmo gene será escolhido o mais longo). Assim, para testar um conjunto de espécies será utilizado a variável NsSites, que permite a variação da taxa ω entre os sítios (entre os códons e aminoácidos da proteína) a fim de verificar quais códons possuem evidências de seleção positiva.

Será estimado a taxa de substituições sinônimas e não-sinônimas (dN/dS), também conhecida como ω , através dos métodos de Nei e Gojobori (1986), e de Yang e Nielsen

(2000). Valores das taxas de substituição $\omega < 1$ e $\omega > 1$ indicam respectivamente seleção purificadora (ou negativa), ou positiva (Darwiniana). Já valores de $\omega = 1$ estariam indicariam evolução neutra.

Os modelos de substituição de códons a serem testados serão os seguintes s: M0 (one ratio), onde um único valor de ω é assumido em todos os códons, já os modelos M1a (neutral) admite diferentes valores de ω , mas nunca maiores do que um, e M2a (selection) admite valores de $\omega > 1$; M8a (beta+ $\omega = 1$) também admite diferentes valores de ω , mas nunca maiores do que um, mas diferentemente de M1a, considera 11 parâmetros; e finalmente M8 (beta & ω) que permite valores de $\omega > 1$. Serão feitas comparações por análises de Máxima Verossimilhança, entre os modelos M1a e M2a, e M8a e M8 para determinar qual melhor modelo de substituição de códons adequa-se aos dados. Uma abordagem bayesiana (BEB) será aplicada para estimar a significância entre as diferenças de dN/dS (Anisimova e cols, 2007; Stern e cols, 2007; Yang, 2006).

A árvore gênica por Máxima Verossimilhança, com 1000 Bootstraps, somente do domínio de ligação ao DNA (DBD) para os parálogos, foi construída no programa MEGA 10, o modelo evolutivo de acordo com os dados foi determinado pela função Model-test.

RESULTADOS E DISCUSSÃO

ORGANIZAÇÃO ESTRUTURAL DOS RECEPTORES NUCLEARES

A superfamília dos NRs possui 37 domínios proteicos identificados, destes, cinco são encontrados na maioria dos receptores, sendo parâmetro orientador na busca dos genes no BIOMART (Figura 06). Não foram encontrados domínio DBD dos parálogos gênicos (NR0B1 e NR0B2), resultado corroborado pelos estudos de (Zhao *et al.*, 2015).

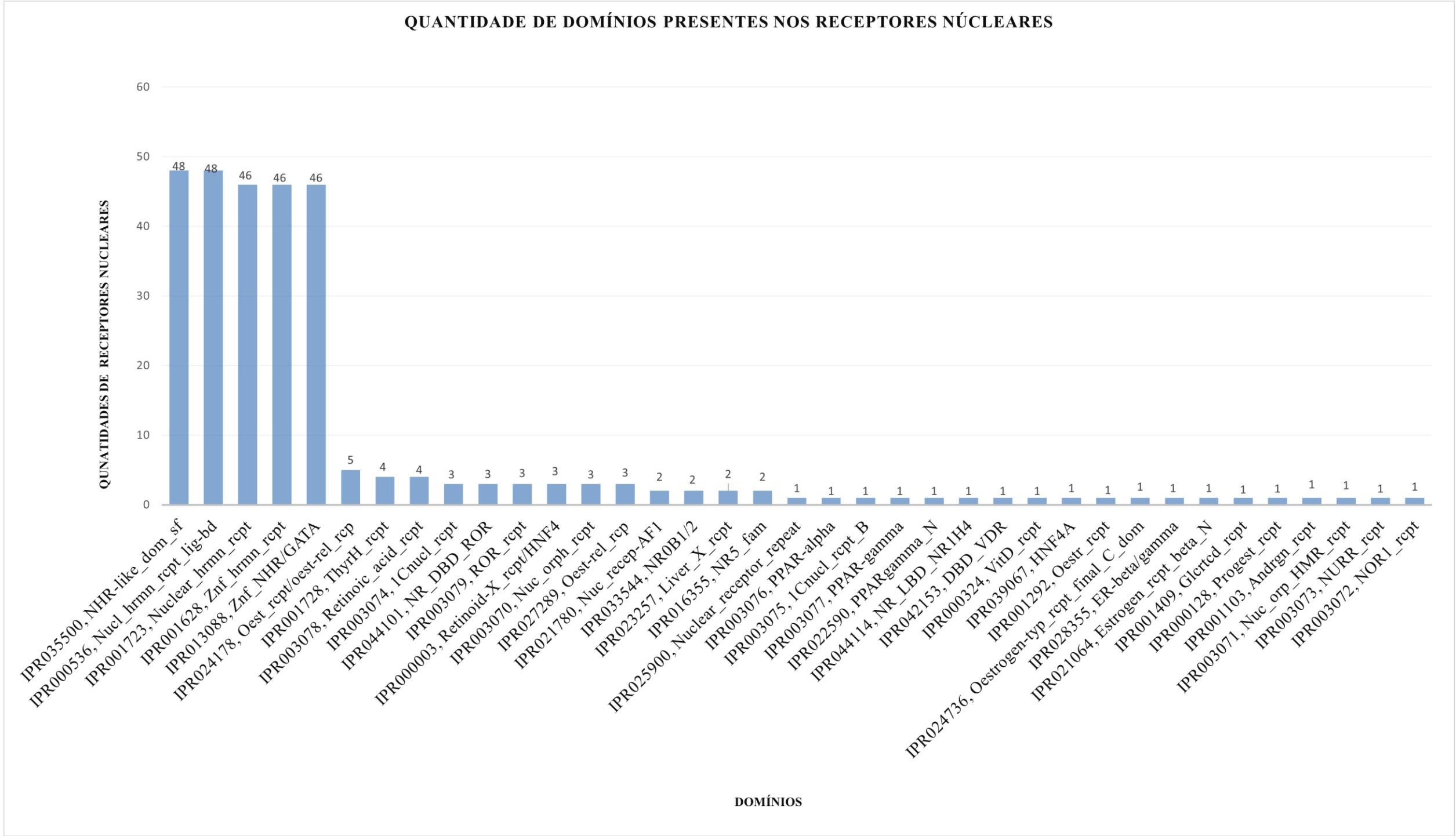
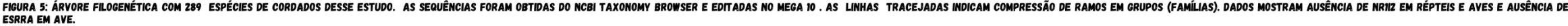


Figura 5. Quantidade de domínios presentes em cada receptor nuclear. Os primeiros receptores estão presentes na maioria dos RNs.

DINÂMICA DE GANHOS E PERDAS

Caracterizamos a família dos genes dos Receptores Nucleares procurando pelas sequências disponíveis nos bancos de dados de genoma animal e nos banco de dados proteicos, incluindo 289 cordados, sendo compostos por: 3 cefalocordados (*Branchiostoma Belcheri*, *Branchiostoma Floridae* e *Branchiostoma lanceolatum*); 2 Urocordados (*Ciona intestinalis* e *Phallusia mammillata*); 4 vertebrados sem mandíbula (Agnatha - *Eptatretus burgeri*, *Lethenteron camtschaticum*, *Myxine glutinosa* e *Petromyzon marinus*); e sendo 280 vertebrados com mandíbulas (Gnathostomata) Veja mais detalhes na figura 05.

-----	Indicam compressão de ramos em grupos (famílias).
	Ausência do gene na espécie e/ou grupo.
	Ausência do ESRRA em aves.
	Ausência do NR12 em aves e répteis.



CORDADOS BASAIS

Recuperamos nas buscas nos bancos genômicos, tanto por identificação de domínios quanto pelas sequências, para os três cefalocordados o número máximo de 22 NRs, Lecroisey et al., (2012), em sua revisão descreveu que o genoma dos anfioxos codificava 33 NRs, com pelo menos um representante de todos os principais grupos NRs, exceto NR1E e NR1I/J. Os autores sugeriram que o número elevado de NRs, esses animais não passaram pelo 2RWGD, significava que os anfioxos sofreram algumas modificações secundárias, evidenciadas no grupo NR1H, visto que os anfioxos teriam 11 NR1H, enquanto humanos possuem 4 genes NR1H funcionais e um pseudogene (ver abaixo), ou autores sugerem que esse fato o pode um ser indício de uma possível diversidade funcional (Lecroisey, Laudet and Schubert, 2012).

Os genes do grupo NR1H, tem como ligantes oxisteróis (derivados de colesterol) e metabólitos do farnesol, hormônio juvenil III, farnesal, acetato de farnesila, ácido farnesóico ou geranilgeraniol. Os NR1Hs precisam formar heterodímeros com o RXRs e os ligantes para exercerem suas funções de fatores de transcrição, encontramos pelo menos 3 tipos de NR1Hs e um RXR nos cefalocordados. Em nossa busca encontramos um número reduzido de receptores nucleares esteroides (ESR1 e ESR2; GR (NR3C1) e MCR (NR3C2)). Lecroisey et al., (2012) discute que embora os cefalocordados possuam os receptores esteroides, os ensaios moleculares revelaram que os receptores ERs são incapazes de se ligarem a qualquer um dos hormônios (e.g. ER, AR e PGR) dos vertebrados. Esses dados sugerem que os receptores esteroides dos cefalocordados se comportam como receptores órfãos, sem um ligante específico. O que reforçaria a hipótese do surgimento desses receptores antes dos seus ligantes específicos que apareceriam somente na linhagem dos vertebrados (Baker, 2019).

Os dados analisados mostram que os urocordados (*Ciona intestinalis*) possuem 17 NRs, sendo que esses genes representavam quase todos os grupos, estando ausente apenas o grupo NR2E, de uma subfamília de receptores nucleares órfãos, expressos em estruturas derivadas do olho e do prosencéfalo. Miglioli et al., (2021) e Lecroisey et al., (2012) referem-se o trabalho de Dehal et al., (2002) que identificou 17 NRs nos Urocordados, e que estes animais teriam um número reduzido de NRs pois não passaram pelo evento de duplicação

2RWGD. Entretanto, isso levanta uma questão, os Urocordados passaram pelo evento 2RWGD? Se levarmos em conta somente os NRs a resposta seria não, e outras famílias de fatores de transcrição (PAX e HOX) sugerem que esses animais divergiram dos demais cordados antes do 2RWGD. Todavia, como já mencionado o genoma desses animais parecem estar evoluindo rapidamente, com duplicação independente de vários fatores de transcrição (por exemplo, Pax2/5/8 no apendicular *Oikopleura dioica*) (Holland, 2013).

AGNATOS/VERTEBRADOS SEM MANDÍBULA

Os vertebrados sem mandíbula foram representados em nosso estudo por 2 espécie de peixe-bruxa (*Eptatretus burger* e *Myxine glutinosa*) e 2 espécies de lampreias

(*Lethenteron camtschaticum* e *Petromyzon marinus*). Juntos, eles formam um grupo irmão dos vertebrados gnatostomados, sendo considerados em muitos estudos bons modelos para investigar as características ancestrais dos vertebrados. Encontramos, ao todo, 39 genes NRs nos agnatas, todos os grupos estão representados com pelo menos uma cópia. Vale destacar, que encontramos 2 ohnologos do grupo NR3C (NR3C1 e NR3C2, os receptores glicocorticóides e mineralocorticoide, respectivamente), enquanto a 3 cópia o receptor do andrógeno (AR/ NR3C4) não foi encontrado em nenhum dos Agnatas. Baker (2019) afirma que o surgimento do receptor do andrógeno (AR/ NR3C4) acontece já nas linhagens dos gnatostomados, mais precisamente nos condrictes (peixes cartilaginosos). A evolução do dos NR3C1 e NR3C2 em um vertebrado sem mandíbula através da duplicação de um receptor esteroide ancestral adicionou aumentou o repertório de ligantes fisiológicos nestes animais, proporcionando regulação de vias fisiológicas complexas, incluindo reprodução,

respostas imunes, homeostase eletrolítica e resposta ao estresse respostas, em vertebrados sem mandíbula e posteriormente em toda a linhagem dos vertebrados ((Donoghue and Purnell, 2009; Donoghue and Keating, 2014).

No *Petromyzon marinus* foram encontrados 32 NRs, enquanto nos Hagfish encontramos 26 NRs. Essa diferença pode ser em consequência da cobertura e do estágio de dos genomas destes organismos. Estudos recentes (Cardoso, Bergqvist and Larhammar, 2020; Ermakova *et al.*, 2020) reforçam a hipótese que o 2RWGD ocorreu antes do split entre agnatos e gnatostomados, as diferenças entre o número de genes entre esses dois grupos devem ser consequência de perdas e expansões gênicas específica de cada linhagem.

GNATOSTOMADOS/ VERTEBRADOS COM MANDÍBULAS

Nessa superclasse foram estudadas 289 espécies (tabela suplementar 1). Alguns estudos prévios indicavam que a composição da superfamília gênica era variável entre os cordados, mas nenhum estudo analisou um número tão grande de vertebrados (Lecroisey, Laudet and Schubert, 2012).

Nos Chondrichthyes, táxon mais basal deste grupo, recuperamos sequencias de 3 espécies (*Amblyraja radiata*, *Callorhinchus milii* e *Scyliorhinus canicula*), encontramos no total 44 NRs. Não recuperamos em nenhuma das espécies os genes ESRRa, ESRRb, Nr1i3 e RORc. Fonseca et al., (2020) estudou os genomas e transcriptomas disponíveis de cinco espécies de (*Chiloscyllium punctatum*, *Scyliorhinus canicula*, *S. torazame*, *Leucoraja erinacea* e *Callorhinchus milii*); e sequenciou o genoma de um pequeno tubarão (*Hydrolagus affinis*), neste estudo eles recuperaram um total de 52 NRs, inclusive três genes não recuperados na nossa busca, possuindo mais uma cópia nos grupos NR1H (LXR/FXR), NR5A (SF-1) e duas cópias a mais dos NR0B (DAX1/SHP). Entretanto, tanto a nossa busca quanto a de Fonseca et al (2020) não recuperou o gene NR1I3 (CAR). Os Chondrichthyes são um grupo que oferecem uma oportunidade excepcional para abordar a diversificação precoce das famílias de genes NR (Fonseca et al., 2020).

Os 86 peixes ósseos (Actinopterygii) estudados por nós, apresentam um número variável de ohnologos, como esperado, pois, parte desta linhagem (teleósteos) passou por mais um evento de duplicação genômica (3RWGD). Fonseca (2020) relata um número máximo de 76 NRs no peixe paulistinha (*Danio rerio*) outros estudos encontraram 73 NRs (Zhao et al., 2015). Enquanto na linhagem dos Lepisosteiformes (peixes lagarto /spotted gar), que não passou pelo 3RWGD (Pasquier et al., 2017) foram encontrados 52 NRs (Fonseca, 2020). Interessantemente o gene NR1H2 (*Oxysterols Receptor LXR-Beta*) está ausente na maioria dos peixes, é encontrado somente em duas espécies, especificamente, nas espécies *Amphilophus citrinellus* e *Erpetoichthys calabaricus*, nas famílias Polypteridae e Cichlidae, respectivamente.

O *Latimeria chalumnae* (celacanto), integrante do grupo Sarcopterygii, peixe com nadadeira lobada, que apresenta todos os 49 NRs, bem como 7 NRs adicionais, possuindo mais uma cópia nos grupos NR5A (SF-1), NR0B (DAX1/SHP) semelhantes aos condrictes, também cópias dos grupos NR1D, NR2F e NR1F (ROR), além de mais duas duplicatas dos NR2A (HNF4) (Tabela suplementar 2 - Mart-export Celacanto 2).

Nas 13 espécies de anfíbios estudadas (tabela suplementar 1) recuperamos, no conjunto, todos os 49 NRs, Fonseca et al., (2020) sugere que estes animais teriam 52 NRs, possuindo mais uma cópia nos grupos NR3B (ERR), e NR1F(ROR), NR2F e NR2A (HNF4), similares ao Celacanto. Em contraponto aos resultados de estudos anteriores, nossos dados ainda demonstram a presença de NR1H2 neste grupo, gene ausente no resultado de Zhao et al., (2015), ainda (Fonseca et al., 2020). encontrou três ohnologos para o grupo NR1H, enquanto nosso trabalho recuperou 4 (Nr1h2, Nr1h3, Nr1h4 e Nr1h5 para *X. tropicalis* (NM_001090085.2)).

Analizamos 39 espécies de répteis (tabela suplementar 1), recuperamos, no conjunto, todos os 49 NRs, Fonseca (2020) sugere que estes animais teriam 50 NRs, com uma cópia a mais nos grupos NR1F(ROR) e NR2B (HNF4) e que no grupo NR2B (RXR) teriam duas duplicatas, enquanto encontramos três ohnologos para este grupo (RXRA, RXRB e RXRG). Não recuperamos, em nenhum dos 39 répteis estudados, o receptor NR1I2 (PXR). Zhao et al., (2015), que indicou a ausência de NR1I2 (PXR) e NR1H2 (LXR), nos testudines, e Fonseca et al., (2020) recuperou somente dois dos genes do grupo NR1I, no lagarto verde (*Anolis carolinensis*).

Estudamos 30 espécies de aves, e recuperamos 46 NRs, no conjunto de espécies da busca pelos 49 NRs. As aves não possuem os receptores ESRRA (NR3B1) e o NR1H2 (LXR) e o NR1I2 (PXR), mesmo grupo de receptores que estão listados como faltantes em Fonseca et al., (2020) e o NR1H2 em Zhao et al., (2015). Este autor descreve que as aves teriam 48 NRs, com uma cópia a mais nos grupos NR1F(ROR), NR2B (HNF4) e NR2F. Além disso, o autor identifica que no grupo NR2B (RXR) teriam duas duplicatas, enquanto encontramos três ohnologos para este grupo, na maioria das aves, e ainda uma possível quarta duplicata, na galinha (ENSGALG00000003759(RORA); ENSGALG00000015150 (RORB); ENSGALG000000033941(RORC); ENSGALG00000001035).

Interessantemente, aves mesmo com a perda de alguns NRs, também parece ter duplicações específicas em sua linhagem. Alguns estudos que comparam o tamanho do genoma de famílias de genes entre as aves e mamíferos, encontraram um número menor de genes nas aves. Um exemplo clássico é que as aves perderam vários genes MHC e imunoglobulinas em contraste com a maioria dos outros vertebrados. Rodrigues Paixão-Côrtes et al., (2013) descreve a perda de dois genes da família PAX (PAX4 e PAX8), alguns autores sugerem que essa redução no número de cópia de genes estaria correlacionada com a evolução do voo (Hughes & Friedman, 2008; Hughes, et al., 2008), mas não encontramos

esse padrão de redução nos NRs, existem perdas de duplicadas, mas também existem retenção de cópias e possivelmente uma duplicação específica.

Analisamos 106 espécies de mamíferos (tabela suplementar 1), recuperamos todos os 49 NRs neste grupo. Entretanto, podemos observar a pseudogenização de um dos NRs em uma linhagem específica. O genoma humano possui 49 NRs, 48 funcionais, e um pseudogene o NR1H5P (Nuclear Receptor Subfamily 1 Group H Member 5, Pseudogene), também conhecido como FXRB. Essa pseudogenização parece estar acontecendo na linhagem dos primatas. Recuperamos a sequência identificada como pseudogene na maioria dos outros primatas estudados, com exceção do bushbaby próximo (*Otolemur garnettii*; ENSOGAG00000030317). Entretanto a sua funcionalidade ainda precisa ser determinada.

NR1H5 é um gene codificador em outros genomas de vertebrados. Especificamente na linhagem dos mamíferos, recuperamos receptores NR1H5 previstos como funcionais: para 19 roedores, 11 carnívoros e 12 em outros mamíferos placentários. Observamos a presença de Nr1h5 em Lampreia (Petromyzontida) e no Lancelet (Cefalocordada), o que confirma a presença e a possível funcionalidade deste receptor na evolução dos cordados.

Fonseca et al., (2020) sugere que o ancestral dos vertebrados com mandíbula possuía ao menos 58 NRs, encontramos nos gnatostomados em torno de 52 - 54 genes, nos condrites (tubarões e arrais) e nos peixes da classe holostei (peixe-lagarto), e as demais linhagens de vertebrados passaram por duplicações e perdas nas linhagens específicas.

É importante ressaltar que os genomas dos peixes cartilaginosos e dos holostei, não passaram pelo 3RWGD e evoluíram a uma taxa mais lenta em comparação com os peixes teleósteos, compartilhando muitas semelhanças com os genomas dos mamíferos. Assim, os condrites e os holostei são fundamentais para entender a evolução das características e inovações dos vertebrados (Fonseca, 2020; Braasch, 2026). O trabalho de Zhao et al., (2015), sugere que o aparecimento de NRs, tais como NR1I2/NR1I3 (PXR/CAR), devido aos eventos de duplicação, colaboraram para a estabelecer a diversidades de funções dos NRs.

SINTENIA CONSERVADA E RECEPTORES NUCLEARES EM TANDEM

Fizemos uma análise de sintenia compartilhada (genes no mesmo cromossomo) e/ou vizinhança conservada (genes lado a lado na mesma ordem) para toda superfamília dos NRs. A sintenia é conservada entre os ortólogos, principalmente em mamíferos, mas de um modo

geral os tetrápodes têm as suas vizinhanças conservadas. É difícil recuperar as sentenias em genomas com cromossomos ou contigs pequenos. Como é o caso do Celacanto e da Lampreia.

A sentenia entre tetrápodes e teleósteos em muitos casos é perdida em consequência do 3RWGD, que leva a uma perda assimétrica de genes e elementos regulatórios. O gene NR1H5 (que é um pseudogene na linhagem dos primatas), possui uma sentenia relativamente conservada, o Celacanto tem um contig com 21 genes, destes recuperamos na mesma sentenia em média 12-13 genes nos tetrápodes, 11 no peixe-lagarto (Holostei) e cerca de seis nos teleósteos (peixe-paulistinha).

A conservação sentenia entre os parálogos (ohnologos) é um evento mais raro no genoma dos vertebrados, podemos observar nos blocos dos genes HOX, onde houve primeiramente uma duplicação em tandem, com compartilhamento de elementos regulatórios pelas duplicatas, e posteriormente cópias desse cromossomo foram geradas por duplicações do genoma completo. Outro exemplo, de sentenia notável e/ou vizinhança conservada foram identificados entre os parálogos dos vertebrados da família PAX, especificamente PAX1 e PAX9 compartilham uma vizinhança gênica de 590 - 896 MYA, inicialmente encontrados PAX1/9-SLC25A, no ouriço-do-mar (*Strongylocentrotus purpuratus*), no anfioxo (cefalocordado) encontramos PAX1/9-SLC25A-FOX-MIPOL e nos vertebrados NKX2.1-NKX2.8-PAX9-SLC25A21-FOXA1-MIPOL e NKX2-4- NKX2-2 -PAX1-FOXA2 (Santagati, 2003; Paixão-Côrtes, 2013; Paixão-Côrtes, 2015).

Os NRs tiveram seus números ampliados no genoma pelas sucessivas duplicações no genoma, ocorreram interessantes eventos de duplicação em tandem, com o evento de duplicação dos THR-RAR-NR1D, todos fazem parte do grupo NR1, mas análises filogenéticas agrupam THR-RARs em um ramo, enquanto os NR1Ds são colocados em um ramo com os PPARs. Na nossa análise sintênica (Figura 7) verificamos que essa duplicação deve ter ocorrido antes da separação dos cefalocordados dos demais cordados (588 MYA), pois o anfioxo possui em um cromossomo THRA-NRD1 e em outro THRB-RARB, nos agnatos onde encontramos contigs pequenos e conseguimos recuperar em sentenia THRB-RARB (peixe-bruxa). Enquanto nos vertebrados com mandíbula recuperamos THRA-NR1D1-RARA e NRD2-THRB-RARB. Outro achado interessante é que entre esses dois cromossomos duplicados também encontramos em sentenia conservada mais dois genes: TOP2A e B (DNA topoisomerase II) e LRRC3C e B (leucine rich repeat containing 3). Não

encontramos a sintenia dos genes em tandem nos Urocordados, Holland (2013) afirma que existe pouca ou nenhuma sintenia entre genomas de tunicados e vertebrados. Esse dado que confirma uma duplicação ancestral em tandem se contrapõe às filogenias dos NRs, onde o foco é as alterações nas sequências, que estão sujeitas as pressões seletivas, principalmente a afinidade com os ligantes alvo.

RARG e VDR (ambos do grupo NR1), também se localizam próximos um do outro no cromossomo nos vertebrados, mas não conseguimos recuperar essa conservação entre outros organismos.

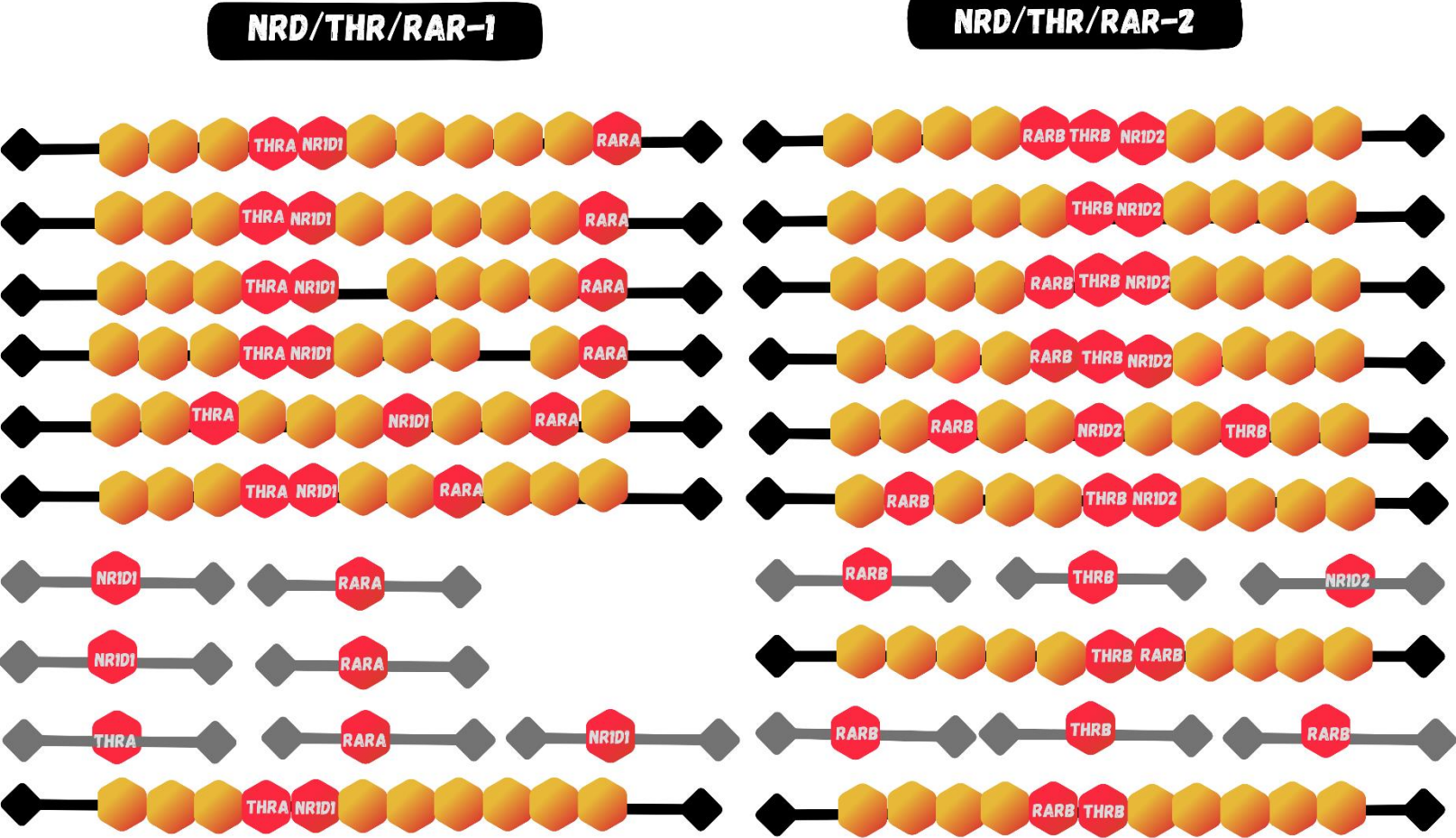
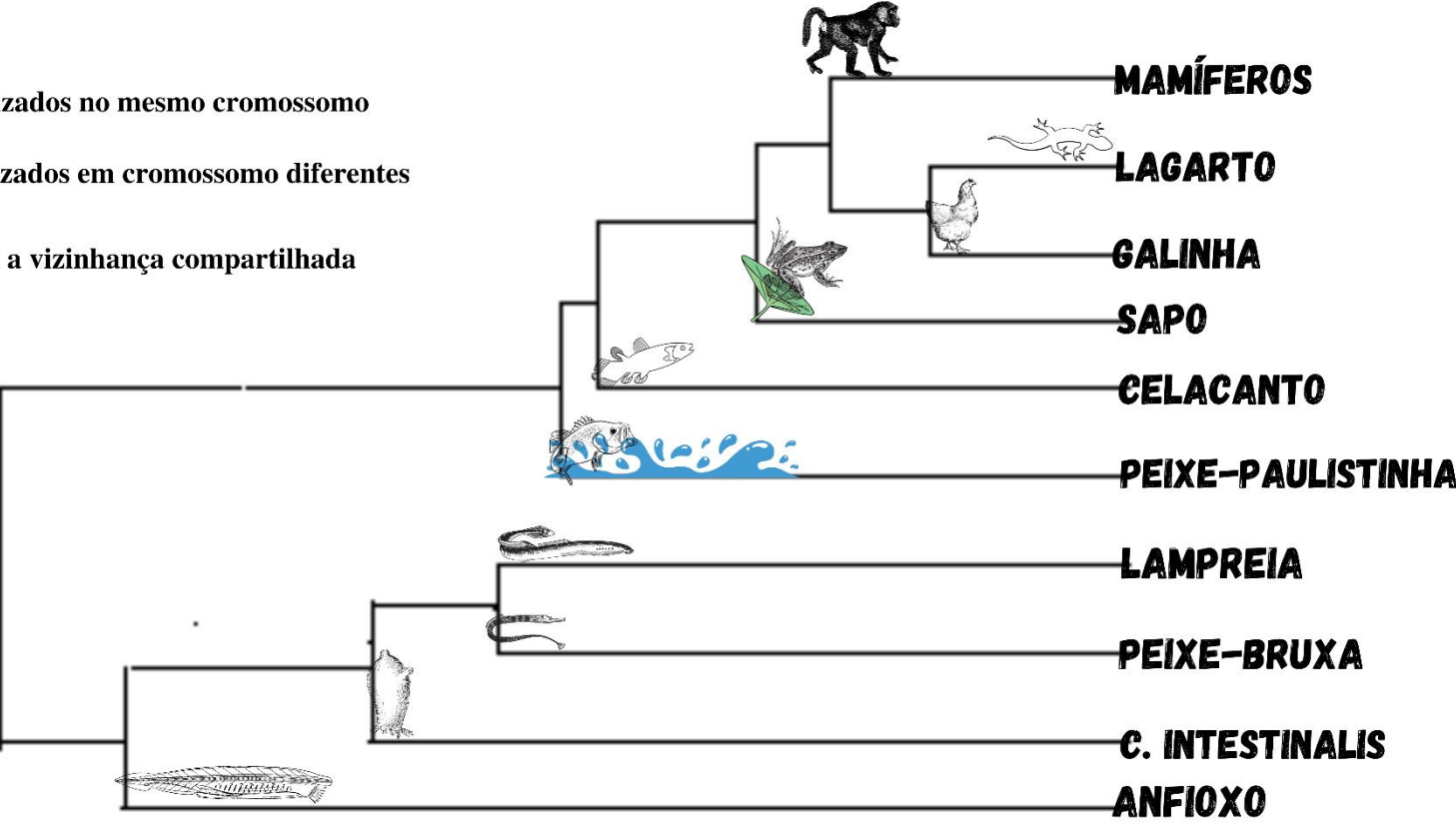
NR5A1 (SF-1) e NR6A1 (GCNF) também parecem ser cópias em tandem, visto que aparecem lado a lado NR5A1-NR6A2, na maioria dos tetrápodes, inclusive no celacanto, e dos teleósteos, mas não conseguimos recuperar nos condrictes, pois o gene NR6A1 estava na extremidade do cromossomo. Entretanto, recuperamos duas cópias em tandem no anfioxo (cefalocordado), não anotadas, bem como também duas cópias no ouriço-do-mar (*Strongylocentrotus purpuratus*) o que pode significar que essa duplicação em tandem é anterior a separação dos equinodermos dos cordados. Se confirmarmos essa duplicação ancestral este dado, também se contrapõe as filogenias dos NRs complicando muito o cenário evolutivo.

A análise de sintenia também foi usada como evidência adicional para a ausência de genes ou duplicação em alguns táxons, conforme descrito na seção anterior. Paixão-Côrtes et al., (2013) usaram os dados de sintenia para sustentar a hipótese da perda dos genes PAX, ocorreu várias vezes em linhagens de vertebrados terrestres

Figura. 7 : Análise de sintenia de RNs em tandem.

Legenda

- Genes localizados no mesmo cromossomo
- Genes localizados em cromossomo diferentes
- Genes e a vizinhança compartilhada



ANÁLISE DAS TAXAS EVOLUTIVAS

A análise superfamília dos NRs, foi realizada utilizando-se 60 espécies de cordados, nos 47 NRS que possuíam o domínio DBD (lista de espécies na tabela suplementar 3), para determinarmos o padrão evolutivo, da região que é compartilhada por todos os NRs, com exceção dos dois genes do grupo NR0B, aplicamos um teste de máxima verossimilhança, para identificar que o melhor modelo de substituição de códons, procurando detectar sinais seleção positiva na sequencias dos diferentes grupos de cordados. A fim de estabelecermos o panorama evolutivo dos NRs, calculamos a taxa de substituições sinônimas e não-sinônimas (dN/DS ou Ômega (ω)). A árvore utilizada nos testes foi construída baseada nas relações filogenéticas estabelecidas em diversos artigos e concatenada pelo servidor Timetree (<http://www.timetree.org/>).

Quando os modelos (M1a/neutro – M2a/seleção e M8a/neutro -M8/seleção) são comparados por máxima verossimilhança (LRT), na maioria dos NRs o modelo mais adequado foi o M1a, que não admite seleção, e os valores de ω , não ultrapassam 1 (Figura 7 e tabela suplementar 4. Somente em dois NRs, NR4A1 (GFRP1; $p=0.04967$) e o RORC (NR1F3; $p=0,00036$) o melhor modelo foi o M8, que é o modelo que admite seleção positiva ($\omega>1$) e variação da pressão seletiva em até 11 classes de ω . Ou seja, indica que embora a pressão seletiva possa variar entre os sítios de aminoácidos. Entretanto, a maior força operando nos DBDs dos NRs é uma onipresente seleção purificadora.

No NR4A1 e no RORC a análise bayesiana (BEB) não detectou nenhum sítio, com a probabilidade à priori, sujeito à seleção positiva. Porém, podemos detectar que alguns sítios estão sofrendo um relaxamento da pressão seletiva. A taxa média de substituições não-sinônimas pela sinônimas (dN/dS), para os DBDs dos genes NR4A1 e RORC são 0.0368 e 0.0325 respectivamente, valores que indicam uma seleção purificadora. Enquanto a análise bayesiana detectou 10% de sítios que estariam sofrendo um relaxamento da pressão seletiva com valores de $\omega= 0,228$ (NR4A1) e $\omega= 0,2293$ RORC, com um relaxamento cerca de 10 vezes maior que a média. Interessantemente, um único sítio detectado possivelmente em seleção positiva $\omega= 1.873$, não é significativo, pois para o gene NR2C2 (TR4) o melhor modelo é o M1 ($p=0.56$)

No entanto, foi observado o relaxamento da restrição funcional em alguns sítios, pode ser fator importante que influencia na plasticidade fenotípica dos seres vivos. (Hunt et al.,

2011), podendo também ser associada ao surgimento de novidades adaptativas na trajetória evolutiva dos cordados (Paixão-Côrtes et al., 2013), visto que, pequenas alterações no DBD podem influenciar a ligação desses fatores de transcrição aos motivos de DNA, dos diversos genes regulados pelos NRs.

Gene	Modelo	dN/dS	Parâmetros estimados	ℓ	p-value
NR4A1	M1	0.0232	p: 0.98678 0.01322 w: 0.01015 1.00000	-3.237.316.304	p=0.9999
	M2	0.0232	p: 0.98678 0.00495 0.00827 w: 0.01015 1.00000 1.00000	-3.237.316.304	
	M8	0.0368	p: 0.10000 0.10000 0.10000 0.10000 0.10000 0.10000 0.10000 0.10000 0.10000 0.10000 0.00001 w: 0.00000 0.00000 0.00000 0.00001 0.00015 0.00101 0.00507 0.02047 0.07272 0.26822 1.00000	-3.212.215.479	p=0.04967077521356889
	M8a	0.0235	p: 0.09868 0.09868 0.09868 0.09868 0.09868 0.09868 0.09868 0.09868 0.09868 0.09868 0.01316 w: 0.00000 0.00000 0.00003 0.00016 0.00062 0.00183 0.00460 0.01057 0.02387 0.06351 1.00000	-3.210.289.210	
RORC	M1	0.0993	p: 0.90907 0.09093 w: 0.00920 1.00000	-4.046.058.610	p=0.9999
	M2	0.0993	p: 0.90907 0.03579 0.05514 w: 0.00920 1.00000 1.00000	-4.046.058.610	
	M8	0.0325	p: 0.10000 0.10000 0.10000 0.10000 0.10000 0.10000 0.10000 0.10000 0.10000 0.10000 0.00001 w: 0.00000 0.00000 0.00000 0.00005 0.00038 0.00187 0.00717 0.02320 0.06887 0.22293 1.00000	-3.840.493.939	p=0.0003604296367352156
	M8a	0.0331	p: 0.10000 0.10000 0.10000 0.10000 0.10000 0.10000 0.10000 0.10000 0.10000 0.10000 0.00001 w: 0.00000 0.00000 0.00000 0.00005 0.00037 0.00187 0.00722 0.02352 0.07018 0.22775 1.00000	-3.834.130.479	

Figura 7. Parâmetros estimados através dos modelos de substituição de códons.

ANÁLISE FILOGENÉTICA

As árvores filogenéticas do domínio DBD foram construídas no MEGA 10, com análise de 1000 bootstrap. Os Alinhamentos das múltiplas sequências foram realizados usando Muscle algoritmos com parâmetros padrão no MEGA 10 e do servidor Guidance 2.

Os NRs foram agrupados em 20 grupos, obedecendo as orientações do comitê de nomenclatura (Committee, 1999). Os alinhamentos foram realizados com base na referência do domínio DBD. Análises adicionais foram realizadas, baseado no alinhamento dos parálogos por cromossomo, necessário para compreender a dinâmica dos NRs que estão em tandem (Figura.8). Foram construídas 19 árvores filogenéticas, seguindo as orientações do sistema unificado de nomenclatura. Na Figura 8. temos como exemplo da relação filogenética do THRA_RARA_NR1D1, mostrando claramente a separação de cada gene, além de evidenciar a presença desses receptores em genomas de cordados basais (*Ciona intestinalis*, *Branchiostoma lanceolatum*, *Branchiostoma Belcheri*). Todas as nossas análises com os DBDs, inclusive com todas as 3000 sequencias dos DBDs (material suplementar XX), recuperaram os grupos independentes de genes, formando os clusters pela nomenclatura, não detectando sinais de conversão gênica. As únicas que sequências que agrupavam, muitas vezes fora dos seus grupos, eram as sequências dos cordados basais (Cefalocordados e Tunicados), pois eram as sequências mais divergentes e muitas vezes com problemas de anotação e a identificação da ortologia correta, além de algumas sequências mais curtas. O suporte dos ramos, com valores baixos, também não foram confiáveis nas nossas árvores. Provavelmente o DBD que é uma região conservada, que foi escolhida porque é o domínio compartilhado pela maioria dos NRs, não é suficientemente informativo para resolver as questões sobre a origem e expansão dos NRs.

A evolução dos receptores nucleares (NRs) começa na base da vida dos metazoários. E aparece primeiramente nos invertebrados, Porífera (esponjas) é o primeiro clado animal onde surgiram NRs funcionais, dois membros da família NR2 (HNF4 -*Amphimedon queenslandica*; RXR - *Suberites domuncula*). Os placozoa (*Trichoplax*), parecem ter, além dos dois NR2Fs o receptor relacionado ao estrogênio (ERR - NR3B), além do próprio receptor do estrogênio (ER - NR3A)(Posnien *et al.*, 2021). Entretanto, é controversa a questão se os NRs de invertebrados realmente e ligam aos hormônios esteroides (Baker, 2019; Miglioli *et al.*, 2021)

A evolução de um ER de ligação ao ligante é proposta antes da divisão de protostomia e deuterostomia, portanto, está localizada na linhagem cnidária ou xenacoelomorpha (estrela roxa). A sinalização de estrogênio totalmente descrita e funcional existe em cefalocordados e vertebrados (Posnien *et al.*, 2021).

A partir destes dados, existem dois cenários diferentes de diversificação de NRs: **1.** baseada em análises filogenéticas, pressupõe que o NR ancestral funcionava como um fator de transcrição constitutivo, sem ligar um ligante, e as duplicatas geradas por WGDs, ou duplicação segmentar (SD) ou duplicação em tandem (TD), adquiriram a capacidade de se ligar aos ligantes de forma secundária e independente, em diferentes momentos da evolução; **2.** O cenário alternativo sugere que os antigos NRs podem ter sido sensores lipídicos, que são receptores com afinidade relativamente baixa para uma variedade de moléculas hidrofóbicas como hemes, retinóides, esteroides, ácidos graxos, eicosanoides e outros lipídios, que são ingeridos com nutrição (Posnien *et al.*, 2021).

A teoria 1 é suportada pelo fato de que os NRs estão se ligando a ligantes estruturalmente diferentes na mesma subfamília e os receptores órfãos estão amplamente espalhados na árvore filogenética, neste contexto, não haveria conexão entre as relações evolutivas dos NRs e a origem de seus ligantes. A teoria 2 é suportada pelo fato de que os dois NRs expressos na esponja (*A. queenslandica*) se ligam, com baixa afinidade, a ácidos graxos de cadeia longa. Assim, ambas as teorias levantam pontos importantes e parece óbvio que LBD e DBD dos receptores apresentam diferentes trajetórias evolutivas, porque estão sob diferentes pressões evolutivas.

Por exemplo, os receptores evolutivamente relacionados da subfamília I, os receptores de hormônio tireoidiano (TRs), os receptores de ácido retinóico (RARs), os receptores ativados por proliferadores de peroxissoma (PPARs) e os receptores de vitamina D (VDRs), ligam-se a ligantes que derivam de vias biossintéticas totalmente diferentes, mas como observamos nas nossas análises sintênicas, pelo menos THR e RAR surgiram de uma duplicação gênica, juntamente com os receptores órfãos NR1Ds.

Posnien *et al.*, (2021) sugere que o DBD foi restrito a regular a rede metabólica, através de sua ligação ao DNA, como fator de transcrição, enquanto o LBD teve que ser flexível e talvez tenha sido adquirido várias vezes de forma independente durante a evolução da característica de ligação do ligante dos NRs.

Figura 9. Grupos de NRs que construímos as árvores filogenéticas, discriminando quais estão em tandem e a localização cromossômica.

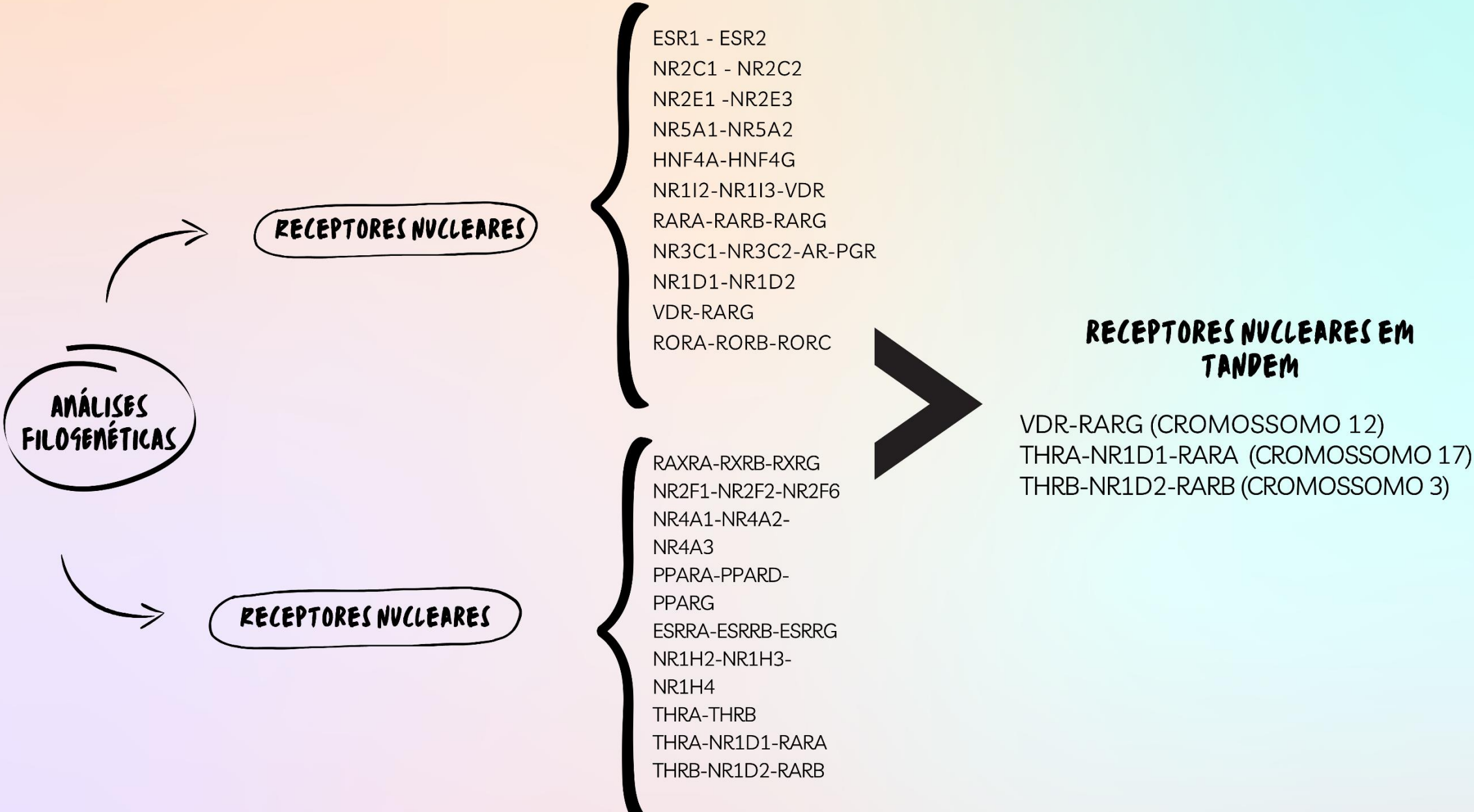


Figura 8. Grupos de NRs que construímos as árvores filogenéticas, discriminando quais estão em tandem e a localização cromossômica.

CONCLUSÃO

Os receptores nucleares (NRs) possivelmente surgiram a partir de uma recombinação quimérica de genes que estavam presentes há mais de 800 milhões de anos atrás. eventos de duplicação de genes, incluindo duplicação de genoma inteiro e duplicações segmentares ou em tandem, foram os principais mecanismos para conduzir a evolução e a expansão da Superfamília dos NRs. Embora a presença de genes NRs tenha sido investigada em uma variedade de organismos, uma ampla pesquisa dos genes NRs de cordados não tinha sido realizada.

Caracterizamos os NRs em 289 cordados, procurando pelas sequências disponíveis nos bancos de dados de genoma animal e nos bancos de dados proteicos. Recuperamos para os cefalocordados 22 NRs, enquanto para os Urocordados foram 17 NRs, nos vertebrados sem mandíbula (agnatas) encontramos, ao todo, 39 NRs. Estudamos um total de 280 espécies de vertebrados com mandíbula (gnatostomados), o maior número de animais investigado para essa classe, e encontramos uma variação do número de cópias dos NRs de 46 a 76 genes em vertebrados. Os dados sugerem que o ancestral dos vertebrados com mandíbula possuía em torno de 52 a 58 NRs.

Nossa análise das taxas evolutivas, para 47 NRs em 60 animais, determinou que o melhor modelo para 45 NRs é o modelo neutro (M1), e que as taxas evolutivas revelam uma extrema seleção negativa, com valores de DN/DS muito menores do que 1. Dois genes, NR4A1 e RORC, têm como o modelo mais adequado o modelo que admite seleção (M8), mas foi incapaz de detectar sítios com seleção positiva (BEB), encontrando somente 10% dos sítios sobre relaxamento da seleção negativa.

Analizamos a sintenia compartilhada para toda superfamília dos NRs, a sintenia é conservada na maioria dos ortólogos nos tetrápodes. Encontramos interessantes eventos de duplicação em tandem, com o evento THR-RAR-NR1D que deve ter ocorrido antes da separação dos cefalocordados dos demais cordados (588 MYA).

Nossos dados de sintenia e de análise filogenética apoiam a hipótese de que os primeiros NRs terem sido sensores lipídicos, de baixa afinidade e através de duplicações, mutações e perdas de genes linhagem ou espécie-específicos ocorreram eventos de neofuncionalização nos NRs emergentes.

REFERÊNCIAS

- Baker, J.L. *et al.* (2016) 'Functional divergence of the nuclear receptor NR2C1 as a modulator of pluripotentiality during hominid evolution', *Genetics*, 203(2), pp. 905–922. Available at: <https://doi.org/10.1534/genetics.115.183889>.
- Baker, M.E. (2019) 'Steroid receptors and vertebrate evolution', *Molecular and Cellular Endocrinology*. Elsevier Ireland Ltd. Available at: <https://doi.org/10.1016/j.mce.2019.110526>.
- Baker, M.E. and Katsu, Y. (2017) 'Evolution of the mineralocorticoid receptor: Sequence, structure and function', *Journal of Endocrinology*. BioScientifica Ltd., pp. T1–T16. Available at: <https://doi.org/10.1530/JOE-16-0661>.
- Bikle, D.D. (2011) 'Vitamin D: An ancient hormone', *Experimental Dermatology*, 20(1), pp. 7–13. Available at: <https://doi.org/10.1111/j.1600-0625.2010.01202.x>.
- Cardoso, J.C.R., Bergqvist, C.A. and Larhammar, D. (2020) 'Corticotropin-Releasing Hormone (CRH) Gene Family Duplications in Lampreys Correlate With Two Early Vertebrate Genome Doublings', *Frontiers in Neuroscience*, 14. Available at: <https://doi.org/10.3389/fnins.2020.00672>.
- Chartier, F.L. *et al.* (1994) *Cloning and sequencing of cDNAs encoding the human hepatocyte nuclear factor 4 indicate the presence of two isoforms in human liver (Hepatic transcription factor; HNF-4; steroid hormone nuclear receptor superfamily)*, *Gene*.
- Chung, A.C.-K. *et al.* (2001) 'Loss of Orphan Receptor Germ Cell Nuclear Factor Function Results in Ectopic Development of the Tail Bud and a Novel Posterior Truncation', *Molecular and Cellular Biology*, 21(2), pp. 663–677. Available at: <https://doi.org/10.1128/mcb.21.2.663-677.2001>.
- Connor, B. *et al.* (2022) 'Biliverdin regulates NR2E3 and zebrafish retinal photoreceptor development', *Scientific Reports*, 12(1). Available at: <https://doi.org/10.1038/s41598-022-11502-3>.
- Coppola, U. and Waxman, J.S. (2021) 'Origin and evolutionary landscape of Nr2f transcription factors across Metazoa', *PLoS ONE*, 16(11 November). Available at: <https://doi.org/10.1371/journal.pone.0254282>.
- Cui, S. *et al.* (2011) 'Nuclear Receptors TR2 and TR4 Recruit Multiple Epigenetic Transcriptional Corepressors That Associate Specifically with the Embryonic β -Type Globin Promoters in Differentiated Adult Erythroid Cells', *molecular and Cellular Biology*, 31(16), pp. 3298–3311. Available at: <https://doi.org/10.1128/mcb.05310-11>.

Cui, Z., Xu, J. and Jiang, W. (2019) 'NR3C2 gene polymorphism is associated with risk of gestational hypertension in Han Chinese women', *Medicine (United States)*, 98(50). Available at: <https://doi.org/10.1097/MD.00000000000018215>.

Davey, R.A. and Grossmann, M. (2016) *Androgen Receptor Structure, Function and Biology: From Bench to Bedside, Androgen Receptor Biology Clin Biochem Rev.*

Dehal, P. *et al.* (2002) *The Draft Genome of Ciona intestinalis: Insights into Chordate and Vertebrate Origins*. Available at: www.jgi.doe.gov/ciona.

Donoghue, P.C.J. and Keating, J.N. (2014) 'Early vertebrate evolution', *Palaeontology*, pp. 879–893. Available at: <https://doi.org/10.1111/pala.12125>.

Donoghue, P.C.J. and Purnell, M.A. (2009) 'The Evolutionary Emergence of Vertebrates from among Their Spineless Relatives', *Evolution: Education and Outreach*, 2(2), pp. 204–212. Available at: <https://doi.org/10.1007/s12052-009-0134-3>.

Ekins, S. *et al.* (2008) 'Evolution of pharmacologic specificity in the pregnane X receptor', *BMC Evolutionary Biology*, 8(1). Available at: <https://doi.org/10.1186/1471-2148-8-103>.

Ermakova, G. v. *et al.* (2020) 'Discovery of four Noggin genes in lampreys suggests two rounds of ancient genome duplication', *Communications Biology*, 3(1). Available at: <https://doi.org/10.1038/s42003-020-01234-3>.

Evans, R.M. and Mangelsdorf, D.J. (2014) 'Nuclear receptors, RXR, and the big bang', *Cell*. Elsevier B.V., pp. 255–266. Available at: <https://doi.org/10.1016/j.cell.2014.03.012>.

Fonseca, E. *et al.* (2017) 'LXR α and LXR β nuclear receptors evolved in the common ancestor of gnathostomes', *Genome Biology and Evolution*, 9(1), pp. 222–230. Available at: <https://doi.org/10.1093/gbe/evw305>.

Fonseca, E. *et al.* (2020) 'Of retinoids and organotins: The evolution of the retinoid x receptor in metazoa', *Biomolecules*, 10(4), pp. 1–16. Available at: <https://doi.org/10.3390/biom10040594>.

Fuchs, B. *et al.* (2014) 'Regulation of polyp-to-jellyfish transition in *Aurelia aurita*', *Current Biology*, 24(3), pp. 263–273. Available at: <https://doi.org/10.1016/j.cub.2013.12.003>.

Gadaleta, R.M. *et al.* (2010) 'Bile acids and their nuclear receptor FXR: Relevance for hepatobiliary and gastrointestinal disease', *Biochimica et Biophysica Acta - Molecular and Cell Biology of Lipids*, pp. 683–692. Available at: <https://doi.org/10.1016/j.bbalip.2010.04.006>.

- Helsen, C. and Claessens, F. (2014) 'Looking at nuclear receptors from a new angle', *Molecular and Cellular Endocrinology*, 382(1), pp. 97–106. Available at: <https://doi.org/10.1016/J.MCE.2013.09.009>.
- Holland, L.Z. (2013) 'Evolution of new characters after whole genome duplications: Insights from amphioxus', *Seminars in Cell and Developmental Biology*. Elsevier Ltd, pp. 101–109. Available at: <https://doi.org/10.1016/j.semcdb.2012.12.007>.
- Hong, F. *et al.* (2019) 'PPARs as nuclear receptors for nutrient and energy metabolism', *Molecules*. MDPI AG. Available at: <https://doi.org/10.3390/molecules24142545>.
- Hughes, A.L. and Friedman, R. (2008) 'Genome size reduction in the chicken has involved massive loss of ancestral protein-coding genes', *Molecular Biology and Evolution*, 25(12), pp. 2681–2688. Available at: <https://doi.org/10.1093/molbev/msn207>.
- Hughes CR. *et al.* (2008) Support for the minimal essential MHC hypothesis: a parrot with a single, highly polymorphic MHC class II B gene. *Immunogenetics* 219–231. DOI: 10.1007/s00251-008-0287-1
- Katsu, Y., Oka, K. and Baker, M.E. (2017) 'Evolution of human, chicken, alligator, frog and zebrafish mineralocorticoid receptors: Allosteric influence on steroid specificity', *bioRxiv* [Preprint]. Available at: <https://doi.org/10.1101/151233>.
- Kitambi, S.S. and Hauptmann, G. (2007) 'The zebrafish orphan nuclear receptor genes nr2e1 and nr2e3 are expressed in developing eye and forebrain', *Gene Expression Patterns*, 7(4), pp. 521–528. Available at: <https://doi.org/10.1016/j.modgep.2006.10.006>.
- Kon, T. *et al.* (2007) 'Phylogenetic position of a whale-fall lancelet (Cephalochordata) inferred from whole mitochondrial genome sequences', *BMC Evolutionary Biology*, 7. Available at: <https://doi.org/10.1186/1471-2148-7-127>.
- Layeghifard, M. *et al.* (2008) 'Evolutionary mechanisms underlying the functional divergence of duplicate genes involved in vertebrates' circadian rhythm pathway', *Gene*, 426(1–2), pp. 65–71. Available at: <https://doi.org/10.1016/J.GENE.2008.08.014>.
- Lecroisey, C., Laudet, V. and Schubert, M. (2012) 'The cephalochordate amphioxus: A key to reveal the secrets of nuclear receptor evolution', *Briefings in Functional Genomics*, 11(2), pp. 156–166. Available at: <https://doi.org/10.1093/bfgp/els008>.

Maeng, S. *et al.* (2012) 'The retinoid X receptor in a marine invertebrate chordate: Evolutionary insights from urochordates', *General and Comparative Endocrinology*, 178(2), pp. 380–390. Available at: <https://doi.org/10.1016/J.YGCEN.2012.06.019>.

Markov, G. v. and Laudet, V. (2011) 'Origin and evolution of the ligand-binding ability of nuclear receptors', *Molecular and Cellular Endocrinology*, 334(1–2), pp. 21–30. Available at: <https://doi.org/10.1016/j.mce.2010.10.017>.

Miglioli, A. *et al.* (2021) 'Nuclear receptors and development of marine invertebrates', *Genes*, 12(1), pp. 1–21. Available at: <https://doi.org/10.3390/GENES12010083>

Ohno, S. (1970) *Evolution by Gene Duplication*, Springer

Palma-Gudiel, H. *et al.* (2015) 'Glucocorticoid receptor gene (NR3C1) methylation processes as mediators of early adversity in stress-related disorders causality: A critical review', *Neuroscience and Biobehavioral Reviews*. Elsevier Ltd, pp. 520–535. Available at: <https://doi.org/10.1016/j.neubiorev.2015.05.016>.

Pasquier, J. *et al.* (2017) 'Evolution of gene expression after whole-genome duplication: New insights from the spotted gar genome', *Journal of Experimental Zoology Part B: Molecular and Developmental Evolution*, 328(7), pp. 709–721. Available at: <https://doi.org/10.1002/jez.b.22770>.

Philip, S. *et al.* (2012) 'Adaptive evolution of the Retinoid X receptor in vertebrates', *Genomics*, 99(2), pp. 81–89. Available at: <https://doi.org/10.1016/J.YGENO.2011.12.001>.

Posnien, N. *et al.* (2021) 'Function and Evolution of Nuclear Receptors in Environmental-Dependent Postembryonic Development', *Frontiers in Cell and Developmental Biology* | www.frontiersin.org, 1, p. 653792. Available at: <https://doi.org/10.3389/fcell.2021.653792>.

Reschly, E. and Krasowski, M. (2006) 'Evolution and Function of the NR1I Nuclear Hormone Receptor Subfamily (VDR, PXR, and CAR) with Respect to Metabolism of Xenobiotics and Endogenous Compounds', *Current Drug Metabolism*, 7(4), pp. 349–365. Available at: <https://doi.org/10.2174/138920006776873526>.

Reschly, E.J. *et al.* (2008) *Ligand specificity and evolution of liver X receptors §*, *J Steroid Biochem Mol Biol*. Available at: <http://ghost.zool.kyoto-u.ac.jp/indexr1.html>.

Robinson-Rechavi, M. *et al.* (2001) 'How many nuclear hormone receptors are there in the human genome?', *Trends in Genetics*, 17(10), pp. 554–556. Available at: [https://doi.org/10.1016/S0168-9525\(01\)02417-9](https://doi.org/10.1016/S0168-9525(01)02417-9).

- Robinson-Rechavi, M., Garcia, H.E. and Laudet, V. (2003) 'The nuclear receptor superfamily', *Journal of Cell Science*, 116(4), pp. 585–586. Available at: <https://doi.org/10.1242/jcs.00247>.
- Rodrigues Paixão-Côrtes, V., Salzano, F.M. and Bortolini, M.C. (2013) 'Evolutionary history of chordate PAX genes: Dynamics of change in a complex gene family', *PLoS ONE*, 8(9), pp. 19–23. Available at: <https://doi.org/10.1371/journal.pone.0073560>.
- Shujiong Feng, S.X.Z.W.Y.Z. (2015) 'Retinoic acid-related orphan receptor ROR β , circadian rhythm abnormalities and tumorigenesis (Review)', *Mol Med.*, 6, pp. 493–500. Available at: <https://doi.org/10.3892/ijmm.2015.2155>.
- Sonoda, J., Pei, L. and Evans, R.M. (2008) 'Nuclear receptors: Decoding metabolic disease', *FEBS Letters*, pp. 2–9. Available at: <https://doi.org/10.1016/j.febslet.2007.11.016>.
- Xia, H., Dufour, C.R. and Giguère, V. (2019) 'ERR α as a bridge between transcription and function: Role in liver metabolism and disease', *Frontiers in Endocrinology*, 10(APR). Available at: <https://doi.org/10.3389/fendo.2019.00206>.
- Zhang, H. *et al.* (2017) 'LXR a and LXR b Nuclear Receptors Evolved in the Common Ancestor of Gnathostomes', 9(1), pp. 222–230. Available at: <https://doi.org/10.1093/gbe/evw305>.
- Zhang, J., Nielsen, R. and Yang, Z. (2004) 'Evaluation of an Improved Branch-Site Likelihood Method for Detecting Positive Selection at the Molecular Level'. Available at: <https://doi.org/10.1093/molbev/msi237>.
- Zhao, Y. *et al.* (2015) 'Families of nuclear receptors in vertebrate models: Characteristic and comparative toxicological perspective', *Scientific Reports*, 5. Available at: <https://doi.org/10.1038/srep08554>.
- Zhao, Y. and Bruemmer, D. (2010) 'NR4A orphan nuclear receptors: Transcriptional regulators of gene expression in metabolism and vascular biology', *Arteriosclerosis, Thrombosis, and Vascular Biology*, pp. 1535–1541. Available at: <https://doi.org/10.1161/ATVBAHA.109.191163>.

DADOS SUPLEMENTARES

1. Presença e ausência dos RNs
2. Mart - export celacanto 2 – Descrição dos genes
3. Domínios
4. Análises das taxas evolutivas
5. Localização cromossômica e fenótipos associados
6. Alinhamento com 69 espécies de cordados estudados
7. Árvores gênicas do
8. s DBD dos NRs