

UNIVERSIDADE FEDERAL DE SANTA CATARINA  
CENTRO DE CIÊNCIAS BIOLÓGICAS  
DEPARTAMENTO DE BIOLOGIA CELULAR, EMBRIOLOGIA E GENÉTICA  
CURSO DE CIÊNCIAS BIOLÓGICAS

Leonardo de Bortoli

**Título: Filogenia de genes da superfamília dos receptores nucleares em espécies  
aquáticas**

Florianópolis

2022

Leonardo de Bortoli

**Título: Filogenia de genes da superfamília dos receptores nucleares em espécies  
aquáticas**

Trabalho Conclusão do Curso de Graduação em  
Florianópolis do Centro de Ciências Biológicas da  
Universidade Federal de Santa Catarina como requisito  
para a obtenção do título de Licenciado em Ciências  
Biológicas  
Orientador: Prof. Dr. Guilherme de Toledo e Silva

Florianópolis

2022

Ficha de identificação da obra elaborada pelo autor, através do Programa de Geração Automática da  
Biblioteca Universitária da UFSC.

Bortoli, Leonardo de  
Filogenia de genes da superfamília dos receptores  
nucleares em espécies da costa brasileira / Leonardo de  
Bortoli ; orientador, Guilherme de Toledo e Silva, 2022.  
52 p.

Trabalho de Conclusão de Curso (graduação) -  
Universidade Federal de Santa Catarina, Centro de Ciências  
Biológicas, Graduação em Ciências Biológicas, Florianópolis,  
2022.

Inclui referências.

1. Ciências Biológicas. 2. Ecotoxicologia. 3.  
Biomarcadores. 4. Evolução. 5. Filogenia Molecular. I.  
Toledo e Silva, Guilherme de. II. Universidade Federal de  
Santa Catarina. Graduação em Ciências Biológicas. III. Título.

Leonardo de Bortoli

**Título: Filogenia de genes da superfamília dos receptores nucleares em espécies  
aquáticas**

Este Trabalho Conclusão de Curso foi julgado adequado para obtenção do Título de “Licenciado em Ciências Biológicas” e aprovado em sua forma final pelo Curso de Ciências Biológicas

Florianópolis, 06 de julho de 2022.

---

Prof.<sup>a</sup> Daniela Cristina de Toni, Dra.  
Coordenadora do Curso

**Banca Examinadora:**

---

Prof. Guilherme de Toledo e Silva, Dr.  
Orientador  
Universidade Federal de Santa Catarina

---

Ariana Nicaretta Amorim, Ma.  
Avaliadora  
Universidade Federal de Santa Catarina

---

Prof. Guilherme Razzera Maciel, Dr.  
Avaliador(a)  
Universidade Federal de Santa Catarina

---

Prof.<sup>a</sup> Andrea Rita Marrero, Dra.  
Avaliador(a)  
Universidade Federal de Santa Catarina

Este trabalho é dedicado aos meus pais, Sandra e Giovani.

## AGRADECIMENTOS

Toda a minha trajetória na Universidade Federal de Santa Catarina não seria possível sozinho e eu gostaria de, nessas linhas, agradecer todos aqueles que fizeram, fazem e espero que continuem fazendo parte de minha vida.

Em primeiro lugar, quero agradecer a minha família, principalmente minha mãe e meu pai, Sandra e Giovani. A dedicação, o carinho, o apoio e o incentivo à curiosidade e a leitura que os dois sempre demonstraram e prezaram em minha criação foram essenciais para toda a minha formação como pessoa, professor e cientista. Cada um, à sua maneira, me mostrou a importância de ser curioso, até mesmo quando a minha curiosidade danificava peças de decoração, roupas de cama ou meu cabelo. Por todo o amor, pela paciência, pelo bom humor, pelo ótimo gosto musical, pelas viagens, pelos genes, centríolos, mitocôndrias e tantas coisas mais, sou eternamente grato aos meus pais. Quero agradecer também meu irmão Guilherme, por dividir a vida e os genes dos nossos pais comigo. Foi meu irmão quem me recebeu em Florianópolis e me ajudou a passar pelos momentos iniciais na universidade, seja levando em festas estranhas da engenharia, Container 2 ou Treze, fazendo lanches duvidosos de madrugada ou me levando para ir comer no grande Xande Lanches. Geneticamente somos 50% iguais, mas somos também 50% diferentes, e apesar das nossas diferenças, muito do que eu sou hoje tem influência de meu irmão e sua (ou nossa) fixação por indie rock e Hermes e Renato. Gostaria de dedicar este trabalho também para minhas avós Maria de Lourdes e Edith, mulheres que amo muito e que sempre me foram muito queridas, que ajudaram na minha criação e a quem devo meu apreço pela natureza, seja pelas flores de Vó Lourdes ou pelas hortas e animais de Vó Edith. Quero agradecer também minha tia e madrinha Márcia, uma das pessoas mais carinhosas e presentes na minha vida e também uma das pessoas mais fortes que conheci, a quem admiro muito por sua resiliência, dedicação e por fazer a melhor lentilha do mundo. Agradeço também meu primo Fernando, que me acompanhou por toda a infância e que hoje me acompanha nos momentos finais da graduação, dividindo momentos de jogatina, tardes assistindo Casimiro ou simplesmente contando como foi o dia. Quero agradecer meu falecido avô Arlindo por todos os momentos e memórias que trago comigo, por me ensinar a pescar, por me incentivar a tentar aprender italiano e por me mostrar a importância de uma sesta após o almoço. A toda a família de Bortoli, muito obrigado, em especial minha Tia Célia por me dar tantos livros e mostrar tantos discos.

Em segundo lugar, quero agradecer a todas as professoras e professores que tive, principalmente à profa. Gláucia, que por suas aulas me mostrou a importância da biologia e a

beleza das moléculas, genes, células e evolução. Agradeço aos professores da graduação que demonstraram cuidado e interesse não só em suas pesquisas, mas em suas aulas, no seu ensino e nos seus alunos. Ao professor Rafael, por ter me dado a oportunidade de entrar no meu primeiro laboratório, por me proporcionar tantas experiências, projetos, conexões, lugares, experimentos, apresentações, erros, acertos e o mais importante, conhecer tantas pessoas incríveis no LIAA, muito obrigado. Quero agradecer ao prof. Guilherme Toledo, meu orientador neste trabalho, que tem um coração enorme e me recebeu durante a pandemia sem nunca ter me visto antes, sempre incentivando a seguir na área da evolução e mostrando como coisas que parecem difíceis são muitas vezes mais fáceis do que parecem. Agradeço ao prof. Guilherme Brito, que mostrou como pode ser incrível acordar às cinco horas da manhã para olhar passarinhos no mato, bem como a beleza da zoologia e taxonomia, que tentei passar durante a escrita deste TCC. Gostaria de agradecer também o prof. Renato, que me acolheu no Projeto Brotar e em suas disciplinas, mostrando que mesmo trabalhando com educação ambiental ou fisiologia animal, há que se ter paixão e empolgação com o que fazemos.

Agradeço a meus amigos, que me fizeram companhia durante todo esse processo e o fizeram ser tão incrível. Em especial, quero agradecer ao Iago, Gabriel e Max, os maiores presentes que Florianópolis me deu. Agradeço à Bruna, Beatriz Rosim e Beatriz Kanasawa, Heloísa, Pietro e Rafael – todos vocês também estavam comigo no começo, meio e fim dessa jornada, seja perto ou longe, sempre fazendo tudo ser mais engraçado e leve. Quero agradecer aos meus amigos do LIAA, Mariana, Gabriel Matos, Cairé, Nicolas e Natan, que me receberam e ensinaram praticamente tudo que sei sobre fazer pesquisa e que alegraram tantos dias com suas músicas, piadas, trocadilhos e fofocas na copa. Aos amigos da turma 17.1 e da República Coisa Boa, muito obrigado por todos os momentos felizes, churrascos, carnavais e conversas de bar. Agradeço às amigas que fiz durante o período de escrita deste projeto, enquanto realizava estágio na Escola Dinâmica, Nina, Cláudia, Agatha e Samanta, professoras incríveis que ensinaram muito sobre a docência e sobre como ser afetuoso nesse processo. Muito obrigado também aos meus queridos alunos, em especial ao 1º ano A, que despertaram o amor à docência dentro de mim e que faziam todas as tardes passarem voando em horas muito preciosas. Agradeço aos meus amigos Klein, Cabral, Sarah, Duda e Roberta que durante os momentos de isolamento da pandemia, fizeram-se tão presentes e próximos. Por fim, quero agradecer meus colegas do Laboratório de Genômica, que me receberam e auxiliaram de prontidão e que foram e são essenciais para o desenvolvimento deste trabalho.

*“O mistério da vida não é um problema a ser resolvido, mas uma realidade a ser experienciada.” (Duna, 1965)*



## RESUMO

Os ambientes aquáticos estão entre os mais ameaçados pela contaminação ambiental resultante de atividades antropogênicas. Entre os poluentes despejados nos corpos hídricos que mais causam preocupação estão os denominados desreguladores endócrinos (DE), que possuem a capacidade de alterar funções fisiológicas dos organismos, causando reações adversas no controle homeostático, desenvolvimento e reprodução. Devido a este cenário, faz-se cada vez mais necessária a identificação de novas moléculas biomarcadoras para o monitoramento ambiental. Dentre os diversos biomarcadores potenciais identificados até o momento, estão os receptores nucleares, uma superfamília de proteínas que atuam em diversas vias metabólicas, incluindo a neutralização de xenobióticos. Esses receptores estão presentes em todos os metazoários e possuem uma arquitetura similar entre si, apresentando regiões conservadas em suas sequências, como o domínio de ligação ao DNA (DBD) e o domínio de ligação ao ligante (LBD), essenciais para o desempenho das suas funções, bem como regiões menos conservadas. Levando em consideração a similaridade entre as sequências dos receptores nucleares, estudos filogenéticos podem ser realizados a fim de elucidar as relações evolutivas desses receptores entre as espécies aquáticas. Dessa forma, o objetivo deste trabalho foi realizar filogenias baseada em sequências de espécies de interesse conservacional, utilizando sequências aminoacídicas completas e sequências dos domínios DBD e LBD, além de criar modelos tridimensionais para esses domínios. Os resultados das análises filogenéticas indicaram topologias coerentes entre os grupos e famílias de receptores nucleares para todas as abordagens, podendo ajudar na identificação de genes ortólogos e futuras moléculas biomarcadoras, auxiliando também no monitoramento ambiental. As análises estruturais geraram modelos tridimensionais muito similares para cada domínio, reforçando a alta conservação dos domínios estudados. De modo geral foi feita uma comparação entre filogenias baseadas em sequências completas e em domínios, demonstrando que para os receptores nucleares é possível utilizar a última como alternativa, facilitando as análises filogenéticas e estruturais e reduzindo o custo computacional.

**Palavras-chave:** Ecotoxicologia. Biomarcadores. Evolução. Filogenia molecular.

## ABSTRACT

Aquatic environments are among the most threatened by contamination resulting from anthropogenic actions. Among the pollutants thrown in bodies of water, endocrinal disruptants are of great concern, since they have the capacity to alter physiological functions, causing adverse reactions in homeostatic control, development and reproduction. Given this scenario, it is urging to identify new biomarker molecules for environmental monitoring. At the moment, amidst the many identified biomarkers, are the nuclear receptors - a protein superfamily which is involved in many metabolic pathways, including xenobiotics neutralization. These receptors can be found in all metazoans and share a similar architecture, showing conserved regions in their sequence, like the DNA binding domain (DBD) and the ligand binding domain (LBD), essential for proper functioning of nuclear receptors, as well as less conserved regions. Considering the similarity between nuclear receptors sequences, phylogenetic studies can be done in order to clarify the evolutionary history of these receptors among aquatic species. Therefore, the main objective of this work was to do phylogenetic analysis based on complete sequences and DBD and LBD sequences, as well as creating tridimensional models for those domains. Results from the phylogenetic analysis showed coherent topologies for nuclear receptors groups and families for both approaches, which may help in the identification of orthologous genes and biomarkers molecules, facilitating environmental monitoring. Structural analysis generated very similar tridimensional models for each domain, reinforcing the known high similarity of those domains. Altogether, a comparison between sequence and domain-based phylogenies was done, showing that for nuclear receptors it is possible to use the latter as an alternative, facilitating phylogenetic and structural analysis and reducing computational costs.

**Keywords:** Ecotoxicology. Biomarkers. Evolution. Molecular phylogenetics.

## LISTA DE FIGURAS

Figura 1 – Esquema de estrutura geral dos receptores nucleares. ....	17
Figura 2 - Conformação tridimensional dos domínios DBD e LBD. ....	18
Figura 3 - <i>Sequence logos</i> de regiões mais conservadas do DBD e LBD. ....	27
Figura 4 - Conformação tridimensional do DBD de diferentes famílias e espécies. ....	29
Figura 5 - Conformação tridimensional do LBD de diferentes famílias e espécies. ....	30
Figura 6 - Filogenia de famílias de RNs baseada em suas sequências completas. ....	33
Figura 7 - Filogenia de famílias de RNs baseada na sequência do domínio de ligação ao DNA. .....	37
Figura 8 - Filogenia de famílias de RNs baseada nas sequências do domínio de ligação ao ligante. ....	39
Figura 9 - Motivos conservados presentes no LBD de diferentes receptores nucleares. ....	47
Figura 10 - Alinhamento entre as sequências do receptor retinoide X de mamíferos. ....	48
Figura 11 - Alinhamento entre as sequências do receptor relacionado ao estrógeno de mamíferos. ....	49
Figura 12 - Alinhamento entre as sequências do receptor X do fígado de mamíferos. ....	50

## **LISTA DE TABELAS**

Tabela 1 – Relação dos receptores nucleares abordados neste trabalho e seus ligantes. .... 24

## LISTA DE ABREVIATURAS E SIGLAS

RN	Receptor nuclear
DBD	Domínio de ligação ao DNA
LBD	Domínio de ligação ao ligante
DE	Disruptor endócrino
PPCPs	Produtos farmacêuticos e de cuidado pessoal
AF-1	Função de ativação 1
AF-2	Função de ativação 2
ER	Receptor de Estrógeno
ERR	Receptor Relacionado ao Estrógeno
HNF4	Fator Nuclear 4 do Hepatócito
LXR	Receptor X do Fígado
FXR	Receptor Farnesoide X
PXR	Receptor Pregnano X
VDR	Receptor de Vitamina D
RAR	Receptor de Ácido Retinoico
RXR	Receptor Retinoide X
EcR	Receptor de Ecdisona
HR96	Receptor Hormonal 96

## SUMÁRIO

<b>1</b>	<b>INTRODUÇÃO .....</b>	<b>15</b>
1.1	OBJETIVOS .....	20
1.1.1	<b>Objetivo Geral.....</b>	<b>20</b>
1.1.2	<b>Objetivos Específicos .....</b>	<b>20</b>
<b>2</b>	<b>MATERIAIS E MÉTODOS .....</b>	<b>22</b>
2.1	ANÁLISE FILOGENÉTICA .....	22
2.2	ANÁLISE DE CONSERVAÇÃO DE AMINOÁCIDOS .....	23
2.3	CRIAÇÃO DE MODELOS TRIDIMENSIONAIS DO DBD E LBD .....	23
<b>3</b>	<b>RESULTADOS E DISCUSSÃO.....</b>	<b>24</b>
3.1	IDENTIFICAÇÃO DE FAMÍLIAS DE RECEPTORES NUCLEARES .....	24
3.2	IDENTIFICAÇÃO E CARACTERIZAÇÃO DOS DOMÍNIOS DBD E LBD ...	26
3.3	FILOGENIA BASEADA EM SEQUÊNCIAS COMPLETAS E EM DOMÍNIOS .....	32
<b>4</b>	<b>CONCLUSÃO.....</b>	<b>43</b>
	<b>REFERÊNCIAS.....</b>	<b>44</b>
	<b>APÊNDICE A – MOTIVOS LXXLL/LLXXL DE DIFERENTES FAMÍLIAS DE RNS .....</b>	<b>49</b>
	<b>APÊNDICE B – ALINHAMENTOS DE SEQUÊNCIAS QUE APRESENTARAM POLITOMIA.....</b>	<b>50</b>

## 1 INTRODUÇÃO

Dentre os ambientes afetados negativamente por ações antropogênicas, os ambientes aquáticos se destacam de maneira grave. Isso se dá principalmente pela contaminação dos corpos hídricos por poluentes químicos vindos das indústrias e domicílios. Entre os poluentes, pode-se destacar xenobióticos como os metais pesados, hidrocarbonetos, antibióticos, pesticidas, remédios veterinários e produtos farmacêuticos e de cuidado pessoal (PPCP), que atualmente possuem pouca informação a respeito de aspectos como persistência, acumulação e toxicidade nos ambientes aquáticos (ARYAL et al., 2020; EBELE; ABOU-ELWafa ABDALLAH; HARRAD, 2017). Estes xenobióticos podem também ser classificados como desreguladores endócrinos (DE), visto que desregulam vias metabólicas ao se ligarem a diversos receptores celulares, podendo causar problemas na reprodução, desenvolvimento, metabolismo e até câncer em humanos e outros animais (TOPOROVA; BALAGUER, 2020; ZHAO et al., 2015).

Devido a este cenário de degradação ambiental, os estudos de ecotoxicidade se fazem cada vez mais necessários. Recentemente, dentro do campo da ecotoxicologia, começou-se a utilizar as ciências ômicas (genômica, transcritômica, proteômica e metabolômica) para corroborar as análises de impacto ambiental (GARCIA-REYERO; PERKINS, 2011). Na ecotoxigenômica, por exemplo, usam-se técnicas que permitem detectar mudanças de expressão gênica nos animais aquáticos antes e após a exposição aos agentes poluentes. Desta forma, é possível saber quais processos celulares estão sendo desregulados a partir da expressão diferencial dos genes envolvidos nessas vias (SNAPE et al., 2004; STEINBERG; STÜRZENBAUM; MENZEL, 2008).

Com a detecção dos genes modulados pela exposição aos xenobióticos, pode-se começar a procurar por moléculas biomarcadoras de poluição nos animais expostos, a fim de facilitar o biomonitoramento dos corpos hídricos. Dentre os organismos afetados, alguns podem servir como sentinelas para o monitoramento ambiental, principalmente grandes mamíferos, aves marinhas e algumas espécies de peixes (BURGER; GOCHFELD, 2004; MOORE, 2008). Alguns fatores são muito importantes para a classificação de mamíferos e aves marinhas como espécies sentinela, como por exemplo, estarem no topo da cadeia alimentar com capacidade de regular as relações tróficas, possuírem longa expectativa de vida, acumularem poluentes e serem relativamente fáceis de encontrar (BURGER; GOCHFELD, 2004; HAZEN et al., 2019; HOEKSTRA et al., 2003). Levando em consideração esses fatores, esses organismos

apresentam grande potencial para atuar como sentinelas no monitoramento ambiental, uma vez que se tornam organismos relevantes biologicamente para identificação de biomarcadores.

Os receptores nucleares (RN) são uma superfamília de proteínas que podem ser utilizadas como potenciais biomarcadores, sendo um dos principais grupos de fatores de transcrição presentes em todos os metazoários, que participam de diversas funções fisiológicas como metabolismo de ácidos graxos, sais biliares e colesterol, controle do desenvolvimento embrionário e diferenciação celular (FRANCIS et al., 2003; HOLZER; MARKOV; LAUDET, 2017). Os principais ligantes naturais desses receptores são geralmente hormônios hidrofóbicos (esteroides como estrógeno, progesterona, glicocorticoides e mineralocorticoides), hormônios da tireoide, ácidos graxos, ácidos retinóicos e sais biliares. Porém, foi observado que moléculas sintéticas como os PPCP e alguns DE podem se ligar aos receptores nucleares de animais marinhos, visto que alguns deles foram desenvolvidos para simular moléculas naturais, as quais se ligam aos receptores causando efeitos adversos (JANOŠEK et al., 2006). Os receptores nucleares que não possuem ligantes conhecidos até o momento são considerados receptores órfãos. O modo como os ligantes (naturais ou sintéticos) se conectam aos receptores é por meio do domínio de ligação do ligante (*ligand binding domain* ou LBD), presente em todos os receptores nucleares (JANOŠEK et al., 2006).

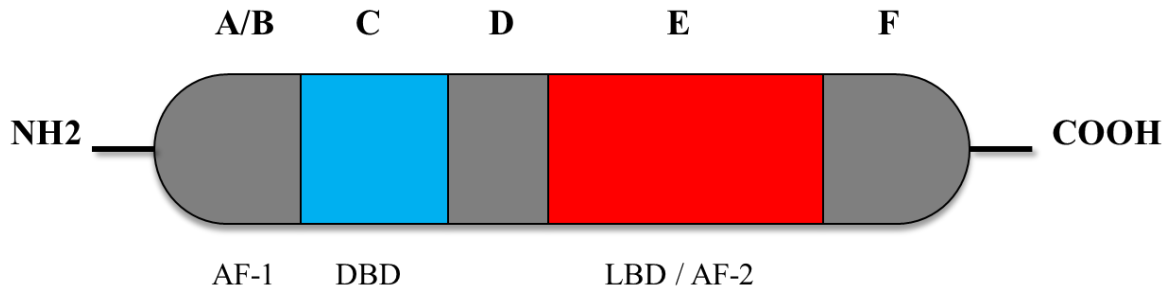
De modo geral, a arquitetura dos receptores nucleares compreende cinco regiões (**Figura 1**). Primeiramente, na porção N-terminal está a região A/B, que possui sequência pouco conservada e onde está localizado o domínio de função de ativação AF-1 (independente de ligante). Em sequência, a porção mais conservada é a região C ou domínio de ligação ao DNA (*DNA binding domain* ou DBD), no qual se encontram dois dedos de zinco e uma *P-box*, responsáveis por reconhecer sequências específicas na fita de DNA, a partir das quais os receptores se ligam e se dimerizam (ROBINSON-RECHAVI; GARCIA; LAUDET, 2003). A região D funciona como uma dobradiça entre as regiões C e E, flexibilizando a estrutura do receptor. A maior região dos receptores nucleares é a E ou LBD, situada na porção C-terminal, é conservada entre os receptores, porém menos que o DBD. É nela que se situa também o AF-2, que depende de ligantes, ao contrário do AF-1. A última porção dos receptores nucleares é a região F, que nem sempre está presente e pouco se sabe a respeito da sua função e estrutura (ROBINSON-RECHAVI; GARCIA; LAUDET, 2003).

Estudos a respeito da filogenia desses receptores podem ser de grande utilidade para facilitar a identificação de genes ortólogos em espécies pouco estudadas. Em relação à homologia genética, os genes podem ser parálogos caso surjam de duplicação gênica ou



ortólogos, caso surjam a partir de eventos de especiação (FITCH, 1970). Apesar de ambos possuírem ancestralidade comum, os genes parálogos geralmente apresentam funções distintas e raramente são sintênicos, enquanto os genes ortólogos quase sempre possuem funções similares ou iguais entre si, além de geralmente serem sintênicos (VALLENDER, 2009).

**Figura 1.** Esquema de estrutura geral dos receptores nucleares.



A arquitetura geral dos receptores apresenta seis regiões de A a F. A região A/B se encontra na porção N-terminal, junto com AF-1. Na porção central, encontra-se a região C (domínio de ligação ao DNA), D e E (domínio de ligação ao ligante e AF-2). Na porção C-terminal, encontra-se a região F. Fonte: Elaborado pelo autor (2022), baseado na ilustração de OWEN et al. (2000).

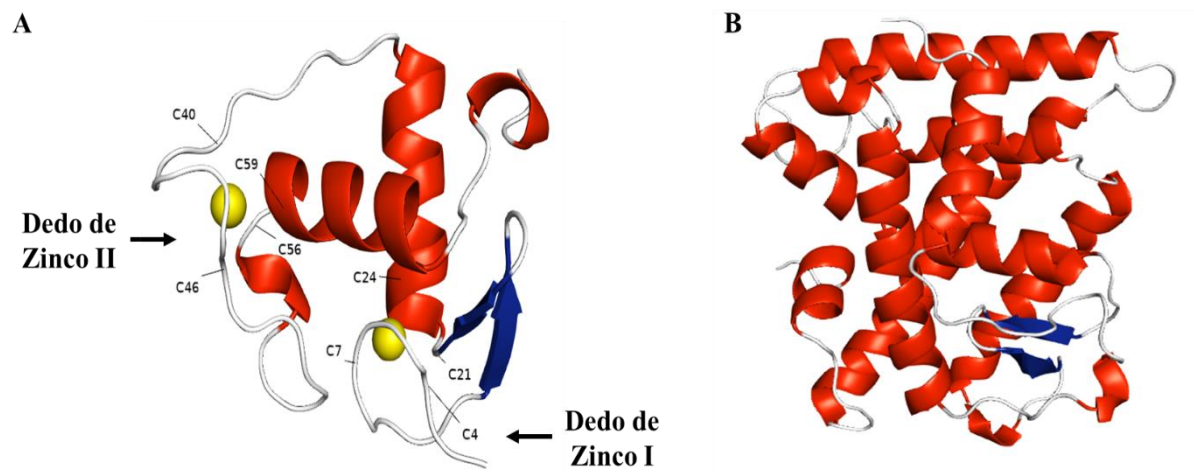
A identificação de genes ortólogos pode ser feita de três maneiras: utilizando métodos baseados em árvores filogenéticas (*tree-based methods*), métodos baseados em grafos (*graph-based methods*) ou métodos híbridos. No primeiro método, árvores filogenéticas contendo vários genes são analisadas a fim de encontrar eventos de duplicação gênica, para então identificar quais genes são ortólogos e quais são parálogos (KRISTENSEN et al., 2011; SJÖLANDER et al., 2011). É possível ainda comparar a filogenia dos genes com uma árvore filogenética representando a evolução das espécies envolvidas, num processo chamado de reconciliação de árvores (*tree reconciliation*), tornando a análise mais confiável (SJÖLANDER, 2010). Na abordagem envolvendo grafos, os genes ortólogos podem ser inferidos por meio da semelhança entre as sequências a serem estudadas e bancos de dados, com base nos chamados *reciprocal BLAST hits* (RBHs; KRISTENSEN et al., 2011; SJÖLANDER, 2010). As sequências que obtiverem a maior semelhança com àquela utilizada como entrada serão inferidas como ortólogos. Por fim, pode-se utilizar métodos híbridos que envolvam ambas as abordagens, a fim de minimizar as desvantagens de cada uma. Por exemplo, a metodologia baseada em árvores filogenéticas é mais confiável, porém depende de uma grande quantidade de informações disponíveis e requer mais tempo e formação do usuário do que a metodologia baseada em gráficos. Esta, por sua vez, é menos sensível e pode inferir genes

parálogos como ortólogos em sequências incompletas ou em processos de perda de genes (CHEN et al., 2007; FULTON et al., 2006).

Como mencionado anteriormente, os receptores nucleares possuem regiões conservadas nas suas sequências, como os domínios DBD e LBD, essenciais para o desempenho das suas funções celulares (**Figura 2**). O domínio de ligação ao DNA é o menor dos dois domínios presentes nos receptores nucleares e o mais conservado, apresentando cerca de 70 a 80 aminoácidos com posições bem conservadas em sua sequência (ZHAO et al., 2015), como por exemplo os oito resíduos de cisteína que compõem os dois dedos de zinco tipo-II presentes nesta região, característicos por sua função de ligação a ácidos nucleicos, interações entre proteínas e ligação a moléculas menores (KRISHNA; MAJUMDAR; GRISHIN, 2003). Ao contrário do padrão clássico observado nesse motivo, os dedos de zinco do DBD não possuem resíduos de histidina envolvidos nos processos de ligação ao zinco. Este domínio atua principalmente ligando-se ao DNA, reconhecendo motivos específicos e iniciando um processo de dimerização com outros receptores nucleares (ROBINSON-RECHAVI; GARCIA; LAUDET, 2003). Até o presente momento, apenas membros da subfamília NR0B foram documentados como não possuindo o DBD, sendo eles DAX1 e SHP, que agem como repressores de outros receptores nucleares por meio do reconhecimento de motivos presentes no LBD (QIAN et al., 2019).

Em contrapartida, o LBD apresenta sequência menos conservada que o DBD, além de possuir também maior comprimento, variando de 170 a 250 aminoácidos aproximadamente (WÄRNMARK et al., 2003; ZHAO et al., 2015). Além disso, é no LBD que é possível encontrar o AF-2, um domínio de ativação dependente de ligante com estrutura bastante conservada, assim como a estrutura geral do próprio LBD (WÄRNMARK et al., 2003), que segue um padrão de 12  $\alpha$ -hélices e 2 a 4 folhas- $\beta$ . Este domínio possui também motivos bem conservados em sua sequência, como o LXXLL/LLXXL, que se acredita estar relacionado ao reconhecimento de moléculas coativadoras do receptor nuclear (MITSIS et al., 2019).

**Figura 2.** Conformação tridimensional dos domínios DBD e LBD.



A) Domínio de ligação ao DNA e seus resíduos conservados de cisteína, os quais atuam no motivo de dedo de zinco, apontados pelas setas. Bolas amarelas representam átomos de zinco. B) Domínio de ligação ao ligante com suas 12  $\alpha$ -hélices e 2 folhas- $\beta$ . Fonte: Elaborado pelo autor (2022).

Tendo em vista a importância desses domínios, as suas sequências podem ter potencial de serem utilizadas em uma metodologia baseada em árvores como alternativa para a construção de uma filogenia da superfamília dos receptores nucleares baseada em suas sequências completas. Visto que as outras regiões (A, B, D e F) dessas proteínas geralmente apresentam bastante variação em suas sequências, é possível diminuir a chance de erros nos alinhamentos das sequências ao retirá-las das análises (OWEN; ZELENT, 2000; SJÖLANDER et al., 2011). Alguns estudos envolvendo filogenia de domínios proteicos demonstraram que este é um método mais simples e confiável de ser aplicado, tanto em larga quanto em pequena escala, inclusive para o DBD e LBD (MITSIS et al., 2019; YANG; DOOLITTLE; BOURNE, 2004; ZHAO et al., 2015). Porém, como os domínios possuem tamanhos, resíduos e funções diferentes, suas taxas de mutação podem também diferir, o que pode resultar em árvores filogenéticas únicas para cada um. Dessa forma, é interessante analisar os resultados de uma análise filogenética baseada em domínios com aqueles de uma análise filogenética geral, se disponível, para efeito de comparação.

As principais aplicações dos métodos baseados em árvores filogenéticas são a detecção de genes ortólogos e a possível inferência de função a esses genes encontrados, partindo da premissa de que este tipo de homologia conserva total ou parcialmente a função dos genes (ALTENHOFF; DESSIMOZ, 2009). Porém, pode-se utilizar também essas metodologias para tentar compreender melhor as relações evolutivas das espécies avaliadas e identificar novos

biomarcadores para os poluentes químicos em espécies pouco estudadas a partir das informações vindas das espécies modelo.

Além disso, sabe-se que a estrutura secundária e terciária de uma proteína tende a se conservar melhor do que a estrutura primária (SJÖLANDER et al., 2011), ou seja, sua estrutura tridimensional (3D) e sua sequência aminoacídica, respectivamente. Desta forma, além de analisar a sequência dos receptores nucleares, é interessante estudar também se esse padrão é mantido nessas moléculas ou se é possível observar mudanças na sua estrutura tridimensional geral, bem como em seus domínios de interesse, DBD e LBD, os quais possuem interfaces de ligação com o DNA e moléculas ligantes. Visto que é a estrutura 3D de uma proteína que determina sua função, analisá-la pode ajudar a explicar as semelhanças e diferenças estruturais entre as famílias de RN e seus domínios. É por meio deste tipo análise também, que é possível visualizar, ilustrar e prever regiões de interesse dessas proteínas, como sítios ativos e motivos (LA; SUTCH; LIVESAY, 2005).

Dessa forma, a identificação de genes ortólogos de receptores nucleares em espécies sentinelas aquáticas e estudos a respeito de sua filogenia podem ajudar a elucidar a relação evolutiva dessas moléculas entre as espécies e a identificar biomarcadores. Com base nisso, a hipótese desse projeto parte da ideia de que é possível realizar uma filogenia dos receptores nucleares de espécies marinhas baseando-se nas sequências dos domínios DBD e LBD, comparando essa abordagem com a filogenia baseada em sequências completas, em conjunto com os dados estruturais desses domínios proteicos.

## 1.1 OBJETIVOS

### 1.1.1 Objetivo Geral

Realizar filogenias dos genes da superfamília dos receptores nucleares baseadas em sequências de seus domínios DBD e LBD e sequências completas de espécies aquáticas de interesse ambiental, bem como criar modelos tridimensionais dos domínios DBD e LBD.

### 1.1.2 Objetivos Específicos

- Identificar os domínios DBD e LBD em todas as sequências de receptores nucleares selecionadas para este trabalho;
- Realizar análises filogenéticas a partir das sequências dos domínios DBD e LBD e de suas sequências completas;

- Analisar a conservação de sítios e motivos de interesse nos domínios DBD e LBD;
- Comparar as estruturas tridimensionais dos domínios e regiões de interesse.

## 2 MATERIAIS E MÉTODOS

### 2.1 ANÁLISE FILOGENÉTICA

Foram utilizados dados de genomas completos e conjuntos de proteínas provenientes de um projeto de pesquisa de nosso grupo (AMORIM, 2021), referentes a 11 espécies de invertebrados e vertebrados aquáticos, e *Homo sapiens* como conjunto de referência. Todos os dados foram coletados do banco de dados público Ensembl ([ensembl.org/index.html](http://ensembl.org/index.html)), para as espécies *Homo sapiens*, *Delphinapterus leucas* (beluga), *Physeter catodon* (cachalote), *Tursiops truncatus* (golfinho-nariz-de-garrafa), *Oncorhynchus mykiss* (truta-arco-íris), *Danio rerio* (peixe-zebra), *Octopus bimaculoides*, *Lottia gigantea*, *Crassostrea gigas* (ostra-do-pacífico), *Daphnia magna*, *Lingula anatina* e *Strongylocentrotus purpuratus* (ouriço-do-mar). Além disso, os dados utilizados no presente trabalho passaram por etapas de processamento, inferência de ortologia e anotação dos genes. A partir disso, as sequências referentes a receptores nucleares de interesse ambiental foram recuperadas e organizadas em bancos de dados específicos para cada família de receptor, sendo elas: Receptor de Ácido Retinoico, Receptor X do Fígado, *Farnesoid X Receptor* ou Receptor de Ácido Biliar, Receptor de Vitamina D, Receptor Pregnano X, Fator Nuclear Hepático 4, Receptor Retinoide X, Receptor de Estrógeno e Receptor Relacionado ao Estrógeno.

Tendo finalizado esta primeira etapa de organização dos dados, foi realizado o alinhamento múltiplo de sequências aminoacídicas de cada família de receptor nuclear, pela plataforma Clustal Omega (SIEVERS et al., 2011), seguindo as especificações recomendadas. Os alinhamentos gerados foram analisados utilizando o *software* Topali v2.5 (MILNE et al., 2009), no qual mais sequências foram filtradas manualmente, principalmente aquelas que possuíam regiões com grandes *gaps* ou segmentos importantes faltando, como as porções N-terminal/C-terminal por exemplo. Após esta etapa de curadoria, outro alinhamento foi feito para cada família, bem como um alinhamento geral abordando todas as famílias estudadas, seguindo a mesma metodologia que o primeiro alinhamento. No Topali v2.5 também foi realizada a análise de melhor modelo de substituição para as análises filogenéticas, habilitando a Distribuição Gama e sítios invariáveis para inferência bayesiana. O melhor modelo de substituição sugerido para todos os alinhamentos foi JTT + G (JONES; TAYLOR; THORNTON, 1992) de acordo com o método de critério de informação bayesiano (BIC - *Bayesian Information Criterion*).

As análises filogenéticas subsequentes foram feitas utilizando o programa MrBayes v3.2.6 (HUELSENBECK; RONQUIST, 2001), com as seguintes especificações: -mcmc

nchains=1 ngen=5000000 samplefreq=5000 diagnfreq=5000 printfreq=5000. A taxa de *burn-in* foi de 25%, como recomendado pelo manual do programa. Para visualizar a árvore filogenética gerada pela análise, foi utilizado o *software* FigTree v1.4.3 ([tree.bio.ed.ac.uk/software/figtree/](http://tree.bio.ed.ac.uk/software/figtree/)).

## 2.2 ANÁLISE DE CONSERVAÇÃO DE AMINOÁCIDOS

Após realizar as análises filogenéticas com as sequências completas, foi feita a procura das regiões relativas aos domínios de ligação ao DNA (DBD - *DNA Binding domain*) e ao de ligação com o ligante (LBD - *Ligand binding domain*) para cada sequência de receptor nuclear. A identificação desses domínios foi possível por meio da plataforma ExPASy - PROSITE (<https://prosite.expasy.org>), o qual possui documentação de diversos domínios proteicos, sítios funcionais e famílias de proteínas (SIGRIST et al., 2013). As porções identificadas das sequências utilizadas como entrada foram separadas em bancos de dados específicos para cada domínio e posteriormente alinhadas pela plataforma Clustal Omega (SIEVERS et al., 2011). O alinhamento então foi submetido ao *software* WebLogo 3 ([weblogo.threeplusone.com](http://weblogo.threeplusone.com); CROOKS et al., 2004) a fim de criar uma *sequence logo* para cada domínio, possibilitando uma representação gráfica dos alinhamentos e identificação de regiões e resíduos mais ou menos conservadas, bem como de motivos nos domínios de interesse.

## 2.3 CRIAÇÃO DE MODELOS TRIDIMENSIONAIS DO DBD E LBD

Com a obtenção das *sequence logos*, foi possível identificar sequências de interesse para realizar a predição de um modelo estrutural dos domínios de ligação ao DNA e de ligação ao ligante. Para o DBD, foram escolhidas 3 sequências: aqu\_NR\_DBD, referente a *Amphimedon queenslandica*, espécie utilizada como *outgroup* na filogenia geral; dre\_ER\_DBD, pertencente ao grupo dos receptores de estrógeno de *D. rerio*; hsa\_VDR\_DBD, receptor de vitamina D de *H. sapiens*, pertencente ao grupo dos receptores relacionados a hormônios tireoides. Para as sequências de LBD, foram selecionadas uma para cada grupo de receptor nuclear, visto que este grupo apresentou maior diversidade de resíduos. As sequências utilizadas foram as seguintes: aqu\_NR\_LBD; hsa\_HNF4\_LBD (*Hepatocyte Nuclear Factor 4*); dre\_ER\_LBD (receptor de estrógeno); dle\_RAR\_LBD (receptor de ácido retinoico). Estas sequências foram então utilizadas como entrada na plataforma SWISS-MODEL (WATERHOUSE et al., 2018) para gerar os modelos de suas estruturas, tentando respeitar um limite de no mínimo 30% de identidade entre a sequência de entrada e o *template* proposto pelo programa. O modelo gerado

foi analisado por fim pelo *software* PyMol 2.0, onde foi possível visualizar e interagir com a estrutura proteica e também gerar figuras do modelo final.



### 3 RESULTADOS E DISCUSSÃO

#### 3.1 IDENTIFICAÇÃO DE FAMÍLIAS DE RECEPTORES NUCLEARES

Os dados provenientes das análises de ortologia anteriores indicaram a presença de diversas famílias de receptores nucleares nos conjuntos proteicos das espécies estudadas, são elas: RAR, LXR, FXR, VDR, PXR, HNF4, RXR, ER e ERR (AMORIM, 2021). Essas sequências foram organizadas em bancos de dados individuais para então passarem por um alinhamento múltiplo, além de processos de filtragem manual de sequências incompletas. Tendo as sequências utilizáveis em mãos, estas foram submetidas à plataforma BLAST (<https://blast.ncbi.nlm.nih.gov/Blast.cgi>) para confirmar sua identidade. Esta análise resultou em dois novos receptores nucleares não identificados na montagem – EcR (Receptor de Ecdisona) e HR96 (Receptor Hormonal 96), ambos presentes apenas em sequências de invertebrados (**Tabela 1**). O receptor de ecdisona foi primeiramente documentado em insetos (KOELLE et al., 1991), o qual desempenha função de controle da metamorfose desses animais quando atua em conjunto com a USP (*Ultraspiracle protein*), homóloga do RXR nos invertebrados, formando o complexo EcR/USP (DHADIALLA, 2005; ORO; MCKEOWN; EVANS, 1990). Até o momento, a maioria dos genes referentes a EcR foram identificados em artrópodes, principalmente insetos e crustáceos, porém alguns genes também foram encontrados em moluscos e anelídeos (LAGUERRE; VEENSTRA, 2010). Nos dados obtidos neste trabalho, foram identificadas sequências de EcR nas espécies *Octopus bimaculoides*, *Lottia gigantea*, *Crassostrea gigas*, *Lingula anatina*, *Daphnia magna* e *Strongylocentrotus purpuratus*, ou seja, moluscos, crustáceos e equinodermos. Porém, de acordo com LAFONT; KOOLMAN (2009), a ocorrência desses receptores fora do grupo dos artrópodes não reflete necessariamente a presença natural de ecdisteroides (hormônios responsáveis pela ecdise em artrópodes) nestes táxons, mas talvez indique uma ação contaminante desses hormônios em outros invertebrados. Outro ponto importante que pode ser levado em consideração é que tanto moluscos quanto anelídeos pertencem ao superfilo Lophotrochozoa, grupo irmão de Ecdysozoa, no qual estão os artrópodes. Portanto, é possível que o ancestral comum destes dois grupos possuísse um receptor similar ao EcR, mas com função diferente, visto que a ecdise é exclusiva de Ecdysozoa, sugerindo talvez um processo de exaptação para esse ganho de função nos ecdisozoários. Isto é reforçado pela presença de sequências identificadas como de EcR em equinodermos, grupo bastante derivado da linhagem animal, indicando que existia um receptor ancestral antes do processo de separação desses grupos.

**Tabela 1.** Relação dos receptores nucleares abordados neste trabalho e seus ligantes.

Nomenclatura	Nome	Abreviação	Ligantes
NR1B1,2	Receptor de ácido retinoico	RAR	ácido trans retinóico
NR1H1	Receptor de ecdisona	EcR	ecdisteroides
NR1H2,3	Receptor X do fígado	LXR	oxisteróis
NR1H4,5	Receptor de ácido biliar	FXR	ácidos biliares
NR1I1	Receptor de vitamina D	VDR	vitamina D
NR1I2	Receptor pregnano X	PXR	xenobióticos
NR1J1	Receptor hormonal 96	HR96	xenobióticos
NR2A1,2	Fator nuclear hepático 4	HNF4	ácidos graxos
NR2B1-3	Receptor retinoide X	RXR	Ácido retinóico
NR3A1,2	Receptor de estrógeno	ER	estrógeno
NR3B1-3	Receptor relacionado ao estrógeno	ERR	x

X = ligante desconhecido.

Em relação ao receptor hormonal 96, acredita-se que este seja um ortólogo dos receptores CAR, PXR e VDR, presentes em vertebrados (BERTRAND et al., 2004), descoberto primeiramente em *Drosophila melanogaster* (FISK; THUMMEL, 1995), e mais recentemente, ortólogos deste receptor foram encontrados em outras espécies de insetos e nematódeos (ANTEBI et al., 2000; VELARDE; ROBINSON; FAHRBACH, 2006). Neste trabalho, foram encontradas sequências identificadas como sendo de HR96 nas espécies *O. bimaculoides*, *L. gigantea*, *C. gigas*, *L. anatina* e *D. magna*, mostrando que esse receptor não é exclusivo dos ecdisozoários, embora sua função nos moluscos ainda não esteja elucidada. Nos insetos, sabe-se que este receptor tem atuação no metabolismo de lipídios e colesteróis (KARIMULLINA et al., 2012), porém, em outros táxons menos estudados seus ligantes permanecem desconhecidos, o que caracteriza o receptor HR96 como um receptor órfão. Além disso, visto que HR96 possui homologia com CAR, PXR e VDR, conhecidos por atuarem também no metabolismo de colesteróis (HORNER et al., 2009), pode-se esperar uma linha de ação parecida nos moluscos, dada a importância desse processo metabólico para a manutenção da integridade celular e sinalização hormonal. Porém, devido à sua presença em diversas vias metabólicas, este receptor também foi caracterizado como promíscuo, pois é ativado por uma série de diferentes ligantes (KARIMULLINA et al., 2012), o que requer mais estudos para explicar melhor seus mecanismos de ação e suas moléculas ativadoras. No que diz respeito à evolução de HR96, a presença deste receptor em Lophotrochozoa indica a existência de um receptor no ancestral

comum entre lofotrocozoários e ecdisozoários, antes da separação entre Prostostomia e Deuterostomia, não sendo então exclusivo de insetos ou ecdisozoários no geral.

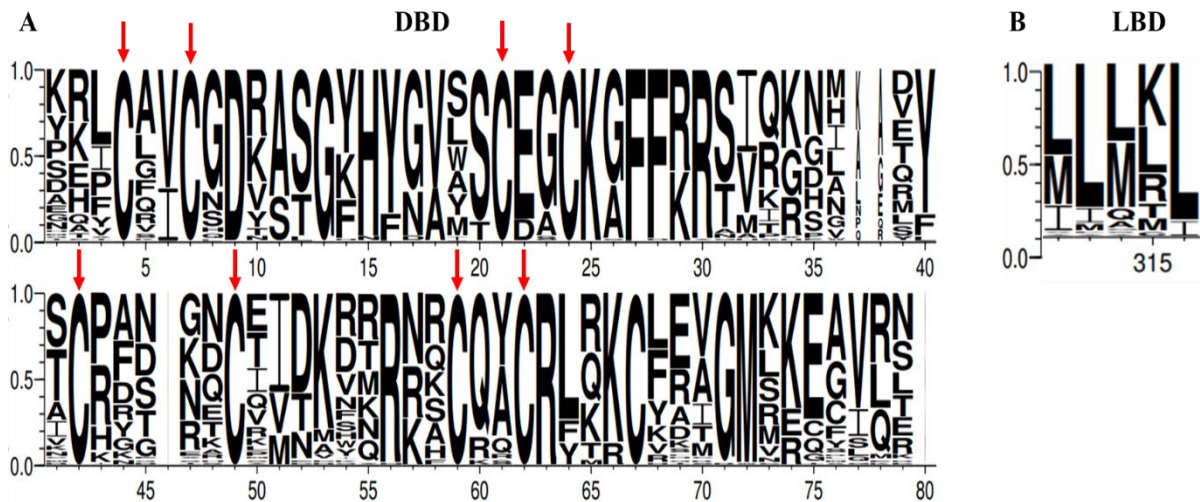
### 3.2 IDENTIFICAÇÃO E CARACTERIZAÇÃO DOS DOMÍNIOS DBD E LBD

A identificação dos domínios de ligação ao DNA e de ligação ao ligante, feita utilizando a plataforma PROSITE, reconheceu ambos os domínios em todas as sequências, com exceção do receptor HNF4 de *T. truncatus* e HR96 de *L. gigantea*, nos quais não foi possível identificar o DBD, provavelmente por algum erro de sequenciamento na região deste domínio. Ambos os domínios possuem tamanho esperado, de 70 a 80 aminoácidos para o DBD e 170 a 250 aminoácidos para o LBD (ZHAO et al., 2015).

Em sequência, o alinhamento e a montagem de *sequence logos* de cada domínio indicaram posições bastante conservadas, principalmente no DBD, como esperado (**Figura 3**). Pode-se dar destaque aos resíduos de cisteína presentes neste domínio, essenciais na composição das duas estruturas de dedo de zinco, responsáveis pela dimerização dos receptores, bem como de interface de interação com elementos de resposta na molécula de DNA (RENAUD; MORAS, 2000). Os dedos de zinco dos DBD encontrados seguem o padrão de quatro cisteínas por estrutura, totalizando oito cisteínas envolvidas na estrutura deste motivo. Além disso, alguns outros resíduos apresentaram alto nível de conservação, como por exemplo, resíduos carregados como histidina (H15), lisina (K25, 66), ácido aspártico (D9) e arginina (R30, 56, 63). Embora provavelmente esses resíduos possuam alguma interação com outras moléculas devido às suas cargas, mais estudos são necessários para indicar suas funções na estrutura deste domínio. Em relação ao LBD, foi possível identificá-lo em todas as sequências utilizadas e este apresentou um perfil menos conservado, como era esperado também. Entre os resíduos mais bem conservados estão os de leucina principalmente, seguidos de aminoácidos polares como arginina, ácido aspártico, glutamina, lisina e ácido glutâmico, que assim como no DBD, podem estar envolvidos em reações de ligação eletrostática com os ligantes ou para auxiliar na conformação da proteína. Novamente, mais estudos são necessários a fim de averiguar a atuação desses resíduos. Porém, em relação à riqueza de resíduos de leucina, uma das razões pode ser a presença do motivo LLxxL/LxxLL (LXD), onde “L” é o resíduo de leucina e “x” pode ser qualquer aminoácido (SAVKUR; BURRIS; BURRIS, 2004). Esse motivo foi detectado primeiramente em proteínas coativadoras de receptores nucleares, com a função de reconhecimento entre as proteínas, porém SAVKUR et al. (2004) o descobriram também nos próprios receptores. Ao observar o alinhamento das sequências do LBD, é possível

identificar em todas as famílias a presença desse motivo, embora haja variações na sua sequência e posição. Uma das causas para essas diferenças na sequência do motivo (e do LBD como um todo) pode ser a especificidade de cada receptor e proteína coativadora, visto que este domínio possui uma gama maior de ligantes, em comparação com o DBD que se liga apenas ao DNA. É principalmente pelas diferenças no LBD que observamos a diversidade de receptores nucleares, pois cada família se liga a um ou mais ligantes diferentes, e para que isso ocorra, elas não podem ser idênticas entre si. Isto é bem ilustrado pela *sequence logo* deste domínio, que permite visualizar com mais clareza regiões com menos diversidade e mais conservação e vice-versa. Porém, como as posições dos motivos podem variar entre as famílias de receptores, visualizar as *sequence logos* de cada família pode facilitar a comparação entre os motivos (APÊNDICE A). Além disso, algumas famílias de RN possuem *gaps* (lacunas) em suas sequências, o que diminui a qualidade da *logo* de forma geral, pois ela ilustra tanto a conservação quanto a frequência relativa dos resíduos, tornando-os mais finos na imagem de acordo com a sua frequência. Portanto, analisar as *sequence logos* das famílias de RN individualmente permite uma melhor resolução das sequências completas, além de poder também inferir qual motivo pertence a qual família. Ao realizar então essa observação individual do LBD, foi possível perceber melhor a distribuição dos motivos até mesmo dentro das próprias famílias de receptores, por exemplo: a família HNF-4 possui o motivo ILLxL nas sequências de invertebrados e o motivo LLLLL nas de vertebrados, indicando claramente um processo mutacional que resultou na troca de uma isoleucina por uma leucina. Esse processo é observado em outras famílias como a do receptor de estrógeno, na qual as sequências de invertebrados apresentam o motivo xLLLL, enquanto em vertebrados o padrão observado é LLxLL. Essas mudanças podem refletir a diferença de ligantes ou coativadores dentro de uma mesma família de receptores, visto que vertebrados e invertebrados se separaram há pelo menos 500 milhões de anos (HOLLAND; CHEN, 2001) e os receptores nucleares surgiram antes disso, dada sua presença em animais bastante primitivos como esponjas (LARROUX et al., 2006).

**Figura 3.** Sequence logos de regiões mais conservadas do DBD e LBD.



Posições conservadas dos domínios DBD e LBD. A) Resíduos de aminoácidos envolvidos na formação dos dedos de zinco. Resíduos de cisteína estão destacados por setas vermelhas acima deles. B) Motivo LxxLL/LLxxL encontrado no LBD. Eixo X indica a probabilidade de ocorrência dos resíduos, numa escala de 0 a 1. O eixo Y indica a posição dos resíduos no alinhamento múltiplo de sequências. A espessura dos símbolos representa a frequência relativa dos resíduos aminoácidos. Fonte: Elaborado pelo autor com auxílio da plataforma Web Logo 3 (2022).

Além disso, sabe-se que os domínios proteicos tendem a se conservar melhor do que o restante da sequência aminoacídica (SJÖLANDER, 2010) e embora as moléculas que interagem com os receptores possam ter mudado durante o curso da evolução, poucas mudanças nos resíduos aminoacídicos podem ser o suficiente para conferir essa alteração de afinidade, preservando ainda a identidade do domínio e do motivo. Outro ponto importante a ser levado em consideração é o de que as proteínas muitas vezes são nomeadas de acordo com a similaridade que possuem com sequências descobertas anteriormente, logo, não necessariamente o receptor de estrógeno da ostra *C. gigas*, por exemplo, se liga ao estrógeno na natureza, mesmo que seu LBD seja bastante similar ao de ER de vertebrados. Estudos (KEAY; BRIDGHAM; THORNTON, 2006; SCOTT, 2012) sugerem que de fato, os moluscos não produzem esteroides similares aos de vertebrados, obtendo-os por meio do ambiente em que vivem, sugerindo que seus receptores de esteroides possuam a função de metabolizar hormônios exógenos ou que sequer se liguem a essas moléculas. Além disso, genes envolvidos na síntese de hormônios esteroides não foram encontrados em moluscos, ao contrário do que é observado em espécies de vertebrados (SCOTT, 2012). Em insetos, essa ausência de homólogos de receptores de esteroides também foi documentada, sugerindo que esses receptores surgiram após a separação entre artrópodes e deuterostômios, com exceção da subfamília ERR (ESCRIVA; DELAUNAY; LAUDET, 2000; OWEN; ZELENT, 2000). Isso é reforçado pela

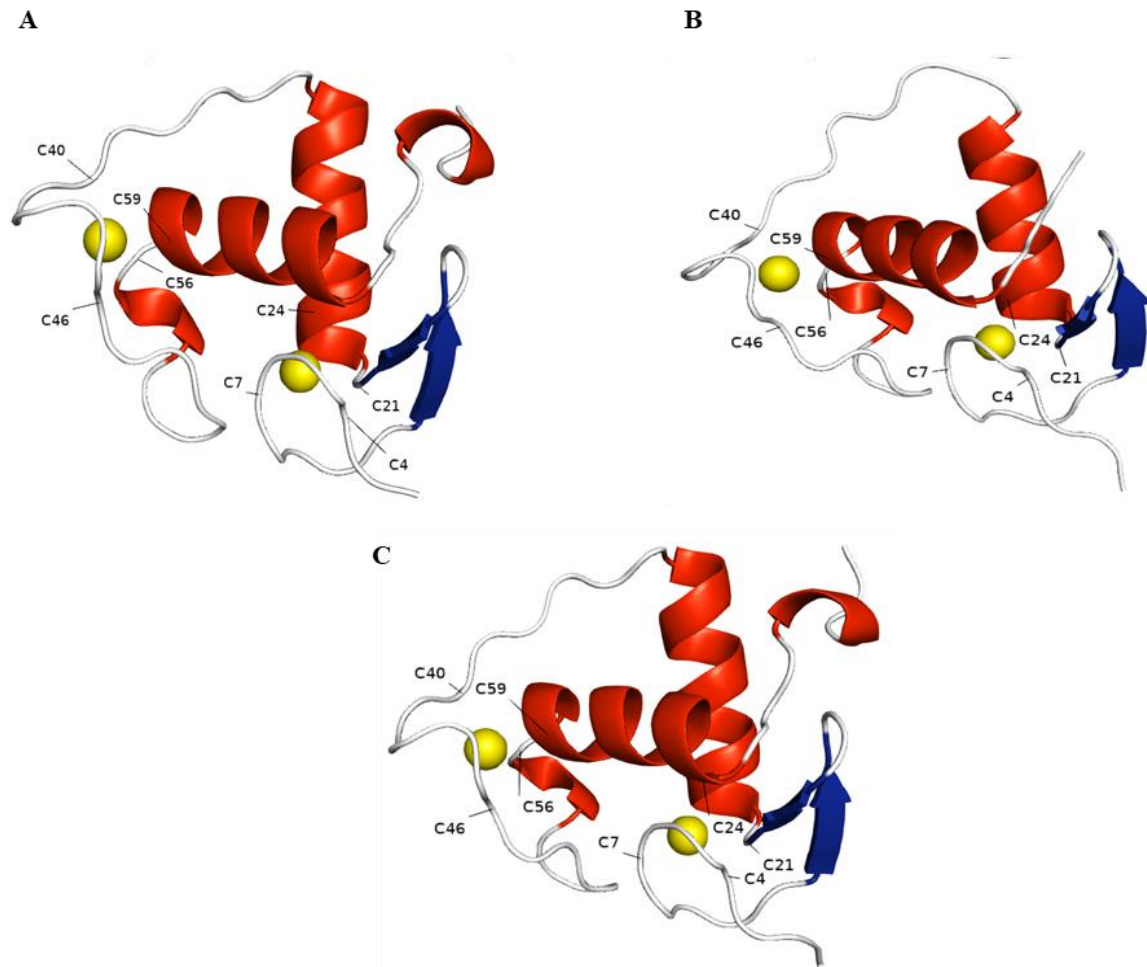
observação de que alguns hormônios presentes em vertebrados podem ser tóxicos para invertebrados, como o próprio estrógeno é para alguns insetos e crustáceos (BOVIER et al., 2018; OKKERMAN; GROSHART; PIJNENBURG, 2001). Vale lembrar também que não são apenas os hormônios naturais que acabam entrando em contato com esses animais e interagindo com seu domínio de ligação, mas também alguns desreguladores endócrinos, que incluem esteroides sintéticos. Porém, visto que a separação entre moluscos e artrópodes ocorreu logo após a divisão entre Prostomia e Deuterostomia, extrapolações de um clado para o outro podem estar equivocadas, fazendo necessário que estudos específicos para cada grupo sejam realizados para esclarecer a produção, absorção e efeito de hormônios esteroides em seus organismos.

Em relação às outras famílias de receptores, a subfamília NR2 que agrupa os receptores RXR e HNF-4 parece manter uma função similar ao longo da evolução, atuando em etapas do desenvolvimento tanto em invertebrados quanto em vertebrados (MIGLIOLI et al., 2021; VOGELER et al., 2016). Curiosamente, foi documentado que essa família de receptores nucleares está entre as primeiras a surgir no curso da evolução (surgindo antes da separação entre cnidários e bilaterais) e também possui a maior similaridade com possíveis receptores ancestrais (BERTRAND et al., 2004; BRIDGHAM et al., 2010). Outro ponto interessante observado em NR2, é o fato de RXR surgir antes dos animais bilaterais, visto que todos os seus parceiros de dimerização são da subfamília NR1 (RAR, LXR, VDR e PXR), exclusiva de Bilateria (MIGLIOLI et al., 2021). Essa observação pode sugerir novamente outro caso de exaptação ou a presença de um receptor ancestral desconhecido ao qual RXR se ligasse, visto que este não se liga a HNF-4, conhecido por realizar homodimerização (JIANG et al., 1995).

No que diz respeito ao LBD desse grupo, todas as sequências obtidas demonstram bastante similaridade tanto a nível de sequência como de estrutura, sugerindo a importância da conservação de sua função no desenvolvimento dos organismos. Outro ponto importante é a descoberta de que receptores ancestrais possuíam comportamento dependente de ligantes, com ácidos graxos sendo os possíveis ligantes ancestrais, similar a HNF-4 (BRIDGHAM et al., 2010). No que diz respeito ao surgimento dos domínios DBD e LBD, acredita-se que ambos tenham surgido juntos, estando presentes no receptor nuclear ancestral (OWEN; ZELEN, 2000). Dessa forma, utilizando uma sequência de receptor nuclear da esponja *Amphimedon queenslandica*, organismo que se situa na base da filogenia dos metazoários, é possível comparar seus domínios com aqueles mais recentes, como de humanos e mamíferos aquáticos. O resultado dessa comparação demonstrou alta semelhança estrutural entre o DBD e LBD de

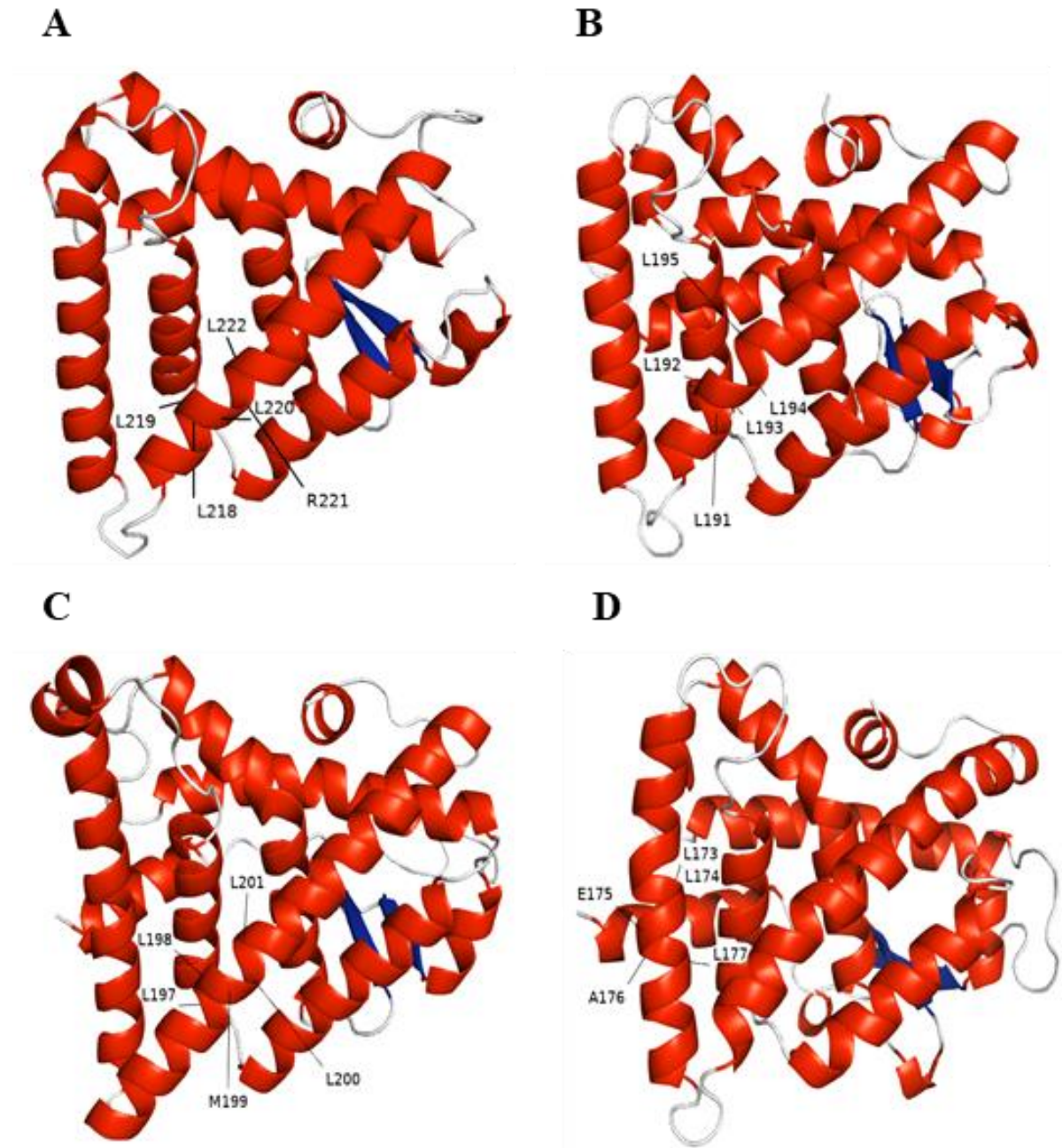
*A. queenslandica* com os domínios de vertebrados, apresentando o padrão clássico de dois dedos de zinco no DBD, bem como as 12  $\alpha$ -hélices, 2 folhas- $\beta$  e o motivo LLxxL no LBD, indicando que todas essas características estavam presentes desde o início da evolução dos receptores nucleares (**Figura 4-5**).

**Figura 4.** Conformação tridimensional do DBD de diferentes famílias e espécies.



Estruturas tridimensionais do DBD de diferentes famílias de NRs. A) Receptor nuclear de *A. queenslandica*. B) Receptor de estrógeno de *D. rerio*. C) Receptor de vitamina D de *H. sapiens*. Fonte: Elaborado pelo autor (2022).

**Figura 5.** Conformação tridimensional do LBD de diferentes famílias e espécies.



Estruturas tridimensionais do LBD de diferentes famílias de NRs. A) Receptor nuclear de *A. queenslandica*. B) Fator nuclear hepático 4 de *H. sapiens*. C) Receptor de estrógeno de *D. rerio*. D) Receptor de ácido retinoico de *D. leucas*. Fonte: Elaborado pelo autor (2022).



### 3.3 FILOGENIA BASEADA EM SEQUÊNCIAS COMPLETAS E EM DOMÍNIOS

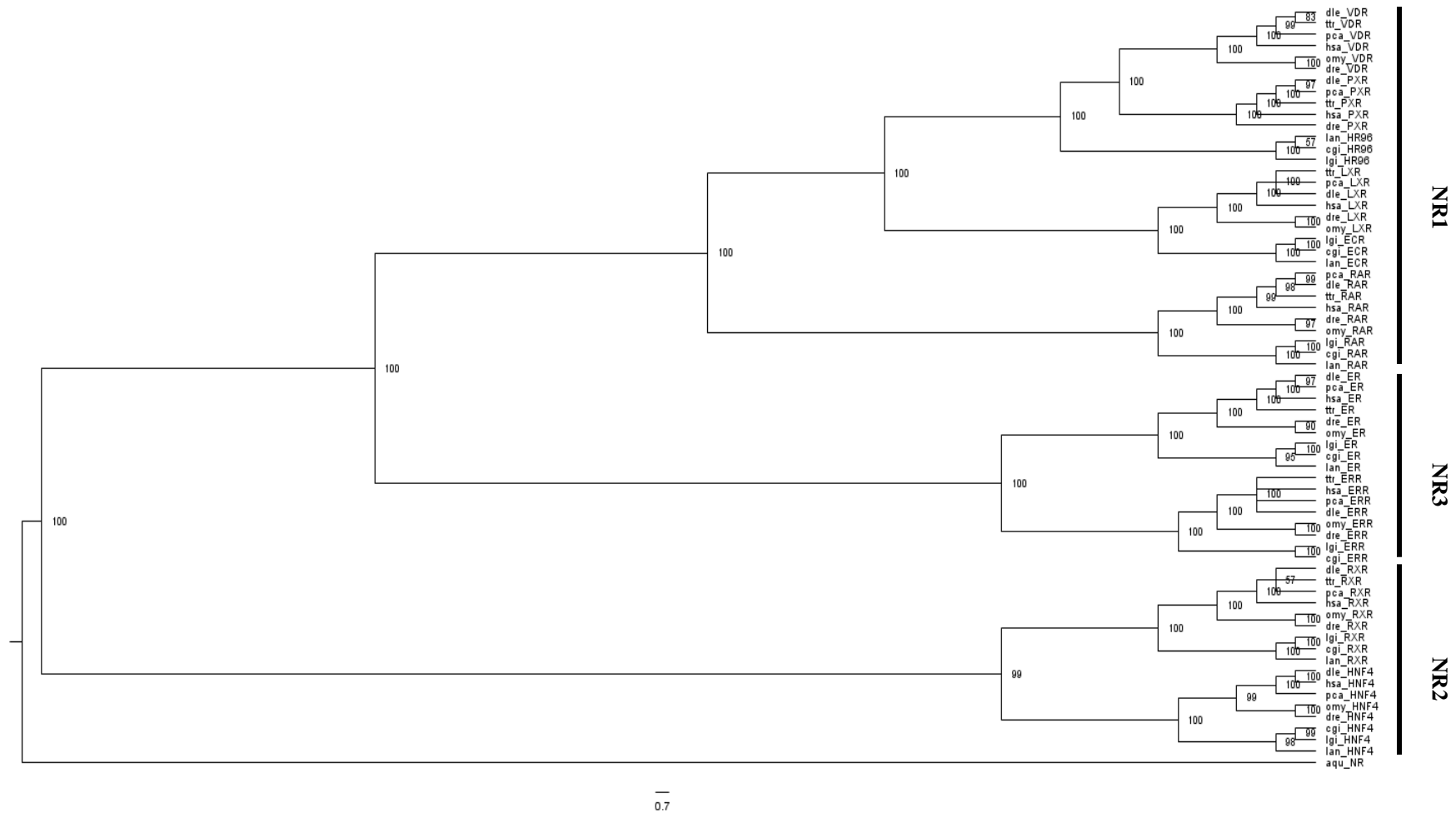
Análises filogenéticas baseadas tanto nas sequências completas dos receptores nucleares, quanto baseadas nas sequências de seus domínios DBD e LBD foram realizadas a fim de comparar as árvores resultantes de cada abordagem. Todas as análises foram feitas utilizando o método de inferência bayesiana, o qual é baseado numa cadeia de Markov Monte Carlo e que além de determinar a topologia da árvore, indica valores como idade e/ou probabilidade dos nós e arestas, bem como taxas de substituição de resíduos (HUELSENBECK; RONQUIST, 2001). As árvores geradas nessas análises podem também ser caracterizadas como filogramas, visto que o comprimento de suas arestas é proporcional ao número de substituições entre as arestas, além de sugerirem uma escala temporal de divergência entre as espécies relativa à raiz da árvore (VERBRUGGEN; THERIOT, 2008). Porém, dado que as taxas de substituição raramente são uniformes entre as arestas, é mais seguro assumir que o filograma seja interpretado da mesma forma que um cladograma, no que diz respeito ao tempo decorrido (DUNN, 2022).

A reconstrução da primeira árvore filogenética foi realizada utilizando sequências completas de RN de espécies aquáticas e *H. sapiens*, incluindo uma sequência da esponja *A. queenslandica*, utilizada como grupo externo (*outgroup*) nas análises. As famílias presentes na análise são aquelas citadas no item 4.1, membros dos grupos NR1 (relacionados a hormônios da tireoide), NR2 e NR3 (este último relacionado a hormônios esteroides). Para as análises filogenéticas, foram removidas espécies que não possuíam dados para todas as famílias de receptores, como *O. bimaculoides*, *D. magna* e *S. purpuratus*, totalizando 10 espécies.

O resultado da primeira análise gerou uma árvore filogenética com grupos bem definidos e coerentes (**Figura 6**). Todos os grupos se aglomeraram entre si, formando clados específicos. Além disso, os receptores EcR e HR96, homólogos de LXR e PXR/VDR respectivamente, situaram-se próximos a seus homólogos em vertebrados, formando grupos irmãos. A filogenia dos receptores seguiu o padrão observado em outros trabalhos (HOLZER; MARKOV; LAUDET, 2017; ROBINSON-RECHAVI; GARCIA; LAUDET, 2003), apresentando apenas alguns casos de politomia nas sequências de RXR, ERR e LXR, ou seja, nós onde não foi possível indicar uma relação dicotômica entre elas. Casos de politomia podem ocorrer por fatores como dados insuficientes ou má interpretação deles e também por eventos de rápida diversificação, tanto de espécies, quanto de caracteres ou genes (LIN; ZHANG; XU, 2011; MADDISON', 1989). Curiosamente, todos os casos de politomia observados nessa primeira reconstrução filogenética envolveram sequências de mamíferos aquáticos e humanos. Como

todas as sequências passaram anteriormente por processos de curadoria, anotação e inferência de ortologia, pode-se supor que as politomias observadas não ocorram devido à má qualidade das sequências. Duas hipóteses surgem então, a) os dados utilizados não foram suficientes, sendo necessário adicionar sequências de outras espécies para explicar melhor a história evolutiva desse grupo, ou b) os receptores envolvidos nos casos de politomia se separaram há pouco tempo, não havendo tempo suficiente para a acumulação de diferenças entre as suas sequências. Visto que para os outros grupos animais abordados, como os peixes ou invertebrados, os resultados foram coerentes, indicando suficiência de dados. Outros ramos da árvore filogenética dão suporte a esse ponto de vista, pois mostram a resolução das relações de parentesco entre receptores de outras famílias para os mamíferos, como é possível observar nos grupos HNF-4, ER, RAR, PXR e VDR. Em sequência, abordando a segunda hipótese, observa-se que a separação entre cetáceos e primatas ocorreu há pelo menos 90 milhões de anos (ARNASON; GULLBERG; JANKE, 1998). No que diz respeito à evolução dos cetáceos, os dados indicam que o grupo se separou dos artiodátilos em torno de 50-60 milhões de anos atrás (ARNASON; GULLBERG; JANKE, 1998; MCGOWEN; SPAULDING; GATESY, 2009), com a divergência entre Mysticeti e Odontoceti ocorrendo há cerca de 34 milhões de anos (MCGOWEN; SPAULDING; GATESY, 2009). Portanto, percebe-se que por mais que os receptores nucleares presentes nos mamíferos tenham um parentesco próximo, não é possível afirmar que divergiram recentemente, mesmo para uma superfamília de proteínas que existe há pelo menos 500 milhões de anos. Porém, dado o alto nível de conservação de sequência dos RN, foram realizados alinhamentos entre as sequências que apresentaram politomia para averiguar se mesmo após um longo tempo de separação desses grupos, esse padrão seria mantido (APÊNDICE B). Os resultados indicaram alta semelhança não só nos domínios DBD e LBD, como esperado, mas em praticamente toda a extensão das sequências, sugerindo que os casos de politomia possam se dar por não haver diferenças suficientes entre elas para estabelecer ramos dicotômicos.

**Figura 6.** Filogenia de famílias de RNs baseada em suas sequências completas.



Análise filogenética baseada nas sequências completas de receptores nucleares das famílias HNF4, RXR, ER, ERR, RAR, EcR, LXR, HR96, PXR e VDR. Os valores nos nós indicam a probabilidade de cada ramo ocorrer, numa escala de 0 a 100. As siglas indicam os grupos aos quais as famílias pertencem. Nomes das espécies estão abreviados da seguinte forma: aqu, *Amphimedon queenslandica*; lan, *Lingula anatina*; lgi, *Lottia gigantea*; cgi, *Crassostrea gigas*; dre, *Danio rerio*; omy, *Oncorhynchus mykiss*, hsa, *Homo sapiens*; dle, *Delphinapterus leucas*; ttr, *Tursiops truncatus*; pca, *Phiseter catodon*. Fonte: Elaborado pelo autor (2022).

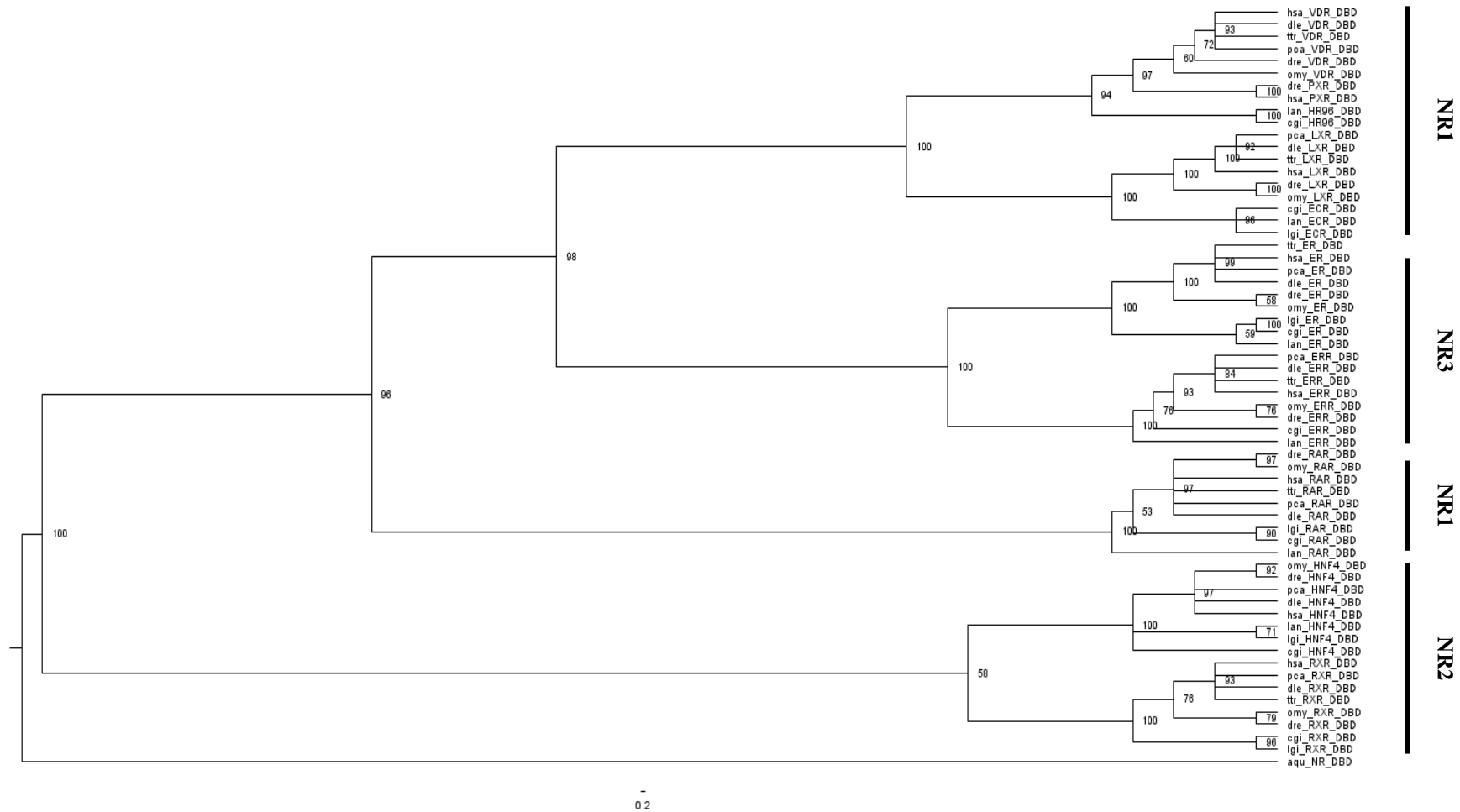


Em sequência, ao realizar análises filogenéticas baseadas nas sequências do DBD e LBD (**Figura 7-8**), observou-se novamente o mesmo padrão da análise anterior, inclusive em relação às politomias, com algumas adições. Entre os dois domínios, o DBD apresentou politomias em todas as sequências de receptores de mamíferos, enquanto o LBD manteve as politomias observadas anteriormente, com adição de sequências da família RAR para mamíferos. Esses dados reforçam a ideia de que não há diferenças suficientes nesses domínios em alguns ramos da filogenia para poder inferir relações mais precisas, principalmente para o DBD (conhecido por ser altamente conservado).

Em ambos os casos, todos os grupos de receptores permaneceram coesos e formaram ramos bem definidos, seguindo topologia igual ou similar à primeira. A única diferença mais significativa foi observada na filogenia do DBD, onde houve uma troca de posição entre RXR e HNF-4, porém, essa mudança não obteve valor de probabilidade alto. O ponto mais importante do grupo NR2, que é sua posição basal na filogenia, foi mantido apesar dessa troca. Em relação aos grupos NR1 e NR3, ambos mantiveram também suas posições e relações internas, reforçando a observação de alta conservação dos domínios de ligação dos RN com exceção da família RAR, que ficou situada entre os receptores de seu grupo e os receptores do grupo NR3. Esse resultado, embora similar entre as diferentes análises, reforça a hipótese de que filogenias baseadas em domínios podem dar resultados confiáveis e precisos, principalmente quando os domínios em questão não apresentam muita variabilidade ao longo do curso evolutivo, como é o caso do DBD e LBD, presentes e conservados em toda a linhagem de metazoários. Porém, em casos em que os domínios de uma proteína apresentem várias combinações diferentes ou estejam presentes apenas em linhagens específicas, torna-se muito mais desafiador inferir parentesco entre essas proteínas baseado apenas nos seus domínios (SJÖLANDER et al., 2011). Além disso, domínios distintos dentro de uma mesma proteína podem apresentar taxas de evolução bastante diferentes entre si (SJÖLANDER et al., 2011), o que pode dificultar uma análise baseada em domínios, dada a dissimilaridade entre eles. Apesar disso, os resultados aqui apresentados demonstram a conservação tanto a nível de sequência quanto de estrutura do DBD e LBD, sugerindo que a abordagem de filogenia baseada em domínios é eficiente e precisa para os receptores nucleares. Entre as principais vantagens de se trabalhar apenas com domínios está a redução de tempo necessário para observação dos dados, facilidade em manusear sequências menores e o menor custo computacional requerido para reconstruir filogenias baseadas em sequências menores.



**Figura 7.** Filogenia de famílias de RNs baseada na sequência do domínio de ligação ao DNA.

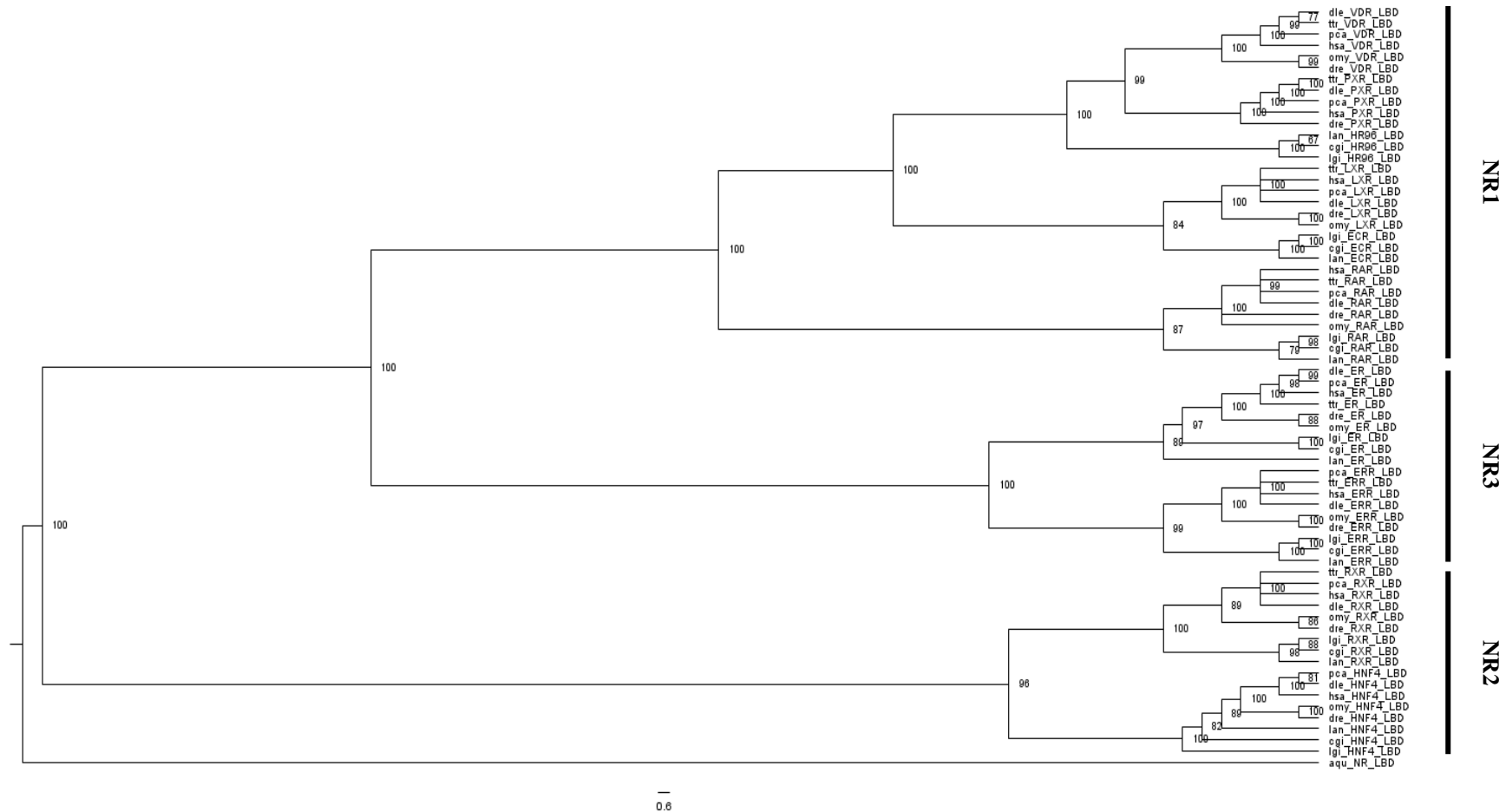


Análise filogenética baseada nas sequências dos DBDs das famílias HNF4, RXR, ER, ERR, RAR, EcR, LXR, HR96, PXR e VDR. Os valores nos nós indicam a probabilidade de cada ramo ocorrer, numa escala de 0 a 100. As siglas indicam o grupo ao qual cada família pertence. Nomes das espécies estão abreviados da seguinte forma: aqu, *Amphimedon queenslandica*; lan, *Lingula anatina*; lgi, *Lottia gigantea*; cgi, *Crassostrea gigas*; dre, *Danio rerio*; omy, *Oncorhynchus mykiss*; hsa, *Homo sapiens*; dle, *Delphinapterus leucas*; ttr, *Tursiops truncatus*; pca, *Phiseter catodon*. Fonte: Elaborado pelo autor (2022).





**Figura 8.** Filogenia de famílias de RNs baseada nas sequências do domínio de ligação ao ligante.



Análise filogenética baseada nas sequências dos LBDs das famílias HNF4, RXR, ER, ERR, RAR, EcR, LXR, HR96, PXR e VDR. Os valores nos nós indicam a probabilidade de cada ramo ocorrer, numa escala de 0 a 100. As siglas indicam o grupo ao qual cada família pertence. Nomes das espécies estão abreviados da seguinte forma: aqu, *Amphimedon queenslandica*; lan, *Lingula anatina*; lgi, *Lottia gigantea*; cgi, *Crassostrea gigas*; dre, *Danio rerio*; omy, *Oncorhynchus mykiss*; hsa, *Homo sapiens*; dle, *Delphinapterus leucas*; ttr, *Tursiops truncatus*; pca, *Phiseter catodon*. Fonte: Elaborado pelo autor (2022).



## 4 CONCLUSÃO

O presente trabalho apresentou a abordagem de filogenias baseadas em domínios como alternativa às metodologias baseadas em sequências completas, que pode ser bastante precisa e eficiente quando os domínios estudados possuem alto nível de conservação como é observado no DBD e LBD dos receptores nucleares. Para realizar o objetivo deste projeto, foram utilizadas sequências de nove espécies aquáticas e de *Homo sapiens*, das quais foram identificados e separados o domínio de ligação ao DNA (DBD) e o domínio de ligação ao ligante (LBD). Com esses dados, foram realizadas três análises filogenéticas, uma para as sequências completas e uma para cada domínio. Além disso, foram construídos modelos tridimensionais do DBD e LBD para receptores de diversas famílias e espécies, a fim de ilustrar as semelhanças e diferenças entre eles, bem como regiões conservadas e motivos de interesse. Os resultados indicaram bastante semelhança na topologia de todas as árvores filogenéticas, reforçando a precisão da abordagem baseada em domínios. As semelhanças se estenderam inclusive nos casos de politomia, que foram muito similares nas análises, ocorrendo exclusivamente em ramos de receptores de mamíferos. Isso pode se dar pela enorme semelhança entre as sequências de RN dessas espécies, como foi observado. Além disso, as topologias resultantes das análises registradas nesse projeto seguem o padrão observado em diversos outros trabalhos na literatura, indicando qualidade dos dados utilizados e precisão da metodologia baseada em domínios para analisar a evolução dos receptores nucleares. Entre as vantagens de se utilizar essa abordagem, tem-se a economia de tempo tanto para observar e manusear os dados, bem como o menor custo computacional para realizar as análises filogenéticas.

Por fim, uma possível sequência para o uso dos dados e resultados presentes neste trabalho, seria a realização de metodologias de atracamento molecular desses receptores nucleares com moléculas sintéticas similares a seus ligantes naturais, a fim de descobrir como se dá a interação dos poluentes com RN de espécies aquáticas. Com os resultados das análises filogenéticas, é possível se ter uma ideia de quais espécies podem estar sendo afetadas nos oceanos e rios e quais famílias de RN podem possuir biomarcadores de poluição, dada a conservação de seus domínios e preservação de suas estruturas proteicas, bem como as relações de parentesco entre eles, além de se ter mais segurança ao extrapolar os dados para outras espécies proximamente relacionadas.

## REFERÊNCIAS

ALTENHOFF, Adrian M.; DESSIMOZ, Christophe. Phylogenetic and Functional Assessment of Orthologs Inference Projects and Methods. **Plos Computational Biology**, [S.L.], v. 5, n. 1, p. 259-280, 16 jan. 2009. Public Library of Science (PLoS). <http://dx.doi.org/10.1371/journal.pcbi.1000262>.

ANTEBI, Adam; YEH, Wen-Hui; TAIT, Danilo; HEDGECOCK, Edward M.; RIDDLE, Donald L.. Daf-12 encodes a nuclear receptor that regulates the dauer diapause and developmental age in *C. elegans*. **Genes & Development**, [S.L.], v. 14, n. 12, p. 1512-1527, 15 jun. 2000. Cold Spring Harbor Laboratory. <http://dx.doi.org/10.1101/gad.14.12.1512>.

ARNASON, Ulfur; GULLBERG, Anette; JANKE, Axel. Molecular Timing of Primate Divergences as Estimated by Two Nonprimate Calibration Points. **Journal Of Molecular Evolution**, [S.L.], v. 47, n. 6, p. 718-727, dez. 1998. Springer Science and Business Media LLC. <http://dx.doi.org/10.1007/pl00006431>.

ARYAL, Niroj; WOOD, Jonae; RIJAL, Ishara; DENG, Dongyang; JHA, Manoj K.; OFORI-BOADU, Andrea. Fate of environmental pollutants: a review. **Water Environment Research**, [S.L.], v. 92, n. 10, p. 1587-1594, 6 ago. 2020. Wiley. <http://dx.doi.org/10.1002/wer.1404>.

BERTRAND, Stéphanie; BRUNET, Frédéric G.; ESCRIVA, Hector; PARMENTIER, Gilles; LAUDET, Vincent; ROBINSON-RECHAVI, Marc. Evolutionary Genomics of Nuclear Receptors: from twenty-five ancestral genes to derived endocrine systems. **Molecular Biology And Evolution**, [S.L.], v. 21, n. 10, p. 1923-1937, 30 jun. 2004. Oxford University Press (OUP). <http://dx.doi.org/10.1093/molbev/msh200>.

BOVIER, Tiziana Francesca; ROSSI, Sergio; MITA, Damiano Gustavo; DIGILIO, Filomena Anna. Effects of the synthetic estrogen 17- $\alpha$ -ethinylestradiol on *Drosophila melanogaster*: dose and gender dependence. **Ecotoxicology And Environmental Safety**, [S.L.], v. 162, p. 625-632, out. 2018. Elsevier BV. <http://dx.doi.org/10.1016/j.ecoenv.2018.07.020>.

BRIDGHAM, Jamie T.; EICK, Geeta N.; LARROUX, Claire; DESHPANDE, Kirti; HARMS, Michael J.; GAUTHIER, Marie E. A.; ORTLUND, Eric A.; DEGNAN, Bernard M.; THORNTON, Joseph W.. Protein Evolution by Molecular Tinkering: diversification of the nuclear receptor superfamily from a ligand-dependent ancestor. **Plos Biology**, [S.L.], v. 8, n.

10, p. 1-13, 5 out. 2010. Public Library of Science (PLoS). <http://dx.doi.org/10.1371/journal.pbio.1000497>.

BURGER, Joanna; GOCHFELD, Michael. Marine Birds as Sentinels of Environmental Pollution. **Ecohealth**, [S.L.], v. 1, n. 3, p. 263-275, 28 maio 2004. Springer Science and Business Media LLC. <http://dx.doi.org/10.1007/s10393-004-0096-4>.

CHEN, Feng; MACKEY, Aaron J.; VERMUNT, Jeroen K.; ROOS, David S.. Assessing Performance of Orthology Detection Strategies Applied to Eukaryotic Genomes. **Plos One**, [S.L.], v. 2, n. 4, p. 1-12, 18 abr. 2007. Public Library of Science (PLoS). <http://dx.doi.org/10.1371/journal.pone.0000383>.

CROOKS, Gavin E.; HON, Gary; CHANDONIA, John-Marc; BRENNER, Steven E.. WebLogo: a sequence logo generator. **Genome Research**, [S.L.], v. 14, n. 6, p. 1188-1190, jun. 2004. Cold Spring Harbor Laboratory. <http://dx.doi.org/10.1101/gr.849004>.

AMORIM, Ariane Nicaretta. **Identificação de grupos de genes ortólogos em modelos vertebrados e invertebrados**: características e análises comparativas. 2021. 140 f. Dissertação (Mestrado) - Curso de Programa de Pós-Graduação em Biologia Celular e do Desenvolvimento, Centro de Ciências Biológicas, Universidade Federal de Santa Catarina, Florianópolis, 2021.

DHADIALLA, T.s.; RETNAKARAN, A.; SMAGGHE, G.. Insect Growth- and Development-Disrupting Insecticides. **Comprehensive Molecular Insect Science**, [S.L.], p. 55-115, 2005. Elsevier. <http://dx.doi.org/10.1016/b0-44-451924-6/00076-4>.

EBELE, Anekwe Jennifer; ABDALLAH, Mohamed Abou-Elwafa; HARRAD, Stuart. Pharmaceuticals and personal care products (PPCPs) in the freshwater aquatic environment. **Emerging Contaminants**, [S.L.], v. 3, n. 1, p. 1-16, mar. 2017. Elsevier BV. <http://dx.doi.org/10.1016/j.emcon.2016.12.004>.

ESCRIVA, Hector; DELAUNAY, Franck; LAUDET, Vincent. Ligand binding and nuclear receptor evolution. **Bioessays**, [S.L.], v. 22, n. 8, p. 717-727, 24 jul. 2000. Wiley. [http://dx.doi.org/10.1002/1521-1878\(200008\)22:83.0.co;2-i](http://dx.doi.org/10.1002/1521-1878(200008)22:83.0.co;2-i).

FISK, G J; THUMMEL, C s. Isolation, regulation, and DNA-binding properties of three Drosophila nuclear hormone receptor superfamily members. **Proceedings Of The National Academy Of Sciences**, [S.L.], v. 92, n. 23, p. 10604-10608, 7 nov. 1995. Proceedings of the National Academy of Sciences. <http://dx.doi.org/10.1073/pnas.92.23.10604>

FITCH, Walter M.. Distinguishing Homologous from Analogous Proteins. **Systematic Zoology**, [S.L.], v. 19, n. 2, p. 99, jun. 1970. Oxford University Press (OUP). <http://dx.doi.org/10.2307/2412448>.

FRANCIS, Gordon A.; FAYARD, Elisabeth; PICARD, Frédéric; AUWERX, Johan. Nuclear Receptors and the Control of Metabolism. **Annual Review Of Physiology**, [S.L.], v. 65, n. 1, p. 261-311, mar. 2003. Annual Reviews. <http://dx.doi.org/10.1146/annurev.physiol.65.092101.142528>.

FULTON, Debra L; LI, Yvonne y; LAIRD, Matthew R; HORSMAN, Benjamin Gs; ROCHE, Fiona M; BRINKMAN, Fiona Sl. Improving the specificity of high-throughput ortholog prediction. **Bmc Bioinformatics**, [S.L.], v. 7, n. 1, p. 1-16, 28 maio 2006. Springer Science and Business Media LLC. <http://dx.doi.org/10.1186/1471-2105-7-270>.

GARCIA-REYERO, Natàlia; PERKINS, Edward J.. Systems biology: leading the revolution in ecotoxicology. **Environmental Toxicology And Chemistry**, [S.L.], v. 30, n. 2, p. 265-273, 23 dez. 2010. Wiley. <http://dx.doi.org/10.1002/etc.401>.

HAZEN, Elliott L; ABRAHMS, Briana; BRODIE, Stephanie; CARROLL, Gemma; JACOX, Michael G; SAVOCA, Matthew s; SCALES, Kylie L; SYDEMAN, William J; BOGRAD, Steven J. Marine top predators as climate and ecosystem sentinels. **Frontiers In Ecology And The Environment**, [S.L.], v. 17, n. 10, p. 565-574, 4 nov. 2019. Wiley. <http://dx.doi.org/10.1002/fee.2125>.

HOEKSTRA, P.F.; O'HARA, T.M.; FISK, A.T.; BORGÅ, K.; SOLOMON, K.R.; MUIR, D.C.G.. Trophic transfer of persistent organochlorine contaminants (OCs) within an Arctic marine food web from the southern Beaufort–Chukchi Seas. **Environmental Pollution**, [S.L.], v. 124, n. 3, p. 509-522, ago. 2003. Elsevier BV. [http://dx.doi.org/10.1016/s0269-7491\(02\)00482-7](http://dx.doi.org/10.1016/s0269-7491(02)00482-7).

HOLLAND, Nicholas D.; CHEN, Junyuan. Origin and early evolution of the vertebrates: new insights from advances in molecular biology, anatomy, and palaeontology. **Bioessays**, [S.L.], v. 23, n. 2, p. 142-151, 17 jan. 2001. Wiley. [http://dx.doi.org/10.1002/1521-1878\(200102\)23:23.0.co;2-5](http://dx.doi.org/10.1002/1521-1878(200102)23:23.0.co;2-5).

HOLZER, Guillaume; MARKOV, Gabriel V.; LAUDET, Vincent. Evolution of Nuclear Receptors and Ligand Signaling. **Current Topics In Developmental Biology**, [S.L.], p. 1-38, 2017. Elsevier. <http://dx.doi.org/10.1016/bs.ctdb.2017.02.003>.

HORNER, Michael A.; PARDEE, Keith; LIU, Suya; KING-JONES, Kirst; LAJOIE, Gilles; EDWARDS, Aled; KRAUSE, Henry M.; THUMMEL, Carl S.. The Drosophila DHR96

nuclear receptor binds cholesterol and regulates cholesterol homeostasis. **Genes & Development**, [S.L.], v. 23, n. 23, p. 2711-2716, 1 dez. 2009. Cold Spring Harbor Laboratory. <http://dx.doi.org/10.1101/gad.1833609>.

HUELSENBECK, J. P.; RONQUIST, F.. MRBAYES: bayesian inference of phylogenetic trees. **Bioinformatics**, [S.L.], v. 17, n. 8, p. 754-755, 1 ago. 2001. Oxford University Press (OUP). <http://dx.doi.org/10.1093/bioinformatics/17.8.754>.

VELARDE, Rodrigo A.; ROBINSON, Gene E.; FAHRBACH, Susan E.. Nuclear receptors of the honey bee: annotation and expression in the adult brain. **Insect Molecular Biology**, [S.L.], v. 15, n. 5, p. 583-595, out. 2006. Wiley. <http://dx.doi.org/10.1111/j.1365-2583.2006.00679.x>.

JANOSEK, J.; HILSCHEOVÁ, K.; BLÁHA, L.; HOLOUBEK, I.. Environmental xenobiotics and nuclear receptors—Interactions, effects and in vitro assessment. **Toxicology In Vitro**, [S.L.], v. 20, n. 1, p. 18-37, fev. 2006. Elsevier BV. <http://dx.doi.org/10.1016/j.tiv.2005.06.001>.

JIANG, G; NEPOMUCENO, L; HOPKINS, K; SLADEK, F M. Exclusive homodimerization of the orphan receptor hepatocyte nuclear factor 4 defines a new subclass of nuclear receptors. **Molecular And Cellular Biology**, [S.L.], v. 15, n. 9, p. 5131-5143, set. 1995. American Society for Microbiology. <http://dx.doi.org/10.1128/mcb.15.9.5131>.

JONES, David T.; TAYLOR, William R.; THORNTON, Janet M.. The rapid generation of mutation data matrices from protein sequences. **Bioinformatics**, [S.L.], v. 8, n. 3, p. 275-282, 1992. Oxford University Press (OUP). <http://dx.doi.org/10.1093/bioinformatics/8.3.275>

KARIMULLINA, Elina; LI, Yangchun; GINJUPALLI, Gautam K.; BALDWIN, William S.. Daphnia HR96 is a promiscuous xenobiotic and endobiotic nuclear receptor. **Aquatic Toxicology**, [S.L.], v. 116-117, p. 69-78, jul. 2012. Elsevier BV. <http://dx.doi.org/10.1016/j.aquatox.2012.03.005>.

KEAY, June; BRIDGHAM, Jamie T.; THORNTON, Joseph W.. The Octopus vulgaris Estrogen Receptor Is a Constitutive Transcriptional Activator: evolutionary and functional implications. **Endocrinology**, [S.L.], v. 147, n. 8, p. 3861-3869, 1 ago. 2006. The Endocrine Society. <http://dx.doi.org/10.1210/en.2006-0363>.

KOELLE, Michael R.; TALBOT, William S.; SEGRAVES, William A.; BENDER, Michael T.; CHERBAS, Peter; HOGNESS, David S.. The drosophila EcR gene encodes an

ecdysone receptor, a new member of the steroid receptor superfamily. **Cell**, [S.L.], v. 67, n. 1, p. 59-77, out. 1991. Elsevier BV. [http://dx.doi.org/10.1016/0092-8674\(91\)90572-g](http://dx.doi.org/10.1016/0092-8674(91)90572-g).

KRISHNA, S. S.. Structural classification of zinc fingers: survey and summary. **Nucleic Acids Research**, [S.L.], v. 31, n. 2, p. 532-550, 15 jan. 2003. Oxford University Press (OUP). <http://dx.doi.org/10.1093/nar/gkg161>.

KRISTENSEN, D. M.; WOLF, Y. I.; MUSHEGIAN, A. R.; KOONIN, E. V.. Computational methods for Gene Orthology inference. **Briefings In Bioinformatics**, [S.L.], v. 12, n. 5, p. 379-391, 19 jun. 2011. Oxford University Press (OUP). <http://dx.doi.org/10.1093/bib/bbr030>.

LA, David; SUTCH, Brian; LIVESAY, Dennis R.. Predicting protein functional sites with phylogenetic motifs. **Proteins: Structure, Function, and Bioinformatics**, [S.L.], v. 58, n. 2, p. 309-320, 30 nov. 2004. Wiley. <http://dx.doi.org/10.1002/prot.20321>.

LAFONT, René; KOOLMAN, Jan. Diversity of Ecdysteroids in Animal Species. **Ecdysone: Structures and Functions**, [S.L.], p. 47-71, 2009. Springer Netherlands. [http://dx.doi.org/10.1007/978-1-4020-9112-4\\_2](http://dx.doi.org/10.1007/978-1-4020-9112-4_2).

LAGUERRE, Michel; VEENSTRA, Jan A.. Ecdysone receptor homologs from mollusks, leeches and a polychaete worm. **Febs Letters**, [S.L.], v. 584, n. 21, p. 4458-4462, 8 out. 2010. Wiley. <http://dx.doi.org/10.1016/j.febslet.2010.10.004>.

LARROUX, Claire; FAHEY, Bryony; LIUBICICH, Danielle; HINMAN, Veronica F.; GAUTHIER, Marie; GONGORA, Milena; GREEN, Kathryn; WORHEIDE, Gert; LEYS, Sally P.; DEGNAN, Bernard M.. Developmental expression of transcription factor genes in a demosponge: insights into the origin of metazoan multicellularity. **Evolution**, [S.L.], v. 8, n. 2, p. 150-173, mar. 2006. Wiley. <http://dx.doi.org/10.1111/j.1525-142x.2006.00086.x>.

LIN, Guan Ning; ZHANG, Chao; XU, Dong. Polytomy identification in microbial phylogenetic reconstruction. **Bmc Systems Biology**, [S.L.], v. 5, n. 3, p. 1-11, dez. 2011. Springer Science and Business Media LLC. <http://dx.doi.org/10.1186/1752-0509-5-s3-s2>.

MADDISON, Wayne. RECONSTRUCTING CHARACTER EVOLUTION ON POLYTOMOUS CLADOGRAMS. **Cladistics**, [S.L.], v. 5, n. 4, p. 365-377, dez. 1989. Wiley. <http://dx.doi.org/10.1111/j.1096-0031.1989.tb00569.x>.

MCGOWEN, Michael R.; SPAULDING, Michelle; GATESY, John. Divergence date estimation and a comprehensive molecular tree of extant cetaceans. **Molecular Phylogenetics And Evolution**, [S.L.], v. 53, n. 3, p. 891-906, dez. 2009. Elsevier BV. <http://dx.doi.org/10.1016/j.ympev.2009.08.018>.



MIGLIOLI, Angelica; CANESI, Laura; GOMES, Isa D. L.; SCHUBERT, Michael; DUMOLLARD, Rémi. Nuclear Receptors and Development of Marine Invertebrates. **Genes**, [S.L.], v. 12, n. 1, p. 83, 11 jan. 2021. MDPI AG. <http://dx.doi.org/10.3390/genes12010083>.

MILNE, I.; LINDNER, D.; BAYER, M.; HUSMEIER, D.; MCGUIRE, G.; MARSHALL, D. F.; WRIGHT, F.. TOPALi v2: a rich graphical interface for evolutionary analyses of multiple alignments on hpc clusters and multi-core desktops. **Bioinformatics**, [S.L.], v. 25, n. 1, p. 126-127, 4 nov. 2008. Oxford University Press (OUP). <http://dx.doi.org/10.1093/bioinformatics/btn575>.

OKKERMAN, P. C.; GROSHART, C. P.. **Chemical study on estrogens**. Haia: Directoraat-Generaal Rijkswaterstaat, 2001.

MITISIS, Thanasis; PAPAGEORGIU, Louis; EFTHIMIADOU, Aspasia; BACOPOULOU, Flora; VLACHAKIS, Dimitrios; CHROUSOS, George; ELIOPOULOS, Elias. A comprehensive structural and functional analysis of the ligand binding domain of the nuclear receptor superfamily reveals highly conserved signaling motifs and two distinct canonical forms through evolution. **World Academy Of Sciences Journal**, [S.L.], p. 264-274, 2 jan. 2020. Spandidos Publications. <http://dx.doi.org/10.3892/wasj.2020.30>.

MOORE, Sue E.. Marine mammals as ecosystem sentinels. **Journal Of Mammalogy**, [S.L.], v. 89, n. 3, p. 534-540, jun. 2008. Oxford University Press (OUP). <http://dx.doi.org/10.1644/07-mamm-s-312r1.1>.

ORO, Anthony E.; MCKEOWN, Michael; EVANS, Ronald M.. Relationship between the product of the *Drosophila* ultraspiracle locus and the vertebrate retinoid X receptor. **Nature**, [S.L.], v. 347, n. 6290, p. 298-301, set. 1990. Springer Science and Business Media LLC. <http://dx.doi.org/10.1038/347298a0>.

OWEN, G. I.; \*, A. Zelent. Origins and evolutionary diversification of the nuclear receptor superfamily. **Cellular And Molecular Life Sciences (Cmls)**, [S.L.], v. 57, n. 5, p. 809-827, 1 maio 2000. Springer Science and Business Media LLC. <http://dx.doi.org/10.1007/s000180050043>.

QIAN, Haihong; HE, Ping; LV, Fenglin; WU, Wei. Genome-wide analysis of LXXLL-mediated DAX1/SHP–nuclear receptor interaction network and rational design of stapled LXXLL-based peptides to target the specific network profile. **International Journal Of Biological Macromolecules**, [S.L.], v. 129, p. 13-22, maio 2019. Elsevier BV. <http://dx.doi.org/10.1016/j.ijbiomac.2019.02.014>.

RENAUD, J. P.; MORAS\*, D.. Structural studies on nuclear receptors. **Cellular And Molecular Life Sciences**, [S.L.], v. 57, n. 12, p. 1748-1769, nov. 2000. Springer Science and Business Media LLC. <http://dx.doi.org/10.1007/pl00000656>.

ROBINSON-RECHAVI, Marc; GARCIA, Hector Escriva; LAUDET, Vincent. The nuclear receptor superfamily. **Journal Of Cell Science**, [S.L.], v. 116, n. 4, p. 585-586, 15 fev. 2003. The Company of Biologists. <http://dx.doi.org/10.1242/jcs.00247>.

SAVKUR, R.s.; BURRIS, T.P.. The coactivator LXXLL nuclear receptor recognition motif. **Journal Of Peptide Research**, [S.L.], v. 63, n. 3, p. 207-212, mar. 2004. Wiley. <http://dx.doi.org/10.1111/j.1399-3011.2004.00126.x>.

SCOTT, Alexander P.. Do mollusks use vertebrate sex steroids as reproductive hormones? Part I: critical appraisal of the evidence for the presence, biosynthesis and uptake of steroids. **Steroids**, [S.L.], v. 77, n. 13, p. 1450-1468, nov. 2012. Elsevier BV. <http://dx.doi.org/10.1016/j.steroids.2012.08.009>.

SIEVERS, Fabian; WILM, Andreas; DINEEN, David; GIBSON, Toby J; KARPLUS, Kevin; LI, Weizhong; LOPEZ, Rodrigo; MCWILLIAM, Hamish; REMMERT, Michael; SÖDING, Johannes. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. **Molecular Systems Biology**, [S.L.], v. 7, n. 1, p. 539, jan. 2011. EMBO. <http://dx.doi.org/10.1038/msb.2011.75>.

SIGRIST, Christian J. A.; CASTRO, Edouard de; CERUTTI, Lorenzo; CUCHE, Béatrice A.; HULO, Nicolas; BRIDGE, Alan; BOUGUELERET, Lydie; XENARIOS, Ioannis. New and continuing developments at PROSITE. **Nucleic Acids Research**, [S.L.], v. 41, n. 1, p. 344-347, 17 nov. 2012. Oxford University Press (OUP). <http://dx.doi.org/10.1093/nar/gks1067>.

SJÖLANDER, Kimmen. Getting Started in Structural Phylogenomics. **Plos Computational Biology**, [S.L.], v. 6, n. 1, p. 1-3, 29 jan. 2010. Public Library of Science (PLoS). <http://dx.doi.org/10.1371/journal.pcbi.1000621>.

SJOLANDER, K.; DATTA, R. S.; SHEN, Y.; SHOFFNER, G. M.. Ortholog identification in the presence of domain architecture rearrangement. **Briefings In Bioinformatics**, [S.L.], v. 12, n. 5, p. 413-422, 28 jun. 2011. Oxford University Press (OUP). <http://dx.doi.org/10.1093/bib/bbr036>.

SNAPE, Jason R.; MAUND, Steve J.; PICKFORD, Daniel B.; HUTCHINSON, Thomas H.. Ecotoxicogenomics: the challenge of integrating genomics into aquatic and

terrestrial ecotoxicology. **Aquatic Toxicology**, [S.L.], v. 67, n. 2, p. 143-154, abr. 2004. Elsevier BV. <http://dx.doi.org/10.1016/j.aquatox.2003.11.011>.

STEINBERG, Christian E.W.; STÜRZENBAUM, Stephen R.; MENZEL, Ralph. Genes and environment — Striking the fine balance between sophisticated biomonitoring and true functional environmental genomics. **Science Of The Total Environment**, [S.L.], v. 400, n. 1-3, p. 142-161, ago. 2008. Elsevier BV. <http://dx.doi.org/10.1016/j.scitotenv.2008.07.023>.

TOPOROVA, Lucia; BALAGUER, Patrick. Nuclear receptors are the major targets of endocrine disrupting chemicals. **Molecular And Cellular Endocrinology**, [S.L.], v. 502, p. 110665, fev. 2020. Elsevier BV. <http://dx.doi.org/10.1016/j.mce.2019.110665>.

VALLENDER, Eric J.. Bioinformatic approaches to identifying orthologs and assessing evolutionary relationships. **Methods**, [S.L.], v. 49, n. 1, p. 50-55, set. 2009. Elsevier BV. <http://dx.doi.org/10.1016/j.ymeth.2009.05.010>.

VERBRUGGEN, Heroen; THERIOT, Edward C.. Building trees of algae: some advances in phylogenetic and evolutionary analysis. **European Journal Of Phycology**, [S.L.], v. 43, n. 3, p. 229-252, ago. 2008. Informa UK Limited. <http://dx.doi.org/10.1080/09670260802207530>.

VOGELER, Susanne; BEAN, Tim P.; LYONS, Brett P.; GALLOWAY, Tamara S.. Dynamics of nuclear receptor gene expression during Pacific oyster development. **Bmc Developmental Biology**, [S.L.], v. 16, n. 1, p. 1-13, 29 set. 2016. Springer Science and Business Media LLC. <http://dx.doi.org/10.1186/s12861-016-0129-6>.

WÄRNMARK, Anette; TREUTER, Eckardt; WRIGHT, Anthony P. H.; GUSTAFSSON, Jan-Åke. Activation Functions 1 and 2 of Nuclear Receptors: molecular strategies for transcriptional activation. **Molecular Endocrinology**, [S.L.], v. 17, n. 10, p. 1901-1909, 1 out. 2003. The Endocrine Society. <http://dx.doi.org/10.1210/me.2002-0384>.

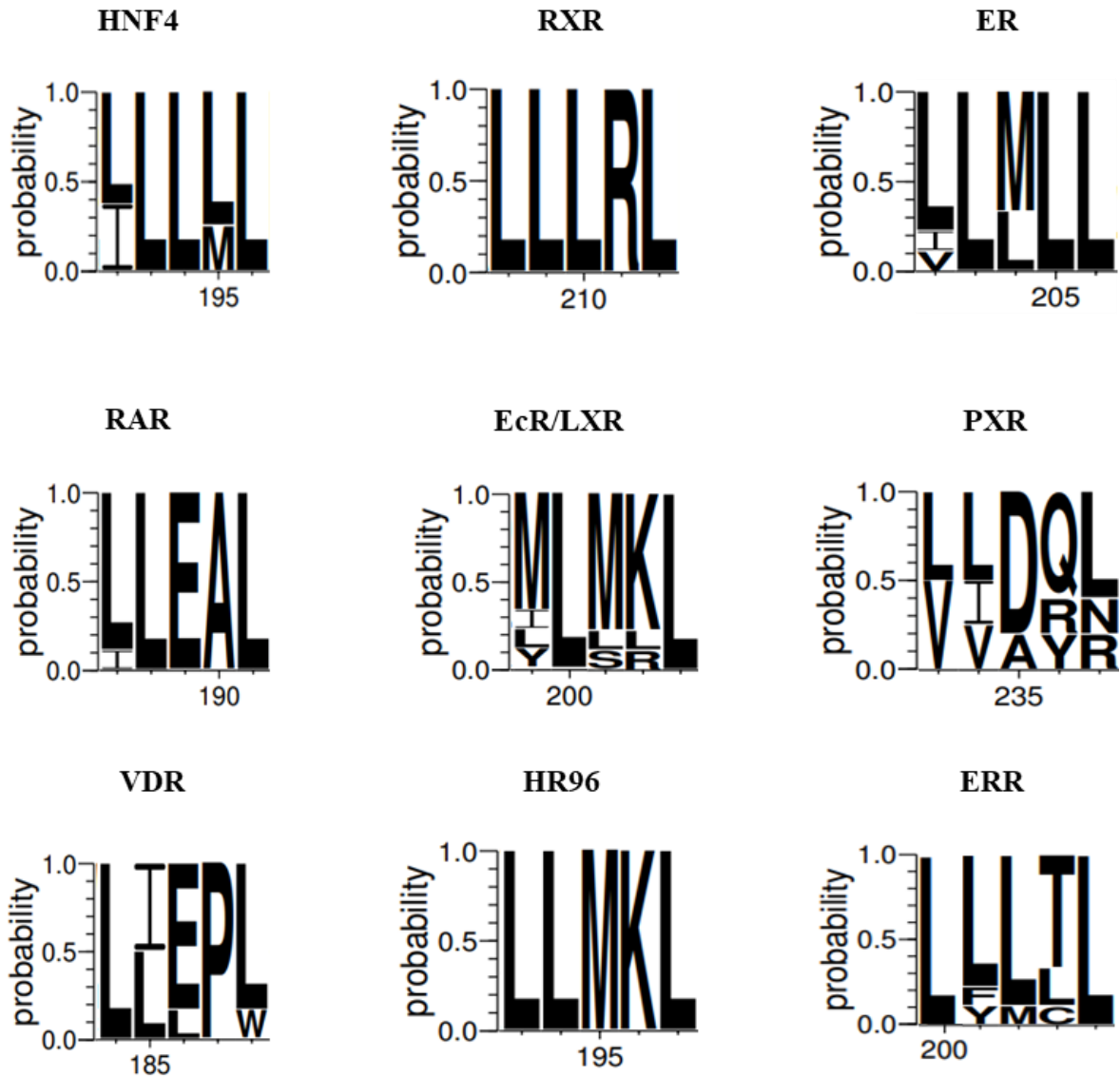
WATERHOUSE, Andrew; BERTONI, Martino; BIENERT, Stefan; STUDER, Gabriel; TAURIELLO, Gerardo; GUMIENNY, Rafal; HEER, Florian T; DE€BEER, Tjaart A P; REMPFER, Christine; BORDOLI, Lorenza. SWISS-MODEL: homology modelling of protein structures and complexes. **Nucleic Acids Research**, [S.L.], v. 46, n. 1, p. 296-303, 21 maio 2018. Oxford University Press (OUP). <http://dx.doi.org/10.1093/nar/gky427>.

YANG, Song; DOOLITTLE, Russell F.; BOURNE, Philip E.. Phylogeny determined by protein domain content. **Proceedings Of The National Academy Of Sciences**, [S.L.], v. 102, n. 2, p. 373-378, 3 jan. 2005. Proceedings of the National Academy of Sciences. <http://dx.doi.org/10.1073/pnas.0408810102>.

ZHAO, Yanbin; ZHANG, Kun; GIESY, John P.; HU, Jianying. Families of Nuclear Receptors in Vertebrate Models: characteristic and comparative toxicological perspective. **Scientific Reports**, [S.L.], v. 5, n. 1, p. 1-9, 25 fev. 2015. Springer Science and Business Media LLC. <http://dx.doi.org/10.1038/srep08554>.

## APÊNDICE A – MOTIVOS LXXLL/LLXXL DE DIFERENTES FAMÍLIAS DE RNS

**Figura 9.** Motivos conservados presentes no LBD de diferentes receptores nucleares.



Resíduos aminoacídicos conservados envolvidos nos motivos LxxLL/LLxxL de diferentes famílias de receptores nucleares. Eixo X indica probabilidade de ocorrência do resíduo, enquanto o eixo Y indica a posição dos resíduos no alinhamento múltiplo de seqüências.

## APÊNDICE B – ALINHAMENTOS DE SEQUÊNCIAS QUE APRESENTARAM POLITOMIA

**Figura 10.** Alinhamento entre as sequências do receptor retinoide X de mamíferos.

hsa_RXR	MSWAARPPFLPQRHAAGQCGPVGVRKEMHCGVASRWRRRRPWLDPAAAAA-AAVAGGEQQ	59
dle_RXR	MSWAARPPFLPQRHAAGQCGPVGVRKEMHCGVASRWRRRRPWLDPAAAAAAVAAGEQQ	60
ttr_RXR	-----	0
pca_RXR	MSWAARPPFLPQRHAAGQCGPVGVRKEMHCGVASRWRRRRPWLDPAAAAAAVAAGEQQ	60
hsa_RXR	TP--EPEPGEAGRDGMGDSGRDSRSPDSSSNPLPQGVPVPPSPPGPPLPSTAPSLGGSG	117
dle_RXR	TPE--PEPGEAGRDGMGDSGRDSRSPDSSSNPLPQGAAPPSPPGPPLPSSAAASLGGSG	118
ttr_RXR	-----SRSPDSSSNPLPQGAAPPSPPGPPLPSSAAASLGGSG	38
pca_RXR	TPEPEPEPGEAGRDGMGDSGRDSRSPDSSSNPLPQGAAPPSPPGPPLPSSAAASLGGSG	120
	*****.*****:*****	
hsa_RXR	APPPPPMPPLGSPFVVISSSMGSPGLPPAPPGFSGPVSSPQINSTVSLPGGGSGPPE	177
dle_RXR	APPPPPMPPLGSPFVVISSSMGSPGLPPAPPGFSGPVSSPQINSTVSLPGGGSGPPE	178
ttr_RXR	APPPPPMPPLGSPFVVISSSMGSPGLPPAPPGFSGPVSSPQINSTVSLPGGGSGPPE	98
pca_RXR	APPPPPMPPLGSPFVVISSSMGSPGLPPAPPGFSGPVSSPQINSTVSLPGGGSGPPE	180
	*****	
hsa_RXR	DVKPPVLGVRGLHCP PPPGGPGAGKRLCAICGDRSSGKHYGVYSCGCKGFFKRTIRKDL	237
dle_RXR	DVKPPVLGVRGLHCP PPPGGPGAGKRLCAICGDRSSGKHYGVYSCGCKGFFKRTIRKDL	238
ttr_RXR	DVKPPVLGVRGLHCP PPPGGPGAGKRLCAICGDRSSGKHYGVYSCGCKGFFKRTIR-DL	157
pca_RXR	DVKPPVLGVRGLHCP PPPGGPGAGKRLCAICGDRSSGKHYGVYSCGCKGFFKRTIRKDL	240
	***** **	
hsa_RXR	TYSCRDNKDCTVDKRQRNRCQYCRYQKCLATGMKREAVQEERQRGKDKDGDGEGAGGAP	297
dle_RXR	TYSCRDNKDCTVDKRQRNRCQYCRYQKCLATGMKREAVQEERQRGKDKDGDGEGAGGAP	298
ttr_RXR	TYSCRDNKDCTVDKRQRNRCQYCRYQKCLATGMKREAVQEERQRGKDKDGDGEGAGGAP	217
pca_RXR	TYSCRDNKDCTVDKRQRNRCQYCRYQKCLATGMKREAVQEERQRGKDKDGDGEGAGGAP	300
	*****	
hsa_RXR	EMPVDRILEAE LAVEQKSDQGVGPGGTAGSGSSPNDPVTNICQAADKQLFTLVWEAKRI	357
dle_RXR	EMPVDRILEAE LAVEQKSDQGVGPGGTAGSGSSPNDPVTNICQAADKQLFTLVWEAKRI	358
ttr_RXR	EMPVDRILEAE LAVEQKSDQGVGPGGTAGSGSSPNDPVTNICQAADKQLFTLVWEAKRI	277
pca_RXR	EMPVDRILEAE LAVEQKSDQGVGPGGTAGSGSSPNDPVTNICQAADKQLFTLVWEAKRI	360
	*****	
hsa_RXR	PHFSSLPLDDQVILLRAGWNELLIASFSHRSIDVRDGILLATGLHVHRNSAHSAGVGAIF	417
dle_RXR	PHFSSLPLDDQVILLRAGWNELLIASFSHRSIDVRDGILLATGLHVHRNSAHSAGVGAIF	418
ttr_RXR	PHFSSLPLDDQVILLRAGWNELLIASFSHRSIDVRDGILLATGLHVHRNSAHSAGVGAIF	337
pca_RXR	PHFSSLPLDDQVILLRAGWNELLIASFSHRSIDVRDGILLATGLHVHRNSAHSAGVGAIF	420
	*****	
hsa_RXR	DR---VLTELVS KM RDMRMDKTELGLRAIILFNPDAGLSNPSEVEVLREKVYASLET	473
dle_RXR	DR---VLTELVS KM RDMRMDKTELGLRAIILFNPDAGLSNPSEVEVLREKVYASLET	474
ttr_RXR	DRSLSRVLTELVS KM RDMRMDKTELGLRAIILFNPDAGLSNPSEVEVLREKVYASLET	397
pca_RXR	DRSLSRVLTELVS KM RDMRMDKTELGLRAIILFNPDAGLSNPSEVEVLREKVYASLET	480
	** *****	
hsa_RXR	YCKQKYPEQQGRFAKLLRLPALRSIGLKCLEHLFFFKLI GDTPIDTFLMEMLEAPHQLA	533
dle_RXR	YCKQKYPEQQGRFAKLLRLPALRSIGLKCLEHLFFFKLI GDTPIDTFLMEMLEAPHQLA	534
ttr_RXR	YCKQKYPEQQGRFAKLLRLPALRSIGLKCLEHLFFFKLI GDTPIDTFLMEMLEAPHQLA	457
pca_RXR	YCKQKYPEQQGRFAKLLRLPALRSIGLKCLEHLFFFKLI GDTPIDTFLMEMLEAPHQLA	540
	*****	

Alinhamento múltiplo de sequências de receptor retinoide X das espécies *D. leucas*, *P. catodon*, *T. truncatus* e *H. sapiens*. Regiões coloridas em azul claro representam a sequência do DBD, enquanto as regiões coloridas em verde representam a sequência do LBD.

**Figura 11.** Alinhamento entre as sequências do receptor relacionado ao estrógeno de mamíferos.

ttr_ERR	MSSQVVGIEPLYIKAEPASPDSPKGSSETETETEPVALAPGPAPTRCLPGHKEEEDGEGAG	60
hsa_ERR	MSSQVVGIEPLYIKAEPASPDSPKGSSETETETEPVALAPGPAPTRCLPGHKEEEDGEGAG	60
pca_ERR	MSSQVVGIEPLYIKAEPASPDSPKGSSETETETEPVALAPGPAPTRCLPGHKEEEDGEGAG	60
dle_ERR	MSSQVVGIEPLYIKAEPASPDSPKGSSETETETEPVALAPGPAPTRCLPGHKEEEDGEGAG	60
	*****	
ttr_ERR	PGEQGGGKLVLSLPPKRLCLVCGDVASGYHYGVASCEACKAFFKRTIQGSI EYSCPASNE	120
hsa_ERR	PGEQGGGKLVLSLPPKRLCLVCGDVASGYHYGVASCEACKAFFKRTIQGSI EYSCPASNE	120
pca_ERR	PGEQGGGKLVLSLPPKRLCLVCGDVASGYHYGVASCEACKAFFKRTIQGSI EYSCPASNE	120
dle_ERR	PGEQGGGKLVLSLPPKRLCLVCGDVASGYHYGVASCEACKAFFKRTIQGSI EYSCPASNE	120
	*****	
ttr_ERR	CEITKRRRKACQACRFTKCLRVGMLKEGVRLDRVRGGRQKYKRRPEVDPLFFPGFPFAGP	180
hsa_ERR	CEITKRRRKACQACRFTKCLRVGMLKEGVRLDRVRGGRQKYKRRPEVDPLFFPGFPFAGP	180
pca_ERR	CEITKRRRKACQACRFTKCLRVGMLKEGVRLDRVRGGRQKYKRRPEVDPLFFPGFPFAGP	180
dle_ERR	CEITKRRRKACQACRFTKCLRVGMLKEGVRLDRVRGGRQKYKRRPEVDPLFFPGFPFAGP	180
	*****	
ttr_ERR	LAVGGGPRKTAAPVNALVSHLLVVEPEKLYAMPDPAGPDGHLPAVATLCDLFDREIVVTI	240
hsa_ERR	LAVAGGPRKTAAPVNALVSHLLVVEPEKLYAMPDPAGPDGHLPAVATLCDLFDREIVVTI	240
pca_ERR	LAVAGGPRKT-APVNALVSHLLVVEPEKLYAMPDPAGPDGHLPAVATLCDLFDREIVVTI	239
dle_ERR	LAVAGGPRKT-APVNALVSHLLVVEPEKLYAMPDPAGPDGHLPAVATLCDLFDREIVVTI	239
	*** . *****	
ttr_ERR	SWAKSIPGFSSLSLSDQMSVLQSVWMEVVLVGV AQRS LPLQDELAF AEDLVLDEEGARAA	300
hsa_ERR	SWAKSIPGFSSLSLSDQMSVLQSVWMEVVLVGV AQRS LPLQDELAF AEDLVLDEEGARAA	300
pca_ERR	SWAKSIPGFSSLSLSDQMSVLQSVWMEVVLVGV AQRS LPLQDELAF AEDLVLDEEGARAA	299
dle_ERR	SWAKSIPGFSSLSLSDQMSVLQSVWMEVVLVGV AQRS LPLQDELAF AEDLVLDEEGARAA	299
	*****	
ttr_ERR	GLGELGAALLQLVRR LQALRLERE EYVLLKALALANS DSVHIEDAEAVEQLREALHEALL	360
hsa_ERR	GLGELGAALLQLVRR LQALRLERE EYVLLKALALANS DSVHIEDAEAVEQLREALHEALL	360
pca_ERR	GLGELGAALLQLVRR LQALRLERE EYVLLKALALANS DSVHIEDAEAVEQLREALHEALL	359
dle_ERR	GLGELGAALLQLVRR LQALRLERE EYVLLKALALANS DSVHIEDAEAVEQLREALHEALL	359
	*****	
ttr_ERR	EYEAGRAGPGGGAERRRAGRLLLLTLP LLRQTAGKVLAHFYGVKLEGVPMHKL FLEMLEA	420
hsa_ERR	EYEAGRAGPGGGAERRRAGRLLLLTLP LLRQTAGKVLAHFYGVKLEGVPMHKL FLEMLEA	420
pca_ERR	EYEAGRAGPGGGAERRRAGRLLLLTLP LLRQTAGKVLAHFYGVKLEGVPMHKL FLEMLEA	419
dle_ERR	EYEAGRAGPGGGAERRRAGRLLLLTLP LLRQTAGKVLAHFYGVKLEGVPMHKL FLEMLEA	419
	***** . *****	
ttr_ERR	MMD 423	
hsa_ERR	MMD 423	
pca_ERR	MMD 422	
dle_ERR	MMD 422	
	***	

Alinhamento múltiplo de sequências de receptor relacionado ao estrógeno das espécies *D. leucas*, *P. catodon*, *T. truncatus* e *H. sapiens*. Regiões coloridas em azul claro representam a sequência do DBD, enquanto as regiões coloridas em verde representam a sequência do LBD.

**Figura 12.** Alinhamento entre as sequências do receptor X do fígado de mamíferos.

hsa_LXR	MSSPTTSSLDTPLPGNGPQPAPSSSPTVKEEGPEPWPGGPDVPGTDEASSACSTDW	60
ttr_LXR	MSTPATSSLDTPLPGNGPST---PSSSPDGKEDGPEPCPG-ADPDVPGTDGANSAS--VV	54
pca_LXR	MSTPATSSLDTPLPGNGPST---PSSSPDGKEDGPEPCPG-ADPDVPGTDGANSAS--VV	54
dle_LXR	MSTPATSSLDTPLPGNGPST---PSSSPDGKEDGPEPCPG-ADPDVPGTDGANSAS--VV	54
	**.*:*****          ***** **.***** ** ***** *.**.	
hsa_LXR	VIPDPEEPPERKRKKGPAPKMLGHELCRVCGDKASGFHYNVLSCEGCKGFFRRSVVRGGA	120
ttr_LXR	VILDTAEEPERKRKKGPAPKMLGDELQVCGDTASGFHYNVLSCEGCKGFFRRSVIRGGA	114
pca_LXR	VILDTAEEPERKRKKGPAPKMLGDELQVCGDTASGFHYNVLSCEGCKGFFRRSVIRGGA	114
dle_LXR	VILDTAEEPERKRKKGPAPKMLGDELQVCGDTASGFHYNVLSCEGCKGFFRRSVIRGGA	114
	** * *****.***.***.*****.***	
hsa_LXR	RRYACRGGGTCQMDAFMRRKCQQCRLRKCKEAGMREQCVLSEEQIRKKKIRKQQQESQSQ	180
ttr_LXR	GRYACRGGGTCQMDAFMRRKCQQCRLRKCKEAGMREQCVLSKEQIRKKKIRKQQQS---	171
pca_LXR	GRYACRGGGTCQMDAFMRRKCQQCRLRKCKEAGMREQCVLSKEQIRKKKIRKQQ-QS---	170
dle_LXR	GRYACRGGGTCQMDAFMRRKCQQCRLRKCKEAGMREQCVLSKEQIRKKKIRKQQQS---	171
	*****:***** *	
hsa_LXR	SQSPVGPQGSSSSASGPGASPGGSEAGSQSGEGEGVQLTAAQELMIQQLVAAQLQCNRK	240
ttr_LXR	--SPTGLGVSSSSASGPGASPGGSDGGGQSGEGEGVQLTAAQELMIQQLVAAQLQCNRK	229
pca_LXR	--SPTGLGVSNSSASGPGAPPGGSDGGGQSGEGEGVQLTAAQELMIQQLVAAQLQCNRK	228
dle_LXR	--SPTGLGVGSSASGPGASPGGSDGGGQSGEGEGVQLTAAQELMIQQLVAAQLQCNRK	229
	**.* ..***** *****:*.*****	
hsa_LXR	SFSDQPKVTPWPLGADPQSRDARQORFAHFTELAIISVQEIIVDFAKQVPGFLQLGREDQI	300
ttr_LXR	SFSDQPKVTPWP-GADPQSRDA-QORFAHFTEL-IISVQEIIVDFAKQVPGFLQLGREDQI	286
pca_LXR	SFSDQPKVTPWPLGADPQSRDARQORFAHFTELAIISVQEIIVDFAKQVPGFLQLGREDQI	288
dle_LXR	SFSDQPKVTPWPLGADPQSRDARQORFAHFTELAIISVQEIIVDFAKQVPGFLQLGREDQI	289
	***** ***** ***** *****	
hsa_LXR	ALLKASTIEIMLLETARRYNHETECITFLKDFITYSKDDFHFRAGLQVEFINPIFEFSRAMR	360
ttr_LXR	ALLKASTIEIMLLETARRYNHETECITFLKDFITYSKDDFHFRAGLQVEFINPIFEFSRAMR	346
pca_LXR	ALLKASTIEIMLLETARRYNHETECITFLKDFITYSKDDFHFRAGLQVEFINPIFEFSRAMR	348
dle_LXR	ALLKASTIEIMLLETARRYNHETECITFLKDFITYSKDDFHFRAGLQVEFINPIFEFSRAMR	349
	*****	
hsa_LXR	RLGLDDAEYALLIAINIFSA DRPNVQEPGRVEALQQPYVEALLSYTRIKRPDQLRFPRM	420
ttr_LXR	RLGLDDAEYALLIAINIFSA DRPNVQEPSRVEALQQPYVDALLSYTRIKRPDQLRFPRM	406
pca_LXR	RLGLDDAEYALLIAINIFSA DRPNVQEPSRVEALQQPYVDALLSYTRIKRPDQLRFPRM	408
dle_LXR	RLGLDDAEYALLIAINIFSA DRPNVQEPSRVEALQQPYVDALLSYTRIKRPDQLRFPRM	409
	*****.*****:*****	
hsa_LXR	LMKLVSLRTLSSVHSEQVFALRLQDKKLPPLLSEIWDVHE	460
ttr_LXR	LMKLVSLRTLSSVHSEQVFALRLQDKKLPPLLSEIWDVHE	446
pca_LXR	LMKLVSLRTLSSVHSEQVFALRLQDKKLPPLLSEIWDVHE	448
dle_LXR	LMKLVSLRTLSSVHSEQVFALRLQDKKLPPLLSEIWDVHE	449
	*****	

Alinhamento múltiplo de sequências de receptor X do fígado das espécies *D. leucas*, *P. catodon*, *T. truncatus* e *H. sapiens*. Regiões coloridas em azul claro representam a sequência do DBD, enquanto as regiões coloridas em verde representam a sequência do LBD.