

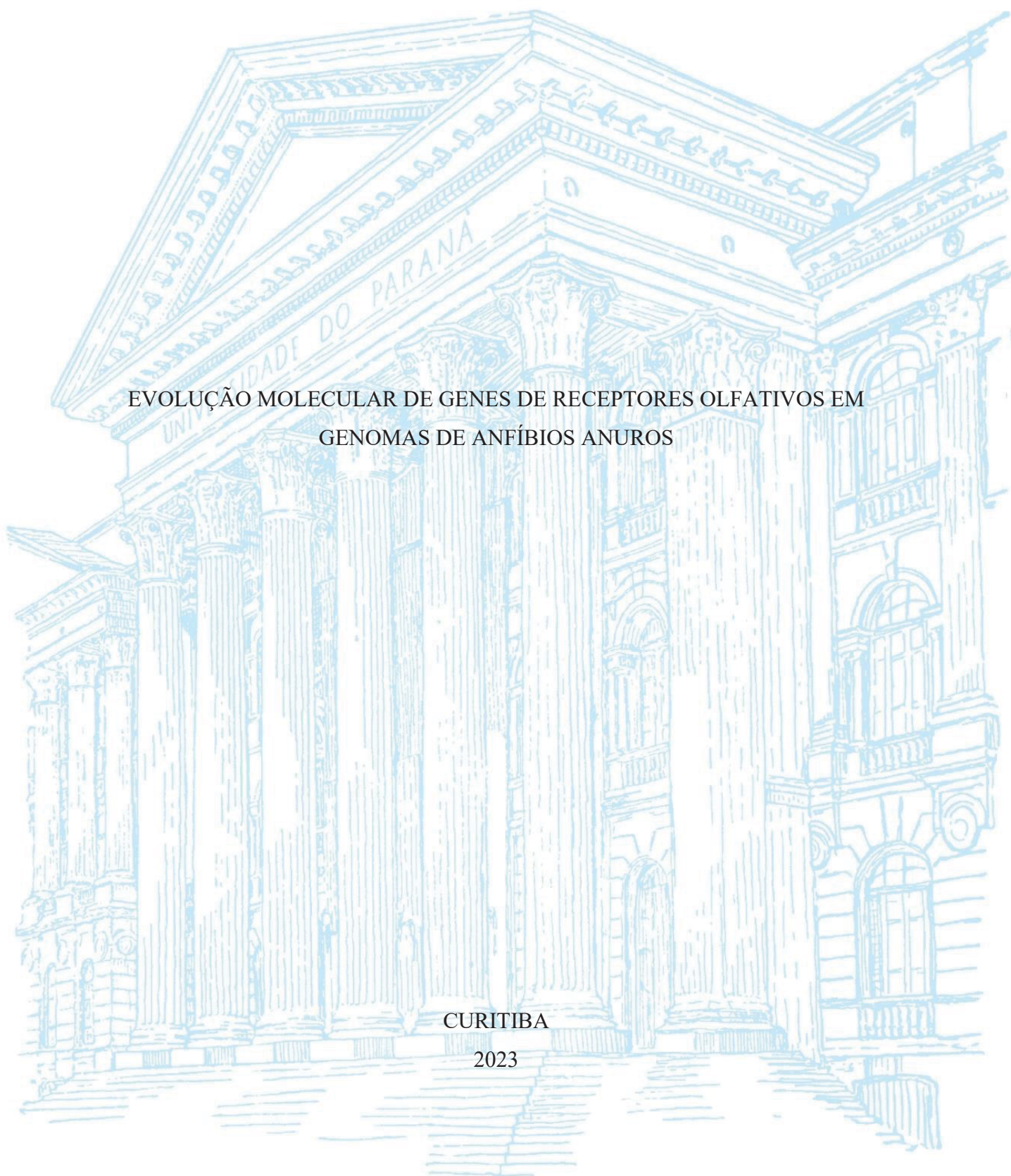
UNIVERSIDADE FEDERAL DO PARANÁ

JOHNNY SOUSA FERREIRA

EVOLUÇÃO MOLECULAR DE GENES DE RECEPTORES OLFATIVOS EM  
GENOMAS DE ANFÍBIOS ANUROS

CURITIBA

2023



JOHNNY SOUSA FERREIRA

EVOLUÇÃO MOLECULAR DE GENES DE RECEPTORES OLFATIVOS EM  
GENOMAS DE ANFÍBIOS ANUROS

Tese apresentada ao curso de Pós-graduação em Genética, Departamento de Genética, Setor de Ciências Biológicas, Universidade Federal do Paraná (UFPR), como requisito parcial à obtenção do título de Doutor em Genética.

Orientador: Prof. Dr. Daniel Pacheco Bruschi

CURITIBA

2023

DADOS INTERNACIONAIS DE CATALOGAÇÃO NA PUBLICAÇÃO (CIP)  
UNIVERSIDADE FEDERAL DO PARANÁ  
SISTEMA DE BIBLIOTECAS – BIBLIOTECA DE CIÊNCIAS BIOLÓGICAS

Ferreira, Johnny Sousa

Evolução molecular de genes de receptores olfativos em genomas de anfíbios anuros / Johnny Sousa Ferreira. – Curitiba, 2023.

1 recurso on-line : PDF.

Tese (Doutorado) – Universidade Federal do Paraná, Setor de Ciências Biológicas, Programa de Pós-graduação em Genética.  
Orientador: Prof. Dr. Daniel Pacheco Bruschi.

1. Olfato. 2. Anuro. 3. Evolução molecular. 4. Genomas. I. Bruschi, Daniel Pacheco, 1987-. II. Universidade Federal do Paraná. Setor de Ciências Biológicas. Programa de Pós-graduação em Genética. III. Título.



MINISTÉRIO DA EDUCAÇÃO  
SETOR DE CIÊNCIAS BIOLÓGICAS  
UNIVERSIDADE FEDERAL DO PARANÁ  
PRÓ-REITORIA DE PESQUISA E PÓS-GRADUAÇÃO  
PROGRAMA DE PÓS-GRADUAÇÃO GENÉTICA -  
40001016006P1

## TERMO DE APROVAÇÃO

Os membros da Banca Examinadora designada pelo Colegiado do Programa de Pós-Graduação GENÉTICA da Universidade Federal do Paraná foram convocados para realizar a arguição da tese de Doutorado de **JOHNNY SOUSA FERREIRA** intitulada: **Evolução molecular de genes de receptores olfativos em genomas de anfíbios anuros**, sob orientação do Prof. Dr. DANIEL PACHECO BRUSCHI, que após terem inquirido o aluno e realizada a avaliação do trabalho, são de parecer pela sua APROVAÇÃO no rito de defesa.

A outorga do título de doutor está sujeita à homologação pelo colegiado, ao atendimento de todas as indicações e correções solicitadas pela banca e ao pleno atendimento das demandas regimentais do Programa de Pós-Graduação.

CURITIBA, 31 de Outubro de 2023.

Assinatura Eletrônica

11/01/2024 16:49:02.0

DANIEL PACHECO BRUSCHI

Presidente da Banca Examinadora

Assinatura Eletrônica

11/01/2024 14:47:58.0

FABRICIUS MAIA CHAVES BICALHO DOMINGOS

Avaliador Externo (UNIVERSIDADE FEDERAL DO PARANÁ)

Assinatura Eletrônica

11/01/2024 19:38:19.0

MARCELO RICARDO VICARI

Avaliador Interno (UNIVERSIDADE ESTADUAL DE PONTA GROSSA)

Assinatura Eletrônica

11/01/2024 14:43:16.0

DESIRRÊ ALEXIA LOURENÇO PETTERS VANDRESEN

Avaliador Interno Pós-Doc (UNIVERSIDADE FEDERAL DO PARANÁ)

---

Setor de Ciências Biológicas, Centro Politécnico - CURITIBA - Paraná - Brasil

CEP 81531-980 - Tel: (41) 3361-1587 - E-mail: ppg-gen@ufpr.br

Documento assinado eletronicamente de acordo com o disposto na legislação federal Decreto 8539 de 08 de outubro de 2015.

Gerado e autenticado pelo SIGA-UFPR, com a seguinte identificação única: 334179

**Para autenticar este documento/assinatura, acesse <https://siga.ufpr.br/siga/visitante/autenticacaoassinaturas.jsp> e insira o código 334179**

Dedico essa tese à minha mãe, Gracilene O. de Sousa e ao meu avô, José Diniz de Sousa, por acreditar que eu sempre faria o meu melhor.

## AGRADECIMENTOS

Agradeço em primeiro lugar a minha família, pelo apoio incondicional. A minha irmã que sempre tem me ajudado, mesmo que eu teime em não pedir sua ajuda. A minha mãe por entender que quando eu decidir sair de São Luís, buscava o crescimento profissional e pessoal, mesmo sabendo que os dias comuns em que eu, minha irmã e ela ficávamos na sala conversando em enquanto assistíamos TV teriam que ser encerrados por um tempo, por todo o amor incondicional e também por me desejar bom dia todos os dias em que estive longe. Me perdoe por estar longe por tanto tempo, mas ainda estou tentando encontrar o meu caminho e fazer da minha vida uma jornada que valha a pena. Estarei sempre aqui. Agradeço a minha irmã, por todo o apoio e consideração, confiança e companheirismo que não mudam, mesmo estando distantes, mas alcançando nossos objetivos e ideias. Pensamos parecido, agimos de forma parecida e, por causa disso, confio plenamente em você.

Agradeço ao meu avô, que sempre foi uma pessoa excepcional e boa, que ajuda a família como um todo. Sei do quão orgulhoso você está, principalmente por ter sido o primeiro da família a ir tão longe na vida acadêmica. Agradeço a minha tia Lu, por acreditar em mim, confiar e apoiar em todos os passos que escolhi dar. Agradeço ao meu afilhado, Enzo, por ver em mim uma pessoa em que ele se espelha e eu sempre espero viver ao nível das suas expectativas e apoiá-lo para que almeje o mesmo crescimento que eu sempre busco para mim.

Agradeço aos meus amigos de infância, Tião e Roberto que, apesar dos anos em que nós não nos reunimos, sempre estamos batendo papo sobre nerdices, quadrinhos, mangás, filmes e séries no nosso grupo pessoal no Whatsapp. Agradeço também à um dos poucos amigos que ainda permanece dos dias da graduação em Biologia na UFMA, Tayllon, que sempre busca bater um papo e botar em dia a conversa nas poucas vezes em que pude voltar à São Luís para celebrar o fim de ano com a família.

Agradeço à todas as pessoas que conheci durante o período pré-pandemia, no início do doutorado. São tantas pessoas que espero não esquecer aqui. Camilla, Amanda, Joana, Julia, Michelle, Fernanda, Ketelen, Franciele, Daniela, Enzo, Aryel, Sebastião. São as pessoas que conhecemos durante um processo tão cansativo e exaustivo como uma

pós-graduação, que estão passando pelos mesmos problemas e alegrias, que nos faz querer continuar nos momentos difíceis.

Agradeço incrivelmente ao meu orientador, que acreditou em mim desde o começo, principalmente quando dizia que amava a parte da Bioinformática, coisa que ainda amo demais. Agradeço pelos puxões de orelha, pela paciência e ajuda, principalmente na etapa final, em que estava tão sobrecarregado de dados que sei que deixei a escrita um pouco de lado. A gente tende a focar naquilo que gosta de fazer e eu sei que meus dados foram refinados à um ponto em que não haveria dúvidas da autenticidade de tudo. Muito obrigado, de verdade. Meu crescimento pessoal e, principalmente profissional, assim como as vezes em que eu emperrava e não conseguia encontrar uma solução ou sentido nas coisas, você me ajudou a continuar.

Agradeço grandemente ao Prof. Dr. Yoshihito Niimura. Considero ele uma das pouquíssimas pessoas que conheci nessa vida que estão acima das outras como ser humano exemplar, não por suas crenças, mas por suas atitudes e genialidade, por entender que ajudar não necessariamente precisa ser algo que vai haver retorno. Ele me ensinou muito e eu aprendi muito com tudo o que vivenciei na Universidade de Miyazaki, nos 3 meses em que estive no Japão. Aliás, eu agradeço a ajuda da Yumi e do Inoue, sei que vocês não falavam inglês perfeitamente, mas nossa comunicação foi sempre clara e precisa. Como todos os japoneses que conheci durante minha estadia em Miyazaki, vocês são um exemplo de educação e cordialidade. Nosso sábado, quando fomos conhecer a Ilha dos Macacos, além do templo e do Parque dos Cavalos, foi um dos dias mais incríveis da minha vida e um divisor de água na minha mente. A vista de cima do monte do Parque foi a coisa mais linda que já vivenciei.

Agradeço a Universidade de Miyazaki por aceitar esse estrangeiro como o primeiro a visitar a Universidade logo no período pós-pandemia. Com a ajuda do Prof. Niimura, pude durante 2 meses desfrutar de um quarto na Casa do Estudante Internacional, que foi mais que suficiente para suprir todas as minhas necessidades básicas. Dito isso, agradeço ao grande Supermercado TopValu, que com suas promoções do fim do dia, me ajudaram a economizar grandemente durante minha estadia. Sem esse supermercado, não sei como eu poderia me alimentar perfeitamente (kkkk). Saudades dos sanduíches ovo, presunto e maionese de 100 ienes, dos empanados de frango e presunto, além da batata chips sabor pizza.

Agradeço ou Toshi, da pousada Aoshima Guesthouse Hooju, que me buscou no aeroporto no dia da minha chegada à Miyazaki. Você e sua esposa foram grandes anfitriões e nunca esquecerei do quão educado vocês eram. Sua pousada era o que se espera de uma casa tipicamente japonesa e durante umas duas semanas, eu não sabia que estava a apenas 250 metros da praia, por conta do tanto de tarefas que precisava completar no pouco tempo que tinha na cidade. Por isso, a localidade também é excepcional.

Agradeço ao Japão, por aceitar que eu vivenciasse um dos grandes sonhos da minha adolescência. Tudo que conheci, tudo que vivenciei, tudo que comi de especial e também cheguei a conhecer, como Akihabara, Shinjuku e Shibuya, foi algo que irei levar para toda a minha vida. Tóquio era as vezes tão parecida com Curitiba, mas ao mesmo tempo era tudo tão diferente e incrível. Lá, eu também pude conhecer um grande amigo alemão, o Gleb, que me ajudou a conhecer bons lugares para procurar games, action figures e mangás.

Agradeço a CAPES pelo bolsa de doutorado, que me ajudou bastante a passar pelo período da pandemia, na qual pude focar em estudar, analisar e sempre redefinir os métodos e ferramentas que utilizei para identificação dos genes dessa tese. O auxílio financeiro durante a minha estadia no Japão também foi essencial para que essa tese finalmente tivesse começado a andar a passos largos. Desse modo, meu agradecimento maior também vai a Ciência brasileira. Esses dados não são meus, mas do Brasil. A vivência multidisciplinar e multicultural que vivi só me fez amar mais ainda o Brasil e como aqui ainda há bastante coisa para ser feita, mas não somos desprovidos profissionais, pesquisadores e cientistas competentes e comprometidos. Com isso na minha mente, sempre busquei ser sempre o mais transparente e imparcial possível nos dados aqui apresentados. Devo isso àqueles que buscam o crescimento desse país.

Por último, agradeço do fundo de tudo o que sou o apoio incondicional e indispensável da minha esposa, Hanna, em todas as etapas dessa escolha louca que foi fazer doutorado e sair da minha zona de conforto. Obrigado por me aguentar, me apoiar e sempre se fazer disposta a auxiliar em todas as etapas dessa jornada, mesmo que eu sempre acabasse insistindo em fazer tudo sozinho. Te amo profundamente. Sem você e nossos gatinhos, Cuscuz, Bagheera, Café e Sakura, tenho certeza de que tudo teria sido muito mais sofrido e difícil.

*“Nós só podemos ver um pouco do futuro, mas o suficiente para perceber que há muito a fazer antes que possamos entender completamente o quão maravilhoso é.”*

**Allan Mathison Turing**

## RESUMO

O olfato é essencial para a sobrevivência dos animais. Moléculas odoríferas espalhadas pelo ambiente são detectadas por receptores olfativos (ORs), os quais constituem uma das maiores famílias multigênicas em vertebrados. A identificação do repertório completo de genes OR de uma espécie é feita através de ferramentas de bioinformática e sua composição varia entre as espécies. Anfíbios constituem um grupo especial, não somente por possuírem vida dupla (larval e adulta), mas por possuírem grupos de genes OR tanto para a detecção de odores encontrados no ar, quanto no ambiente aquático. Utilizando *scripts* e ferramentas de bioinformática, nós identificamos o repertório completo de genes OR de 28 espécies de anuros e analisamos a dinâmica evolutiva desses genes em um contexto evolutivo, assim como seus padrões de organização cromossômica. Todas as espécies mantêm genes dos grupos OR do Tipo 1 e Tipo 2, com exceção de  $\zeta$  (zeta), exclusivamente observados em peixes, e a composição varia grandemente entre espécies, sendo o genoma de *Engystomops pustulosus* mais abundante em cópias de genes OR no grupo  $\gamma$  (gamma), evidenciando um alto nível de expansão de número de cópias seguida com uma ampla dispersão de *singletons* e *clusters* de sequências entre os cromossomos deste cariótipo. Análises envolvendo espécies do gênero *Xenopus* revelaram a presença de 534 cópias ortólogas altamente conservadas a pelo menos 48 milhões de anos (OGGs – *ortholog gene groups*), que correspondem a quantidade de genes OR potencialmente observadas no ancestral do gênero *Xenopus*. Nossos dados revelam o papel de rearranjos intragenômicos e o papel do fracionamento de genes como origem dos padrões de perdas e ganhos de OGGs na evolução do repertório de ORs em espécies de origem diplóide e aloploplóide. Por fim, nosso estudo adiciona a identificação inédita do repertório OR de 13 espécies de anuros e auxilia na compreensão da dinâmica evolutiva dessa família multigênica em anfíbios anuros.

Palavras-chave: anura; olfato; cheiro; sapo; gene.

## ABSTRACT

Olfaction is essential for the survival of animals. Odorous molecules scattered throughout the environment are detected by olfactory receptors (ORs), which constitute one of the largest multigene families in vertebrates. Identifying the complete repertoire of OR genes in a species is done through bioinformatics tools, and its composition varies among species. Amphibians are a special group, not only because they have a dual life (larval and adult), but also because they have OR gene groups for detecting odors both in the air and in the aquatic environment. Using bioinformatics scripts and tools, we identified the complete repertoire of OR genes in 28 species of anurans and analyzed the evolutionary dynamics of these genes in an evolutionary context, as well as their chromosomal organization patterns. All species retain genes from Type 1 and Type 2 OR groups, with the exception of  $\zeta$  (zeta), which is exclusively observed in fishes, and the composition varies greatly among species. The genome of *Engystomops pustulosus* is particularly rich in copies of genes from the  $\gamma$  (gamma) group, demonstrating a high level of copy number expansion followed by widespread dispersal of singletons and sequence clusters across the chromosomes of this karyotype. Analyses involving species of the *Xenopus* genus revealed the presence of 534 highly conserved orthologous copies that have been maintained for at least 48 million years (OGGs - orthologous gene groups), corresponding to the number of OR genes potentially present in the ancestral *Xenopus* genus. Our data reveal the role of intragenomic rearrangements and gene fractionation in the origin of patterns of gain and loss of OGGs in the evolution of the OR repertoire in diploid and allopolyploid amphibian species. Finally, our study provides the previously undocumented identification of the OR repertoire in 13 anuran species and contributes to our understanding of the evolutionary dynamics of this multigene family in anuran amphibians.

Key words: anura; smell; scent; frog; gene.

## LISTA DE FIGURAS

- Figura 1. Representação esquemática da estrutura proteica de uma proteína OR de acordo com Touhara et al. (2016) e Zozulya et al. (2001). Motivos específicos estão destacados na cor cinza e os números representam cada um dos domínios transmembrana (TMs). Legenda: IC – Região Intracelular; EC – Região Extracelular. 16
- Figura 2. Classificação dos genes OR envolvendo mamíferos (Classe 1 e Classe 2) e vertebrados em geral (Tipo I e Tipo II) considerando o critério filogenético. .... 26
- Figura 3. Esquema representativo do protocolo de identificação de genes OR para vertebrados de acordo com Niimura (2013a; 2013b). .... 31
- Figura 4. Relacionamento filogenético entre as 28 espécies de anuros analisadas. Dendograma estimado na plataforma TimeTree: A Resource for Timelines, Timetrees, and Divergence Times 50 (<http://www.timetree.org/>). A espécie *Platypectrum ornatum* não foi incluída neste dendrograma por ausência de dados recuperados pela plataforma. .... 136
- Figura 5. Árvore filogenética inferida pelo IQ-TREE a partir das cópias dos OR genes intactos de espécies das famílias Bufonidae, Bombinatoridae, Dendrobatidae e Dicroglossidae. Cores:  $\alpha$  (alfa) – verde;  $\beta$  (beta) – rosa;  $\gamma$  (gamma) – azul;  $\delta$  (delta) – laranja;  $\epsilon$  (epsilon) – roxo;  $\eta$  (eta) – azul-claro; outgroup – preto..... 138
- Figura 6. Árvore filogenética inferida pelo IQ-TREE a partir das cópias dos OR genes intactos de espécies das famílias Hylidae, Leptodactylidae, Eleutherodactylidae, Pyxicephalidae e Limnodynastidae. Cores:  $\alpha$  (alfa) – verde;  $\beta$  (beta) – rosa;  $\gamma$  (gamma) – azul;  $\delta$  (delta) – laranja;  $\epsilon$  (epsilon) – roxo;  $\eta$  (eta) – azul-claro; outgroup – preto..... 139
- Figura 7. Árvore filogenética inferida pelo IQ-TREE a partir das cópias dos OR genes intactos de espécies das famílias Megophryidae, Ranidae e Scaphiopodidae. Cores:  $\alpha$  (alfa) – verde;  $\beta$  (beta) – rosa;  $\gamma$  (gamma) – azul;  $\delta$  (delta) – laranja;  $\epsilon$  (epsilon) – roxo;  $\eta$  (eta) – azul-claro; outgroup – preto. .... 140
- Figura 8. Árvore filogenética inferida pelo IQ-TREE a partir das cópias dos OR genes intactos de espécies da família Pipidae. Cores:  $\alpha$  (alfa) – verde;  $\beta$  (beta) – rosa;  $\gamma$  (gamma) – azul;  $\delta$  (delta) – laranja;  $\epsilon$  (epsilon) – roxo;  $\eta$  (eta) – azul-claro; outgroup – preto. .... 141
- Figura 9. Número de genes OR pertencentes a 28 genomas de anfíbios. Barras vermelhas, amarelas e azuis representam genes intactos, truncados e pseudogenes,

respectivamente. O número em parênteses abaixo de cada barra a quantidade dos genes intactos, truncados e pseudogenes, respectivamente. A porcentagem de pseudogenes em comparação com o total de genes OR no repertório é indicado por barras cinzas. .... 144

Figura 10. Distribuição dos genes OR nos cromossomos das espécies *Bufo bufo*, *Bufo gargarizans*, *Eleutherodactylus coqui* e *Engystomops pustulosus*. Legendas das cores:  $\alpha$  (verde),  $\beta$  (rosa),  $\gamma$  (azul),  $\delta$  (laranja),  $\epsilon$  (roxo),  $\eta$  (azul claro). Cores mais escuras dos respectivos grupos representam os pseudogenes e o amarelo representa truncados em geral. .... 149

Figura 11. Distribuição dos genes OR nos cromossomos das espécies *Hymenochirus boettgeri*, *Leptobrachium ailaonicum*, *Leptobrachium leishanense* e *Pyxicephalus adpersus*. Legendas das cores:  $\alpha$  (verde),  $\beta$  (rosa),  $\gamma$  (azul),  $\delta$  (laranja),  $\epsilon$  (roxo),  $\eta$  (azul claro). Cores mais escuras dos respectivos grupos representam os pseudogenes e genes em amarelo representam truncados em geral. .... 151

Figura 12. Distribuição dos genes OR nos cromossomos das espécies *Rana temporaria*, *Xenopus tropicalis*, *Xenopus borealis* e *Xenopus laevis*. Legendas das cores:  $\alpha$  (verde),  $\beta$  (rosa),  $\gamma$  (azul),  $\delta$  (laranja),  $\epsilon$  (roxo),  $\eta$  (azul claro). Cores mais escuras dos respectivos grupos representam os pseudogenes e genes em amarelo representam truncados em geral. Sub genomas L e S em *X. borealis* e *X. tropicalis* são exibidos com seus respectivos números. .... 154

## LISTA DE TABELAS

Tabela 1. Número de genes ORs identificados dos genomas de diversos organismos de acordo com a recente literatura.....	21
Tabela 2. Número de sequências funcionais de genes OR em diferentes espécies utilizadas.....	30
Tabela 3. Genomas de anfíbios anuros disponíveis no NCBI utilizados. ....	32
Tabela 4. Número de genes OR em 28 genomas de espécies de anfíbios anuros. ....	142
Tabela 5. Quantidade total de clusters encontrados nos cromossomos das 12 espécies aqui identificadas. ....	147
Tabela 6. Os cinco maiores clusters de genes OR identificados nos cromossomos de cada espécie. ....	154
Tabela 7. Comparação da quantidade de genes intactos, truncados e pseudogenes identificados neste trabalho, em Lu et al. (2021) e Han et al. (2022). Os traços determinam espécies aonde os genes OR não foram identificados em determinado trabalho. Legendas: I – Intactos; T – truncados; P – pseudogenes.....	158

# Sumário

<b>1</b>	<b>INTRODUÇÃO</b>	14
1.1.	<i>O olfato e sua importância na sobrevivência do indivíduo</i>	14
1.2	<i>Os receptores olfativos</i>	15
1.3	<i>O processo de identificação das moléculas odoríferas pelos ORs</i>	17
1.4	<i>A diversidade de genes ORs em vertebrados</i>	20
1.5	<i>Evolução dos receptores olfativos em vertebrados</i>	23
1.6	<i>A estrutura do sistema olfativo em vertebrados</i>	25
1.7	<i>A classificação dos receptores olfativos em vertebrados</i>	26
1.8	<i>Os genes OR em anfíbios anuros</i>	27
<b>2.</b>	<b>OBJETIVOS</b>	29
<b>3.</b>	<b>MATERIAIS E MÉTODOS</b>	29
3.1	<i>Identificação dos genes para receptores olfativos em espécies de anfíbios</i>	29
3.2	<i>Identificação dos genes intactos (funcionais)</i>	33
3.2.1	<i>O processo de obtenção dos dados e buscas TBLASTN</i>	33
3.2.2	<i>Obtenção dos melhores hits por região genômica</i>	34
3.2.3	<i>Filtragem das sequências por tamanho</i>	35
3.2.4	<i>Alinhamento das sequências</i>	36
3.2.5	<i>Árvore filogenética e escolha do ATG inicial</i>	36
3.3	<i>Identificação dos pseudogenes e genes truncados</i>	37
3.4	<i>Distribuição cromossômica dos genes OR</i>	41
<b>4.</b>	<b>RESULTADOS</b>	43
4.1	<b>Capítulo I – “Tracking the diversity and chromosomal distribution of the olfactory receptor gene repertoires of three anurans species”</b>	44
4.2	<b>Capítulo II – “Intrachromosomal rearrangement and gene fractionation modulated the repertoire of olfactory receptors genes in post-polyploid species of <i>Xenopus</i> (Anura; Pipidae)”</b>	91
4.3	<b>Capítulo III – Análise genômica comparativa do repertório de genes OR de 28 espécies de anuros: novos achados, lições e desafios no estudo desta família multigênica..</b>	135
4.3.1	<i>Relação do repertório com a proximidade filogenética das espécies</i>	136
4.3.2	<i>Total de genes OR nos genomas de anfíbios anuros</i>	142
4.3.3.	<i>Presença de pseudogenes e a relação com a quantidade de cópias intactas...</i>	145
4.3.4	<i>Presença de truncados</i>	146
4.3.5	<i>Distribuição cromossômica dos genes OR</i>	146
<b>5.</b>	<b>DISCUSSÃO</b>	157
5.1	<i>Diferenças metodológicas na identificação dos genes OR</i>	157

5.2	<i>Relação do repertório com a proximidade filogenética das espécies</i> .....	160
5.3	<i>A presença de clusters de genes OR nos cromossomos</i> .....	160
<b>6.</b>	<b>CONSIDERAÇÕES FINAIS</b> .....	162
<b>7.</b>	<b>REFERÊNCIAS</b> .....	164



## 1 INTRODUÇÃO

### *1.1. O olfato e sua importância na sobrevivência do indivíduo*

Virtualmente, todas as decisões comportamentais inerentes nos animais são baseadas em informações obtidas, transmitidas e interpretadas pelos seus sistemas sensoriais. Dados quimio-sensoriais são analisados de diversas formas. Durante os últimos 50 a 70 anos, os estudos que envolvem esse campo têm se concentrado na tentativa de entender como a adaptabilidade evolutiva tem agido na estruturação dos diversos sistemas sensoriais com base na correlação dos diferentes campos biológicos como a ecologia, fisiologia e ontogenia (DANGLES ET AL., 2009).

Esses estudos comparativos que analisam esses sistemas se tornam atrativos pela identificação de possíveis padrões evolutivos que, conseqüentemente, estão relacionados aos traços comportamentais dos animais e que podem constituir dados quantitativos e mensuráveis (DANGLES ET AL., 2009). Por exemplo, estudos ontogenéticos que relacionam determinado sistema sensorial podem ser úteis para entender como as forças evolutivas agem na captação de diferentes pistas químicas no ambiente, mas precisam ser aliados à processos ecológicos-chave, como a busca por recursos, detecção de possíveis predadores e também aos processos inerentes de reprodução da espécie (DANGLES ET AL., 2009).

Desse modo, o reconhecimento e filtragem dos diversos estímulos externos, aliados às propriedades ontogenéticas e estruturais teve diversos caminhos evolutivos nos diferentes grupos. Desses sistemas, **o sistema olfativo** compõe um dos sistemas cruciais no sucesso evolutivo e adaptativo de uma espécie. A evolução sensorial olfativa envolve a multiplicação e diversificação rápida de genes envolvidos na maquinaria do reconhecimento das diferentes moléculas odoríferas, que são identificadas pelos produtos desses genes expressos, principalmente, na forma de **receptores olfativos**. Estima-se que haja entre 400 a 500 milhares de odores no ambiente (TOUHARA ET AL., 2016) e por consequência, sua detecção e interpretação sensorial pelos organismos representa um grande desafio.

Os odores estão espalhados tanto dissolvidos no ambiente aquático, quanto dispersos de forma volátil no ar, sendo que neste último se encontra em maior diversidade. Para nós, humanos, a sensação do odor está também atrelada a outros sentidos, como a visão e a audição, mas para a maioria dos animais, entretanto, o olfato é essencial para sua sobrevivência. Assim, acredita-se que a diversidade de organismos possui variadas

quantidades de receptores de acordo com o ambiente em que estão inseridos, e a capacidade de adaptar-se a esses ambientes específicos tem impacto direto sobre a **evolução do repertório de genes de receptores olfativos** nas espécies.

## 1.2 Os receptores olfativos

A identificação primária dos genes de receptores olfativos foi feita por Buck & Axel (1991) que utilizaram PCR para identificar sequências tendo como base regiões conservadas de membros da superfamília das proteínas de acoplamento G (GPCR) e identificaram a família multigênica que codifica para os receptores de odores em roedores. Esses receptores proteicos foram chamados de receptores olfativos ou olfatórios (ORs, do inglês, *Olfactory Receptors*), pois eram primariamente expressos nos neurônios sensoriais olfativos (OSNs, do inglês, *Olfactory Sensory Neurons*) e por reconhecerem moléculas odoríferas. Essa grande descoberta lhes conferiu um prêmio Nobel em 2004 e, nas décadas seguintes, diversos estudos se concentraram na descoberta, mapeamento e estudo da dinâmica funcional desses receptores.

Os receptores olfatórios identificados foram categorizados como proteínas de acoplamento G (GPCR, do inglês, *G Protein-Coupled Receptors*) pertencentes à família das rodopsinas. Estudos anteriores já haviam identificado caminhos da ligação da molécula odorante com receptores transmembrana específicos dessa família, que desencadeiam um aumento na presença de Adenil-ciclase dependente de proteínas de ligação GTP (ou proteínas G), que por sua vez, são responsáveis pela cascata de sinalizações intracelulares através de axônios do neurônio olfativo até o bulbo cerebral (BUCK & AXEL, 1991, DRYER, 2000). De acordo com similaridades nas sequências proteicas, análises filogenéticas e identificação de motivos proteicos através da abordagem HMM (*Hidden Markov Models*), essa superfamília de proteínas é categorizada em ao menos cinco famílias, dentre as quais podemos citar a família das rodopsinas (FREDIKSSON ET AL., 2003).

Todas as proteínas GPCR, incluindo as proteínas ORs que pertencem a família das rodopsinas, compartilham sete domínios transmembrana alpha-hélice (TMs, do inglês, *Transmembrane Domains*) (Figura 1), um sítio glicosilado conservado na porção N-terminal e uma ligação dissulfeto na região extracelular. Em contrapartida, as proteínas



OR em um mesmo *cluster* genômico não necessariamente são indicativos de origem evolutiva comum em um mesmo clado filogenético ou mesmo co-expressos na mesma zona no epitélio olfativo. Um estudo feito há 20 anos por Serizawa et al. (2000), já demonstrou que cópias duplicadas de genes OR idênticos sequer eram expressos em um mesmo neurônio.

### 1.3 O processo de identificação das moléculas odoríferas pelos ORs

Dentro do grupo das proteínas GPCR, podemos destacar outros receptores como os relacionados à hormônios peptídeos, lipídeos, nucleotídeos, além das já conhecidas opsinas, que são baseadas em proteínas foto sensitivas (NIIMURA 2009b).

Quando falamos do processo de identificação de moléculas quimio sensoriais, tomando a visão como exemplo, os genes expressam proteínas nas células foto receptivas da retina que absorvem luz e iniciam a cascata de foto transdução. Essas proteínas são expressas em duas células fotorreceptoras: cones e bastonetes, os quais não somente se distinguem na forma, mas no tipo de opsina presente. Enquanto os bastonetes possuem rodopsinas para visão noturna, cones possuem as chamadas opsinas cones para detecção de luz diurna (TERAKITA, 2005). Vertebrados possuem cinco tipos de genes opsinas ancestrais (RH1, RH2, LWS, SWS1 e SWS2), os quais se diferenciam pela sensibilidade aos diferentes comprimentos de luz e que conseqüentemente resultam em diferentes pigmentos visuais formados pelas proteínas através da absorção da luz em diferentes comprimentos. Além disso, as diferenças nos comprimentos de onda são tanto controladas pela estrutura da proteína de opsina, quanto por outro componente visual: o cromatóforo (BOWMAKER, 2008; SCHOTT ET AL., 2022).

Durante a evolução dos vertebrados, as opsinas foram perdidas ou duplicadas independentemente entre os diferentes grupos, resultando na presença de tanto somente uma opsina, como no caso de algumas espécies de cetáceos, como *Mesoplodon bidens*, *Physeter macrocephalus* e *Kogia breviceps*, além de outras espécies da família das Balaenidae e Balaenopteridae, todas conhecidas por mergulhar a profundidades de pelo menos 100 metros (MEREDITH ET AL., 2013), quanto em até 38 cópias de RH1 em *Diretmus argenteus*, um peixe nativo de Portugal (MUSILOVA ET AL., 2019). Em anfíbios, por outro lado, podemos encontrar cinco das quatro opsinas (RH1, LWS, SWS1

e SWS2) em até oito diferentes tipos de fotorreceptores para a absorção de luz com pigmentos verde e azul (SIDDIQI ET AL., 2004; GOVARDOVSKII & REUTER, 2014).

A tabela classificatória dos diferentes comprimentos de onda já está inteiramente esclarecida e o mesmo pode ser dito sobre o paladar. Primariamente, os sabores são classificados pela percepção, que inclui intensidade e hedonismo (agradável ou desagradável), quanto pela qualidade: doce, amargo, azedo, salgado e umami (BACHMANOV ET AL., 2014). Estudos anteriores já identificaram os receptores para os sabores doce e umami (T1R), amargo (T2R) e salgado (ENaC), mas os receptores para o sabor azedo ainda são desconhecidos. Os receptores conhecidos são bastante conservados entre vertebrados (ortólogos) e, assim como as opsinas, sofreram duplicações gênicas, deleções (contrações) e pseudogenizações (por inserções e substituição de bases) que contribuíram para a variação na quantidade de cópias, padrão de expressão gênica e regiões da sequência entre diferentes espécies (BACHMANOV ET AL., 2014).

No entanto, quando falamos de olfato, não há um consenso quanto a classificação dos odores, apesar dos esforços em identificar os chamados “odores primários”, dos quais os outros odores poderiam ser reproduzidos pela mistura apropriada, visto a grande diversidade de moléculas presente nos ambientes (NIIMURA, 2009b). Essa questão foi debate entre alguns pesquisadores, mesmo antes da identificação dos genes OR. Por exemplo, em 1963, Amoore propôs os seguintes odores primários: canforado, etéreo, almíscar, menta, floral, pútrido e pungente. Com o passar dos anos e, após a identificação dos genes OR, a grande diversidade de receptores estimada em torno dos odores no ambiente levou essa classificação ao desuso.

Mesmo nos dias atuais, ainda não há um mapeamento completo da quantidade de odores ou mesmo a classificação satisfatória desses. Os estudos se concentraram em odores específicos e sua identificação por determinado grupo de vertebrados. Por exemplo, diversos estudos mais antigos mostram que peixes detectam baixas concentrações de um odor que se assemelha ao cheiro de flores, o álcool  $\beta$ -fene etílico (NEURATH, 1949; TEICHMAN, 1959; NEVIT ET AL., 1994). De todo modo, estes estudos tiveram mais avanço em humanos, observando-se que a família de genes OR utiliza de um método combinatório de diferentes receptores para identificar odores (MALNIC ET AL., 1999; GLEZER & MALNIC, 2019).

Assim, estudos que buscam identificar experimentalmente os ligantes específicos de determinadas proteínas OR utilizam uma combinação que envolve a modificação da superfície do receptor para otimizar a superfície de expressão com a co-expressão de outras proteínas, como a proteína de transporte receptora 1 (RTP1) e também a Ric-8B (PETERLIN ET AL., 2014), os quais potencializam a superfície de expressão e amplificam o sinal após a ativação da subunidade específica olfativa  $G\alpha$  (ou  $G\alpha_{olf}$ ) (VON DANNECKER ET AL., 2005; MACHADO ET AL., 2017). Além disso, é possível fazer a fusão dos 20 primeiros aminoácidos de rodopsina com a região terminal N de uma proteína OR para amplificar essa superfície de expressão em diversos receptores (KRAUTWURST ET AL., 1998). No geral, esses métodos ajudam a correlacionar os ligantes específicos com a expressão dos genes OR em diversos trabalhos (MAINLAND ET AL., 2014; GONZALEZ-KRISTELLER ET AL., 2015; MAINLAND ET AL., 2015; GEITHE ET AL., 2015).

Apesar disso, o mecanismo de detecção geral do ponto de vista anatômico já está bastante esclarecido, pois a anatomia dos vertebrados se manteve bastante conservada por mais de 500 milhões de anos (STODDART, 1980; DAHANUKAR ET AL., 2005; MENASHE & LANCET, 2006; HOOVER, 2010) e reflete os desafios olfativos através da diversidade filogenética entre espécies para detecção do amplo espectro de sinais químicos relativos à sua própria história ecológica.

A detecção de moléculas odoríferas em vertebrados tem início nos neurônios olfativos presentes no epitélio olfativo. Nesse tecido, um único dendrito se estende e termina em uma ponta ciliada. Esses cílios representam o primeiro local ativo no caminho olfativo e contêm as milhares de proteínas OR. Quando uma molécula odorífera se liga a uma proteína OR, ocorre a ativação da cascata de transdução das proteínas de acoplamento G, resultando na elevação dos níveis intracelulares de nucleotídeos cíclicos (ou fosfatos Inositol em algumas espécies). Esse processo culmina na despolarização da membrana de neurônios sensoriais olfativos e no início da propagação de potenciais de ação. No bulbo olfativo, os axônios dos sensores olfativos liberam sinapses com neurônios de retransmissão nos glomérulos. Nessa região, a informação primária é processada e retransmitida no centro do prosencéfalo. O fato das proteínas OR estarem localizadas na superfície entre o ambiente e o sistema nervoso central as torna um passo crítico, essencial e inicial no processamento da informação daquela molécula odorífera (DRYER, 2000).

#### 1.4 A diversidade de genes ORs em vertebrados

Nas últimas três décadas que se sucederam à descoberta da família multigênica dos genes OR por Buck & Axel (1991), os genes OR foram identificados nas principais linhagens de vertebrados: agnatos (FREITAG ET AL., 1999; LIBANTS ET AL., 2009), peixes (ALIOTO & NGAI, 2005; SARAIVA & KORSCHING, 2007; NIIMURA, 2009a; JOHNSTONE ET AL., 2012; WANG ET AL., 2021), anfíbios (NIIMURA, 2009a; LIU ET AL., 2021); aves (NEF ET AL., 1996; SIN ET AL., 2019; STEIGER ET AL., 2009) e, principalmente, mamíferos (ZOZULYA ET AL., 2001; NIIMURA & NEI, 2003; NIIMURA, 2013; NIIMURA ET AL., 2014; NIIMURA ET AL., 2018; FREEMAN ET AL., 2020).

O repertório de genes OR de um genoma é composto por cópias de genes intactos, presumidamente funcionais e por pseudogenes, cópias degeneradas destes genes que possuem mutações de mudança de quadro de leitura (*frameshifts*) e/ou mutações que geram stop códons prematuros (*nonsense mutation*). Apesar de inferidas como cópias funcionais, poucos estudos até o momento envolveram a validação da expressão gênica dessas cópias, sendo, portanto, essa distinção baseada apenas na sequência codificante sem considerar a análise das regiões regulatórias que coordenam a transcrição destas cópias (TOUHARA ET AL., 2016). Uma terceira categoria de cópias tem também sido relatada nas descrições dos repertórios de genes OR de diferentes genomas, as chamadas cópias (genes) truncadas. Essas cópias parecem representar muito mais um artefato metodológico proveniente das atuais metodologias de sequenciamento genômico por *short reads* do que um atributo biológico propriamente dito, uma vez que se trata de sequências de genes OR parciais que se encontram no fim dos *contigs*. A adição de novas metodologias de sequenciamento e montagem dos genomas tem reduzido consideravelmente o número de cópias truncadas nos repertórios de genes OR, o que pode corroborar essa observação de serem apenas de artefato metodológico (LU ET AL., 2021; HAN ET AL., 2022).

A diversidade do repertório de genes OR varia drasticamente entre os diferentes grupos taxonômicos (Tabela 1). Aspectos ecológicos e evolutivos são fatores que interferem na composição do repertório de genes de uma espécie. Por exemplo, dois

estudos analisaram a composição de genes OR em primatas do Novo Mundo e Velho Mundo e atribuíram as diferenças encontradas à aquisição de visão tricromática. Enquanto Gilad et al. (2004) analisaram a proporção de pseudogenização de determinado subconjunto de genes OR de diferentes espécies e encontraram deterioração do repertório olfativo de forma concomitante com a total aquisição da visão tricromática em diferentes linhagens, Niimura et al. (2018) não somente corroboraram com os resultados de Gilad et al. (2004), como observaram concordância na proporção de genes OR com outros aspectos morfológicos, anatômicos e ecológicos, mais precisamente relacionados à diferenças na morfologia nasal, comportamento diurno ou noturno e a alimentação por folhas ou frutos.

Tabela 1. Número de genes ORs identificados dos genomas de diversos organismos de acordo com a recente literatura.

<b>Espécie</b>	<b>Grupo</b>	<b>Intactos</b>	<b>Truncados</b>	<b>Pseudogenes</b>	<b>Autores</b>
<i>Bufo gargarizans</i>	Anfíbios	765	-	248	Lu et al. (2021)
<i>Rhinella marina</i>	Anfíbios	797	-	356	Lu et al. (2021)
<i>Nanorana parkeri</i>	Anfíbios	598	-	214	Lu et al. (2021)
<i>Pyxicephalus adspersus</i>	Anfíbios	711	-	405	Lu et al. (2021)
<i>Xenopus tropicalis</i>	Anfíbios	854	-	326	Lu et al. (2021)
<i>Xenopus laevis</i>	Anfíbios	592	-	320	Lu et al. (2021)
<i>Ambystoma Mexicanum</i>	Anfíbios	472	-	711	Lu et al. (2021)
<i>Microacaecilia unicolor</i>	Anfíbios	813	-	181	Lu et al. (2021)
<i>Rhinatrema bivittatum</i>	Anfíbios	1133	-	312	Lu et al. (2021)
<i>Cyprinus carpio</i>	Peixes	33	-	6	Wang et al. (2021)
<i>Siniperca chuatsi</i>	Peixes	123	-	29	Wang et al. (2021)
<i>Bos grunniens</i>	Mamíferos	262	11	114	Zhou et al. (2022)
<i>Cricetomys ansorgei</i>	Mamíferos	1060	-	2525	Freeman et al. (2020)
<i>Oceanodroma leucorhoa</i>	Aves	163	51	224	Sin et al. (2019)
<i>Coilia nasus</i>	Peixes	13	-	39	Zhu et al. (2017)
<i>Sillago sihama</i>	Peixes	112	26	3	Ye et al. (2023)
<i>Homo sapiens</i>	Mamíferos	403	0	438	Han et al. (2022)
<i>Bos taurus</i>	Mamíferos	1006	0	1212	Han et al. (2022)
<i>Balaenoptera physalus</i>	Mamíferos	54	5	271	Han et al. (2022)
<i>Hyaena hyaena</i>	Mamíferos	771	24	303	Han et al. (2022)
<i>Sphaenodon punctatus</i>	Répteis	345	2	153	Han et al. (2022)
<i>Naja naja</i>	Répteis	374	1	216	Han et al. (2022)
<i>Cathartes aura</i>	Aves	52	52	77	Han et al. (2022)

<i>Jacana jacana</i>	Aves	67	30	66	Han et al. (2022)
<i>Branchiostoma belcheri</i>	Anfioxo	8	1	40	Han et al. (2022)
<i>Petromyzon marinus</i>	Lampreia-marinha	46	0	17	Han et al. (2022)
<i>Marmota marmota</i>	Mamíferos	1034	14	570	Han et al. (2022)

Outro caso de correlação evolutiva com ecologia é observado entre o número de pseudogenes identificados em pinguins (LU ET AL., 2016). Análises comparativas da proporção de genes OR do repertório de duas espécies de pinguins, *Aptenodytes forsteri* e *Pygoseclis adeliae*, com outras 4 espécies Pelecaniformes, uma de Procellariiformes e uma de Gaviiformes, verificaram uma alta porcentagem de pseudogenes OR em pinguins em comparação com as outras aves aquáticas quanto aos genes ortólogos em comum. Além disso, essas evidências genéticas foram suportadas pela redução anatômica do bulbo olfativo que pinguins exibem em comparação com as outras espécies de aves aquáticas (LU ET AL., 2016).

Apesar de algumas tentativas de correlacionar o repertório de ORs com aspectos adaptativos tenham sido recentemente sugeridas, uma questão ainda em discussão é a seguinte: até que ponto podemos atribuir a evolução da presença de genes OR sob o ponto de vista ecológico e anatômico, visto que a capacidade de entender as pressões seletivas nessa família multigênica ainda é muito limitada? E se existem, as correlações negativas podem realmente ser atribuídas à possíveis processos ecológicos, aquisições ou perdas anatômicas divergentes? Yohe et al. (2022) estudaram morcegos da família Phyllostomidae, segundo maior grupo em número de espécies da ordem Chiroptera e que possui uma grande diversidade ecológica, numa tentativa de verificar assinaturas moleculares e morfológicas da adaptação olfativa relacionadas com a dieta de espécies de morcegos e observaram que, apesar da excepcional variação em ambas as morfologias olfativas e a variação no tamanho e na composição do repertório dos genes OR, as taxas de evolução molecular se mostraram independentes do tipo de dieta alimentar de cada linhagem evolutiva.

Já o estudo de Martínez et al. (2023) buscou investigar a relação entre cornetos nasais olfativos (cornetos cobertos pelo epitélio olfativo) e variações entre o repertório de genes OR entre roedores, baseado em estudos anteriores que sugeriam uma possível conexão entre o tamanho desses cornetos com a performance olfativa, assim como entre o tamanho dos cornetos e performance na conservação do calor e umidade (VAN

VALKENBURGH ET AL., 2011). Esses estudos demonstraram correlação entre traços ecológicos e áreas de superfície dos cornetos como um padrão geral. Apesar da evidente variação anatômica entre os grupos estudados, os dados desses autores não recuperaram nenhuma correlação significativa entre essas variações e a evolução do repertório de genes OR nessas espécies e a área de superfície absoluta (considerada pela área total de todos os cornetos) e relativa (a separação dos cornetos em regiões de acordo com as funções olfativa e respiratória) dos cornetos nasais olfativos que foi, neste caso, atribuído a outros possíveis fatores variantes entre as diferentes espécies: a não uniformidade de cobertura, tanto das proteínas OR nos cornetos como do epitélio olfativo e o grau de conectividade neural.

Em partes, esta dificuldade de correlacionar os dados genômicos de composição de genes OR de uma espécie decorre diretamente do fato de que os mecanismos envolvendo o olfato ainda não são totalmente esclarecidos (VOSSHALL, 2015) e receptores olfativos parecem se ligar a moléculas odoríferas de maneira combinatória (KURIAN ET AL., 2021) e não por uma correlação direta, fato que levou a Hipótese do Código Combinatório de receptores para odorantes (*Combinatorial Receptor code for odors*) (MALNIC ET AL., 1999). Assim, se uma única molécula odorífera pode ser detectada tanto por um receptor especialista, quanto por múltiplos receptores operando independentemente ou de forma combinatória (MALNIC ET AL., 1999; HAYDEN & TEELING, 2014; YOHE & BRAND, 2018), testar as pressões seletivas sobre um determinado gene, ou grupo de genes, passa a ser um grande quebra-cabeças para o qual ainda estamos no princípio de sua resolução.

### *1.5 Evolução dos receptores olfativos em vertebrados*

É presumido que os primeiros genes OR emergiram de um dos genes GPCR do tipo das rodopsinas que estava presente no ancestral comum de todas as espécies do clado Bilateria (NIIMURA, 2009a).

Por volta de pelo menos 200 milhões de anos atrás, os genes OR tiveram um pico de duplicação gênica que corresponde ao período em que os répteis dominavam o ambiente terrestre (FUCHS ET AL., 2001), ligado à irradiação genética provocada pela mudança adaptativa dos eventos de especiação que se sucederam com o domínio do

ambiente terrestre. A Ordem dos Coelacanthiformes, que inclui os celacantos, é representada como um estado transicional entre peixes e tetrápodes por possuir estruturas parecidas com as patas, principalmente devido aos fósseis dos ancestrais (JANVIER, 2007). Atualmente há duas espécies viventes, *Latimeria chalumnae* e *Latimeria menadoensis*, das quais a primeira é bastante relevante nos estudos da evolução do olfato por possuir genes OR observados em anfíbios e mamíferos, mas não em peixes (FREITAG ET AL., 1998).

Nos estudos iniciais da evolução dos genes OR, logo após sua descoberta, Freitag et al. (1998), fizeram afirmações de que os genes OR que aparecem somente em peixes (na época, chamados de “*fish-like*”) estavam tanto em baixa quantidade e diversidade entre diferentes espécies, enquanto os genes OR presentes em mamíferos (na época, chamados de “*mammalian-like*”) estavam em grande quantidade e diversidade de sequências (DRYER, 2000). De fato, análises comparativas entre vertebrados confirmaram essa afirmação (NIIMURA & NEI, 2005): por exemplo, peixes detectam quatro principais moléculas dissolvidas em água: aminoácidos, esteroides gonadais, ácidos biliares e prostaglandinas, dos quais vertebrados terrestres por não serem voláteis (LABERGE & HARA, 2001).

Assim, separa-se a evolução do sistema olfativo dos vertebrados em três grandes momentos de acordo com genômica comparativa: (1) a mudança que envolveu a transição para o ambiente terrestre pelos ancestrais anfíbios; (2) o declínio de diversos genes OR nos primatas e; (3) a expansão e delineamento atual dos receptores de feromônios na radiação atual dos roedores (KAMBERE & LANE, 2007).

Niimura & Nei (2005) estudaram a dinâmica evolutiva dos genes olfativos em diferentes vertebrados e encontraram evidências filogenéticas de que o mais recente ancestral comum entre peixes e todos os tetrápodes possuía pelo menos nove genes para receptores olfativos, cuja expansão da família multigênica atual permite a divisão em nove grupos filogenéticos relacionados a esses genes ancestrais que são recuperados como grupos monofiléticos em árvores de sequências.

Os trabalhos de Niimura & Nei (2005a; 2005b) mostram a grande capacidade que a família de genes OR possui de contrair e expandir em quantidade ao longo de pequenas escalas evolutivas. Um exemplo disso é o comparativo evolutivo entre humanos e roedores (NIIMURA & NEI, 2005a). Roedores possuem cerca de 1400 genes OR, dos quais somente 25% são considerados pseudogenes, em comparação com humanos, que

possuem mais de 50% de pseudogenes dos 800 genes OR conhecidos. Apesar desse grande diferencial, foi possível notar que a diferença na quantidade desses genes entre as duas espécies possivelmente se deve a duplicações e rearranjos cromossômicos, fato que é facilmente observado pela detecção de ortólogos em análises filogenéticas e análises de sintenia.

### *1.6 A estrutura do sistema olfativo em vertebrados*

De maneira geral, o sistema olfativo de grande parte dos vertebrados é bastante similar. O sistema é composto de um epitélio olfativo periférico, que inclui uma mucosa que serve como detecção de moléculas odoríferas e, um bulbo olfativo, que serve como o centro de processamento dos odores que vem do epitélio olfativo e de centros de processamento terminais (os chamados córtices olfatórios) (WEISS ET AL., 2021; ACHE & YOUNG, 2005; TANIGUCHI ET AL., 2011). Além disso, tetrápodes possuem a separação do sistema olfativo em subsistemas, na qual o sistema olfativo acessório é composto do epitélio vomeronasal e de um bulbo olfativo acessório (FLEISCHER ET AL., 2020). Anteriormente, o epitélio vomeronasal era tido como um tecido exclusivo na detecção de feromônios (EISTHEN 1992; 1997), no entanto, novas evidências fornecidas pelo trabalho de MOHRHARDT ET AL. (2018) demonstraram que o epitélio olfativo principal também pode possuir essa função.

Foi a presença do epitélio vomeronasal em larvas de anfíbios que desbancou a teoria anterior de que esse tecido do subsistema olfativo era considerado uma grande adaptação de tetrápodes terrestres na transição da água para a terra, visto que anatomicamente, este epitélio é somente encontrado nessa linhagem (BERTMAR, 1981). Mais ainda, o epitélio vomeronasal não somente é encontrado nas três linhagens de anfíbios, como também sua presença no estágio adulto é considerada um traço ancestral do seu estágio de vida larval (JUNGBLUT ET AL., 2012; BENZEKRI & REISS, 2012).

Grande parte dos anuros, inclusive, exhibe uma separação do epitélio olfativo principal que se acredita ser dedicado para a detecção de moléculas dissolvidas em água (REISS & EISTHEN, 2008). Diferentemente de *Xenopus*, na qual o epitélio sensorial está no *diverticulum lateral*, como mencionado anteriormente, a grande maioria dos anuros

aparenta ter receptores olfativos em órgãos olfativos periféricos (e.g. *recessus olfactorius*) (BELANGER & CORKUM, 2009; JUNGBLUT ET AL., 2021; WEISS ET AL., 2021).

### 1.7 A classificação dos receptores olfativos em vertebrados

Há dois sistemas de classificação para genes OR em vertebrados, baseados em similaridade e filogenia de sequências (Figura 2). O primeiro divide os genes em dois grandes grupos de acordo com Niimura (2009a; 2012): Tipo I, contendo seis subgrupos ( $\alpha$  – alfa,  $\beta$  – beta,  $\gamma$  – gamma,  $\delta$  – delta,  $\varepsilon$  – epsilon,  $\zeta$  – zeta) e Tipo II contendo somente um subgrupo ( $\eta$  – eta). Desses grupos,  $\delta$ ,  $\varepsilon$ ,  $\zeta$  (Tipo I) e  $\eta$  (Tipo II) detectam odorantes dissolvidos em água e estão presentes nos genomas de peixes. Já em répteis, aves e mamíferos são encontradas somente sequências homólogas dos grupos  $\alpha$  e  $\gamma$  como constituintes de seu repertório de genes OR. Anfíbios, por outro lado, possuem tanto genes do Tipo I quanto Tipo II (HAYDEN & TEELING, 2004; NIIMURA, 2009a; 2009b).

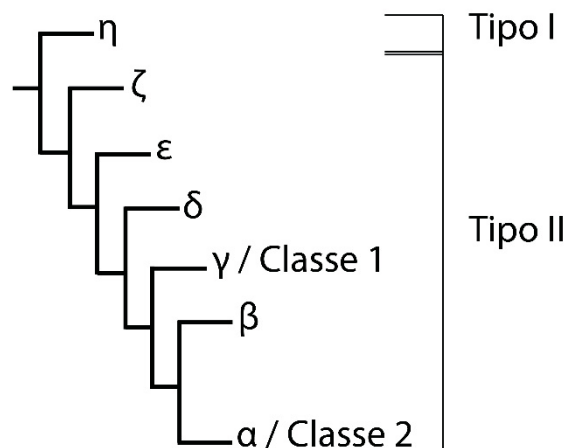


Figura 2. Classificação dos genes OR envolvendo mamíferos (Classe 1 e Classe 2) e vertebrados em geral (Tipo I e Tipo II) considerando o critério filogenético.

Como répteis, aves e mamíferos possuem somente dois grupos de genes propostos na classificação de Niimura (2009a), o grupo  $\alpha$  e  $\gamma$ , estudos focaram na subdivisão desses dois grupos, na qual os genes do grupo  $\alpha$  estão categorizados como ORs de Classe 1

enquanto os do grupo  $\gamma$  são categorizados de Classe 2. Além disso, essas classes são subdivididas em famílias numeradas, sendo para a Classe 1, OR51, OR52, OR55 e OR56 e para a Classe 2, OR1/3/7, OR2/13, OR4, OR5/8/9, OR6, OR10, OR11, OR12 e OR14 (HAYDEN & TEELING, 2004). A Classe 1 é considerada mais antiga (OLENDER ET AL., 2004), pois genes dessa classe também podem ser encontrados em peixes em menor número e, por conta disso, predizem identificar pequenos odorantes dissolvidos em água (FREITAG ET AL., 1998). Por outro lado, a Classe 2 é a mais numerosa e composta de genes especialização na detecção da vasta quantidade de odorantes presentes no ar e compõem majoritariamente o repertório dos mamíferos, inclusive os marinhos (HAYDEN ET AL., 2010).

### 1.8 Os genes OR em anfíbios anuros

O primeiro trabalho envolvendo a identificação putativa de genes OR em anfíbios foi feito por Freitag et al. (1995), que se basearam na suposição de haver grande homologia entre as sequências da família de genes OR já identificadas em outras espécies. Suas análises foram realizadas em *Xenopus laevis* e levaram a identificação de duas classes, que na época foram chamadas de Classe 1 e Classe 2, correspondendo aos atuais grupos de gene do Tipo I,  $\alpha$  e  $\gamma$  (NIIMURA, 2009a). Estudos morfológicos da cavidade nasal de *Xenopus* revelam a existência de três subsistemas olfativos: o órgão vomeronasal, o *diverticulum lateral* e o *diverticulum medial*. Freitag e colaboradores (1995) mostraram que, a partir de uma análise de hibridização *in situ*, genes do grupo  $\alpha$  são exclusivamente expressos no *diverticulum lateral* nas cavidades nasais para odorantes aquáticos, enquanto a genes do grupo  $\gamma$  são expressos no *diverticulum medial* (FREITAG ET AL., 1995).

Com a publicação do sequenciamento completo do genoma de *X. tropicalis* e o maior desenvolvimento de ferramentas de bioinformática, Niimura & Nei (2005) reavaliaram o repertório de ORs desta espécie, ampliando o número de genes identificados. Subseqüentes reavaliações da quantidade de genes do seu repertório foram realizadas à medida que novas versões do genoma desta espécie foram sendo disponibilizadas (NIIMURA 2009a; LU ET AL., 2021; HAN ET AL., 2022). Esses trabalhos de caracterização dos repertórios de ORs fomentaram subseqüentes trabalhos relacionados à análise da expressão dos genes de OR, da especificidade dos receptores à

determinados ligantes em *X. tropicalis* e *X. laevis* (MANZINI, 2010), e estrutura molecular e morfológica dos sistemas olfativos dessas espécies-modelo (MANZINI & KORSCHING, 2011; WEISS ET AL., 2020; WEISS ET AL., 2021).

Somente em 2020, Yohe et al. (2020), estudando a taxa de evolução dos genes OR junto com variações nas estruturas morfológicas quimio sensoriais identificou novos genes para as espécies de anfíbios *Ambystoma mexicanum* e *Rhinatrema bivittatum*, enquanto Lu et al. (2021) ao descrever a montagem de genoma de *Bufo gargarizans*, identificou o repertório completo de diversas espécies de anfíbios, tanto em quantidade quanto ao nível de agrupamento filogenético utilizando a classificação de Niimura (2009a). Mais recentemente, Han et al. (2022) construíram e disponibilizaram o banco de dados Chordata Olfactory Receptor Database (CORD, <https://cord.ihuman.shanghaitech.edu.cn>), um banco de dados de anotação de genes OR de 1695 espécies de cordados. Neste banco de dados é possível obter e extrair o repertório completo de genes OR das espécies, modelos estruturais, redes de similaridade de sequência, além das ferramentas utilizadas para identificação dos genes. Nesta proposta, Han e colaboradores (2022), forneceram inicialmente a anotação de 19 genomas de anfíbios, representando um avanço no conhecimento dos repertórios de ORs nesses organismos. Atualmente, este banco de dados constitui o maior acervo de genes OR. Embora tenham ampliado o conhecimento de mais 10 espécies de anfíbios em comparação com a identificação de Lu et al. (2021), Han et al. (2022) não disponibilizaram as informações quanto aos agrupamentos filogenéticos dos ORs das espécies dos 19 genomas presentes.

Uma vez consolidada a posição evolutiva dos anfíbios do ponto de vista dos sistemas quimio receptivos, estudos mais recentes ainda se atêm a descrição e análise morfológica do sistema olfativo único em comparação com pesquisas envolvendo mamíferos (WEISS ET AL., 2021). Mais ainda, enquanto há análises envolvidas na identificação dos genes OR, essa identificação comparativa entre espécies ainda não leva em consideração um comparativo da estrutura morfológica das diferentes espécies, justamente por existirem muitas espécies de anfíbios com habitats e histórias de vida diferentes.

Assim, pouco se sabe sobre a expressão dos genes OR quanto à sua presença nas estruturas nasais dedicadas em anfíbios, os quais concentram-se na composição ciliar observada nos vários tipos de tecidos epiteliais, com somente poucos estudos focados na

transição da detecção molecular dos odorantes. Essas lacunas no conhecimento sobre o ORs de anfíbios representam importantes desafios que deverão ser vencidos nas próximas décadas de pesquisa e poderão resultar em uma melhor compreensão deste complexo sistema sensorial deste grupo taxonômico.

## 2. OBJETIVOS

### *Geral:*

Identificar o repertório de genes de receptores olfativos em anfíbios anuros sob o ponto de vista evolutivo, com foco nos aspectos genéticos e contribuições ecológicas que modularam suas trajetórias evolutivas.

### *Específicos:*

- Investigar grupos ortólogos altamente conservados que sejam linhagem-específica e se esses genes preditos estiveram evolutivamente envolvidos com rearranjos intragenômicos (intra e intercromossômicos);
- Avaliar o número e diversidade de cópias de genes de receptores olfativos presentes no genoma de anfíbios com diferenças ecológicas e comportamentais;
- Analisar o papel do modelo de evolução *Birth and Death* na evolução desta família multigênica;
- Avaliar a integridade e contexto molecular de genes de receptores olfativos por meio da quantidade de pseudogenes e truncados;
- Verificar se existe uma evolução divergente na diversidade e densidade desses grupos de genes entre genomas de anfíbios terrestres e aquáticos;

## 3. MATERIAIS E MÉTODOS

### *3.1 Identificação dos genes para receptores olfativos em espécies de anfíbios*

Este trabalho se ateve ao protocolo previamente proposto e aprimorado por Niimura & Nei (2005), Niimura (2009), Niimura (2013a) e Niimura (2013b) que define os parâmetros básicos da identificação dos receptores olfativos, separados na identificação de mamíferos e vertebrados em geral, mas que divergem em pequenos parâmetros. Assim, foi possível não somente a automatização de grande parte da identificação de genes OR, mas como o aprimoramento de *scripts* desenvolvidos nas linguagens *Perl* e *Python*, que foram utilizados e os quais já se propunham a tal objetivo.

Todos os *scripts* proprietários criados para este fim estão disponíveis no link <https://github.com/johnnymundi/> sob os repositórios de nome “Phylogenetic-Scripts” e “bioinformatic\_perl\_scripts”.

A abordagem consiste na recuperação de sequências através de buscas por homologia, análise de diversos motivos específicos e classificação através de árvores filogenéticas que diferenciam os receptores olfativos de outros membros do grupo dos receptores quimio sensoriais e da superfamília das rodopsinas acopladas a proteínas G.

Desse modo, todas as sequências conhecidas de genes de receptores olfativos funcionais de vertebrados foram extraídas do trabalho de Niimura (2009), com exceção dos genes OR humanos, que foram extraídos do *Database The Human Olfactory Data Explorer (HORDE)* ([genome.weizmann.ac.il/horde](http://genome.weizmann.ac.il/horde)) (OLENDER ET AL., 2013) (Tabela 2).

Tabela 2. Número de sequências funcionais de genes OR em diferentes espécies utilizadas.

Nome comum	Nome científico	Número de genes OR	Candidatos CD-Hit
Lampreia-marinha	<i>Petromyzon marinus</i>	32	14
Tubarão-elefante	<i>Callorhinchus milli</i>	1	1
Peixe-zebra	<i>Danio rerio</i>	154	44
Peixe-arroz	<i>Oryzias latipes</i>	68	21
Esgana-gato	<i>Gasterosteus aculeatus</i>	102	37
Torafugu	<i>Takifugu rubripes</i>	47	19

Baiacu verde pintado	<i>Tetraodon nigroviridis</i>	11	9
Rã	<i>Xenopus tropicalis</i>	824	110
Lagarto	<i>Anolis carolinensis</i>	112	35
Galinha	<i>Gallus gallus</i>	211	37
Humano	<i>Homo sapiens</i>	391	92
TOTAL		1953	419

O processo para identificação de genes funcionais e não-funcionais está ilustrada na Figura 2. Todos os genomas de anuros utilizados foram baixados do site National Center for Biotechnology Information (NCBI) ([www.ncbi.nlm.nih.gov](http://www.ncbi.nlm.nih.gov)) e consistem das espécies apresentadas na Tabela 3.

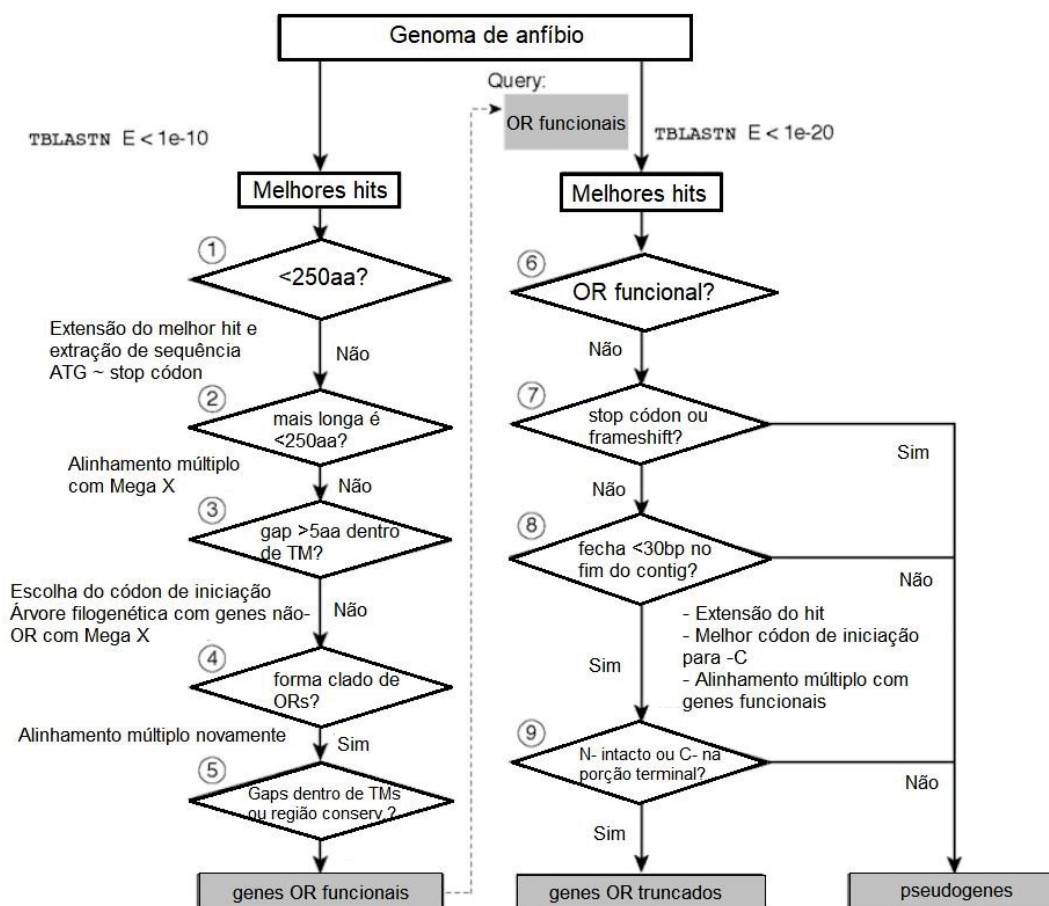


Figura 3. Esquema representativo do protocolo de identificação de genes OR para vertebrados de acordo com Niimura (2013a; 2013b).

Tabela 3. Genomas de anfíbios anuros disponíveis no NCBI utilizados.

Species	Genome Id	Genome Coverage	Assembly Level	Contig N50
<i>Bombina variegata variegata</i>	GCA_905336975.1	45x	Scaffold	1857
<i>Bufo bufo</i>	GCA_905171765.1	64x	Chromosome	3959750
<i>Bufo gargarizans</i>	GCA_014858855.1	103x	Chromosome	1738317
<i>Eleutherodactylus coqui</i>	GCA_019857665.1	240x	Chromosome	10856
<i>Engystomops pustulosus</i>	GCA_019512145.1	103x	Chromosome	305357
<i>Glandirana rugosa</i>	GCA_018402905.1	154x	Scaffold	11002
<i>Hymenochirus boettgeri</i>	GCA_019447015.1	110x	Chromosome	783846
<i>Leptobranchium ailaonicum</i>	GCA_018994145.1	78x	Chromosome	821125
<i>Leptobranchium leishanense</i>	GCA_009667805.1	80.3x	Chromosome	1946319
<i>Limnodynastes dumerilii</i>	GCA_011038615.1	156x	Scaffold	10550
<i>Lithobates catesbeianus</i>	GCA_002284835.2	66x	Scaffold	5415
<i>Nanorana parkeri</i>	GCA_000935625.1	83x	Scaffold	32798
<i>Oophaga pumilio</i>	GCA_009801035.1	136x	Scaffold	77984
<i>Phrynoglossus myanhessei</i>	GCA_022657655.1	25x	Contig	374416
<i>Phyllomedusa bahiana</i>	GCA_025379985.1	111x	Contig	78483
<i>Pipa carvalhoi</i>	GCA_021901965.1	36x	Contig	145463
<i>Pipa parva</i>	GCA_019650415.1	15x	Scaffold	6215
<i>Platyplectrum ornatum</i>	GCA_016617825.1	140x	Scaffold	4707
<i>Pyxicephalus adspersus</i>	GCA_004786255.1	189x	Chromosome	30445
<i>Rana temporaria</i>	GCA_905171775.1	63x	Chromosome	6.264.810
<i>Rhinella marina</i>	GCA_900303285.1	141x	Contig	3374
<i>Scaphiopus couchii</i>	GCA_009364435.1	25x	Scaffold	362
<i>Scaphiopus holbrookii</i>	GCA_009364455.1	39x	Scaffold	514
<i>Spea bombifrons</i>	GCA_009364475.1	35.45x	Scaffold	21170366
<i>Spea multiplicata</i>	GCA_009364415.1	21x	Scaffold	30692
<i>Xenopus borealis</i>	GCA_024363595.1	50x	Chromosome	21213
<i>Xenopus laevis</i>	GCA_017654675.1	28x	Chromosome	22.451.507
<i>Xenopus tropicalis</i>	GCA_000004195.4	111.5x	Chromosome	14.634.335

Para diminuição do tempo computacional das buscas por homologia e seleção de um repertório suficientemente diverso de ORs, o seguinte critério foi utilizado: cada sequência de OR funcional utilizada como *query* em cada espécie foi agrupada de acordo com um dos grupos filogenéticos os quais foram previamente identificados por Niimura (2009) ( $\alpha$ ,  $\beta$ ,  $\gamma$ ,  $\delta$ ,  $\epsilon$ ,  $\zeta$ ,  $\eta$ ). Utilizamos o programa CD-Hit (FU ET AL., 2012; LI & GODZIK, 2006) para a escolha das sequências candidatas considerando um nível de 60% de identidade. Assim, um total de 419 sequências candidatas foram selecionadas (Figura 2)

Os critérios de busca foram definidos “a priori” para cada um dos passos do *pipeline* aplicado (Figura 2), sendo que todas as etapas foram realizadas utilizando o sistema operacional Linux. Desse modo, uma explicação teórica das etapas será fornecida seguido das eventuais linhas de comando e processos utilizados para determinado fim, sendo que comentários dentro das linhas de comandos serão sinalizados com ‘#’ e, para fins de padronização, os nomes dos arquivos representativos do genoma e *queries* serão em inglês. Desse modo, a *pipeline* ocorre do seguinte modo:

### 3.2 Identificação dos genes intactos (funcionais)

#### 3.2.1 O processo de obtenção dos dados e buscas TBLASTN

Buscas por TBLASTN (ALTSCHUL ET AL., 1990) foram realizadas contra um dos genomas específicos utilizando 419 *queries* candidatas dos genes OR funcionais da Tabela 2. Para a identificação de possíveis genes OR, um *E-value* de  $1e-05$  foi estabelecido como critério para inclusão de acordo com Young et al. (2002), Niimura (2013a), Niimura & Nei (2005).

Para este fim, os arquivos executáveis do BLAST+v.2.14.1 do NCBI foram baixados e instalados através do link <https://ftp.ncbi.nlm.nih.gov/blast/executables/blast+/LATEST/>. Todos os genomas disponíveis no formato ‘.fna’ foram baixados do banco de dados Assembly do NCBI, tendo como critério RefSeq Category: Representative. Assim, os comandos para execução do TBLASTN local envolvem a criação de um banco de dados a partir dos genomas e realização do TBLASTN:

```
# Criação do bando de dados:
```

```
makeblastdb -in genome_name.fna -dbtype nucl -parse_seqids
```

```
# Execução do tblastn local:
```

```
tblastn -query query_ORgene_sequences.fasta -db genome_name.fna -out genome_name.tblastn -evalue 1e-05
```

### 3.2.2 Obtenção dos melhores hits por região genômica

Como muitas *queries* atingem uma mesma região genômica, a qual é estabelecida como um determinado intervalo entre o início e o fim de determinado *hit* de homologia, um intervalo de *hits* estabelecido por diferentes *queries*, conseqüentemente, é mais provável de conter um OR. Desse modo, *scripts* em *Perl* foram utilizados para recuperar as sequências que apresentavam sobreposição, sendo essas sequências consideradas os “melhores *hits*”:

```
# organização dos resultados do arquivo '.tblastn' para seleção dos
melhores hits:

getgenomepos.pl -t genome_name.tblastn > genome_name.genomepos;

# separação dos melhores hits de acordo com o e-value para regiões
genômicas distintas:

dobesthit.pl genome_name.genomepos;

# Como muitos hits representativos podem ter a mesma query como hit, é
necessário remover queries que aparecem repetidamente por mais de 40
vezes para determinada contig ou região genômica. Pra isso, o comando
AWK é utilizado no arquivo '.genomepos':

awk          '$1          ==          "XetrORs347.3"          {print}'
GCA_018994145.1_ASM1899414v1_genomic2.besthit |wc;

# exemplo de remoção de determinada query repetitiva:

awk          '$1!="query_name"          {print}'          genome_name.genomepos          >
genome_name.genomepos

# realização da escolha do melhor hit novamente, que gerará um arquivo
genome_name.besthit:

dobesthit.pl genome_name.genomepos;
```

### 3.2.3 Filtragem das sequências por tamanho

Genes não-funcionais serão excluídos da matriz de genes funcionais quando as sequências forem menores que 250 aminoácidos, pois esse comprimento é curto do que qualquer outro gene de receptor olfativo conhecido. Além disso, cada hit extraído será estendido em ambas as direções no genoma e, a partir de um códon ATG, será extraída a sequência mais longa até algum códon de parada (*stop códon*) (TAA, TAG ou TGA). Se estas sequências tiverem menos de 250 aminoácidos, serão excluídas e se possuírem aminoácidos desconhecidos devido à incerteza quanto ao nucleotídeo correspondente, também serão excluídas. Por fim, haverá a separação das sequências nos seus determinados grupos filogenéticos a partir da *query* representativa do *hit*.

```
# comando para extração dos melhores hits que gera um arquivo
'genome_name.250.fa' que contém as sequências codificantes mais longas
e maiores de 250 aminoácidos. Pra isso, é necessária a presença do
arquivo 'genome_name.besthit' e o genoma 'genome_name.fna' na pasta:
do_getseq.pl genome_name genome_name.fna;
```

```
# Para a separação do arquivo 'genome_name.250.fa' nos seus respectivos
grupos filogenéticos, um ambiente virtual Python foi criado e a pasta
continha os arquivos 'genome_name.250.fa', 'genome_name.besthit' e o um
arquivo 'OR_group.txt' no formato .JSON contendo um dicionário Python
na qual a nome da key é o nome do grupo filogenético e os valores são
uma lista com os nomes das queries utilizadas no TBLASTN:
```

```
python get_ORgroup.py genome_name
```

```
# O comando acima gera um arquivo 'genome_name.group_name.idlist' que
contém somente os headers das sequências. Desse modo, outro comando é
necessário para separar em um novo arquivo as sequências do arquivo
principal 'genome_name.250.fa' correspondentes aos headers do
'genome_name.group_name.idlist':
```

```
grepfasta.pl genome_name.group_name.idlist genome_name.250.fa >
genome_name.group_name.250.fa;
```

### 3.2.4 Alinhamento das sequências

O alinhamento das sequências restantes será realizado com o MAFFT 7.52 (KATO ET AL., 2019) e a posição das setes regiões transmembrana (TMs) será feita através do alinhamento com a rodopsina bovina, de acordo com especificações de Man et al. (2004). Para evitar discrepâncias, nós estabelecemos um *gap* de cinco ou mais aminoácidos dentro dos domínios como critério de exclusão da sequência.

```
# o comando abaixo gera duas pastas 'olfalign_group_name' e
olfalign2_group_name', na qual a última contém os arquivos gerados pelo
MAFFT nos formatos postscript e .pdf, um arquivo .fasta contendo as
sequências junto com queries representativas que foram alinhadas com o
MAFFT, além de um arquivo .txt contendo os headers das sequências não
deletadas, pois não possuíam gaps:
```

```
do_align.pl genome_name genome_name.fna .
```

```
# uma vez de posse do arquivo fasta, é necessário a remoção das queries
representativas e sequências que foram deletadas que não estão presentes
no arquivo .txt contendo os headers:
```

```
grepfasta.pl genome_name.group_name.align.del.notdeleted
genome_name.group_name.align.del.fa > genome_name.group_name.fa;
```

### 3.2.5 Árvore filogenética e escolha do ATG inicial

A escolha de um códon ATG inicial foi realizada visualmente para aquelas sequências que possuem mais de dois códons iniciais na região montante ao primeiro domínio transmembrana. Uma árvore de sequências exploratória por máxima verossimilhança foi gerada com as sequências, utilizando os seguintes genes recuperados do GenBank *Database* como grupos externos: isoforma 1 do receptor adrenérgico alfa-1A (NP\_000671.2), receptor adrenérgico beta 1, (NP\_000675.1), receptor adenosina A2b

(NP\_000667.1), receptor 1B de 5-hidroxitriptamina (serotonina) (NP\_000854.1), receptor 1F de 5-hidroxitriptamina (serotonina) (NP\_000857.1), receptor 6 de 5-hidroxitriptamina (serotonina) (NP\_000862.1), receptor H2 de histamina Hrh2 (NP\_001010973.1) e receptor 4 de somatostatina (NP\_001043.2). Esses genes são suficientemente próximos do grupo dos receptores olfativos entre os genes pertencentes a superfamília das rodopsinas GPCR. Assim o programa IQ-TREE 1.6.12 (NGUYEN ET AL., 2015) produzirá uma árvore de máxima verossimilhança, utilizando o modelo de substituição calculado pelo software Mega X 10.1.8 (KUMAR ET AL., 2018). Visto que os genes OR formam um clado monofilético com alto valor de *Bootstrap* (acima de 90%), com critério de informação AIC aplicado, as sequências que não foram recuperadas nesse clado foram descartadas, sendo as sequências restantes consideradas como genes intactos (funcionais).

# após a obtenção das sequências que não foram excluídas por *gaps*, uma árvore de máxima verossimilhança foi obtida utilizando o programa IQ-TREE após a junção das sequências com *queries* representativas:

```
iqtree -s genome_name.group_name.fa -m JTT+G4 -bb 1000
```

# Uma vez excluídas sequências adicionais que não foram excluídas em análises anteriores, o arquivo `.fasta` foi importado para um editor na qual a escolha do ATG inicial foi feita de forma manual. As sequências restantes são consideradas funcionais e exportadas para uma pasta final após a exclusão das *queries* representativas. Assim, o arquivo principal contendo todas as sequências dos grupos filogenéticos unidas é obtido com o comando que gerará o arquivo `.idlist` e um arquivo `.fasta` com todos os genes `genome_name.func.fasta`:

```
mk_funcOR_nfa.pl genome_name
```

### 3.3 Identificação dos pseudogenes e genes truncados

Para identificação de pseudogenes e genes truncados, um TBLASTN foi realizado contra o genoma utilizando todas as sequências funcionais identificadas com um *E-value*

de  $1e-20$  por duas razões: 1) o melhor *E-value* correspondente ao melhor *hit* para uma região genômica correspondente à um gene não receptor olfativo é de  $1e-18$  ou maior; 2) todos os melhores *hits* com *E-value* abaixo de  $1e-20$  obtidas de *queries* de receptores olfativos são mais similares entre genes de receptores olfativos do que qualquer outro gene não-olfativo (NIIMURA & NEI, 2005).

Após a exclusão dos genes funcionais do resultado, todas as sequências restantes foram classificadas como pseudogenes ou genes truncados. Consideramos uma cópia como pseudogene quando a sequência apresentava códons de parada, *frameshifts* (mutações por mudança de matriz de leitura) e/ou *gaps* dentro das regiões conservadas, enquanto um gene foi considerado como truncado quando uma sequência parcial de um receptor olfativo funcional estava localizada no fim de um *contig*.

A metodologia para diferenciação de cada gene restante seguiu exatamente como descrito em Niimura (2013a), Niimura (2013b), Niimura & Nei (2005) e snedo resumidamente feita da seguinte forma:

Extração dos melhores *hits* segundo o seguinte critério: ausência de códons de parada (stop códon) e identificação de *frameshifts*, considerando a distância entre o fim da sequência e o fim do *contig* contendo uma sequência menor que 30 pb.

As sequências restantes foram classificadas em três categorias: *C-missing*, *N-missing* e *NC-missing*. *C-missing* será classificado quando sua porção montante estiver no *contig*, mas sua porção jusante estiver ausente, enquanto a categoria *N-missing* corresponde ao inverso da *C-missing*. Por fim, *NC-missing* será quando ambas as porções montante e jusante da sequência estiverem ausentes.

Para sequências *C-missing*, realizamos a extensão no genoma e extração da sequência do códon ATG até o códon mais a jusante. Para sequências *N-missing*, a extração ocorreu do códon mais a montante até o códon de parada. Por fim, para as *NC-missing*, realizamos a extração do códon mais a montante até o mais a jusante.

Para cada sequência *N-missing*, escolhemos o códon de iniciação (ATG) apropriado e um alinhamento múltiplo foi realizado com as restantes. As sequências que não possuíam um N- ou C- intacto na porção terminal foram excluídas. Sequências com mais de cinco *gaps* no TM1 e no TM7 também foram excluídas para C- e N-, respectivamente. As sequências restantes foram consideradas genes truncados.

# A identificação dos genes truncados exige a obtenção de um arquivo que contém a análise N50 do genoma. Desse modo, o comando abaixo gera esse arquivo:

```
detectN.pl -n genome_name > genome_name.detectN
```

# Uma vez de posse do arquivo, a pasta deverá conter os arquivos 'genome\_name.detectN', o 'genome\_name.fna', além de uma pasta chamada 'diretorio' contendo os genes funcionais mais o arquivo genome\_name.besthit:

```
mk_truncOR.pl genome_name diretorio genome_name.fna
```

# Esse comando acima gera arquivos no formato '.nfa' (formato não tradicional, criado para fins de separação de sub produtos dos *scripts*) contendo as sequências truncadas separadas de acordo com suas respectivas categorias: *C-missing*, *N-missing*, *NC-missing*. Desse modo, para verificar se há sequências nos arquivos:

```
grep -c "^>" *.nfa
```

# Uma vez que existam sequências nos arquivos, é necessário juntá-los em um único arquivo para posterior separação em grupos filogenéticos:

```
cat genome_name.X_Mstop.fa genome_name.Xext.fa genome_name.noC_M.fa
genome_name.noN_stop.fa > genome_name.Tcand2.fa
```

# O processo de separação em grupos filogenéticos gera um arquivo '.idlist' contendo os *headers* das sequências de acordo com seus respectivos grupos, levando em consideração o arquivo OR\_group.txt, de forma semelhante à separação dos intactos:

```
python get_ORgroupTrunc.py genome_name;
```

# Uma vez gerados os arquivos '.idlist' de cada grupo, é preciso verificar quais os arquivos possuem sequências para separá-las do arquivo principal genome\_Name.Tcand2.fa:

```
grepfasta.pl genome_name.group_name.idlist genome_name.Tcand2.fa >
genome_name.Tcand2.fa;
```

# Após isso, é preciso utilizar *queries* representativas para gerar um arquivo ``.mafft'` e analisar manualmente as sequências para possíveis *gaps* em um editor:

```
cat genome_name.group_name.Tcand2.fa query_group_name.fa >
genome_name.group_name.Tcand2.fa
```

```
mafft -auto genome_name.group_name.Tcand2.fa >
genome_name.group_name.mafft
```

# Uma vez finalizada a análise das sequências, as *queries* representativas são excluídas e o restante das sequências é unida em um arquivo só:

```
cat *.Tcand2.fa > genome_name.truncated.fasta;
```

# Por fim, utiliza-se um comando para gerar o arquivo ``.idlist'` contendo os *headers* de todas as sequências:

```
mk_truncOR_nfa.pl genome_name;
```

Após excluir todas os genes truncados das sequências obtidas da primeira extração, as restantes foram consideradas pseudogenes, excluindo sequências com menos de 50 aminoácidos de comprimento.

```
# Para obtenção dos pseudogenes, a pasta precisa conter os arquivos
genome_name.besthit, genome_name.func.fasta,
genome_name.truncated.fasta e o genoma genome_name.fna:
get_pseudogenes.pl genome_name .
```

# O arquivo gerado de nome `genome_name.pseudo2.idlist` contém os *headers* das sequências que precisam também ser extraídas do genoma, o que gera um arquivo de formato ``.cutfain5'` e traduzidas com os seguintes comandos:

```
mkcutfain5pl genome_name;
```

```
cutfasta.pl genome_name.fna genome_name.cutfain5 > genome_name.seq.nfa
```

# o arquivo genome\_name.seq.nfa contém as sequências de nucleotídeos dos pseudogenes que é traduzida com o comando:

```
translate.pl genome_name.seq.nfa > genome_name.pseudogenes.fa;
```

# a separação em grupos segue a linha de separação dos genes intactos e truncados e também envolve a existência do arquivo OR\_group.txt na mesma pasta:

```
python get_ORgroup.py genome_name
```

# os arquivos '.idlist' gerados são separados em seus respectivos grupos filogenéticos a partir do arquivo genome\_name.pseudogenes.fa com o comando:

```
grepfasta.pl genome_name.group_name.idlist >
genome_name.group_name.pseudogenes.fa;
```

# por fim, utiliza-se o comando que gera o '.idlist' 'geral' e faz a união dos grupos em um só arquivo '.fasta':

```
mk_pseudoOR_nfa.pl genome_name;
```

Após essas etapas, todos os genes do repertório de um genoma foram identificados e separados em seus respectivos grupos filogenéticos utilizando uma árvore de sequências inferidas por máxima verossimilhança utilizando o programa IQ-TREE 1.6.12 com modelo de evolução JTT.

### 3.4 Distribuição cromossômica dos genes OR

*Clusters* de genes OR foram mapeados nos cromossomos de 12 espécies de anuros, que possuem o nível de montagem cromossômica do genoma, usando o pacote Chromomap v0.4.1 (ANAND & LOPEZ, 2022) no R. Para identificação dos *clusters*, nós utilizamos um *script* em *Perl* chamado *clustering2.pl* que partiu da premissa de que um

*cluster* genômico era definido como qualquer gene OR vizinho (incluindo pseudogenes e genes truncados) separados por uma distância menor do que 500Kb, independente da presença de outros genes dentro do *cluster* (YOUNG ET AL., 2022; NIIMURA & NEI, 2003). Na identificação dos *clusters*, *singletons*, ou seja, genes OR que se encontravam isolados de outros por mais de 500Kb, não foram considerados. A nomenclatura de cada *cluster* identificado pelo *script* foi feita de acordo com o número do cromossomo em que o *cluster* se encontra, seguido pelo índice numérico em ordem crescente. Desse modo, se determinado cromossomo possuía mais de um *cluster*, os *clusters* foram identificados de 1 até  $1 + n$  *clusters*.

## 4. RESULTADOS

Os resultados obtidos foram organizados em três capítulos conforme descrito a seguir:

**Capítulo I – “Tracking the diversity and chromosomal distribution of the olfactory receptor gene repertoires of three anurans species”** – Manuscrito aceito para publicação no periódico *Molecular Evolution Journal*. Artigo descreve comparativamente o repertório de ORs de três espécies de anfíbios anuros com genomas em nível cromossômico de montagem e discute a distribuição genômica de *clusters* de ORs no genoma de *Engystomops pustulosus*, *Bufo bufo* e *Hymenochirus boettgeri*, assim como a identificação de *clusters* de genes ortólogos entre as três espécies. Insights sobre a relação da quantidade de genes com história de vida das espécies são abordados.

**Capítulo II – “Intrachromosomal rearrangement and gene fractionation modulated the repertoire of olfactory receptors genes in post-polyploid species of *Xenopus* (Anura; Pipidae)”** – Manuscrito formatado para submissão no periódico *Molecular Phylogenetics and Evolution*. O artigo reavalia o repertório de *Xenopus tropicalis* e *X. laevis*, e a identificação dos genes OR de *X. borealis*. Neste manuscrito, as análises identificam os grupos de genes ortólogos (do inglês, *Orthologous Gene Group*) entre as três espécies e as possíveis implicações evolutivas que envolveram o repertório de genes OR entre espécies que possuem subgenomas resultantes de duplicação do genoma.

**Capítulo III – Análise genômica comparativa do repertório de genes OR de 28 espécies de Anuros: novos achados, lições e desafios no estudo desta família multigênica.** Dados gerados a partir de 27 espécies de anfíbios anuros e analisam a organização filogenética dos genes OR, assim como a distribuição genômica dos *clusters* nas diferentes espécies.

**4.1 Capítulo I – “Tracking the diversity and chromosomal distribution of the olfactory receptor gene repertoires of three anurans species”**

## Tracking the diversity and chromosomal distribution of the olfactory receptor gene repertoires of three anurans species

Johnny Sousa Ferreira<sup>1</sup>, Daniel Pacheco Bruschi<sup>1,\*</sup>

<sup>1</sup> Laboratório de Citogenética Evolutiva e Conservação Animal (LabCECA). Departamento de Genética.

Universidade Federal do Paraná (UFPR), Paraná, Brazil.\*

E-mail address: danielbruschi@ufpr.br (D.P. Bruschi)

### Abstract:

Olfaction is a crucial capability for most vertebrates and is realized through olfactory receptors in the nasal cavity. The enormous diversity of olfactory receptors has been created by gene duplication, following a birth-and-death model of evolution. The olfactory receptor genes of the amphibians have received relatively little attention up to now, although recent studies have increased the number of species for which data are available. This study analyzed the diversity and chromosomal distribution of the OR genes of three anuran species (*Engystomops pustulosus*, *Bufo bufo* and *Hymenochirus boettgeri*). The OR genes were identified through searches for homologies, and sequence filtering and alignment using bioinformatic tools and scripts. A high diversity of OR genes was found in all three species, ranging from 917 in *B. bufo* to 1194 in *H. boettgeri*, and a total of 2076 OR genes in *E. pustulosus*. Six OR groups were recognized using an evolutionary gene tree analysis. While *E. pustulosus* has one of the highest numbers of genes of the gamma group (which detect airborne odorants) yet recorded in an anuran, *B. bufo* presented the smallest number of pseudogene sequences ever identified, with no pseudogenes in either the beta or epsilon groups. Although *H. boettgeri* shares many morphological adaptations for an aquatic lifestyle with *Xenopus*, and presented a similar number of genes related to the detection of water-soluble odorants, it had comparatively far fewer genes related to the detection of airborne odorants. This study is the first to describe the complete OR repertoire of the three studied species and represents an important contribution to the understanding of the evolution and function of the sense of smell in vertebrates.

**Key words:** smell, adaptation, anurans, olfaction, bioinformatics.

### Introduction

Most vertebrates depend on five senses – smell, taste, hearing, sight, and touch – to detect and perceive phenomena in their external environment. One of these senses, olfaction, is essential to the survival of many vertebrates, by helping to decipher signs of potential predators, locate food sources, attract mates, identify conspecifics, support parental care, and drive many different types of social behavior (Kurian et al 2021; Yohe et al. 2022). Even among phylogenetically-proximate species, however, olfactive behavior can vary considerably, as a function of differences in lifestyle, selection pressures, and ecological constraints (Yohe et al. 2021). In fish, for example, olfactory receptor genes are specialized for the detection of water-soluble odorants, whereas in terrestrial organisms, the genes that determine the capacity to detect airborne odorants tend to be much more extensive (Niimura and Nei 2005b; Niimura 2009a, b; Sin et al. 2022; Yohe et al. 2020).

Olfactory Receptors (ORs) are present in the nasal cavity, arranged in two major organs in most tetrapods – the olfactory epithelium and the vomeronasal organ (Buck and Axel 1991; Jungblut et al. 2021). The perception of odors requires the binding of the odorant molecules to G protein-coupled receptors (GPCRs), which are encoded by a diverse, multigene family with numerous OR genes. This enormous diversity of OR genes is created mostly by gene duplication events, which follow a birth-and-death evolutionary pathway (Nei and Rooney 2005; Eirín-López et al. 2012). As a consequence of this evolutionary model, the genome typically presents a diverse mixture of intact paralogs, pseudogenes or partial OR genes, clustered within specific genomic regions (Nei and Rooney 2005). Based on gene trees, the ORs of vertebrates are classified as either type I, which includes six well-supported clades ( $\alpha$  – alpha,  $\beta$  – beta,  $\gamma$  – gamma,  $\delta$  – delta,  $\epsilon$  – epsilon, and  $\zeta$  – zeta) or type II, which has one group,  $\eta$  – eta (Niimura 2009a). Niimura & Nei (2005b) identified three other type II groups – theta, kappa, and lambda – although they were later considered to be non-ORs by Niimura (2009a), but are still used to separate ORs from non-ORs in identification analyses. The olfactory receptors encoded by alpha and gamma groups identify airborne odorants, while delta ( $\delta$ ) to eta ( $\eta$ ) detect water-soluble odorants. OR-genes from beta group can detect both airborne and water-soluble odorants (Niimura 2009a). It is interesting to note that

the amphibians are the only group that have functional OR genes of both types, I and II (Niimura 2009a; Lu et al. 2021; Han et al. 2022).

Comparative genome studies have shown a degree of variability in the number of copies and the richness of the sequences of each OR gene group in closely-related species (Silva et al. 2020; Wang et al. 2020) and there is mounting evidence of the role of ecological constraints in the evolutionary dynamics of this multigene family (Yohe et al. 2021; Yohe et al. 2022). As the OR gene repertoire is highly adaptive and interacts directly with environmental signals, recent studies have focused increasingly not just on the diversity of genes, but also the proportion of intact genes to pseudogenes (including partial copies) from an ecological perspective (Silva et al. 2020; Yole et al. 2021).

In vertebrates, the OR genes have been well studied in primates (Niimura et al. 2014; Niimura et al. 2018), rodents (Degl'Inocenti and D'Errico 2017; van der Linden et al. 2018; Freeman et al. 2020), and fish (Lv et al. 2019; Liu et al. 2021; Wang et al. 2021). Amphibians occupy a strategic position in the tree of life, representing a link between fish and the tetrapods, that is, between an aquatic and a predominantly terrestrial lifestyle. While amphibians have distinct and well-developed regions of sensory epithelia, there are important differences between the larval and adult stages (Jungblut et al. 2021), and their unique life cycle, in which they must adapt to both aquatic and terrestrial environments, implies a complex evolutionary pathway that makes this group an intriguing model for the investigation of OR gene repertoires. Despite this potential, the OR gene repertoire of the anuran genome is still poorly known (Lu et al. 2021; Han et al. 2022), and *Xenopus tropicalis* was, for many years, the only representative of the anurans available for comparative genomic analyses of the complete OR gene repertoire (Niimura 2009a, b; Lu et al. 2021). However, Han et al. (2022) recently published OR gene repertoires for 15 different anuran species, which greatly amplified perspectives on general patterns in this class of vertebrates. The present study complements these data with descriptions of the complete OR gene repertoires of three anuran species – *Engystomops pustulosus* (Leptodactylidae), *Bufo bufo* (Bufonidae) and *Hymenochirus boettgeri* (Pipidae) – and the chromosomal arrangements of the OR gene clusters. The major differences found among these species in the number and configuration of OR genes provide important insights into the evolution of olfaction in these, and other anuran species.

## Materials and Methods

### *Study species and the identification of the OR gene repertoire*

To understand the diversity of OR genes in *Engystomops pustulosus*, *Bufo bufo*, and *Hymenochirus boettgeri* genomes, homologous sequences were searched for in the genome assemblies of each species, downloaded from GenBank (<http://www.ncbi.nlm.nih.gov/genbank/>). The genome assembly of each species is summarized in Table 1.

The identification of the OR genes followed the approach described in Niimura and Nei (2005b) and Niimura (2009a, 2013a, b), which consists of searches for homologies, and the analysis and filtering of sequences using transmembrane motifs and sequence trees. These criteria were used to distinguish sequences of OR genes from those of the other chemosensory receptors and the superfamily of GPCR rhodopsins.

The first step in the analysis was to search for homologous sequences in the genomes of the study species using TBLASTN (Altschul et al., 1990). The query dataset (Table 2) for the TBLASTN search was made up of the sequences of the intact OR genes of 11 vertebrates species published by Niimura (2009a) and Olender et al. (2013). This dataset was divided into specific phylogenetic groups following the arrangement adopted by Niimura (2009a), with the best candidates for query sequences being identified using a 60% homology cut-off point in CD-HIT (Huang et al. 2010) (Table 2) to reduce computation time. An E-value of  $1 \times 10^{-5}$  was established as the cut-off point for the TBLASTN analyses.

A custom script was prepared in the GenomicRanges package of the R software to separate the lowest E-value hits for a given genomic region, with all the hits of less than 250 amino acids being excluded. The remaining sequences were assigned to a best hit file. The complete coding region of each of the remaining sequences was obtained with ORFfinder (<http://www.ncbi.nlm.nih.gov/gorf/gorf.html>) and the nucleotide sequences were translated on the ExPASy web server (<http://web.expasy.org/translate/>). The remaining sequences were aligned using the MAFFT webserver (Katoh et al. 2002; Kuraku 2013) and the Transmembrane domains (TMs) were assigned following Man et al. (2004). Sequences with more than five gaps within the TM were excluded, and a new alignment was then obtained, with the initial ATG being selected manually.

The translated sequences were aligned with the following non-olfactory GPCRs as outgroups: the alpha-1-adrenergic receptor isoform 1 (GenBank id, NP\_000671.2), beta-1 adrenergic receptor (NP\_000675.1), adenosine receptor A2b (NP\_000667.1), 5-hydroxytryptamine receptor 1B (NP\_000854.1), 5-hydroxytryptamine receptor 1F (NP\_000857.1), 5-hydroxytryptamine receptor 6

(NP\_000862.1), and the histamine H2 receptor (NP\_001010973.1). A maximum-likelihood (ML) tree was constructed based on the JTT model of amino acid evolution using the Akaike Information Criterion (AIC-c), as implemented in ModelFinder (Kalyaanamoorthy et al. 2017), which was obtained from PhyloSuite (Zhang et al. 2020), with an RAxML online (Stamatakis 2014) analysis. All the sequences plotted outside the monophyletic clade of the OR genes were discarded, with the remaining sequences being intact OR genes.

The level of conservation of the specific protein motifs of the OR sequences of different groups was evaluated to validate the intact sequences recovered from the genomes using the Multiple Em for Motif Elicitation (MEME) tool (Bailey and Elkan 1994) available at the MEME suite 5.4.1 server (<https://meme-suite.org/meme/>; Bailey et al. 2015). An ungapped matrix of the translated sequences was used for each genome, with the following parameters being applied – an initial search discovery of 20 motifs, followed by any number of repetitions, and an optimal motif width of at least 50 base pairs (bps).

The truncated sequences were assigned to one of three categories: (i) C-missing gene sequences (missing the downstream region), (ii) N-missing sequences (missing the upstream region), and (iii) NC-missing sequences (missing both the downstream and the upstream regions). These sequences were identified from the best hits in which the distance between the end of the sequence and the end of the contig containing the sequence was less than 30 bps.

To identify pseudogenes, the intact and truncated OR genes were removed from the best hit file, and the remaining sequences were considered to be pseudogenes if the E-value was over  $1 \times 10^{-20}$ , following Niimura and Nei (2005a), and had a minimum length of 50 bps, and contained disrupting mutations, such as indels, that result in frameshifts.

### *Classification of the OR genes, orthology inference and distribution of the chromosomes clusters*

The primary classification of the OR groups considered the most representative version of a given best hit sequence as an intact, truncated or pseudogene sequence. Following this classification, an RAxML (Stamatakis 2014) analysis was run to confirm whether the sequence belonged to a given OR group. This analysis considered the JTT model of amino acid evolution determined by ModelFinder (Kalyaanamoorthy et al. 2017) and was based on the Akaike Information Criterion (AIC-c), obtained in PhyloSuite (Zhang et al. 2020). The sequences were compared with those of known OR genes from *Xenopus tropicalis*, with fish zeta and theta–kappa sequences (the latter being used to separate the non-ORs from the classification), and non-olfactory GPCRs as the outgroups. All the analyses included a Bootstrap routine with 500 replicates.

The OR gene clusters were mapped on the chromosomes of each species using the ChromoMap v0.4.1 package (Anand and Lopez 2022) in R. The map contains information on each OR gene identified in the study and shows its location on the chromosomes of each species. The details are presented in Only Resource Tables S2–S4, which contain information on the number of OR genes located in each chromosome region. However, some OR genes were located in unplaced contigs, which means that their precise location on the chromosome is currently unknown. Unplaced contigs are partial DNA sequences that have been identified but not yet assembled into a chromosome sequence, so that their location on the chromosome cannot be determined accurately until more data become available.

A synteny network approach across species was also implemented using intact OR gene copies to evaluate conserved collinearity and genomic context (synteny blocks) within this multigene family and help to understand the evolutionary dynamic of the tandem-duplicated array, segmental or genome duplication, and gene transposition events through evolution (Zhao et al., 2019; Zhao et al., 2021). The analyses were performed using the *syntenet* approach (Almeida-Silva et al., 2023), an R/Bioconductor package available on (<https://bioconductor.org/packages/syntenet>) (accessed in July 2023). This package uses DIAMOND (Buchfink et al., 2015) to generate an all-vs-all reciprocal similarity search for OR database used based on the McScanX algorithm (Wang et al., 2012) to identifier paralogs and syntenic collinear blocks using whole-genome homology and gene positional information. The anchor pairs recovered were visualized with Circos (Krzywinski et al., 2009).

## **Results**

### *Evidence from the gene trees*

The complete olfactory receptor repertoire varied considerably among the different species. A total of 2076 OR genes were recovered from *E. pustulosus*, including 1080 intact genes, 919

pseudogenes, and 77 truncated genes (Fig. 1). In *B. bufo*, on the other hand, only 917 OR genes were recovered, including 745 intact genes and 172 pseudogenes, but no truncated genes, whereas 1194 OR genes were recovered from *H. boettgeri*, including 507 intact genes, 668 pseudogenes, and 19 truncated genes (Fig. 1). The *B. bufo* genome was also unusual due to the complete lack of  $\beta$  and  $\epsilon$  pseudogene sequences, despite the good assembly and coverage (64x) of this species, and its N50 value of over 600,000 (Online Resource Table S2).

The MEME analysis applied to the all intact OR genes of each genome confirmed a high level of conservation of the family-specific motifs that distinguish the ORs from the other GPCRs (Fig. 2) around the seven TMs. The most conserved motif in all three species was MA(S)YDRYV(L)AIC, which was found at the end of TM3, followed by KA(T)FSTCSS(A)H at the end of TM7, with two motifs, PM(L)LNPL(I)IYSL and LH(Q)T(S)PMYFFL(I) within the TM7 (Fig. 2). A CI(L/M)T(I/A)Q motif was also observed in the extracellular region between TM2 and TM3. The most significant motifs are shown in Online Resource (Table S5).

The evolutionary gene tree based on the ML approach recognized six monophyletic OR gene groups in the genomes of the three study species, with well-supported (>98% bootstrap values)  $\alpha$ ,  $\beta$ ,  $\gamma$ ,  $\delta$ , and  $\epsilon$  clades (type I) and other sequences in an  $\eta$  clade, classified as type II OR genes (Figs. 3 and 4). No zeta ( $\zeta$ ) homolog sequences were recovered from any of the species analyzed (Fig. 4). The phylogenetic tree of each OR group is presented separately in the Online Resource Fig. S2–S4. Species-specific subclades of  $\delta$  genes were observed in the *E. pustulosus* and *H. boettgeri* genomes (Fig. 4a), and similar  $\eta$  gene groups were found in the genomes of *E. pustulosus* and *B. bufo* (Fig. 4b). This may represent recent species-specific gene duplication, which is reflected in the short branch length.

### *Distribution of the orthologous gene clusters at chromosomes*

The mapping of the chromosomal arrangement of the members of the different OR group revealed clusters of OR genes scattered throughout the chromosomes. In the present study, a genomic cluster was defined as any neighbouring OR genes (including pseudogenes and truncated genes) separated by a distance of less than 500 kb, regardless of the presence of other genes within the cluster (Young et al. 2002; Niimura and Nei 2003). For simplicity, singleton OR genes were also considered to be clusters. The clusters were named according to the number of the chromosome on which they were located, together with an index number for each species (see Only Resource Tables S2–S4). Based on this definition, 170 clusters were identified in *E. pustulosus* (including 96 singletons), 59 in *B. bufo* (12 singletons), and 82 clusters in *H. boettgeri* (32 singletons).

The  $\gamma$  group genes, which were found on every chromosome in the *E. pustulosus* ( $2n=22$ ) and *H. boettgeri* ( $2n=18$ ) karyotypes, were distributed randomly, with clusters of sequences arranged along the entire chromosome (Figs. 5a and 5c, respectively). In *E. pustulosus*, in particular, some chromosome pairs have equidistant, interspersed clusters of  $\gamma$  group genes, as observed in pairs 5–11 (Fig. 5a). In the *B. bufo* ( $2n=22$ ) genome, by contrast, the  $\gamma$  group genes are found mostly in the subterminal regions of some chromosomes (e.g., pairs 1, 4, 6–9, 11) while in pairs 2, 3, and 10, there were interstitial clusters of  $\gamma$  group genes (Fig. 5b), while no OR genes were identified on pair 5. Despite the large number of  $\gamma$  gene clusters found on some chromosomes, the fine-scale analysis of each genomic cluster often revealed phylogenetically-related gene copies (subclades) in the gene tree (Figs. S1–S4). This pattern was not observed in all genomic clusters, however, given that many clusters include copies of genes that are related distantly in the gene tree (Fig. S1, Tables S2–S4).

The subtelomeric regions of chromosome 3 of *B. bufo* and chromosome 2 of *H. boettgeri* presented a sequential genomic arrangement with closed clusters from all six clades, i.e.,  $\alpha$ ,  $\beta$ ,  $\gamma$ ,  $\delta$ ,  $\epsilon$ , and  $\eta$  (Figs. 4b and 4c). The sequential order and size of each OR group in this genomic arrangement varied between the two species, although clusters of intact copies were often followed by pseudogene copies (Figs. 5b and 5c). This arrangement was not observed in the *E. pustulosus* genome, however, in which a dispersive pattern of different OR gene groups was observed among the different chromosomes (Fig. 5a). In this species, the  $\alpha$  clusters were located on chromosomes 1 and 3, while the  $\delta$  clusters were located on chromosomes 2, 4, 5, and 9, a singleton  $\epsilon$  cluster was exclusive to chromosome 11, and an  $\eta$  cluster was located on chromosome 2 (Fig. 4a). While some functional  $\beta$  genes were identified, they were all found within contigs whose location could not be determined, so their chromosomal distribution remains unclear (Table S2).

Paralogs and syntenic colinear blocks in chromosomes appear as interconnected sequences in our circle plot results (Online Resource Fig. S5). More prominent size clusters observed in our circle plot results reflect the tandem-duplicated events that occurred during the evolution of the OR genes in each

genome (Online Resource Fig 5S). The synteny network clusters reinforced the genomic pattern of cluster gene organization observed at the chromosome level. They supported the profile of the copy duplications of  $\gamma$  group genes and massive dispersion across the chromosomes across species (Online Resource Fig 5S). The synteny network clusters profiled gene copy number (0, 1, 2, and  $\geq 3$ ) observed in synteny network inferences across species reinforced signature of the extensive gene duplications of the gamma group in all species analyzed, evidenced by significant clusters with two or more copies (orange and red columns in heatmap figure, Online Resource Fig 5S).

## Discussion

The present study provides the first identification and analysis of the the whole OR gene repertoire of three anuran amphibians, which represents an important advance in the understanding of the organization of this multigene family in the anuran genome. Several shared sets of OR genes were identified in the three study species, which have undergone distinct evolutionary processes during their adaptive history, after being transmitted vertically through the evolutionary lineage (Lu et al. 2021; Niimura 2009a, b). This was revealed by the maintenance of the gene groups present in the repertoire of OR genes in the anuran genome. The OR genes of the  $\delta$  and  $\epsilon$  (type I) and  $\eta$  (type II) groups, which codify the ORs that detect water-soluble odorants, are present only in teleost fish and amphibians (Niimura 2009a), and appear to represent the retention of ancestral traits in the anurans, their retraction in the Caudata, and possible loss in the caecilians (Lu et al. 2021). These OR groups are well represented in the olfactory repertoires of all three study species (*E. pustulosus*, *H. boettgeri*, and *B. bufo*), which is consistent with the scenario observed in the genomes of other anurans (Niimura 2009a, b; Lu et al. 2021). Species-specific clades of intact genes were detected in the *E. pustulosus* and *B. bufo* genomes, which suggest the recent amplification of copies. In general, the anuran genome has a high proportion (61–100%) of intact vs. disrupted (or truncated) copies of the genes of the  $\delta$ ,  $\epsilon$ , and  $\eta$  groups, as observed here in *E. pustulosus* and *B. bufo*, and in other anurans by Lu et al. (2021). In *H. boettgeri*, however, this proportion declined to 40–42%. This contradicts the expected scenario of a more stable repertoire of these OR groups, as observed in the closely related frogs *Xenopus tropicalis* and *Xenopus laevis*, that, like *H. boettgeri*, are well adapted for an aquatic lifestyle (Duellman and Trueb 1986). Selection pressures acting on aquatic frogs are thought to favor the preservation of the OR genes related to the detection of water-soluble odorants, although the considerable number of pseudogenes found in the *H. boettgeri* genome demands further attention. In particular, there was an expansion of  $\beta$  pseudogenes on chromosomes 2 and 8 of this aquatic species. Beta group genes are found in aquatic and terrestrial vertebrates (Niimura and Nei 2005) and are used for the detection of both airborne and water-soluble odorants. A similar increase in the  $\beta$  pseudogenes has been found in other aquatic species, such as *X. tropicalis* and *X. laevis*, and may be the result of adaptive strategies related to the evolution of an aquatic lifestyle.

The  $\alpha$ ,  $\beta$ , and  $\gamma$  OR-groups are considered to be specialized for the detection of airborne odorants, and as it is assumed that there are more airborne than aquatic odorants, a much greater number of functional genes would be expected to be adapted for the detection of these substances. As expected, the  $\gamma$  group had the highest number of intact and pseudogene copies in the genomes of the three study anurans, with a total of almost one thousand functional genes. The  $\gamma$  group has undergone a significant expansion in mammals, birds (Yohe et al. 2020; Sin et al. 2022; Zhou et al. 2022), salamanders, and caecilians (Lu et al. 2021). The present study confirmed that *E. pustulosus*, together with *Rhinella marina* and *Bufo gargarizans*, are the anurans with the largest numbers of  $\gamma$  group genes identified up to now (Lu et al. 2021). While *E. pustulosus* and *H. boettgeri* had similar proportions of intact and pseudogenes, *B. bufo* had the highest proportion of intact genes (81%), similar to that of *Microcaecilia unicolor*, as reported by Lu et al. (2021).

*Bufo bufo* had the smallest proportion of pseudogenes identified in any of the anuran species studied up to now. It had no truncated genes and lacked pseudogenes of the  $\beta$  or  $\gamma$  groups. Niimura and Nei (2007), Khan et al. (2015), and Vandeweghe et al. (2016) proposed that the number of functional ORs in the genome of an anuran species is related to its life history. Many species of the family Bufonidae are relatively terrestrial in comparison with other anurans (Duellman and Trueb 1986; Escoriza and Hassine 2019). In this case, *B. bufo* should have a large number of  $\alpha$ ,  $\beta$ , and  $\gamma$  group genes. Despite its tolerance of dehydration (Jorgensen, 1994), this toad still depends on water for breeding and the development of the tadpoles, which implies that it would have retained some genes related to olfaction in the aquatic medium.

The correlation between expansion/contraction in OR genes repertoire and niche adaptation is a promising frontier to compression about the evolution of this multigene family in vertebrates (Hughes et

al., 2018; Yohe et al., 2022). For example, the OR repertoire of the mammalian with a solitary lifestyle exhibits a strong signature of gene expansion when compared with species with a social lifestyle (Hughes et al. 2018). However, the correlation between changes in different OR classes to the species tree and organismal lifestyle represents a significant challenge in Anura species due to the high diversity in the ecological and behavioral traits directly related to local adaptations in each evolutionary lineage. Evaluations at the intrageneric or intrafamilial level may provide a more realistic scenario about the adaptive process and ecological constraints that modulate the composition of OR gene repertoire in Anurans genomes.

### *Chromosomal organization of the OR genes reflects the conservation of ancient patterns in the Anura and a synteny break in the *E. pustulosus* genome*

The subtelomeric regions of chromosome 3 of *B. bufo* and chromosome 2 of *H. boettgeri* have clusters of each of the OR groups ( $\alpha$ ,  $\beta$ ,  $\gamma$ ,  $\delta$ ,  $\epsilon$ , and  $\eta$ ). A similar feature has already been observed in *X. tropicalis* (Niimura 2009a) and *B. gargarizans* (Lu et al. 2021). Long-term chromosomal synteny in the anuran lineage has been revealed by a number of different analytical approaches (Sun et al. 2015; Liedtke et al. 2018; Li et al. 2019; Lu et al. 2021). This extensive chromosome synteny based on genome-scale ortholog alignment indicates that chromosome 2 is homologous in *X. tropicalis*, *X. laevis*, *H. boettgeri*, and *E. pustulosus* (Mudd 2019) and homologous to chromosome 3 of *B. gargarizans* (Lu et al. 2021) and *B. bufo* (Streicher et al. 2021). The arrangement of OR genes in the subterminal region of chromosomes 2 and 3 is consistent with this hypothesis of homology, and confirms the preservation of large segments of the ancestral chromosomes throughout the evolutionary history of the Anura (Mudd 2019; Lu et al. 2021).

The evidence presented here indicates that *E. pustulosus* is an exception among the anuran lineages analyzed up to now, given that a synteny break was observed in the subterminal region of chromosome 2, a detail not found in any other anurans for which the chromosomal distribution of the OR genes is known. Interestingly, no large clusters of OR genes were observed in this species, but rather, a spread of genes across different genomic regions. This dispersive chromosomal distribution may be the result of either chromosomal rearrangements (Guo and Kim 2007; Nozawa and Nei 2007; Conceição and Aguadé 2008) or the effects of transposable elements (Niimura and Nei 2005; McKenzie and Kronauer 2018). These two processes are not mutually exclusive and may have operated in synergy. Transposable elements may provide a substrate for ectopic recombination and unequal crossover (Yang et al. 2008), as well as contributing to the dispersal of a number of the singletons found in the chromosomes of *E. pustulosus* and the observed synteny break. Further, more detailed studies will be needed to assess the interplay between transposable elements and the OR genes, in order to better understand the genetic mechanisms involved in this process.

The  $\gamma$  group genes were dispersed widely among the chromosomes in the genomes analyzed here. This distribution pattern is expected due to the abundance of clusters of this OR group in the anuran genome, as also observed in *B. gargarizans* (Lu et al. 2021) and *X. tropicalis* (Niimura 2009a, b). The most interesting case was that of the *E. pustulosus* genome, which presented a pan-genomic distribution of clusters of functional  $\gamma$  group genes.

## Conclusions and future perspectives

Over the past two decades, the progressive identification of OR genes have led to the establishment of a growing number of databases that provide OR profiles for systematic comparisons across a wide range of organisms, including anurans (Han et al. 2022, Marengo et al. 2016). Research on the OR genes of amphibians has provided valuable insights into the genetics of their sense of smell, and the evolution of the olfactory system in this group of vertebrates, which may have been influenced by the number and different types of ORs, selective pressures and ecological constraints, related to their varying modes of life. The systematic, comparative analysis of these genes can provide important clues on the evolution of olfactory receptors and their function. The mapping of the genes that codify these receptors is still incipient, however, and further research will help to fill the existing knowledge gaps on the evolution of OR multigene repertoires, in particular for detailed comparisons among different animal groups, and a fundamental perspective on the diversity of OR genes and the evolution of chemosensory mechanisms. The present study provided a wealth of information on the chromosomal organization and dynamics of this rapid-evolving multigene family in anurans.

## Acknowledgements

We would like to thank the Brazilian Coordination for Higher Education Personnel Training (CAPES/PROAP – Finance Code 001 and CAPES-Print) for the scholarships provided to JFS. DPB is grateful the Brazilian National Council for Scientific and Technological Development (CNPq #303646/2021-7 and CNPq #401024/2022-9). We are grateful to Dr. Fabricio Almeida-Silva for helping with *syntenet* package and anonymous reviewers for their valuable criticisms that helped to improve this paper.

## References

- Abaffy T, Malhotra A, Luetje CW (2007) The molecular basis for ligand specificity in a mouse olfactory receptor: a network of functionally important residues. *J of Bio Chem* 282(2):1216–1224.
- Almeida-Silva F, Zhao T, Ullrich KK, Schranz ME, Peer YV (2023) *syntenet*: an R/Bioconductor package for the inference and analysis of synteny networks. *Bioinformatics*, 39(1), btac806.
- Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. *J of Mol Bio* 215(3):403–410.
- Anand L, Lopez, CMR (2022) ChromoMap: R package for interactive visualization of multi-omics data and annotation of chromosomes. *BMC Bioinfo* 23(33):1–9.
- Bailey TL, Elkan, C (1994) Fitting a mixture model by expectation maximization to discover motifs in biopolymers. *Proc Int Conf Intell Syst Mol Bio* 2:28–36.
- Bailey TL, Johnson J, Grant CE, Noble WS (2015) The MEME Suite. *Nuc Ac Res* 43(W1):W39–W49.
- Barrasso DA, Cajade R, Nenda SJ, Baloriani G, Herrera R (2009) Introduction of the American bullfrog *Lithobates catesbeianus* (Anura: Ranidae) in natural and modified environments: an increasing conservation problem in Argentina. *South Am J Herp* 4:69–75.
- Buck L, Axel R (1991) A novel multigene family may encode odorant receptors: a molecular basis for odor recognition. *Cell* 65:175–187.
- Conant GC, Wolfe KH (2008) Turning a hobby into a job: how duplicated genes find new functions. *Nat Rev Genet* 9:938–950.
- Conceição I C, Aguadé M (2008) High incidence of interchromosomal transpositions in the evolutionary history of a subset of OR genes in *Drosophila*. *J Mol Evol* 66: 325–332.
- Degl'Innocenti A, D'Errico A (2017) Regulatory features for odorant receptor genes in the mouse genome. *Front Genet* 8(9):1–8.
- Duellman WE, Trueb L (1986) *Biology of Amphibians*. Johns Hopkins University Press, Baltimore.
- Eirín-López JM, Rebordinos L, Rooney AP, Rozas J. (2012) The birth-and-death evolution of multigene families revisited. *Genome Dynamics* 7:170–196.
- Emms D, Kelly S (2019) OrthoFinder: phylogenetic orthology inference for comparative genomics. *Genome Bio* 20(1):1–14. <https://doi.org/10.1186/s13059-019-1832-y>.
- Escoriza D, Hassine J (2019) North-western Africa amphibians. In: Escoriza D, Hassine J (eds) *Amphibians of North Africa*. Academic Press, pp 75–229.
- Faivovich J, Haddad CFB, Garcia PCA, Frost DR, Campbell JA, Ward, W (2005) Systematic review of the frog family Hylidae, with special reference to Hylinae: phylogenetic analysis and taxonomic revision. *Bull Am Mus nat* 294:1–240.
- Fei L, Ye C-Y, Huang Y-A, Liu M-Y (1999) *Atlas of Amphibians of China*. Henan Science and Technical Press, Zhengzhou.
- Ficetola GF, Coïc C, Detaint M, Berroneau M, Lorvelec O, Miaud C (2006) Pattern of distribution of the American bullfrog *Rana catesbeiana* in Europe. *Bio Inv* 9:767–772.
- Fleischer J, Breer H, Strotmann J (2009) Mammalian olfactory receptors. *Front Cell Neurosci* 3:9.
- Fleming RI, Mackenzie CD, Cooper A, Kennedy MW (2009) Foam nest components of the túngara frog: a cocktail of proteins conferring physical and biological resilience. *Proc R Soc* 276(1663):1787–1795.
- Freeman AR, Ophir AG, Sheeran MJ (2020) The giant pouched rat (*Cricetomys ansorgei*) olfactory receptor repertoire. *PLOS One* 15(4):e0221981.
- Frost DR, Grant T, Faivovich J, Bain RH, Haas A, Haddad CFB, De Sá RO, Channing A, Wilkinson M, Donnellan SC, Raxworthy CJ, Campbell JA, Blotto BL, Moler P, Drewes RC, Nussbaum RA, Lynch JD, Green DM, Wheeler WC (2006) The amphibian tree of life. *AMNH* 2006(297):1–300.
- Glusman G, Yanai I, Rubin I, Lancet D (2001) The complete human olfactory subgenome. *Genome Res* 11(5):685–702.
- Guo S, Kim J (2007) Molecular evolution of *Drosophila* odorant receptor genes. *Mol Biol Evol* 24: 1198–1207.

- Han W, Wu Y, Zeng L, Zhao S (2022) Building the Chordata Olfactory Receptor Database using more than 400,000 receptors annotated by Genome2OR. *Sci China Life Sci* 65:2539–2551.
- Hayden S, Teeling E (2014) The molecular biology of vertebrate olfaction. *Anat Rec* 297:2216–2226.
- Heyer WR, Rand AS (1977) Foam Nest Construction in the Leptodactylid Frogs *Leptodactylus pentadactylus* and *Physalaemus pustulosus* (Amphibia, Anura, Leptodactylidae). *J Herpetol* 11:225–228.
- Huang Y, Niu B, Gao Y, Fu L, Li W (2010) CD-HIT Suite: a web server for clustering and comparing biological sequences. *Bioinformatics* 26:680–682.
- Hughes GM, Boston ESM, Finarelli JA, Murphy WJ, Higgins DG, Teeling EC (2018) The Birth and Death of olfactory receptors gene families in mammalian niche adaptation. *Molecular Biology and Evolution* 35(6): 1390-1406.
- Jane T (2019) Development of central respiratory control in anurans: the role of neurochemicals in the emergence of air-breathing and the hypoxic response. *RSPNB* 270:103266. <https://doi.org/10.1016/j.resp.2019.103266>.
- Joger U (1982) Zur herpetofaunistik kameruns (II) [On the herpetofaunistics of Camaroon (II)]. *Bonner Zoologische Beltrage* 33:313–342.
- Jorgensen CB (1994) Water economy in a terrestrial toad (*Bufo bufo*), with special reference to cutaneous drinking and urinary bladder function. *Comp Biochem Physiol* 109A(2):311–324.
- Jungblut LD, Reiss JO, Pozzi, AG (2021) Olfactory subsystems in the peripheral olfactory organ of anuran amphibians. *Cell Tissue Res* 383(1):289–299.
- Kalyaanamoorthy S, Minh BQ, Wong TKF, Haeseler A, Jermiin LS (2017) ModelFinder: fast model selection for accurate phylogenetic estimates. *Nat Methods* 14:587–589.
- Katoh K, Rozewicki J, Yamada KD (2019) Brief *Bioinformatics* 20(4):1160–1166.
- Khan I, Yang Z, Maldonado E, Li C, Zhang G, Gilbert MTP, Jarvis ED, O'Brien SJ, Johnson WE, Antunes A (2015) Olfactory receptor subgenomes linked with broad ecological adaptation in Sauropsida. *Mol Bio Evol* 32:2832–2843.
- Kosmala GK, Brown GP, Shine R, Christina K (2020) Skin resistance to water gain and loss has changed in cane toads (*Rhinella marina*) during their Australian invasion. *Ecol Evol* 00:1–9.
- Krzywinski MI, Schein JE, Birol I, Connors J, Gascoyne R, Horsman D, Jones SJ, Marra MA (2009) Circos; an information aesthetic for comparative genomics. *Genome Res* 19(9):1639–1645. <https://doi.org/10.1101/gr.092759.109>.
- Kurak, S, Zmasek CM, Nishimura O, Katoh K (2013) aLeaves facilitates on-deman exploration of metazoan gene family trees on MAFFT sequence alignment server with enhanced interactivity. *Nucleic Acids Res* 41(W1):W22–W28.
- Leivas PT, Moura MO, Fávoro LF (2012) The reproductive biology of the invasive *Lithobates catesbeianus* (Amphibia: Anura). *J Herpeto* 46(2):153–161.
- Li J, Yu H, Wang W, Fu C, Zhang W, Han F, Wu H (2019) Genomic and transcriptomic insights into molecular basis of sexually dimorphic nuptial spines in *Leptobrachium leishanense*. *Nat Commun* 10:5551. <https://doi.org/10.1038/s41467-019-13531-5>.
- Li W, Godzik A (2006) Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics* 22(13):1658–1589.
- Liedtke HC, Gower DJ, Wilkinson M, Gomez-Mestre I (2018) Macroevolutionary shift in the size of amphibian genomes and the role of life history and climate. *Nat Ecol Evol* 2:1792–1799.
- Liu H, Chen C, Lv M, Liu N, Hu Y, Zhang H, Enbody ED, Gao Z, Andersson L, Wang W (2021) A chromosome-level assembly of blunt snout bream (*Megalobrama amblycephala*) genome reveals an expansion of olfactory receptor genes in freshwater fish. *Mol Bio Evol* 38(10):4238–4251.
- Lu B, Jiang J, Wu H, Chen X, Song X, Liao W, Fu J (2021) A large genome with chromosome-scale assembly sheds light on the evolutionary success of a true toad (*Bufo gargarizans*). *Mol Ecol* 21:1256–1273.
- Lu S-Q, Yang D-T (1995) A study of relationships among ranid frogs of the genera *Nanorana* and *Altirana* in the Transhimalaya Mountains of China. *Asiat Herpetol Res* 6:73–77.
- Lv L, Liang X, He S (2019) Genome-wide identification and characterization of olfactory receptor genes in Chinese Perch, *Siniperca chuatsi*. *Genes* 10(2):178.
- Lynch JD (2006) The tadpoles of frogs and toads found in the lowlands of northern Colombia. *Rev Acad Colomb Cienc* 30(116):443–457.
- Man R, Gilad Y, Lancet D (2004) Prediction of the odorant binding site of olfactory receptor proteins by human-mouse comparisons. *Protein Sci* 13:240–254.
- Marengo L, Wang R, McDougal R, Olender T, Twik M, Bruford E, Liu X, Zhang J, Lancet D, Shepherd G, Crasto C (2016) ORDB, HORDE, ODORactor and other on-line knowledge resources of olfactory receptor-odorant interactions. *Database* 2016:1–10.

- McKenzie S K, Kronauer D J C (2018) The genomic architecture and molecular evolution of odorant receptors. *Genome research* 28:1757-1765.
- Medeiros CI, Both C, Kaefer IL, Cechin SZ (2016) Reproductive phenology of the American Bullfrog in subtropical Brazil: photoperiod as a main determinant of seasonal activity. *An Acad Bras Cienc* 88(3):1909–1921.
- Mombaerts P (2004) Genes and ligands for odorant, vomeronasal and taste receptors. *Nat Rev Neurosci* 5:263–278.
- Mudd AB (2019) Comparative genomics and chromosome evolution. Dissertation, University of California, Berkeley.
- Murrell B, Wertheim JO, Moola S, Weighill T, Scheffler K, Pond SLK (2012) Detecting individual sites subject to episodic diversifying selection. *PLOS Genetics* 8(7):1–10.
- Nei M, Rooney AP (2005) Concerted and birth-and-death evolution of multigene families. *Annu Rev Genet* 39:121–152.
- Nieuwkoop PD, Faber J (1994) Normal table of *Xenopus laevis* (Daudin): a systematical & chronological survey of the development from the fertilized egg til the end of metamorphosis. Garland Science, New York.
- Niimura Y (2009a) On the origin of evolution of vertebrate olfactory receptor genes: comparative genome analysis among 23 chordate species. *Genome Bio Evol* 1:34–44.
- Niimura Y (2009b) Evolutionary dynamics of olfactory receptor genes in chordates: interaction between environment and genomic contents. *Hum Genomics* 4(2):107–118.
- Niimura Y (2012) Olfactory receptor multigene Family in vertebrates: from the viewpoint of evolutionary genomics. *Curr Genomics* 13:103–114.
- Niimura Y (2013a) Identification of chemosensory receptor genes from vertebrate genomes. *Methods Mol Bio* 1068:95–105.
- Niimura Y (2013b) Identification of olfactory receptor genes from mammalian genome sequences. *Methods Mol Bio* 1003:39–49.
- Niimura Y, Matsui A, Touhara K. (2014) Extreme expansion of the olfactory receptor gene repertoire in African elephants and evolutionary dynamics of orthologous gene groups in 13 placental mammals. *Genome Res* 24(9):1485–1496.
- Niimura Y, Nei M (2003) Evolution of olfactory receptor genes in the human genome. *PNAS* 100(21):12235–12240.
- Niimura Y, Nei M (2005a) Evolutionary changes of the number of olfactory receptor genes in the human and mouse lineages. *Gene* 346:23–28.
- Niimura Y, Nei M (2005b) Evolutionary dynamics of olfactory receptor genes in fishes and tetrapods. *PNAS* 102(17):6039–6044.
- Niimura Y, Nei M (2007) Extensive gains and losses of olfactory receptor genes in mammalian evolution. *PLoS One* 2:e708.
- Niimura Y, Nei M (2008) The evolution of animal chemosensory receptor gene repertoires: roles of chance and necessity. *Nat Rev Genet* 9:951–963.
- Niimura Y, Nei M (2006) Evolutionary dynamics of olfactory and other chemosensory receptor genes in vertebrates. *J Hum Genet* 51:505–517.
- Nozawa M, Nei M (2007) Evolutionary dynamics of olfactory receptor genes in *Drosophila* species. *Proc Natl Acad Sci* 104: 7122–7127.
- Okada Y (1927) Frogs in Japan. *Copeia* 158:161–166.
- Olender T, Nativ N, Lancet D (2013) HORDE: comprehensive resource for olfactory receptor genomics. *Methods Mol Biol* 1003:23–38.
- Phillippy AM (2017) New advances in sequence assembly. *Genome Res* 27:11–13.
- Pond SLK, Frost SDW (2005) Not so different after all: a comparison of methods for detecting amino acid sites under selection. *Mol Biol Evol* 22(5):1208–1222.
- Ren X, Chang S, Laframboise A, Green W, Dubuc R, Zielinski B (2009) Projections from the accessory olfactory organ into the medial region of the olfactory bulb in the sea lamprey (*Petromyzon marinus*): a novel vertebrate sensory structure? *J Comp Neurol* 516:105–116.
- Rodel MO (2000) Herpetofauna of West Africa: Amphibians of the western savannah. Frankfurt am Main, Germany.
- Ron S, Narváez A, Romero G (2014) Reproduction and spawning behavior in the frog, *Engystomops pustulatus* (Shreve 1941). *Amphib Reptile Conserv* 8:25–32.
- Ryan MJ (1985) The Túngara Frog: a study in sexual selection and communication. University of Chicago Press, Chicago.
- Ryan MJ, Bernal XE, Rand AS (2007) Patterns of mating call preferences in túngara frogs, *Physalameus pustulosus*. *J Evol Bio* 20(6):2235–2247.

- Ryan MJ, Rand AS (2003) Mate recognition, in túngara frogs: a review of some studies of brain, behavior, and evolution. *Acta Zool Sin* 49:721–726.
- Saito S., Ohkita M, Saito CT, Takahashi K, Tominaga M, Ohta T (2016) Evolution of heat sensors drove shifts in thermosensation between *Xenopus* species adapted to different thermal niches. *J Biol Chem* 291(21):11446–11459.
- Sanabria EA, Quiroga LB, Acosta JC (2005) Introducción de *Rana catesbeiana* (rana toro), em ambientes precordilleranos de la provincia de San Juan, Argentina. *Multequina* 14:65–68.
- Schoegel LM, Daszak P, Cunningham AA, Speare R, Hill B (2010) Two amphibian diseases, chytridionmycosis and ranaviral disease, are now globally notifiable to the World Organization for Animal Health (OIE): an assessment. *Dis Aquat Org* 92(2–3):101–108.
- Seebacher F, Alford RA (1999) Movement and microhabitat use of a terrestrial amphibian (*Bufo marinus*) on a tropical island: seasonal variation and environmental correlates. *J Herpetol* 33(2):208–214.
- Segeman B (2020) The most frequently used sequencing technologies and assembly methods in different time segments of the bacterial surveillance and RefSeq genome databases. *Front Cell Infect Microbiol* 10:527102.
- Silva ET, Reis EP, Feio RN, Filho OPR (2009) Diet of the invasive frog *Lithobates catesbeianus* (Shaw, 1802) (Anura: Ranidae) in Viçosa, Minas Gerais state, Brazil. *S Am J Herpetol* 4(3):286–294.
- Silva MC, Chibucos M, Munro JB, Daugherty S, Coelho MM, Silva JC (2020) Signature of adaptive evolution in olfactory receptor genes in Cory's Shearwater supports molecular basis of smell in procellariiform seabirds. *Sci Rep* 10:1–11.
- Sin SYW, Cloutier A, Nevitt G, Edwards SV (2019) Olfactory receptor subgenome and expression in a highly olfactory procellariiform seabird. *Genetics* 220(2):1–16.
- Sin SYW, Cloutier A, Nevitt G, Edwards SV (2022) Olfactory receptor subgenome and expression in a highly olfactory procellariiform seabird. *Genetics* 220(2):1–16. <https://doi.org/10.1093/genetics/iyab210>.
- Stamatakis A (2014) RAxML Version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 30(9):1312–1313.
- Steiger SS, Fidler AE, Kempnaers B (2009) Evidence for increased olfactory receptor gene repertoire size in two nocturnal bird species with well-developed olfactory ability. *BMC Evol Biol* 9:1–11.
- Still MB, Lea AM, Hofmann HA, Ryan MJ (2019) Multimodal stimuli regulate reproductive behavior and physiology in male túngara frogs. *Horm Behav* 115:1–11.
- Streicher JW, Wellcome Sanger Institute Tree of Life programme (2021) The genome sequence of the common toad, *Bufo bufo* (Linnaeus, 1758). *Wellcome Open Res* 6:281.
- Sun J, Lu F, Luo Y, Bie L, Xu L, Wang Y (2023) OrthoVenn3: an integrated platform for exploring and visualizing orthologous data across genomes. *Nucleic Acids Research* 23: W397–W403.
- Sun Y, Xiong Z, Xiang X, Liu S, Zhou W, Tu X, Zhong L, Wang L, Wu D, Zhang B, Zhu C, Yang M, Chen H, Li F, Zhou L, Feng S, Huang C, Zhang G, Irwin D, Hillis DM, Murphy RW, Yang H, Che J, Wang J, Zhang Y. (2015) Whole-genome sequence of the Tibetan frog *Nanorana parkeri* and the comparative evolution of tetrapod genomes. *PNAS* 112(11):E1257–E1262.
- Touhara K (2007) Deorphanizing vertebrate olfactory receptors: recent advances in odorant-response assays. *Neurochem Int* 51(2–4):132–139.
- Van der Linden C, Jakob S, Gupta P, Dulac C, Santoro SW (2018) Sex separation induces differences in the olfactory sensory repertoires of male and female mice. *Nat Commun* 4(9):5081.
- Vandeweghe MW, Mangum SF, Gabaldón T, Castoe TA, Ray DA, Hoffman FG (2016) Contrasting patterns of evolutionary diversification in the olfactory repertoires of reptile and bird genomes. *Genome Biol Evol* 8(3):470–480. <https://doi.org/10.1093/gbe/evw013>.
- Wang H, Chen L, Dong C, Chen B, Li B, Li X, Xu P (2021) Genome-wide identification and characterization of olfactory receptor genes in common carp (*Cyprinus carpio*). *Gene* 777:145468.
- Wang K, Tian S, Galindo-González J, Dávalos LM, Zhang Y, Zhao H (2020) Molecular adaptation and convergent evolution of frugivory in Old World and neotropical fruit bats. *Mol Biol* 29:4366–4381.
- Yang S, Arguello JR, Li X, Ding Y, Zhou Q, Chen Y, Zhang Y, Zhao R, Brunet F, Peng L, Long M, Wang W (2008) Repetitive element-mediated recombination as a mechanism for new gene origination in *Drosophila*. *PLoS Genet* 4: e3.
- Yohe LR, Fabbri M, Hanson M, Bhullar, BS (2020) Olfactory receptor gene evolution is unusually rapid across Tetrapoda and outpaces chemosensory phenotypic change. *Curr Zool* 66(5):505–514.
- Yohe LR, Fabbri M, Lee D, Davies KTJ, Yohe TP, Sánchez MKR, Rengifo EM, Hall RP, Mutumi G, Hedrick BP, Sadier A, Simmons NB, Sears KE, Dumont E, Rossiter SJ, Bhullar BS, Dávalos LM (2022) Ecological constraints on highly evolvable olfactory receptor genes and morphology in neotropical bats. *Evol* 76(10):2347–2360.

- Yohe LR, Leiser-Miller LB, Kaliszewska ZA, Donat P, Santana SE, Dávalos LM (2021) Diversity of olfactory receptor repertoires is associated with dietary specialization in a genus of frugivorous bat. *G3* 11(10):1–7.
- Young JM, Friedman C, Williams EM, Ross JA, Tonnes-Priddy L, Trask BJ (2002) Different evolutionary processes shaped the mouse and human olfactory receptor gene families. *Hum Mol Genet* 11(5):535–546.
- Young JM, Shykind BM, Lane RP, Tonnes-Priddy L, Ross JA, Walker M, Williams EM, Trask BJ (2003) Odorant receptor expressed sequence tags demonstrate olfactory expression of over 400 genes, extensive alternative splicing and unequal expression levels. *Genome Biol* 4(11):R71. <https://doi.org/10.1186/gb-2003-4-11-r71>.
- Young JM, Trask BJ (2002) The sense of smell: genomics of vertebrate odorant receptors. *Hum Mol Genet* 11(10):1153–1160.
- Zhang D, Gao F, Jakovlić I, Zou H, Zhang J, Li WX, Wang GT (2020) PhyloSuite: an integrated and scalable desktop platform for streamlined molecular sequence data management and evolutionary phylogenetics studies. *Mol Ecol Resour* 20(1):348–355.
- Zhang X, Firestein S (2002) The olfactory receptor gene superfamily of the mouse. *Nat Neurosci* 5(2):124–133.
- Zhang X, Rodriguez I, Mombaerts P, Firestein S (2004) Odorant and vomeronasal receptor genes in two mouse genomes assemblies. *Genomics* 83:802–811.
- Zhang Y, Hoon MA, Chandrashekar J, Mueller KL, Cook B, Wu C, Zuker CS, Ryba JP (2003) Coding sweet, bitter, and umami tastes: different receptor cells sharing similar signaling pathways. *Cell* 112:293–301.
- Zhao T, Schranz E M (2019) Network-Based Microsynteny Analysis Identifies Major Differences and Genomic Outliers in Mammalian and Angiosperm Genomes. *Proceedings of the National Academy of Sciences* 116 (6): 2165–2174.
- Zhao T, Zwaenepoel A, Xue JY, Kao SM, Li Z, Schranz ME, Peer YV (2021) Whole-Genome Microsynteny-Based Phylogeny of Angiosperms. *Nature Communications* 12 (1): 1–14.
- Zhou C, Liu Y, Zheng X, Shang K, Cheng M, Wang L, Yang N, Yue B (2022) Characterization of olfactory receptor repertoires provides insights into the high-altitude adaptation of the yak based on the chromosome-level genome. *Int J Biol Macromol* 209:220–230.
- Zozulya S, Echeverri F, Nguyen T (2001) The human olfactory receptor repertoire. *Genome Biol* 1(6): research0018.1.

## Competing Interests and Funding

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Figure Captions

**Fig. 1.** Numbers of OR genes identified for each of the three species. Total number of genes are represented by “I”, “T”, and “P” for intact genes, truncated genes and pseudogenes, respectively. Dark grey bars represent the fraction (%) of pseudogenes.  $\gamma$  (gamma) and  $\beta$  (beta) group positions are inverted for didactic purposes.

**Fig. 2.** Protein sequence structure representation of an OR from each of the species here analysed. Rectangular blocks represent conserved motifs according to MEME motif discovery. Each rectangular block represents the best non-overlapping tiling of motif matches on the sequence. The position of the block shows where a motif has matched the sequence. The width and height of a block represent the width of the motive relative to the length of the sequence and an indication of the significance of the match as taller blocks are more significant, respectively. Horizontal lines in grey below the sequence represent the predicted locations of a transmembrane domains (TM1–7) and intracellular loops (IC1–3) and extracellular loops (EC1–3) are shown as well. Light orange circles above a particular block represent the position of the five most representative motifs in terms of e-value.

**Fig. 3.** Radial tree of the Maximum Likelihood phylogeny from RAxML of the OR gene groups in the three species. Node colours are:  $\alpha$  (alpha): green;  $\beta$  (beta): pink;  $\gamma$  (gamma): blue;  $\delta$  (delta): orange;  $\epsilon$  (epsilon): purple;  $\zeta$  (zeta): red;  $\eta$  (eta): light blue; *X. tropicalis* theta: light green; *X. tropicalis* kappa: yellow; outgroup: black.

**Fig. 4.** Maximum Likelihood phylogeny from RAxML of the OR gene groups. (a) Detail of OR groups alpha, beta, delta and epsilon phylogeny tree. (b) Detail of the OR group eta with kappa and theta phylogeny tree. Polygons: blue circle: *Engystomops pustulosus*; red triangle: *Bufo bufo*; green square: *Hymenochirus boettgeri*. Bootstrap values of the basal nodes from the main OR groups are shown, including any phylogenetic clade with only one species over 70% bootstrap are highlight with the shade of the species. Node colours are:  $\alpha$  (alpha): green;  $\beta$  (beta): pink;  $\gamma$  (gamma): blue;  $\delta$  (delta): orange;  $\epsilon$  (epsilon): purple;  $\zeta$  (zeta): red;  $\eta$  (eta): light blue; *X. tropicalis* theta: light green; *X. tropicalis* kappa: yellow; outgroup: black.

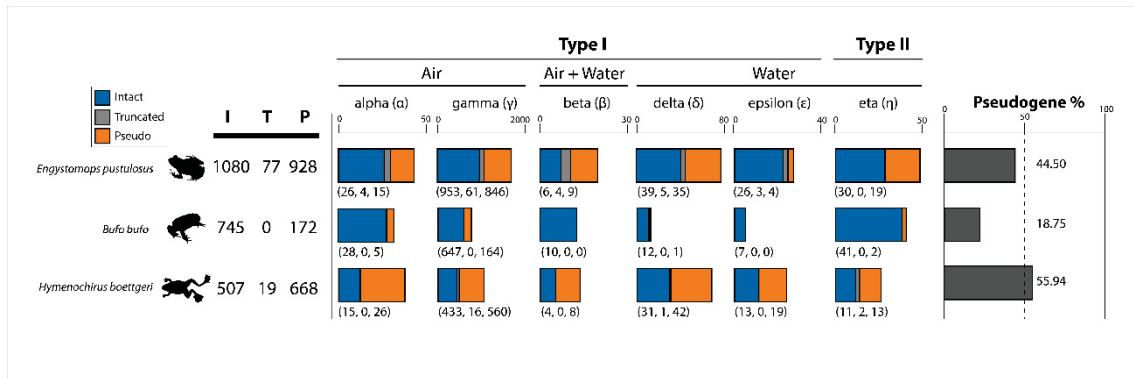
**Fig. 5.** OR gene distribution throughout chromosomes for each species. Coloured squares represent each of the OR groups, with F for intact (functional) genes, P for pseudogenes and T for truncated genes. (a) *Engystomops pustulosus*; (b) *Bufo bufo*; (c) *Hymenochirus boettgeri*. Chromosome colours are:  $\alpha$  (alpha): green;  $\beta$  (beta): pink;  $\gamma$  (gamma): blue;  $\delta$  (delta): orange;  $\epsilon$  (epsilon): purple;  $\zeta$  (zeta): red;  $\eta$  (eta): light blue; *X. tropicalis* theta: light green; *X. tropicalis* kappa: yellow

Table 1. Genome assembly information for *Engystomops pustulosus*, *Bufo bufo* and *Hymenochirus boettgeri*.

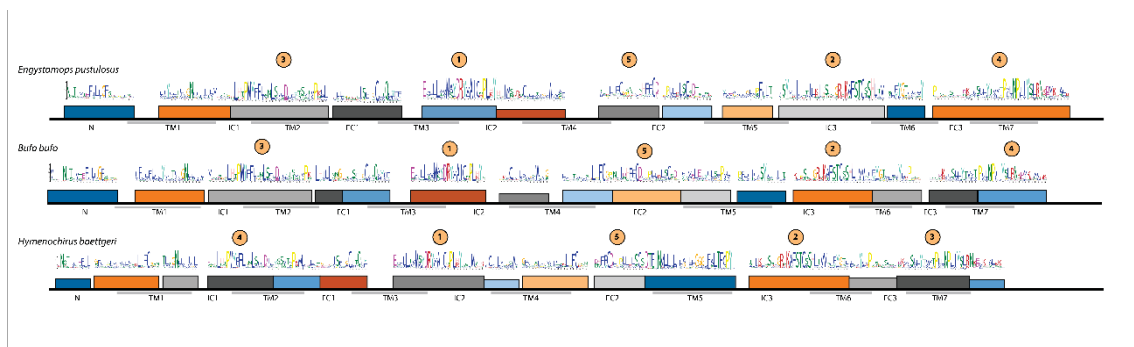
Species	Assembly version	Assembly level	Genome Coverage	Contig N50
<i>Engystomops pustulosus</i>	GCA_019512145.1	Chromosome	103x	305.357
<i>Bufo bufo</i>	GCA_905171765.1	Chromosome	64x	3.959.750
<i>Hymenochirus boettgeri</i>	GCA_019447015.1	Chromosome	110x	783.846

Table 2. Functional OR sequences from species used as queries in TBLASTN search against the genome of each species.

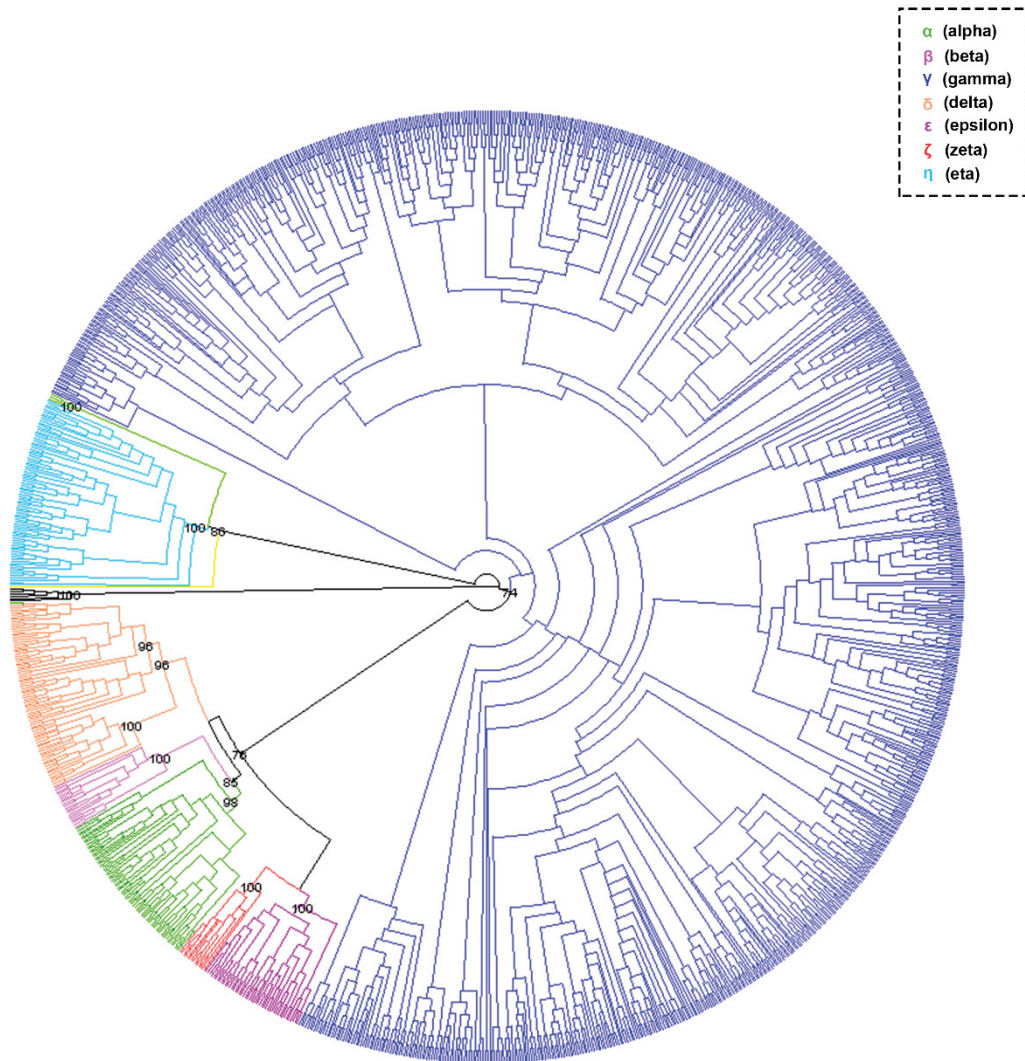
Common name	Scientific name	Number of ORs	CD-Hit candidates
Sea lamprey	<i>Petromyzon marinus</i>	32	14
Elephant shark	<i>Callorhynchus milli</i>	1	1
Zebra fish	<i>Danio rerio</i>	154	44
Medaka	<i>Oryzias latipes</i>	68	21
Stickeback	<i>Gasterosteus aculeatus</i>	102	37
Fugu	<i>Takifugu rubripes</i>	47	19
Spotted green puffer fish	<i>Tetraodon nigroviridis</i>	11	9
Western clawed frog	<i>Xenopus tropicalis</i>	824	110
Lizard	<i>Anolis carolinensis</i>	112	35
Chicken	<i>Gallus gallus</i>	211	37
Human	<i>Homo sapiens</i>	391	92
TOTAL		1953	419



**Figure 1.** Numbers of OR genes identified for each of the three species. Total number of genes are represented by “I”, “T”, and “P” for intact genes, truncated genes and pseudogenes, respectively. Dark gray bars represent the fraction (%) of pseudogenes.  $\gamma$  (gamma) and  $\beta$  (beta) group positions are inverted for didactic purposes.



**Figure 2.** Protein sequence structure representation of an OR from each of the species here analysed. Rectangular blocks represent conserved motifs according to MEME motif discovery. Each rectangular block represents the best non-overlapping tiling of motif matches on the sequence. The position of the block shows where a motif has matched the sequence. The width and height of a block represent the width of the motive relative to the length of the sequence and an indication of the significance of the match as taller blocks are more significant, respectively. Horizontal lines in gray below the sequence represent the predicted locations of a transmembrane domains (TM1–7) and intracellular loops (IC1–3) and extracellular loops (EC1–3) are shown as well. Light orange circles above a particular block represent the position of the five most representative motifs in terms of e-value.



**Figure 3. Radial tree of the Maximum Likelihood phylogeny from RAxML of the OR gene groups in the three species. Node colours are:  $\alpha$  (alpha): green;  $\beta$  (beta): pink;  $\gamma$  (gamma): blue;  $\delta$  (delta): orange;  $\epsilon$  (epsilon): purple;  $\zeta$  (zeta): red;  $\eta$  (eta): light blue; *X. tropicalis* theta: light green; *X. tropicalis* kappa: yellow; outgroup: black.**

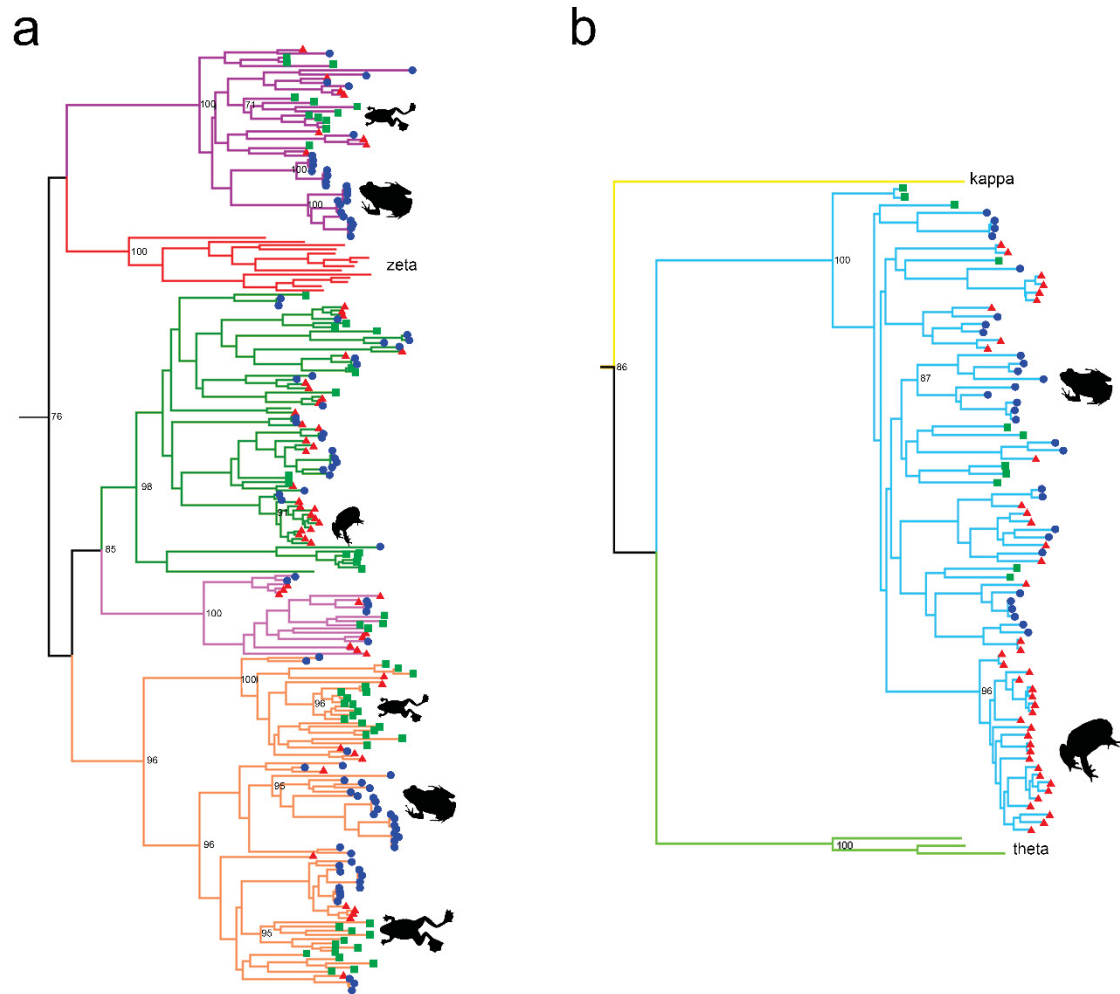
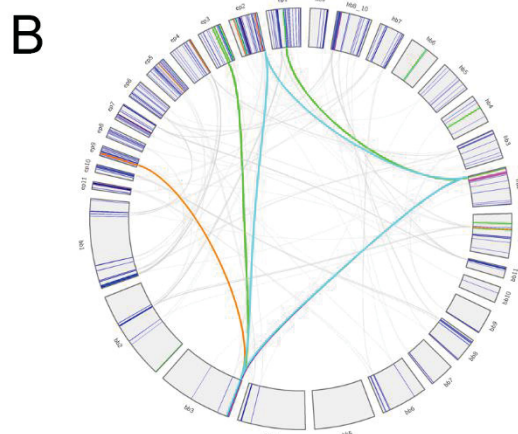
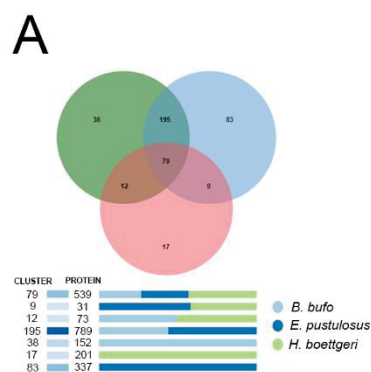


Figure 4. (a) Detail of OR groups alpha, beta, delta and epsilon phylogeny tree. (b) Detail of the OR group eta with kappa and theta phylogeny tree. Polygons: blue circle: *Engystomops pustulosus*; red triangle: *Bufo bufo*; green square: *Hymenochirus boettgeri*. Bootstrap values of the basal nodes from the main OR groups are shown, including any phylogenetic clade with only one species over 70% bootstrap are highlight with the shade of the species. Node colours are:  $\alpha$  (alpha): green;  $\beta$  (beta): pink;  $\gamma$  (gamma): blue;  $\delta$  (delta): orange;  $\epsilon$  (epsilon): purple;  $\zeta$  (zeta): red;  $\eta$  (eta): light blue; *X. tropicalis* theta: light green; *X. tropicalis* kappa: yellow; outgroup: black.



**Figure 5. OR gene distribution throughout chromosomes for each species. Coloured squares represent each of the OR groups, with F for intact (functional) genes, P for pseudogenes and T for truncated genes. (a) *Engystomops pustulosus*; (b) *Bufo bufo*; (c) *Hymenochirus boettgeri*.**



**Figure 6. (a) A Venn diagram of ortholog gene clusters between the three species. (b) all chromosomes by species and the corresponding ortholog gene pairs linked for all OR groups. Link and OR group colours are:  $\alpha$  (alpha): green;  $\beta$  (beta): pink;  $\gamma$  (gamma): grey (links) and blue (chromosomes);  $\delta$  (delta): orange;  $\eta$ (eta): light blue. Chromosomes are represented by lowercase letters of initials from species name followed by the respective chromosome number.**

**Supplementary Information**

Table S1 The repertoire of olfactory receptors (ORs) genes for three anuran genomes

Species	Assembly_version	Functional	Pseudogenes	Truncated	Total	%Functional	%Pseudogenes
<i>Engystomops pustulosus</i>	GCA_019512145.1	1080	919	77	2076	0,52023	0,44268
<i>Bufo bufo</i>	GCA_905171765.1	745	172	0	917	0,81243	0,18757
<i>Hymenochirus boettgeri</i>	GCA_019447015.1	507	668	19	1194	0,42462	0,55946

Table S2 Number of OR genes located in the chromosomes and in unplaced contigs for three species in study

<i>Engystomops pustulosus</i> OR group	unplaced OR genes		placed OR genes		TOTAL		
	Intact	Pseudogene	Truncated	Intact		Pseudogene	Truncated
alpha	22	13	4	4	2	0	45
beta	6	9	4	4	0	0	19
gamma	510	454	60	60	392	1	1874
delta	34	29	5	5	6	0	79
epsilon	21	4	3	3	0	0	33
eta	4	1	0	0	18	0	49
<b>TOTAL</b>	<b>597</b>	<b>510</b>	<b>76</b>	<b>76</b>	<b>418</b>	<b>1</b>	<b>2099</b>
<i>Bufo bufo</i> OR group	unplaced OR genes		placed OR genes		TOTAL		
alpha	0	0	0	0	5	0	33
beta	0	0	0	0	0	0	10
gamma	10	2	0	0	162	0	811
delta	0	0	0	0	1	0	13
epsilon	0	0	0	0	0	0	7
eta	0	0	0	0	2	0	43
<b>TOTAL</b>	<b>10</b>	<b>2</b>	<b>0</b>	<b>0</b>	<b>170</b>	<b>0</b>	<b>917</b>
<i>Hymenochirus boettgeri</i> OR group	unplaced OR genes		placed OR genes		TOTAL		
alpha	2	11	0	0	15	0	41
beta	0	1	0	0	7	0	12
gamma	98	158	16	16	402	0	1009
delta	30	38	1	1	4	0	74
epsilon	12	19	0	0	0	0	32

eta	0	0	0	0	11	13	2	26
TOTAL	142	227	17	365	441	2	1194	

Table S3 OR genomic clusters in the genomes of the species in study. OR genomic clusters that contain at least five functional genes or pseudogenes are listed.

Cluster name	Chromosome	Intact	Pseudogene	Truncated	F / (F + P + T)	Start	End	Length, kb
<i>Engystomops pustulosus</i>								
1.1	CM033641.1	3	2	0	0,60	149204679	149467913	263,23
1.2	CM033641.1	6	24	0	0,20	215187162	216412922	1225,76
1.3	CM033641.1	12	5	0	0,71	216927177	217265656	338,47
1.4	CM033641.1	29	14	0	0,67	218091367	219700370	1609,00
1.5	CM033641.1	10	11	0	0,48	227579617	228628404	1048,78
2.1	CM033642.1	26	18	0	0,59	68101	864753	796,65
2.3	CM033642.1	5	4	0	0,56	35852121	35989217	137,09
3.1	CM033643.1	6	1	0	0,86	30618430	30619437	1
4.1	CM033644.1	3	5	1	0,33	580076	671108	91,03
4.2	CM033644.1	102	98	0	0,51	2965111	6107160	3142,04
4.3	CM033644.1	2	9	0	0,18	9613257	10263627	650,37
4.4	CM033644.1	13	8	0	0,62	49887816	50230776	342,96
4.5	CM033644.1	8	2	0	0,80	68393951	68945887	551,93
5.1	CM033645.1	3	5	0	0,38	81779083	81816529	37,44
6.1	CM033646.1	16	15	0	0,52	26346557	26882550	535,99
6.2	CM033646.1	9	5	0	0,64	119956816	120263462	306,64
6.3	CM033646.1	9	11	0	0,45	130171029	130585852	414,82
6.4	CM033646.1	5	3	0	0,63	132935646	133171005	235,35
7.1	CM033647.1	21	34	0	0,38	9745	1474170	1464,42
7.2	CM033647.1	29	21	0	0,58	7725299	8689299	964,00
7.3	CM033647.1	4	3	0	0,57	11334645	11510225	175,58
9.1	CM033649.1	3	2	0	0,60	67552652	67604725	52,07
9.2	CM033649.1	3	2	0	0,60	79511306	79552893	41,58

10.1	CM033650.1	16	12	0	0,57	3073841	3390586	316,74
10.2	CM033650.1	10	6	0	0,63	13394236	13633074	238,83
10.3	CM033650.1	4	1	0	0,80	18338058	18598501	260,44
11.1	CM036351.1	5	4	0	0,56	9288565	9503707	215,14
11.2	CM036351.1	4	3	0	0,57	26334787	26442735	107,94
11.3	CM036351.1	5	0	0	1,00	55060930	55460896	399,96
<i>Bufo bufo</i>								
1.1	LR991667.1	5	0	0	1,00	115338906	115737327	398,42
1.2	LR991667.1	6	1	0	0,86	119566182	120023609	457,42
1.3	LR991667.1	6	0	0	1,00	141811886	142401698	589,81
1.4	LR991667.1	7	0	0	1,00	625585428	625982969	397,54
1.5	LR991667.1	8	1	0	0,89	674076280	674511471	435,19
1.6	LR991667.1	5	0	0	1,00	768282252	768404802	122,55
1.7	LR991667.1	25	14	0	0,64	776653674	779542010	2888,33
1.8	LR991667.1	7	0	0	1,00	814203172	814450058	246,88
1.9	LR991667.1	186	102	0	0,65	817298847	835053210	17754,36
2.1	LR991668.1	20	1	0	0,95	271031369	272499982	1468,61
2.2	LR991668.1	30	6	0	0,83	292419603	295681113	3261,51
2.3	LR991668.1	9	1	0	0,90	298208344	298787363	567,01
3.1	LR991669.1	6	0	0	1,00	595158997	595378550	219,55
3.2	LR991669.1	41	2	0	0,95	686737753	689358672	2620,91
3.3	LR991669.1	13	2	0	0,87	690523970	691586478	1062,50
3.4	LR991669.1	7	0	0	1,00	692440374	692780864	340,49
3.5	LR991669.1	37	3	0	0,93	699871041	702212457	2341,41
4.1	LR991670.1	24	10	0	0,71	29815764	33175577	3359,81
4.2	LR991670.1	23	2	0	0,92	34142697	36697341	2554,64
6.1	LR991672.1	7	0	0	1,00	29000336	29248626	248,29

6.2	LR991672.1	14	2	0	0,88	64392650	65884543	1491,89
6.3	LR991672.1	8	0	0	1,00	324345490	324964018	618,52
7.1	LR991673.1	60	5	0	0,92	15707605	18157656	2450,05
8.1	LR991674.1	9	1	0	0,90	10983174	11176647	193,47
8.2	LR991674.1	19	2	0	0,90	196316811	197102037	785,22
8.3	LR991674.1	15	2	0	0,88	203639448	204174493	535,04
8.4	LR991674.1	8	0	0	1,00	210035119	210156854	121,73
9.1	LR991675.1	14	2	0	0,88	7118025	7630706	512,68
11.1	LR991677.1	35	3	0	0,92	84759402	86584998	1825,59
11.2	LR991677.1	29	1	0	0,97	90496983	91150321	653,33
<i>Hymenochirus boettgeri</i>								
1.1	CM033469.1	2	3	0	0,40	192571047	192683481	112,43
1.2	CM033469.1	23	6	0	0,79	270137483	270896573	759,09
2.1	CM033470.1	0	5	0	0,00	292700467	292920299	219,83
2.2	CM033470.1	6	5	0	0,55	336297790	336878283	580,49
2.3	CM033470.1	9	10	0	0,47	345104087	345654827	550,74
2.4	CM033470.1	11	2	0	0,85	348102987	348562962	459,97
3.1	CM033471.1	14	12	0	0,54	238252178	238852135	599,95
3.2	CM033471.1	8	10	0	0,44	244543110	245508120	965,01
3.3	CM033471.1	37	44	0	0,46	269886686	272031362	2144,67
7.1	CM033475.1	27	49	0	0,36	13337839	15126426	1788,58
7.2	CM033475.1	4	5	0	0,44	30524178	30750000	225,82
7.3	CM033475.1	2	3	0	0,40	187347243	187679048	331,8
8_10.1	CM033476.1	2	9	0	0,18	24325199	24719383	394,18
8_10.2	CM033476.1	4	4	0	0,50	105912703	106094087	181,38
8_10.3	CM033476.1	35	34	0	0,51	297721989	299165528	1443,53
8_10.4	CM033476.1	11	7	0	0,61	299837581	300587270	749,68

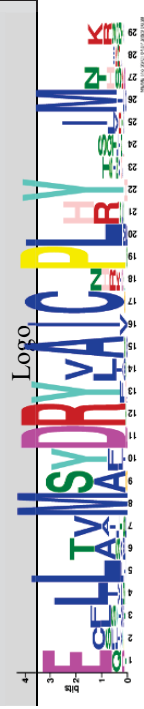
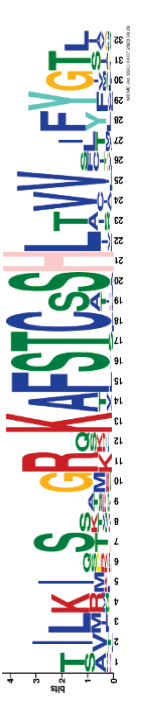
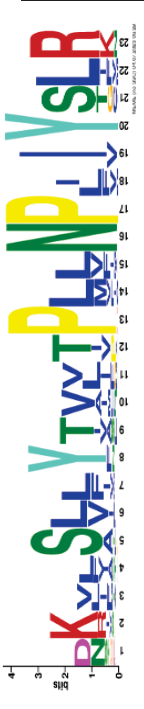
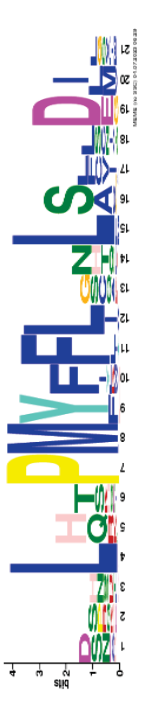

8_10.5	CM033476.1	0	17	0	0,00	302056900	302396226	339,32
8_10.6	CM033476.1	9	18	0	0,33	305174707	305596704	421,99
8_10.7	CM033476.1	14	4	0	0,78	315846518	316710282	863,76
9.1	CM033477.1	47	106	0	0,31	1128	4148410	4147,28
9.2	CM033477.1	33	25	0	0,57	5154697	6486654	1331,95
9.3	CM033477.1	8	5	0	0,62	9004244	9794905	790,66
9.4	CM033477.1	3	2	0	0,60	34611479	34819906	208,42

Table S4 Total number of clusters in the genomes of the species in study, including singletons.

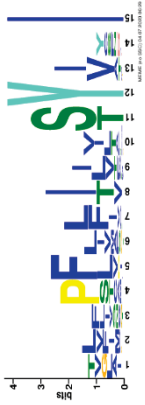
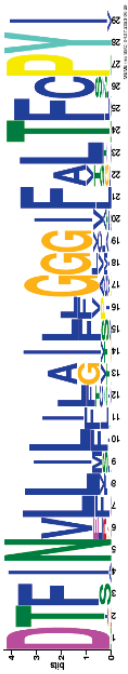

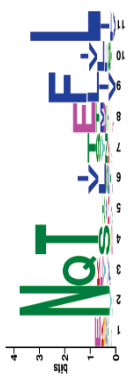

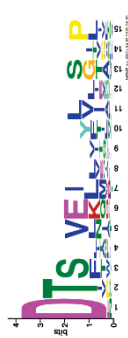

Chromosome_id	Chromosome number	Number of ORs	Singletons
<i>Engystomops pustulosus</i>			
CM033641.1	1	25	13
CM033642.1	2	29	20
CM033643.1	3	12	7
CM033644.1	4	12	6
CM033645.1	5	24	13
CM033646.1	6	15	9
CM033647.1	7	13	5
CM033648.1	8	11	8
CM033649.1	9	13	7
CM033650.1	10	11	6
CM033651.1	11	5	2
<i>Bufo bufo</i>			
LR991667.1	1	14	2
LR991668.1	2	9	3
LR991669.1	3	7	1
LR991670.1	4	5	2
LR991671.1	5	-	-
LR991672.1	6	5	2
LR991673.1	7	2	0
LR991674.1	8	9	1
LR991675.1	9	2	0
LR991676.1	10	1	0
LR991677.1	11	5	1
<i>Hymenochirus boettgeri</i>			
CM033469.1	1	12	4
CM033470.1	2	15	6
CM033471.1	3	8	3
CM033472.1	4	4	3
CM033473.1	5	8	4
CM033474.1	6	3	3
CM033475.1	7	9	2
CM033476.1	8_10	18	7
CM033477.1	9	5	0



Table S5 Sequence logos and information of the 20 most representative putative conserved motifs according to MEME motif discovery.

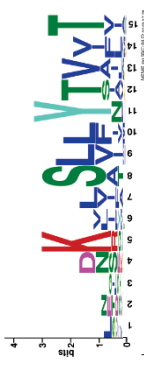
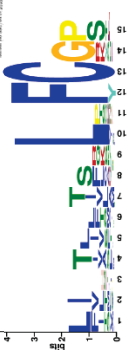
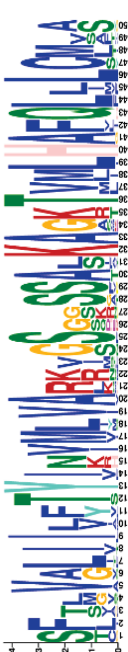
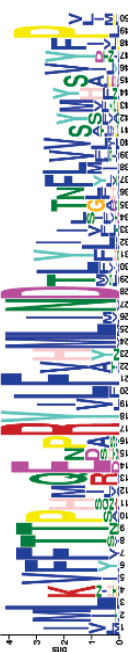
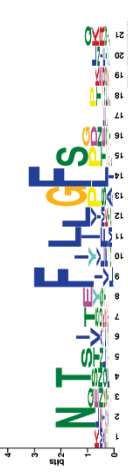


<i>Engystomops pustulosus</i>					
Motif number	Logo	Name	E-value	Sites	Width
1		ECFLLTVMSYDRYVAICNPLHYTSMNKK	3.8e-5892	507	29
2		TILKIQSSAGRQKAFSTCSSHLTVVSIFYGTL	2.2e-5052	435	32
3		BKVLSLLYTVVTPLLNPJYSLR	1.1e-3728	475	23
4		DSHLHTPMYFFLGNLSLLDIL	1.7e-3077	496	21
5		BHFFCDLPLLKLSCS	9.4e-1731	458	16

6	<p>Sequence logo for motif 6. The y-axis represents information content in bits (0 to 4). The x-axis represents positions 1 to 21. The sequence GFSB L P S L Q J P L F L L F L L I Y J is shown below the logo.</p>	9.5e-1463	417	21
7	<p>Sequence logo for motif 7. The y-axis represents information content in bits (0 to 4). The x-axis represents positions 1 to 15. The sequence GGKTISFQGCITQLY is shown below the logo.</p>	2.4e-1285	475	15
8	<p>Sequence logo for motif 8. The y-axis represents information content in bits (0 to 4). The x-axis represents positions 1 to 21. The sequence GFLYSLVHVLLLLSKLTFCGSN is shown below the logo.</p>	1.3e-996	410	21
9	<p>Sequence logo for motif 9. The y-axis represents information content in bits (0 to 4). The x-axis represents positions 1 to 11. The sequence NKEIKKALKKL is shown below the logo.</p>	1.0e-783	448	11
10	<p>Sequence logo for motif 10. The y-axis represents information content in bits (0 to 4). The x-axis represents positions 1 to 15. The sequence YSSVTVPKMLFDLLT is shown below the logo.</p>	3.7e-610	284	15
11	<p>Sequence logo for motif 11. The y-axis represents information content in bits (0 to 4). The x-axis represents positions 1 to 11. The sequence TLTGNLLIILL is shown below the logo.</p>	5.6e-528	354	11

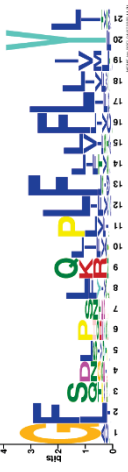
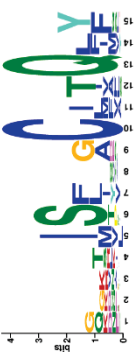
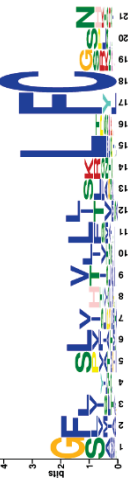
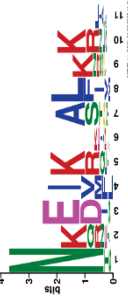
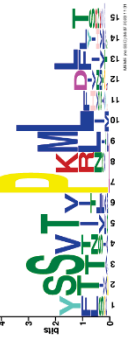
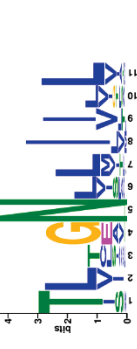
12	 <p>Sequence logo for motif 12. The x-axis represents positions 1-13. The y-axis represents bits (0 to 4). The most prominent peaks are at positions 2 (P), 3 (F), 4 (L), 11 (S), and 12 (V).</p>	TLFPFLIJVSYIYI	6.3e-625	333	15
13	 <p>Sequence logo for motif 13. The x-axis represents positions 1-28. The y-axis represents bits (0 to 4). The most prominent peaks are at positions 1 (D), 2 (E), 3 (N), 4 (M), 5 (L), 6 (L), 7 (I), 8 (A), 9 (I), 10 (L), 11 (G), 12 (G), 13 (G), 14 (F), 15 (A), 16 (L), 17 (T), 18 (F), 19 (C), 20 (P), 21 (Y), and 22 (I).</p>	DTFINVLLILAILLGGGIFALTFCPYI	1.3e-835	76	29
14	 <p>Sequence logo for motif 14. The x-axis represents positions 1-14. The y-axis represents bits (0 to 4). The most prominent peaks are at positions 3 (Y), 4 (P), 11 (S), 12 (S), and 13 (S).</p>	IFIYLPSSGYSLNV	6.1e-473	332	15
15	 <p>Sequence logo for motif 15. The x-axis represents positions 1-11. The y-axis represents bits (0 to 4). The most prominent peaks are at positions 1 (N), 2 (Q), 3 (T), 4 (T), 5 (I), 6 (T), 7 (E), 8 (F), 9 (I), and 10 (L).</p>	ENQTTITEFIL	8.9e-385	337	11
16	 <p>Sequence logo for motif 16. The x-axis represents positions 1-11. The y-axis represents bits (0 to 4). The most prominent peaks are at positions 1 (C), 2 (L), 3 (V), 4 (S), 5 (I), and 6 (L).</p>	VCVLLVLGSWL	2.7e-347	384	11
17	 <p>Sequence logo for motif 17. The x-axis represents positions 1-11. The y-axis represents bits (0 to 4). The most prominent peaks are at positions 1 (D), 2 (T), 3 (S), 4 (V), 5 (E), 6 (L), 7 (V), 8 (Y), 9 (L), 10 (J), and 11 (S).</p>	DTSLVEJLVYLSIP	2.3e-331	244	15
18	 <p>Sequence logo for motif 18. The x-axis represents positions 1-11. The y-axis represents bits (0 to 4). The most prominent peaks are at positions 1 (I), 2 (T), 3 (N), 4 (V), 5 (P), 6 (L), 7 (M), 8 (L), 9 (G), 10 (T), and 11 (T).</p>	LTTNVPILLQTJN	1.3e-288	134	15

19		KDARSKALYTCCTTHLJVISLY	3.2e-181	51	21
20		HPSVMTLGFGEMTSIKYLYCVLVLVGYVVLILLSNSTVIAVI	2.8e-123	22	41
<i>Bufo bufo</i>					
Motif number	Logo	Name	E-value	Sites	Width
1		ECLLLAVMAYDRYVAICNPLHYS	1.1e-4586	745	23
2		KIPSSTGRQKAFSTCSSHLTVVSI	3.3e-3856	685	24
3		VSTSPHLHTPMYFFLSNLSFLDLITYTSVIVPK	1.4e-4253	704	32
4		PLLNPIJYSLRNKDVKKALKK	3.3e-3469	741	21

5	<p>Sequence logo for motif 5. The y-axis represents information content in bits (0 to 4). The x-axis represents positions 1 to 16. The sequence is N P H F E D P L K T A S D T.</p>	NVIBHFFCDLPPLLKLSCSDT	693	8.6e-2946	21
6	<p>Sequence logo for motif 6. The y-axis represents information content in bits (0 to 4). The x-axis represents positions 1 to 12. The sequence is I S F G T Q M Y F F.</p>	KTISFIGCITQMYFF	694	2.4e-1256	15
7	<p>Sequence logo for motif 7. The y-axis represents information content in bits (0 to 4). The x-axis represents positions 1 to 19. The sequence is L F J L F L L Y L L T L C G N L L I L L.</p>	LFJLFLLYLLTLCGNLLIIL	619	1.0e-1270	21
8	<p>Sequence logo for motif 8. The y-axis represents information content in bits (0 to 4). The x-axis represents positions 1 to 21. The sequence is M E E K N Q T T V T E F I L L G F S N S P.</p>	MEEKNQTTVTEFILLGFSNSP	371	8.2e-964	21
9	<p>Sequence logo for motif 9. The y-axis represents information content in bits (0 to 4). The x-axis represents positions 1 to 13. The sequence is F Y G T I F V Y L R P T S G.</p>	FYGTIFVYLRPTSG	599	9.6e-956	15
10	<p>Sequence logo for motif 10. The y-axis represents information content in bits (0 to 4). The x-axis represents positions 1 to 13. The sequence is P F J L I S Y I Y I H I S T.</p>	PFJLISYIYIIST	548	6.9e-745	15
11	<p>Sequence logo for motif 11. The y-axis represents information content in bits (0 to 4). The x-axis represents positions 1 to 14. The sequence is K R L C I L L A I G S W L L G.</p>	KRLCILLAIGSWLLG	583	5.4e-635	15



12	 <p>Sequence logo for motif 12. The x-axis represents positions 1 to 15. The y-axis represents bits (0 to 4). The most conserved residues are K, S, Y, T, V, Y, T at positions 1-7.</p>	LNQDKVLSLLYTVVT	1.3e-644	513	15
13	 <p>Sequence logo for motif 13. The x-axis represents positions 1 to 15. The y-axis represents bits (0 to 4). The most conserved residues are T, S, L, F, G, S at positions 1-6.</p>	LIHTLLTSRLPFCGP	9.8e-509	420	15
14	 <p>Sequence logo for motif 14. The x-axis represents positions 1 to 50. The y-axis represents bits (0 to 4). The motif is highly conserved across all positions.</p>	SFTLVALILFTYINVM LVARKVGGSSALKA GKTVM LHFQLLCMVA	1.8e-452	41	50
15	 <p>Sequence logo for motif 15. The x-axis represents positions 1 to 50. The y-axis represents bits (0 to 4). The motif is highly conserved across all positions.</p>	IMLK VFFTTPHMQEBPR YILFVHMLINDTLY JILTNFLVVS YMSVYFPV	9.3e-367	40	50
16	 <p>Sequence logo for motif 16. The x-axis represents positions 1 to 21. The y-axis represents bits (0 to 4). The most conserved residues are N, T, S, Y, I, F, H, G, S at positions 1-9.</p>	KNQTSITEFIJLGFSGLPKLQ	6.3e-275	226	21
17	 <p>Sequence logo for motif 17. The x-axis represents positions 1 to 8. The y-axis represents bits (0 to 4). The most conserved residues are L, H, I, L, L, N, N, G at positions 1-8.</p>	LHILLNNG	1.1e-112	156	8
18	 <p>Sequence logo for motif 18. The x-axis represents positions 1 to 15. The y-axis represents bits (0 to 4). The most conserved residues are F, T, V, Z, L, E, I, Y, J, L, S, F, P, V, L at positions 1-15.</p>	FTVZLEIYJLSFPVL	8.4e-110	125	15

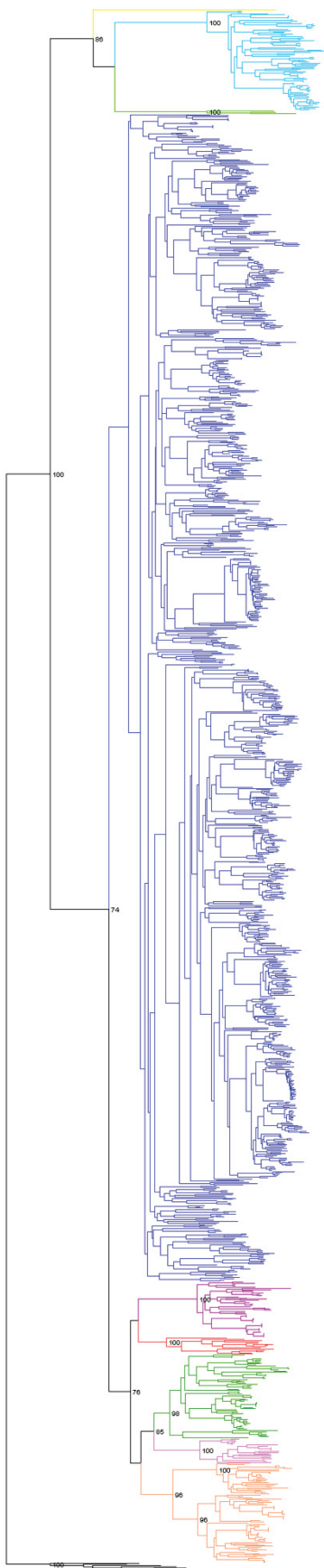
19		WLNEIAMYIAAGLIY	2.9e-090	37	15
20		FTYKVDVVTRTVFLITSLLCFTFFFYFIA	4.0e-089	20	29
<i>Hymenochirus boettgeri</i>					
Motif number	Logo	Name	E-value	Sites	Width
1		ECFLLTVMSYDRYVAICNPLHYTSIMNKK	3.8e-5892	507	29
2		TILKIQSSAGRQKAFSTCSSHLTVVSIFYGTL	2.2e-5052	435	32
3		BKVLSLLYTVVTPLLNPJYSLR	1.1e-3728	475	23
4		DSHLHTPMYFFLGNLSLLDIL	1.7e-3077	496	21

5		BHFFCDLPLLLKLSCS	9.4e-1731	458	16
6		GFSBLSLQJPLFLFLLIYJ	9.5e-1463	417	21
7		GGKTISFQGCITQLY	2.4e-1285	475	15
8		GFLYSLVHVLLLSKLTFCGSN	1.3e-996	410	21
9		NKEIKKALKKL	1.0e-783	448	11
10		YSSVTVPKMLFDLLT	3.7e-610	284	15
11		TLTGNLLIILL	5.6e-528	354	11

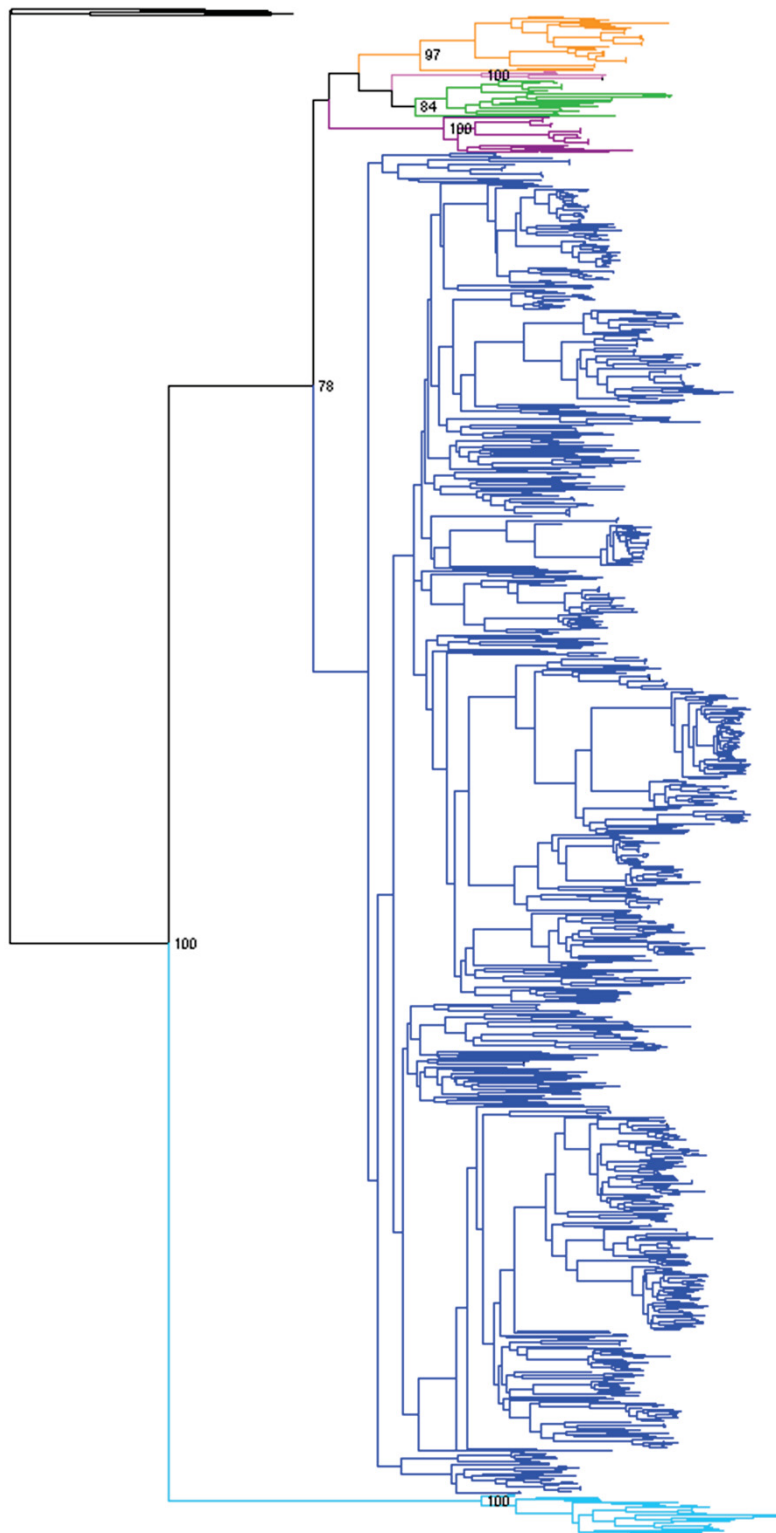


12		TLFPFLIJVSYIYI	6.3e-625	333	15
13		DTFINVLLILAILLGGGIFALTFCPYI	1.3e-835	76	29
14		IFIYLPSSGYSLNV	6.1e-473	332	15
15		ENQTTITEFIL	8.9e-385	337	11
16		VCVLLVLGSWL	2.7e-347	384	11
17		DTSLVEJLVYLSIP	2.3e-331	244	15
18		LTTNVPILLQTJIN	1.3e-288	134	15

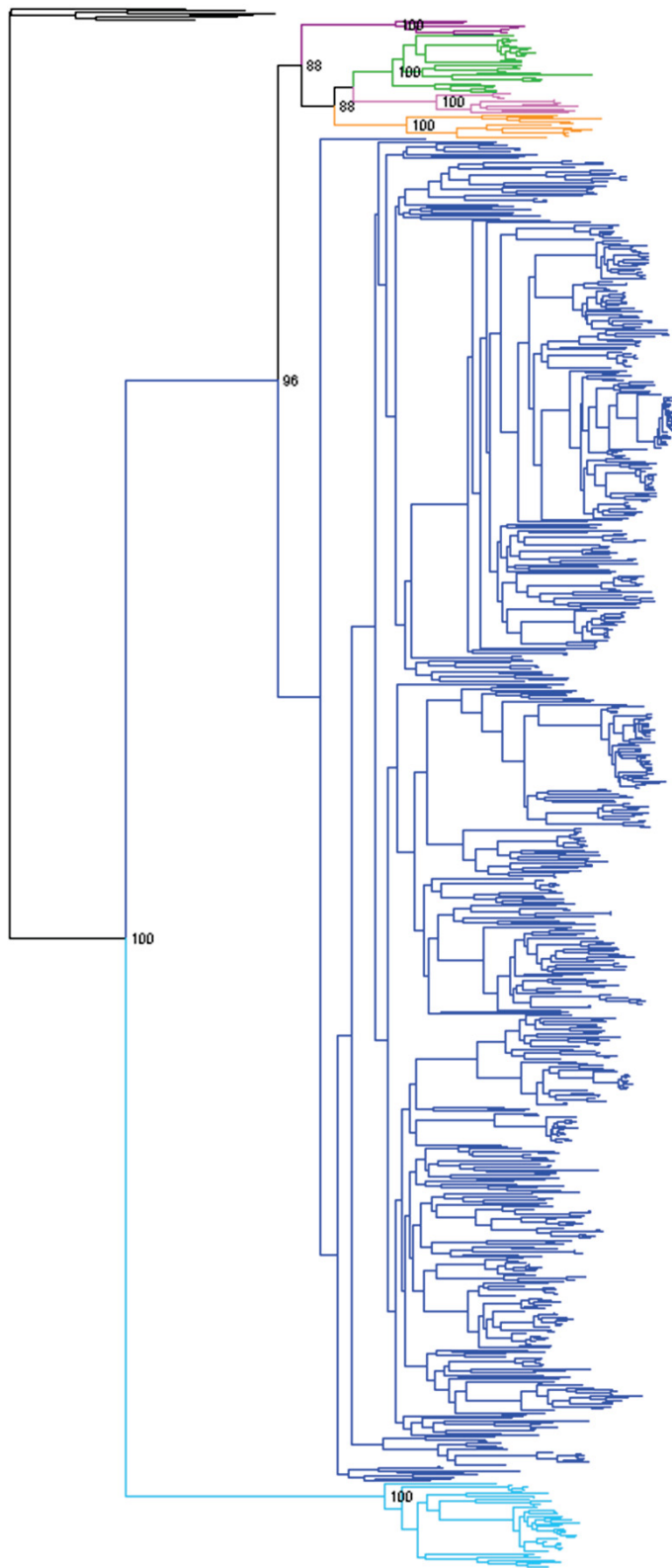
<p>19</p> 	<p>KDARSKALYTCCTTHLJVISLY</p>	<p>3.2e-181</p>	<p>51</p>	<p>21</p>
<p>20</p> 	<p>HPSVMTLGFGEMTSIKYLYCVLVLVGYVLIILSNSTVIAVI</p>	<p>2.8e-123</p>	<p>22</p>	<p>41</p>



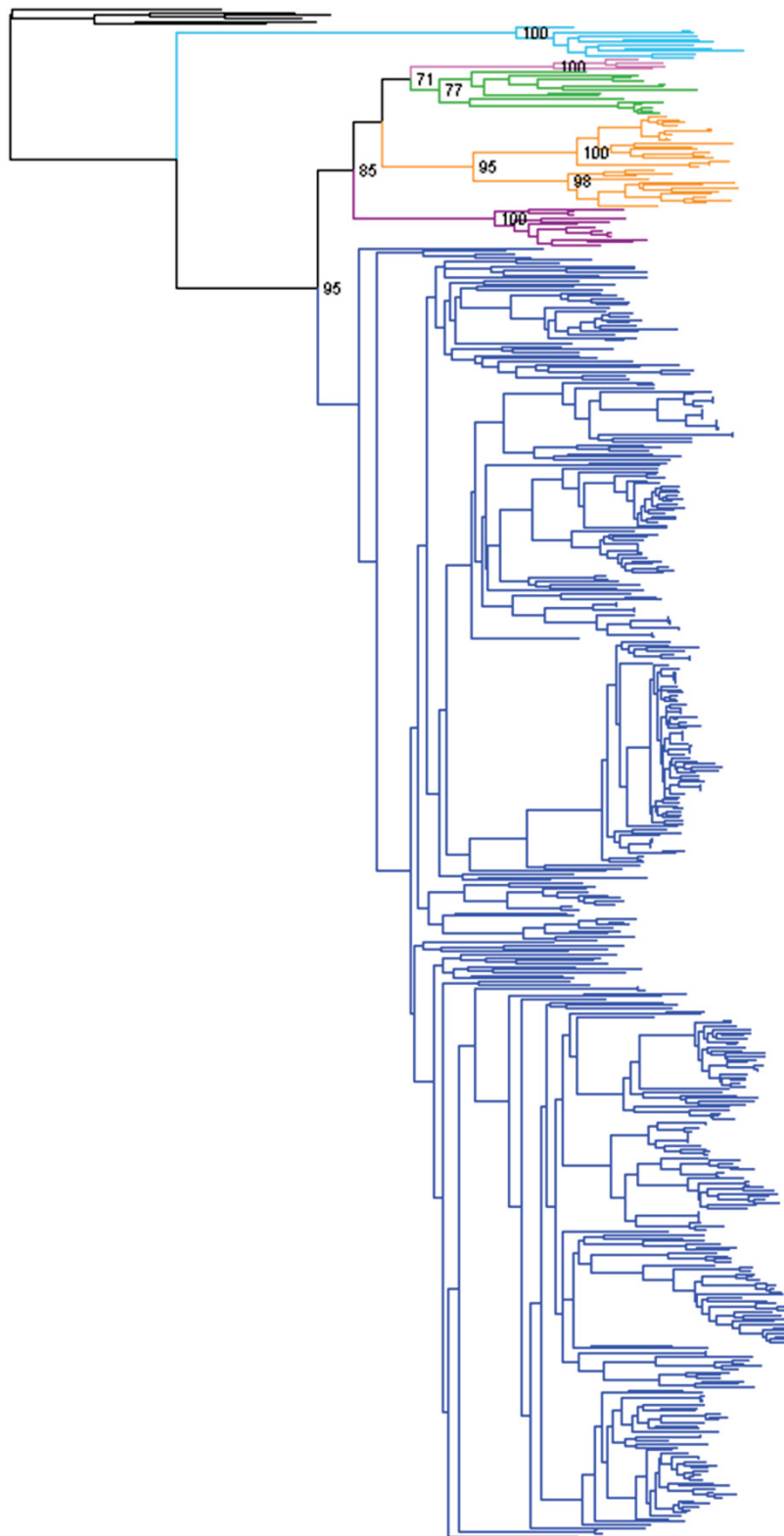
**Figure 1S. Phylogenetic tree of the Maximum Likelihood phylogeny from RAxML of the OR gene groups in the three species. Node colours are:  $\alpha$  (alpha): green;  $\beta$  (beta): pink;  $\gamma$  (gamma): blue;  $\delta$  (delta): orange;  $\epsilon$  (epsilon): purple;  $\zeta$  (zeta): red;  $\eta$  (eta): light blue; *X. tropicalis* theta: light green; *X. tropicalis* kappa: yellow; outgroup: black.**



**Figure 2S. Phylogenetic tree of the Maximum Likelihood phylogeny from RAxML of *Engystomops pustulosus* intact OR genes. Node colours are:  $\alpha$  (alpha): green;  $\beta$  (beta): pink;  $\gamma$  (gamma): blue;  $\delta$  (delta): orange;  $\epsilon$  (epsilon): purple;  $\zeta$  (zeta): red;  $\eta$  (eta): light blue; outgroup: black.**



**Figure 3S. Phylogenetic tree of the Maximum Likelihood phylogeny from RAxML of *Bufo bufo* intact OR genes. Node colours are:  $\alpha$  (alpha): green;  $\beta$  (beta): pink;  $\gamma$  (gamma): blue;  $\delta$  (delta): orange;  $\epsilon$  (epsilon): purple;  $\zeta$  (zeta): red;  $\eta$  (eta): light blue; outgroup: black.**





4.2 **Capítulo II** – “Intrachromosomal rearrangement and gene fractionation modulated the repertoire of olfactory receptors genes in post-polyploid species of *Xenopus* (Anura; Pipidae)”

## **Intrachromosomal rearrangements and gene fractionation have modulated the repertoire of olfactory receptor genes in the post-allopolyploid species of the African frog genus *Xenopus* (Anura; Pipidae)**

Johnny Sousa Ferreira<sup>1</sup>, Yoshihito Niimura<sup>2</sup>; Daniel Pacheco Bruschi<sup>1,\*</sup>

<sup>1</sup> Laboratory of Evolutionary Cytogenetics and Animal Conservation (LabCECA).

Department of Genetics. Federal University of Paraná (UFPR), Curitiba, Paraná, Brazil.

<sup>2</sup> Department of Veterinary Science. University of Miyazaki, Miyazaki, Japan.

\* E-mail address: danielbruschi@ufpr.br (D.P. Bruschi)

### **Abstract**

Clawed frogs represent an extraordinary model of biology evolution. Species of this genus have similar morphological traits adapted to an aquatic lifestyle and different evolutionary pathways that shaped their chromosomal structure. To clarify the olfactory receptor gene structure of the genus, we reevaluated the complete OR repertoire of *Xenopus tropicalis* and *Xenopus laevis* and described for the first time the repertoire of the allotetraploid species, *Xenopus borealis*. We identified 534 OGGs that represent the complete set of OR putatively presented in the MRCA of these species and quantify the gain/loss across tree branches and verified intense gene expansion of an exclusively OGG-5 in *X. tropicalis*. Also, we verified 144 OGGs that are composed of one gene per species that reflects a pattern of vertical transmission. Regarding the chromosomal distribution of OR clusters, we observed a HOGC composed of  $\eta$ ,  $\epsilon$ ,  $\delta$ ,  $\alpha$ , and  $\beta$  groups that is maintained in chromosome 2 from *X. tropicalis* and both 2S from *X. laevis* and *X. borealis*, but presents a 715Kb inversion of  $\epsilon$  and  $\delta$  and a total loss of  $\eta$  cluster genes, while *X. laevis* presents a complete inversion of 4.8Mb of this HOGC. Finally, although we observed ~25% of all OR genes in the S subgenome in both *X. laevis* and *X. borealis* species, chromosome 3L has ~27% of the total genes in these species. Our findings corroborate with birth-and-death evolutionary processes modulating the organization of OR genes in the post-allopolyploid *Xenopus* species, particularly in the intrachromosomal rearrangements and gene fractionation that occurred during the diploidization process.

### **Key-words**

Polyploid species, diploidization, anura, sensory genes.

### **Introduction**

Olfactory Receptors (ORs) are G protein-coupled receptors expressed by sensory neurons located in the olfactory epithelium of the nasal cavity, which bind to odorants for environmental perception. This complex sensorial system is broadly

adaptive and is subject to a range of challenges from processes such as predation, reproduction, and foraging patterns (Niimura et al., 2018; Yohe et al., 2021). Given this, the genes that encode olfactory receptors have undergone relatively high rates of evolution (Yohe et al., 2020). A large-scale study of the major vertebrate lineages has shown that most OR genes evolve at rates an order of magnitude higher than non-OR genes, irrespective of the taxonomic group (Yohe et al., 2020; Yohe et al., 2022). This makes the OR genes the largest multigene family in the vertebrate genome. The primary mechanism of diversification appears to be repeated, unequal crossovers, which have generated duplications, followed by an accumulation of mutations that diversify through the production of paralogous copies. This generates clusters of paralogous copies of the OR genes, which contribute to the expansion of this multigene family (Niimura, 2012). Paralogous copies provide the substrate for evolutionary processes, following the birth-and-death model (Nei and Rooney, 2005), in which duplicate genes accumulate novel mutations that may lead to (i) a loss of function (pseudogenization) or (ii) the acquisition of new functions (neofunctionalization). In the latter case, the substitution of codons and positive selection may contribute to the diversification of the potential of the system to bind to new odorant ligands (Niimura, 2012; Yohe et al., 2020), leading to an increase in the spectrum of genes in the OR repertoire of the species. A substantial set of empirical evidence supports this interpretation of the process (Dong et al., 2009; Hughes et al., 2018; Wang et al., 2023). The birth-death process may also lead to fluctuations (expansion/contraction) in the composition and size of the OR gene repertoire between phylogenetically related species, reflecting a typical species-specific pattern (Nei and Rooney, 2005).

The functional OR genes can be divided into seven phylogenetic groups – alpha ( $\alpha$ ), beta ( $\beta$ ), gamma ( $\gamma$ ), delta ( $\delta$ ), epsilon ( $\epsilon$ ), zeta ( $\zeta$ ) (collectively, Type I) and eta ( $\eta$ ) (Type II) – based on the gene-tree approach (Niimura, 2009a). The olfactory receptors encoded by the  $\alpha$  and  $\gamma$  groups identify airborne odorants, while the others ( $\delta$  through  $\eta$ ) detect water-soluble odorants, and the OR genes of the  $\beta$  group can detect both airborne and water-soluble odorants. Amphibians are unique vertebrates, with genes capable of identifying both airborne (type I) and water-soluble odorants (type II) (Niimura 2009a, 2009b, Lu et al., 2021, Han et al., 2022).

The clawed frogs of the genus *Xenopus* include two subgenera, *Xenopus* (25 species) and *Silurana* (four species), which represent an extraordinary model of biological evolution. The frogs of this genus have an aquatic lifestyle, and are native to sub-Saharan Africa (Frost, 2023). Although most *Xenopus* species have similar morphological traits, some have followed distinct evolutionary pathways. The subgenera can be differentiated by their ancestral ploidy. The subgenus *Silurana* has retained the predicted ancestral diploid number of 20 chromosomes (*Silurana tropicalis*), with other, polyploid species being derived from this baseline (Evans, 2008). By contrast, the species of the subgenus *Xenopus* are polyploids derived from an ancestor with 18 chromosomes, which originated from the fusion of chromosome pairs 9 and 10, following its divergence from *Silurana* (Session et al., 2016). This means that the subgenus *Xenopus* includes species that originated from events of allopolyploidization, which probably occurred through successive backcrossing (for a review, see Evans et al., 2015), leading to what currently may be recognized as two or more subgenomes (Session et al., 2016). For example, *Xenopus laevis* and *X. borealis* are allopolyploid species derived from a  $2n=18$  ancestor whose whole genome was duplicated about 17–18 Ma, resulting in a tetraploid genome of 36 chromosomes (Session et al., 2016). Despite their origin, these subgenomes are disomic, due to their rare or complete lack of recombination (Session et al., 2016). Related to this diploidization process (Wolfe, 2001), the absence of any recombination of the subgenomes, the differential mobilization of the transposons, and intra-chromosomal rearrangements resulted in the homologous chromosomes of each subgenome being differentiated cytogenetically (Matsuda et al., 2015; Session et al., 2016). In this case, the subgenomes of *X. laevis* and *X. borealis* are referred to as the “L” (Large) and “S” (Small) subgenomes, being differentiated by contrasts in the level of gene expression, alternative splicing patterns, epigenetic features, and their repetitive DNA content (Liu et al., 2014; Session et al., 2016; Elurbe et al., 2017; Mei et al., 2017; Cheng et al., 2018; Furman et al., 2018; Edger et al., 2019; Zhao et al., 2021; Jia et al., 2021; Schiavinato et al., 2021).

The different theoretical approaches all agree that the evolution of allopolyploid species toward disomic inheritance results in a differential intergenomic evolutionary pathway which frequently involves the loss of genes, different levels of purification of the coding genes between the subgenome copies, and intra-chromosomal

rearrangements (Wolfe, 2001). In this context, it seems reasonable to postulate that this complex intragenomic scenario may modulate the evolution of the OR gene repertoire of the allopolyploid *Xenopus* species. *A priori*, it also seems reasonable to expect that the rates of pseudogenization will be biased between the two subgenomes due to (i) differential levels of purifying selection acting on each orthologous copy (Furman et al., 2018) and (ii) differential rates of gain or loss in syntenic loci. To confirm these expectations, a framework of phylogenetic and map-based approaches was designed to compare the evolutionary dynamics of the OR gene repertoire of *X. laevis* and *X. borealis* with that of the diploid species *X. tropicalis* (= *Silurana tropicalis*). These three species were selected for this study because they are closely-related, in phylogenetic terms, which reduces potential adaptive biases, given that all three species have evolved under similar selective pressures for an aquatic lifestyle through the occupation of broadly similar ecological niches. The OR gene repertoires of the studied *Xenopus* species are discussed from the perspective of their evolutionary history, which provides valuable insights into the genetic mechanisms that have modulated the evolution of these genes in the species that had an allopolyploid origin.

## Materials and Methods

### *Identification of the OR Genes from Genome Sequences*

The latest genome assemblies, for *Xenopus tropicalis* (GCA\_000004195.4), *Xenopus laevis* (GCA\_017654675.1), and *Xenopus borealis* (GCA\_024363595.1), were obtained from GenBank (<https://www.ncbi.nlm.nih.gov/genbank/>). The OR genes were identified from the whole-genome sequences using the method described by Niimura, (2013a, 2013b), with some slight modifications. The TBLASTN searches (Altschul et al., 1990) were conducted using an *E* value of  $1 \times 10^{-5}$  derived from the intact OR genes obtained from 11 vertebrate species by Niimura (2009a) and Olender et al. (2013). This dataset was divided into phylogenetic groups based on Niimura (2009a), with the best candidates for query sequences (Table 2) being identified in CD-HIT (Huang et al. 2010) using a 60% homology cut-off point, to reduce computation time.

### *Gene tree reconstruction and description of the OR repertoire*

The Neighbour-Joining (NJ) trees (Saitou and Nei 1987) were constructed with the Poisson correction (PC) distance in the LINTREE program (Takezaki et al., 1995: <http://www.personal.psu.edu/nxm2/software.htm>). Multiple alignments of the translated amino acid sequences were generated in MAFFT v.7.520 (Katoh et al., 2019: <http://mafft.cbrc.jp/alignment/software/>). Each OR gene identified in the analyses was assigned to one of the phylogenetic OR groups of type I ( $\alpha$ ,  $\beta$ ,  $\gamma$ ,  $\delta$ ,  $\epsilon$ , and  $\zeta$ ) or type II, which has a single group ( $\eta$ ) according to Niimura (2009a).

#### *Identification of the Orthologous Gene Groups (OGGs) in the Xenopus species*

The intact OR genes identified from the genomes of the three species were assigned to OGGs following a phylogenetic approach (Niimura et al., 2014). For this, an OGG was defined as a group of genes that originated from a single gene in the Most Recent Common Ancestor (MRCA) of the studied species, with the total number of genes in an OGG varying in accordance with the number of duplications occurring between the MRCA and the target species. In the present case, the MRCA of all the *Xenopus* species and previously identified OR sequences from *Pipa parva* (unpublished data) were considered to be the outgroup. The ideal OGG was considered to be a group that contains one gene in each of the *Xenopus* species, with *X. tropicalis* as the most basal branch (Session et al., 2016). Given this, the NJ trees were constructed using all the intact genes allocated to each OR group ( $\alpha$ ,  $\beta$ ,  $\gamma$ ,  $\delta$ ,  $\epsilon$ ,  $\zeta$ ,  $\eta$ ), with one gene tree being reconstructed for each major clade identified. Candidate OGGs were extracted from each gene tree considering the monophyletic groups that contain at least one gene and are supported with a bootstrap value of at least 70% (Niimura et al., 2014).

#### *Estimation of the number of gains/losses in the OGGs*

The number of genes gained or lost on each branch of the phylogenetic tree with only intact OR genes was estimated by examining the presence/absence of a gene from each OGG per OR group in a given species under the ‘‘Dollo principle of parsimony’’ (Farris, 1977; Niimura et al., 2018). In this case, the absence of a gene from an OGG in a given species was considered to be irreversible (Matsui et al., 2010). When more than one gene was observed in a given species in a clade that encompasses an OGG, the

number of gains was determined according to the bootstrap value. If more than one gene was observed in different subclades separated by a bootstrap value of less than 70%, it was counted as a gain, because a low bootstrap value indicates a recent duplication.

### *Genomic mapping of the OR groups on the chromosomes*

The distance (in base pairs) between neighbouring pairs of OR genes (including pseudogenes) was calculated and a histogram was compiled to analyze the most common distances and define the boundaries of the genomic clusters observed in each species (Supplementary Figures 1–3). Based on this histogram, a genomic cluster of OR genes was defined as a distance of <500 Kb, according to the results obtained here, and the findings of previous studies (Niimura and Nei, 2003; Young et al., 2002). Each OGG was plotted on the chromosomes to analyze the level of conservation of the genomic blocks containing OR genes that were orthologous between the chromosomes of each species. The OGGs were plotted and classified according to their OR group. The OR gene clusters were mapped on the chromosomes of each study species using the ChromoMap v0.4.1 package (Anand and Lopez, 2022) in the R environment.

## **Results**

### *Identification of the OR genes in the Xenopus genome*

Through extensive searches of homologs, more than 2000 potentially functional genes were identified in the three *Xenopus* study species analyzed here, together with more than 700 pseudogenes (Figure 1). Each OR sequence was classified as either intact, pseudo or truncated. Truncated genes were detected only in the *X. borealis* genome (Figure 1). *Xenopus tropicalis* has the largest repertoire of OR genes (1240), followed by *X. laevis* (1048), and *X. borealis*, with 842. The phylogenetic gene tree recovered sequences of each of the six OR groups in all the three genomes evaluated (Figure 2). A similar number of intact OR genes was recorded in the OR groups of the different species, except for the  $\alpha$  group in *X. laevis* (which contained almost 45% of all the  $\alpha$  genes) and a major expansion of the  $\gamma$  group in the *X. tropicalis* genome, which represented 44% of all the  $\gamma$  genes of the *Xenopus* species studied here. The proportion

of pseudogenes was calculated as a percentage of the total genes found in each genome or subgenome, with a percentage of 29.58% being recorded in *X. laevis*, 21.49% in *X. borealis*, and 20.16% in *X. tropicalis* (Figure 1).

#### *Identification of the OGGs and estimates of the gains/losses in Xenopus*

A total of 534 OGGs were identified in the present study, including all the six OR groups known to occur in *Xenopus* (Supplementary Table 1). The  $\gamma$  group has the majority of the OGGs, with 457 (85% of the total), while OGGs containing only one gene (singletons) were observed in all the species (Supplementary Table 1). *Xenopus tropicalis* has 20 singleton OGGs, which are present in the  $\beta$  and  $\gamma$  groups. *Xenopus laevis* has 17 singletons, the majority of which are found in the  $\alpha$ ,  $\beta$ , and  $\gamma$  groups, while *X. borealis* has 24 singletons, all in the  $\gamma$  group. More than a quarter (144 or 27%) of all the OGGs are composed of one gene per species, which reflects a pattern of vertical transmission from the MRCA. Table 1 shows the number of OR genes per group that compose the seven largest OGGs. A more detailed analysis of the OGGs identified the orthologous copies shared exclusively by the *X. laevis* and *X. borealis* genomes, as observed in the third and sixth largest OGGs recovered here (Table 1). Although the data lack the resolution needed to determine exactly whether these OGGs represent gains in *X. laevis* and *X. borealis* or losses from *X. tropicalis*, the genetic signature observed here was consistent with the recent expansion of paralogous copies (including pseudogenes) of these OGGs in the *X. tropicalis* genome.

Considerable recent gene expansion was identified in one OGG of the  $\gamma$  group (OGG gamma-5), which is composed of 209 genes (162 intact genes and 47 pseudogenes). The relatively short branches in the gene tree reinforce the hypothesis that gene expansion occurred recently in *X. tropicalis* (Figure 2). The OGG values inferred in the species tree analysis were optimized to identify the signature of gain/loss across the overall evolutionary trajectory of the three study species. Based on the results of the present study, *X. tropicalis* has the largest number of gains/losses of OR genes, which would account for the current size of the OR gene repertoire in this species, which represents more than 50% of all the gains observed in the genus *Xenopus*, with 607 gains versus 408 in *X. laevis*, *X. borealis* and their MRCA, in addition to 40% of all the losses (151 vs. 220; Figure 3). Exclusive OGGs (found in only one species) were

also more frequent in *X. tropicalis* (41 OGGs, with a total of 255 genes, distributed in the  $\beta$ ,  $\gamma$ ,  $\delta$ , and  $\eta$  groups). *Xenopus laevis* and *X. borealis* have a similar number of exclusive OGGs, with 24 OGGs (31 genes in total, distributed in the  $\alpha$ ,  $\beta$ ,  $\gamma$ , and  $\eta$  groups) in *X. laevis* and 25 OGGs (26 genes in total, all in the  $\gamma$  group) in *X. borealis*.

#### *Genomic distribution of OR genes in the Xenopus species*

The chromosomal mapping revealed OR clusters on all the chromosomes of the karyotype of the different *Xenopus* species, except chromosome 6 (Figure 4). As expected, the  $\gamma$  group genes were found scattered throughout the chromosomes. The subtelomeric region of chromosome 2 of all three species had a sequential genomic arrangement containing a large cluster of OR genes, known as a High Order Genomic Cluster (HOGC), which encompasses OR genes from five groups –  $\eta$ ,  $\epsilon$ ,  $\delta$ ,  $\alpha$ , and  $\beta$  (Figure 5). In *X. tropicalis*, the HOGC is a genomic block of approximately 5 Mb (Figure 5). The order and size of each OR group in this genomic arrangement varied slightly among the different species and their subgenomes, which indicates the occurrence of intra-chromosomal micro-rearrangements. The order of the groups is maintained in *X. tropicalis* and in chromosome 2S of both *X. laevis* and *X. borealis*, although a number of rearrangements are apparent in chromosome 2L (Figure 5). In *X. borealis*, chromosome 2L has a 715Kb inversion of  $\delta$  and  $\epsilon$ , and what appears to be the loss of a genomic block from the tandem array of the  $\eta$  genes. In *X. laevis*, this chromosome has a complete inversion of approximately 4.8 Mb of the genomic segment, in which the OR gene groups are in inverted order –  $\beta$ ,  $\alpha$ ,  $\delta$ ,  $\epsilon$ , and  $\eta$ . In addition, one intact  $\eta$  gene and two pseudogenes of the  $\epsilon$  group were clustered on chromosome 5L, which appears to represent a recent rearrangement (Figure 5). In general, copies of the pseudogenes tend to be intercalated between the intact genes, although some chromosomes have a very small number of OR genes, such as chromosome 4S of both *X. laevis* and *X. borealis*, which each have only one intact OR gene, while chromosome 4 of *X. tropicalis* has six intact genes and three pseudogenes. Similarly, *X. tropicalis* chromosome 10 has eight intact genes, and no pseudogenes (Figure 5).

While chromosome 3 has the largest number of OR genes in all the study species analyzed here, a comparison of the number of genes found in each subgenome indicates

that the S subgenome contains ~25% of all the OR genes in *X. laevis* and *X. borealis*. While 785 were found in the L subgenome of *X. laevis* and 613 in that of *X. borealis*, only 263 and 210 genes were found in the S subgenomes of the two species, respectively. In all, chromosome 3L has ~27% of the total number of genes found in the genomes of the two species, that is, 297 in *X. laevis* and 205 in *X. borealis* (see Supplementary Table 6) and a similarly high percentage was also observed on chromosome 3 of *X. tropicalis*, which has as many as 34% of all the genes (424) located on these chromosomes.

#### *Relationship between the genomic clusters and the OGGs*

The OGGs of the  $\gamma$  group were scattered among almost all the chromosomes, although the largest clusters are found in chromosomes 2, 3, 7, and 9 in all the studied species (Table 2; Supplementary Table 3–6). Most clusters do not represent the same number of genes from a given OGG, which means that varying numbers of genes from different OGGs may comprise a given cluster. For example, while cluster Xtro7.1 of chromosome 7 in *X. tropicalis* presents all 209 genes of the gamma-5 OGG clustered together in a tandem array, at least 10 other OGGs can be observed in this cluster, which together comprise a total of 253 genes, including those of OGG-5 (Table 2).

## **Discussion**

### *Highly-conserved evolution of the OR gene repertoire in the Xenopus species*

The analyses presented here assessed the entire OR gene repertoires of *X. tropicalis*, *X. laevis*, and *X. borealis*. The use of the most recent genome assemblies of *X. tropicalis* and *X. laevis* allowed an increase in the number of sequences of intact genes and pseudogenes in comparison with the previous versions of the genome (Niimura, 2009a; Lu et al., 2021; Han et al., 2022), as well as eliminated the truncated copies observed in these versions. Truncated genes were observed in *X. borealis*, however, which appears to reflect the level of the genome assembly contiguity available for this species. These results reinforce the importance of chromosome-scale genomes,

such as those available for *X. laevis* and *X. tropicalis*, for a more comprehensive evaluation of the OR gene repertoire, and reliable comparisons among species.

The OR gene repertoires of the three studied species are all similar in size and in the composition of the OR groups, even though the species of allopolyploid origin, i.e., *X. laevis* and *X. borealis*, which both have genomes of approximately 2.75 Gb (Session et al., 2016; Evans et al., 2022), have much larger genomes than the diploid species *X. tropicalis*, with 1.45 Gb (Hellsten et al., 2010). These three species share similar ecological niches and a series of morphological and physiological adaptations to an aquatic lifestyle (Tinsley et al., 1996; Irisarri et al., 2011; Evans et al., 2015). Given the exposure of their chemical-sensorial systems to similar selective pressures in the aquatic environment, a high level of conservation in the OR gene repertoire would be expected in these species, as observed here, with vertical transmission from the MRCA, dated to at least 48 Ma (Session et al., 2016). In fact, a large proportion of the Orthologous Gene Groups (OGGs) were highly conserved in all three species. Overall, of the OGGs predicted in the analyses presented here, that were putatively functional in the MRCA of the genus *Xenopus*, at least 146 have persisted over the long term in the genomes of all three study species. This indicates that intense purifying selection pressured these loci, preventing the loss of genes even during the diversification of the species.

While *Xenopus tropicalis* has the most extensive OR gene repertoire of any *Xenopus* species for which data are available, as corroborated by the estimates of gains/losses in the evolutionary tree presented here, a closer examination of the data revealed that this was the result of a major expansion in the number of copies of the  $\gamma$  OR gene group. This observation is consistent with the presence of the exclusive OGG-5 ( $\gamma$ ) in this species, which includes 209 copies (162 of intact genes and 47 of pseudogenes). This group has recently undergone high rates of tandem duplication, which has resulted in a well-arranged cluster of genes in a tandem array. No copies of this OGG are found in any other chromosome region, which indicates extensive rounds of tandem duplication in this genomic region. Recent research has shown that the  $\alpha$  and  $\gamma$  groups are evolving at one of the fastest rates documented in any tetrapod genome (Yohe et al., 2020; Silva et al., 2020). A similar rapid expansion of the  $\gamma$  group has been observed in chicken (*Gallus gallus*) and zebra fish (*Danio rerio*) lineages, with a strong signature of positive selection in paralogous copies, which suggests that most avian  $\gamma$

group sequences are subject to codon-purifying selection pressure (Steiger et al., 2009). The large number of OR genes in the  $\gamma$  group, present as singletons or genomic clusters throughout the genome, except on chromosome 6, points to a massive species-specific expansion in the *X. tropicalis* genome, which may reflect species-specific chemosensory adaptations to a new environmental scenario.

*Intra-chromosomal rearrangements in post-allopolyploid species modulate the diversification of the subgenomes*

The *Xenopus* karyotype has high levels of chromosomal homology, which were revealed here by the syntenic blocks of OR clusters that have been conserved for at least 48 Ma in all three species. The subterminal region of chromosome 2 bears a High Order Genomic Cluster (HOGC) composed of  $\eta$ ,  $\epsilon$ ,  $\delta$ ,  $\alpha$ , and  $\beta$  OR genes. This HOGC is present on chromosomes 2L and 2S in both allopolyploid species, although several intra-chromosomal micro-rearrangements and segmental deletions are found between the two subgenomes. In *X. laevis*, for example, paracentric inversions involving a genomic segment of approximately 4.7 Mb have inverted the HOGC completely in chromosome 2L, in comparison with 2S, while the latter chromosome also presents a signature of extensive loss of the paralogous copies. In *X. borealis*, on the other hand, the paracentric inversion involved only the  $\delta$  and  $\epsilon$  clusters of the HOGC on chromosome 2L, which resulted in the internal reorganization of the HOGC in comparison with chromosome 2 of *X. tropicalis*. The occurrence of the paracentric inversion in the HOGC on chromosome 2L in both *X. laevis* and *X. borealis* may also have contributed to a reduction in the recombination of these genomic segments during meiosis, which would have accelerated the diploidization process in these species.

The results of the present study indicate that the HOGC may have suffered the effects of the gene fractionation observed between the L and S subgenomes of *X. borealis* and *X. laevis* following the chromosomal inversions. Biased deletions of functional and pseudo OR genes were observed in the S subgenome in comparison with the L subgenome of both allopolyploid species. It is possible that the suppression of recombination in the HOGC region, associated with the more relaxed selective pressure on one of the redundant copies, accounts for the asymmetrical loss of genes in the OR gene clusters (and in the size of the HOGC) observed between the 2L and 2S genomes.

This process is known as biased genome fractionation (Langham et al., 2004), in which extra duplicated gene copies that originated during allopolyploidy or hybridization are expressed preferentially at lower levels than the other homologous copies in one of the subgenomes. Gene fractionation is a common process in hybrids and post-allopolyploid species (Mandáková and Lysak, 2018; Freeling et al., 2015, Cheng et al., 2018), and represents a strategy to avoid an imbalance in gene dosage that may reduce the fitness of the organism, if it is dosage sensitive. This bias of gene expression toward one copy (or subgenome) maintains an intense level of selective pressure on this copy, while the poorly-expressed paralogous copy may experience relaxed selective pressure, and as a consequence, accumulate mutations that result in its degeneration or even its eventual loss from the genome (Freeling et al., 2012). The differential rates of gene retention and gene expression, and the unequal epigenetic markers observed between the L and S subgenomes in *X. laevis* (Session et al., 2016; Elurbe et al., 2017; Furman et al., 2018) and *X. borealis* (Furman et al., 2018; Suda et al., 2022) are consistent with the influence of genome fractionation mechanisms. Differences in the content of Transposable Element (TEs) between subgenomes may also contribute to genome fractionation in post-allopolyploid species via epigenetic regulation (Freeling et al., 2015; Renny-Byfield et al., 2017), which may then impact gene expression in the different subgenomes (Cheng et al., 2016; Lannes et al., 2019; Gill et al., 2021; Giraud et al., 2021). In both *X. laevis* (Session et al., 2016) and *X. borealis* (Suda et al., 2022), differences between subgenomes in the TE content may contribute differentially to the expression of the duplicated copies of the genes found in each subgenome, reinforcing the hypothesis of genome fractionation.

At a fine scale, a sequential pattern of HOGC organization was observed, in which each OR gene group is composed of a mixture of the copies of intact genes and pseudogenes, which is compatible with the theoretical predictions on the generation of repetitive genes in the birth-and-death model (Nei and Rooney, 2005). Furthermore, despite the overall conservation of these genomic clusters and the expected rapid expansion of the  $\gamma$  OR genes, some exclusive OR genes (from different OR groups) were found in all three of the species analyzed here. These features, including the large number of pseudogenes, point to a high ratio of birth-and-death processes (Nei and Rooney, 2005), which may have arisen, putatively, as a result of the varied strategies applied by the different *Xenopus* species during the aquatic-terrestrial transition, which

underpins the evolution of their OR genes. This process also appears to have happened at different rates in different OGGs or OR genes, as observed in the relatively rapid duplication of the gamma-5 OGG. A similar pattern has been observed in other tetrapod species, involving primarily the 5/8/9 and OR 6 OR subfamilies, which may evolve at a relatively rapid pace in comparison with the other subfamilies (Yohe et al., 2020). In fact, significant differences in evolutionary rates are also found among subfamilies of OR genes in humans (Wang et al., 2023).

Overall, the findings of the present study clearly indicate that the evolution of the OR repertoire of the diploid and post-allopolyploid *Xenopus* species has been modulated primarily by birth-and-death evolutionary processes. In particular, the analyses revealed the critical role of intra-chromosomal rearrangements and gene fractionation, which occurred during the process of diploidization in the post-allopolyploidy *X. laevis* and *X. borealis* genomes, and provides important insights into the evolution of the OR gene repertoire in these species.

## References

- Altschul, S.F., Gish, W., Miller, W., Myers, E.W., & Lipman, D.J., 1990. Basic local alignment search tool. *J. Mol. Bio.* 215(3), 403–410.
- Cheng, F., Wu J., Cai X., Liang J., Freeling M., Wang, X., 2018. Gene retention, fractionation and subgenome differences in polyploid plants. *Nat. Plants.* 4(5), 258–268. doi:10.1038/s41477-018-0136-7.
- Cheng, F., Sun, C., Wu, J., Schnable, J., Woodhouse, M.R., Liang, J., Cai, C., Freeling, M., 2016. Epigenetic regulation of subgenome dominance following whole genome triplication in *Brassica rapa*. *New Phytol.* 211(1), 288–299. doi: 10.1111/nph.13884.
- Dong, D., He, G., Zhang, S., Zhang, Z., 2009. Evolution of olfactory receptor genes in primates dominated by birth-and-death process. *Genome Biol. Evol.* 1, 258–64. <https://doi.org/10.1093/gbe/evp026> PMID: 20333195.
- Edger, P.P., Poorten, T., VanBuren, R., Hardigan, M.A., Colle, M., McKain, M.R., Smith, R.D., Teresi, S., Nelson, A.D.L., Wai, C.M., Alger, E.I., Bird, K.A., Yocca, A.E., Pumphlin, N., Ou, S., Ben-Zvi, G., Brodt, A., Baruch, K., Swale, T., Shiue, L., Acharya, C.B., Cole, G.S., Mower, J.P., Childs, K.L., Jiang, N., Lyons, E., Freeling, M.,

- Puzey, J.R., Knapp, S.J., 2019. Origin and evolution of the octoploid strawberry genome. *Nat. Genet.* 51(3), 541–547. doi: 10.1038/s41588-019-0356-4.
- Elurbe, D. M., Paranjpe, S. S., Georgiou, G., Van Kruijsbergen, I., Bogdanovic, O., Gibeaux, R., Heald, R., Lister, R., Huynen, M. A., Van Heeringen, S. J., Veenstra, G. J. C., 2017. Regulatory remodeling in the allo-tetraploid frog *Xenopus laevis*. *Genome Bio.* 18, 1–18. doi : 10.1186/s13059-017-1335-7
- Evans, B.J., 2008. Genome Evolution and speciation genetics of allopolyploid clawed frogs (*Xenopus* and *Silurana*). *Front Biosci.* 13, 4687–706.
- Evans, B.J., Carter, T.F., Greenbaum, E., Gvoždík, V., Kelley, D.B., McLaughlin, P.J., Pauwels, O.S.G., Portik, D.M., Stanley, E.L., Tinsley, R.C., Tobias, M.L., Blackburn, D.C., 2015. Genetics, morphology, advertisement calls, and historical records distinguish six new polyploid species of African Clawed Frog (*Xenopus*, Pipidae) from West and Central Africa. *Plos One* 10(2), e0412823.
- Evans, B.J., Mudd, A.B., Bredeson, J.V., Furman, B.L.S., Wasonga, D.V., Lyons, J.B., Harland, R.M., Rokhsar, D.S., 2022. New insights into *Xenopus* sex chromosome genomics from the Marsabit clawed frog *X. borealis*. *J. Evol. Bio.* 35(12), 1777–1790.
- Farris, J.S., 1977. Phylogenetic analysis under Dollo's law. *Syst. Zool.* 26(1), 77–88.
- Freeling, M., Scanlon, M.J., Fowler, J.E., 2015. Fractionation and subfunctionalization following genome duplications: mechanisms that drive gene content and their consequences. *Curr. Opin. Genet. Dev.* 35, 110–118.  
<http://dx.doi.org/10.1016/j.gde.2015.11.002>.
- Frost, D. R., 2023. Amphibian Species of the World: an Online Reference. Version 6.2.  
<https://amphibiansoftheworld.amnh.org/index.php>. doi.org/10.5531/db.vz.0001. (accessed 13 October 2023)
- Furman, B. L. S., Evans, B. J., 2018. Divergent evolutionary trajectories of two young, homomorphic, and closely related sex chromosome systems. *Genome Biol. Evol.* 10, 742–755.
- Gill, R.A., Scossa, F., King, G.J., Golicz, A.A., Tong, C., Snowdon, R.J., Fernie, A.R., Liu, S., 2021. On the role of transposable elements in the regulation of gene expression

and subgenomic interactions in crop genomes. *Crit. Rev. Plant. Sci.* 40(2), 157–189.  
<https://doi.org/10.1080/07352689.2021.1920731>.

Han, W., Wu, Y., Zeng, L., Zhao, S., 2022. Building the Chordata olfactory receptor database using more than 400,000 receptors annotated by Genome2OR. *Sci. China Life Sci.* 65, 2539–2551.

Hellsten, U., Harland, R.M., Gilchrist, M.J., Hendrix D., Jurka J., Kapitonov V., Ovcharenko, I., Putnam, N.H., Shu, S., Taher, L., Blitz, I.L., Blumberg, B., Dichamnn, D.S., Dubchak, I., Amaya, E., Detter, J.C., Fletcher, R., Gerhard, D.S., Goodstein, D., Graves, T., Grigoriev, I.V., Grimwood, J., Kawashima, T., Lindquist, E., Lucas, S.M., Mead, P.E., Mitros, T., Ogino, H., Ohta, Y., Poliakov, A.V., Pollet, N., Robert, J., Salamov, A., Sater, A.K., Schumtz, J., Terry, A., Vize, P.D., Warren, W.C., Wells, D., Grainger, R., Grammer, T., Khokha, M.K., Richardson, P.M., Rokhsar, D.S., 2010. The genome of the Western clawed frog *Xenopus tropicalis*. *Science* 328(5978), 633–636.

Huang, Y., Niu, B., Gao, Y. Fu, L., Li, W., 2010. CD-HIT Suite: a web server for clustering and comparing biological sequences. *Bioinformatics* 26, 680–682.

Hughes, G.M., Boston, E.S.M., Finarellim J.A., Murphym W.J., Higgins, D.G., Teeling, E.C. 2018. The Birth and Death of Olfactory Receptor Gene Families in Mammalian Niche Adaptation. *Mol. Biol. Evol.* 35(6), 1390– 406.  
<https://doi.org/10.1093/molbev/msy028>.

Irisarri, I., Vences, M., Mauro, D.S., Glaw, F., Zardoya, R., 2011. Reversal to air-driven sound production revealed by a molecular phylogeny of tongueless frogs, family Pipidae. *BMC Evol. Biol.* 27(11), 114–124.

Jia, J., Xie, Y., Cheng, J., Kong, C., Wang, M., Gao, L., Zhao, F., Guo, J., Wang, K., Li, G., Cui, D., 2021. Homology-mediated inter-chromosomal interactions in hexaploid wheat lead to specific subgenome territories following polyploidization and introgression. *Genome Biol.* 22(1), 1–21. doi:10.1186/s13059-020-02225-7.

Katoh ,K., Rozewicki, J., Yamada ,K.D., 2019. MAFFT online service: multiple sequence alignment, interactive sequence choice and visualization. *Brief. Bioinformatics.* 20(4), 1160–1166. <https://doi.org/10.1093/bib/bbx108>.

- Langham, R.J., Walsh, J., Dunn, M., Ko, C., Goff, S.A., Freeling, M., 2004. Genomic duplication, fractionation and the origin of regulatory novelty. *Genetics*. 166(2), 935–945. doi: 10.1534/genetics.166.2.935.
- Liu, S., Liu, Y., Yang, X., Tong, C., Edwards, D., Parkin, I.A.P., Zhao, M., Ma, J., Yu, J., Huang, S., Wang, X., Wang, J., Lu, K., Fang, Z., Bancroft, I., Yang, T.J., Hu, Q., Wang, X., Yue, Z., Li, H., Yang, L., Wu, J., Zhou, Q., Wang, W., King, G.J., Pires, J.C., Lu, C., Wu, Z., Sampath, P., Wang, Z., Guo, H., Pan, S., Yang, L., Min, J., Zhang, D., Jin, D., Li, W., Belcram, H., Tu, J., Guan, M., Qi, C., Du, D., Li, J., Jiang, L., Batley, J., Sharpe, A.G., Park, B.S., Rupearo, P., Cheng, F., Waminal, N.E., Huang, Y., Dong, C., Wang, L., Li, J., Hu, Z., Zhuang, M., Huang, Y., Huang, J., Shi, J., Mei, D., Liu, J., Lee, T.H., Wang, J., Jin, H., Li, Z., Li, X., Zhang, J., Xiao, L., Zhou, Y., Liu, Z., Liu, X., Qin, R., Tnag, X., Liu, W., Wang, Y., Zhang, Y., Lee, J., Kin, H.H., Denoeud, F., Xu, X., Liang, X., Hua, W., Wang, X., Wang, J., Chalhou, B., Paterson, A.H., 2014. The *Brassica oleracea* genome reveals the asymmetrical evolution of polyploid genomes. *Nat. Commun.* 5(1), 3930. Doi: 10.1038/ncomms4930.
- Lu, B., Jiang, J., Wu, H., Chen, X., Song, X., Liao, W., Fu, J., 2021. A large genome with chromosome-scale assembly sheds light on the evolutionary success of a true toad (*Bufo gargarizans*). *Mol. Ecol. Resour.* 21, 1256–1273.
- Matsuda, Y., Uno, Y., Kondo, M., Gilchrist, M. J., Zorn, A. M., Rokhsar, D. S., Schmid, M., Taira, M., 2015. A new nomenclature of *Xenopus laevis* chromosomes based on the phylogenetic relationship to *Silurana/Xenopus tropicalis*. *Cytogenet. Genome Res.* 145, 187-191.
- Matsui, A., Go, Y., Niimura, Y., 2010. Degeneration of olfactory receptor gene repertoires in Primates: no direct link to full trichromatic vision. *Mol Bio. Evol* 27(5), 1192–1200.
- Mei, W., Liu, S., Schnable, J.C., Yeh, C.-T., Springer, N.M., Schnable, P.S., Barbazuk, W.B., 2017. A comprehensive analysis of alternative splicing in paleopolyploid maize. *Front. Plant Sci.* 8, 694. Doi:10.3389/ fpls.2017.00694.
- Nei, M., Rooney, A.P., 2005. Concerted and Birth-and-Death evolution of multigene families. *Annu. Rev. Genet.* 39, 121–152.

- Niimura, Y., 2009a. On the origin of evolution of vertebrate olfactory receptor genes: comparative genome analysis among 23 chordate species. *Genome Bio. Evol.* 1, 34–44.
- Niimura, Y., 2009b. Evolutionary dynamics of olfactory receptor genes in chordates: interaction between environment and genomic contents. *Hum. Genomics.* 4(2), 107–118.
- Nimura, Y., 2012. Olfactory receptor multigene family in vertebrates: from the viewpoint of evolutionary genomics. *Curr. Genomics.* 13, 103–114.
- Niimura, Y., 2013a. Identification of chemosensory receptor genes from vertebrate genomes. *Methods Mol. Bio.* 1068, 95–105.
- Niimura, Y., 2013b. Identification of olfactory receptor genes from mammalian genome sequences. *Methods Mol. Bio.* 1003, 39–49.
- Niimura, Y., Matsui, A., Touhara, K., 2014. Extreme expansion of the olfactory receptor gene repertoire in African elephants and evolutionary dynamics of orthologous gene groups in 13 placental mammals. *Genome Res.* 24(9), 1485–1496.
- Niimura, Y., Matsui, A., Touhara, K., (2018). Acceleration of olfactory receptor gene loss in primate evolution: possible link to anatomical change in sensory systems and dietary transition. *Mol. Biol. Evol.* 35(6), 1437–1450.
- Niimura, Y., Nei, M., 2003. Evolution of olfactory receptor genes in the human genome. *PNAS.* 100(21), 12235–12240.
- Olender, T., Nativ, N., Lancet, D., 2013. HORDE: comprehensive resource for olfactory receptor genomics. *Methods Mol. Bio.* 1003, 23–38.
- Renny-Byfield, S., Rodgers-Melnick, E., Ross-Ibarra, J., 2017. Gene fraction and function in the Ancient subgenomes of Maize. *Mol. Bio. Evol.* 34(8), 1825-1832. <https://doi.org/10.1093/molbev/msx121>
- Saitou, N., Nei, M., 1987. The neighbor-joining method: a new method for Reconstructing Phylogenetic Trees. *Mol. Bio. Evol.* 4(4), 406–425.
- Schiavinato, M., Bodrug-Schepers, A., Dohm, J.C., Himmelbauer, H., 2021. Subgenome evolution in allotetraploid plants. *Plant J.* 106(3), 672–688. doi:10.1111/tpj.15190.
- Session, A.M., Uno, Y., Kwon, T., Chapman, J.A., Toyoda, A., Takahashi, S., Fukui, A., Hikosaka, A., Suzuki, A., Kondo, M., van Heeringen, S.J., Quigley, I., Heinz, S.,

Ogino, H., Ochi, H., Hellsten, U., Lyons, J.B., Simakov, O., Putnam, N., Stites, J., Kuroki, Y., Tanaka, T., Michiue, T., Watanabe, M., Bogdanovic, O., Lister, R., Georgiou, G., Paranjpe, S.S., van Kruijsbergen, I., Shengquiang, S., Carlson, J., Kinoshita, T., Ohta, Y., Mawaribuchi, S., Jenkins, J., Grimwood, J., Schmutz, J., Mitros, T., Mozaffari, S.V., Suzuki, Y., Haramoto, Y., Yamamoto, T.S., Takagi, C., Heald, R., Miller, K., Haudenschield, C., Kitzman, J., Nakayama, T., Izutsu, Y., Robert, J., Fortriede, J., Burns, K., Lotay, V., Karimi, K., Yasuoka, Y., Dichmann, D.S., Flajnik, M.F., Houston, D.W., Shendure, J., Dupasquier, L., Vize, P.D., Zorn, A.M., Ito, M., Marcotte, E.M., Wallingford, J.B., Ito, Y., Asashima, M., Ueno, N., Matsuda, Y., Veenstra, G.J.C., Fujiyama, A., Harland, R.M., Taira, M., Rokhsar, D.S., 2016. Genome evolution in the allotetraploid frog *Xenopus laevis*. *Nature*. 538, 336–343.

Silva, M.C., Chibucos, M., Munro, J.B., Daugherty, S., Coelho, M.M., Silva, J.C., 2020. Signature of adaptive evolution in olfactory receptor genes in Cory's Shearwater supports molecular basis for smell in procellariiform seabirds. *Sci. Rep.* 10(543), 1–11. <https://doi.org/10.1038/s41598-019-56950-6>.

Steiger, S.S., Kuryshev, V.Y., Stensmyr, M.C., Kempnaers, B., Mueller, J.C., 2009. A comparison of reptilian and avian olfactory receptor gene repertoires: Species-specific expansion of group  $\gamma$  in birds. *BMC Genomics*. 10(446), 1–10. <https://doi.org/10.1186/1471-2164-10-446>.

Suda, K., Hayashi, S.R., Tamura, K., Takamatsu, N., Ito, M., 2022. Activation of DNA transposons and evolution of piRNA genes through interspecific hybridization in *Xenopus* frogs. *Front. Genet.* 13, 1–9.

Takezaki, N., Rzhetsky, A., Nei, M., 1995. Phylogenetic test of the molecular clock and linearized trees. *Mol. Bio. Evol.* 12(5), 823–833. doi: 10.1093/oxfordjournals.molbev.a040259.

Tinsley, R.C., Loumont, C., Kobel, H.R., 1996. Geographical distribution and ecology, in: Tinsley, R.C., Kobel, H.R. (Eds.), *The Biology of Xenopus*. University Press, Oxford, pp. 35–59.

Wang, Y., Sun, Y., Joseph, P.V., 2023. Diverse evolutionary rates and gene duplication patterns among families of functional olfactory receptor genes in humans. *PLOS One*. 18(4), e0282575. DOI:<https://doi.org/10.1371/journal.pone.0282575>.

Wolfe, K.H., 2001. Yesterday's polyploids and the mystery of diploidization. *Nat. Rev. Genet.* 2(5), 333–341.

Yohe, L. R., Leiser-Miller, L.B., Kaliszewska, Z.A., Donat, P., Santana, S.E., Dávalos, L.M., 2021. Diversity in olfactory receptor repertoires is associated with dietary specialization in a genus of frugivorous bat. *G3*. 11(10), jkab260.

Yohe, L.R., Fabbri, M., Hanson, M., Bhullar, B.S., 2020. Olfactory receptor gene evolution is unusually rapid across Tetrapoda and outpaces chemosensory phenotypic change. *Current Zool.* 66(5), 505–514.

Yohe, L.R., Fabbri, M., Lee, D., Davies, K.T.J., Yohe, T.P., Sánchez, M.K.R., Rengifo, E.M., Hall, R.P., Mutumi, G., Hedrick, B.P., Sadier, A., Simmons, N.B., Sears, K.E., Dumont, E., Rossiter, S.J., Bhullar, B.S., Dávalos, L.M., 2022. Ecological constraints on highly evolvable olfactory receptor genes and morphology in neotropical bats. *Evolution*. 76(10), 2347–2360.

Young, J. M., Friedman, C., Williams, E. M., Ross, J. A., Tonnes-Priddy, L. Trask, B. J., (2002). Different evolutionary processes shaped the mouse and human olfactory receptor gene families. *Hum. Mol. Genet.* 11(5), 535–546.

Zhao, Y., Dong, L., Jiang, C., Wang, X., Xie, J., Rashid, M.A.R., Liu, Y., Li, M., Bu, Z., Wang, H., Ma, X., Sun, S., Wang, X., Bo, C., Zhou, T., Kong, L., 2020. Distinct nucleotide patterns among three subgenomes of bread wheat and their potential origins during domestication after allopolyploidization. *BMC Biol.* 18(1), 188.  
doi:10.1186/s12915-020-00917-x.

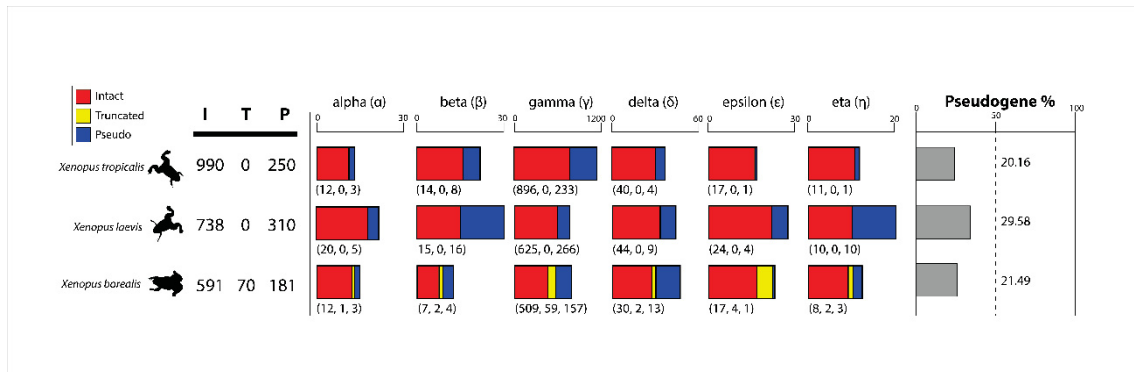


Figure 1. Number of OR genes in the genome sequence from *Xenopus* species. “I”, “T”, and “P” represent the number of intact genes, truncated genes, and pseudogenes, respectively. An intact gene was defined as a sequence starting from an initiation codon and ending with a stop codon that did not contain any disrupting mutations. A pseudogene was defined as a sequence with a nonsense mutation, frameshift, deletion within conserved regions, or some combination thereof. A truncated gene was defined as a partial, intact sequence located at a contig end. An intact gene was assumed to be functional, while a truncated gene was presumed to be either a functional gene or a pseudogene. The fraction of OR pseudogenes was calculated as the number of OR pseudogenes divided by the total number of OR genes.

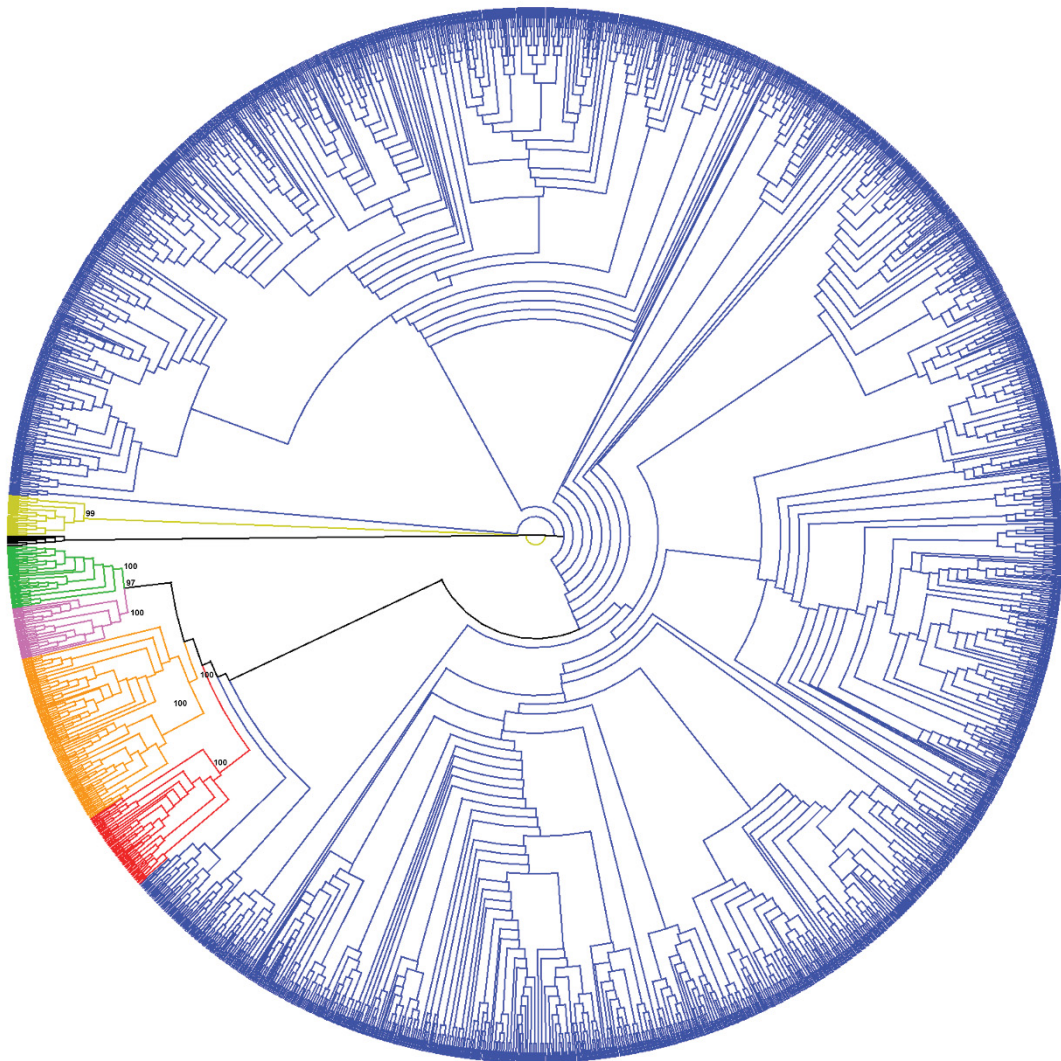


Figure 2. phylogenetic tree for all the intact genes from *Xenopus* species. Coloured branches indicated as follow: alpha ( $\alpha$ ) – green; beta ( $\beta$ ) – pink; gamma ( $\gamma$ ) – blue; delta ( $\delta$ ) – orange; epsilon ( $\epsilon$ ) – red; eta ( $\eta$ ) – yellow.

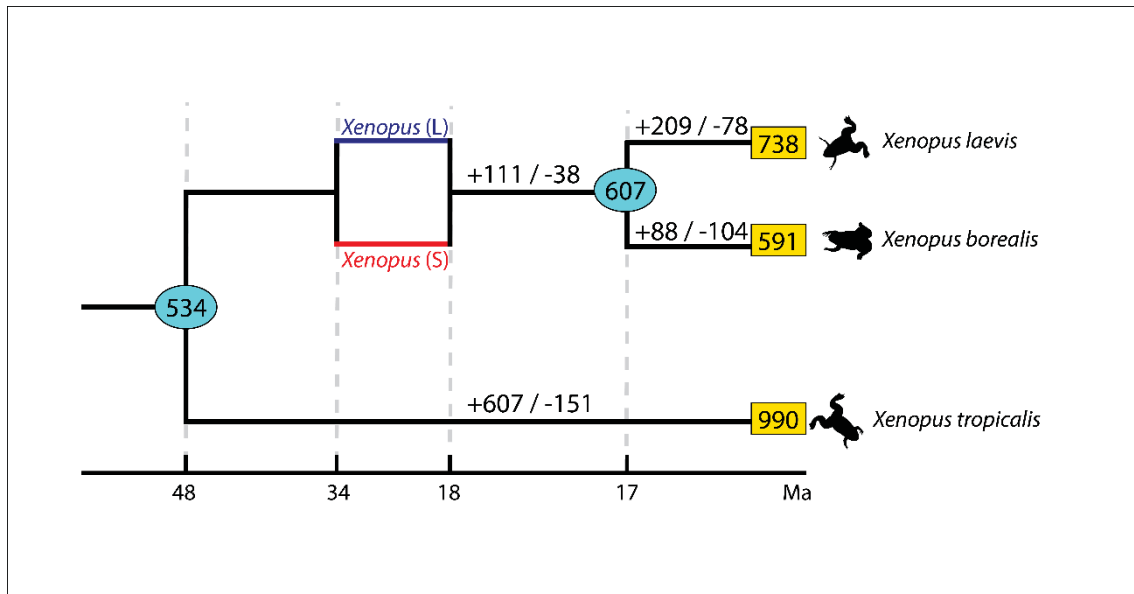


Figure 3. Gain/Loss of OR genes during the evolution of *Xenopus* species. Each number in a yellow box indicates the number of intact OR genes in an extant species. Each number in a cyan oval represents the number of functional OR genes in an ancestral node estimated by the reconciled-tree method (Niimura and Nei 2007). Estimated numbers of gene gains and gene losses in each branch are also shown.

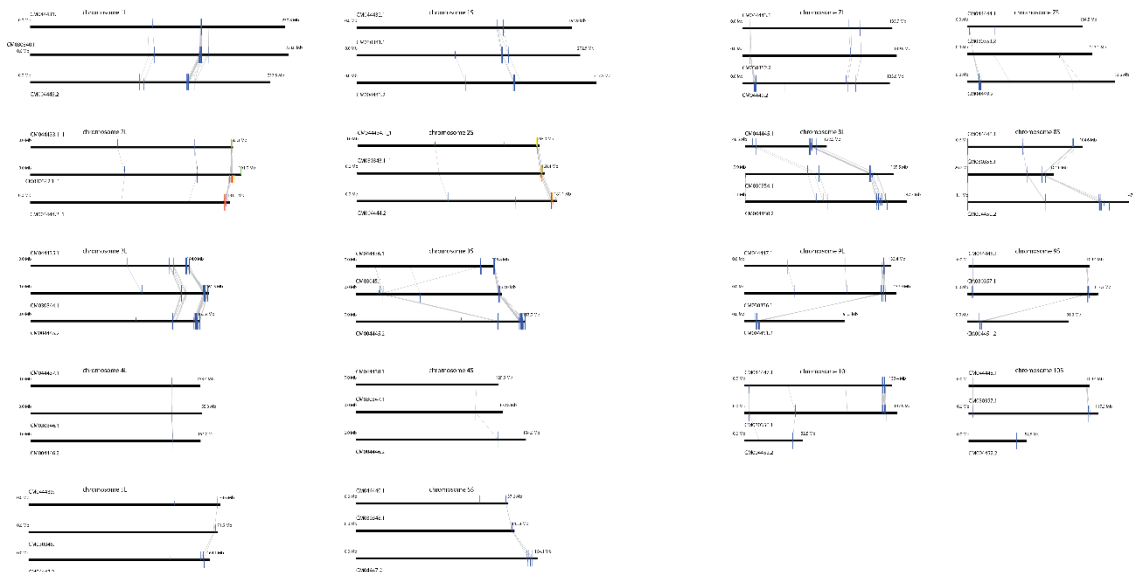


Figure 4. Distribution of OR genes on *Xenopus* chromosomes. Vertical bars above and below the chromosomes indicate the locations of OR functional genes and pseudogenes, respectively. Genomic clusters containing five or more OR genes (including pseudogenes) are represented by boxes. Singletons are represented by triangles. Each vertical bar is colored according to the OGG OR group to which the OR gene: alpha ( $\alpha$ ) – green; beta ( $\beta$ ) – pink; delta ( $\delta$ ) – orange; epsilon ( $\epsilon$ ) – red; eta ( $\eta$ ) – yellow.

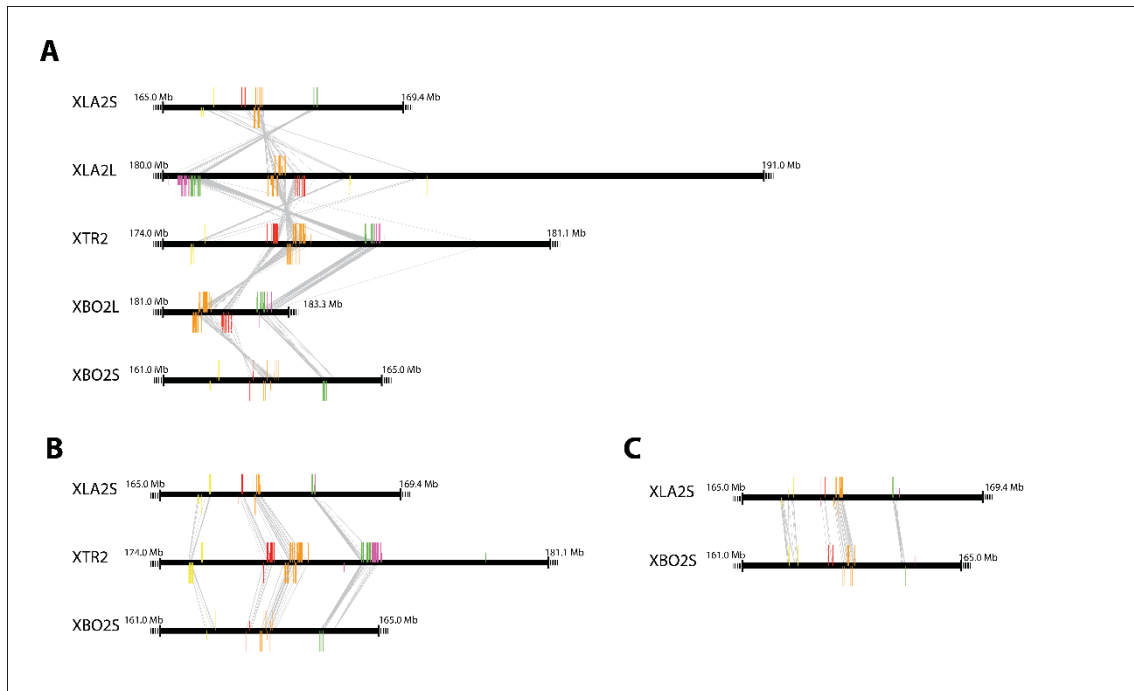


Figure 5. Distribution of OR genes on *Xenopus* chromosome. The position of each OR gene is represented by a colored vertical bar above or below a horizontal line, the former indicating an intact gene and the latter a pseudogene. (A) Comparative cytogenetic map of XLA (*Xenopus laevis*), XTR (*Xenopus tropicalis*) and XBO (*Xenopus borealis*) OR genes on chromosome 2, including all subgenomes. (B) Comparative of OR gene distribution of chromosome 2S from XLA and XBO and XTR. (C) Comparative OR gene distribution between chromosomes 2S from XLA and XBO. Each vertical bar is colored according to the OGG OR group to which the OR gene: alpha ( $\alpha$ ) – green; beta ( $\beta$ ) – pink; delta ( $\delta$ ) – orange; epsilon ( $\epsilon$ ) – red; eta ( $\eta$ ) – yellow.

Table 1. List of the seven greatest OGGs in *Xenopus* in ascending order

OGG name	<i>X. tropicalis</i>	<i>X. laevis</i>	<i>X. borealis</i>	TOTAL
gamma-5	209	0	0	209
gamma-350	71	17	14	102
gamma-234	40	21	17	78
gamma-273	0	39	16	55
gamma-432	45	5	0	50
gamma-163	0	25	17	42
alpha-9	8	14	9	31

Table 2. List of the five greatest clusters in *Xenopus* chromosomes

<i>Xenopus tropicalis</i>				
Clusters	Intact	Pseudogene	Truncated	Total
3.6	233	45	0	278
7.1	199	54	0	253
9.1	84	14	0	98
3.2	54	21	0	75
2.4	57	5	0	62
<i>Xenopus laevis</i>				
Clusters	Intact	Pseudogene	Truncated	Total
3L.5	106	38	0	144
3L.7	55	21	0	76
3S.5	57	7	0	64
2L.4	45	11	0	56
9_10L.6	47	7	0	54
<i>Xenopus borealis</i>				
Clusters	Intact	Pseudogene	Truncated	Total
3L.5	75	17	15	107
3S.5	46	14	9	69
3L.7	45	13	3	61
9_10L.6	42	42	12	96
2L.4	25	9	2	36

## Supplementary Information

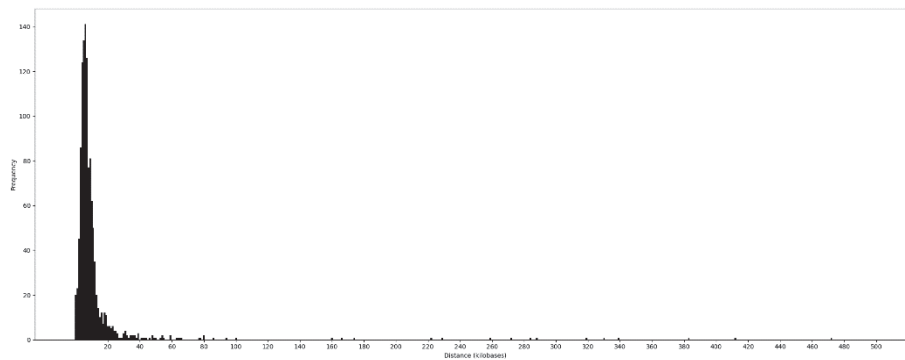


Figure 1S. Distribution of distances between two neighbouring genes in *Xenopus tropicalis*.

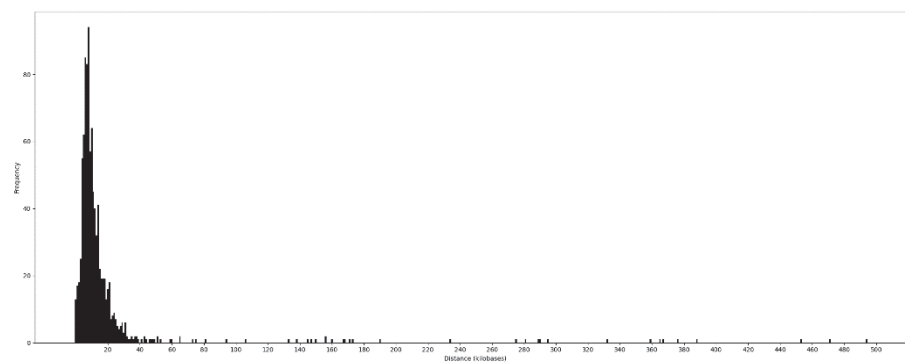


Figure 2S. Distribution of distances between two neighbouring genes in *Xenopus laevis*.

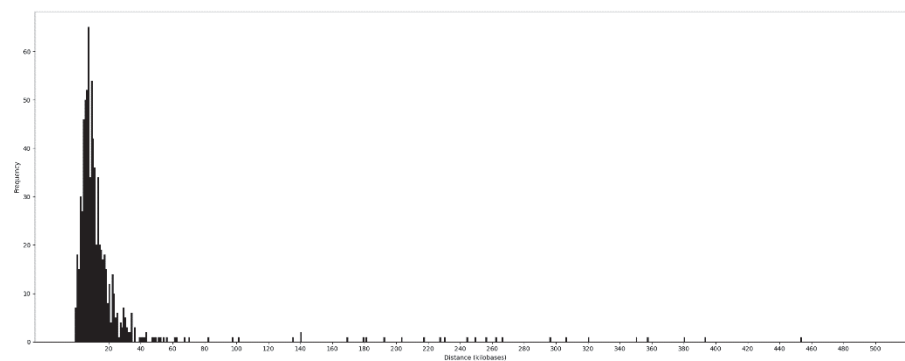


Figure 3S. Distribution of distances between two neighbouring genes in *Xenopus borealis*.

Table 1S. List of all OGGs identified for *Xenopus* species.

OGG number	Intact	Pseudogene	Truncated	Total
alpha-1	3	2	0	5
alpha-2	1	0	0	1
alpha-3	3	0	0	3
alpha-4	2	0	0	2
alpha-5	1	0	0	1
alpha-6	3	0	0	3
alpha-7	3	0	0	3
alpha-8	3	2	0	5
alpha-9	23	7	1	31
alpha-10	2	0	0	2
beta-1	5	9	0	14
beta-2	2	3	0	5
beta-3	1	4	0	5
beta-4	2	0	0	2
beta-5	3	0	0	3
beta-6	3	1	1	5
beta-7	2	0	0	2
beta-8	3	0	0	3
beta-9	1	0	0	1
beta-10	3	5	0	8
beta-11	3	3	1	7
beta-12	5	2	0	7
beta-13	1	0	0	1
beta-14	1	1	0	2
beta-15	1	0	0	1
delta-1	3	0	0	3
delta-2	3	0	0	3
delta-3	3	0	0	3
delta-4	3	0	0	3
delta-5	4	1	0	5
delta-6	5	0	0	5
delta-7	3	0	0	3
delta-8	3	0	0	3
delta-9	2	1	0	3
delta-10	3	0	0	3
delta-11	5	0	1	6
delta-12	4	0	0	4
delta-13	10	0	0	10
delta-14	9	3	0	12
delta-15	3	1	0	4
delta-16	3	2	0	5
delta-17	3	0	0	3
delta-18	9	0	0	9
delta-19	2	1	0	3

delta-20	3	1	0	4
delta-21	4	3	1	8
delta-22	3	3	0	6
delta-23	2	1	0	3
delta-24	6	0	0	6
delta-25	5	4	0	9
delta-26	10	2	0	12
delta-27	5	1	0	6
delta-28	4	2	0	6
epsilon-1	4	0	0	4
epsilon-2	3	2	0	5
epsilon-3	2	0	0	2
epsilon-4	5	1	0	6
epsilon-5	3	0	0	3
epsilon-6	3	1	0	4
epsilon-7	4	0	0	4
epsilon-8	8	1	0	9
epsilon-9	3	1	2	6
epsilon-10	3	0	2	5
epsilon-11	3	0	0	3
epsilon-12	3	0	0	3
epsilon-13	10	0	0	10
epsilon-14	2	0	0	2
epsilon-15	2	0	0	2
eta-1	3	0	0	3
eta-2	2	0	0	2
eta-3	4	2	0	6
eta-4	5	3	0	8
eta-5	3	0	0	3
eta-6	3	8	1	12
eta-7	3	0	0	3
eta-8	3	1	0	4
eta-9	3	0	1	4
gamma-1	3	1	0	4
gamma-2	3	0	0	3
gamma-3	2	2	0	4
gamma-4	3	0	0	3
gamma-5	162	47	0	209
gamma-6	3	1	0	4
gamma-7	3	0	0	3
gamma-8	3	0	0	3
gamma-9	3	0	0	3
gamma-10	4	0	0	4
gamma-11	1	0	0	1
gamma-12	1	0	0	1
gamma-13	3	3	0	6
gamma-14	1	0	0	1

gamma-15	6	7	0	13
gamma-16	2	1	0	3
gamma-17	3	9	0	12
gamma-18	2	0	0	2
gamma-19	2	0	0	2
gamma-20	2	0	0	2
gamma-21	2	2	0	4
gamma-22	3	2	0	5
gamma-23	2	0	0	2
gamma-24	3	2	0	5
gamma-25	2	0	0	2
gamma-26	2	1	1	4
gamma-27	2	0	0	2
gamma-28	2	0	0	2
gamma-29	2	0	0	2
gamma-30	2	0	0	2
gamma-31	4	1	0	5
gamma-32	3	3	0	6
gamma-33	2	0	0	2
gamma-34	3	0	0	3
gamma-35	1	0	0	1
gamma-36	4	2	0	6
gamma-37	3	0	0	3
gamma-38	4	1	0	5
gamma-39	3	0	0	3
gamma-40	3	0	0	3
gamma-41	3	0	0	3
gamma-42	16	5	0	21
gamma-43	2	1	0	3
gamma-44	2	0	0	2
gamma-45	10	4	2	16
gamma-46	3	0	0	3
gamma-47	2	0	0	2
gamma-48	2	2	0	4
gamma-49	6	3	1	10
gamma-50	3	0	0	3
gamma-51	2	0	0	2
gamma-52	1	0	0	1
gamma-53	3	4	0	7
gamma-54	2	2	0	4
gamma-55	2	0	0	2
gamma-56	5	4	0	9
gamma-57	16	10	0	26
gamma-58	6	2	3	11
gamma-59	2	0	0	2
gamma-60	2	0	0	2
gamma-61	1	0	0	1

gamma-62	2	0	0	2
gamma-63	3	0	0	3
gamma-64	2	1	0	3
gamma-65	2	1	0	3
gamma-66	2	4	0	6
gamma-67	3	0	0	3
gamma-68	3	4	0	7
gamma-69	1	0	0	1
gamma-70	3	0	0	3
gamma-71	3	0	0	3
gamma-72	3	0	0	3
gamma-73	1	0	0	1
gamma-74	2	8	0	10
gamma-75	2	1	0	3
gamma-76	3	2	0	5
gamma-77	1	0	0	1
gamma-78	3	0	0	3
gamma-79	13	4	3	20
gamma-80	2	0	0	2
gamma-81	8	3	1	12
gamma-82	4	0	0	4
gamma-83	3	0	0	3
gamma-84	3	0	0	3
gamma-85	3	0	0	3
gamma-86	3	1	0	4
gamma-87	3	0	0	3
gamma-88	3	0	0	3
gamma-89	3	1	1	5
gamma-90	2	1	0	3
gamma-91	3	0	0	3
gamma-92	3	1	0	4
gamma-93	3	7	0	10
gamma-94	2	0	0	2
gamma-95	10	5	0	15
gamma-96	2	0	0	2
gamma-97	4	1	0	5
gamma-98	4	1	0	5
gamma-99	4	3	0	7
gamma-100	5	0	0	5
gamma-101	5	2	0	7
gamma-102	3	0	0	3
gamma-103	3	1	0	4
gamma-104	3	0	0	3
gamma-105	4	2	0	6
gamma-106	2	0	0	2
gamma-107	4	0	0	4
gamma-108	2	0	0	2

gamma-109	5	3	0	8
gamma-110	3	0	0	3
gamma-111	3	0	0	3
gamma-112	3	1	0	4
gamma-113	3	3	0	6
gamma-114	23	4	1	28
gamma-115	7	2	0	9
gamma-116	3	0	0	3
gamma-117	3	2	0	5
gamma-118	6	3	0	9
gamma-119	7	1	0	8
gamma-120	4	0	0	4
gamma-121	6	1	0	7
gamma-122	2	0	0	2
gamma-123	4	0	0	4
gamma-124	2	0	0	2
gamma-125	6	0	2	8
gamma-126	3	0	0	3
gamma-127	1	1	0	2
gamma-128	2	3	0	5
gamma-129	1	0	0	1
gamma-130	6	0	0	6
gamma-131	4	1	0	5
gamma-132	5	2	1	8
gamma-133	1	0	0	1
gamma-134	5	1	0	6
gamma-135	2	0	0	2
gamma-136	3	0	0	3
gamma-137	6	2	0	8
gamma-138	8	0	0	8
gamma-139	1	0	0	1
gamma-140	5	6	1	12
gamma-141	1	0	0	1
gamma-142	1	3	0	4
gamma-143	2	1	0	3
gamma-144	3	2	0	5
gamma-145	3	1	0	4
gamma-146	3	1	0	4
gamma-147	3	2	1	6
gamma-148	3	0	0	3
gamma-149	1	0	0	1
gamma-150	1	0	0	1
gamma-151	2	0	0	2
gamma-152	4	6	0	10
gamma-153	3	2	0	5
gamma-154	3	0	0	3
gamma-155	3	1	0	4

gamma-156	4	0	0	4
gamma-157	4	2	0	6
gamma-158	3	1	0	4
gamma-159	3	20	0	23
gamma-160	4	1	0	5
gamma-161	3	8	0	11
gamma-162	3	6	0	9
gamma-163	5	37	1	43
gamma-164	2	13	0	15
gamma-165	2	0	0	2
gamma-166	4	1	0	5
gamma-167	1	1	0	2
gamma-168	3	3	0	6
gamma-169	3	0	0	3
gamma-170	3	0	0	3
gamma-171	6	1	0	7
gamma-172	3	0	0	3
gamma-173	1	0	0	1
gamma-174	3	5	0	8
gamma-175	3	0	0	3
gamma-176	3	1	0	4
gamma-177	3	0	0	3
gamma-178	2	0	0	2
gamma-179	7	10	1	18
gamma-180	8	3	0	11
gamma-181	5	0	0	5
gamma-182	6	0	0	6
gamma-183	4	3	0	7
gamma-184	1	0	0	1
gamma-185	1	0	0	1
gamma-186	3	1	0	4
gamma-187	2	1	0	3
gamma-188	2	1	0	3
gamma-189	4	2	0	6
gamma-190	5	1	0	6
gamma-191	2	0	0	2
gamma-192	3	0	0	3
gamma-193	4	1	0	5
gamma-194	3	2	0	5
gamma-195	2	0	0	2
gamma-196	2	0	0	2
gamma-197	2	0	0	2
gamma-198	3	4	0	7
gamma-199	3	0	0	3
gamma-200	5	2	0	7
gamma-201	3	0	0	3
gamma-202	3	0	0	3

gamma-203	4	0	0	4
gamma-204	3	3	0	6
gamma-205	3	1	0	4
gamma-206	3	1	0	4
gamma-207	4	0	0	4
gamma-208	3	0	0	3
gamma-209	3	0	0	3
gamma-210	2	1	0	3
gamma-211	7	4	0	11
gamma-212	3	1	0	4
gamma-213	1	0	0	1
gamma-214	2	1	0	3
gamma-215	6	6	0	12
gamma-216	3	1	0	4
gamma-217	2	0	0	2
gamma-218	3	0	0	3
gamma-219	1	1	0	2
gamma-220	9	0	0	9
gamma-221	3	2	0	5
gamma-222	3	0	0	3
gamma-223	2	0	0	2
gamma-224	4	0	0	4
gamma-225	4	2	0	6
gamma-226	8	0	0	8
gamma-227	3	0	0	3
gamma-228	5	2	0	7
gamma-229	3	1	0	4
gamma-230	1	0	0	1
gamma-231	4	1	0	5
gamma-232	4	0	0	4
gamma-233	1	0	0	1
gamma-234	78	20	7	105
gamma-235	2	0	0	2
gamma-236	2	0	0	2
gamma-237	3	0	0	3
gamma-238	2	0	0	2
gamma-239	3	1	1	5
gamma-240	4	0	0	4
gamma-241	3	0	0	3
gamma-242	1	1	0	2
gamma-243	7	4	0	11
gamma-244	2	1	0	3
gamma-245	6	1	0	7
gamma-246	3	0	0	3
gamma-247	3	2	0	5
gamma-248	6	3	0	9
gamma-249	7	0	0	7

gamma-250	3	0	0	3
gamma-251	4	0	0	4
gamma-252	2	3	0	5
gamma-253	5	1	0	6
gamma-254	3	2	0	5
gamma-255	3	0	0	3
gamma-256	4	6	0	10
gamma-257	5	2	0	7
gamma-258	5	2	0	7
gamma-259	8	3	0	11
gamma-260	2	0	0	2
gamma-261	2	0	0	2
gamma-262	3	1	0	4
gamma-263	1	0	0	1
gamma-264	2	1	0	3
gamma-265	1	0	0	1
gamma-266	3	2	0	5
gamma-267	2	1	0	3
gamma-268	3	0	0	3
gamma-269	2	0	0	2
gamma-270	3	0	0	3
gamma-271	2	0	0	2
gamma-272	2	0	0	2
gamma-273	55	11	6	72
gamma-274	1	0	0	1
gamma-275	3	0	0	3
gamma-276	5	0	0	5
gamma-277	3	0	1	4
gamma-278	23	2	2	27
gamma-279	6	0	1	7
gamma-280	9	1	2	12
gamma-281	3	1	0	4
gamma-282	2	0	0	2
gamma-283	2	0	0	2
gamma-284	7	1	1	9
gamma-285	2	0	0	2
gamma-286	9	0	0	9
gamma-287	3	0	0	3
gamma-288	5	0	0	5
gamma-289	3	1	0	4
gamma-290	3	4	0	7
gamma-291	2	1	0	3
gamma-292	2	0	0	2
gamma-293	3	0	0	3
gamma-294	4	0	0	4
gamma-295	2	0	0	2
gamma-296	3	0	1	4

gamma-297	4	2	0	6
gamma-298	3	0	0	3
gamma-299	2	0	0	2
gamma-300	12	2	1	15
gamma-301	2	0	0	2
gamma-302	3	0	0	3
gamma-303	2	0	0	2
gamma-304	2	0	0	2
gamma-305	2	0	0	2
gamma-306	4	0	0	4
gamma-307	5	3	1	9
gamma-308	6	1	0	7
gamma-309	3	0	0	3
gamma-310	2	0	0	2
gamma-311	4	3	0	7
gamma-312	3	1	0	4
gamma-313	5	2	0	7
gamma-314	2	0	0	2
gamma-315	1	0	0	1
gamma-316	3	2	0	5
gamma-317	3	0	0	3
gamma-318	3	0	0	3
gamma-319	1	0	0	1
gamma-320	7	2	0	9
gamma-321	3	0	0	3
gamma-322	4	2	0	6
gamma-323	3	2	0	5
gamma-324	3	1	1	5
gamma-325	2	0	0	2
gamma-326	5	0	1	6
gamma-327	3	0	0	3
gamma-328	4	5	0	9
gamma-329	3	1	0	4
gamma-330	3	1	0	4
gamma-331	2	2	0	4
gamma-332	3	0	0	3
gamma-333	2	0	0	2
gamma-334	3	1	0	4
gamma-335	3	1	0	4
gamma-336	5	3	0	8
gamma-337	4	3	0	7
gamma-338	5	2	0	7
gamma-339	3	0	0	3
gamma-340	5	6	0	11
gamma-341	3	0	0	3
gamma-342	6	3	0	9
gamma-343	1	0	0	1

gamma-344	4	2	1	7
gamma-345	6	4	0	10
gamma-346	4	1	0	5
gamma-347	2	0	0	2
gamma-348	8	2	1	11
gamma-349	5	2	0	7
gamma-350	101	19	2	122
gamma-351	2	0	0	2
gamma-352	3	0	0	3
gamma-353	3	1	0	4
gamma-354	6	2	1	9
gamma-355	2	0	0	2
gamma-356	4	1	0	5
gamma-357	5	1	0	6
gamma-358	4	2	0	6
gamma-359	3	2	0	5
gamma-360	8	2	0	10
gamma-361	2	0	0	2
gamma-362	3	0	0	3
gamma-363	4	0	0	4
gamma-364	3	2	0	5
gamma-365	4	3	0	7
gamma-366	12	4	0	16
gamma-367	3	0	0	3
gamma-368	3	2	0	5
gamma-369	1	0	0	1
gamma-370	3	2	0	5
gamma-371	3	4	0	7
gamma-372	2	0	0	2
gamma-373	2	1	0	3
gamma-374	4	1	0	5
gamma-375	7	2	0	9
gamma-376	13	2	0	15
gamma-377	11	3	2	16
gamma-378	17	2	0	19
gamma-379	1	0	0	1
gamma-380	3	1	0	4
gamma-381	1	0	0	1
gamma-382	1	0	0	1
gamma-383	2	0	0	2
gamma-384	3	0	0	3
gamma-385	1	0	0	1
gamma-386	2	0	0	2
gamma-387	2	0	0	2
gamma-388	3	1	1	5
gamma-389	1	0	0	1
gamma-390	1	0	0	1

gamma-391	4	0	1	5
gamma-392	6	4	0	10
gamma-393	2	0	0	2
gamma-394	4	0	0	4
gamma-395	5	1	0	6
gamma-396	7	0	0	7
gamma-397	19	8	0	27
gamma-398	3	0	0	3
gamma-399	2	0	0	2
gamma-400	1	0	0	1
gamma-401	2	1	0	3
gamma-402	2	0	0	2
gamma-403	2	0	0	2
gamma-404	2	1	0	3
gamma-405	2	1	0	3
gamma-406	2	0	0	2
gamma-407	3	2	0	5
gamma-408	1	0	0	1
gamma-409	4	2	0	6
gamma-410	3	2	0	5
gamma-411	3	0	0	3
gamma-412	3	0	2	5
gamma-413	3	2	0	5
gamma-414	3	0	0	3
gamma-415	3	2	0	5
gamma-416	2	1	0	3
gamma-417	3	0	0	3
gamma-418	1	0	0	1
gamma-419	1	1	0	2
gamma-420	3	1	0	4
gamma-421	2	0	0	2
gamma-422	1	0	0	1
gamma-423	1	0	0	1
gamma-424	3	0	0	3
gamma-425	2	2	0	4
gamma-426	5	0	0	5
gamma-427	10	1	0	11
gamma-428	10	0	0	10
gamma-429	8	4	0	12
gamma-430	5	1	0	6
gamma-431	1	0	0	1
gamma-432	50	12	0	62
gamma-433	4	1	1	6
gamma-434	1	0	0	1
gamma-435	1	0	0	1
gamma-436	3	0	0	3
gamma-437	4	0	0	4

gamma-438	4	1	0	5
gamma-439	1	0	0	1
gamma-440	2	1	0	3
gamma-441	1	0	0	1
gamma-442	1	0	0	1
gamma-443	2	0	0	2
gamma-444	3	0	0	3
gamma-445	11	5	0	16
gamma-446	5	0	0	5
gamma-447	5	0	0	5
gamma-448	3	0	0	3
gamma-449	2	3	0	5
gamma-450	4	1	0	5
gamma-451	1	1	0	2
gamma-452	1	0	0	1
gamma-453	2	0	0	2
gamma-454	2	2	1	5
gamma-455	1	1	0	2
gamma-456	3	0	0	3
gamma-457	2	0	0	2

Table 2S. Number of clusters and singletons per chromosome in *Xenopus*.

<i>Xenopus tropicalis</i>			
GenBank Id	Chromosome number	No of OR clusters	Singletons
CM004443	1	6	1
CM004444	2	6	1
CM004445	3	8	1
CM004446	4	2	1
CM004447	5	7	2
CM004448	-	-	-
CM004449	7	5	1
CM004450	8	10	1
CM004451	9	3	1
CM004452	10	2	0
<i>Xenopus laevis</i>			
GenBank Id	Chromosome number	No of OR clusters	Singletons
CM030340	1L	5	0
CM030341	1S	4	0
CM030342	2L	7	1
CM030343	2S	3	1
CM030344	3L	7	1
CM030345	3S	6	1
CM030346	4L	1	0
CM030347	4S	1	1

CM030348	5L	4	2
CM030349	5S	2	1
CM030350	6L	-	-
CM030351	6S	-	-
CM030352	7L	5	1
CM030353	7S	4	3
CM030354	8L	9	1
CM030355	8S	4	0
CM030356	9_10L	7	1
CM030357	9_10S	3	0
<i>Xenopus borealis</i>			
GenBank Id	Chromosome number	No of OR clusters	Singletons
CM044431	1L	6	0
CM044432	1S	4	0
CM044433	2L	3	0
CM044434	2S	5	2
CM044435	3L	6	0
CM044436	3S	5	2
CM044437	4L	1	0
CM044438	4S	1	1
CM044439	5L	3	2
CM044440	5S	2	1
CM044441	6L	-	-
CM044442	6S	-	-
CM044443	7L	5	1
CM044444	7S	3	2
CM044445	8L	7	0
CM044446	8S	4	0
CM044447	9_10L	6	1
CM044448	9_10S	3	0

Table 3S. Number of intact, pseudogene and truncated genes per cluster in *Xenopus tropicalis*.

Cluster	Intact	Pseudogene	Truncated	Total
Xtro1.1	0	1	0	1
Xtro1.2	2	0	0	2
Xtro1.3	1	1	0	2
Xtro1.4	8	3	0	11
Xtro1.5	33	8	0	41
Xtro1.6	6	3	0	9
Xtro2.1	3	0	0	3
Xtro2.2	1	4	0	5
Xtro2.3	11	1	0	12
Xtro2.4	57	5	0	62
Xtro2.5	26	10	0	36

Xtro2.6	0	1	0	1
Xtro3.1	0	1	0	1
Xtro3.2	54	21	0	75
Xtro3.3	28	4	0	32
Xtro3.4	4	0	0	4
Xtro3.5	2	0	0	2
Xtro3.6	233	45	0	278
Xtro3.7	3	1	0	4
Xtro3.8	28	0	0	28
Xtro4.1	0	1	0	1
Xtro4.2	6	2	0	8
Xtro5.1	0	1	0	1
Xtro5.2	1	0	0	1
Xtro5.3	1	2	0	3
Xtro5.4	2	1	0	3
Xtro5.5	3	0	0	3
Xtro5.6	1	2	0	3
Xtro5.7	3	0	0	3
Xtro7.1	199	54	0	253
Xtro7.2	1	0	0	1
Xtro7.3	2	0	0	2
Xtro7.4	5	2	0	7
Xtro7.5	5	1	0	6
Xtro8.1	6	3	0	9
Xtro8.2	1	0	0	1
Xtro8.3	6	0	0	6
Xtro8.4	4	0	0	4
Xtro8.5	29	3	0	32
Xtro8.6	33	8	0	41
Xtro8.7	0	8	0	8
Xtro8.8	11	29	0	40
Xtro8.9	2	1	0	3
Xtro8.10	35	5	0	40
Xtro9.1	84	14	0	98
Xtro9.2	40	4	0	44
Xtro9.3	1	0	0	1
Xtro10.1	5	0	0	5
Xtro10.2	3	0	0	3

Table 4S. Number of intact, pseudogene and truncated genes per cluster in *Xenopus laevis*.

Cluster	Intact	Pseudogene	Truncated	Total
Xla1L.1	2	1	0	3
Xla1L.2	2	0	0	2
Xla1L.3	9	8	0	17

Xla1L.4	23	7	0	30
Xla1L.5	6	5	0	11
Xla1S.1	0	2	0	2
Xla1S.2	3	0	0	3
Xla1S.3	7	2	0	9
Xla1S.4	3	2	0	5
Xla2L.1	3	1	0	4
Xla2L.2	4	1	0	5
Xla2L.3	28	19	0	47
Xla2L.4	47	7	0	54
Xla2L.5	3	2	0	5
Xla2L.6	1	3	0	4
Xla2L.7	1	0	0	1
Xla2S.1	0	1	0	1
Xla2S.2	26	9	0	35
Xla2S.3	6	2	0	8
Xla3L.1	3	2	0	5
Xla3L.2	25	25	0	50
Xla3L.3	25	6	0	31
Xla3L.4	1	1	0	2
Xla3L.5	106	38	0	144
Xla3L.6	0	1	0	1
Xla3L.7	57	7	0	64
Xla3S.1	0	1	0	1
Xla3S.2	6	3	0	9
Xla3S.3	11	4	0	15
Xla3S.4	1	2	0	3
Xla3S.5	55	21	0	76
Xla3S.6	10	0	0	10
Xla4L.1	4	2	0	6
Xla4S.1	1	0	0	1
Xla5L.1	1	2	0	3
Xla5L.2	3	2	0	5
Xla5L.3	1	0	0	1
Xla5L.4	1	0	0	1
Xla5S.1	0	1	0	1
Xla5S.2	2	1	0	3
Xla7L.1	24	15	0	39
Xla7L.2	0	1	0	1
Xla7L.3	1	1	0	2
Xla7L.4	7	4	0	11
Xla7L.5	4	2	0	6
Xla7S.1	19	7	0	26
Xla7S.2	1	0	0	1
Xla7S.3	0	1	0	1
Xla7S.4	0	1	0	1
Xla8L.1	0	3	0	3

Xla8L.2	1	0	0	1
Xla8L.3	1	1	0	2
Xla8L.4	2	0	0	2
Xla8L.5	5	0	0	5
Xla8L.6	10	2	0	12
Xla8L.7	25	12	0	37
Xla8L.8	6	23	0	29
Xla8L.9	30	7	0	37
Xla8S.1	3	1	0	4
Xla8S.2	4	1	0	5
Xla8S.3	15	3	0	18
Xla8S.4	3	0	0	3
Xla9_10L.1	2	1	0	3
Xla9_10L.2	2	1	0	3
Xla9_10L.3	4	3	0	7
Xla9_10L.4	1	0	0	1
Xla9_10L.5	13	5	0	18
Xla9_10L.6	45	11	0	56
Xla9_10L.7	10	4	0	14
Xla9_10S.1	1	1	0	2
Xla9_10S.2	8	4	0	12
Xla9_10S.3	4	4	0	8

Table 5S. Number of intact, pseudogene and truncated genes per cluster in *Xenopus borealis*.

Cluster	Intact	Pseudogene	Truncated	Total
Xbo1L.1	2	0	0	2
Xbo1L.2	1	0	1	2
Xbo1L.3	3	0	0	3
Xbo1L.4	4	0	0	4
Xbo1L.5	27	8	0	35
Xbo1L.6	7	2	1	10
Xbo1S.1	1	1	0	2
Xbo1S.2	3	0	0	3
Xbo1S.3	9	3	0	12
Xbo1S.4	3	0	0	3
Xbo2L.1	3	0	0	3
Xbo2L.2	5	0	0	5
Xbo2L.3	46	14	9	69
Xbo2S.1	0	1	0	1
Xbo2S.2	0	1	0	1
Xbo2S.3	4	3	0	7
Xbo2S.4	20	5	0	25
Xbo2S.5	5	2	0	7
Xbo3L.1	4	1	1	6
Xbo3L.2	25	9	2	36
Xbo3L.3	19	5	2	26

Xbo3L.4	2	1	0	3
Xbo3L.5	75	17	15	107
Xbo3L.6	18	6	3	27
Xbo3S.1	9	4	0	13
Xbo3S.2	1	0	0	1
Xbo3S.3	9	2	3	14
Xbo3S.4	1	0	0	1
Xbo3S.5	45	13	3	61
Xbo4L.1	6	1	0	7
Xbo4S.1	1	0	0	1
Xbo5L.1	0	1	0	1
Xbo5L.2	3	4	1	8
Xbo5L.3	0	1	0	1
Xbo5S.1	1	0	0	1
Xbo5S.2	3	1	0	4
Xbo7L.1	28	3	4	35
Xbo7L.2	0	1	0	1
Xbo7L.3	1	2	0	3
Xbo7L.4	8	0	1	9
Xbo7L.5	3	3	0	6
Xbo7S.1	10	1	1	12
Xbo7S.2	1	0	0	1
Xbo7S.3	1	0	0	1
Xbo8L.1	0	0	3	3
Xbo8L.2	3	0	0	3
Xbo8L.3	5	0	0	5
Xbo8L.4	15	2	0	17
Xbo8L.5	20	8	1	29
Xbo8L.6	7	22	1	30
Xbo8L.7	24	5	3	32
Xbo8S.1	5	0	0	5
Xbo8S.2	2	2	0	4
Xbo8S.3	8	2	0	10
Xbo8S.4	2	1	0	3
Xbo9_10L.1	1	2	0	3
Xbo9_10L.2	1	0	2	3
Xbo9_10L.3	2	3	1	6
Xbo9_10L.4	1	0	0	1
Xbo9_10L.5	13	4	2	19
Xbo9_10L.6	42	12	0	54
Xbo9_10S.1	2	0	0	2
Xbo9_10S.2	4	1	4	9
Xbo9_10S.3	4	1	1	6

Table 6S. Number of OR genes in each chromosome in *Xenopus laevis* and *Xenopus borealis*.

GenBank Id	Chromosome number	Intact	Pseudogene	Truncated	Intact %	Pseudo %	Total
CM004443	1	50	16	-	75,8	24,2	66
CM004444	2	98	21	-	82,4	17,6	119
CM004445	3	352	72	-	83,0	17,0	424
CM004446	4	6	3	-	66,7	33,3	9
CM004447	5	11	6	-	64,7	35,3	17
CM004448	-	-	-	-	-	-	-
CM004449	7	212	57	-	78,8	21,2	269
CM004450	8	127	57	-	69,0	31,0	184
CM004451	9	125	18	-	87,4	12,6	143
CM004452	10	8	0	-	100,0	0,0	8

**TOTAL 1239**

*Xenopus laevis*

GenBank Id	Chromosome number	Intact	Pseudogene	Truncated	Intact %	Pseudo %	Total
CM030340	1L	42	21	-	66,7	33,3	63
CM030341	1S	13	6	-	68,4	31,6	19
CM030342	2L	87	33	-	72,5	27,5	120
CM030343	2S	32	12	-	72,7	27,3	44
CM030344	3L	217	80	-	73,1	26,9	297
CM030345	3S	83	31	-	72,8	27,2	114
CM030346	4L	4	2	-	66,7	33,3	6
CM030347	4S	1	0	-	100,0	0,0	1
CM030348	5L	6	4	-	60,0	40,0	10
CM030349	5S	2	2	-	50,0	50,0	4
CM030350	6L	-	-	-	-	-	-
CM030351	6S	-	-	-	-	-	-
CM030352	7L	36	23	-	61,0	39,0	59
CM030353	7S	20	9	-	69,0	31,0	29
CM030354	8L	80	48	-	62,5	37,5	128
CM030355	8S	25	5	-	83,3	16,7	30
CM030356	9_10L	77	25	-	75,5	24,5	102
CM030357	9_10S	13	9	-	59,1	40,9	22

4.3 **Capítulo III** – Análise genômica comparativa do repertório de genes OR de 28 espécies de anuros: novos achados, lições e desafios no estudo desta família multigênica.

#### 4.3.1 Relação do repertório com a proximidade filogenética das espécies

Das 28 espécies de anuros com o repertório de genes OR identificado, temos as seguintes espécies por família: Bombinatoridae (*B. variegata variegata*), Scaphiropodidae (*S. couchii*, *S. holbrookii*, *S. bombifrons*, *S. multiplicata*), Megophryidae (*L. ailaonicum*, *L. leishanense*), Dendrobatidae (*O. pumilio*), Bufonidae (*R. marina*, *B. bufo*, *B. gargarizans*), Leptodactylidae (*E. pustulosus*), Hylidae (*P. bahiana*), Eleutherodactylidae (*E. coqui*), Limnodynastidae (*L. dumerilii*, *P. ornatum*), Pyxicephalidae (*P. adspersus*), Ranidae (*L. catesbeianus*, *R. temporaria*, *G. rugosa*), Dicroglossidae (*P. myanhessei*, *N. parkeri*), Pipidae (*H. boettgeri*, *P. carvalhoi*, *P. parva*, *X. borealis*, *X. laevis*, *X. tropicalis*). Cada árvore filogenética gerada pelo IQ-TREE contendo somente os genes intactos de cada uma das espécies está nas Figuras 5–8.

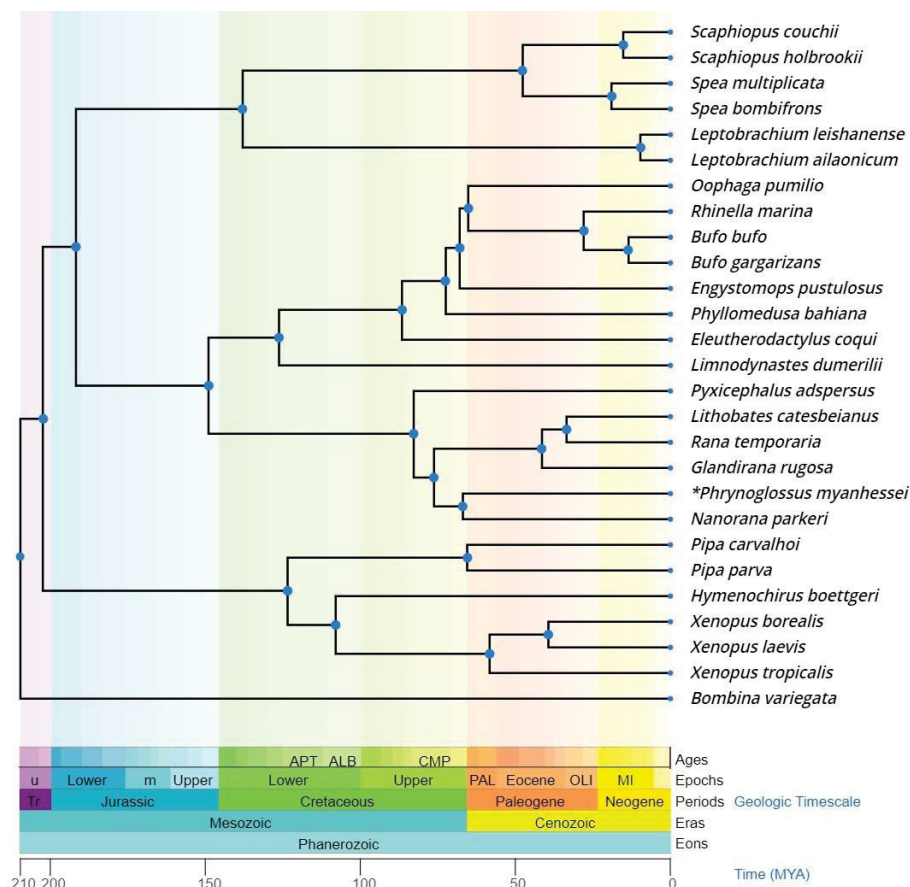


Figura 4. Relacionamento filogenético entre as 28 espécies de anuros analisadas. Dendrograma estimado na plataforma TimeTree: A Resource for Timelines, Timetrees, and Divergence Times 50 (<http://www.timetree.org/>). A espécie *Platypectrum ornatum* não foi incluída neste dendrograma por ausência de dados recuperados pela plataforma.

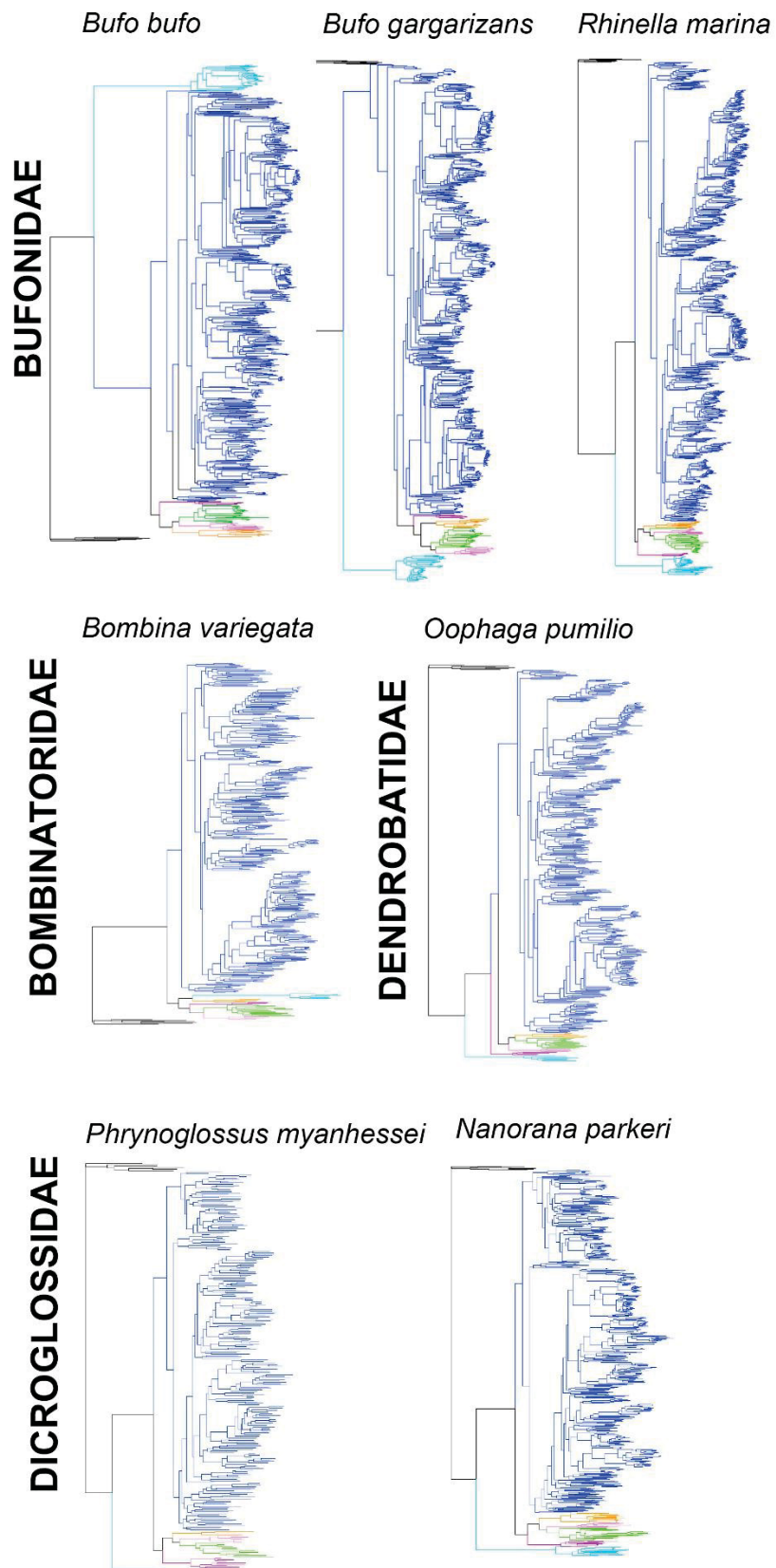


Figura 5. Árvore filogenética inferida pelo IQ-TREE a partir das cópias dos OR genes intactos de espécies das famílias Bufonidae, Bombinatoridae, Dendrobatidae e Dicroglossidae. Cores:  $\alpha$  (alfa) – verde;  $\beta$  (beta) – rosa;  $\gamma$  (gamma) – azul;  $\delta$  (delta) – laranja;  $\epsilon$  (epsilon) – roxo;  $\eta$  (eta) – azul-claro; outgroup – preto.

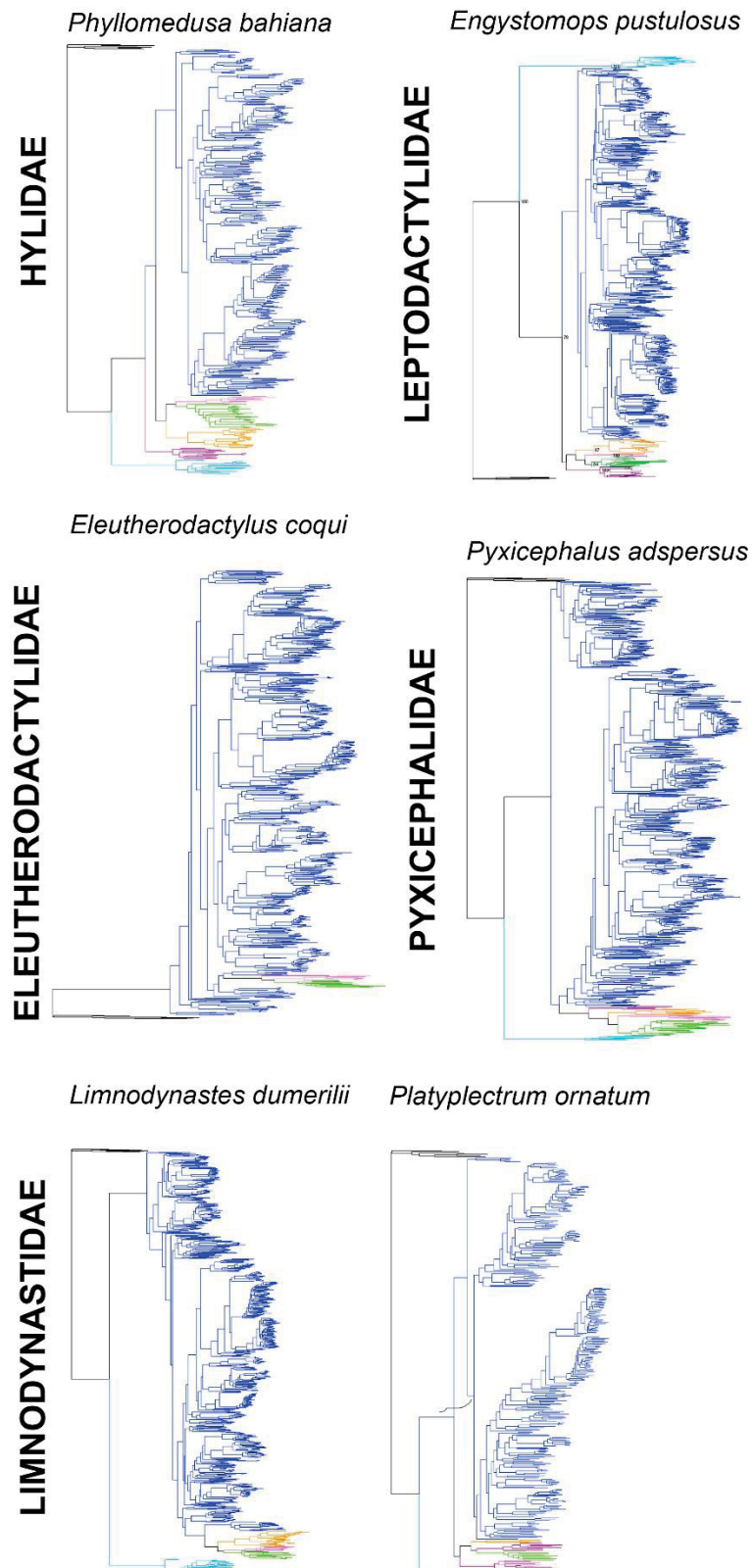


Figura 6. Árvore filogenética inferida pelo IQ-TREE a partir das cópias dos OR genes intactos de espécies das famílias Hylidae, Leptodactylidae, Eleutherodactylidae, Pyxicephalidae e

Limnodynastidae. Cores:  $\alpha$  (alfa) – verde;  $\beta$  (beta) – rosa;  $\gamma$  (gamma) – azul;  $\delta$  (delta) – laranja;  $\epsilon$  (epsilon) – roxo;  $\eta$  (eta) – azul-claro; outgroup – preto.

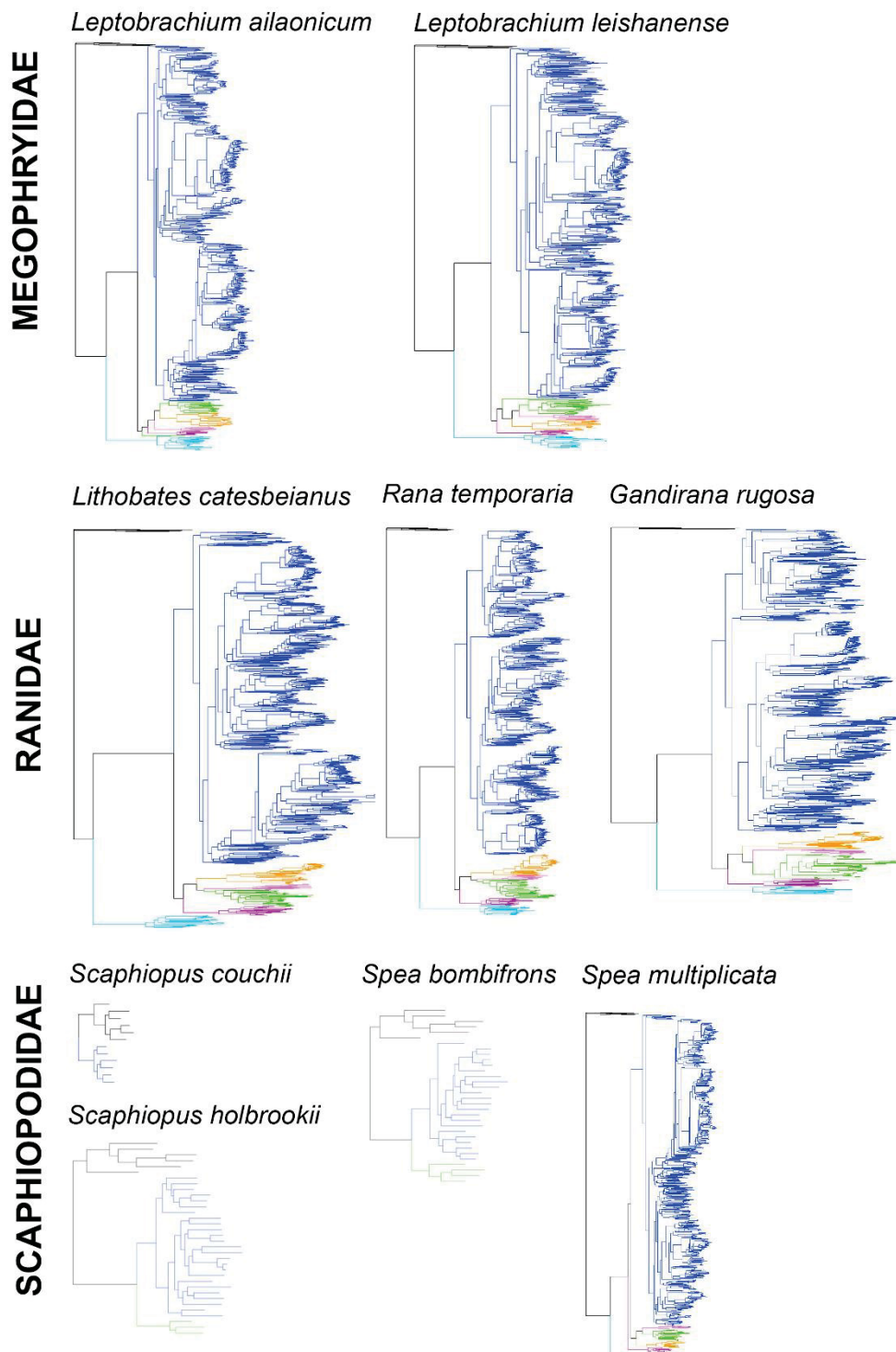


Figura 7. Árvore filogenética inferida pelo IQ-TREE a partir das cópias dos OR genes intactos de espécies das famílias Megophryidae, Ranidae e Scaphiopoidea. Cores:  $\alpha$  (alfa) – verde;  $\beta$

(beta) – rosa;  $\gamma$  (gamma) – azul;  $\delta$  (delta) – laranja;  $\epsilon$  (epsilon) – roxo;  $\eta$  (eta) – azul-claro; outgroup – preto.

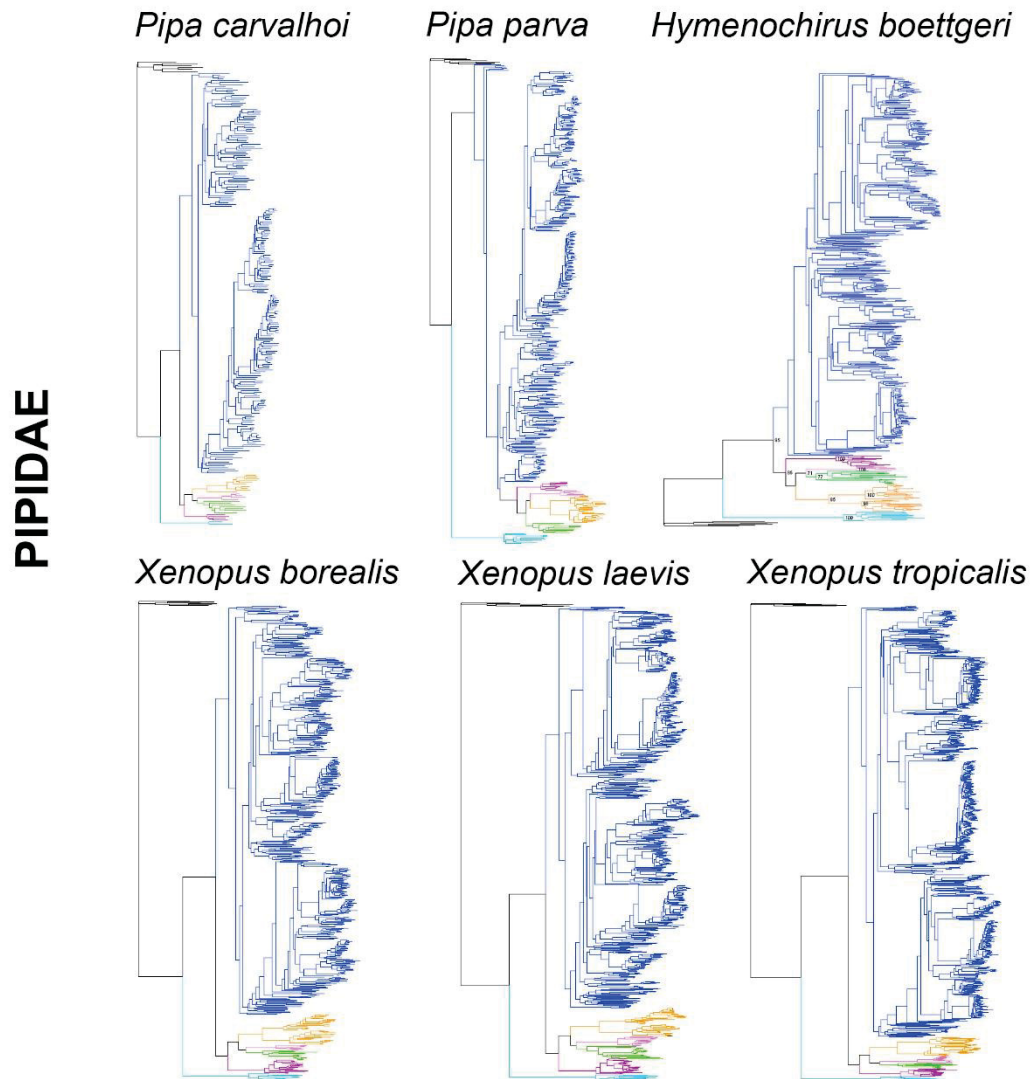


Figura 8. Árvore filogenética inferida pelo IQ-TREE a partir das cópias dos OR genes intactos de espécies da família Pipidae. Cores:  $\alpha$  (alfa) – verde;  $\beta$  (beta) – rosa;  $\gamma$  (gamma) – azul;  $\delta$  (delta) – laranja;  $\epsilon$  (epsilon) – roxo;  $\eta$  (eta) – azul-claro; outgroup – preto.

#### 4.3.2 Total de genes OR nos genomas de anfíbios anuros

A Tabela 4 mostra o número de genes OR de cada uma das 28 espécies de anfíbios anuros analisadas para o qual o genoma sequenciado está disponível. O número de genes funcionais (intactos), genes truncados e pseudogenes é mostrado separadamente.

Foram identificados mais de 28.000 genes OR entre todas as espécies aqui estudadas sendo que 12 espécies contabilizaram um total de mais de mil genes em seu repertório (Figura 5–9). *Engystomops pustulosus* foi a espécie que possuiu mais genes OR intactos, seguido de *Glandirana rugosa*, com 1047 genes intactos e *Lithobates catesbeianus* com 986 (Tabela 4). Todos os grupos filogenéticos foram recuperados para todas as espécies, sejam na presença de intactos, quanto truncados e pseudogenes, com exceção do grupo ζ (zeta), que permanece como um grupo exclusivo dos peixes. Como esperado, espécies na qual o nível de contiguidade e cobertura da montagem do genoma foi bastante baixo, em especial *Scaphiopus couchii* e *S. holbrookii*, apresentaram um número de genes truncados e pseudogenes mais elevado em comparação com os genes intactos (Tabela 4).

Tabela 4. Número de genes OR em 28 genomas de espécies de anfíbios anuros.

Espécie	Intactos	Truncados	Pseudogenes	TOTAL	Pseudo %
<i>Bombina variegata variegata</i>	403	137	151	691	21,85
<i>Bufo bufo</i>	745	0	172	917	18,76
<i>Bufo gargarizans</i>	805	0	226	1031	21,92
<i>Eleutherodactylus coqui</i>	688	75	160	923	17,33
<i>Engystomops pustulosus</i>	1080	77	928	2085	44,51
<i>Glandirana rugosa</i>	1047	70	340	1457	23,34
<i>Hymenochirus boettgeri</i>	507	19	668	1194	55,95
<i>Leptobrachium ailaonicum</i>	684	1	282	967	29,16
<i>Leptobrachium leishanense</i>	754	3	241	998	24,15
<i>Limnodynastes dumerilii</i>	873	47	377	1297	29,07
<i>Lithobates catesbeianus</i>	986	110	415	1511	27,47
<i>Nanorana parkeri</i>	601	4	327	932	35,09
<i>Oophaga pumilio</i>	423	44	132	599	22,04
<i>Phrynoglossus myanhenssei</i>	269	212	136	617	22,04

<i>Phyllomedusa bahiana</i>	491	6	486	983	49,44
<i>Pipa carvalhoi</i>	241	224	182	647	28,13
<i>Pipa parva</i>	440	91	74	605	12,23
<i>Platyplectrum ornatum</i>	336	124	281	741	37,92
<i>Pyxicephalus adspersus</i>	725	126	350	1201	29,14
<i>Rana temporaria</i>	862	0	189	1051	17,98
<i>Rhinella marina</i>	865	5	292	1162	25,13
<i>Scaphiopus couchii</i>	6	198	497	701	70,90
<i>Scaphiopus holbrookii</i>	28	287	480	795	60,38
<i>Spea bombifrons</i>	26	316	452	794	56,93
<i>Spea multiplicata</i>	853	44	204	1101	18,53
<i>Xenopus borealis</i>	591	70	181	842	21,50
<i>Xenopus laevis</i>	738	0	310	1048	29,58
<i>Xenopus tropicalis</i>	990	0	250	1240	20,16

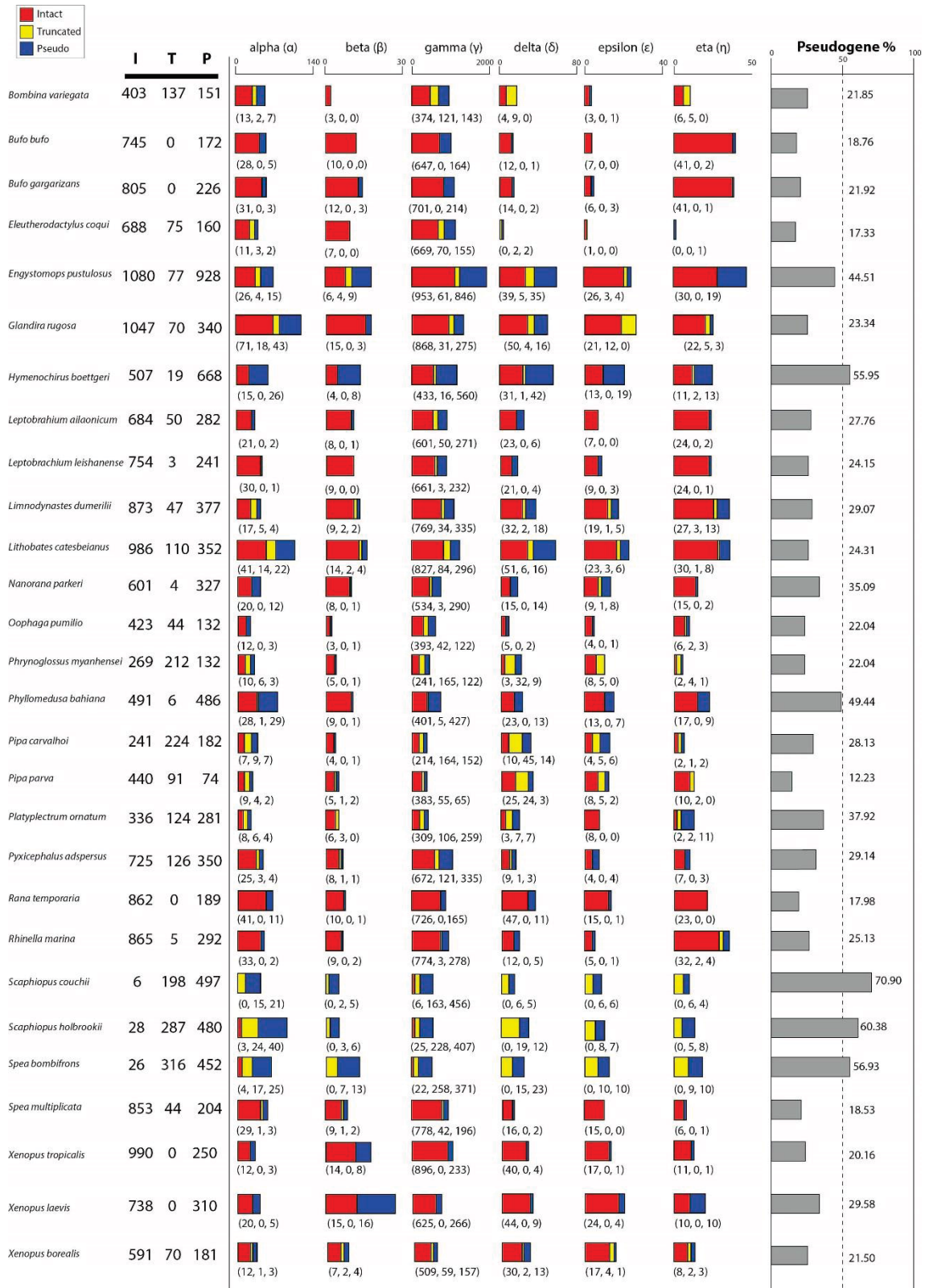


Figura 9. Número de genes OR pertencentes a 28 genomas de anfíbios. Barras vermelhas, amarelas e azuis representam genes intactos, truncados e pseudogenes, respectivamente. O número em parênteses abaixo de cada barra a quantidade dos genes intactos, truncados e

pseudogenes, respectivamente. A porcentagem de pseudogenes em comparação com o total de genes OR no repertório é indicado por barras cinzas.

#### 4.3.3. *Presença de pseudogenes e a relação com a quantidade de cópias intactas*

A presença de pseudogenes OR nos genomas é um resquício evolutivo e possivelmente confirmação de que ocorreram diversas duplicações e degeneração de cópias que permitiram os diversos organismos de se adaptar a novos ambientes e detectar uma variedade de odores. Desse modo, alguns pseudogenes podem ter se formado como resultado desse processo, resultantes do acúmulo de mutações (inserções, deleções, substituições de nucleotídeos) em suas sequências nucleotídicas, de deleções via recombinação, ou rearranjos cromossômicos.

Os genomas de anfíbios apresentaram uma média de 25% de pseudogenes em comparação com o repertório total de genes OR. A alta quantidade de pseudogenes em determinados grupos filogenéticos é sempre esperada, principalmente para aqueles em que pode ser observada uma alta razão da expansão de número de cópias, como os genes do grupo  $\gamma$  (gamma), o grupo que mais apresenta variabilidade de sequências em todas as espécies de tetrápodes terrestres com repertório de ORs já descrito.

Quando observamos espécies com relações filogenéticas mais próximas, como as três espécies do gênero *Xenopus* aqui estudadas, observamos uma baixa quantidade de pseudogenes em comparação com os genes intactos, que pode ser indicativa de pressões evolutivas semelhantes, apesar de *Xenopus laevis* e *Xenopus borealis* serem espécies que passaram pelo processo de aloploidização (Capítulo II).

*Engystomops pustulosus* apresentou uma maior quantidade de pseudogenes em relação as outras espécies, apesar da proporção entre a quantidade total de genes *versus* pseudogenes ter sido maior em *Scaphiopus couchii*, o que pode ser associado ao nível de montagem deste genoma. Alternativamente, nossos dados de mapeamento cromossômico das cópias de ORs no genoma de *E. pustulosus* revelou um alto nível de dispersão de cópias entre os cromossomos (ver Capítulo I), o que pode sugerir impacto de rearranjos intercromossômicos ou da dispersão via elementos genéticos móveis como mecanismos genéticos promotores deste alto nível de pseudogenização observado nesta espécie.

#### 4.3.4 Presença de truncados

A presença de cópias de ORs truncados esteve principalmente relacionada ao nível de montagem do genoma, pois uma vez que genomas com *contigs* mais curtos são mais propensos a apresentar fragmentos de genes que podem ser interpretados como cópias truncadas. Isso foi observado nos genomas de *S. couchii*, *S. holbrooki* e *Spea multiplicata*, todos em nível de montagem de *contigs*. Não só a montagem, como também o nível de cobertura parece interferir na capacidade de recuperar as cópias intactas desses genes repetitivos. Desse modo, à medida que novas versões do genoma são lançadas e que preenchem *gaps* anteriores, esse número é esperado diminuir ou mesmo chegar a zero (referências que mostrem essa redução, terminando em nosso capítulo II), ao exemplo das novas versões de montagem do genoma de *Xenopus tropicalis* e *X. laevis*.

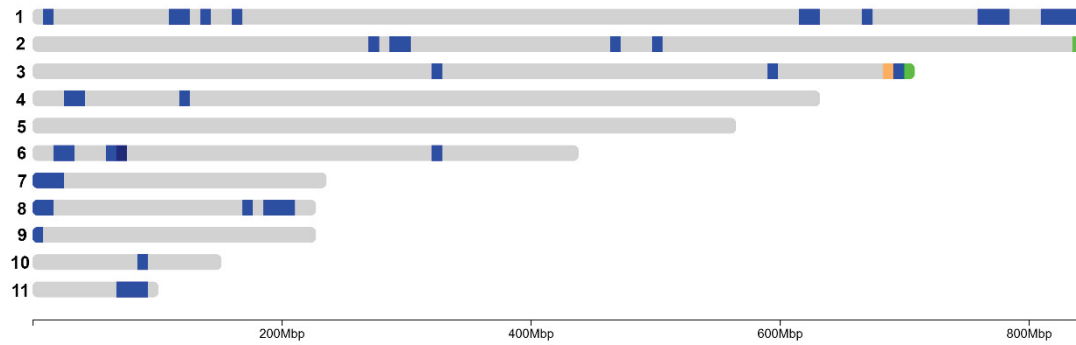
#### 4.3.5 Distribuição cromossômica dos genes OR

Das 28 espécies de anfíbios estudadas, 12 possuem o nível cromossômico de montagem do genoma. Desse modo, as Figuras 10–12 mostram a distribuição física dos genes nos cromossomos dessas espécies. *Clusters* de genes  $\gamma$ , assim como *singletons*, estão espalhados em quase todos os cromossomos de todas as espécies. A média de comprimento de *clusters* é de 2.9Mb. Baseado na definição de *cluster* que nossa metodologia estabeleceu, excluindo *singletons*, houve uma média de 16 *clusters* (Tabela 5). Além disso, não foram observados genes intactos intercalados com pseudogenes no mesmo *cluster*, ou seja, uma vez estabelecido o *cluster*, este era composto somente ou de genes intactos ou pseudogenes (Tabela 6).

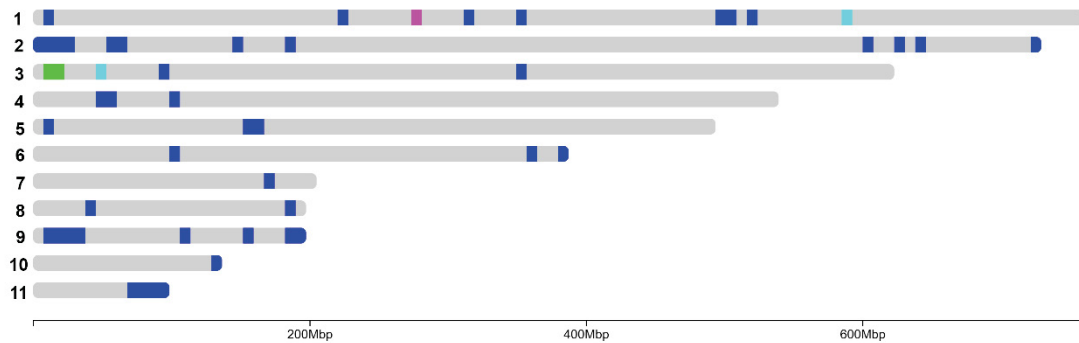
Tabela 5. Quantidade total de clusters encontrados nos cromossomos das 12 espécies aqui identificadas.

<b>Species</b>	<b>Cariótipo</b>	<b>Clusters</b>
<i>Bufo bufo</i>	2n=22	21
<i>Bufo gargarizans</i>	2n=22	23
<i>Eleutherodactylus coqui</i>	2n=26	2
<i>Engystomops pustulosus</i>	2n=22	12
<i>Hymenochirus boettgeri</i>	2n=18	21
<i>Leptobrachium ailaonicum</i>	2n=26	24
<i>Leptobrachium leishanense</i>	2n=26	16
<i>Pyxicephalus adspersus</i>	2n=26	25
<i>Rana temporaria</i>	2n=26	20
<i>Xenopus borealis</i>	2n=18	8
<i>Xenopus laevis</i>	2n=18	12
<i>Xenopus tropicalis</i>	2n=20	12

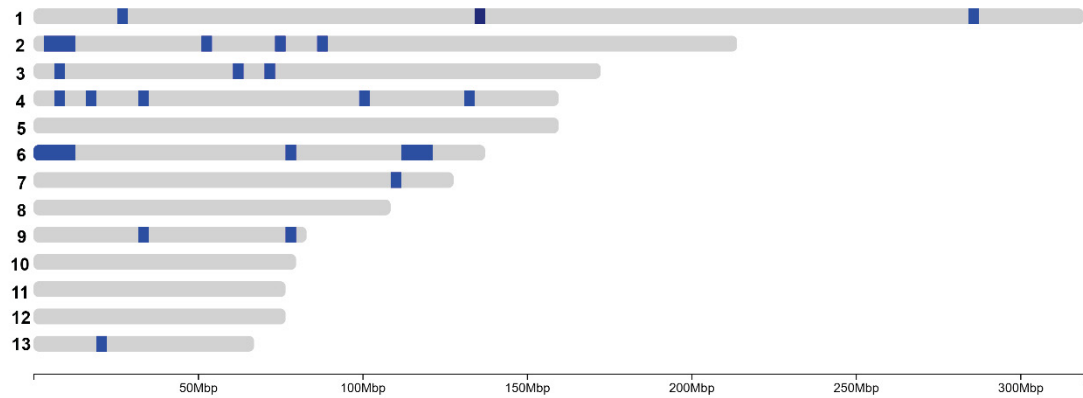
## *Bufo bufo*



## *Bufo gargarizans*



## *Eleutherodactylus coqui*



## *Engystomops pustulosus*

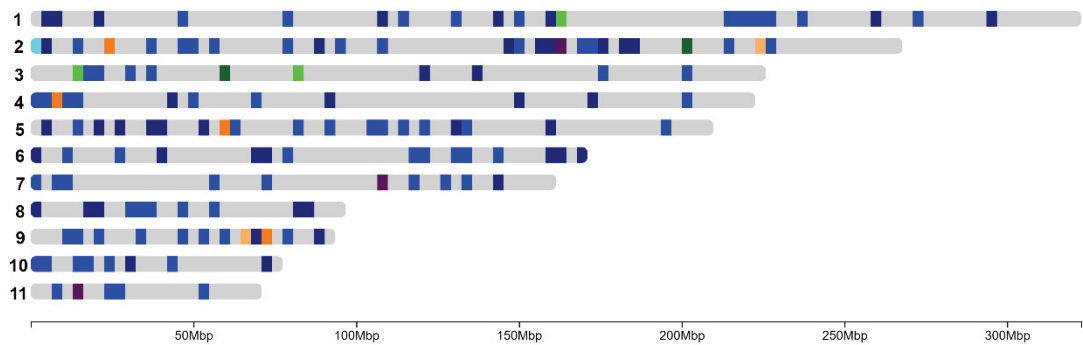
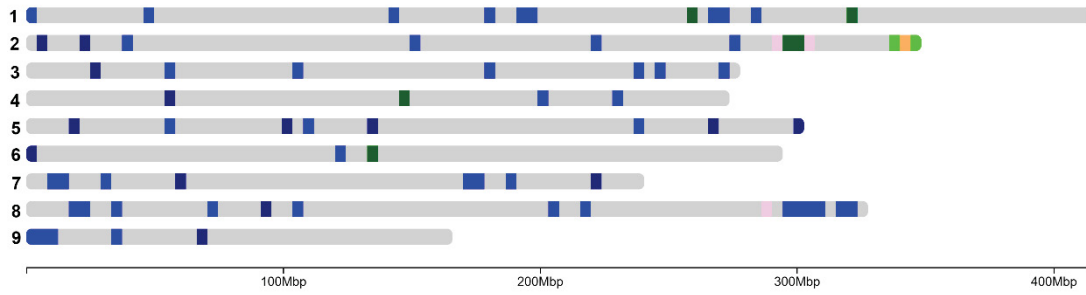
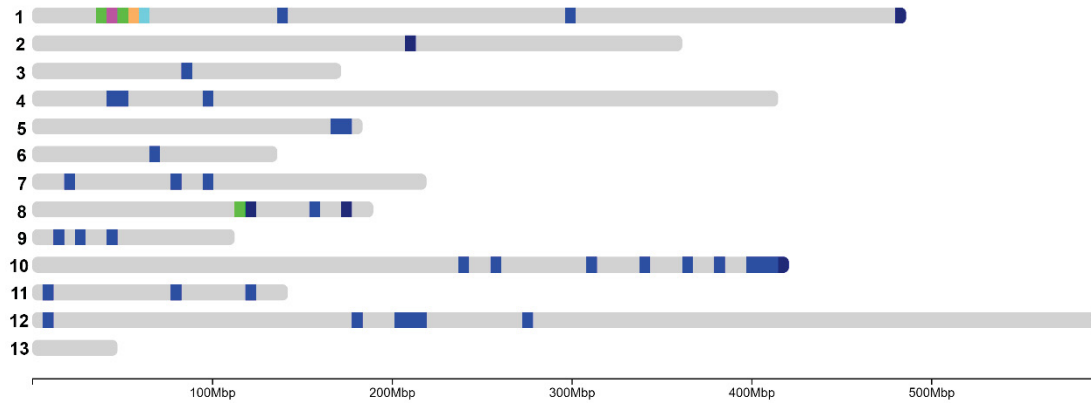


Figura 10. Distribuição dos genes OR nos cromossomos das espécies *Bufo bufo*, *Bufo gargarizans*, *Eleutherodactylus coqui* e *Engystomops pustulosus*. Legendas das cores:  $\alpha$  (verde),  $\beta$  (rosa),  $\gamma$  (azul),  $\delta$  (laranja),  $\epsilon$  (roxo),  $\eta$  (azul claro). Cores mais escuras dos respectivos grupos representam os pseudogenes e o amarelo representa truncados em geral.

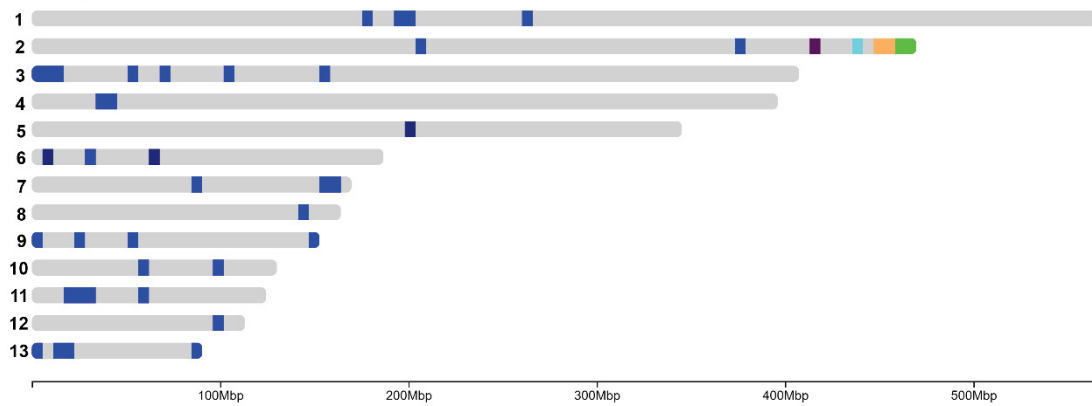
### *Hymenochirus boettgeri*



### *Leptobrachium ailaonicum*



### *Leptobrachium leishanense*



### *Pyxicephalus adspersus*

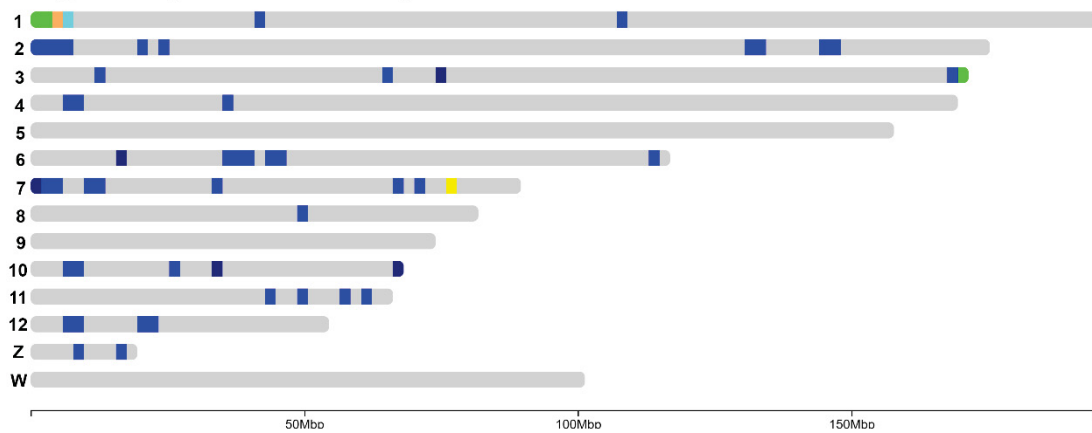


Figura 11. Distribuição dos genes OR nos cromossomos das espécies *Hymenochirus boettgeri*, *Leptobrachium ailaonicum*, *Leptobrachium leishanense* e *Pyxicephalus adspersus*. Legendas das cores:  $\alpha$  (verde),  $\beta$  (rosa),  $\gamma$  (azul),  $\delta$  (laranja),  $\varepsilon$  (roxo),  $\eta$  (azul claro). Cores mais escuras dos

respectivos grupos representam os pseudogenes e genes em amarelo representam truncados em geral.

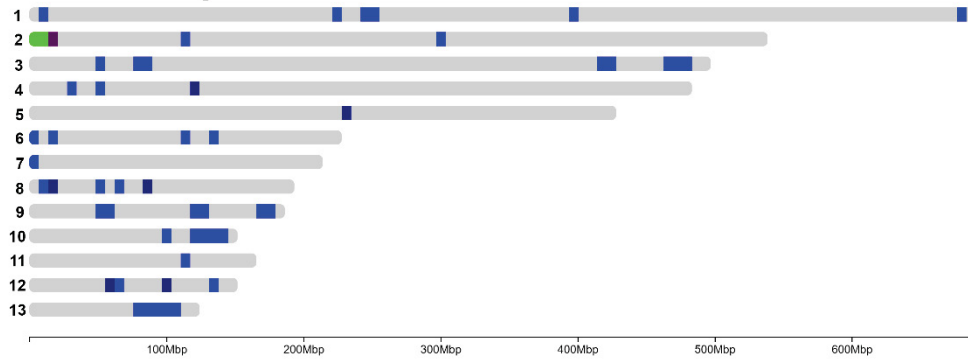
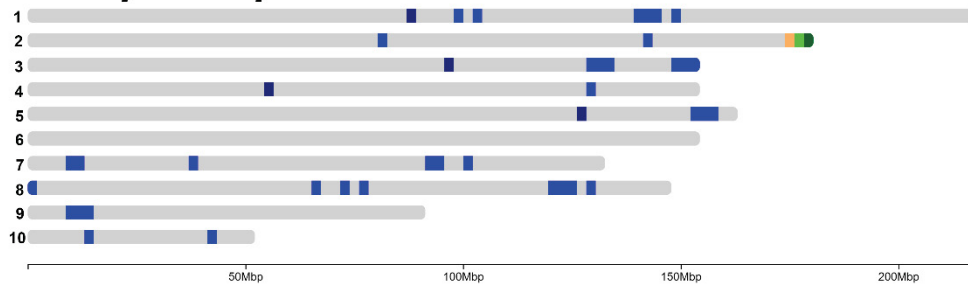
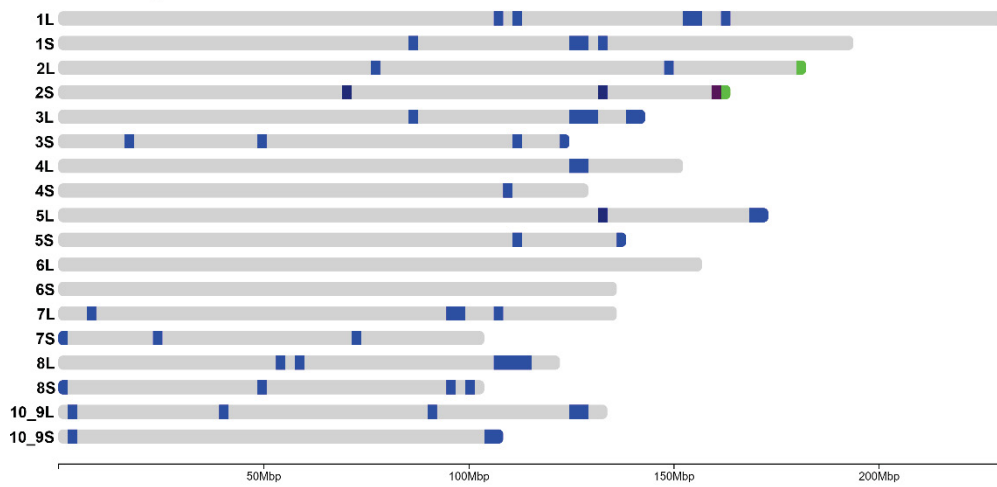
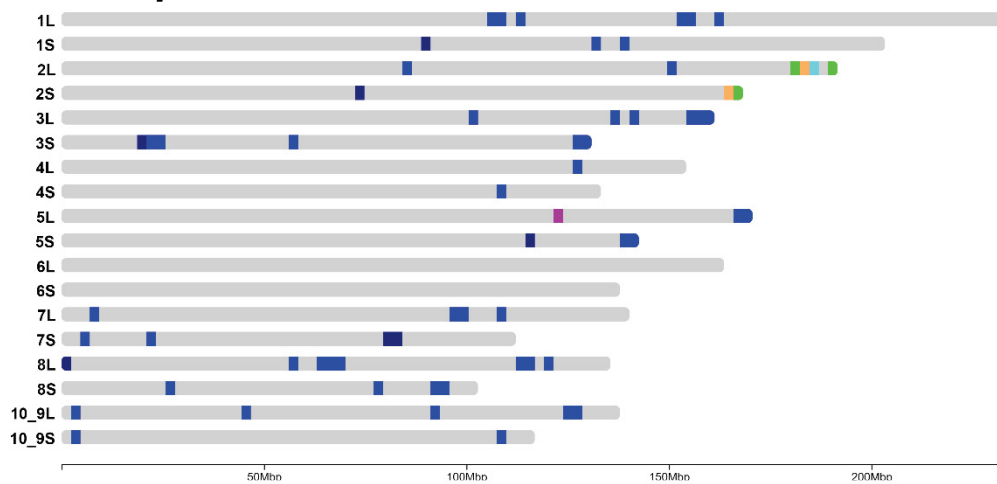
***Rana temporaria******Xenopus tropicalis******Xenopus borealis******Xenopus laevis***

Figura 12. Distribuição dos genes OR nos cromossomos das espécies *Rana temporaria*, *Xenopus tropicalis*, *Xenopus borealis* e *Xenopus laevis*. Legendas das cores:  $\alpha$  (verde),  $\beta$  (rosa),  $\gamma$  (azul),  $\delta$  (laranja),  $\epsilon$  (roxo),  $\eta$  (azul claro). Cores mais escuras dos respectivos grupos representam os pseudogenes e genes em amarelo representam truncados em geral. Sub genomas L e S em *X. borealis* e *X. tropicalis* são exibidos com seus respectivos números.

Os maiores *clusters* foram observados em *B. bufo* (13.6Mb), *B. gargarizans* (11Mb), *L. ailaonicum* (9.3Mb) e *L. leishanense* (8.2Mb). Além disso, todos os maiores *clusters* em todas as espécies são compostos de genes intactos do grupo  $\gamma$  (gamma), com exceção de *P. adspersus* e *X. tropicalis*, na qual o maior *cluster* é composto de pseudogenes, sendo 2.04Mb e 3.09Mb, respectivamente, apesar do segundo *cluster* nessas duas espécies possuir tamanho aproximado, com 2.03Mb e 3.01Mb, respectivamente (Tabela 6).

Houver bastante proximidade de *clusters* que continham genes dos grupos  $\alpha$  (alfa),  $\beta$  (beta),  $\delta$  (delta),  $\epsilon$  (épsilon) e  $\eta$  (eta) nas regiões subterminais, principalmente no cromossomo 2 (*H. boettgeri*, *L. leishanense*, *R. temporária*, as três espécies de *Xenopus*), cromossomo 1 em *L. ailaonicum* e cromossomo 3 em *B. gargarizans*. Uma vez observado certa proximidade entre esses grupos, nas três espécies de *Xenopus*, *H. boettgeri*, *P. adspersus* e *L. leishanense* é possível observar *clusters* compostos dos grupos  $\alpha$  (alfa),  $\delta$  (delta) e  $\eta$  (eta), sendo que os grupos  $\beta$  (beta) e  $\epsilon$  (épsilon) tendem a ter genes OR dispersos em poucos cromossomos, como um *cluster* de  $\beta$  (beta) de 600Kb em *B. gargarizans*, e *cluster* de pseudogenes  $\beta$  (beta) no cromossomo 8 de *H. boettgeri* e poucos genes dispersos nos cromossomos 7 e 11 de *E. pustulosus* (Figuras 6 e 7).

Por fim, *E. pustulosus* foi a espécie que mais apresentou dispersão de *clusters* e *singletons* em todos os cromossomos, que possivelmente são indicativos de muitos rearranjos cromossômicos.

Tabela 6. Os cinco maiores clusters de genes OR identificados nos cromossomos de cada espécie.

<i>Bufo bufo</i>							
<i>Cluster</i>	<i>Cromossomo</i>	<i>Intactos</i>	<i>Pseudogenes</i>	<i>Truncados</i>	<i>Início</i>	<i>Fim</i>	<i>Comprimento em bases</i>
1.1	LR991667.1	135	0	0	817298847	830966729	13667883
1.2	LR991667.1	0	49	0	827002784	832142124	5139341
4.1	LR991670.1	30	6	0	29815764	33030528	3214765
2.1	LR991668.1	29	0	0	293023459	295681113	2657655
1.3	LR991667.1	0	13	0	776735444	778851827	2116384
<i>Bufo gargarizans</i>							
2.1	CM026466.1	136	0	0	2939498	13977536	11038039
8.1	CM026472.1	71	0	0	186984119	190013251	3029133
2.2	CM026466.1	32	0	0	23645254	25796153	2150900
11.1	CM026475.1	37	0	0	79551436	81621074	2069639
3.1	CM026467.1	24	0	0	14807370	16670909	1863540
<i>Eleutherodactylus coqui</i>							
3.1	CM034088.1	20	0	0	71047393	72200665	1153273
3.2	CM034088.1	4	0	0	61333618	61854043	520426
<i>Engystomops pustulosus</i>							
4.1	CM033644.1	102	0	0	2988471	6107160	3118690
4.2	CM033644.1	0	98	0	2965111	5991763	3026653
1.1	CM033641.1	29	0	0	218091367	219700370	1609004
7.1	CM033647.1	21	0	0	79904	1474170	1394267
1.2	CM033641.1	0	24	0	215235960	216412922	1176963
<i>Hymenochirus boettgeri</i>							
9.1	CM033477.1	47	0	0	88858	4148410	4059553
3.1	CM033471.1	0	44	0	269886686	272017823	2131138
7.1	CM033475.1	0	49	0	13337839	15126426	1788588
8_10.1	CM033476.1	35	0	0	297721989	299165528	1443540
8_10.2	CM033476.1	0	34	0	297756833	299134339	1377507
<i>Leptobrachium ailaonicum</i>							
10.1	CM032243.1	141	0	0	404191432	413511698	9320267
10.2	CM032243.1	0	82	0	404155817	410579299	6423483
4.1	CM032237.1	118	0	0	46362123	50755409	4393287
4.2	CM032237.1	0	24	0	44861978	47844773	2982796
4.3	CM032237.1	0	22	0	48400026	50737530	2337505
<i>Leptobrachium leishanense</i>							
3.1	CM019065.1	170	0	0	5017392	13302591	8285200
3.2	CM019065.1	0	47	0	5010157	10130526	5120370
4.1	CM019066.1	119	0	0	36444309	40523743	4079435
1.1	CM019063.1	0	45	0	195753735	197700329	1946595
4.2	CM019066.1	0	13	0	36517246	38186606	1669361
<i>Pyxicephalus adspersus</i>							
2.1	CM016417.1	0	56	0	5227280	7273386	2046107
6.1	CM016421.1	58	0	0	37210494	39248177	2037684
2.2	CM016417.1	66	0	0	3166484	4691971	1525488
2.3	CM016417.1	0	50	0	3175671	4665960	1490290
7.1	CM016422.1	68	0	0	10132799	11587393	1454595
<i>Rana temporaria</i>							
3.1	LR991682.1	152	0	0	472652512	479266922	6614411
3.2	LR991682.1	0	72	0	472968990	478980350	6011361
1.1	LR991680.1	44	0	0	244593151	246851055	2257905

3.3	LR991682.1	58	0	0	469473036	471401881	1928846
3.4	LR991682.1	0	8	0	469515033	471424735	1909703
<i>Xenopus borealis</i>							
3L.1	CM044435.1	75	0	0	140083582	141772446	1688865
3S.1	CM044436.1	0	13	0	124367719	125589076	1221358
1L.1	CM044431.1	27	0	0	154515658	155684442	1168785
3L.2	CM044435.1	0	17	0	140526413	141655909	1129497
8S.1	CM044446.1	8	0	0	95779533	96814074	1034542
<i>Xenopus laevis</i>							
3L.1	CM030344.1	106	0	0	156073421	158390810	2317390
3L.2	CM030344.1	0	38	0	156164571	158271368	2106798
3S.1	CM030345.1	55	0	0	128936791	130599668	1662878
3S.2	CM030345.1	0	22	0	128927450	130581765	1654316
1L.1	CM030340.1	23	0	0	154326728	155489236	1162509
<i>Xenopus tropicalis</i>							
3.1	CM004445.2	0	45	0	147793062	150889567	3096506
3.2	CM004445.2	233	0	0	147844074	150858815	3014742
7.1	CM004449.2	199	0	0	10278198	12594200	2316003
9.1	CM004451.2	84	0	0	10237300	11879321	1642022
1.1	CM004443.2	33	0	0	142370897	143552701	1181805

## 5. DISCUSSÃO

### 5.1 Diferenças metodológicas na identificação dos genes OR

Neste trabalho foram apresentados o repertório completo de genes OR de 28 espécies de anfíbios anuros em diferentes níveis de montagem do genoma, dos quais 13 são identificações inéditas em comparação com a análise recente de Han et al. (2022).

Han et al. (2022) criaram um conjunto de módulos em *Python*, chamado Genome2OR, que foi baseado no método de busca alternativo ao *blastp*, chamado *nhmmer*, que se baseia em buscas em torno de perfis de HMM (do inglês, *Profile Hidden Markov models*). Desse modo, os *scripts* utilizam perfis de HMM previamente classificados de acordo com sete clados evolucionários (anfioxo, peixes agnatos, peixes gnatostomados, anfíbios, répteis, aves e mamíferos) que foram identificados através de buscas de receptores olfativos humanos contra o banco de dados de nucleotídeos do NCBI com um nível de restringência *E-value*  $1e-60$ . Assim, uma vez categorizados com uma identidade de sequência de 80%, os grupos foram alocados nesses sete grandes clados.

Mais ainda, a utilização dos cinco módulos resumidamente faz buscas contra genomas de cordados utilizando essas sequências de perfis HMM categorizadas em nucleotídeos e fazem a triagem que separa o resultado de acordo com as categorias de genes intactos, truncados e pseudogenes. Assim, com sucesso, eles identificaram 765.248 genes de receptores olfativos distribuídos em 1694 genomas.

Esta tese utiliza em grande parte os *scripts* criados de acordo com a metodologia de Niimura (2012a; 2012b), os quais sofreram determinadas adaptações para que se adequassem a identificação de vertebrados em geral. De forma semelhante a Han et al. (2022), nós utilizamos dos resultados das buscas homólogas tendo como base genes OR previamente identificados dos diferentes grupos, mas que foram categorizados de acordo com os grupos filogenéticos de acordo com Niimura (2009a). Desse modo, as diferenças na identificação se basearam na maneira como a triagem das sequências identificadas foi feita. Nós tentamos utilizar os módulos do Genome2OR desenvolvido por Han et al. (2022), sem sucesso, principalmente devido à erros em alguns dos *scripts* que parecem ter sido encontrados também por outros pesquisadores e que não foram resolvidos pelos autores em seu repositório público no Github.

De parte disso e, do fato dos autores também citarem a dificuldade em obter os *scripts* na qual a metodologia disponibilizada por Niimura (2012) como sendo de não conhecimento do público, nós utilizamos a metodologia do Prof. Dr. Niimura, que gentilmente nos disponibilizou os *scripts*, mas como auxiliou e participou da criação de novos *scripts* durante o processo de identificação. Assim, era esperado encontrar diferenças na composição dos genes intactos, pseudogenes e truncados com relação aos repertórios observados por Han et al. (2022) em maior grau, como no caso da grande quantidade de genes truncados em *S. holbrookii* (1659 em comparação com nossa identificação de 287) e *S. bombifrons* (1272 em comparação com nossa identificação de 316) por Han et al. (2022), e em menor grau, como no caso da identificação dos genes intactos de *R. temporaria* por nós (862) em comparação com os pesquisadores chineses (520) (Tabela 7).

Tabela 7. Comparação da quantidade de genes intactos, truncados e pseudogenes identificados neste trabalho, em Lu et al. (2021) e Han et al. (2022). Os traços determinam espécies aonde os genes OR não foram identificados em determinado trabalho. Legendas: I – Intactos; T – truncados; P – pseudogenes.

Espécie	Este estudo			Lu et al. (2021)			Han et al. (2022)		
	I	T	P	I	T	P	I	T	P
<i>Bombina variegata</i>	403	137	151	-	-	-	-	-	-
<i>Bufo bufo</i>	745	0	172	-	-	-	-	-	-
<i>Bufo gargarizans</i>	805	0	226	765	-	248	738	0	237
<i>Eleutherodactylus coqui</i>	688	75	160	-	-	-	-	-	-
<i>Engystomops pustulosus</i>	1080	77	928	-	-	-	-	-	-
<i>Glandirana rugosa</i>	1047	70	340	-	-	-	-	-	-
<i>Hymenochirus boettgeri</i>	507	19	668	-	-	-	-	-	-
<i>Leptobrachium ailaonicum</i>	684	50	282	-	-	-	-	-	-
<i>Leptobrachium leishanense</i>	754	3	241	-	-	-	708	0	245
<i>Limnodynastes dumerilii</i>	873	47	377	-	-	-	849	154	390
<i>Lithobates catesbeianus</i>	986	110	352	-	-	-	946	157	428
<i>Nanorana parkeri</i>	601	4	327	598	-	221	616	179	5
<i>Oophaga pumilio</i>	423	44	132	-	-	-	431	12	192
<i>Phrynoglossus myanhessei</i>	269	212	136	-	-	-	-	-	-
<i>Phyllomedusa bahiana</i>	491	6	486	-	-	-	-	-	-
<i>Pipa carvalhoi</i>	241	224	182	-	-	-	-	-	-
<i>Pipa parva</i>	440	91	74	-	-	-	-	-	-
<i>Platyplectrum ornatum</i>	336	124	281	-	-	-	-	-	-
<i>Pyxicephalus adspersus</i>	725	126	350	711	-	405	726	2	536

<i>Rana temporaria</i>	862	0	189	-	-	-	520	48	315
<i>Rhinella marina</i>	865	5	292	797	-	356	832	2	330
<i>Scaphiopus couchii</i>	6	198	497	-	-	-	6	675	15
<i>Scaphiopus holbrookii</i>	28	287	480	-	-	-	31	1659	86
<i>Spea bombifrons</i>	26	316	452	-	-	-	27	1272	105
<i>Spea multiplicata</i>	853	44	204	-	-	-	702	33	232
<i>Xenopus borealis</i>	591	70	181	-	-	-	-	-	-
<i>Xenopus laevis</i>	738	0	310	592	-	320	656	85	349
<i>Xenopus tropicalis</i>	990	0	250	854	-	326	1021	0	230

Em todo caso, a anotação de genes OR acaba sendo um desafio menos problemático devido ao fato da maioria dos genes OR de cordados possuir somente um éxon (*SEG*) (BUCK & AXEL, 1991; NEF ET AL., 1992), com a exceção de peixes Ciclídeos da ordem dos Perciformes, que possuem um número entre 2 a 4 éxons (AZZOUZI ET AL., 2014). Por conta disso e, do fato de haver a presença de padrões nas sequências conservadas das proteínas OR que possibilitam uma anotação mais precisa para determinado genoma, nos últimos anos, diversos autores partiram de princípios de identificação de ORs semelhantes para obter o repertório de OR em diferentes espécies.

Por exemplo, Hayden et al. (2010) desenvolveram um método baseado em TFASTX (PEARSON ET AL., 1997), que foi chamado de OR (*Olfactory Receptor Family Assigner*), que foi amplamente utilizado nos trabalhos de Yohe et al. (2020; 2021; 2022) e identifica genes OR categorizados de acordo com as subfamílias da Classe 1 e Classe 2, que são utilizados na identificação de genes de aves e mamíferos.

Outros autores utilizam da metodologia da separação das sequências obtidas a partir de PCR, que consiste em separar as sequências obtidas pela presença ou ausência de mutações de quadro de leitura e códons de parada para obter um número aproximado do repertório de determinado genoma, tanto após uma busca contra um banco de dados, quando pela análise direta das sequências obtidas (STEIGER ET AL., 2009; VIEYRA, 2011; QIAN ET AL., 2020).

Nosso método de identificação dos genes OR, como é baseado em buscas homólogas utilizando TBLASTN, acaba se tornando um processo mais demorado do que métodos que utilizam perfis de HMM (Han et al., 2022), mas não menos eficaz (Tabela 5). Assim, diferenças quanto a composição no repertório das diferenças espécies aqui estudadas não reflete possíveis erros na estimativa de identificação, principalmente dos

genes intactos. De todo modo, nosso trabalho amplia em diversas espécies a quantidade de genes identificados, além de categorizá-los filogeneticamente nos principais grupos de genes OR conhecidos e também em analisar sua dispersão física nos diferentes cromossomos nas espécies em que o nível de montagem do genoma é apropriado pra tal.

### 5.2 Relação do repertório com a proximidade filogenética das espécies

Nosso trabalho teve como foco a identificação do repertório das espécies aqui estudadas, com foco na análise evolutiva que está presente na dinâmica da família multigênica dos genes OR. Apesar disso, a composição de genes intactos e pseudogenes, junto com sua distribuição física nos diferentes cromossomos, é levada em consideração quando analisamos espécies próximas filogeneticamente.

Devido à falta da disponibilidade maior de genomas de outros anfíbios, a família Hylidae, maior representante da Ordem Anura, com 1050 espécies (FROST, 2023), possui somente um representante aqui analisado, *Phyllomedusa bahiana*. Além disso, estudamos espécies com diferentes graus de parentesco evolutivo, das quais possuem maior proximidade entre si espécies das famílias Pipidae, Scaphiropodidae, Bufonidae e Ranidae (Figura 7).

A família Pipidae é a que possui mais representantes, tanto no número de espécies, quanto na disponibilidade de diferentes avaliações de seu repertório (NIIMURA, 2009a; 2009b; LU ET AL., 2021; HAN ET AL., 2022). As espécies dessa família são as únicas completamente morfologicamente adaptadas para o ambiente aquático (CANNATELLA & TRUEB, 1988) e estão distribuídos na África subsaariana e também na América do Sul (FROST, 2023). Além disso, duas espécies aqui estudadas possuem genomas poliplóides (*X. borealis* e *X. laevis*), sendo atualmente reconhecidos dois sub genomas derivados de um ancestral  $2n=18$  que passou por um evento de duplicação completa do genoma acerca de 17 a 18 Ma (SESSION ET AL., 2016).

### 5.3 A presença de clusters de genes OR nos cromossomos

Nós identificamos *clusters* de genes OR nas espécies que possuem nível de montagem cromossômica, estabelecendo uma distância máxima de 500Kb. Muitos dos trabalhos envolvendo a distribuição cromossômica de genes OR foi feita com base no repertório de genes humanos. Desse modo, trabalhos anteriores envolvendo a identificação de *clusters* envolveu desde dois genes OR consecutivos com menos de 1Mb como pertencentes ao mesmo *cluster* (GLUSMAN ET AL., 2001; ZHANG & FIRESTEIN, 2002), assim como menos de 500Kb (NIIMURA & NEI, 2003). Nós optamos por estabelecer um limite de 500Kb depois de verificar que a presença entre dois genes OR vizinhos acima de 450–500Kb era menos frequente (ver capítulo II).

Mudd (2019) em sua tese sobre genômica comparativa e distribuição de genes ortólogos entre diferentes espécies de anuros, analisou a distribuição de genes ortólogos um-a-um ao longo de diferentes espécies e verificou uma ampla conservação da região pericentromérica nos genomas, mas observou potenciais translocações seguidas de fusões entre regiões diferentes em diversos cromossomos das espécies *P. adspersus* e *E. pustulosus*, que possivelmente explicam a grande dispersão de genes  $\gamma$  (gamma) observados em *E. pustulosus* no nosso trabalho. Além disso, nossas análises mostraram conservação sintênica dos cromossomos 2 quanto a distribuição dos outros grupos OR nas espécies de *Xenopus* que foi evidenciada por Mudd (2019) entre outros genes ortólogos.

É certo que o ancestral comum de todos os tetrápodes possuía dois grupos de genes ( $\alpha$  e  $\gamma$ ) que adquiriram a habilidade de detectar odores voláteis e ao longo da evolução e divisão das espécies atuais, esses dois grupos expandiram grandemente, provavelmente pela importância que o olfato possui nas espécies terrestres (NIIMURA, 2009a; NIIMURA 2009b). Uma amplificação, que corresponde a expansão rápida de uma família gênica é funcionalmente importante e sugere cenários adaptativos. Como a rápida expansão desse grupo foi observada logo após a expansão dos tetrápodes, a manutenção de todos os grupos OR nos diferentes grupos de vertebrados sugere a importância da identificação variada dos genes OR atrelada ao ambiente em que a espécie está inserida. Desse modo, espécies aquáticas mantiveram grupos especializados na identificação das moléculas presentes no ambiente, assim como houve o ciclo de expansão por duplicação gênica e divergência funcional ao mesmo tempo em que pseudogenes eram formados, acumulando mutações deletérias e sendo excluídos do genoma, numa dinâmica que

estruturou os diferentes repertórios aqui observados, de acordo com o modelo Birth and Death (HUGHES & NEI, 1989).

De fato, os genes do grupo  $\gamma$  (gamma) foram os mais numerosos em todas as espécies e essa quantidade é evidenciada pela dispersão em *clusters* e genes isolados em quase todos os cromossomos aqui estudados. Apesar dessa alta dispersão desse grupo, pudemos observar que os outros grupos têm a maior tendência em sofrer menos rearranjos, se estabelecendo nos cromossomos 1, 2 e 3 na maior parte das espécies. As diferentes composições quantitativas desses diferentes genes possivelmente se mantêm por diferentes taxas evolutivas de duplicação, que mantêm genes funcionais exibindo diferentes níveis de similaridade, com poucos ou nenhum pseudogene intercalado nas regiões cromossômicas.

Genes formados por duplicação gênica tendem a manter essencialmente a mesma função e, no caso dos genes OR, essa dinâmica se mantêm em flutuações constantes de acordo com o previsto pelo modelo de Birth and Death, apesar dessa dinâmica ainda representar uma identificação combinatória que discrimina os diferentes odores (KAJIYA ET AL., 2001; MALNIC ET AL., 1999; SAITO ET AL., 2009)

## 6. CONSIDERAÇÕES FINAIS

O presente trabalho avaliou o repertório de genes de receptores olfatório de 28 espécies de anfíbios Anuros, aumentando o número de espécies com o repertório de ORs conhecidos. Nossas principais contribuições neste estudo foram:

A identificação pela primeira vez de 13 novos repertórios completos de genes OR descritos em anuros visto que, com o crescente avanço das técnicas de montagem dos genomas e a maior disponibilidade de espécies em bancos de dados públicos, há ainda uma carência de dados a respeito das famílias gênicas que são especializadas em funções relacionadas a detecção de moléculas químico sensoriais em espécies de anuros.

Nós também desenvolvemos diversos *scripts*, não somente voltados para a automatização da identificação, filtragem e refinamento de dados relacionados ao estudo dos receptores olfativos, mas que podem facilmente ser modificados para o estudo de outros genes ou famílias multigênicas de forma escalável e de acesso público.

O capítulo I do nosso trabalho disponibilizou a identificação inédita do repertório de genes OR das espécies *E. pustulosus*, *B. bufo* e *H. boettgeri*, além de fornecer informações a respeito da dinâmica evolutiva entre as três espécies, assim como a organização cromossômica dos genes identificados.

O capítulo II deste trabalho forneceu novas informações quanto a dinâmica evolutiva dos genes OR em torno dos genomas de espécies do gênero *Xenopus* filogeneticamente próximas, além da identificação dos OGGs, (*Ortholog Gene Groups*) que foram utilizados para desvendar parte da evolução dos genes OR nessas três espécies e também como a dinâmica desses genes foi estabelecida em espécies aloploplóides, assim como sua distribuição nos diferentes subgenomas das espécies *X. borealis* e *X. laevis*.

Nosso capítulo 3 possui os dados da identificação do repertório de genes OR de todas as espécies de anuros que possuíam genomas disponíveis em bancos de dados públicos até o fim de 2022, além de informações quanto ao mapeamento desses genes nos cromossomos de espécies no nível de montagem estabelecido para tal. Nossas futuras perspectivas envolvem a identificação dos OGGs entre todos os genomas aqui estudados, além de incluir novos genomas à medida que estes forem disponibilizados, com o intuito de desvendar informações a respeito da evolução desses genes desde o ancestral de todos os anfíbios anuros até os dias atuais, de forma que possamos fazer o comparativo com outros parâmetros ecológicos, morfológicos e de história de vida das diferentes espécies de anuros.

## 7. REFERÊNCIAS

- ABAFFY, T.; MALHOTRA, LUETJE, C.W. The molecular basis for ligand specificity in a mouse olfactory receptor: a network of functionally important residues. **J. of Bio. Chem.**, v. 282, n. 2, p. 1216–1224, 2007.
- ACHE, B.W.; YOUNG, J.M. Olfaction: diverse species, conserved principles. **Neuron**, v. 48, p. 417–430, 2005.
- ALIOTO, T.S.; NGAI, J. The odorant receptor repertoire of teleost fish. **BMC Genomics**, v. 6, n. 173, p. 1–14, 2005.
- ALMEIDA-SILVA, F.; ZHAO, T.; ULLRICH, K.K.; SCHRANZ, M.E.; PEER, Y.V. Syntenet: an R/Bioconductor package for the inference and analysis of synteny networks. **Bioinformatics**, v. 39, n. 1, p. btac806, 2023.
- ALTSCHUL, S.F.; GISH, W.; MILLER, W.; MYERS, E.W.; LIPMAN, D.J. Basic local alignment search tool. **J. of Mol. Bio.**, v. 215, n. 3, p. 403–410, 1990.
- ANAND, L.; LOPEZ, C.M.R. ChromoMap: R package for interactive visualization of multi-omics data and annotation of chromosomes. **BMC Bioinformatics**, v. 23, n. 33, p. 1–9, 2022.
- ARZABE, C.; SKUK, G.; BEIBER, M. *Pipa carvalhoi*. **The IUCN red list of threatened species**. Disponível em <http://dx.doi.org/10.2305/IUCN.UK.2010-2.RLTS.T58160A11728418.en>. Acesso em: 19 julho 2019.
- AZZOUZI, N.; BARLOY-HUBLER, F.; GALIBERT, F. Inventory of the cichlid olfactory receptor gene repertoires: identification of olfactory genes with more than one coding exon. **BMC Genomics**, v. 15, p.586, 2014.
- BACHMANOV, A.A.; BOSAK, N.P.; LIN, C.; MATSUMOTO, I.; OHMOTO, M.; REED, D.R.; NELSON, T.M. Genetics of taste receptors. **Curr. Pharm. Des.**, v. 20, n. 16, p. 2669–2683, 2014.
- BAILEY, N. **The elements of stochastic processes**. New York: Wiley, 1964.
- BAILEY, T.L.; ELKAN, C. Fitting a mixture model by expectation maximization to discover motifs in biopolymers. **Proc. Int. Conf. Intell. Syst. Mol. Bio.**, v. 2, p. 28–36, 1994.
- BAILEY, T.L.; JOHNSON, J.; GRANT, C.E.; NOBLE, W.S. The MEME Suite. **Nuc. Ac. Res.**, v. 43, n. W1, p. W39–W49, 2015.

- BARGMANN, C.I. **Chemosensation in *C. elegans***. Wormbook: California, 2006.
- BARRASSO, D.A.; CAJADE, R.; NENDA, S.J.; BALORIANI, G.; HERRERA, R. Introduction of the American bullfrog *Lithobates catesbeianus* (Anura: Ranidae) in natural and modified environments: an increasing conservation problem in Argentina. **South Am. J. Herp.**, v. 4, p. 69–75, 2009.
- BARRON, F.; WOODS, C.; KUHN, K.; BISPHO, J.; HOWARD, M.; CLOUTHIER, D.E. Downregulation of *dlx5* and *Dlx6* expression by *Hand2* is essential for initiation of tongue morphogenesis. **Development**, v. 138, p. 2249–2259, 2011.
- BELANGER, R.M.; CORKUM, L.D. Review of aquatic sex pheromones and chemical communication in anurans. **J. Herpetol.**, v. 43, p. 184–191, 2009.
- BENZEKRI, N.A.; REISS, J.O. Olfactory metamorphosis in the coastal tailed frog *Ascaphus truei* (Amphibia, Anura, Leiopelmatidae). **J. Morphol.**, v. 273, p. 68–87, 2012.
- BERTMAR, G. Evolution of vomeronasal organs in vertebrates. **Evolution**, v. 35, p. 359–366, 1981.
- BOWMAKER, J.K. Evolution of vertebrate visual pigments. **Vision Research**, v. 48, n. 20, p. 2022–2041, 2008.
- BUCK, L.; AXEL, R. A novel multigene family may encode odorant receptors: a molecular basis for odor recognition. **Cell**, v. 65, p. 175–187, 1991.
- CHENG, F.; SUN, C.; WU, J.; SCHNABLE, J.; WOODHOUSE, M.R.; LIANG, J.; CAI, C.; FREELING, M. Epigenetic regulation of subgenome dominance following whole genome triplication in *Brassica rapa*. **New Phytol.**, v. 211, n. 1, p. 288–299, 2016. DOI: 10.1111/nph.13884.
- CHENG, F.; WU, J.; CAI, X.; LIANG, J.; FREELING, M.; WANG, X. Gene retention, fractionation and subgenome differences in polyploid plants. **Nat. Plants.**, v. 4, n. 5, p. 258–268, 2018. DOI:10.1038/s41477-018-0136-7.
- CONANT, G.C.; WOLFE, K.H. Turning a hobby into a job: how duplicated genes find new functions. **Nat. Rev. Genet.**, v. 9, p. 938–950, 2008.
- CONCEIÇÃO, I.C.; AGUADÉ, M. High incidence of interchromosomal transpositions in the evolutionary history of a subset of OR genes in *Drosophila*. **J. Mol. Evol.**, v. 66, p. 325–332, 2008.
- DAHANUKAR, A.; HALLEN, E.A.; CARLSON, J.R. Insect chemoreception. **Curr. Opin. Neurobiol.**, v. 15, p. 423–430, 2005.

DANGLES, O.; IRSCHICK, D.J.; CHITTKA.; CASAS, J. Variability in sensory ecology: expanding the bridge between physiology and evolutionary biology. **Q. Rev. Biol.**, 84:51–74, 2009.

DEGL'INNOCENTI, A.; D'ERRICO, A. Regulatory features for odorant receptor genes in the mouse genome. **Front. Genet.**, v. 8, n. 9, p. 1–8, 2017.

DONG, D.; HE, G.; ZHANG, S.; ZHANG, Z. Evolution of olfactory receptor genes in primates dominated by birth-and-death process. **Genome Biol. Evol.**, v. 1, p. 258–64, 2009. DOI: <https://doi.org/10.1093/gbe/evp026> PMID: 20333195.

DRYER, L. Evolution of odorant receptors. **BioEssays**, v. 22, p. 803–810, 2000.

DUELLMAN, W.E.; TRUEB, L. **Biology of Amphibians**. Baltimore: Johns Hopkins University Press, 1986.

EDGER, P.P.; POORTEN, T.; VANBUREN, R.; HARDIGAN, M.A.; COLLE, M.; MCKAIN, M.R.; SMITH, R.D.; TERESI, S.; NELSON, A.D.L.; WAI, C.M.; ALGER, E.I.; BIRD, K.A.; YOCCA, A.E.; PUMPLIN, N.; OU, S.; BEN-ZVI, G.; BRODT, A.; BARUCH, K.; SWALE, T.; SHIUE, L.; ACHARYA, C.B.; COLE, G.S.; MOWER, J.P.; CHILDS, K.L.; JIANG, N.; LYONS, E.; FREELING, M.; PUZEY, J.R.; KNAPP, S.J. Origin and evolution of the octoploid strawberry genome. **Nat. Genet.**, v. 51, n. 3, p. 541–547, 2019. DOI: 10.1038/s41588-019-0356-4.

EIRÍN-LÓPEZ, J.M.; REBORBINOS, L.; ROONEY, A.P.; ROZAS, J. The birth-and-death evolution of multigene families revisited. **Genome Dynamics**, v. 7, p. 170–196, 2012.

EISTHEN, H.L. Evolution of vertebrate olfactory systems. **Brain Behav. Evol.**, v. 40, p. 222–233, 1997.

EISTHEN, H.L. Phylogeny of the vomeronasal system and of receptor cell types in the olfactory and vomeronasal epithelia of vertebrates. **Microsc. Res. Tech.**, v. 23, p. 1–21, 1992.

ELURBE, D.M.; PARANJPE, S.S.; GEORGIU, G.; VAN KRUIJSBERGEN, I.; BOGDANOVIC, O.; GIBEAUX, R.; HEALD, R.; LISTER, R.; HUYNEN, M.A.; VAN HEERINGEN, S.J.; VEENSTRA, G.J.C. Regulatory remodeling in the allo-tetraploid frog *Xenopus laevis*. **Genome Bio.**, v. 18, p. 1–18, 2017. DOI : 10.1186/s13059-017-1335-7

EMMS, D.; KELLY, S. OrthoFinder: phylogenetic orthology inference for comparative genomics. **Genome Bio.**, v. 20, n. 1, p. 1–14, 2019.

DOI:<https://doi.org/10.1186/s13059-019-1832-y>.

ESCORIZA, D.; HASSINE, J. North-western Africa amphibians. *In*: ESCORIZA, D.; HASSINE, J. (Eds). **Amphibians of North Africa**. Academic Press, 2019.

EVANS, B.J. Genome Evolution and speciation genetics of allopolyploid clawed frogs (*Xenopus* and *Silurana*). **Front Biosci.**, v. 13, p. 4687–706, 2008.

EVANS, B.J. ; CARTER, T.F. ; GREENBAUM, E. ; GVOŽDÍK, V. ; KELLEY, D.B. ; MCLAUGHLIN, P.J. ; PAUWELS, O.S.G. ; PORTIK, D.M. ; STANLEY, E.L. ; TINSLEY, R.C. ; TOBIAS, M.L. ; BLACKBURN, D.C. Genetics, morphology, advertisement calls, and historical records distinguish six new polyploid species of African Clawed Frog (*Xenopus*, Pipidae) from West and Central Africa. **Plos One**, v. 10, n. 2, p. e0412823, 2015.

EVANS, B.J.; MUDD, A.B.; BREDESON, J.V.; FURMAN, B.L.S.; WASONGA, D.V.; LYONS, J.B.; HARLAND, R.M.; ROKHSAR, D.S. New insights into *Xenopus* sex chromosome genomics from the Marsabit clawed frog *X. borealis*. **J. Evol. Bio.**, v. 35, n. 12, p. 1777–1790, 2022.

FAIVOVICH, J.; HADDAD, C.F.B.; GARCIA, P.C.A.; FROST, D.R.; CAMPBELL, J.A. WARD, W. Systematic review of the frog family Hylidae, with special reference to Hylinae: phylogenetic analysis and taxonomic revision. **Bull. Am. Mus. Nat.**, v. 294, p. 1–240, 2005.

FARRIS, J.S. Phylogenetic analysis under Dollo's law. **Syst. Zool.**, v. 26, n. 1, p. 77–88, 1997.

FEI, L.; YE, C.; HUANG, Y.; LIU, M. Atlas of Amphibians of China. Zhengzhou: Henan Science and Technical Press, 1999.

FICETOLA, G.F.; COIĆ, C.; DETAINT, M.; BERRONEAU, M.; LORVELEC, O.; MIAUD, C. Pattern of distribution of the American bullfrog *Rana catesbeiana* in Europe. **Bio. Inv.**, v. 9, p. 767–772, 2006.

FLEISCHER J.; STROTMANN, R.A.; JANKENS, K.M. Compartmentalization of the mammalian olfactory system. *In*: FRITZSCH, B (Ed.). **The sense: a comprehensive reference**. Iowa: Elsevier, 2020, p. 493–515.

- FLEISCHER, J.; BREER, H.; STROTMANN, J. Mammalian olfactory receptors. **Front. Cell Neurosci.**, v. 3, p. 9, 2009.
- FLEMING, R.I.; MACKENZIE, C.D.; COOPER, A.; KENNEDY, M.W. Foam nest components of the túngara frog: a cocktail of proteins conferring physical and biological resilience. **Proc. R. Soc.**, v. 276, n. 1663, p. 1787–1795, 2009.
- FREDRIKSSON, R.; LAGERSTRÖM, M.C.; LUNDIN, L.; SCHIÖTH, H.B. The G-protein-coupled receptors in the human genome form five main families. Phylogenetic analysis, paralogon groups, and fingerprints. **Mol. Pharmacol.**, v. 63(6), p. 1256–1272, 2003.
- FREELING, M.; SCANLON, M.J.; FWLER, J.E. Fractionation and subfunctionalization following genome duplications: mechanisms that drive gene content and their consequences. **Curr. Opin. Genet. Dev.**, v. 35, p. 110–118, 2015. DOI:<http://dx.doi.org/10.1016/j.gde.2015.11.002>.
- FREEMAN, A.R.; OPHIR, A.G.; SHEERAN, M.J. The giant pouched rat (*Cricetomys ansorgei*) olfactory receptor repertoire. **PLOS One**, v. 15, n. 4, p. e0221981, 2020.
- FREITAG, J.; BECK, A.; LUDWIG, G.; VON BUCHHOLTZ, L.; BREER, H. On the origin of the olfactory receptor family: receptor genes of the jawless fish (*Lampetra fluviatilis*). **Gene**, v. 226, n. 2, p. 165–174, 1999.
- FREITAG, J.; KRIEGER J.; STROTMANN, J.; BREER, H. Two classes of olfactory receptors in *Xenopus laevis*. **Neuron**, v. 15, p. 1383–1392, 1995.
- FREITAG, J.; LUDWIG, G.; ANDREINI, I.; RÖSSLER, P.; BREER, H. Olfactory receptors in aquatic and terrestrial vertebrates. **J. Comp. Physiol. A.**, v. 183, p. 635–650, 1998.
- FROST, D.R. **Amphibian species of the world: an online reference**. Version 6.0. Disponível em <http://research.amnh.org/herpetology/amphibia/index.html>. Acesso em: 19 outubro 2023.
- FROST, D.R.; GRANT, T.; FAIVOVICH, J.; BRAIN, R.H.; HAAS, A.; HADDAD, C.F.B.; DE SA, R.O.; CHANNING, A.; WILKINSON, M. DONNELLAN, S.C.; RAXWORTHY, C.J.; CAMPBELL, J.A.; BLOTTO, B.L.; MOLER, P.; DREWES, R.C.; NUSSBAUM, R.A.; LYNCH, J.D.; GREEN, D.M.; WHEELER, W.C. The amphibian tree of life. **AMNH**, v. 2006, n. 297, p. :1–300, 2006.
- FUCHS, T.; GLUSMAN, G.; HORN-SABAN, S.; LANCET, D.; PILPEL, Y. The human olfactory subgenome: from sequence to structure and evolution. **Hum. Genet.**, v. 108, p. 1–13, 2001.

- FURMAN, B.L.S.; EVANS, B.J. Divergent evolutionary trajectories of two young, homomorphic, and closely related sex chromosome systems. **Genome Biol. Evol.**, v. 10, p. 742–755, 2018.
- GEITHE, C.; ANDERSEN, G.; MALKI, A.; KRAUTWURST, D. A butter aroma recombinant activates human Class-I odorant receptors. **J. Agric. Food Chem.**, v. 63, n. 43, p. 9410–9420, 2015.
- GILAD, Y.; PRZEWORSKI, M.; LANCET, D. Loss of olfactory receptor genes coincides with the acquisition of full trichromatic vision primates. **PLOS One**, v. 2, p. E5, 2004.
- GILL, R.A.; SCOSSA, F.; KING, G.J.; GOLICZ, A.A.; TONG, C.; SNOWDON, R.J.; FERNIE, A.R.; LIU, S. On the role of transposable elements in the regulation of gene expression and subgenomic interactions in crop genomes. **Crit. Rev. Plant. Sci.**, v. 40, n. 2, p. 157–189, 2021. DOI:<https://doi.org/10.1080/07352689.2021.1920731>.
- GLEZER, I.; MALNIC, B. Chapter 5 - Olfactory receptor function. *In*: DOTY, R. (Ed). **Smell and Taste**. Elsevier, 2019.
- GLUSMAN, G.; YANAI, I.; RUBIN, I.; LANCET, D. The complete human olfactory subgenome. **Genome Res.**, v. 11, n. 5, p. 685–702, 2001.
- GONZALEZ-KRISTELLER, D.C.; NASCIMENTO, J.B.P.; GALANTE, P.A.F.; MALNIC, B. Identification of agonists for a group of human odorant receptors. **Front. Pharmacol.**, v. 6, n. 35, p. 1–8, 2015.
- GOVARDOVSKII, V.I.; REUTER, T. Why do green rods of frog and toad retinas look green? **J. Comp. Physiol. A. Neuroethol. Sens. Neural Behav. Physiol.**, v. 200, n. 9, p. 823–835, 2014.
- GUO, S.; KIM, J. Molecular evolution of Drosophila odorant receptor genes. **Mol. Biol. Evol.**, v. 24, p. 1198–1207, 2007.
- HAN, W.; WU, Y.; ZENG, L.; ZHAO, S. Building the Chordata Olfactory Receptor Database using more than 400.000 receptors annotated by Genome2OR. **Sci. China Life Sci.**, v. 65, n. 12, p. 2539–2551, 2022.
- HASSENKLÖVER, T.; MANZINI, I. Olfactory wiring logic in amphibians challenges the basic assumptions of the unbranched axon concept. **The Journal of Neuroscience**, v. 33, n. 44), p. 17247–17252, 2013.

HAYDEN, S.; BEKAERT, M.; CRIDER, T.A.; MARIANI, S.; MURPHY, W.J.; TEELING, E.C. Ecological adaptation determines functional mammalian olfactory subgenomes. **Genome Res.**, v. 20, p. 1–9, 2010.

HAYDEN, S.; TEELING, E.C. The molecular biology of vertebrate olfaction. **Anat. Rec.**, v. 297, p. 2216–2226, 2014.

HELLSTEN, U.; HARLAND, R.M.; GILCHRIST, M.J.; HENDRIX, D.; JURKA, J.; KAPITONOV, V.; OVCHARENKO, I.; PUTNAM, N.H.; SHU, S.; TAHER, L.; BLITZ, I.L.; BLUMBERG, B.; DICHAMNN, D.S.; DUBCHAK, I.; AMAYA, E.; DETTER, J.S.; FLETCHER, R.; GERHARD, D.S.; GOODSTEIN, D.; GRAVES, T.; GRIGORIEV, I.V.; GRIMWOOD, J.; KAWASHIMA, T.; LINDQUIST, E.; LUCAS, S.M.; MEAD, P.E.; MITROS, T.; OGINO, H.; OHTA, Y.; POLIAKOV, A.V.; POLLET, N.; ROBERT, J.; SALAMOV, A.; SATER, A.K.; SCHUMTZ, J.; TERRY, A.; VIZE, P.D.; WARREN, W.C.; WELLS, D.; GRAINGER, R.; GRAMMER, T.; KHOKHA, M.K.; RICHARDSON, P.M.; ROKHSAR, D.S. The genome of the Western clawed frog *Xenopus tropicalis*. **Science**, v. 328, n. 5978, p. 633–636, 2010.

HEYER, W.R.; RAND, A.S. Foam Nest Construction in the Leptodactylid Frogs *Leptodactylus pentadactylus* and *Physalaemus pustulosus* (Amphibia, Anura, Leptodactylidae). **J. Herpetol.**, v. 11, p. 225–228, 1977.

HOOVER, K.C. Smell with inspiration: the evolutionary significance of olfaction. **Yearbook of Physical Anthropology**, v. 53, p. 63–74, 2010.

HUANG, Y.; NIU, B.; GAO, Y.; FU, L.; LI, W. CD-HIT Suite: a web server for clustering and comparing biological sequences. **Bioinformatics**, v. 26, p. 680–682, 2010.

HUGHES, G.M.; BOSTON, E.S.M.; FINARELLIM, J.A.; MURPHY, W.J.; HIGGINS, D.G.; TEELING, E.C. The Birth and Death of Olfactory Receptor Gene Families in Mammalian Niche Adaptation. **Mol. Biol. Evol.**, v. 35, n. 6, p. 1390–406, 2018. DOI:<https://doi.org/10.1093/molbev/msy028>.

HUGHES, G.M.; BOSTON, E.S.M.; FINARELLI, J.A.; MURPHY, W.J.; HIGGINS, D.G.; TEELING, E.C. The Birth and Death of olfactory receptors gene families in mammalian niche adaptation. **Mol. Biol. Evol.**, v. 35, n. 6, p. 1390–1406, 2018.

IRISSARRI, I.; VENCES, M.; MAURO, D.S.; GLAW, F.; ZARDOYA, R. Reversal to air-driven sound production revealed by a molecular phylogeny of tongueless frogs, family Pipidae. **BMC Evol. Biol.**, v. 27, n. 11, p. 114–124, 2011.

- JANE, T. Development of central respiratory control in anurans: the role of neurochemicals in the emergence of air-breathing and the hypoxic response. **RSPNB**, v. 270, p. 103266, 2019. DOI:<https://doi.org/10.1016/j.resp.2019.103266>.
- JANVIER, P. Living primitive fishes and fishes from deep time. *In*: MCKENZIE, D.J.; FARREL, A.P.; BRAUNER, C.J. (Eds). Primitive fishes. Londo: Elsevier, 2007, p. 53.
- JIA, J.; XIE, Y.; CHENG, J.; KONG, C.; WANG, M.; GAO, L.; ZHAO, F.; GUO, J.; WANG, K.; LI, G.; CUI, D. Homology-mediated inter-chromosomal interactions in hexaploid wheat lead to specific subgenome territories following polyploidization and introgression. **Genome Biol.**, v. 22, n. 1, p. 1–21, 2021. DOI:10.1186/s13059-020-02225-7.
- JOGER, U. Zur herpetofaunistik kameruns (II) [On the herpetofaunistics of Camaroon (II)]. **Bonner Zoologische Beltrage**, v. 33, p. 313–342, 1982.
- JOHNSTONE, K.A.; LUBIENIECKI, K.P.; KOOP, B.F.; DAVIDSON, W.S. Identification of olfactory receptor genes in Atlantic salmon *Salmo salar*. **J. Fish. Biol.**, v. 81, n. 2, p. 559e575, 2012.
- JORGENSEN, C.B. Water economy in a terrestrial toad (*Bufo bufo*), with special reference to cutaneous drinking and urinary bladder function. **Comp. Biochem. Physiol.**, v. 109A2, p. 311–324, 1994.
- JUNGBLUT, L.D.; POZZI, A.G.; PAZ, D.A. A putative functional vomeronasal system in anuran tadpoles. **J. Anat.**, v. 221, p. 364–372, 2012.
- JUNGBLUT, L.D.; REISS, J.O.; POZZI, A.G. Olfactory subsystems in the peripheral olfactory organ of anuran amphibians. **Cell. Tissue Res.**, v. 383, n. 1, p. 289–299, 2021.
- KAJIYA, K.; INAKI, K.; TANAKA, M.; HAGA, T.; KATAOKA, H.; TOUHARA, K. Molecular bases of odor discrimination: reconstitution of olfactory receptors that recognize overlapping sets of odorants. **J. Neurosci.**, v. 21, n. 16, p. 6018e6025, 2001.
- KALYAANAMOORTHY, S.; MINH, B.Q.; WONG, T.K.F.; HAESLER, A.; JERMIIN, L.S. ModelFinder: fast model selection for accurate phylogenetic estimates. **Nat. Methods**, v. 14, p. 587–589, 2017.
- KAMBERE, M.B.; LANE, R.P. Co-regulation of a large and rapidly evolving repertoire of odorant receptor genes. **BMC Neurosci.**, v. 8, n. S2, p. 1–16, 2007.
- KATOH, K.; MISAWA, K.; KUMA, K.; MIYATA, T. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. **Nucl. Acids. Res.**, v. 30, n. 4, p. 3059–3066, 2002.

- KHAN, I.; YANG, Z.; MALDONADO, E.; LI, C.; ZHANG, G.; GILBERT, M.T.P.; JARVIS, E.D.; O'BRIEN, S.J.; JOHNSON, W.E.; ANTUNES, A. Olfactory receptor subgenomes linked with broad ecological adaptation in Sauropsida. **Mol. Bio. Evol.**, v. 32, p. 2832–2843, 2015.
- KOSMALA, G.K.; BROWN, G.P.; SHINE, R.; CHRISTINA, K. Skin resistance to water gain and loss has changed in cane toads (*Rhinella marina*) during their Australian invasion. **Ecol. Evol.**, v. 00, p. 1–9, 2020.
- KRAUTWURST, D.; YAU, K.W.; REED, R.R. Identification of ligands for olfactory receptors by functional expression of a receptor library. **Cell.**, v. 95, n. 7, p. 917–926, 1998.
- KRZYWINSKI, M.I.; SCHEIN, J.E.; BIROL, I.; CONNORS, J.; GASCOYNE, R.; HORSMAN, D.; JONES, S.J.; MARRA, M.A. Circos; an information aesthetic for comparative genomics. **Genome Res.**, v. 19, n. 9, p. 1639–1645, 2009.  
DOI:<https://doi.org/10.1101/gr.092759.109>.
- KUMAR, S.; STECHER, G.; LI, M.; KNYAZ, C.; TAMURA, K. MEGA X: molecular evolutionary genetics analysis across computing platforms. **Mol. Biol. Evol.**, v. 35, n. 6, p. 1547–1549, 2018.
- KURAKU, S.; ZMASEK, C.M.; NISHIMURA, O.; KATOH, K. aLeaves facilitates on-demand exploration of metazoan gene family trees on MAFFT sequence alignment server with enhanced interactivity. **Nucleic Acids. Res.**, v. 41, n. W1, p. W22–W28, 2013.
- KURIAN, S.M.; NARESSI, R.G.; MANOEL, D.; BARWICH, A.; MALNIC, B.; SARAIVA, L.R. Odor coding in the mammalian olfactory epithelium. **Cell and Tissue Res.**, v. 383, p. 445–456, 2021.
- LABERGE, F.; HARA, T.J. Neurobiology of fish olfaction: a review. **Brain Res. Rev.**, v. 36, p. 46–59, 2001.
- LANE, R.P.; CUTFORTH, T.; YOUNG, J.; ATHANASIOU, M.; FRIEDMAN, C.; ROWEN, L.; EVANS, G.; AXEL, R.; HOOD, L.; TRASK, B.J. Genomic analysis of orthologous mouse and human olfactory loci. **Proc. Nat. Acad. Sci.**, v. 98, p. 7390–7395, 2001.
- LANGHAM, R.J.; WALSH, J.; DUNN, M.; KO, C.; GOFF, S.A.; FREELING, M. Genomic duplication, fractionation and the origin of regulatory novelty. **Genetics**, v. 166, n. 2, p. 935–945, 2004. DOI: 10.1534/genetics.166.2.935.

LARKIN, D.M.; PAPE, G.; DONTU, R.; AUVIL, L.; WELGE, M.; LEWIN, H.A. Breakpoint regions and homologous synteny blocks in chromosomes have different evolutionary histories. **Genome Research**, v. 19, p. 770–777, 2009.

LEIVAS, P.T.; MOURA, M.O.; FÁVARO, L.F. The reproductive biology of the invasive *Lithobates catesbeianus* (Amphibia: Anura). **J. Herpeto.**, v. 46, n. 2, p. 153–161, 2012.

LI, J.; YU, H.; WANG, W.; FU, C.; ZHANG, W.; HAN, F.; WU, H. Genomic and transcriptomic insights into molecular basis of sexually dimorphic nuptial spines in *Leptobrachium leishanense*. **Nat. Commun.**, v. 10, p. 5551, 2019.

DOI:<https://doi.org/10.1038/s41467-019-13531-5>.

LI, W.; GODZIK, A. Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. **Bioinformatics**, v. 22, n. 13, p. 1658–1589, 2006.

LIBANTS, S.; CARR, K.; WU, H.; TEETER, J.H.; CHUNG-DAVIDSON, Y.W.; ZHANG, Z.; WILKERSON, C.; LI, W. The sea lamprey *Petromyzon marinus* genome reveals the early origin of several chemosensory receptor families in the vertebrate lineage. **BMC Evol. Biol.**, v. 9, n. 180, p. 1–14, 2009.

LIEDTKE, H.C.; GOWER, D.J.; WILKINSON, M.; GOMEZ-MESTRE, I. Macroevolutionary shift in the size of amphibian genomes and the role of life history and climate. **Nat. Ecol. Evol.**, v. 2, p. 1792–1799, 2018.

LIU, H.; CHEN, C.; LV, M.; LIU, N.; HU, Y.; ZHANG, H.; ENBODY, E.D.; GAO, Z.; ANDERSSON, L.; WANG, W. A chromosome-level assembly of blunt snout bream (*Megalobrama amblycephala*) genome reveals an expansion of olfactory receptor genes in freshwater fish. **Mol. Bio. Evol.**, v. 38, n. 10, p. 4238–4251, 2021.

LIU, S.; LIU, Y.; YANG, X.; TONG, C.; EDWARDS, D.; PARKIN, I.A.P.; ZHAO, M.; MA, J.; YU, J.; HUANG, S.; WANG, X.; WANG, J.; LU, K.; FANG, Z.; BANCROFT, I.; YANG, T.J.; HU, Q.; WANG, X.; YUE, Z.; LI, H.; YANG, L.; WU, J.; ZHOU, Q.; WANG, W.; KING, G.J.; PIRES, J.C.; LU, C.; WU, Z.; SAMPATH, P.; WANG, Z.; GUO, H.; PAN, S.; YANG, L.; MIN, J.; ZHANG, D.; JIN, D.; LI, W.; BELCRAM, H.; TU, J.; GUAN, M.; QI, C.; DU, D.; LI, J.; JIANG, L.; BATLEY, J.; SHARPE, A.G.; PARK, B.S.; RUPEARO, P.; CHENG, F.; WAMINAL, N.E.; HUANG, Y.; DONG, C.; WANG, L.; LI, J.; HU, Z.; ZHUANG, M.; HUANG, Y.; HUANG, J.; SHI, J.; MEI, D.; LIU, J.; LEE, T.H.; WANG, J.; JIN, H.; LI, Z.; LI, X.; ZHANG, J.; XIAO, L.; ZHOU, Y.; LIU, Z.; LIU, X.; QIN, R.; TNAG, X.; LIU, W.;

- WANG, Y.; ZHANG, Y.; LEE, J.; KIN, H.H.; DENOEUDE, F.; XU, X.; LIANG, X.; HUA, W.; WANG, X.; WANG, J.; CHALHOUB, B.; PATERSON, A.H. The *Brassica oleracea* genome reveals the asymmetrical evolution of polyploid genomes. **Nat. Commun.**, v. 5, n. 1, p. 3930, 2014. DOI:10.1038/ncomms4930.
- LU, B.; JIANG, J.; WU, H.; CHEN, X.; SONG, X.; LIAO, W.; FU, J. A large genome with chromosome-scale assembly sheds light on the evolutionary success of a true toad (*Bufo gargarizans*). **Mol. Ecol. Resour.**, v. 21, p. 1256–1273, 2021.
- LU, Q.; WANG, K.; LEI, F.; YU, D.; ZHAO, H. Penguins reduced olfactory receptor genes common to other waterbirds. **Scientific Reports**, v. 6, p. 1–9, 2016.
- LU, S.; YANG, D. A study of relationships among ranid frogs of the genera *Nanorana* and *Altirana* in the Transhimalaya Mountains of China. **Asiat. Herpetol. Res.**, v. 6, p. 73–77, 1995.
- LV, L.; LIANG, X.; HE, S. Genome-wide identification and characterization of olfactory receptor genes in Chinese Perch, *Siniperca chuatsi*. **Genes**, v. 10, n. 2, p. 178, 2019.
- LYNCH, J.D. The tadpoles of frogs and toads found in the lowlands of northern Colombia. **Rev. Acad. Colomb. Cienc.**, v. 30, n. 116, p. 443–457, 2006.
- MACHADO, C.F.; NAGAI, M.H.; LYRA, C.S.; REIS-SILVA, T.M.; XAVIER, A.M.; GLEZER, I.; FELICIO, L.F.; MALNIC, B. Conditional deletion of *Ric-8b* in olfactory sensory neurons leads to olfactory impairment. **J. Neurosci.**, v. 37, n. 50, p. 12202–12213, 2017.
- MAINLAND, J.D.; LUNDSTRÖM, J.N.; REISERT, J.; LOWE, G. From molecule to mind: an integrative perspective on odor intensity. **Trends Neurosci.**, v. 37, n. 8, p. 443–454, 2014.
- MAINLAND, J.D.; ZHOU, Y.R.L.; LIU, W.L.L.; MATSUNAMI, H. Human olfactory receptor responses to odorants. **Scientific Data**, v. 2, p. 1–9, 2015.
- MALNIC, B.; HIRONO, J.; SATO, T.; BUCK, L.B. Combinatorial receptor codes for odors. **Cell**, v. 96, p. 713–723, 1999.
- MAN, R.; GILAD, Y.; LANCET, D. Prediction of the odorant binding site of olfactory receptor proteins by human-mouse comparisons. **Protein Sci.**, v. 13, p. 240–254, 2004.

- MANZINI, I.; KORSCHING, S. The peripheral olfactory system of vertebrates: molecular, structural and functional basics of the sense of smell. **E-Neuroforum**, v. 2, p. 68–77.
- MANZINI, I.; SCHILD, D.; MENINI, A. Olfactory coding in larvae of the African Clawed Frog *Xenopus laevis*. In: SIMON, S.A.; NICOLELIS, M.A.L. (Eds). **Frontiers in Neuroscience**. Boca Raton: CRC Press, 2010.
- MARENCO, L.; WANG, R.; MCDUGAL, R.; OLENDER, T.; TWIK, M.; BRUFORD, E.; LIU, X.; ZHANG, J.; LANCET, D.; SHEPHERD, G.; CRASTO, C. ORDB, HORDE, ODORactor and other on-line knowledge resources of olfactory receptor-odorant interactions. **Database**, v, 2016, p. 1–10, 2016.
- MARTINEZ, Q.; COURCELLE, M.; DOUZERY, E.; FABRE, P. When morphology does not fit the genomes: the case of rodent olfaction. **Biol. Lett.**, v. 19, n. 4, p. 1–7, 2023.
- MATSUDA, Y.; UNO, Y.; KONDO, M.; GILCHRIST, M.J.; ZORN, A.M.; ROKHSAR, D.S.; SCHMID, M.; TAIRA, M. A new nomenclature of *Xenopus laevis* chromosomes based on the phylogenetic relationship to *Silurana/Xenopus tropicalis*. *Cytogenet. Genome Res.*, v. 145, p. 187–191, 2015.
- MATSUI, A.; GO, Y.; NIIMURA, Y. Degeneration of olfactory receptor gene repertoires in Primates: no direct link to full trichromatic vision. **Mol. Bio. Evol.**, v. 27, n. 5, p. 1192–1200, 2010.
- MCKENZIE, S.K.; KRONAUER, D.J.C. The genomic architecture and molecular evolution of odorant receptors. **Genome Research**, v. 28, p. 1757-1765, 2018.
- MEDEIROS, C.I.; BOTH, C.; KAEFER, I.L.; CECHIN, S.Z. Reproductive phenology of the American Bullfrog in subtropical Brazil: photoperiod as a main determinant of seasonal activity. **An. Acad. Bras. Cienc.**, v. 88, n. 3, p. 1909–1921, 2016.
- MEI, W.; LIU, S.; SCHNABLE, J.C.; YEH, C.T.; SPRINGER, N.M.; SCHNABLE, P.S.; BARBAZUK, W.B. A comprehensive analysis of alternative splicing in paleopolyploid maize. **Front. Plant Sci.**, v. 8, p. 694, 2017. DOI:10.3389/fpls.2017.00694.
- MENASHE, I.; LANCET, D. Variations in the human olfactory receptor pathway. *Cell Mol. Life Sci.*, v. 63, p. 1485–1493, 2006.

- MEREDITH, R.W.; GATESY, J.; EMERLING, C.A.; YORK, V.M.; SPRINGER, M.S. Rod monochromacy and the coevolution of cetacean retinal opsins. **PLOS Genetics**, v. 9, n. 4, p. 1–12, 2013.
- MEZLER, M.; FLEISCHER, J.; BREER, H. Characteristic features and ligand specificity of the two olfactory receptor classes from *Xenopus laevis*. **The Journal of Experimental Biology**, v. 204, p. 2987–2997, 2001.
- MOHRHARDT J.; NAGEL, M.; FLECK, D.; BEN-SHAUL, Y.; SPEHR, M. Signal of detection and coding in the accessory olfactory system. **Chem. Senses**, v. 43, n. 9, p. 667–695, 2018.
- MOMBAERTS, P. Genes and ligands for odorant, vomeronasal and taste receptors. **Nat. Rev. Neurosci.**, v. 5, p. 263–278, 2004.
- MUDD, A.B. **Comparative genomics and chromosome evolution**. Orientador: Daniel Rokhsar Chair. 2019. 65 f. Tese (Doutorado em Molecular and Cell Biology) - University of California, Berkeley, 2019.
- MURRELL, B.; WERTHEIM, J.O.; MOOLA, S.; WEIGHILL, T.; SCHEFFLER, K.; POND, S.L.K. Detecting individual sites subject to episodic diversifying selection. **PLOS Genetics**, v. 8, n. 7, p. 1–10, 2012.
- MUSILOVA, Z.; CORTESI, F.; MATSCHINER, M.; DAVIES, W.I.L.; PATEL, J.S.; STIEB, S.M.; BUSSEROLLES, F.; MALMSTROM, M.; TORRESEN, O.K.; BROWN, C.J.; MOUNTFORD, J.K.; HANEL, R.; STENKAMP, D.L.; JAKOBSEN, K.S.; CARLETON, K.L.; JENTOFT, S.; MARSHALL, J.; SALZBURGER, W. Vision using multiple distinct rod opsins in deep-sea fishes. **Science**, v. 364, n. 6440, p. 588–592.
- NEF, P.; HERMANS-BORGMEYER, I.; ARTIÈRES-PIN, H.; BEASLEY, L.; DIONNE, V.E.; HEINEMANN, S.F. Spatial pattern of receptor expression in the olfactory epithelium. **Proc Natl. Acad. Sci.**, v. 89, p. 8948–8952, 1992.
- NEF, S.; ALLAMAN, I.; FIUMELLI, H.; DE CASTRO, E.; NEF, P. Olfaction in Birds: differential embryonic expression. Of nine putative odorant receptor genes in the avian olfactory system. **Mech. Dev.**, v. 55, n. 1, p. 65–77, 1996.
- NEI, M.; ROONEY, A.P. Concerted and Birth-and-Death evolution of multigene families. **Annu. Rev. Genet.**, v. 39, p. 121–152, 2005.
- NEURATH, H. Über die leistung des geruchssinnes bei elritzen. **Z. Vgl. Physiol.**, v. 31, p. 609–626, 1949.

- NEVITT, G.A.; DITTMAN, A.H.; QUINN, T.P.; MOODY, W.J. Evidence for a peripheral olfactory memory in imprinted salmon. **Proc. Natl. Acad. Sci.**, v. 91, p. 4288–4292.
- NIEUWKOOP, P.D.; FABER, J. **Normal table of *Xenopus laevis* (Daudin): a systematical & chronological survey of the development from the fertilized egg til the end of metamorphosis**. New York: Garland Science, 1994.
- NIIMURA, Y. Evolutionary dynamics of olfactory receptor genes in chordates: interaction between environment and genomic contents. **Hum. Genomics.**, v. 4, n. 2, p. 107–118, 2009b.
- NIIMURA, Y. Identification of chemosensory receptor genes from vertebrate genomes. *In*: CRASTO, C.J. (Ed). **Olfactory receptors: methods and protocols**. Methods in Molecular Biology. New Jersey: Humana Totowa, 2013b, p. 95–105.
- NIIMURA, Y. Identification of olfactory receptor genes from mammalian genome sequences. *In*: CRASTO, C.J. (Ed). **Olfactory receptors: methods and protocols**. Methods in Molecular Biology. New Jersey: Humana Totowa, 2013a, p. 39–49.
- NIIMURA, Y. Olfactory receptor multigene family in vertebrates: from the viewpoint of evolutionary genomics. **Curr. Genomics.**, v. 13, p. 103–114, 2012.
- NIIMURA, Y. On the origin of evolution of vertebrate olfactory receptor genes: comparative genome analysis among 23 chordate species. **Genome Bio. Evol.**, v. 1, p. 34–44, 2009a.
- NIIMURA, Y.; MATSUI, A.; TOUHARA, K. Acceleration of olfactory receptor gene loss in primate evolution: possible link to anatomical change in sensory systems and dietary transition. **Mol. Biol. Evol.**, v. 35, n. 6, p. 1437–1450, 2018.
- NIIMURA, Y.; MATSUI, A.; TOUHARA, K. Extreme expansion of the olfactory receptor gene repertoire in African elephants and evolutionary dynamics of orthologous gene groups in 13 placental mammals. **Genome Res.**, v. 24, n. 9, p. 1485–1496, 2014.
- NIIMURA, Y.; NEI, M. Evolution of olfactory receptor genes in the human genome. **PNAS**, v. 100, n. 21, p. 12235–12240, 2003.
- NIIMURA, Y.; NEI, M. Evolutionary changes of the number of olfactory receptor genes in the human and mouse lineages. **Gene**, v. 346, p. 23–28, 2005a.
- NIIMURA, Y.; NEI, M. Evolutionary dynamics of olfactory and other chemosensory receptor genes in vertebrates. **J. Hum. Genet.**, v. 51, p. 505–517, 2006.

- NIIMURA, Y.; NEI, M. Evolutionary dynamics of olfactory receptor genes in fishes and tetrapods. **PNAS**, v. 102, n. 17, p. 6039–6044, 2005b.
- NIIMURA, Y.; NEI, M. Extensive gains and losses of olfactory receptor genes in mammalian evolution. **PLOS One.**, v. 2, n. 8, p. e708, 2007a.
- NIIMURA, Y.; NEI, M. Extensive gains and losses of olfactory receptor genes in mammalian evolution. **PLoS One**, v. 2, p. e708, 2007.
- NIIMURA, Y.; NEI, M. The evolution of animal chemosensory receptor gene repertoires: roles of chance and necessity. **Nat. Rev. Genet.**, v. 9, p. 951–963, 2008.
- NOZAWA, M.; NEI, M. Evolutionary dynamics of olfactory receptor genes in *Drosophila* species. **Proc. Natl. Acad. Sci.**, v. 104, p. 7122–7127, 2007.
- OKADA, Y. Frogs in Japan. **Copeia**, v.158, p. 161–166, 1927.
- OLENDER, T.; FELDMESSER, E.; ATAROT, T.; EISENSTEIN, M.; DORON, L. The olfactory receptor universe—from whole genome analysis to structure and evolution. **Genet. Mol. Res.**, v. 3, p. 545–553, 2004.
- OLENDER, T.; NATIV, N.; LANCET, D. HORDE: comprehensive resource for olfactory receptor genomics. **Methods Mol. Bio.**, v. 1003, p. 23–38, 2013.
- PALOMAR, G.; AHMAD, F.; VASEMAGI, A.; MATSUBA, C.; NICIEZA, A.G.; CANO, J.M. Comparative high-density linkage mapping reveals conserved genome structure but variation in levels of heterochiasmy and location of recombination cold spots in the Common frog. **G3**, v. 7, p. 637–645, 2017.
- PETERLIN, Z.; FIRESTEIN, S.; ROGERS, M. The state of the art of odorant receptor deorphanization: a report from the orphanage. **J. Gen. Physiol.**, v.143, n. 5, p. 527–542.
- PHILLIPPY, A.M. New advances in sequence assembly. **Genome Res.**, v. 27, p. 11–13, 2017.
- POND, S.L.K.; FROST, S.D.W. Not so different after all: a comparison of methods for detecting amino acid sites under selection. **Mol. Biol. Evol.**, v. 2, n. 5, p. 1208–1222, 2005.
- QIAN, J.; MANG, D.; LV, G.; YE, J.; LI, Z.; CHU, B.; SUN, L.; LIU, Y.; ZHANG, L. Identification and expression profile of olfactory receptor genes based on *Apriona germari* (Hope) antennal transcriptome. **Frontiers in Physiol.**, v. 11, p. 1–14, 2020.
- REISS, J.O.; EISTHEN, H.L. Comparative anatomy and physiology of chemical senses in amphibians. *In*: THEWISSEN, J.G.M. (Ed). **Sensory evolution on the threshold –**

**adaptations in secondarily aquatic vertebrates.** California: University of California Press, 2008, p. 42–63.

REN, X.; CHANG, S.; LAFRAMBOISE, A.; GREEN, W.; DUBUC, R.; ZIELINSKI, B. Projections from the accessory olfactory organ into the medial region of the olfactory bulb in the sea lamprey (*Petromyzon marinus*): a novel vertebrate sensory structure? **J. Comp. Neurol.**, v. 516, p. 105–116, 2009.

RENNY-BYFIELD, S.; RODGERS-MELNICK, E.; ROSS-IBARRA, J. Gene fraction ad function in the Ancient subgenomes of Maize. **Mol. Bio. Evol.**, v. 34, n. 8, p. 1825–1832, 2017. DOI:<https://doi.org/10.1093/molbev/msx121>

RODEL, M.O. Herpetofauna of West Africa: Amphibians of the western savannah. Germany: Frankfurt am Main, 2000.

RON, S.; NARVÁEZ, A.; ROMERO, G. Reproduction and spawning behavior in the frog, *Engystomops pustulatus* (Shreve 1941). **Amphib. Reptile Conserv.**, v. 8, p. 25–32, 2014.

RUBANOV, L.I.; SELIVERSTOV, A.V.; ZVERKOV, O.A.; LYUBETSKY, A. A method for identification of highly conserved elements and evolutionary analysis of superphylum Alveolata. **BMC Bioinformatics**, v. 17, n. 385, p. 1–16, 2016.

RYAN, M.J. The Túngara Frog: a study in sexual selection and communication. Chicago: University of Chicago Press, 1985.

RYAN, M.J.; BERNAL, X.E.; RAND, A.S. Patterns of matting call preferences in túngara frogs, *Physalameus pustulosus*. **J. Evol. Bio.**, v. 20, n. 6, p. 2235–2247, 2007.

RYAN, M.J.; RAND, A.S. Mate recognition, in túngara frogs: a review of some studies of brain, behavior, and evolution. **Acta. Zool. Sin.**, v. 49, p. 721–726, 2003.

SAITO, H.; CHI, Q.; ZHUANG, H.; MATSUNAMI, H.; MAINLAND, J.D. Odor coding by a mammalian receptor repertoire. **Sci Signal**, v. 2, n. 60, p. ra9, 2009.

SAITO, S.; OHKITA, M.; SAITO, C.T.; TAKAHASHI, K.; TOMINAGA, M.; OHTA, T. Evolution of heat sensors drove shifts in thermosensation between *Xenopus* species adapted to different thermal niches. **J. Biol. Chem.**, v. 291, n. 21, p. 11446–11459, 2016.

SAITOU, N.; NEI, M. The neighbor-joining method: a new method for Reconstructing Phylogenetic Trees. **Mol. Bio. Evol.**, v. 4, n. 4, p. 406–425, 1987.

SANABRIA, E.A.; QUIROGA, L.B.; ACOSTA, J.C. Introducción de *Rana catesbeiana* (rana toro), em ambientes precordilleranos de la provincia de San Juan, Argentina.

**Multequina**, v. 14, p. 65–68, 2005.

SARAIVA, L.R.; KORSCHING, S.I. A novel olfactory receptor gene family in teleost fish. **Genome Res.**, v. 17, n. 10, p. 1448–1457, 2007.

SCHIAVINATO, M.; BODRUG-SCHEPERS, A.; DOHM, J.C.; HIMMELBAUER, H. Subgenome evolution in allotetraploid plants. **Plant J.**, v. 106, n. 3, p. 672–688, 2021. DOI:10.1111/tpj.15190.

SCHMIDT, R. Synteny – recent advances and future prospects. **Curr. Opin. Plant. Biol.**, v. 3, p. 97–102, 2000.

SCHOEGEL, L.M.; DASZAK, P.; CUNNINGHAM, A.A.; SPEARE, R.; HILL, B. Two amphibian diseases, chytridionmycosis and ranaviral disease, are now globally notifiable to the World Organization for Animal Health (OIE): an assessment. **Dis. Aquat. Org.**, v. 92, n. 2–3, p. 101–108, 2010.

SCHOTT, R.K.; PEREZ, L.; KWIATKOWSKI, M.A.; IMHOFF, V.; GUMM, J.M. Evolutionary analyses of visual opsin genes in frogs and toads: diversity, duplication, and positive selection. **Ecology and Evolution**, v. 12, n. 2, p. e8595, 2022.

SEEBACHER, F.; ALFORD, R.A. Movement and microhabitat use of a terrestrial amphibian (*Bufo marinus*) on a tropical island: seasonal variation and environmental correlates. **J. Herpetol.**, v. 33, n. 2, p. 208–214, 1999.

SEGERMAN, B. The most frequently used sequencing technologies and assembly methods in different time segments of the bacterial surveillance and RefSeq genome databases. **Front. Cell. Infect. Microbiol.**, v. 10, p. 527102, 2020.

SERIZAWA, S.; ISHII, T.; NAKATANI, H.; TSUBOI, A.; NAGAWA, F.; ASANO, M.; SUDO, K.; SAKAGAMI, J.; SAKANO, H.; IJIRI, T.; MATSUDA, Y.; SUZUKI, M.; YAMAMORI, T.; IWAKURA, Y.; SAKANO, H. Mutually exclusive expression of odorant receptor transgenes. **Nature Neuroscience**, v. 3, p. 687–693, 2000.

SESSION, A.M.; UNO, Y.; KWON, T.; CHAPMAN, J.A.; TOYODA, A.; TAKAHASHI, S.; FUKUI, A.; HIKOSAKA, A.; SUZUKI, A.; KONDO, M.; VAN HEERINGEN, S.J.; QUIGLEY, I.; HEINZ, S.; OGINO, H.; OCHI, H.; HELLSTEN, U.; YONS, J.B.; SIMAKOV, O.; PUTNAM, N.; STITES, J.; KUROKI, Y. ET AL. Genome evolution in the allotetraploid frog *Xenopus laevis*. **Nature**, v. 538, p. 336–343, 2016.

SESSIONS, S.K. Evolutionary cytogenetics in salamanders. **Chromosome Res.**, v. 16, p. 183–201, 2008.

SHEPHERD, G.M. Discrimination of molecular signals by the olfactory receptor neuron. **Neuron**, v. 13, p. 771–790, 1994.

SIDDIQI, A.; CRONIN, T.W.; LOEW, E.R.; VOROBYEV, M.; SUMMERS, K.

Interspecific and intraspecific of color signals in the strawberry poison frog

*Dendrobates pumilio*. **Journal of Experimental Biology**, v. 207, p. 2471–2485, 2004.

SIEPEL, A.; BEJERANO, G.; PEDERSEN, J.S.; HINRICHS, A.S.; HOU, M.; ROSENBLOOM, K.; CLAWSON, H.; SPIETH, J.; HILLIER, L.W.; RICHARDS, S.; WEINSTOCK, G.M.; WILSON, R.K.; GIBBS, R.A.; KENT, W.J.; MILLER, W.; HAUSSLER, D. Evolutionarily conserved elements in vertebrate, insect, worm, and yeast genomes. **Genome**, v. 15, p. 1034–1050, 2005.

SILVA, E.T.; REIS, E.P.; FEIO, R.N.; FILHO, O.P.R. Diet of the invasive frog

*Lithobates catesbeianus* (Shaw, 1802) (Anura: Ranidae) in Viçosa, Minas Gerais state, Brazil. **S. Am. J. Herpetol.**, v. 4, n. 3, p. 286–294, 2009.

SILVA, M.C.; CHIBUCOS, M.; MUNRO, J.B.; DAUGHERTY, S.; COELHO, M.M.;

SILVA, J.C. Signature of adaptive evolution in olfactory receptor genes in Cory's

Shearwater supports molecular basis for smell in procellariiform seabirds. **Sci. Rep.**, v.

10, n. 543, p. 1–11, 2020. DOI:<https://doi.org/10.1038/s41598-019-56950-6>.

SIN, S.Y.W.; CLOUTIER, A.; NEVITT, G.; EDWARDS, S.V. Olfactory receptor

subgenome and expression in a highly olfactory procellariiform seabird. **Genetics**, v.

220, n. 2, p. 1–16, 2022. DOI:<https://doi.org/10.1093/genetics/iyab210>.

SMITH, J.J.; VOSS, S.R. Gene order data from a model amphibian (*Ambystoma*): new perspectives on vertebrate genome structure and evolution. **BMC Bioinformatics**, v. 7, n. 219, p. 1–12, 2006.

STAMATAKIS, A. RAxML Version 8: A tool for phylogenetic analysis and post-

analysis of large phylogenies. **Bioinformatics**, v. 30, n. 9, p. 1312–1313, 2014.

STEIGER, S.S.; FIDLER, A.E.; KEMPENAERS, B. Evidence for increased olfactory

receptor gene repertoire size in two nocturnal bird species with well-developed

olfactory ability. **BMC Evol. Biol.**, v. 9, p. 1–11, 2009.

STEIGER, S.S.; KURYSHEV, V.Y.; STENSMYR, M.C.; KEMPENAERS, B.;

MUELLER, J.C. A comparison of reptilian and avian olfactory receptor gene

repertoires: Species-specific expansion of group  $\gamma$  in birds. **BMC Genomics**, v. 10, n.

446, p. 1–10, 2009. DOI:<https://doi.org/10.1186/1471-2164-10-446>.

STODDART, D.M. **The ecology of vertebrate olfaction**. London: Chapman and Hall, 1980.

STREICHER, J.W. The genome sequence of the common toad, *Bufo bufo* (Linnaeus, 1758). **Wellcome Open Res.**, v. 6, p. 281, 2021.

SUDA, K.; HAYASHI, S.R.; TAMURA, K.; TAKAMATSU, N. ITO, M. Activation of DNA transposons and evolution of piRNA genes through interspecific hybridization in *Xenopus* frogs. **Front. Genet.**, v. 13, p. 1–9, 2022.

SUN, Y.; XIONG, Z.; XIANG, X.; LIU, S.; ZHOU, W.; TU, X.; ZHONG, L.; WANG, L.; WU, D.; ZHANG, B.; ZHU, C.; YANG, M.; CHEN, H.; LI, F.; ZHOU, L.; FENG, S.; HUANG, C.; ZHANG, G.; IRWIN, D.; HILLIS, D.M.; MURPHY, R.W.; YANG, H.; CHE, J.; WANG, J.; ZHANG, Y. Whole-genome sequence of the Tibetan frog *Nanorana parkeri* and the comparative evolution of tetrapod genomes. **PNAS**, v. 112, n. 11, p. E1257–E1262, 2015.

TAKEZAKI, N.; RZHETSKY, A.; NEI, M. Phylogenetic test of the molecular clock and linearized trees. **Mol. Bio. Evol.**, v. 12, n. 5, p. 823–833, 1995. DOI: 10.1093/oxfordjournals.molbev.a040259.

TANG, H.; LYONS, E.; PEDERSEN, B.; SCHNABLE, J.C.; PATERSON, A.H.; FREELING, M. Screening synteny blocks in pairwise genome comparisons through integer programming. **BMC Bioinformatics**, v. 12, n. 102, p. 1–11, 2011.

TANIGUCHI, K.; SAITO, S.; OIKAWA, T.; TANIGUCHI, K. Phylogenic aspects of the amphibian dual olfactory system. **The Journal of Veterinary Medical Science**, v. 70, n. 1, p. 1–9, 2007.

TANIGUCHI, K.; SAITO, S.; TANIGUCHI, K. Phylogenic outline of the olfactory system in vertebrates. **J. Vet. Med. Sci.**, v. 73, p. 139–147, 2011.

TEICHMAN, H. Über die leistung des gurechssinnes beim Aal (*Anguilla anguilla*). **L. Z. Vgl. Physiol.**, v. 42, p. 206–254.

TERAKITA, A. The opsins. **Genome Biol.**, v.6, n. 3, p. 213, 2005.

TINSLEY, R.C.; LOUMONT, C.; KOBEL, H.R. Geographical distribution and ecology. *In*: TINSLEY, R.C., KOBEL, H.R. (Eds). **The Biology of Xenopus**. Oxford: University Press, 1996.

TOUHARA, K. Deorphanizing vertebrate olfactory receptors: recent advances in odorant-response assays. **Neurochem. Int.**, v. 51, v. 2–4, p. 132–139, 2007.

TOUHARA, K. Structure, expression, and function of olfactory receptors. *In*: FIRESTEIN, S.; BEAUCHAM, G.K. (Eds). **The Senses: A comprehensive reference**. New York: Academic Press, 2008.

TOUHARA, K.; NIIMURA, Y.; IHARA, S. Chapter 3 – Vertebrate odorant receptors. *In*: ZUFALL, F.; MUNGER, S.D. (Eds). **Chemosensory transduction: the detection of odors, tastes, and other chemostimuli**. Florida: Academic Press, 2016.

VAN DER LINDEN, C.; JAKOB, S.; GUPTA, P.; DULAC, C.; SANTORO, S.W. Sex separation induces differences in the olfactory sensory repertoires of male and female mice. **Nat. Commun.**, v. 4, n. 9, p. 5081, 2018.

VAN VALKENBURGH, B.; CURTIS, A.; SAMUELS, J.X.; BIRD, D.; FULKERSON, B.; MEACHEN-SAMUELS, J.; SLATER, G.J. Aquatic adaptations in the nose of carnivorans: evidence from the turbinates. **J. Anat.**, v. 218, p. 298–310, 2011.

VANDEWEGE, M.W.; MANGUM, S.F.; GABALDÓN, T.; CASTOE, T.A.; RAY, D.A.; HOFFMAN, F.G. Contrasting patterns of evolutionary diversification in the olfactory repertoires of reptile and bird genomes. **Genome Biol. Evol.**, v. 8, n. 3, p. 470–480, 2016. DOI:<https://doi.org/10.1093/gbe/evw013>.

VIEYRA, M.L. Olfactory receptor genes in terrestrial, freshwater, and sea turtles: evidence for a reduction in the number of functional genes in aquatic species. **Chelonian Conserv. Biol.**, v. 10, n. 2, p. 181–187, 2011.

VON DANNECKER, L.E.C.; MERCADANTE, A.F.; MALNIC, B. Ric-8B, an olfactory putative GTP exchange factor, amplifies signal transduction through the olfactory-specific G-protein G $\alpha$ olf. **J. Neurosci.**, v. 25, n. 15, p. 3793–3800, 2005.

VOSSHALL, L.B. Laying a controversial smell theory to rest. **Proc. Natl. Acad. Sci.**, v. 112, p. 6525–6526, 2015.

WANG, H.; CHEN, L.; DONG, C.; CHEN, B.; LI, B.; LI, X.; XU, P. Genome-wide identification and characterization of olfactory receptor genes in common carp (*Cyprinus carpio*). **Gene**, v. 777, p. 1–9, 2021.

WANG, K.; TIAN, S.; GALINDO-GONZÁLEZ, J.; DÁVALOS, L.M.; ZHANG, Y.; ZHAO, H. Molecular adaptation and convergent evolution of frugivory in Old World and neotropical fruit bats. **Mol. Biol.**, v. 29, p. 4366–4381, 2020.

WANG, Y.; SUN, Y.; JOSEPH, P.V. Diverse evolutionary rates and gene duplication patterns among families of functional olfactory receptor genes in humans. **PLOS One**, v. 18, n. 4, p. e0282575, 2023. DOI:<https://doi.org/10.1371/journal.pone.0282575>.

- WEISS, L.; JUNGBLUT, L.D.; POZZI, A.G.; ZIELINSKI, B.S.; O'CONNEL, L.A.; HASSENKLÖVER, T.; MANZINI, I. **J. Comp. Neurol.**, v. 528, p. 2239–2253.
- WEISS, L.; MANZINI, I.; HASSENKLÖVER, T. Olfaction across the water-air interface in anuran amphibians. **Cell and Tissue Res.**, v. 383, p. 301–325, 2021.
- WOLFE K.H. Yesterday's polyploids and the mystery of diploidization. **Nat. Rev. Genet.**, v. 2, n. 5, p. 333–341, 2001.
- YANG, S.; ARGUELLO, J.R.; LI, X.; DING, Y.; ZHOU, Q.; CHEN, Y.; ZHANG, Y.; ZHAO, R.; BRUNET, F.; PENG, L.; LONG, M.; WANG, W. Repetitive element-mediated recombination as a mechanism for new gene origination in *Drosophila*. **PLoS Genet.**, v. 4, p. e3, 2008.
- YE, M.; LIN, X.; ZHANG, Y.; HUANG, Y.; LI, G.; TIAN, C. Genome-wide identification and characterization of olfactory receptor genes in Silver Sillago (*Sillago sihama*). **Animals**, v. 13, n. 7, p. 1232–1248, 2023.
- YOHE, L. R.; LEISER-MILLER, L.B.; KALISZEWSKA, Z.A.; DONAT, P.; SANTANA, S.E.; DÁVALOS, L.M. Diversity in olfactory receptor repertoires is associated with dietary specialization in a genus of frugivorous bat. **G3**, v. 11, n. 10, p. jkab260, 2021.
- YOHE, L.R.; BRAND, P. Evolutionary ecology of chemosensation and its role in sensory drive. **Curr. Zool.**, v. 65, p. 525–533, 2018.
- YOHE, L.R.; FABBRI, M.; HANSON, M.; BHULLAR, B.S. Olfactory receptor gene evolution is unusually rapid across Tetrapoda and outpaces chemosensory phenotypic changes. **Curr. Zool.**, v. 66, n. 5, p. 505–514, 2020.
- YOHE, L.R.; FABBRI, M.; HANSON, M.; BHULLAR, B.S. Olfactory receptor gene evolution is unusually rapid across Tetrapoda and outpaces chemosensory phenotypic change. **Curr. Zool.**, v. 66, n. 5, p. 505–514, 2020.
- YOHE, L.R.; FABBRI, M.; LEE, D.; DAVIES, K.T.J.; YOHE, T.P.; SÁNCHEZ, M.K.R.; RENGIFO, E.M.; HALL, R.P.; MUTUMI, G.; HEDRICK, B.P.; SADIÉ, A.; SIMMONS, N.B.; SEARS, K.E.; DUMONT, E.; ROSSITER, S.J.; BHULLAR, B.A.; DÁVALOS, L.M. Ecological constraints on highly evolvable olfactory receptor genes and morphology in neotropical bats. **Evolution**, v. 76, n. 10, p. 2347–2360, 2022.

YOHE, L.R.; LEISER-MILLER, L.B.; KALISZEWSKA, Z.A.; DONAT, P.; SANTANA, S.S.; DÁVALOS, L.M. Diversity in olfactory receptor repertoires is associated with dietary specialization in a genus of frugivorous bat. **G3**, v. 11, n. 10, p. jkab260, 2021.

YOUNG, J.M.; FRIEDMAN, C.; WILLIAMS, E.M.; ROSS, J.A.; TONNES-PRIDDY, L.; TRASK, B.J. Different evolutionary processes shaped the mouse and human olfactory receptor gene families. **Hum. Mol. Genet.**, v. 11, n. 5, p. 535–546, 2002.

YOUNG, J.M.; SHYKIND, B.M.; LANE, R.P.; TONNES-PRIDDY, L.; ROSS, J.A.; WALKER, M.; WILLIAMS, E.M.; TRASK, B.J. Odorant receptor expressed sequence tags demonstrate olfactory expression of over 400 genes, extensive alternative splicing and unequal expression levels. **Genome Biol.**, v. 4, n. 11, p. R71, 2003. DOI:<https://doi.org/10.1186/gb-2003-4-11-r71>.

YOUNG, J.M.; TRASK, B.J. The sense of smell: genomics of vertebrate odorant receptors. **Hum. Mol. Genet.**, v. 11, n. 10, p. 1153–1160, 2002.

ZHANG, D.; GAO, F.; JAKOVLIĆ, I.; ZOU, H.; ZHANG, J.; LI, W.X.; WANG, G.T. PhyloSuite: an integrated and scalable desktop platform for streamlined molecular sequence data management and evolutionary phylogenetics studies. **Mol. Ecol. Resour.**, v. 20, n. 1, p. 348–355, 2020.

ZHANG, X.; FIRESTEIN, S. The olfactory receptor gene superfamily of the mouse. **Nat. Neurosci.**, v. 5, n. 2, p. 124–133, 2002.

ZHANG, Y.; HOON, M.A.; CHANDRASHEKAR, J. MUELLER, K.L.; COOK, B.; WU, C.; ZUKER, C.S.; RYBA, J.P. Coding sweet, bitter, and umami tastes: different receptor cells sharing similar signaling pathways. **Cell**, v. 112, p. 293–301, 2003.

ZHAO, T.; SCHRANZ, E.M. Network-Based Microsynteny Analysis Identifies Major Differences and Genomic Outliers in Mammalian and Angiosperm Genomes. **Proceedings of the National Academy of Sciences**, v. 116, n. 6, p. 2165–2174, 2019.

ZHAO, T.; ZWAENEPOEL, A.; XUE, J.Y.; KAO, S.M.; LI, Z.; SCHRANZ, M.E.; PEER, Y.V. Whole-Genome Microsynteny-Based Phylogeny of Angiosperms. *Nature Communications*, v. 12, n. 1, p. 1–14, 2021.

ZHAO, Y.; DONG, L.; JIANG, C.; WANG, X.; XIE, J.; RASHID, M.A.R.; LIU, Y.; LI, M.; BU, Z.; WANG, H.; MA, X.; SUN, S.; WANG, X.; BO, C.; ZHOU, T.; KONG,

L. Distinct nucleotide patterns among three subgenomes of bread wheat and their potential origins during domestication after allopolyploidization. **BMC Biol.**, v. 18, n. 1, p. 188, 2020. DOI:10.1186/s12915-020-00917-x.

ZHOU, C.; LIU, Y.; ZHENG, X.; SHANG, K.; CHENG, M.; WANG, L.; YANG, N.; YUE, B. Characterization of olfactory receptor repertoires provides insights into the high-altitude adaption of the yak based on the chromosome-level genome. **Int. J. Biol. Macromol.**, v. 209, p. 220–230, 2022.

ZHU, G.; WANG, L.; TANG, W.; WANG, X.; WANG, C. Identification of olfactory receptor genes in the Japanese grenadier anchovy *Coilia nasus*. **Genes Genom.**, v. 39, p. 521–532, 2017.

ZOZULYA, S.; ECHEVERRI, F.; NGUYEN, T. The human olfactory receptor repertoire. **Genome Biol.**, v. 1, n. 6, p. research0018.1, 2001.