

UNIVERSIDADE FEDERAL DO PARANÁ

FERNANDA LINO DE SOUZA

ANÁLISE COMPARATIVA DE DNAs SATÉLITES EM *XENOPUS TROPICALIS* E
PIPA CARVALHOI.

CURITIBA

2024

FERNANDA LINO DE SOUZA

ANÁLISE COMPARATIVA DE DNAS SATÉLITES EM *XENOPUS TROPICALIS* E
PIPA CARVALHOI.

Monografia apresentada ao curso de Graduação
em Ciências Biológicas, Setor de Ciências
Biológicas, Universidade Federal do Paraná, como
requisito parcial à obtenção do título de Bacharel
em Ciências Biológicas

Orientador: Prof. Dr. Kaleb Pretto Gatto

CURITIBA

2024

AGRADECIMENTOS

Sou grata a Universidade Federal do Paraná que desde 2019 me acolheu e se tornou minha segunda casa, no qual pude me desenvolver pessoal e profissionalmente, me ofertando diferentes oportunidades de crescimento. Agradeço aos professores que me deram todo suporte educacional e emocional para finalizar um ciclo, no qual encerro com a bagagem repleta de conhecimentos com sede de expor para o mundo tudo que aqui me foi ensinado.

Ao meu orientador, Prof. Dr. Kaleb Pretto Gatto, que me conduziu nesta jornada, com todo o apoio para alcançar o meu melhor. Aos meus amigos, que durante toda a graduação estavam ao meu lado, nas dificuldades e nas alegrias, principalmente a Rafaela Caroline Balduino, que desde o primeiro dia de aula jamais me deixou sozinha, e ao Henrique Gonçalves por compartilhar comigo medos e incertezas, mas sempre com pensamento positivo que no final tudo daria certo. Sou grata ao meu namorado, Willian, por em momentos onde pensei que não aguentaria, me deu o suporte que eu precisava para seguir em frente, incentivando para meu crescimento e aplaudindo as minhas conquistas..

Aos meus familiares, no qual expresso o maior amor e gratidão por terem tornado a finalização deste ciclo possível. Apesar de todas as dificuldades enfrentadas ao longo destes anos, jamais me desampararam e me deixaram desistir do meu sonho. Dedico especialmente ao meu pai, Mauro (*in memoriam*), que infelizmente não pode acompanhar o final desta trajetória, mas que enquanto pode, investiu na minha educação e a minha mãe, Rosi, que mesmo com suas próprias lutas, esteve ao meu lado. Aos meus irmãos, João Pedro e Mariana, que compreenderam a minha ausência em determinados momentos para que este momento enfim chegasse. E aos meus filhos, Maria Alice e Gabriel, que eram a minha maior alegria em meio a correria.

Agradeço a PRAE, pois, sem o auxílio estudantil, a continuação e finalização da graduação não seria possível. Levo comigo não somente os conhecimentos adquiridos, mas amizades e experiências dos melhores anos da minha vida.

A educação é a única coisa que ninguém jamais poderá tirar de você.
(Souza, Mauro, 2019)

RESUMO

DNAs satélites (DNAsats) estão dispostos em tandem no genoma de organismo eucariotos, organizados em monômeros de mais de 100bp e localizados em regiões de heterocromatina constitutiva. Espécies próximas filogeneticamente, tendem a compartilhar DNAs satélites, com diferenças no número de cópias ou na estrutura, refletindo na dinâmica evolutiva de cada espécie ao longo do tempo. O presente trabalho teve como objetivo geral a elaboração de testes para a hipótese da biblioteca de DNAs satélites entre espécies divergentes da família Pipidae (*Xenopus* e *Pipa*), tendo as espécies *Xenopus tropicalis* e *Pipa carvalhoi* como alvo de pesquisa. Os genomas de *X. tropicalis* e *P. carvalhoi* foram analisados para identificar DNAs satélites. As reads provenientes do genoma de *P. carvalhoi* e de *X. tropicalis* foram analisadas no pipeline *RepeatExplorer2* para identificar *clusters* com características de sequências de DNAsats. Em *X. tropicalis* 13 contigs finais de DNAsats foram identificados após etapas de filtragem para falsos positivos. Já para *P. carvalhoi*, foram identificados 9 contigs finais. Nas análises foram obtidos 2 satélites compartilhados entre as duas espécies, mas com baixa taxa de proporção e de identidade entre eles. Para o restante dos DNAs satélites específicos identificados em cada espécie, variações foram observadas para porcentagem de identidade entre repeats da mesma família, no número de cópias presentes nas montagens do genoma e para a proporção genômica ocupada por esses DNAs satélites. Especialmente no caso de *X. tropicalis*, houve variação na localização dos DNAsats entre os cromossomos. Devido a isso, foram observados tanto em *Pipa carvalhoi* quanto em *Xenopus tropicalis* diferentes padrões de homogeneização das famílias de DNAsat. Análises filogenéticas e mapeamento citogenético em estudos futuros podem ajudar a esclarecer os mecanismos de evolução que impactam a diversidade de DNAsats entre espécies.

Palavras-chave: DNA satélite; Sequenciamento de nova geração; Anura; Pipidae; Genoma.

ABSTRACT

Satellite DNAs (satDNAs) are arranged in tandem in the genome of eukaryotic organisms, organized in monomers of more than 100bp and located in regions of constitutive heterochromatin. Phylogenetically close species tend to share satellite DNAs, with differences in copy number or structure, reflecting the evolutionary dynamics of each species over time. The general aim of this study was to test the hypothesis of a satellite DNA library between divergent species of the Pipidae family (*Xenopus* and *Pipa*) and the species *Xenopus tropicalis* and *Pipa carvalhoi* as the research targets. The genomes of *X. tropicalis* and *P. carvalhoi* were analyzed to identify satellite DNAs. The reads from the genomes of *P. carvalhoi* and *X. tropicalis* were analyzed in the RepeatExplorer2 pipeline to identify clusters with the characteristics of satellite DNA sequences. In *X. tropicalis*, 13 final DNAsats contigs were identified after filtering for false positives. For *P. carvalhoi*, 9 final contigs were identified. In the analyses, 2 satellites shared between the two species were obtained, but with a low proportion and identity rate between them. For the rest of the specific satellite DNAs identified in each species, variations were observed in the percentage of identity between repeats from the same family, in the number of copies present in the genome assemblies and in the genomic proportion occupied by these satellite DNAs. Especially in the case of *X. tropicalis*, there was variation of location of the satDNAs amongst the chromosomes. Because of this, different patterns of homogenization of the satDNAs families were observed in both *Pipa carvalhoi* and *Xenopus tropicalis*. Phylogenetic analyses and cytogenetic mapping in future studies may help to clarify the mechanisms of evolution that impact the diversity of satDNAs between species.

Keywords: Satellite DNA; Next generation sequencing; Anura; Pipidae; Genome.

LISTA DE FIGURAS

FIGURA 1 - ÁRVORE FILOGENÉTICA PIPIDAE.....	9
FIGURA 2 - ESPÉCIME DE <i>Pipa Carvalhoi</i>	10
FIGURA 3 - ESPÉCIME DE <i>Xenopus tropicalis</i>	10
FIGURA 4 - PROPORÇÃO DE READS DE DNAs REPETITIVOS DE <i>X. tropicalis</i> ...	13
FIGURA 5 - READS DE <i>X. tropicalis</i>	14
FIGURA 6 - ALINHAMENTO DE UNIDADE MONOMÉRICA CANÔNIMA DO DNA SATÉLITE CENTROMÉRICO DE <i>Xenopus tropicalis</i>	18
FIGURA 7 - PROPORÇÃO DE READS DE DNAs REPETITIVOS DE <i>P. carvalhoi</i> ...	25
FIGURA 8 - READS DE <i>P. carvalhoi</i>	26
FIGURA 9 - DNAs SATÉLITES COMPARTILHADOS ENTRE <i>X. tropicalis</i> e <i>P. carvalhoi</i>	26
FIGURA 10 - GRÁFICO DOTPLOT DE ALINHAMENTOS DE SCAFFOLDS DO GENOMA DE <i>P. carvalhoi</i> QUE CONTÉM SEQUÊNCIAS DE DNAs SATÉLITES....	28

LISTA DE GRÁFICOS

GRÁFICO 1 - DISTRIBUIÇÃO DOS DNASAT XtrSat04-206 E XtrSat07-1299.....	17
GRÁFICO 2 - DISTRIBUIÇÃO DOS DNASAT XtrSat01-1030 E XtrSat02-1609.....	19
GRÁFICO 3 - DISTRIBUIÇÃO DOS DNASAT XtrSat03-276 E XtrSat05-544.....	20
GRÁFICO 4 - DISTRIBUIÇÃO DOS DNASAT XtrSat12-181, XtrSat11-702 E XtrSat13-564.....	21
GRÁFICO 5 - DISTRIBUIÇÃO DOS DNASAT XtrSat06-481, XtrSat08-2.290, XtrSat09-1189 E XtrSat10-454.....	22
GRÁFICO 6 - ANÁLISE DE SIMILARIDADE DE DNASAT EM <i>X. tropicalis</i>	23
GRÁFICO 7 - ANÁLISE DE SIMILARIDADE DE DNASAT EM <i>P. carvalhoi</i>	29

LISTA DE TABELAS

TABELA 1 - DISTRIBUIÇÃO CROMOSSÔMICA, NÚMEROS DE CÓPIAS E NOMENCLATURA DE DNAs SATÉLITES CARACTERIZADOS PARA <i>Xenopus tropicalis</i>	14
TABELA 1 - DISTRIBUIÇÃO CROMOSSÔMICA, NÚMEROS DE CÓPIAS E NOMENCLATURA DE DNAs SATÉLITES CARACTERIZADOS PARA <i>Pipa carvalhoi</i>	27

SUMÁRIO

1	INTRODUÇÃO.....	1
2	OBJETIVOS.....	2
2.1	Objetivo geral.....	2
2.2	Objetivo específico.....	2
3	REVISÃO BIBLIOGRÁFICA.....	2
3.1	DNA repetitivo.....	3
3.2	DNAs satélites.....	5
3.3	Anfíbios Anuros da família Pipidae.....	8
4	MATERIAIS E MÉTODOS.....	10
4.1	Aquisição dos dados e sequenciamento do genoma de <i>P. carvalhoi</i>	11
4.2	Caracterização <i>de novo</i> de DNAs satélites.....	11
4.3	Mapeamento <i>in silico</i> de DNAs satélites em <i>X. tropicalis</i> e <i>P. carvalhoi</i>	12
4.4	Analise de similaridade de DNAs satélites.....	12
5	RESULTADOS.....	12
5.1	DNAs satélites no genoma de <i>Xenopus tropicalis</i>	13
5.2	DNAs satélites no genoma de <i>Pipa carvalhoi</i>	23
6	DISCUSSÃO.....	29
6.1	Perfis de DNAs satélites de <i>X. tropicalis</i> e <i>P. carvalhoi</i>	29
6.2	Homogeneidade X heterogeneidade dos DNAs satélites de <i>X. tropicalis</i> e <i>P. carvalhoi</i>	32
7	CONCLUSÃO.....	35
	REFERÊNCIAS.....	36
	ANEXO 1 – DISTRIBUIÇÃO DE VALORES DE BITScore DAS BUSCAS POR BLAST DAS SEQUÊNCIAS CONSENSO DE DNA SATÉLITES NA MONTAGEM DO GENOMA DE <i>X. TROPICALIS</i>.....	42
	ANEXO 2 – DISTRIBUIÇÃO DE VALORES DE BITScore DAS BUSCAS POR BLAST DAS SEQUÊNCIAS CONSENSO DE DNA SATÉLITES NA MONTAGEM DO GENOMA DE <i>P. CARVALHOI</i>.....	43

1. INTRODUÇÃO

DNAs repetitivos compõem uma porção considerável dos genomas de eucariotos, sendo que mais de 40% do genoma humano é composto por sequências repetitivas (IHGSC, 2001; López-Flores e Garrido-Ramos, 2012). Dentre os DNAs repetitivos em eucariotos há aqueles que são classificados como dispersos pelo genoma, como por exemplo os elementos de transposição (TEs) e algumas famílias multigênicas como dos genes de RNA transportador, e os DNAs repetidos *in tandem*, onde as unidades repetitivas estão umas justapostas as outras. Dentre os DNAs repetitivos *in tandem* estão abarcadas algumas famílias multigênicas, como aquelas de RNAs ribossomais, e os DNAs microssatélites, minissatélites e satélites (Charlesworth et al., 1994; López-Flores e Garrido-Ramos, 2012; Plohl et al., 2012).

DNAs satélites em geral estão localizados em regiões de heterocromatina constitutiva nos cromossomos de eucariotos, como nos centrômeros e nas regiões teloméricas (Plohl et al., 2012). Entretanto, algumas unidades repetitivas podem ocupar regiões eucromáticas em algumas espécies, como no besouro *Tribolium castaneum* (Feliciello et al., 2015), na mosca-das-frutas *Drosophila melanogaster* (Kuhn et al., 2012) e no anuro *Cycloramphus bolitoglossus* (Bueno et al., 2021). Sequências eucromáticas de DNAs satélites estão relacionadas com aspectos funcionais em diferentes genomas, pois podem modular a expressão de genes que estão próximos dessas unidades repetitivas (Feliciello et al., 2015). Além disso, DNAs satélites também participamativamente do estabelecimento de regiões de heterocromatina, bem como também participamativamente da organização e biologia dos centrômeros (Plohl et al., 2008; Pezer et al., 2012).

Geralmente, perfis de DNAs satélites são espécie específicos, ou então compartilhados apenas entre espécies muito próximas filogeneticamente (Plohl et al., 2012). São poucos os exemplos de DNAs satélites que são compartilhados entre espécies muito divergentes, como o satélite BIV160 em moluscos bivalves (Plohl et al., 2010). Entre anuros, os DNAs satélites são pouco estudados (Targueta et al. 2018), entretanto há muitas montagens de genomas e *reads* provenientes de experimentos do sequenciamento de genomas de diversas espécies, estão atualmente disponíveis no NCBI, dentre elas *Xenopus tropicalis* e *Pipa carvalhoi*. Essas duas espécies estão alocadas na família Pipidae e divergiram cerca de 190

milhões de anos atrás, durante a separação do supercontinente Gondwana, que levou a formação da África e da América do Sul (Canatella 2015).

O alto grau de divergência entre *X. tropicalis* e *P. carvalhoi*, aliado a proximidade filogenética entre essas espécies (i.e. família Pipidae) caracterizam como ótimos modelos para testar a hipótese de bibliotecas de DNAs satélites (Ugarkovic e Plohl 2002). Tal hipótese explica que espécies viventes atuais, por compartilharem um ancestral em comum, também poderiam compartilhar algumas famílias de DNAs satélites, porém o número de cópias de unidades repetitivas de cada família de DNA satélite pode variar entre as espécies que compartilham ancestrais em comum (Ugarkovic e Plohl, 2002). Dessa forma, o presente trabalho procurou evidências que apoiem a hipótese da biblioteca de DNAs satélites por meio de caracterização e buscas *in silico* desses DNAs satélites nos genomas de *X. tropicalis* e de *P. carvalhoi*.

2. OBJETIVOS

2.1 Objetivo Geral

- Elaboração de testes para a hipótese da biblioteca de DNAs satélites entre espécies divergentes da família Pipidae (*Xenopus tropicalis* e *Pipa carvalhoi*).

2.2 Objetivos específicos

- Caracterizar novos DNAsat em *X. tropicalis*;
- Caracterizar novos DNAsat de *P. carvalhoi*;
- Localizar os *repeats* de DNAsat no genoma montado de *X. tropicalis*;
- Avaliar grau de homogeneidade das sequências de DNAsat de *X. tropicalis* e *P. carvalhoi*.

3. REVISÃO BIBLIOGRÁFICA

3.1 DNA repetitivo

Sequências de DNA repetitivo compõem regiões de heterocromatina em muitos genomas eucarióticos (revisão de Ugarkovic e Plohl, 2002; revisão de Charlesworth, Sniegowski, Stephan, 1994). Embora anteriormente, os pesquisadores considerassem o DNA repetitivo como DNA lixo (“junk DNA”) devido ao desconhecimento de suas funções, atualmente há uma ampla gama de informações dada a significância no genoma eucariótico, como por exemplo, a formação de compartimentos genômicos, importantes para as etapas da divisão celular (revisão de Lopez-Flores, Garrido-Ramos, 2012; Ugarkovic e Plohl, 2002). Ao longo do genoma, há sequências de DNA repetitivo que estão dispostas em tandem e as que estão dispersas pelo genoma (revisão de Charlesworth, Sniegowski, Stephan, 1994).

Sequências dispersas pelo genoma - elementos transponíveis (TE), são caracterizadas como sequências que possuem a habilidade de se mover e inserir novas cópias de si mesmas em novas localizações genômicas (revisão de Charlesworth, Sniegowski, Stephan, 1994; revisão de Lopez-Flores, Garrido-Ramos, 2012; revisão de Biscotti, Olmo, Heslop-Harrison, 2015). Dessa forma, a inserção destes elementos ao longo do genoma pode causar mutações que podem levar a mudanças fenotípicas com variação quantitativa de características. TEs também podem causar mutações com impacto evolutivo e no desenvolvimento dos organismos, fornecendo meios para regulação de genes ou até mesmo sendo cooptados pelo genoma hospedeiro (revisão de Charlesworth, Sniegowski, Stephan, 1994; revisão de Lopez-Flores, Garrido-Ramos, 2012). Por exemplo, em *Leishmania* spp. retrotransposons extintos e curtos em tamanho de sequência foram cooptados para a coordenação pós-transcricional da regulação gênica (revisão de Charlesworth, Sniegowski, Stephan, 1994; revisão de Lopez-Flores, Garrido-Ramos, 2012). Os TEs são divididos em 2 classes - Classe I ou retrotransposons, no qual ocorre a reinserção de uma cópia de DNA complementar (cDNA) sintetizada a partir de um RNA transcrito a partir do retrotransponson — e — Classe II ou transposons de DNA, cuja transposição ocorre diretamente, sem o uso de um RNA intermediário (revisão de Biscotti, Olmo, Heslop-Harrison, 2015; revisão de Charlesworth, Sniegowski, Stephan, 1994). A presença e diversidade de TEs, bem como sua

abundância são variáveis para cada espécie, visto que algumas espécies possuem maior presença de determinada classe, e outras possuem maior abundância de outra (revisão de Biscotti, Olmo, Heslop-Harrison, 2015; revisão de Charlesworth, Sniegowski, Stephan, 1994).

As repetições em tandem ocorrem em diferentes escalas, cujo mecanismos genéticos como *crossing-over* desigual e mutação por substituição de bases podem afetar o seu tamanho (revisão de Charlesworth, Sniegowski, Stephan, 1994). Ao contrário dos TEs, as repetições em tandem são divididas em três classes principais: microssatélites, minissatélites e DNAs satélites (revisão de Charlesworth, Sniegowski, Stephan, 1994; revisão de Lopez-Flores, Garrido-Ramos, 2012). Microssatélites (STRs) são sequências de arranjos curtos de 2-5bp, e devido à natureza repetitiva, apresentam maiores taxas de mutação comparada com outros tipos de DNA repetitivo (revisão de Charlesworth, Sniegowski, Stephan, 1994; revisão de Lopez-Flores, Garrido-Ramos, 2012; Gymrek, 2017). Os STRs podem ser encontrados em locais de codificação e de não codificação de proteínas, bem como em regiões regulatórias do genoma (revisão de Lopez-Flores, Garrido-Ramos, 2012). Ademais, o deslizamento durante a replicação do DNA, com adição ou deleção de pares de base, bem como o *crossing over* desigual, contribuem para o número e polimorfismo nas populações (revisão de Charlesworth, Sniegowski, Stephan, 1994; revisão de Lopez-Flores, Garrido-Ramos, 2012). Apesar da sua instabilidade, a taxa de mutação varia entre os diferentes STRs, e isso está associado ao número de unidades de repetição, e ao comprimento da unidade de repetição (revisão de Lopez-Flores, Garrido-Ramos, 2012).

Por outro lado, os minissatélites (VNTRs) são sequências de arranjos longos de 15bp com tamanho de arranjos variáveis (revisão de Charlesworth, Sniegowski, Stephan, 1994; revisão de Lopez-Flores, Garrido-Ramos, 2012). Os processos de mutação de VNTRs, são semelhantes aos de STRs, entretanto, são diferenciados em 2 tipos de repetições em tandem devido ao seu tamanho, distribuição e função nos genomas (revisão de Charlesworth, Sniegowski, Stephan, 1994; revisão de Lopez-Flores, Garrido-Ramos, 2012). Os VNTRs podem ser encontrados em diferentes locais do genoma, sendo regiões subteloméricas, pericentroméricas e centroméricas (revisão de Lopez-Flores, Garrido-Ramos, 2012). Devido a taxa de

mutação, os VNTRs são polimórficos e por isso são comumente utilizados para testes de paternidade e estudos forenses (revisão de Charlesworth, Sniegowski, Stephan, 1994; revisão de Lopez-Flores, Garrido-Ramos, 2012).

Outro tipo de repetições em tandem, são os DNA satélite (DNAsat), podendo ser similares em comprimento com VTNRs e STRs, com repetições de 150-400bp comunitades de repetição ricas em AT e sendo variáveis. (revisão de Charlesworth, Sniegowski, Stephan, 1994; Lima, Ruiz-Ruano, 2022).

3.2 DNAs satélites

Os DNAsat estão organizados em tandem, em unidades monoméricas com mais de 100pb nas regiões de heterocromatina constitutiva, organizadas em arranjos que podem ser maiores que 100 megabases (Mb), principalmente próximas aos centrômeros e telômeros, que durante o ciclo celular permanecem condensadas (revisão de Charlesworth, Sniegowski, Stephan, 1994; revisão de Plohl et al, 2008; revisão de Plohl et al, 2007). Dentro do genoma, o DNAsat pode compreender cerca de 50% do total de DNA, e isso contribui para que existam variações do tamanho do genoma de diferentes eucariotos (revisão de Plohl et al, 2007).

Tais porções genômicas foram descobertas através de um pico no perfil de ultracentrifugação - devido a diferença entre A+T - de CsCl, antiga técnica utilizada para analisar a sedimentação em equilíbrio em gradientes de densidade de CsCl do DNA, separando diferentes moléculas de DNA de acordo com as suas propriedades físicas (Kit et al, 1961; Ruiz-Ruano et al, 2016; revisão de Plohl el at, 2007; revisão de Lower et al, 2018). Atualmente, este método foi substituído por técnicas de cinética de renaturação de DNA, digestão do DNA genômico com endonucleases de restrição, eletroforese e análises bioinformáticas de coleções de sequências curtas de DNA produzidas pelo Sequenciamento de Nova Geração (NGS) (Ruiz-Ruano et al, 2016; revisão de Plohl, Mestrovic, Mravinac, 2012; revisão de Plohl el at, 2007; revisão de Lower et al, 2018). Na década de 70, os DNAsats eram considerados como uma porção genômica inútil, devido à ausência de potencial de codificação, e a falta de evidência para função, essa teoria perdurou por um longo tempo (revisão

de Plohl, Mestrovic, Mravinac, 2012). No entanto, novos estudos foram realizados buscando entender o potencial funcional dos DNAsat, com a questão de que: se não há funcionalidade por que essas frações genômicas não são perdidas ou diminuídas através de mecanismos de diminuição da heterocromatina (revisão de Plohl, Mestrovic, Mravinac, 2012)? Com o avanço dos estudos, os DNAsats foram associados em processos de funcionamento e organização de cromossomos, processos de divisão celular, regulação da expressão gênica, bem como tendo um papel em eventos de especiação (revisão de Plohl et al, 2007; Rosic, Kohler, Erhardt, 2014; Ferree, Barbash, 2009; revisão de Rosic, Kohler, Erhardt, 2014; McNulty, Sullivan, Sullivan, 2017; Pezer et al, 2012). Por exemplo, DNAsat centromérico alfa satélite, em humanos, participa de interações funcionais e para isso é necessário que ele contenha sequências *motifs* que serão reconhecidas por componentes proteicos, como por exemplo, o CENP-B box, que ao se ligar a proteína CENP-B (variante de histona centromérica) confere identidade ao centrômero e facilita a formação do cinetocoro (revisão de Plohl et al, 2007; Yoda et al, 2008). Também os DNAsats podem estar envolvidos com mecanismos de regulação da transcrição, como a iniciação da metilação da histona H3, que é um pré-requisito para a formação e manutenção da heterocromatina (revisão de Plohl et al, 2007). Estudos em *Drosophila* inferem que a especificidade das sequências de DNAsat afeta a segregação dos cromossomos em híbridos de fêmeas de *Drosophila simulans* e machos de *Drosophila melanogaster* relacionado ao cromossomo X (Ferree, Barbash, 2009).

Os DNAsats têm um papel importante na evolução dos genomas, já que a sua dinâmica evolutiva é imprescindível para a compreensão da diversidade do genoma (Belyayev et al, 2019). As unidades repetitivas de DNAsat evoluem em concerto (*concerted evolution*), ou seja, de modo não independente, no qual o resultado é uma homogeneização das unidades repetitivas do mesmo DNAsat dentro do genoma de uma espécie (Ugarkovic e Plohl, 2002; revisão de Dover, 1986). Dessa forma, novas variantes poderão ser geradas (por ação de mutação) e espalhadas para outros sítios do genoma, levando a fixação dessas variantes em uma população reprodutiva, por meio de um processo conhecido como “*molecular drive*”, que dependerá de mecanismos moleculares como: (i) crossing over desigual responsável pela mudança no número de cópias e (ii) conversão gênica (Ugarkovic

e Plohl, 2002; revisão de Dover, 1986). Este processo resulta em uma alta similaridade entre as unidades repetitivas em uma mesma espécie do que entre espécies diferentes (Belyayev et al, 2019; revisão de Dover, 1986). Para a dispersão pelo genoma, diferentes processos estão envolvidos, como crossing over desigual, conversão gênica, inserção de DNA extracromossômico, recombinação ectópica bem como também pela ação de mecanismos de transposição (Ugarkovic and Plohl, 2002; revisão de López-Flores e Garrido-Ramos, 2012, Plohl et al, 2010). Ademais, outros processos como o deslizamento da DNA polimerase durante a replicação, replicação por círculo rolante, mecanismo semelhante a conversão, podem aumentar o número de cópias de DNAsat no genoma (Ugarkovic e Plohl, 2002).

Dentro de um genoma, há diferentes sequências de DNAsat, que se originam a partir de uma biblioteca de DNAsat ancestral. Duas espécies espécies aparentadas compartilham sequências de DNAsat, que eram presentes na biblioteca ancestral, porém cada espécie pode fixar um número de cópias diferentes para cada família de DNAsats, através de mutações, expansão e contração dos *clusters* de DNAsats substituindo uma repetição dominante por outra de menor., representação (Ugarkovic e Plohl, 2002; Plohl et al, 2008; revisão de Thakur, Packiaraj, Henikoff, 2021). De acordo com o modelo da Biblioteca de DNAsat de Fry e Salser et al (1977), há a formação de um perfil específico de DNAsat para cada espécie relacionada (revisão de Thakur, Packiaraj, Henikoff, 2021; Plohl et al, 2008). Estes perfis podem ser espécie-específicos, no entanto entre espécies relacionadas existem famílias de DNAsat compartilhadas entre gênero, família e ordem que irão, em geral, variar no número de cópias de unidades repetitivas (Garrido-Ramos, 2017). Desse modo, a alteração do número de cópias de DNAsat pode ocasionar grandes mudanças evolutivas em um dado genoma (Plohl, Mestrovic, Mravinac, 2012). Sendo assim, alterações nos perfis de DNAsat tem um grande impacto na estabilidade genômica, já que tais mudanças levam a uma variabilidade genética ao alterar os perfis de DNAsat de uma espécie (Plohl, Mestrovic, Mravinac, 2012, Mestrovic et al, 1998).

Dentre os vertebrados, o genoma dos anfíbios contém uma maior variabilidade de tamanho, devido à alta quantidade de DNA repetitivo (Roco et al, 2021). Por isso, são comumente utilizados como modelos de estudos de genética

evolutiva (Zettera et al, 2020). Em Anura, os DNAsat foram estudados no gênero *Xenopus*, *Bufo* e *Rana*, com a melhor caracterização de família de DNAsat do S1 de *European brown frog*, cuja espécie possui estrutura e distribuição cromossômica específica (Amor et al, 2010; revisão de Targueta et al. 2018). O DNAsat Pcp190 foi mapeado em diferentes genomas de anuros, sendo inicialmente isolado em *Physalaemus cuvieri*, com origem associada ao DNAr 5S dada a similaridade com o locus do gene RNAr 5S. Pcp190 também foi identificado em diferentes espécies do gênero *Physalaemus*, *Leptodactylus latrans*, *Crossodactylus gaudichaudii* e *Pseudis tocantins*, reforçando a ideia de que se trata de um DNAsat antigo (Gatto et al, 2008). A presença de Pcp190 em diferentes espécies de anuros vai contra o que é comumente encontrado a respeito de DNAsat antigos, no qual é possível a identificação do DNAsat espécie específico ou em espécies relacionadas filogeneticamente (Gatto et al, 2018). Devido a essas características, mudanças de perfis de DNAsat nos genomas podem ter um importante impacto na evolução das espécies que os possuem. Adicionalmente, em função da possibilidade de os mesmos apresentarem aspectos funcionais, bem como serem sequências altamente dinâmicas evolutivamente, o estudo de DNAsat associados após e/ou durante a ocorrência de eventos de especiação pode auxiliar em um melhor entendimento da evolução dos genomas de diferentes espécies, sejam elas próxima ou distivamente relacionadas filogeneticamente (revisão de Ugarkovic, Pohl, 2002).

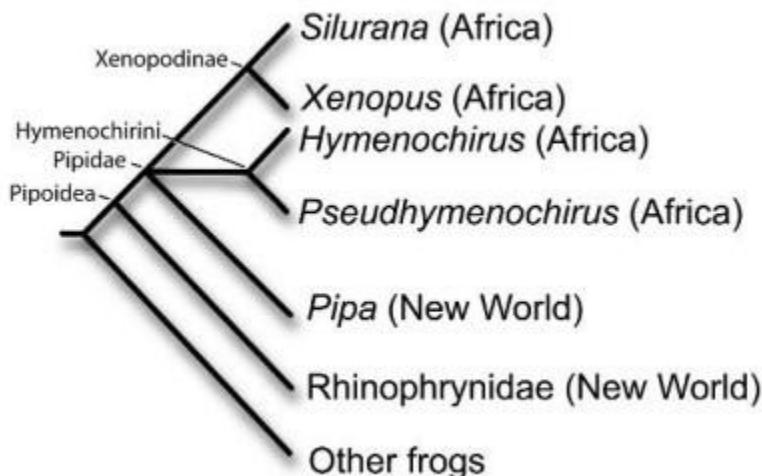
3.3 Anfíbios Anuros da família Pipidae

Anfíbios anuros são considerados um grupo diverso e filogeneticamente antigo, sendo utilizados como modelos de estudos biológicos, como é o caso da família Pipidae, que teve origem durante o Jurássico Médio, cerca de 170Ma atrás, atualmente contendo 41 espécies (Cannatella, 2015; Evans, 2008; Pyron e Wiens 2011; Frost, 2024).

A família Pipidae é dividida em 4 gêneros: *Xenopus*, *Hymenochirus*, *Pseudohymenochirus* e *Pipa*, no qual apenas o último é encontrado na América do Sul e os demais na África subsaariana (Bewick et al, 2012). A relação filogenética dos Pipídeos não está resolvida, existindo controvérsias devido às diferenças de relações obtidas com análises dos dados moleculares e de dados fenotípicos, uma

vez que, as topologias das árvores obtidas com os dados fósseis, morfológicos e moleculares são contraditórias e podem exibir diferentes graus de homoplasia que resultam em diferentes arranjos filogenéticos (Cannatella, 2015; Mezzasalma et al, 2015). Entretanto, há um consenso em relação ao gênero, estabelecida através de análises de dados morfológicos e moleculares, onde cada um é entendido como monofilético, sendo *Xenopus* e *Silurana* considerado um clado, e *Hymenochirus* e *Pseudohymenochirus* outro clado, com ambos os clados formando um grupo irmão com o gênero *Pipa*. Adicionalmente, a família Pipidae corresponde a um grupo irmão da família Rhinophrynididae (FIGURA 1) (Bewick et al, 2011; Irisarri et al, 2011).

FIGURA 1 – ÁRVORE FILOGENÉTICA PIPIDAE



FONTE: Bewick et al, 2011.

Os sapos da família Pipidae divergiram dezenas de milhões de anos atrás. A teoria mais aceita é relacionada à deriva continental. Com a abertura do Oceano Atlântico, separando a América do Norte da África Ocidental, cerca de 190Ma atrás, pode ter desencadeado a divergência entre Rhinophrynididae e os Pipidae. Posteriormente, cerca de 100Ma atrás, com a separação da América do Sul e da África, pode ter ocorrido a divergência de *Pipa* dos pipídeos africanos (Bewick et al, 2011).

Membros da família Pipidae possuem adaptações para o ambiente aquático como o corpo achatado dorso-ventralmente, membros posteriores com membranas interdigitais e direcionados para frente e olhos localizados dorsalmente, também,

não possuem língua (Molina, 2007). Dentre os Pipideos, a espécie *Pipa carvalhoi* (FIGURA 2) tem como habitat a região Sudeste e Nordeste do Brasil, ocupando sistemas terrestres e aquáticos (IUCN, 2024). Já *Xenopus tropicalis* (FIGURA 3) é encontrado no continente Africano, na África Ocidental (IUCN, 2024). A respeito do cariotípico da família Pipidae, tanto para *Pipa carvalhoi* quanto para *Xenopus tropicalis* é dado como $2n=20$, apesar de membros do gênero *Xenopus* possuirem diferentes variações no número do cariotípico (Zattera et al, 2020).

FIGURA 2 – ESPÉCIME DE *Pipa Carvalhoi*.



FONTE: Arquivo pessoal de Juliane Petry de Carli Monteiro.

FIGURA 3 - ESPÉCIME DE *Xenopus tropicalis*.



Fonte: Elena Reljic (2023)

4. MATERIAIS E MÉTODOS

4.1 Aquisição dos dados e sequenciamento

Reads do genoma de *X. tropicalis* foram obtidas a partir do repositório *Sequence Read Archive* (SRA) (número de acesso: SRX123910). Já as *reads* do genoma de *P. carvalhoi* foram as mesmas geradas no Laboratório de Citogenética Evolutiva e Conservação Animal do Departamento de Genética da Universidade Federal do Paraná, em parceria com o prof. Dr. Daniel Pacheco Bruschi. Brevemente, o DNA genômico de *P. carvalhoi* foi obtido a partir de amostras de fígado utilizando o protocolo de Medeiros et al. (2013). O genoma foi sequenciado com uma cobertura de 46x em uma plataforma *Illumina HiSeq-4000*. A preparação das biblioteca de sequenciamento foi realizada empregando *TruSeq DNA Library Prep Kit* (Illumina, Inc).

4.2 Caracterização *de novo* de DNAs satélites

As *reads* dos genomas de *X. tropicalis* e *P. carvalhoi* foram inicialmente processadas para retirada dos adaptadores e de *reads* de baixa qualidade usando o *software Trimmomatic* (Bolger et al. 2014). Selecionando o “*Performance initial Illumina clip*”, selecionando “*TruSeq3 (Paired-ended, for MiSeq and HiSeq)*” em “*Adapter Sequence to use*”, por fim na opção de “*Trimmomatic operation*” com número de bases para calcular a média de 4, qualidade média de janela de 15, comprimento de leitura de 30, qualidade mínima necessária para manter uma base de 3 para *leading* e *trailing*. Também, foi selecionada a opção de “*Phred33*” para “*Quality score end*”. Após essa etapa, as *reads* de *X. tropicalis* foram usadas como *input* no *software RepeatExplorer2* (Novák et al. 2013). Os *clusters* de sequências que apresentaram sinais clássicos de DNAsat (e.g. estrutura dos grafos circulares e com os vértices muito próximos) foram selecionados para análises posteriores e também para a construção de uma biblioteca customizada para realizar análises comparativas. De posse da biblioteca customizada, a mesma foi utilizada para a caracterização de DNAsat de *P. carvalhoi*. Inicialmente, as *reads* de *P. carvalhoi* também foram processadas pelo *software Trimmomatic* e utilizadas então como

input no *software RepeatExplorer2*, juntamente com a biblioteca customizada de DNAsat de *X. tropicalis* para análise comparativa. Para a exclusão de falsos positivos (i.e. regiões ricas em famílias multigênicas) os DNAsats caracterizados pelo *software RepeatExplorer2* tanto de *P. carvalhoi* quanto de *X. tropicalis*, foram utilizados em buscas por blastn, com valor de bitscore estipulado em 300 para exclusão de falsos positivos (ANEXO 1, ANEXO 2) (Altschul et al. 1990) contra coleção de sequências nucleotídicas do *National Center for Biotechnology Information* (NCBI).

4.3 Mapeamento *in silico* de DNAs satélites em *X. tropicalis* e *P. carvalhoi*

De posse dos DNAsat caracterizados pelo *software RepeatExplorer2*, os mesmos foram mapeados *in silico* contra as montagens dos genomas de *X. tropicalis* (número de acesso: GCA_000004195.4) e *P. carvalhoi* (número de acesso: GCA_021901965.1). O mapeamento *in silico* foi realizado por meio do *software BLAST* (Altschul et al. 1990) e os dados foram filtrados a partir da tabela de resultados com base em gráficos de distribuição de valores Bitscore. A visualização dos dados de mapeamento *in silico* de *X. tropicalis* (espécie que tem um genoma montado em nível de cromossomos) foi obtida por meio do *software R* (R Core Team 2023) empregando o pacote karyoplotR (Gel e Serra 2017). No caso de *P. carvalhoi*, a confirmação da organização *in tandem* e visualização dessa organização dos DNAsats será obtida por meio da construção de gráficos de alinhamento *dot plot* usando o *software Gepard* (KRUMSIEK et al., 2007).

4.4 Análises de similaridade de DNAs satélites

De posse dos resultados de BLAST contra os genomas de *X. tropicalis* e *P. carvalhoi*, os valores de porcentagem de identidade das *queries* em relação às unidades repetitivas das montagens dos genomas foram utilizados para gerar gráficos de *boxplot*. Com a distribuição desses valores de identidade pode-se inferir quais famílias de DNAs satélites apresentam maior homogeneidade nos genomas de *X. tropicalis* e *P. carvalhoi*. Os gráficos *boxplot* foram confeccionados no *software*

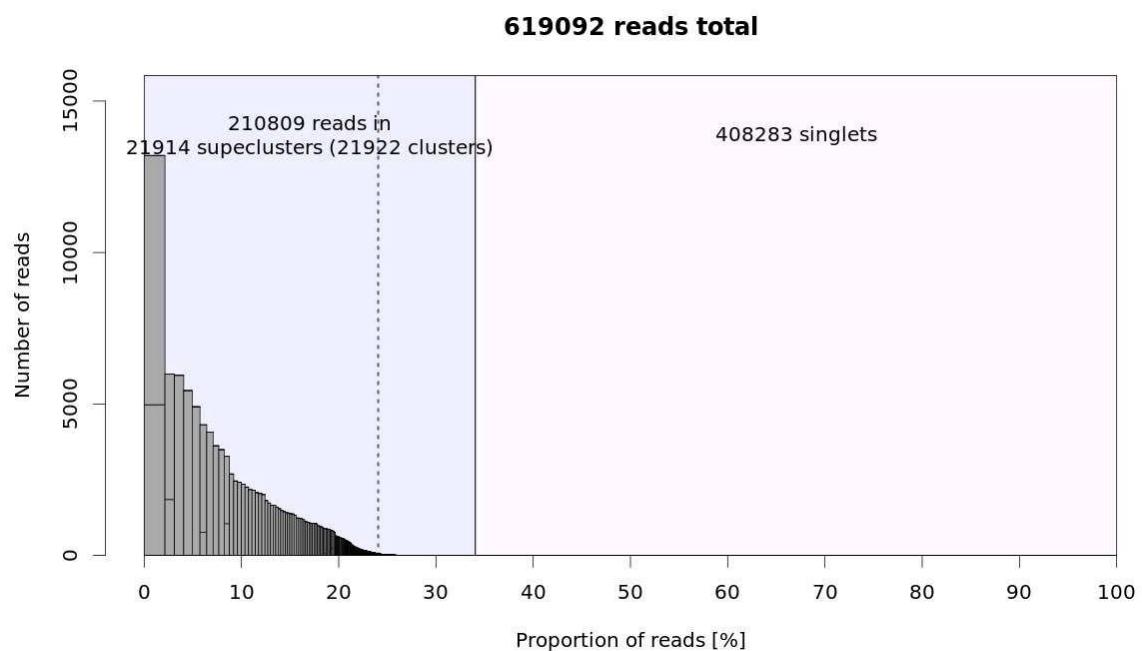
R (R Core Team 2023) empregando o pacote *ggplot2* (WICKHAM 2016). DNAs sat mais homogêneos foram delimitados como aqueles que apresentaram menor dispersão dos valores de identidade no gráfico *boxplot*.

5. RESULTADOS

5.1 DNAs satélites no genoma de *Xenopus tropicalis*

A partir da extração dos dados do repositório *Sequence Read Archive* (SRA) de *Xenopus tropicalis*, e seleção dos dados para a análise pela Plataforma *RepeatExplorer2*, detectou-se uma proporção de pouco mais de 25% das *reads* pertencendo a porção repetitiva do genoma dessa espécie que foram atribuídas a algum *cluster/supercluster* de sequências repetitivas (FIGURA 4). Com base na análise do *RepeatExplorer2*, sendo 21 *clusters* de alta confiança e 42 de baixa confiança, totalizando 53 contigs (FIGURA 5). Destes contigs, foram excluídos aqueles que tinham similaridade com organelas e os que a sua sequência consenso eram inferior a 50bp de comprimento, somando 46 contigs. No que diz a respeito aos contigs identificados com genoma organelar, a filtragem levou em consideração a identificação destes por meio da análise do *RepeatExplorer*, usando o banco de dados default do pipeline (Novák et al, 2021). Buscas por BLASTn contra a coleção de sequências nucleotídicas do NCBI não revelaram similaridades significantes com famílias multigênicas ou genes de cópia única entre os 46 contigs anteriormente citados.

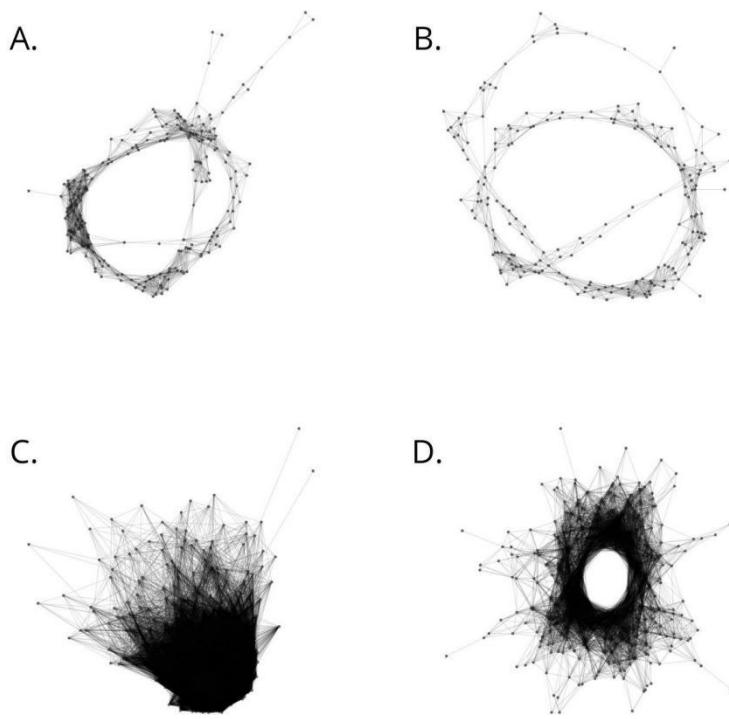
FIGURA 4 - PROPORÇÃO DE *READS* DE DNAs REPETITIVOS DE *X. tropicalis*.



FONTE: *RepeatExplorer2*.

LEGENDA: Gráfico resultante da clusterização das *reads* provenientes do genoma de *X. tropicalis* relacionadas à elementos de DNAs repetitivos obtidos em cada *supercluster*. No eixo Y temos o número de *reads* atribuídas para cada *supercluster* de DNA repetitivo e no eixo X indica a proporção em relação ao total de *reads* analisadas.

FIGURA 5 – READS DE *X. tropicalis*



FONTE: *RepeatExplorer* 2.

LEGENDA: Gráficos das *reads* geradas pelo *RepeatExplorer* 2. A. Contig 107; B. Contig 108, C. Contig 51, D. Contig 70. Cada ponto sinalizado no gráfico é uma *read*, as linhas que conectam estes pontos são chamadas de vértices. Estão conectadas as *reads* que possuem sobreposição, ou seja, *reads* que se alinham e, em geral, pertencem a um mesmo cluster de região repetitiva.

Buscas por BLASTn dos 46 contigs de DNAsats resultantes realizadas contra a montagem do genoma de *X. tropicalis*, usadas para filtragem dos dados, de acordo com a distribuição de valores Bitscore (ANEXO I), sendo o valor de corte estipulada como acima de 300, resultaram em uma matriz final de 13 contigs. Estes 13 contigs apresentaram variação tanto no número de cópias quanto na distribuição entre os 10 cromossomos de *X. tropicalis*. Os DNAsat foram nomeados segundo o método utilizado por Ruiz-Ruano et al. (2016), como sigla da espécie + Sat, XtrSat, seguido do número do DNAsat mais o tamanho da sua sequência consenso, sendo um total de 42 DNAsat (TABELA 1).

TABELA 1: DISTRIBUIÇÃO CROMOSSÔMICA, NÚMEROS DE CÓPIAS E NOMENCLATURA DE DNAs SATÉLITES CARACTERIZADOS PARA *Xenopus tropicalis*.

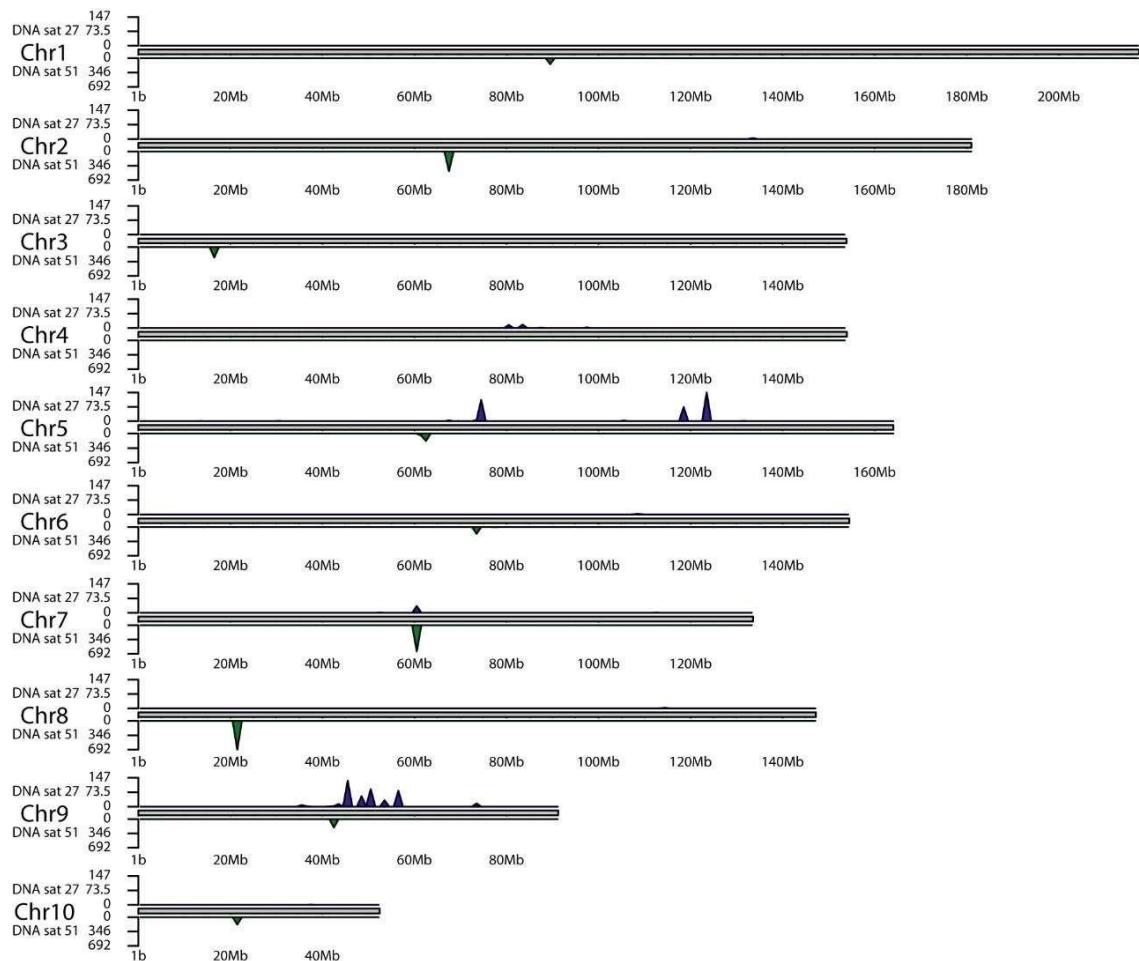
Contig	Cromossomo	Tamanho da unidade repetitiva (bp)	Nomenclatura proposta	Número de cópias
107	1	1030	XtrSat01-1030	26
108		1609	XtrSat02-1609	2
136		276	XtrSat03-276	961
51		206	XtrSat04-206	144
70		544	XtrSat05-544	105
107	2	1030	XtrSat01-1030	2
108		1609	XtrSat02-1609	5
183		481	XtrSat06-481	33
27		1299	XtrSat07-1299	4
51		206	XtrSat04-206	477
85		290	XtrSat08-290	119
108	3	1609	XtrSat02-1609	26
134		1189	XtrSat09-1189	121
51		206	XtrSat04-206	248
136	4	276	XtrSat03-276	6
27		1299	XtrSat07-1299	37
51		206	XtrSat04-206	228
108	5	1609	XtrSat02-1609	67
27		1299	XtrSat07-1299	351
51		206	XtrSat04-206	223
70		544	XtrSat05-544	478
107	6	1030	XtrSat01-1030	132
108		1609	XtrSat02-1609	2
136		276	XtrSat03-276	247
173		454	XtrSat10-454	361
27		1299	XtrSat07-1299	5
51		206	XtrSat04-206	164
108	7	1609	XtrSat02-1609	3
27		1299	XtrSat07-1299	36
51		206	XtrSat04-206	628
107	8	1030	XtrSat01-1030	1
108		1609	XtrSat02-1609	76
117		702	XtrSat11-702	321
165		181	XtrSat12-181	472
27		1299	XtrSat07-1299	4
35		564	XtrSat13-564	187
51		206	XtrSat04-206	692
27	9	1299	XtrSat07-1299	431
51		206	XtrSat04-206	203
27	10	1299	XtrSat07-1299	1
35		564	XtrSat13-564	7
51		206	XtrSat04-206	176

Fonte: a autora (2024)

A partir dos resultados das buscas por blastn dos DNAsats contra a montagem do genoma de *Xenopus tropicalis*, foi possível determinar a localização dos contigs nos cromossomos dessa espécie. Os dados foram analisados conforme a densidade de cada unidade repetitiva em janelas de 10000 bp ao longo do genoma de referência, demonstrando o quanto o DNAsat ocupa no genoma.

No XtrSat04-206 (contig 51), é possível observar a presença do mesmo em apenas um *cluster* específico em todos os cromossomos de *X. tropicalis* (GRÁFICO 1). Por outro lado, o XtrSat07-1299 (contig 27) apresenta uma variação em respeito à localização e em quantidade de repetições em determinada região dos cromossomos. Nos cromossomo 7 e 9, é possível visualizar que ambos DNAsat estão identificados próximos um do outro, na região de 60Mb e 40Mb respectivamente. A maior densidade observada do XtrSat07-1299 está presente nos cromossomos 5 e 9, o que indica que nas regiões de 120Mb e 40Mb há uma maior concentração de unidades repetitivas respectivamente. O mesmo é valido para o XtrSat04-206, mas nas regiões de 60Mb e 20Mb nos cromossomos 7 e 8, com uma densidade de 692. Interessantemente, XtrSat04-206 mostrou alta similaridade com o DNAsat centromérico de *X. tropicalis*, já caracterizado anteriormente (Brederson et al., 2024) (FIGURA 6).

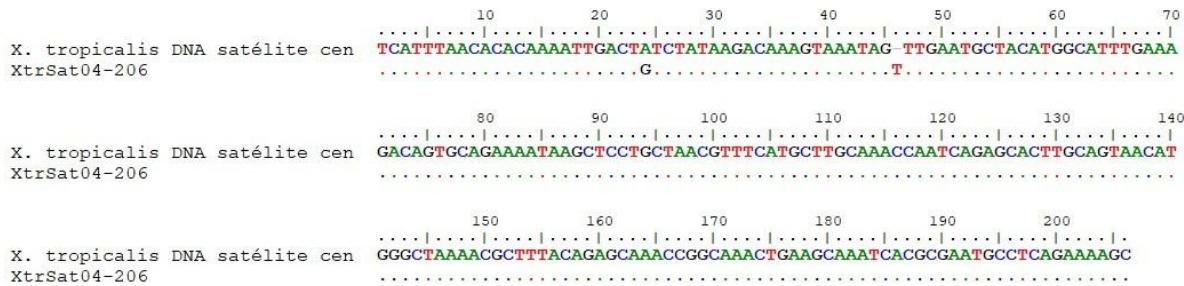
GRÁFICO 1 - DISTRIBUIÇÃO DOS DNASAT XtrSat04-206 E XtrSat07-1299.



FONTE: A autora (2024).

LEGENDA: Densidade dos DNAs satélites XtrSat04-206 (Contig 51), representado em verde e XtrSat07-1299 (Contig 27), representado em azul, na montagem do genoma de *Xenopus tropicalis* ao longo de janelas de 10 Mb. O eixo Y em cada cromossomo representa a densidade de unidades repetitivas e o eixo X de cada cromossomo representa as coordenadas genômicas.

FIGURA 6 – ALINHAMENTO DE UNIDADE MONOMÉRICA CANÔNICA DO DNA SATÉLITE CENTROMÉRICO DE *Xenopus tropicalis*.

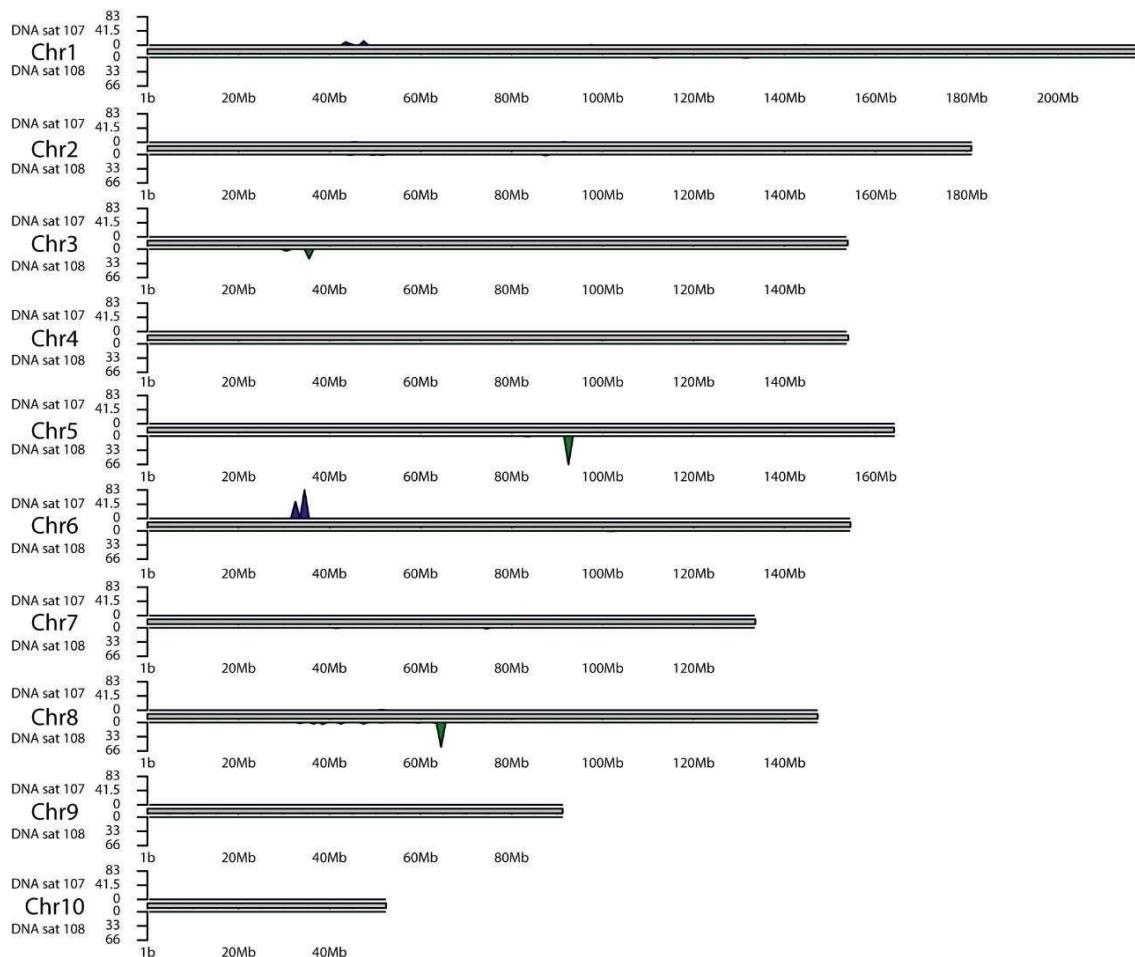


Fonte: A autora (2024).

LEGENDA: Alinhamento entre as sequências do DNA satélite centromérico de *Xenopus tropicalis* descrito por Bredenson et al. (2024) e o DNA satélite XtrSat04-206 obtido no presente estudo. Os pontos indicam 100% de similaridade entre as sequências.

Analizando os XtrSat01-1030 e XtrSat02-1609, diferentemente dos DNAsat anteriores, é possível observar que sua densidade é menor, distribuídos com maior grau nos cromossomos 5 e 8 para o XtrSat02-1609 e no cromossomo 6 para o XtrSat01-1030, nas regiões próximas de 100Mb, 60Mb e 40Mb respectivamente (GRÁFICO 2). Ademais, nos outros cromossomos é observado um valor de densidade menor, demonstrando uma menor quantidade de unidades repetitivas.

GRÁFICO 2 – DISTRIBUIÇÃO DOS DNASAT XtrSat01-1030 E XtrSat02-1609.

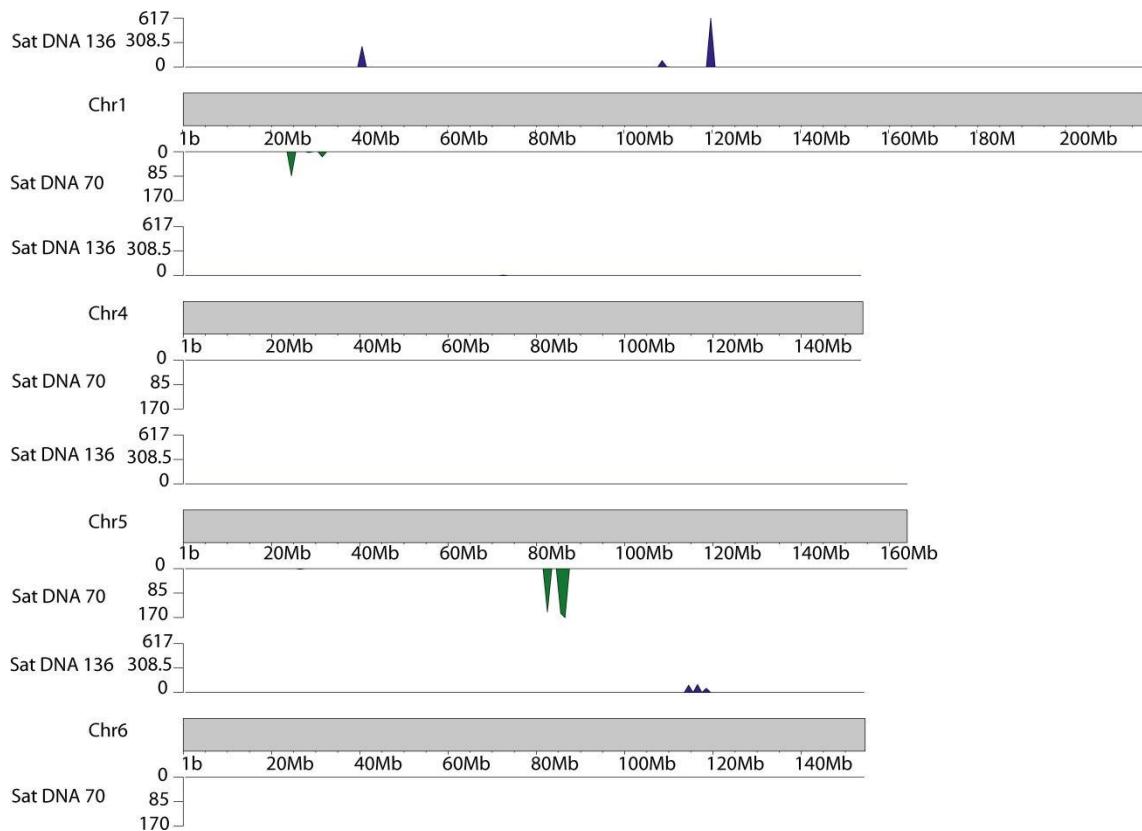


FONTE: A autora (2024).

LEGENDA: Densidade dos DNAs satélites XtrSat01-1030 (DNA sat 107), representado em azul e XtrSat02-1609 (DNA sat 108), representado em verde na montagem do genoma de *Xenopus tropicalis* ao longo de janelas de 10 Mb. O eixo Y em cada cromossomo representa a densidade de unidades repetitivas e o eixo X de cada cromossomo representa as coordenadas genômicas.

Para outros satélites menos abundantes, como o XtrSat03-276 e o XtrSat05-544 (GRÁFICO 3), observa-se que ambos estão presentes no cromossomo 1, nas regiões de 40Mb, 110Mb e 120Mb para o primeiro, e entre 20Mb e 40Mb para o segundo. No cromossomo 4 há a presença somente do XtrSat03-276, na região entre 70Mb e 75Mb, mas em baixo número de cópias. Por outro lado, no cromossomo 5 há somente unidades repetitivas do XtrSat05-544, no intervalo entre 25Mb e 30Mb, em baixa repetição, e entre 80Mb e 90Mb, com alta quantidade de repetições. No cromossomo 6, é visualizado baixas unidades de repetição de XtrSat03-276, na região de 110Mb a 120Mb.

GRÁFICO 3 - DISTRIBUIÇÃO DOS DNASAT XtrSat03-276 E XtrSat05-544.

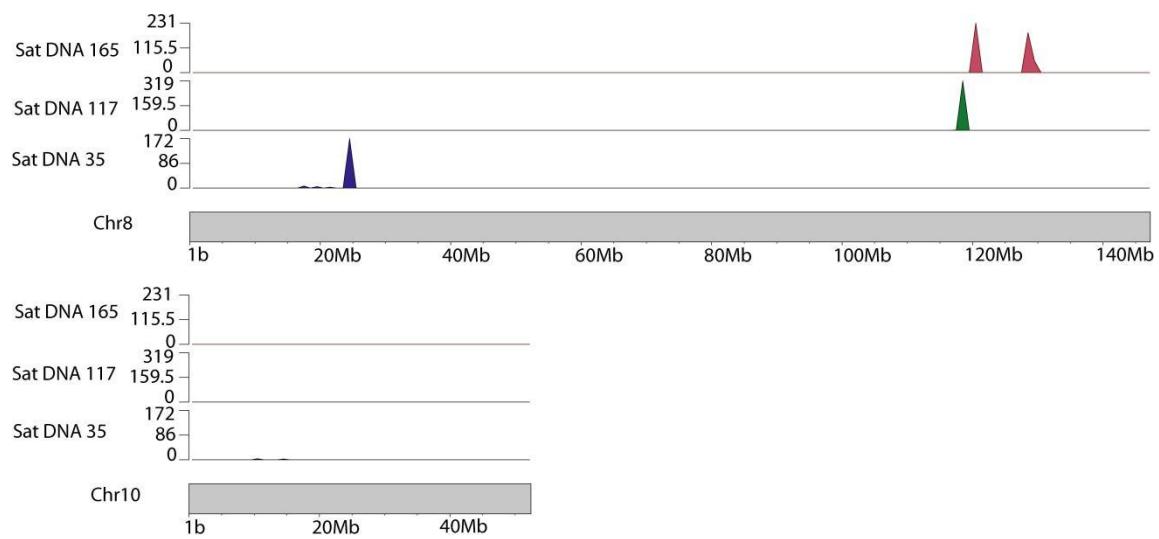


FONTE: A autora (2024).

LEGENDA: Densidade dos DNAs satélites XtrSat03-276 (Sat DNA 136), representado em verde e XtrSat05-544 (Sat DNA 70), representado em azul, na montagem do genoma de *Xenopus tropicalis* ao longo de janelas de 10 Mb. O eixo Y em cada cromossomo representa a densidade de unidades repetitivas e o eixo X de cada cromossomo representa as coordenadas genômicas.

No cromossomo 8, há a presença de XtrSat12-181, XtrSat11-702 e XtrSat13-564 (GRÁFICO 4), sendo identificadas unidades repetitivas na região de 120Mb e 135Mb para os dois primeiros DNAsat, e 15Mb e 25Mb para o XtrSat13-564. É possível visualizar no cromossomo 10 a presença em baixa repetição do XtrSat13-564 na região de 10Mb e 15Mb. É identificada no cromossomo 2 a presença do XtrSat06-481 e do XtrSat08-290, na região entre 80Mb e 100Mb, enquanto no cromossomo 3 há somente a presença do XtrSat09-1189 na região de 145Mb (GRÁFICO 5). Por fim, no cromossomo 6, o XtrSat10-454 é visualizado na região de 120Mb.

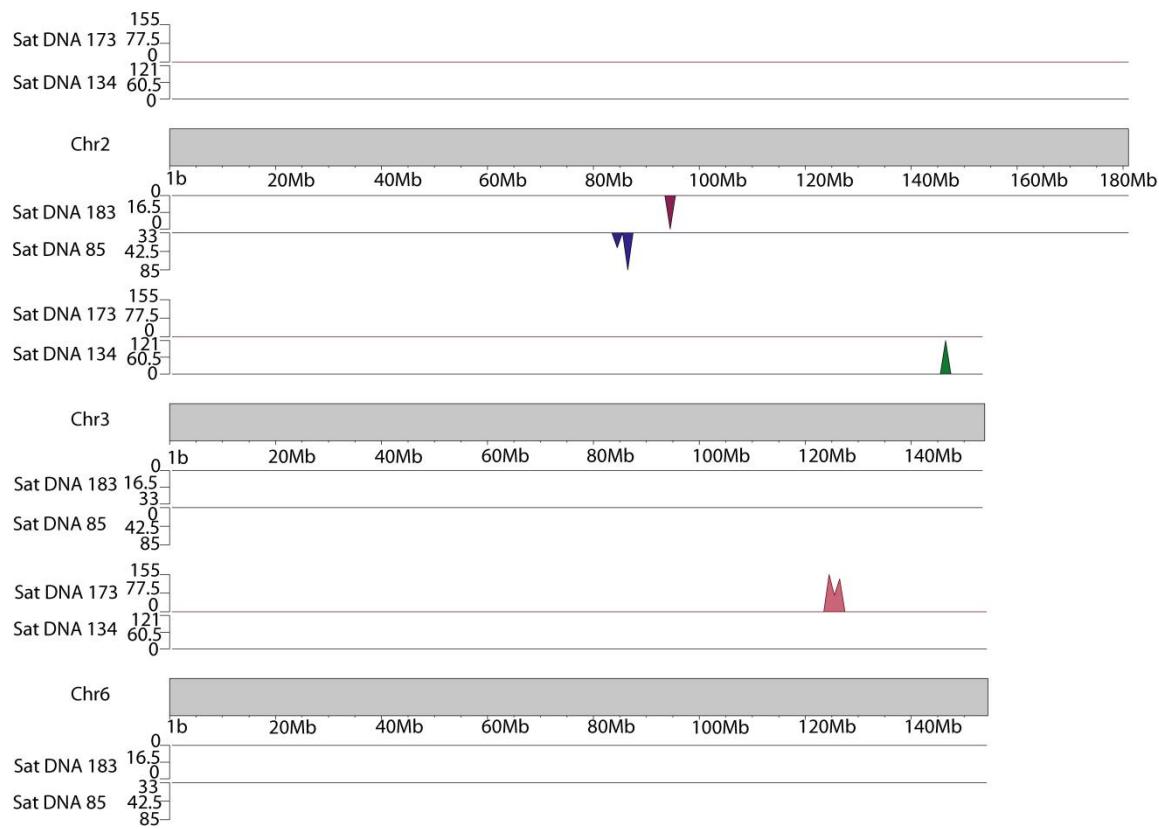
GRÁFICO 4 - DISTRIBUIÇÃO DOS DNASAT XtrSat12-181, XtrSat11-702 E XtrSat13-564



FONTE: A autora (2024).

LEGENDA: Densidade dos DNAs satélites XtrSat12-181 (Sat DNA 165) representado em vermelho, XtrSat11-702 (Sat DNA 117), representado em verde, e XtrSat13-564 (Sat DNA 35), representado em azul, na montagem do genoma de *Xenopus tropicalis* ao longo de janelas de 10 Mb. O eixo Y em cada cromossomo representa a densidade de unidades repetitivas e o eixo X de cada cromossomo representa as coordenadas genômicas.

GRÁFICO 5 - DISTRIBUIÇÃO DOS DNASAT XtrSat06-481, XtrSat08-2.290, XtrSat09-1189 E XtrSat10-454

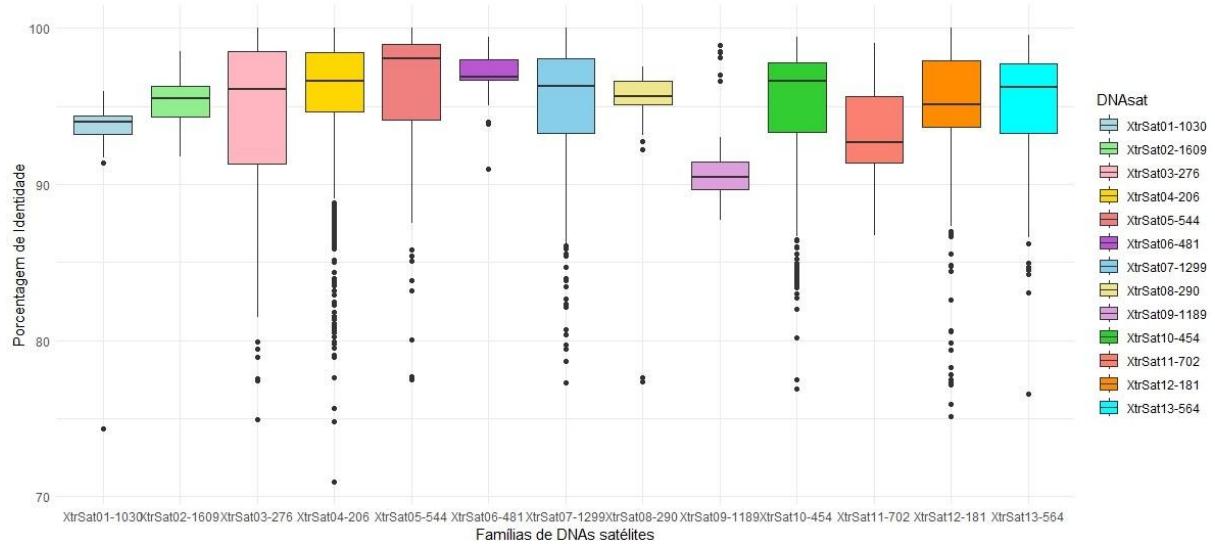


FONTE: A autora (2024).

LEGENDA: Densidade dos DNAs satélites XtrSat10-454 (Sat DNA 173), representado em vermelho XtrSat09-1189 (Sat DNA 134), representado em verde XtrSat06-481 (Sat DNA 183) e XtrSat08-290 (Sat DNA 85), representado em azul, na montagem do genoma de *Xenopus tropicalis* ao longo de janelas de 10 Mb. O eixo Y em cada cromossomo representa a densidade de unidades repetitivas e o eixo X de cada cromossomo representa as coordenadas genômicas.

A partir das buscas por BLAST, utilizando as sequências consenso dos DNAsats, as distribuições das porcentagens de identidade das unidades repetitivas de cada DNAsat no genoma de *X. tropicalis* variaram consideravelmente (GRÁFICO 6).

GRÁFICO 6 - ANÁLISE DE SIMILARIDADE DE DNAsat EM *X. tropicalis*



Fonte: a autora (2024)

Legenda: Distribuição da porcentagem de identidade observada para cada DNAsat de *X. tropicalis* em gráfico de boxplot. O eixo Y representa a Porcentagem de identidade dos DNAsat e o eixo X representa as Famílias de DNAsat.

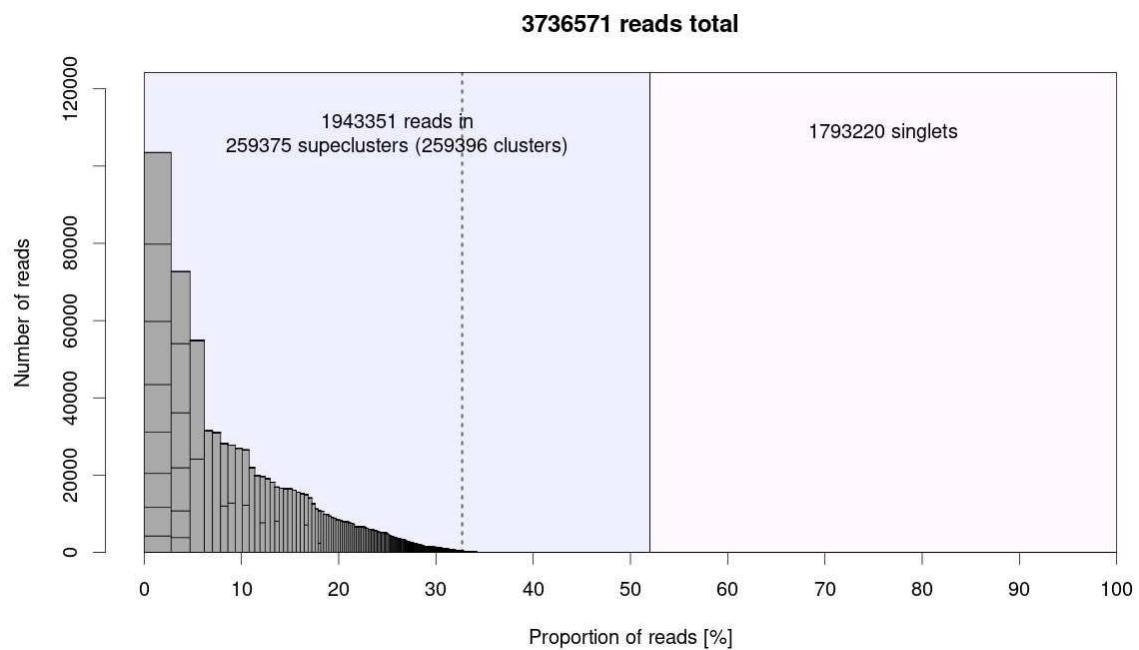
Através do Gráfico 6 é possível visualizar que há diferenças nos graus de identidade entre os satélites investigados. Observando o XtrSat01-1030 (Contig 107), o XtrSat02-1609 (Contig 108), o XtrSat06-481 (Contig 183), o XtrSat08-290 (Contig 85) e XtrSat09-1189 (Contig 134), há um grau de identidade maior entre as unidades repetitivas, observado pela dispersão das porcentagens de identidades. Por outro lado, outros DNAsats, como por exemplo, o XtrSat03-276 (Contig 136), XtrSat04-206 (Contig 51) e XtrSat09-1189 (Contig 134), apresentaram uma maior dispersão das porcentagens de identidade, incluindo diversas unidades repetitivas que se apresentam como *outliers*.

5.2. DNAs Satélites no genoma de *Pipa carvalhoi*

A partir do processamento de dados na plataforma *RepeatExplorer 2* de *Pipa carvalhoi* foi identificado que pouco mais de 30% das *reads* analisadas são relacionadas à proporção da fração repetitiva do genoma dessa espécie, sendo atribuídas a algum *cluster/supercluster* de sequências repetitivas (FIGURA 7). O *pipeline RepeatExplorer2* levou a identificação de 15 satélites de alta confiança e 30 satélites de baixa confiança, totalizando 45 DNAsat (FIGURA 8). Destes, 24

foram selecionados para posterior análise, os demais não foram contabilizados devido à sua similaridade com organelas e que possuíam sua sequência consenso inferior a 50bp. Utilizando a biblioteca customizada de DNAsat de *X. tropicalis* como análise comparativa, foram obtidos 2 contigs de *P. carvalhoi* com similaridade de 0,14% com o Contig 110 (Contig 224 de *P. carvalhoi*) e outro com 1,60% de similaridade com o Contig 199 (Contig 251 de *P. carvalhoi*) - cujos dados não foram contabilizados anteriormente nas análises de *X. tropicalis* devido ao seu baixo bitscore nas buscas por BLAST contra o genoma dessa espécie. Entretanto, ambas as sequências mostraram porções de alinhamento considerável entre si (FIGURA 9). Adicionalmente, ambos os satélites estão em baixo número de cópias tanto em *X. tropicalis* quanto em *P. carvalhoi*, já que os DNAsats dos contigs 110 e 224 tem uma proporção no genoma de 0,031% e 0,019%, respectivamente. Já os DNAsats dos contigs 199 e 251 apresentaram uma proporção nos genomas de 0,010% e 0,015%, respectivamente. Entretanto, o DNAsat 224 — que apresentou similaridade com o contig 110 de *X. tropicalis* - não está contabilizado nas análises posteriores devido ao seu bitscore estar abaixo da linha de corte de 300 previamente estabelecida, estando entre a faixa de 39,2 e 89,7 nas buscas por BLAST contra o genoma de *P. carvalhoi*. Situação similar também ocorre para o DNAsat 251, que obteve similaridade com o contig 199 de *X. tropicalis*, estando seu bitscore entre 41 e 104. Nas buscas por BLASTn contra a coleção de sequências nucleotídicas do NCBI, não foram identificadas similaridades significantes com as famílias multigênicas ou genes de cópia única entre os 24 contigs citados.

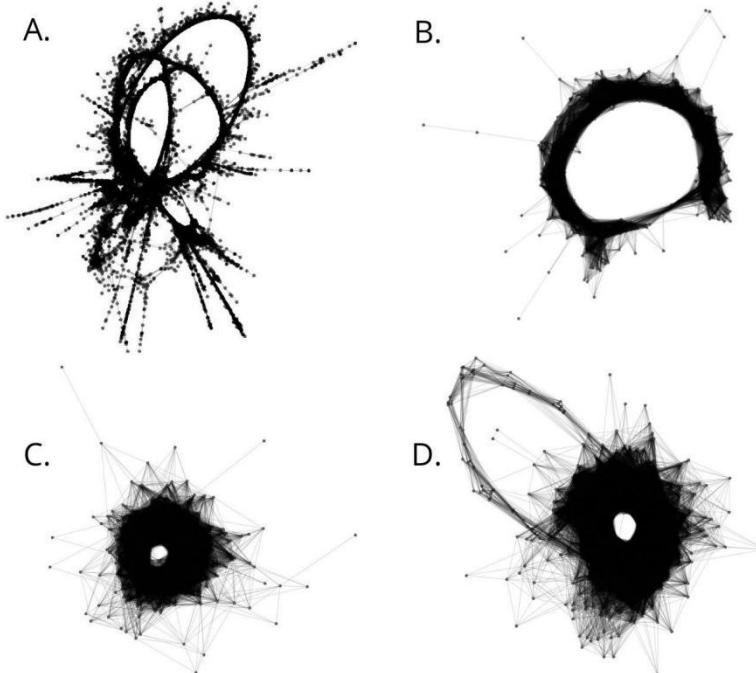
FIGURA 7 - PROPORÇÃO DE *READS* DE DNAs REPETITIVOS DE *P. carvalhoi*.



FONTE: *RepeatExplorer2*.

LEGENDA: Gráfico resultante da clusterização das *reads* provenientes do genoma de *P. carvalhoi* relacionadas à elementos de DNAs repetitivos obtidos em cada *supercluster*. No eixo Y temos o número de *reads* atribuídas para cada *supercluster* de DNA repetitivo e no eixo X indica a proporção em relação ao total de *reads* analisadas.

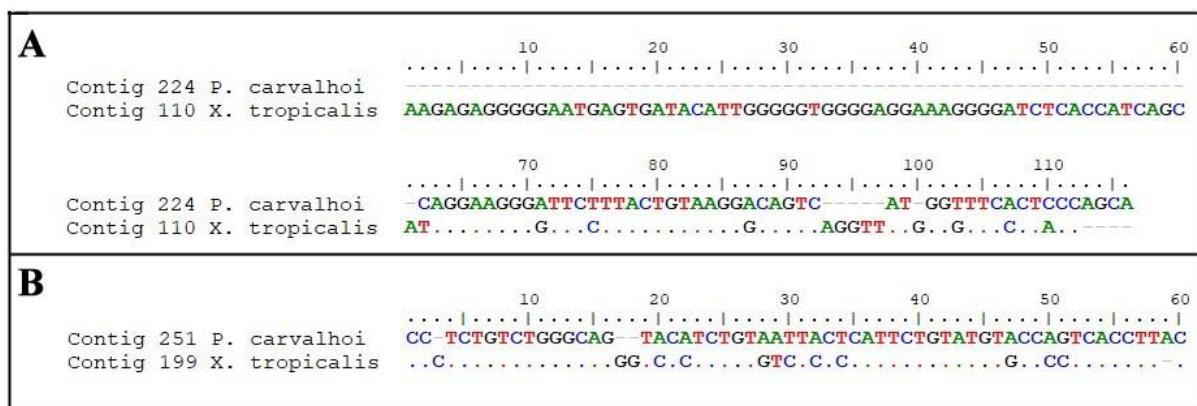
FIGURA 8 – READS DE *P. carvalhoi*



FONTE: *RepeatExplorer* 2.

LEGENDA: Gráficos das *reads* geradas pelo *RepeatExplorer* 2. A. Contig 7; B. Contig 221, C. Contig 239, D. Contig 253. Cada ponto sinalizado no gráfo é uma *read*, as linhas que conectam estes pontos são chamadas de vértices. Estão conectadas as *reads* que possuem sobreposição, ou seja, *reads* que se alinham e, em geral, pertencem a um mesmo cluster de região repetitiva.

FIGURA 9 – DNAs SATÉLITES COMPARTILHADOS ENTRE *X. tropicalis* e *P. carvalhoi*.



FONTE: A autora.

LEGENDA: Alinhamento entre as sequências de DNAs satélites dos contigs 224 de *P. carvalhoi* e 110 de *X. tropicalis* (**A**) e dos contigs 251 de *P. carvalhoi* e 199 de *X. tropicalis* (**B**). Os pontos indicam 100% de similaridade entre as sequências.

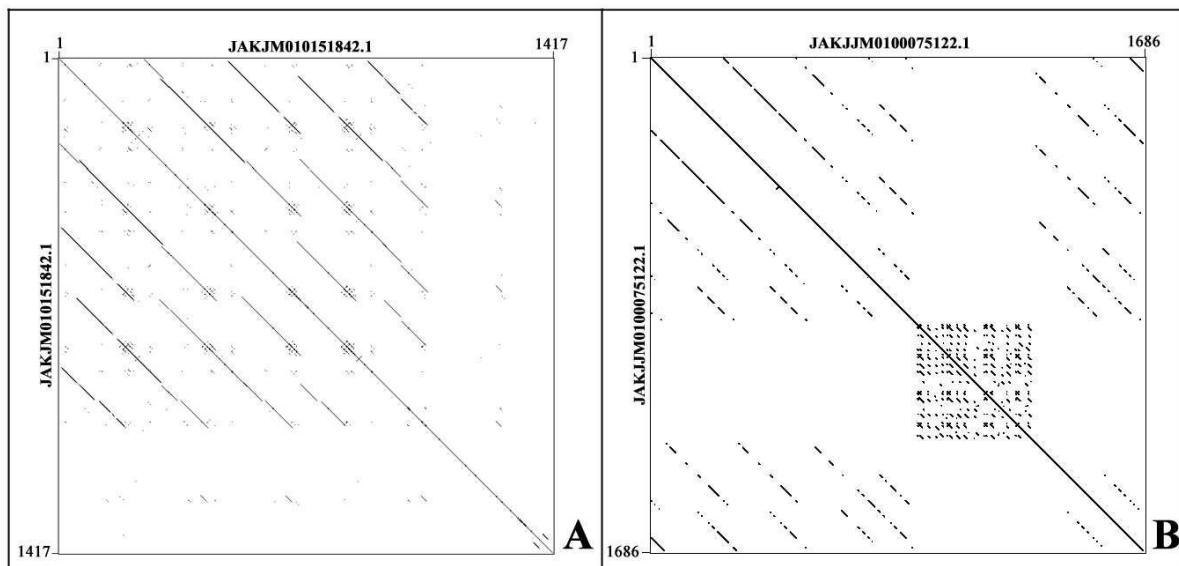
Através das buscas de BLASTn com os 24 contigs resultantes, foram utilizadas para filtragem de dados os valores de bitscore, sendo este com a linha de corte de 300, resultando em 9 contigs (ANEXO 2). Esses contigs resultantes apresentam poucas variações em seu número de cópias, resultando em 287 cópias distribuídas ao longo do genoma entre os 9 contigs (TABELA 2). Para o caso dos satélites PcarSat01-235 e PcarSat08-247, análises de *dotplot* confirmam a natureza *in tandem* dessas sequências no genoma de *P. carvalhoi* (FIGURA 10). Para outras sequências de DNAsats não foi possível a obtenção de gráficos *dotplot*, seja pelo baixo número de cópias de unidades repetitivas recuperadas após o BLAST, seja pelas unidades repetitivas serem grandes (> 1kb), não sendo possível a recuperação de mais de uma unidade repetitiva completa nos *scaffolds* da montagem do genoma de *P. carvalhoi*.

TABELA 2: DISTRIBUIÇÃO CROMOSSÔMICA, NÚMEROS DE CÓPIAS E NOMENCLATURA DE DNAs SATÉLITES CARACTERIZADOS PARA *Pipa carvalhoi*.

Contig	Tamanho da unidade repetitiva (bp)	Nomenclatura proposta	Número de cópias	Quantidade de scaffolds
152	235	PcarSat01-235	7	4
221	744	PcarSat02-744	44	35
237	2298	PcarSat03-2298	23	22
239	220	PcarSat04-220	3	3
253	1074	PcarSat05-1074	8	5
31	1828	PcarSat06-1828	69	66
7	2998	PcarSat07-2998	111	99
82	247	PcarSat08-247	20	13
9	198	PcarSat09-198	2	1

FONTE: A autora (2024)

FIGURA 10 – GRÁFICOS DOTPLOT DE ALINHAMENTOS DE SCAFFOLDS DO GENOMA DE *P. Carvalhoi* QUE CONTÉM SEQUÊNCIAS DE DNAs SATÉLITES.

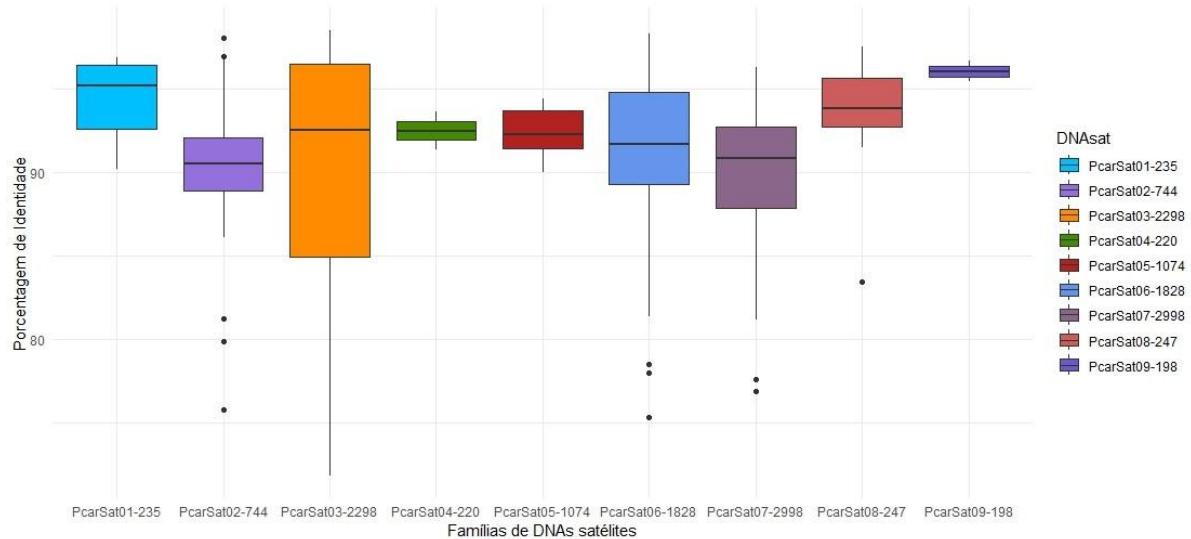


FONTE: A autora.

LEGENDA: Autoalinhamentos *dotplot* de *scaffolds* do genoma de *P. carvalhoi* que possuem os satélites PcarSat01-235 (A) e PcarSat08-247 (B) indicando a organização *in tandem* desses satélites.

Da mesma maneira que em *X. tropicalis*, nas buscas por BLAST com a utilização das sequencias consenso de *P. carvalhoi*, para cada DNAsat foi distribuído a porcentagem de identidade sendo identificado variações nas unidades repetitivas (GRÁFICO 7).

GRÁFICO 7 - ANÁLISE DE SIMILARIDADE DE DNASAT EM *P. carvalhoi*.



Fonte: a autora (2024)

Legenda: Distribuição da Porcentagem de identidade dos DNAsat de *P. carvalhoi* em gráfico de boxplot. O eixo Y representa a Porcentagem de identidade dos DNAsat e o eixo X representa as Famílias de DNAsat.

Mediante a distribuição dos dados de Porcentagem de identidade visualizados no Gráfico 5, de maneira comparativa é visível a disparidade nos graus de identidade de PcarSat03-2298 (Contig 237) e PcarSat09-198 (Contig 9), onde o grau de identidade é maior no segundo devido a dispersão das porcentagens de identidade ser menor. O mesmo pode ser dito para o PcarSat04-220 (Contig 239), PcarSat05-1074 (Contig 253), enquanto os demais — PcarSat01-235 (Contig 152), PcarSat02-744 (Contig 221), PcarSat06-1828 (Contig 31), PcarSat07-2998 (Contig 7), PcarSat08-247 (Contig 82) - apresentam uma maior dispersão das porcentagens de identidade, consequentemente um menor grau de identidade.

6. DISCUSSÃO

6.1. Perfis de DNAs satélites de *X. tropicalis* e *P. carvalhoi*

Fry e Salser (1977) em seu modelo de Biblioteca de DNAsat evidenciaram que diferentes DNAs satélites coexistem no genoma de espécies relacionadas, entretanto, são amplificadas em diferentes proporções. Este estudo infere que a

partir da variação no número de cópias dentro da biblioteca de DNAsat ancestral há a formação de perfis espécie-específicos. Ou seja, alguns DNAsats irão passar por processos de amplificação ou retração, levando a fixação de diferentes números de cópias dos DNAsat em duas linhagens aparentadas (Ugarkovic, Plohl, 2002; Goes et al, 2022). Um exemplo bastante prático é o de Lima e Ruiz-Ruano (2022) no estudo de DNAsat de 37 espécies do gênero *Drosophila*, em que inferem que as famílias de DNAsat são compartilhadas entre espécies intimamente relacionadas, suportando a hipótese de biblioteca de DNAsat. Entretanto, afirmam ainda, que há famílias de DNAsat específicas para espécies devido aos processos de amplificação das unidades repetitivas ou de abundância em algumas linhagens específicas.

Os resultados obtidos da pesquisa *in silico* das espécies alvo do projeto demonstram que não há um grande compartilhamento de famílias de DNAsats entre elas. Esse perfil de DNAsats espécie-específico para *X. tropicalis* e *P. carvalhoi* pode ser explicado devido ao tempo de divergência entre essas duas espécies, de cerca de 190 milhões de anos com a separação do supercontinente Gondwana (Cannatella et al, 2015). Dessa forma, pode ser possível que a ação de eventos estocásticos pode ter modulado a fixação de diferentes famílias de DNAsats nas duas espécies devido à alta dinâmica evolutiva que DNAsats apresentam (Ugarkovic, Plohl, 2002; Plohl et al., 2012). Apesar de os dados demonstrarem que há dois DNAsat compartilhados entre *X. tropicalis* e *P. carvalhoi*, as suas proporções no genoma de ambos é baixa, bem como a porcentagem de similaridade entre as sequências das duas espécies. Tal situação pode ser explicada pela dinâmica evolutiva das sequências de DNAsat (revisão de Charlesworth, Sniegowski, Stephan, 1994). Essa divergência ocorre de maneira gradual principalmente pelo acúmulo de substituição de nucleotídeos, sendo detectada em diferentes níveis taxonômicos (Ugarkovic, Plohl, 2002).

Apesar de perfis espécie-específicos em *X. tropicalis* e *P. carvalhoi* de DNAsats poderem ter uma maior influência de processos estocásticos na sua origem e fixação, não pode ser excluída a ação de pressões seletivas no estabelecimento de novas famílias de DNAsats (Plohl et al. 2012). Por exemplo, no estudo de Brederson et al (2024), é evidenciado que espécies filogeneticamente próximas possuem DNAsat centroméricos diferentes e isto é visualizado com a

análise do DNAsat centromérico que interage com a variante de histona Cenp-a em *X. tropicalis* e *X. laevis*. Cenp-a é reconhecida como uma variante da histona H3 específica do centrômero, onde em *X. tropicalis* é observado que a sequência do satélite centromérico é diferente da que é apresentada em *X. laevis*. Isso sugere que as sequências satélites dos centrômeros evoluem rapidamente, sendo que nesses casos a seleção natural atuou para fixar diferentes tipos de DNAsats centroméricos nessas duas espécies (Brederson et al, 2024).

Há também evidências de que TEs estão envolvidos na origem de biblioteca de DNAsat, contribuindo diretamente para a formação do DNAsat *de novo* por meio de mecanismos como duplicação, inserção adjacente a repetições existentes, e amplificação através de eventos de recombinação (Zattera, Bruschi, 2022). Por exemplo, no estudo de Talbot e Hennikof (2020) foi evidenciado que os TEs contribuem para a evolução dos neocentrômeros ao influenciar a formação *de novo* de centrômeros, devido a deposição de marcadores epigenéticos em uma região cromossômica específica, que, subsequentemente, é enriquecida com sequências repetitivas. Além de elementos de transposição, outras famílias de DNAs repetitivos, como famílias multigênicas, também podem dar origem à novos DNAsats através de eventos de duplicação e estes adquirirem aspectos funcionais, sendo assim mantidos por seleção natural em espécies bastante divergentes. Um exemplo de funcionalidade adquirida por um DNAsat vem dos estudos com a família PcP190 em anuros. Esse DNAsat possui uma região conservada, de aproximadamente 120pb, correspondente a região transcritora do DNAr 5S, inferido como sua origem devido ao grau de similaridade (Vittorazzi et al. 2011), bem como uma região hipervariável tanto em tamanho quanto em composição de nucleotídeos entre espécies ou até mesmo em uma mesma espécie (Gatto et al, 2018; Targueta et al. 2018). Recentemente, foi evidenciado por meio de análise de transcriptomas que o DNAsat PcP190 de *Rhinella marina* e *Engystomops pustulosus* (duas espécies distantes filogeneticamente) é transcrito tanto na sua região mais conservada quanto na região hipervariável, e, dessa forma, desempenha algum papel funcional no genoma dessas espécies e, provavelmente, nas outras espécies de anuros da superfamília Hyloidea em que o PcP190 está presente nos genomas (Pompeo et al. 2024).

Apesar das evidências aqui apresentadas não indicarem uma conformidade significativa com a hipótese de bibliotecas de DNAsats, as mesmas não dão suporte de que ela seja totalmente inválida, já que há duas sequências de DNAsats compartilhadas entre *X. tropicalis* e *P. carvalhoi*. Em *P. carvalhoi* houve limitação quanto às análises e seleção dos dados para rodar no *RepeatExplorer 2*, visto que, uma das limitações do *pipeline* é do número de leituras utilizadas para a análise, devido ao espaço em disco no servidor *Galaxy* do software *RepeatExplorer2* (Novak et al, 2021). Por isso, pode ser possível que com uma maior amostragem de *reads* seria possível identificar um maior número de DNAsats compartilhados entre *P. carvalhoi* e *X. tropicalis*. Ainda, sabendo que há disponibilidade de dados genômicos de outros membros dentro da família Pipidae, como para *Hymenochirus boettgerii*, que forma grupo irmão com *Xenopus* e *Pipa*, seria possível analisar os perfis de DNAsat de uma outra espécie pertencente ao grupo, buscando identificar uma maior similaridade com as duas espécies alvos do presente trabalho (Bewick et al, 2012).

6.2. Homogeneidade X heterogeneidade dos DNAs satélites de *X. tropicalis* e *P. carvalhoi*

Ao comparar uma família repetitiva em diferentes espécies, é observado que as sequências são semelhantes dentro de uma mesma espécie do que entre espécies, sugerindo que as famílias de DNA repetitivo não evoluem de forma independente, processo conhecido como evolução em concerto (Liao, 1999). Em um nível intraespecífico, as sequências repetitivas apresentam pouca variação, por outro lado em um nível interespecífico, a evolução em concerto leva a uma divergência entre os DNAs satélites em diferentes espécies (Robles et al, 2004). A atuação da evolução em concerto juntamente com a ação do processo de “*molecular drive*” levará a uma homogeneização das repetições em uma família que se espalhará por uma população, levando à sua fixação (Dover, 1986). A homogeneização das repetições, como por exemplo, por *crossing-over* desigual e por conversão gênica, pode espalhar novas mutações adquiridas por uma unidade repetitiva para todos os *repeats* de um *cluster* de DNAsats, seguido da fixação dessa nova variante em uma população, sem depender de fatores externos ou preferências por certas variantes, diferentemente de fatores estocásticos (Dover, 1982; Dover, 1986).

Analizando os dados de *X. tropicalis*, através da porcentagem de identidade entre as unidades repetitivas de cada DNAsat analisado, é possível inferir que os satélites XtrSat01-1030, XtrSat02-1609, XtrSat06-481 e XtrSat08-290 são mais homogêneos que os demais satélites. Da mesma forma que em *P. carvalhoi* os DNAs satélites PcarSat04-220, PcarSat05-1074 e o PcarSat09-198 também são considerados mais homogêneos que os outros DNAsat da espécie. Segundo Dover (1986), a distribuição das *clusters* de sequências repetitivas em vários cromossomos pode indicar que há níveis consideráveis de troca de informação em cada família de repetições, ou seja, há uma menor taxa de homogeneização, já que a frequência de recombinação homóloga é muito maior que a frequência de recombinação ectópica. Em estudos de lagartos do gênero *Varanus*, a baixa homogeneização é correlacionada com a distribuição das sequências em diferentes cromossomos, limitando assim a troca genética entre as cópias (Prakhongcheep et al, 2017). Em *X. tropicalis* ao se analisar o XtrSat04-206 com XtrSat01-1030, observa-se que há uma diferença significativa em sua amplitude de distribuição das porcentagens de identidade no GRÁFICO 6. Essa diferença de amplitude implica em uma maior homogeneidade no XtrSat01-1030 do que comparado com o XtrSat04-206. Ao se avaliar é visto que o DNAsat XtrSat01-1030 apresenta um menor número de cópias (presença nos cromossomos 1, 2, 6 e 8) que o DNAsat XtrSat04-206 (presença em todos os 10 cromossomos). De modo geral, em *X. tropicalis*, pode ser observado que DNAs satélites que estão presentes em mais de um cromossomo apresentam uma menor homogeneidade que aqueles que estão presentes em vários cromossomos.

Em *P. carvalhoi* também foi observada uma variação na distribuição das porcentagens de identidade entre a sequência consenso de DNAsats e as unidades repetitivas presentes na montagem do genoma dessa espécie. Entretanto, o padrão observado para *Xenopus* (i.e. satélites com maior número de cópias e presentes em vários cromossomos sendo menos homogêneos que satélites com menores números de cópias e localização em poucos cromossomos) não foi observado. Por exemplo, as unidades repetitivas do satélite PcarSat09-198 são muito mais similares entre si do que aquelas do DNAsat PcarSat03-2298 (GRÁFICO 7). Porém, há uma diferença no número de cópias recuperadas na montagem do genoma de *P. carvalhoi*, tendo sido recuperadas apenas duas cópias em um *scaffold* satélite

PcarSat09-188 enquanto que o satélite PcarSat03-2298 foram recuperadas 23 cópias em 22 *scaffolds* diferentes. Pode ser possível que o número amostral menor de PcarSat09-198 esteja enviesando a observação de que esse seria mais homogêneo que PcarSat03- 2298. Entretanto, a proporção do genoma inferida como tendo sequências do satélite PcarSat09-198 foi maior (0,53%) do que aquela obtida para PcarSat03-2298 (0,017%). Dessa forma, pode ser inferido que PcarSat09-198 tem um maior número de cópias, porém as mesmas são tão homogêneas que, durante o processo de montagem do genoma de *P. carvalhoi*, resultou na presença desse satélite em apenas um *scaffold*. É comum obter erros de montagem de regiões repetitivas do genoma, principalmente quando a montagem é realizada com base em bibliotecas sequenciadas com coberturas mais baixas (revisão de MASOUDI-NEJAD et al., 2013). Diferentemente de *X. tropicalis*, a montagem do genoma de *P. carvalhoi* não se encontra em nível cromossômico, sendo muito fragmentada principalmente devido a baixa cobertura de sequenciamento empregada (GAMA et al., 2022). De tal modo, utilizando estes valores de porcentagem de identidade (BLAST) e de proporção do satélite no genoma (*RepeatExplorer2*), seria possível inferir que PcarSat09-198 apresenta na verdade um maior número de cópias que o PcarSat03- 2298 e que pode ser possível que o primeiro poderia estar presente em mais cromossomos. Análises futuras envolvendo o mapeamento desse DNA satélite, bem como outras análises com base em uma maior amostragem das *reads* do genoma de *P. carvalhoi* poderão esclarecer melhor essa contradição observada entre o número de cópias desses DNAsats e os valores de identidade obtidos nas buscas por BLAST.

A evolução em concerto explica a homogeneidade das sequências repetitivas em uma mesma espécie, entretanto, não aborda adequadamente a origem, diversificação e manutenção das diferenças encontradas em espécies parentadas (revisão de Thankur, Packiaraj, Henikoff, 2021). Para tal, as análises filogenéticas podem fornecer um caminho para superar as limitações impostas pela evolução em concerto, rastreando processos evolutivos que influenciam estas sequências (revisão de Thankur, Packiaraj, Henikoff, 2021; de Lima, Hanlon, Gerton, 2020). Adicionalmente, as análises filogenéticas, juntamente com análises de diversidade nucleotídica de cada DNAsat de *X. tropicalis* e de *P. carvalhoi* podem contribuir ainda mais também para entender melhor os padrões de homogeneização das

sequências, ou seja, se haveriam *repeats* que sejam ou não cromossomo-específicos, assim como observado para o DNAsat centromérico humano alfa-satélite (Willard, 1985). Juntamente com a análises filogenéticas, o mapeamento citogenético de diferentes indivíduos da espécie auxiliaria na correlação de mudanças evolutivas, avaliando impactos estruturais e funcionais das sequências ao longo do genoma, bem como avaliar o potencial dessas sequências servirem como marcadores citogenéticos a fim de serem utilizadas no reconhecimento de homologias cromossômicas interespecíficas (revisão de Thankur, Packiaraj, Henikoff, 2021; Targueta et al., 2018). Ademais, com o isolamento de sequências de DNAs satélites de múltiplas outras espécies da família Pipidae, pode ser possível esclarecer a ação dos diferentes mecanismos de evolução que impactam na diversidade de DNAsat, auxiliando assim no entendimento dos processos de evolução genômica (revisão de Thankur, Packiaraj, Henikoff, 2021; Plohl, 2012).

7. CONCLUSÃO

Com base nos resultados, a análise comparativa dos DNAs satélites de *Xenopus tropicalis* e *Pipa carvalhoi* demonstrou uma baixa frequência de DNAsat compartilhados. Isso sugere que ambos evoluem de maneira independente, sob pressão de forças evolutivas, como por exemplo, mutação e recombinação genética. Os diferentes graus de homogeneidade observados podem estar relacionados pelo número de *clusters* cromossômicos e quantidade de cópias da sequência, onde, pelo menos em *X. tropicalis*, satélites que estão identificados em poucos cromossomos tendem a ter um grau maior de homogeneização quando comparado com os que são identificados em mais cromossomos. Da mesma forma, pode-se considerar a quantidade de cópias obtidas em cada DNAsat, onde um menor número de cópias leva a uma menor variação de porcentagem de identificação, levando a maior homogeneização.

REFERÊNCIAS

ALTSCHUL, S.F.; GISH, W.; MILLER, W.; MYERS, E.W. LIPMAN, D.J. Basic Local Alignment Search Tool. *J Mol Biol*, v. 215, n. 3 p. 403-410. 1990. Acesso em: 20 de junho de 2024.

BELYAYEV, A.; JANDOVÁ, M.; JOSEFIOVA, J.; KALENDAR, R.; MAHELKA, V.; MANDÁK, B.; KRAK, K. Natural History of a Satellite DNA Family: From the Ancestral Genome Component to Species-Specific Sequences, Concerted and Non-Concerted Evolution. *Inter J Mol Sci*. 2019. Disponível em: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6429384/>. Acesso em: 10 de junho de 2024.

BELYAYEV, A.; JANDOVÁ, M.; JOSEFIOVA, J.; KALENDAR, R.; MAHELKA, V.; MANDÁK, B.; KRAK, K. The major satellite DNA families of the diploid *Chenopodium album* aggregate species: Arguments for and against the “library hypothesis”. *Plos One*, v.15, 2020. Disponível em: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7591062/>. Acesso em: 20 de junho de 2024.

BEWICK, A.J.; CHAIN, F.J.J.; HELED, J.; EVANS, B.J. The Pipid Root. *Syst. Biology*. v.5, p.913-926, 2012. Acesso em: 16 de junho de 2024.

BOLGER, A.M.; LOHSE, M.; USADEL, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*, v. 30: p. 2114-2120, 2014.

BREDEKERSON, J.V.; MUDD, A.B. MEDINA-RUIZ, S.; MITROS, T.; SMITH, O.K.; MILLER, K.E.; LYONS, J.B.; BATRA, S.S.; PARK, J.; BERKOFF, K.C.; PLOTT, C.; GRIMWOOD, J.; SCHUTZ, J.; AGUIRRE-FIGUEROA, G.; KHOJHA, M.K.; LANE, M.; PHILIPP, I.; LASLO, M.; HANKEN, J.; KENDIVEL, G.; BUISINA, N.; SACHS, L.M.; BUCHHOLZ, D.R.; KWON, T.; SMITH-PARKER, H.; GRIDI-PAPP, M.; RYAN, M.J.; DENTON, R.D.; MALONE, J.H.; WALLINGFORD, J.B.; STRAIGHT, A.F.; HEALD, R.; HOCKEMEYER, D.; HARLAND, R.M.; ROKHSAR, D.S. Conserved chromatin and repetitive patterns reveal slow genome evolution in frogs. *Nature communication*. 2024. Acesso em: 11 de junho de 2024.

BUENO, G.P.; GATTO, K.P.; GAZOLLA, C.B.; LEIVAS, P.T.; STRUETT, M.M.; MOURA, M.; BRUSCHI, D.P. *Plos one*. 2021. Disponível em: <https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0245128>. Acesso em: 05 de novembro de 2024.

CANNATELLA, D. *Xenopus in Space and Time: Fossils, Node Calibrations, Tip-Dating, and Paleobiogeography*. *Cytogenet Genome Res*. 2015. Disponível em: <https://karger.com/cgr/article/145/3-4/283/62763/Xenopus-in-Space-and-Time-Fossils-Node>. Acesso em 13 de junho de 2024.

CHALESWORTH, B.; SNIEGOWSKI, P.; STEPHAN, W. The evolutionary dynamics of repetitive DNA in eucaryotes. *Nature*, v. 371, 1994. Acesso em: 05 de julho de 2024.

DOVER, G.A. Molecular Drive in multigene families: how biological novelties arise, spread and are assimilated. *TIG*. 1986. Acesso em: 15 de junho de 2024.

EVANS, B.J. Genome evolution and speciation genetics of clawed frogs (*Xenopus* and *Silurana*). *Frontiers in Bioscience*. v.1, p.4687-4706, 2008. Acesso em: 11 de junho de 2024.

FELLICIELO, I.; AKRAP, I.; UGARKOVIC, D. Satellite DNA Modulates Gene Expression in the Beetle *Tribolium castaneum* after Heat Stress. *PLOS Genetics*, v.11, n.9, 2015. Disponível em: <https://journals.plos.org/plosgenetics/article?id=10.1371/journal.pgen.1005466>. Acesso em: 05 de novembro de 2024.

FERRE, P.M.; BARBASH, D.A. Species-specific heterochromatin prevents mitotic chromosome segregation to cause hybrid lethality in *Drosophila*. *PLoS Biol*, v. 7, n. 10, 2009. Disponível em: <https://pubmed.ncbi.nlm.nih.gov/19859525/>. Acesso em 06 de outubro de 2024.

FROST, DARREL R. *Amphibian Species of the World: an Online Reference*. Version 6.2 (Date of access). Database eletrônico disponível em: <https://amphibiansoftheworld.amnh.org/index.php>. American Museum of Natural History, New York, USA. Doi.org/10.5531/db.vz.0001. 2024. Acesso em 16 de julho de 2024.

FRY, K.; SALSER, W. Nucleotide sequences of HS- α satellite DNA from kangaroo rat *dipodomys ordii* and characterization of similar sequences in other rodent. *Cell*, v.12, p.1069-1084, 1977. Acesso em: 17 de junho de 2024.

GAMA, J.M.; LUDWIG, A.; GAZOLLA, C.B.; GUIZELINI, D.; RECCO-PIMENTEL, S.M.; BRUSCHI, D.P. A genomic survey of LINE elements in Pipidae aquatic frogs shed light on Rex-elements evolution in these genomes. *Mol Phylogenet Evol* 168: 107393. Acesso em: 05 de dezembro de 2024.

GARRIDO-RAMOS, M.A. Satellite DNA: An Evolving Topic. *Genes*. 2017, Disponível em: <https://www.mdpi.com/2073-4425/8/9/230>. Acesso em: 12 de junho de 2024.

GEL, B.; SERRA, E. (2017). *karyoplotR*: and R/Bioconductor package to plot customizable genomes displaying arbitrary data. *Bioinformatics*, 33: 3088-3090.

GOES, C.A.G.; DOS SANTOS, R.Z.; AGUIAR, W.R.C.; ALVES, D.C.V.; SILVA, D.M.Z.A.; FORESTI, F.; OLIVEIRA, C.; UTSUNOMIA, R.; PORTO-FORESTI, F. Revealing the Satellite DNA History in *Psalidodon* and *Astyanax* Characid Fish by Comparative Satellitomics. *Front Genet*, 2022. Disponível em: <https://pmc.ncbi.nlm.nih.gov/articles/PMC9253505/>. Acesso em: 13 de outubro de 2024.

GYMREK, M. A genomic view of short tandem repeats. *Current opinion in Genetic & Development*, v.44, p.9-16. 2017. Disponível em: <https://www.sciencedirect.com/science/article/abs/pii/S0959437X16301538>. Acesso em: 11 de outubro de 2024.

I.H.G.S.C. (2001). Initial sequencing and analysis of the human genome. Acesso em: 05 de dezembro de 2024.

IRISSARI, I.; VENCES, M.; SAN MAURO, D.; GLAW, F.; ZARDOYA, R. Reversal to air-driven sound production revealed by a molecular phylogeny of tongueless frogs,

family Pipidae. Biomed Central: Evolutionary Biology. 2011. Acesso em: 11 de junho de 2024.

KRUMSIEK, J.; ARNOLD, R.; RATTEI, T. Gepard: a rapid and sensitive tool for creating dotplots on genome scale. Bioinformatics 23: 1026-1028. 2007. Acesso em: 24 de novembro de 2024.

KUHN, G.C.S.; KUTTLER, H.; MOREIRA-FILHO, O.; HESLOP-HARRISON, J.S. The 1.688 Repetitive DNA of *Drosophila*: Concerted Evolution at Different Genomic Scales and Association with Genes. Molecular Biology and Evolution, v.29, 2012. Disponível em: <https://academic.oup.com/mbe/article/29/1/7/1748314>. Acesso em: 05 de novembro de 2024.

KUMAR, S.; STECHER, G.; LI, M. MEGA X: Molecular Evolutionary Genetics Analysis across computing platforms. Mol Biol Evol, v. 35, p. 1547-1549, 2018.

LI, H.; HANDSAKER, B.; WYSOKER, A.; FENNEL, T.; RUAN, J. The Sequence Alignment/Map format and SAMtools. Bioinformatics, v. 25, p. 2078-2079, 2009.

LIAO, D. Concerted Evolution: Molecular Mechanism and Biological Implications. Am. J. Hum. Genetic. v.64, p. 24-30. 1999. Acesso em: 18 de novembro de 2024.

LIMA, L.G.; HANLON, S.L. GERTON, J. Origins and Evolutionary Patterns of the 1.688 Satellite DNA Family in *Drosophila* Phylogeny. G3 Genes, v.10, p.4129-4146. Disponível em: <https://academic.oup.com/g3journal/article/10/11/4129/6048629>. Acesso em: 02 de dezembro de 2024.

LIMA, L.G.; RUIZ-RUANO, F.J. In-Depth Satellitome Analyses of 37 *Drosophila* Species Illuminate Repetitive DNA Evolution in the *Drosophila* Genus. Genome Biol Evo, v. 14, n. 5, 2022. Disponível em: <https://pubmed.ncbi.nlm.nih.gov/35511582/>. Acesso em 01 de outubro de 2024.

LOPES-FLORES, I.; GARRIDO-RAMOS, M.A. The repetitive DNA content of Eucaryotic , Genomes. Genome Dyn, v.7, p. 1-28. 2012. Acesso em: 05 de junho de 2024.

LOWER, S.S.; MCGURK, M.P.; CLARK, A.G.; BARBASH, D.A. Satellite DNA evolution: old ideas, new approaches. Current Opinion in Genetics & Development. V. 49, p. 70-78, 2018. Disponível em: <https://www.sciencedirect.com/science/article/abs/pii/S0959437X1730120X?via%3Dhub>. Acesso em: 01 de outubro de 2024.

MASOUDI-NEJAD, A.; NARIMANI, Z.; HOSSEIKHAN, N. Next Generation Sequencing and Sequence Assembly: methodologies and algorithms. Springer-Verlag, Heidelberg, Alemanha, pp. 41 - 54. 2013. Acesso em: 05 de dezembro de 2024.

MCNULTY, S.M.; SULLIVAN, L.L.; SULLIVAN, B. Human Centromeres Produce Chromosome-Specific and Array-Specific Alpha Satellite Transcripts that Are Complexed with CENP-A and CENP-C. Dev Cell, v. 42, n. 3, p. 226-240, 2017. Disponível em: <https://pubmed.ncbi.nlm.nih.gov/28787590/>. Acesso em: 08 de outubro de 2024.

MEHROTRA, S.; GOYAL, V. Repetitive sequences in Plant nuclear DNA: Types, distribution, evolution and function. Genomics Proteomics Bioinformatics, v.12, p.

164-171, 2014. Disponível em: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4411372/>. Acesso em: 08 de julho de 2024.

MESTROVIĆ, N.; PLOHL, M.; MRAVINAC, B.; UGARKOVIĆ, D. Evolution of satellite DNAs from the genus *Palorus*—experimental evidence for the “library” hypothesis. *Mol Biol Evol.* V.15, n. 8, 1062-1068, 1998.

MEZZASALMA, M; GLAW, F.; ODIERNA, G.; PETRACCIOLI, A.; GUARINO, F.M. Karyological analyses of *Pseudhymenochirus merlini* and *Hymenochirus boettgeri* provide new insights into the chromosome evolution in the anuran family Pipidae. *Zoologischer Anzeiger*, p.47-53, 2015. Acesso em: 11 de junho de 2024.

NOVÁK, P.; NEUMANN, P.; PECH, J.; STEINHAISL, J.; MACAS, J. RepeatExplorer: a Galaxy-based web server for genome-wide characterization of eukaryotic repetitive elements from next-generation sequence reads. *Bioinformatics*, v. 29, p. 792-793, 2013.

PELICANO, P.I.M. Morfologia do Plexo Braquial e das Extremidades Digitais de *Pipa carvalhoi*, Miranda Ribeiro, 1937 (AMPHIBIA, ANURA, PIPIDAE). 2007. 34 páginas. Dissertação de Mestrado. Universidade de Brasília, Brasília, 2007. Acesso em: 15 de julho de 2024.

PEZER, Z.; BRAJKOVIC, J.; FELICIELLO, I.; UGARKOVIC, D. Satellite DNA-Mediated Effects on Genome Regulation. In.: Garrido-Ramos MA (ed). *Genome Dynamics: Repetitive DNA*. Karger, Basel, pp. 153-169. 2012. Acesso em: 15 de novembro de 2024.

PEZER, Z.; BRAJKOVIC, J.; FELICIELLO, I.; UGARKOVIC, D. Satellite DNA-mediated effects on genome regulation. *Genome Dyn.*, v. 7, p. 153-169, 2012. Disponível em: <https://pubmed.ncbi.nlm.nih.gov/22759818/>. Acesso em: 08 de outubro de 2024.

PLOHL M, LUCHETTI A, MESTROVIC N, MANTOVANI B. Satellite DNAs between selfish and functionality: Structure, genomics and evolution of tandem repeats in centromeric (hetero)chromatin. *Gene* v. 409: 72-82. 2008. Acesso em: 20 de novembro de 2024.

PLOHL, M.; LUCHETTI, A.; MESTROVIC, N.; MANTOVANI, B. Satellite DNA between selfishness and functionality: Structure, genomics and evolution of tandem repeats in centromeric (hetero) chromatin. *Gene*, p.72-82, 2008. Acesso em 15 de junho de 2024.

PLOHL, M.; MESTROVIC, N.; MRAVINAC, B. Satellite DNA Evolution. *Genome Dyn.* v.7, p.126-152, 2012. Acesso em: 17 de junho de 2024.

PLOHL, M.; PETROVIC, V.; LUCHETTI, A.; RICCI, A.; SATOVIC, E.; PASSAMONTI, M; MANTOVANI, B. Long-term conservation vs high sequence divergence: the case of an extraordinarily old satellite DNA in bivalve mollusks. *Heredity*, p. 543-551, 2010. Acesso em: 15 de junho de 2024.

PLOHL, M.; PETROVIC, V.; LUCHETTI, A.; RICCI, A.; SATOVIC, E.; PASSAMONTI, M.; MANTOVANI, B. Long-term conservation vs high sequence divergence: the case

of an extraordinarily old satellite DNA in bivalve mollusks. *Heredity* 104: 543-551. 2010 Acesso em: 16 de novembro de 2024.

POMPEO, J. N.; GATTO, K. P.; BALDO, D.; LOURENÇO, L. B. Evidence for the Transcription of a Satellite DNA Widely Found in Frogs. *Genes* 15: 1572, 2024. doi: 10.3390/genes15121572. Acesso em: 05 de dezembro de 2024.

PRAKHONGCHEEP, O.; THAPANA, W.; SUNTRONPONG, A.; SINGCHAT, W.; PATTANATANANG, K.; PHATCHARAKULLAWARAWAT, R.; MUANGMAI, N.; PEYACHOKNAGUL, S.; MATSUBARA, K.; EZAZ, T.; SRIKULNATH, K. Lack of satellite DNA species-specific homogenization and relationship to chromosomal rearrangements in monitor lizards (Varanidae, Squamata). *BMC Evolutionary Biology*, n. 193, 2017. Disponível em: <https://bmcecoevol.biomedcentral.com/articles/10.1186/s12862-017-1044-6>. Acesso em: 03 de dezembro de 2024.

PYRON, R.A.; WIENS, J.J. A large-scale phylogeny of Amphibia including over 2800 species, and a revised classification of extant frogs, salamanders, and caecilians. *Mol Phylogenet Evol*, v. 61, p. 543-583, 2011. Acesso em: 12 de junho de 2024.

R CORE TEAM (2023). R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. <<https://www.R-project.org/>>.

ROBLES, F.; HERRÁN, R.; LUDWIG, A.; REJÓN, C.R.; REJÓN, M.R.; GARRIDO-RAMOS, M.A. Evolution of ancient satellite DNAs in sturgeon genomes. *Gene*, v.338, p. 133-142. 2004. Acesso em: 18 de novembro de 2024.

ROVIC, S.; KOHLER, F.; ERDHARDT, S. Repetitive centromeric satellite RNA is essential for kinetochore formation and cell division. *J Cell*, v. 207, n. 3, 2014. Disponível em: <https://pubmed.ncbi.nlm.nih.gov/25365994/>. Acesso em: 05 de outubro de 2024.

ROZAS, J.; FERRER-MATA, A.; SÁNCHEZ-DELBARRIO, J.C. DnaSP 6: DNA Sequence Polymorphism Analysis of Large Data Sets. *Mol Biol Evol*, v. 34, p. 3299-3302, 2017.

RUIZ-RUANO, F.J.; LÓPEZ-LEÓN, M.D.; CABRERO, J.; CAMACHO, J.P.M. High-throughput analysis of the satellitome illuminates satellite DNA evolution. *Sci Rep*, v. 6: 28333. 2016. Acesso em: 16 de novembro de 2024.

TALBET, P.B.; HENIKOFF, S. What's make a centromere? *Experimental Cell Research*, v. 389, 2020. Disponível em: <https://www.sciencedirect.com/science/article/pii/S001448272030094X?via%3Dihub>. Acesso em: 20 de novembro de 2024.

TARGUETA, C.P.; VITTORAZZI, S. E.; GATTO, K.P.; BRUSCHI, D.P.; VEIGA-MENONCELLO, A.C.P.; RECCO-PIMENTEL, S.M.; LOURENÇO, L.B. Anuran Cytogenetics: an overview. In.: NORRIS, C. and MILLER, C. *An Essential Guide to Cytogenetics*. Nova Science Publishers, pp. 1 - 64, 2018.

THAKUR, J.; PACKIARAJ, J.; HENIKOFF, S. Sequence, Chromatin and Evolution of Satellite DNA. *Inter J. Mol. Sci.* 2021. Disponível em: <https://www.mdpi.com/1422-0067/22/9/4309>. Acesso em: 12 de junho de 2024.

THOMPSON, J.D.; HIGGINS, D.G.; GIBSON, T.J. CLUSTALW: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res*, v. 22, p. 4673-4680, 1994.

UGARKOVIC, D.; PLOHL, M. Variation in satellite DNA profiles - cause and effects. *The Embo Journal*, v. 21, n. 22, p. 5955-5959, 2002. Acesso em: 05 de junho de 2024.

VITTORAZZI, S. E.; LOURENÇO, L. B.; DEL-GRANDE, M. L.; RECCO-PIMENTEL, S. M. Satellite DNA derived from 5S rDNA in *Physalaemus cuvieri* (Anura, Leiuperida). *Cytogenet Genome Res*, v. 134: 101-107, 2011. Acesso em 15 de novembro de 2024.

WICKHAM, H. *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag, NewYork, ISBN 978-3-319-24277-4. 2016. Acesso em: 16 de novembro de 2024.

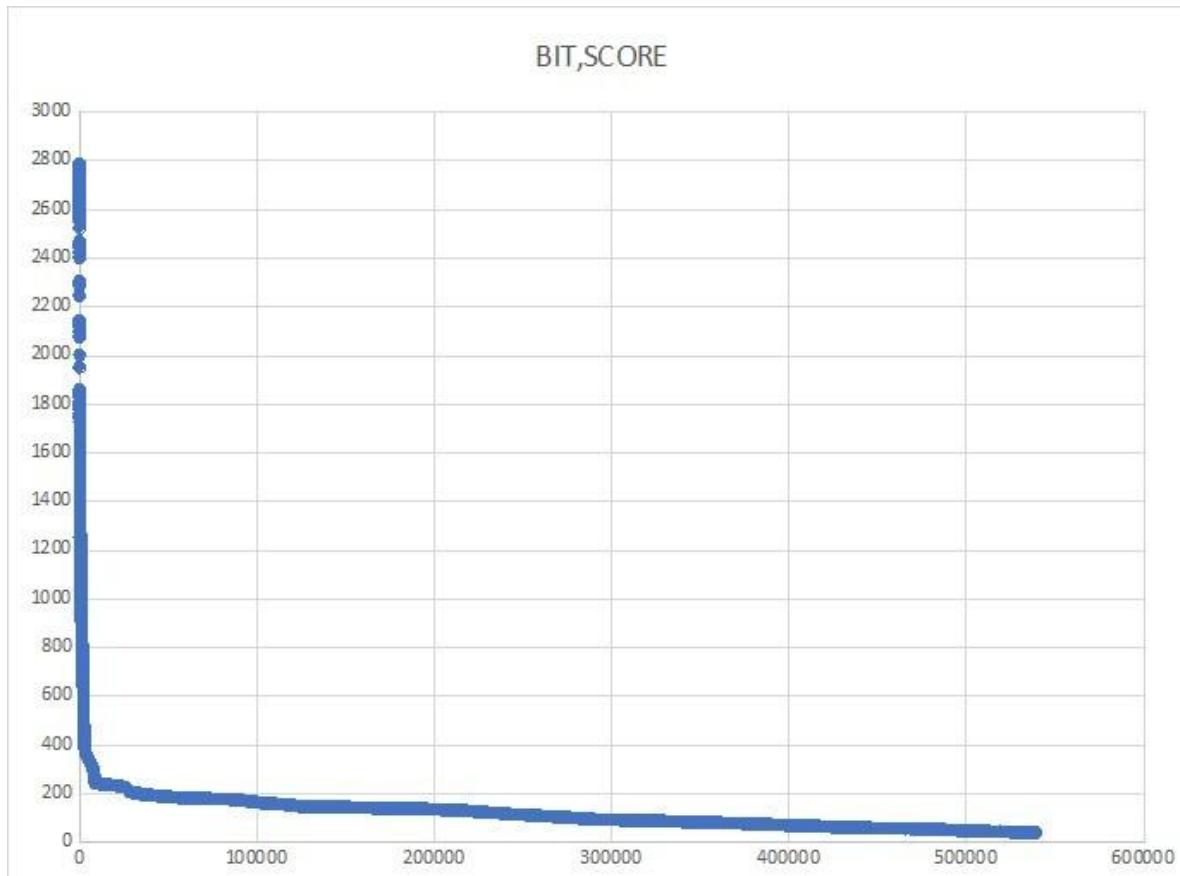
WILLARD, H.F. Chromosome-specific organization of human alpha satellite DNA. *Am J Hum Genet*, v. 37, p. 524-532, 1985. Disponível em: <https://pmc.ncbi.nlm.nih.gov/articles/PMC1684601/>. Acesso em: 25 de novembro de 2024.

YODA, K.; ANDO, S. OKUDA, A.; KIKUCHI, A.; OKAZAKI, T. In vitro assembly of the CENP-B/ Alpha-satellite DNA/core histone complex: CENP-B causes nucleososome positioning. *Genes Cells*, v. 3, n. 8, 1998. Disponível em: <https://pubmed.ncbi.nlm.nih.gov/9797455/>. Acesso em: 05 de outubro de 2024.

ZATTERA, M.L.; BRUSCHI, D.P. Transposable Elements as a Source of Novel Repetitive DNA in the Eukaryote Genome. *Cells*, v. 11, 2022. Disponível em: <https://www.mdpi.com/2073-4409/11/21/3373c>. Acesso em: 02 de dezembro de 2024.

ZATTERA, M.L.; BRUSCHI, D.P. Transposable elements as a source of novel repetitive DNA in the eukaryote genome. *Cells*, v.11, n.22, 2022. Disponível em: <https://www.mdpi.com/2073-4409/11/21/3373>. Acesso em: 25 de novembro de 2024.

ANEXO 1 - DISTRIBUIÇÃO DE VALORES DE BITScore DAS BUSCAS
POR BLAST DAS SEQUÊNCIAS CONSENSO DE DNA SATÉLITES NA
MONTAGEM DO GENOMA DE *X. tropicalis*.



ANEXO 2 - DISTRIBUIÇÃO DE VALORES DE BITScore DAS BUSCAS POR BLAST DAS SEQUÊNCIAS CONSENSO DE DNA SATÉLITES NA MONTAGEM DO GENOMA DE *P. carvalhoi*.

