



**LARA BEATRIZ OLIVEIRA**

**IDENTIFICAÇÃO DE SEQUÊNCIAS DE DNA SATÉLITE  
NO GENOMA DE ESPÉCIES DE *Piper* L.**

**LAVRAS – MG**

**2023**

**LARA BEATRIZ OLIVEIRA**

**IDENTIFICAÇÃO DE SEQUÊNCIAS DE DNA SATÉLITE  
NO GENOMA DE ESPÉCIES DE *Piper* L.**

Monografia apresentada à Universidade Federal de Lavras, como parte das Exigências do Curso de Ciências Biológicas, para a obtenção do título de Bacharel.

Prof<sup>ª</sup>. Dr<sup>ª</sup>. Giovana Augusta Torres

Orientadora

Doutoranda Liliana Rocivalda Gomes Leitão

Co-orientadora

**LAVRAS – MG**

**2023**

**LARA BEATRIZ OLIVEIRA**

**IDENTIFICAÇÃO DE SEQUÊNCIAS DE DNA SATÉLITE NO GENOMA DE  
ESPÉCIES DE *Piper* L.**

**IDENTIFICATION OF SATELLITE DNA SEQUENCES IN THE GENOME OF  
PIPER L. SPECIES**

Monografia apresentada à Universidade Federal de Lavras, como parte das exigências do Curso de Ciências Biológicas, para a obtenção do título de Bacharel.

APROVADA em 25 de maio de 2023.

Profa. Giovana Augusta Torres - UFLA

Liliana Rocivalda Gomes Leitão - UFLA

Marcelo Antônio da Trindade - UFLA

À minha mãe Eliane por me incentivar a seguir os meus sonhos

e me apoiar sempre.

Ao meu pai José Marcos por me ensinar que o conhecimento

é a maior virtude humana.

Dedico

## AGRADECIMENTOS

Agradeço pelo apoio que meus pais me deram desde o início da minha formação superior, assim como todo o sacrifício que eles fizeram para que eu pudesse completar minha graduação.

Aos professores da graduação que compartilharam seus conhecimentos não somente acadêmicos, mas lições valiosas que vou levar comigo nos próximos passos da minha vida.

Aos bons amigos que fiz na graduação é que tornaram a experiência da faculdade única e especial.

À Liliana por estar comigo e me ajudar nos meus trabalhos da Iniciação Científica e com os Estágios da graduação. Obrigada pelo tempo, paciência e lições compartilhadas.

À Profa. Giovana, por me orientar, me ensinar e me apoiar. Graças a sua dedicação poderei concluir essa etapa tão importante.

Ao Tiago que me recebeu no laboratório de Citogenética Vegetal e se dispôs a me ensinar as práticas aplicadas ao estudo dos cromossomos vegetais.

À Coordenadoria da Graduação em Ciências Biológicas, por estar disponível a ajudar os discentes e esclarecer todas as dúvidas.

Aos membros da Banca, por participarem dessa etapa fundamental da minha formação.

Ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) por ter concedido a bolsa que subsidiou a elaboração dessa monografia.

A Universidade Federal de Lavras, pela qualidade do ensino, estrutura e profissionais de qualidade que são disponibilizados aos seus estudantes.

*“A verdadeira coragem é ir atrás dos seus sonhos mesmo  
quando todos dizem que ele é impossível”.*

*(Cora Coralina)*

## RESUMO

As espécies de *Piper* L. endêmicas das regiões orientais da África e Ásia são popularmente utilizadas pelas culturas locais como matéria-prima na gastronomia e na medicina tradicional devido suas propriedades culinárias e terapêuticas. Dentre as espécies, três se destacam por apresentarem alto valor sociocultural nessas regiões: *Piper capense*, *Piper sarmentosum* e *Piper caninum*. Estudos genômicos, incluindo a caracterização da fração repetitiva, são escassos no gênero *Piper* e ausentes nessas três espécies. Na fração repetitiva, os DNAs satélites são elementos fundamentais, visto que participam de modulações gênicas que interferem nos fenótipos das plantas. Essas sequências podem ser identificadas e caracterizadas por meio da plataforma *Repeat Explorer*, a partir de dados de sequenciamento genômico. Para melhor compreensão da organização genômica de *P. capense*, *P. sarmentosum* e *P. caninum* foram caracterizadas as sequências satélites de seus respectivos genomas. Os dados de sequenciamento genômico de baixa cobertura dessas três pimenteiras foram extraídos do banco de dados *European Nucleotide Archive* (ENA) e essas sequências foram pré-processadas no *Repeat Explorer*. A ferramenta *Tarean* foi usada na identificação dos satélites e a identidade e similaridade das sequências foi determinada por meio do alinhamento MAFFT no *software* UGENE. A abundância dos satélites foi visualizada na forma de gráfico *Bubbles* construído no *software* R. Cinco DNA satélites foram identificados – PcaSat1, PcpSat2, PcpSat3, PcpSat4 e PsaSat5 – e se observou heterogeneidade no padrão de abundância, devido a variações no número de *reads* mapeados contra o genoma de cada espécie. PcaSat1 foi o único satélite compartilhado em *P. caninum* e *P. sarmentosum*, com valores de 1094 e 33 *reads*, respectivamente. Os demais satélites foram exclusivos de cada espécie. Concluiu-se que as sequências de DNA satélite de *P. capense*, *P. sarmentosum* e *P. caninum* são sequências recentes, uma vez que foram exclusivas de cada espécie.

**Palavras-chave:** *Piper capense*. *Piper sarmentosum*. *Piper caninum*. Sequência satélite. *Repeat Explorer*.

## ABSTRACT

The species of *Piper* L. endemic to the eastern regions of Africa and Asia are popularly used by local cultures as raw material in gastronomy and traditional medicine due to their culinary and therapeutic properties. Among the species, three stand out as having high sociocultural value in these regions: *Piper capense*, *Piper sarmentosum*, and *Piper caninum*. Studies about the genomes of *Piper*, including repetitive fraction, are rare and absent in these three species.. In the repetitive fraction, the satellite DNAs are fundamental elements, since they participate in gene modulations that interfere in the phenotypes of plants. These sequences can be identified and characterized using the Repeat Explorer platform from genome sequencing data. To better understand the genomic organization of *P. capense*, *P. sarmentosum* and *P. caninum*, the satellite sequences of their respective genomes were characterized. Low coverage genomic DNA of these three pepper plants were extracted from the European Nucleotide Archive (ENA) database and these sequences were preprocessed in Repeat Explorer. The Tarean tool was used in satellite identification and the identity and similarity of the sequences was determined by MAFFT alignment in UGENE software. The abundance of the satellites was visualized in the form of a Bubbles plot constructed in R software. Five DNA satellites were identified - PcaSat1, PcpSat2, PcpSat3, PcpSat4 and PsaSat5 - and heterogeneity in the pattern of abundance was observed, due to variations in the number of reads mapped against the genome of each species. PcaSat1 was the only satellite shared in *P. caninum* and *P. sarmentosum*, with values of 1094 and 33 reads, respectively. The other satellites were unique to each species. It was concluded that the satellite DNA sequences of *P. capense*, *P. sarmentosum* and *P. caninum* are recent sequences, since they were unique to each species.

**Key words:** *P. capense*. *P. sarmentosum*. *P. caninum*. Satellite DNA. Repeat Explorer.

## SUMÁRIO

|            |   |           |
|------------|---|-----------|
| <b>1</b>   | <b>INTRODUÇÃO</b> .....                                       | <b>10</b> |
| <b>2</b>   | <b>REFERENCIAL TEÓRICO</b> .....                              | <b>11</b> |
| <b>2.1</b> | <b>Família <i>Piperaceae</i></b> .....                        | <b>11</b> |
| <b>2.2</b> | <b>Gênero <i>Piper</i> L.</b> .....                           | <b>13</b> |
| <b>2.3</b> | <b>Importância econômica e cultural de <i>Piper</i></b> ..... | <b>14</b> |
| <b>2.4</b> | <b>Análise de DNA satélite</b> .....                          | <b>15</b> |
| <b>3</b>   | <b>MATERIAL E MÉTODOS</b> .....                               | <b>18</b> |
| <b>3.1</b> | <b>Pré-processamento dos dados</b> .....                      | <b>18</b> |
| <b>3.2</b> | <b>Identificação das sequências satélites</b> .....           | <b>19</b> |
| <b>3.3</b> | <b>Caracterização e nomenclatura dos satélites</b> .....      | <b>19</b> |
| <b>4</b>   | <b>RESULTADOS E DISCUSSÃO</b> .....                           | <b>19</b> |
| <b>5</b>   | <b>CONCLUSÃO</b> .....  | <b>22</b> |
|            | <b>REFERÊNCIAS</b> .....                                      | <b>23</b> |

## 1 INTRODUÇÃO

As espécies de *Piper* L. endêmicas das regiões orientais da África e Ásia são popularmente utilizadas pelas culturas locais como matéria-prima em pratos gastronômicos e na medicina tradicional devido a suas propriedades culinárias e terapêuticas (ZIMUDZI, 2008; SEIDEMANN, 2005; SALLEH et al., 2011). Dentre as espécies, três se destacam por apresentarem alto valor sociocultural nessas regiões, sendo muito utilizadas por povos locais: *Piper capense*, *Piper sarmentosum* e *Piper caninum*.

*P. capense*, encontrada em países da África e ilhas do Oceano Índico, tem preparações feitas com suas folhas para o tratamento de infecções de pele, distúrbios abdominais e poliomielite. Tais propriedades medicinais são advindas dos óleos essenciais que são extraídos de suas folhas e sementes. Esses óleos essenciais são ricos em hidrocarbonetos monoterpênicos e contêm principalmente  $\alpha$ -pineno,  $\beta$ -pineno, canfeno e sabineno em sua composição (ZIMUDZI, 2008).

*P. sarmentosum* é encontrada nas regiões tropicais do Sudeste Asiático, Nordeste da Índia e Sul da China, até as ilhas Andaman, onde estão em conservação *ex situ* (SEIDEMANN, 2005). Na culinária asiática, suas folhas e frutos são ingredientes utilizados nos mais variados pratos (SOLOMON, 1998). As folhas de *P. sarmentosum* também são usadas em medicamentos tradicionais, por conter o antioxidante naringenina que apresenta atividades antituberculose e antiplasmodiais (HUSSAIN et al., 2012).

*P. caninum*, conhecido localmente como “cabai hutan” ou “lada hantu” é encontrado principalmente no Sudeste Asiático, como na Tailândia e Malásia (SALLEH et al., 2011). Os óleos essenciais extraídos de suas folhas e caules são ricos em safrol,  $\beta$ -cariofileno,  $\beta$ -pineno e germacreno D. Esse óleo possui elevada atividade antimicrobiana, podendo ser útil na conservação de alimentos.

Devido ao alto potencial socioeconômico, torna-se importante aprofundar o conhecimento sobre a genética dessas pimenteiras, uma vez que são ainda pouco estudadas. Nesse sentido, os DNA satélites são elementos de alto valor por estarem presentes em regiões heterocromáticas funcionais do DNA, participando de modulações gênicas que interferem no fenótipo das plantas (GARRIDO-RAMOS, 2017).

Com a disponibilidade cada vez mais crescente de dados de sequenciamento genômico e de ferramentas de bioinformática especializadas na caracterização da fração repetitiva do genoma é possível obter informações mais completas sobre os tipos e abundância de sequências de DNA repetitivo, tanto de elementos transponíveis quanto de

satélite presentes nos genomas eucariotos (GARRIDO-RAMOS, 2015). Portanto, visando contribuir para o entendimento da organização genômica de *P. capense*, *P. sarmentosum* e *P. caninum*, objetivou-se identificar e caracterizar sequências de DNA satélites que compõem a fração repetitiva de seus genomas.

## 2 REFERENCIAL TEÓRICO

### 2.1 Família Piperaceae

A família Piperaceae pertence ao grupo das angiospermas basais, linhagens evolutivas de angiospermas que surgiram antes da divisão em monocotiledôneas e eudicotiledôneas (RAVEN et al., 2007). As angiospermas basais apresentam flores diferentes das eudicotiledôneas por possuírem angiospermia constantemente por secreção e não por fusão pós-genital, gineceu apocárpico, pétalas ausentes, várias peças florais, filotaxia floral espiralada ou variando entre espiralada e verticilada (SANTOS et al., 2012) (GUIMARÃES et al., 2009).

Alocam-se nesse grupo das angiospermas basais, espécies do Grado ANITA e as Magnoliídeas, que possuem a maior quantidade de características plesiomórficas das angiospermas (RAPINI, 2012). O Grado ANITA é composto pelas ordens Amborellales, Nymphaeales e Austrobaileyales, que se encontram como grupos irmãos sucessivos das demais angiospermas (MENDES e CHAVES, 2015).

No clado das Magnoliídeas estão as ordens Magnoliales (Magnoliaceae, Annonaceae), Laurales (Lauraceae), Canellales e Piperales (Piperaceae e Aristolochiaceae) (GUIMARÃES et al., 2009). Esse grupo porta características morfológicas variadas, com folhas com venação geralmente peninérvia, coriáceas e com estômatos paracíticos. As flores apresentam muitas peças periânticas dispostos de forma espiralada ou em número de três; estames frequentemente laminares, com filete pouco diferenciado da antera e conectivo geralmente bem desenvolvido (JUDD et al., 2009).

A família Piperaceae pertence à ordem Piperales e é composta por cinco gêneros *Ottonia*, *Peperomia*, *Piper*, *Pothomorphe* e *Sarcorrhachis*. Esses gêneros estão distribuídos nas regiões tropicais e subtropicais e os mais abundantes em número de espécies identificadas são *Piper* e *Peperomia*, com 265 e 166 espécies, respectivamente (ARAÚJO, 2023).

As piperáceas são bastante diversificadas, apresentando espécies com características botânicas desde árvores delgadas, arbustos, trepadeiras, até ervas terrestres e epífitas. As folhas possuem polimorfismos entre os gêneros, variando em tamanho, comprimento, forma e tipo de enervação. A estrutura floral é composta por flores bem pequenas, sésseis ou pediceladas nas plantas eretas, ou inflorescências pendentes (TEBBS, 1993).

Os frutos são tipo baga ou dupáceos, às vezes com estilos proeminentes, principalmente em *Piper*, sendo carnudos na maturidade. Com exceção de *Peperomia*, onde o ovário é unitegmico, os óvulos são solitários, ortotrópicos, basais, crassinucelares e bitegmicos. O saco uterino é tetraspórico. As sementes são minúsculas, com um endosperma esparso e um perisperma abundante. O embrião é pouco distinto e difícil de distinguir na semente madura (CRONQUIST, 1981).

As piperáceas se caracterizam por apresentar um metabolismo secundário diversificado dentro do grupo, sendo os óleos essenciais ricos em diferentes compostos (HEGNAUER, 1990). As células que armazenam esses óleos essenciais são esféricas e distribuídas por todo o parênquima vegetal. As células de óleos contêm monoterpenos, sesquiterpenos e fenilpropanos em concentrações próximas.

No Brasil, os gêneros *Ottonia*, *Peperomia* e *Piper* estão distribuídos em todas as regiões do país com muitas espécies endêmicas do território brasileiro. Nesses gêneros existem espécies nativas e exóticas com uma ampla adaptação, sendo capazes de conquistar e dominar os biomas brasileiros, permitindo uma ampla distribuição geográfica no país (GUIMARÃES et al., 2023).

De acordo com dados da flora brasileira no portal Re flora, existem 467 espécies de piperáceas das quais 298 são endêmicas, sendo a região Sudeste a mais abundante com 269 espécies, seguida do Norte com 235 espécies, Nordeste com 156 espécies, Sul com 138 espécies e Centro-Oeste com 117 espécies. Quanto aos domínios fitogeográficos, a Mata Atlântica abriga um total de 293 espécies, seguida da Amazônia com 239 espécies, o Cerrado 102 espécies, a Caatinga 29 espécies e Pantanal e o Pampa com 4 e 2 espécies, respectivamente (GUIMARÃES et al., 2023).

## 2.2 Gênero *Piper* L.

O gênero *Piper* L. abrange cerca de 1000 espécies em todo o mundo, sendo um dos mais diversos explorados pelo homem e conhecidas como pimenteiras ou videiras de pimenteiras. É composta de ervas, arbustos e cipós, a maioria com alta capacidade de dominância dentro de seu habitat. *Piper* tem uma distribuição pantropical, sendo mais frequentemente encontrado no sub-bosque das florestas tropicais de terras baixas, mas também podem ocorrer em clareiras e em zonas de vida de altitude mais elevada, como florestas nubladas (SANDEEP et al., 2019).

Grande parte das espécies de *Piper* L, cerca de 700, se encontra nas Américas, enquanto cerca de 300 espécies estão no Sul Asiático. Existem grupos menores de espécies do Pacífico Sul (cerca de 40 espécies) e da África (cerca de 15 espécies). Os grupos americanos, asiáticos e do Pacífico Sul parecem ser monofiléticos e a afinidade das espécies africanas não é clara (SANDEEP et al., 2019).

No Brasil, existem 297 espécies de *Piper* distribuídas em todo o território, sendo 192 endêmicas, com 48 variedades aceitas e 31 endêmicas. O Norte brasileiro possui o maior índice de espécies do gênero (182 spp.), cujo estado do Amazonas abriga 140 espécies. No que se refere aos domínios fitogeográficos, a Amazônia é o bioma mais rico em diversidade de espécies, com 185 espécies, seguido da Mata Atlântica com 164, o Cerrado com 68, a Caatinga com 18, e Pantanal e Pampa com 4 e 2, respectivamente (GUIMARÃES et al., 2023).

Com relação às características botânicas, as espécies podem ser arvoretas, arbustos ou subarbustos com caules lignificados, intumescidos com ramos eretos e nodosos. As folhas são alternadas membranáceas, cartáceas ou coriáceas quando *in natura*, sésseis ou pecioladas, algumas peltadas, com uma bainha curta ou estendendo-se até a lâmina. As lâminas apresentam tamanho, forma e consistência variadas, apresentando ou não glândulas translúcidas ou opacas, possuem padrão de enervação acródromo, broquidódromo, campidódromo ou eucampidódromo (YUNCKER, 1958).

A inflorescência ocorre por meio de espigas ou racemos opostos às folhas ou nas axilas foliares e podem apresentar pedúnculo com brácteas presentes. As flores são bissexuais, pequenas e protegidas por bráctea floral de variados formatos. O fruto é denominado drupa, recoberto por um pericarpo fino, delgado, obpiramidal, sulcado-tetragonal, obovoide ou trigonal, liso ou papiloso, estigmas 3-4(-9) com estiletes presentes ou não (YUNCKER, 1958).

As espécies de *Piper* são grandes produtoras de óleos essenciais, constituídos por diferentes compostos, incluindo monoterpenos, sesquiterpenos e fenilpropanoides. Piperina e chavicina são importantes metabólitos secundários provenientes das pimenteiras, sendo isolados pela primeira vez na pimenta-do-reino (*Piper nigrum*) e relatados como tendo atividades antibióticas. A piperina possui atividade antibacteriana contra uma grande gama de bactérias (JARAMILLO; MANOS, 2001).

Em Camarões, os óleos essenciais são extraídos das folhas frescas e sementes de *Piper capense*. Esses óleos são ricos em hidrocarbonetos monoterpenos e contêm principalmente  $\alpha$ -pineno,  $\beta$ -pineno, camfeno e sabineno. De suas raízes, sesquiterpeno capentina e vários neo-lignanós foram isolados. O extrato bruto apresentou importantes propriedades antibacteriana *in vitro* contra *Staphylococcus aureus*, *Streptococcus pyogenes* e *Corynebacterium xerosis* (ZIMUDZI, 2008).

No óleo essencial extraído de *Piper caninum* foram identificados 48 compostos nas folhas (77,9%) e no caule (87,0%) que se caracterizavam por elevadas proporções de fenilpropanoide, como o safrol, com 17,1% nas folhas e 25,5% no caule. Esses óleos possuem propriedades químicas de interesse para as indústrias terapêuticas, nutracêuticas e fabricantes de alimentos, além de apresentarem potencial antimicrobiano e antioxidantes (SALLEH et al., 2011).

A espécie *Piper sarmentosum* é produtora de óleo essencial, porém nenhuma pesquisa envolvendo a composição de seus óleos foi realizada até o momento, visto que os estudos se concentram em realizar a identificação de compostos feitos a partir de suas folhas, que são muito utilizados na medicina tradicional asiática (SEIDEMANN, 2005).

### **2.3 Importância econômica e cultural de *Piper***

No setor econômico, *Piper nigrum* é sem dúvida a espécie mais conhecida e comercializa ao redor do mundo, com grandes plantações cultivadas em diferentes países. Tal popularidade remonta desde antes das grandes navegações, onde os países do oriente utilizavam as pimentas culturalmente em sua culinária e medicina ancestral (SALOMÃO, 2014).

Na Ásia e em alguns países da África, folhas de *Piper betle* (outro exemplo importante do gênero) são utilizadas como mastigatório para a cura da rouquidão. A pimenta-longa da Índia (*Piper longum*) é um condimento indispensável na culinária local.

Na América Central, *Piper darienense* possui frutos com um efeito entorpecente, nos quais são usados como analgésicos para a dor de dente (TEBBS, 1993).

Soluções feitas das folhas de *Piper capense* são remédio contra dores abdominais, poliomielite, problemas dermatológicos, e na recuperação de mulheres que deram à luz. Em Camarões, as folhas são ingredientes para um preparado que auxilia no tratamento da epilepsia. Da raiz é feito um creme terapêutico para combate da paralisia que acomete os membros de pacientes que sofreram hemorragia cerebral. Dentro da medicina veterinária, um bálsamo feito dos frutos é utilizado para o tratamento de doenças nos animais domésticos como a raiva, a hérnia, a anaplasmose, colapso retal, pernas inchadas do gado e gastroenterite (ZIMUDZI, 2008).

Na culinária asiática, as folhas de *Piper sarmentosum* são ingredientes indispensáveis na confecção de inúmeros pratos locais. A tradição de enrolar carne bovina nas folhas de pimenteira surgiu no Oriente Médio, sendo introduzida na Índia pelos persas. A carne é cozida junto com várias especiarias como alho, molho de soja, erva-limão e cebola. Depois de pronto o embrulho de carne bovina pode ser servido com macarrão, arroz, frutas picadas, sopa de peixe e ervas verdes (DAVIDSON, 1999).

Na Malásia, as folhas e frutos de *Piper caninum* são utilizadas na gastronomia local como tempero essencial em diversos pratos tradicionais. Suas folhas servem de unguento para tratamento da rouquidão e problemas na garganta. Os frutos podem ser utilizados para fabricação de aromatizantes em conjunto com outras especiarias como o anis estrelado, *Illicium verum* (SALLEH et al., 2011).

## **2.4 Análise de DNA satélite**

Os elementos repetitivos são constituídos por longas cadeias repetidas de nucleotídeos e compõem a maior fração de elementos dentro dos organismos eucarióticos (até 98% em algumas espécies). Esses elementos são classificados em duas grandes famílias, as “Repetições dispersas” e as “Repetições em tandem”. As repetições dispersas são compostas por transposons e retrotransposons, ao passo que as repetições em tandem são formadas por longas matrizes de DNA satélites, que são subdivididos em satélites, minissatélites e microssatélites. As repetições dispersas possuem ferramentas de propagação próprias, que fazem a dispersão e localização desses elementos de maneiras específicas dentro do genoma (RICHARD et al., 2008).

O termo "DNA satélite" foi usado para designar sequências do genoma altamente repetitivas organizadas em tandem e cunhado por Singer em 1982. O termo inclui uma ampla diversidade de sequências que traduzem uma fração bastante variável do genoma eucariótico. Os satélites diferem uns dos outros por sua localização, comprimento da região de repetição, abundância e sequência de nucleotídeos (PECH et al., 1979).

Desde sua descoberta, o DNA satélite tem sido um dos elementos mais intrigantes do genoma. Com a descoberta do DNA repetitivo em tandem, em 1960, novas pesquisas foram sendo desenvolvidas sobre estrutura, função, organização e evolução do DNA satélite. Passamos a compreender muitas informações a respeito dos fatores moleculares, cromossômicos, biológicos e populacionais que interferem na rota evolutiva dos DNA satélites, gerando uma série de hipóteses relacionadas à origem, propagação e evolução de sequências satélites (LÓPEZ-FLORES; GARRIDO-RAMOS, 2012).

Com a criação das tecnologias de sequenciamento em massa, como o Sequenciamento de Nova Geração (NGS), e o desenvolvimento de plataformas computacionais como o *Repeat Explorer*, a pesquisa sobre DNA repetitivo, inclusive das sequências satélite, começou a ganhar destaque na biologia molecular, filogenética e citogenética vegetal. Desse modo, o Sequenciamento de Nova Geração (NGS), em conjunto com a análise *in silico* de alto rendimento dos dados de sequenciamento, estimulou a análise da porção repetitiva dos genomas eucarióticos (GARRIDO-RAMOS, 2017). O *Repeat Explorer* é um programa da *web* baseado em *Galaxy* com o objetivo de caracterizar elementos repetitivos presentes nos genomas eucarióticos a partir de sequenciamentos de baixa cobertura de sequenciadores de próxima geração (NOVÁK et al., 2013).

Uma aproximação primordialmente diferente para análise de repetição global que possibilite a identificação de repetição de novo e apropriada para utilizar leituras não montadas foi elaborada por Novak et al. (2010). Fundamenta-se na descoberta e quantificação de semelhanças entre leituras de sequências individuais particularmente obtidas por sequenciamento de próxima geração de DNA genômico cortado aleatoriamente (NOVAK et al., 2010).

O *software* caracteriza-se como um *pipeline* computacional que estabelece uma versão aperfeiçoada de agrupamento baseada em gráficos e várias ferramentas adicionais para posterior análise das repetições observadas (GOECKS et al., 2010). As leituras de sequência são pré-processadas por meio das ferramentas incluídas na plataforma (NOVAK et al., 2010).

O componente fundamental do *Repeat Explorer* é a análise de *clustering*, que efetua comparações de similaridade total de leituras de sequências curtas formando gráficos de grupo de leituras oriundas de elementos repetitivos. Tais grupos podem ser definidos de acordo com sua quantificação no genoma e similaridade com elementos repetitivos conhecidos e domínios proteicos conservados (NOVAK et al., 2010).

A união de NGS com a análise computacional, permite uma profunda análise genômica a respeito da porção repetitiva do DNA possibilitando a descoberta de novas famílias de satDNA, cujo isolamento era impossível com os métodos antigos (MELTERS et al., 2013). Essa técnica se mostrou eficaz na análise da composição repetida de genomas vegetais (MACAS et al., 2011) e animais (PAGAN et al., 2012).

O *Tarean* é uma ferramenta do *Repeat Explorer* utilizada para a identificação e análise de DNA satélites devido a sua capacidade de detectar frações repetidas em tandem. Por formarem matrizes muito longas, as sequências satélites apresentam uma organização genômica difícil de ser caracterizada em análises computacionais baseadas em genomas inteiros, logo, o *Tarean* é capaz de detectar essas sequências em leituras curtas não montadas (NOVÁK et al., 2017).

Os *clusters* encontrados são compostos pelos mais variados tipos de repetições, visto que o *Tarean* é capaz de identificar outros elementos repetitivos do DNA além dos satélites como, por exemplo, retrotransposons, transposons, DNAr, entre outros. A *pipeline* executa uma análise para procura de estruturas circulares típicas de repetições em tandem (satélites). As leituras dos *clusters* são convertidas em *k-mers*, onde os mais abundantes irão formar sequências que simbolizam os monómeros dos satélites (NOVÁK et al., 2017).

Pesquisas feitas dentro do gênero *Piper* L. para caracterização de DNA satélites ainda não foram realizadas. Entretanto, estudos com sequências satélites em Poaceae, família próxima de Piperaceae, servem como exemplo da relevância de se analisar as sequências satélites do genoma.

Para duas espécies de *Deschampsia*, a ferramenta TAREAN identificou 34 famílias de satDNA incluídas em 21 superfamílias, grande parte delas compartilhadas entre as duas espécies. Algumas dessas sequências apresentaram homologia com outras sequências satélites já conhecidas dentro da família Poaceae. Os satélites encontrados nas duas espécies de gramíneas possuem elevada homologia de sequências, porém sua distribuição nos cromossomos foi diferente entre *D. cespitosa* e *D. antarctica*. Dessa

forma, foram identificados múltiplos potenciais marcadores cromossômicos para o gênero *Deschampsia* (GONZÁLEZ et al., 2020).

### 3 MATERIAL E MÉTODOS

#### 3.1 Obtenção e pré-processamento dos dados

As sequências genômicas de *Piper capense*, *Piper sarmentosum* e *Piper caninum* foram retiradas do banco de dados *European Nucleotide Archive* (ENA), cujo o código de identificação para *Piper capense* (Pcp): SRS9011470, *Piper sarmentosum* (Psa): SRR14578844 e *Piper caninum* (Pca): SRR14578849. Os genomas das três pimenteiras foram sequenciados através do Illumina MiSeq, gerando sequências *paired end* a partir da extração de DNA das folhas. A coleta dos genomas de *P. capense* foi realizada em Camarões na África e de *P. caninum* em Queensland na Austrália, já para *P. sarmentosum* o local de coleta do seu material genético não foi informado. O tamanho dos genomas das espécies não está descrito na literatura. Essas sequências foram pré-processadas na plataforma *Repeat Explorer* para análise de *clustering* (Novák et al., 2013). A primeira etapa do pré-processamento consistiu em adicionar prefixos e sufixos em cada sequência, adotando como prefixo uma sigla que identificasse cada espécie (Pcp para *P. capense*, Psa para *P. sarmentosum* e Pca para *P. caninum*) e como sufixo “/1” para identificação da fita *forward* e “/2” para identificação da fita *reverse*.

Em seguida, realizou-se uma filtragem das sequências em um intervalo de qualidade pré-determinado de 20-90 pb. Posteriormente, foi feito o corte dessas sequências para a retirada de regiões de baixa confiabilidade cujas medidas para o corte foram individuais para cada genoma. Para *P. capense* o corte foi realizado no intervalo de 20-265 pb, para *P. sarmentosum* de 80-285 pb e *P. caninum* de 90-275 pb. Após o corte, ficaram disponíveis para a análise 147961 reads para Pcp, 183057 reads para Psa e 123590 reads para Pca. Por fim, ocorreu a etapa de entrelaçamento que uniu as sequências *Forward* e *Reverse* em um só conjunto de dados. Entre cada etapa do pré-processamento foi realizado a verificação do controle de qualidade dos *reads* adotando como padrão a escala de *Score Phred*.

### 3.2 Identificação das sequências satélites

As sequências satélites foram identificadas pela ferramenta TAREAN, adotando uma configuração de *Cluster size* de 0.001 (NOVÁK et al., 2017). Uma vez identificado os satélites, as estruturas de repetição (alta e baixa confiabilidade) foram analisadas através de gráficos *Dotplot* na plataforma *Dotmatcher – EMBOSS*, na qual as sequências foram confrontadas contra si, gerando gráficos do tipo *Dotplot* para confirmação dos arranjos em tandem (LI et al., 2015). Após a confirmação dos satélites, foram feitas buscas por homologies com sequências previamente publicadas em bancos de dados de nucleotídeos do NCBI, através da ferramenta BLASTn.

### 3.3 Caracterização e nomenclatura dos satélites

A identidade e similaridade entre as sequências satélites foram determinadas pelo alinhamento local múltiplo MAFFT (KATO; STANDLEY, 2013) no *software* UGENE (OKONECHNIKOV et al., 2012). As matrizes de identidade foram avaliadas de acordo com Ruiz-Ruano et al. (2016), na qual similaridades entre sequências satélites superiores a 95% indicam a mesma variante; similaridades entre 80% e 95% indicam diferentes variantes de uma mesma família e similaridade inferior a 80% revelam sequências distintas cuja proximidade ocorre no nível de superfamília.

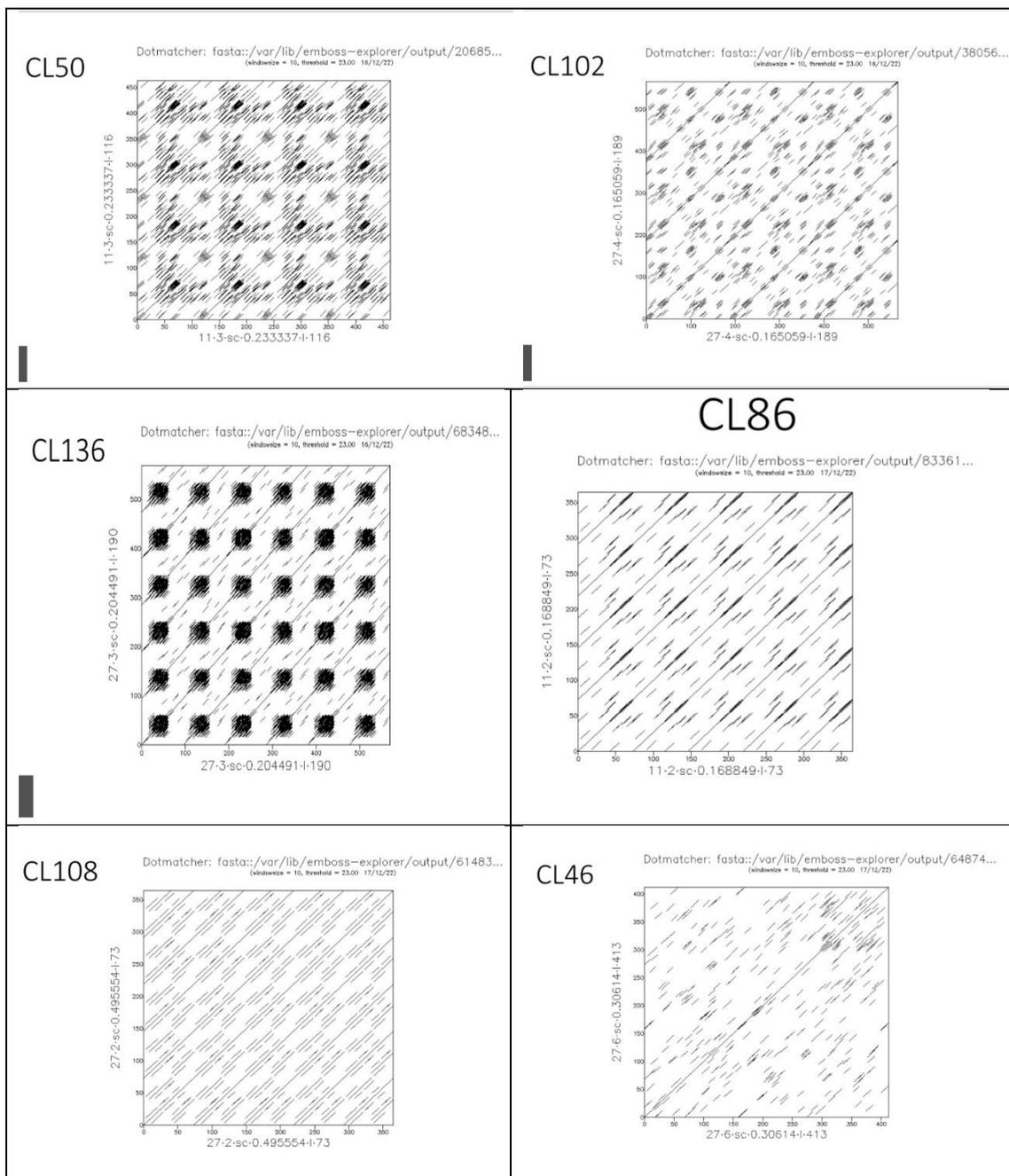
A nomenclatura adotada para as sequências satélites constituiu de um prefixo utilizado para representar cada espécie, seguido do sufixo "Sat" e posterior número para identificar cada satélite. Os DNA satélites foram identificados por comprimento e abundância de *reads*. A abundância foi quantificada por meio do mapeamento dos monômeros obtidos contra o sequenciamento do genoma de cada espécie, na ferramenta *Map to Reference* do UGENE. A abundância dos satélites foi visualizada na forma de gráfico *Bubbles* construído no *software* R, usando o pacote GGplot2 (R CORE TEAM, 2023).

## 4 RESULTADOS E DISCUSSÃO

A análise do *Tarean* identificou 17 sequências satélites, dos quais 11 eram de *Piper capense*, três de *Piper sarmentosum* e três de *Piper caninum*. Após a análise do *Dotmatcher*, foram confirmados três satélites para *P. capense*, um satélite para *P.*

*sarmentosum* e um satélite para *P. caninum* (Figura 1). Nenhum dos satélites encontrados nas três espécies apresentou homologia com as sequências de nucleotídeos depositadas no NCBI.

Figura 1 – Gráficos Dotplot confeccionados pelo *Dotmatcher*. CL50: Satélite PcpSat2, CL102: Satélite PcpSat3, CL136: Satélite PcpSat4, CL86: Satélite PsaSat5, CL108: Satélite PcaSat1, CL46: Não representa uma sequência satélite.



Fonte: Da autora (2023).

Todos os cinco satélites identificados foram classificados a nível de superfamílias por possuírem similaridades entre si e entre espécies inferiores a 80%. Nenhuma família e variante foram identificadas. O comprimento dos satélites variou de 40 pb em PcaSat1 a 304 pb em PsaSat1 e se observou heterogeneidade no padrão de abundância dos satélites, devido a variações no número de *reads* mapeados contra o genoma de cada espécie (Tabela 1).

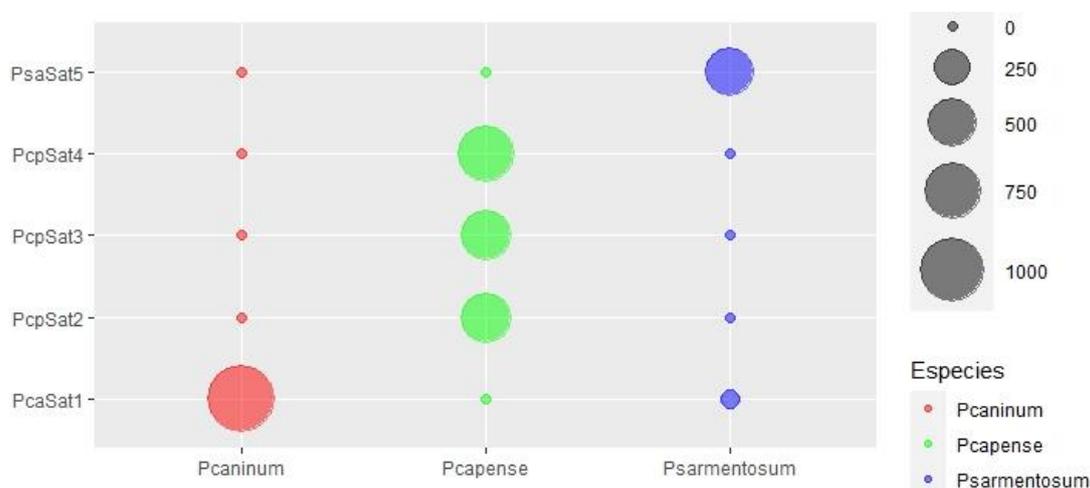
Tabela 1 – Nomenclatura, comprimento e número de *reads* das sequências satélites em *Piper caninum* (Pca), *Piper capense* (Pcp) e *Piper sarmentosum* (Psa).

| DNA satélite | Comprimento (pb) | Pca                    | Pcp | Psa |
|--------------|------------------|------------------------|-----|-----|
|              |                  | número de <i>reads</i> |     |     |
| PcaSat1      | 40               | 1.094                  | 0   | 33  |
| PcpSat2      | 152              | 0                      | 546 | 0   |
| PcpSat3      | 187              | 0                      | 568 | 0   |
| PcpSat4      | 88               | 0                      | 755 | 0   |
| PsaSat5      | 304              | 0                      | 0   | 523 |

Fonte: Da autora (2023).

PcaSat1 foi o único satélite compartilhado em *P. caninum* e *P. sarmentosum*, com valores de 1.094 e 33 *reads*, respectivamente. Os demais satélites foram exclusivos, sendo PcpSat2 (546 *reads*), PcpSat3 (568 *reads*) e PcpSat4 (755 *reads*) exclusivos de *P. capense* e PsaSat5 exclusivo de *P. sarmentosum* (Tabela 1, Figura 2).

Figura 2 - Bubble Chart de número de *reads* das sequências satélites de *P. caninum* (vermelho), *Piper capense* (verde) e *Piper sarmentosum* (azul).



Fonte: Da autora (2023).

Dos cinco DNA satélites identificados nesse trabalho, três satélites foram exclusivos de *P. capense* e um exclusivo de *P. sarmentosum*. Esses resultados indicam que são sequências recentes, que surgiram após a diversificação. Enquanto, o satélite PcaSat1 foi identificado em *P. caninum* e *P. sarmentosum*, demonstrando que se trata de uma sequência mais antiga, por ter sido compartilhada nessas espécies.

DNA satélites no gênero *Piper* e nas angiospermas basais ainda não foram relatados, sendo *Poaceae*, a família mais próxima de *Piper* na qual estudo sobre DNA satélites foram realizados. DNA satélites presentes no genoma de duas espécies de *Deschampsia* spp. pertencentes ao grupo das gramíneas, foram identificados. A maioria dos satDNAs encontrados estão presentes em ambas as espécies e mostraram alta homologia de sequência, mas sua distribuição nos cromossomos foi diferente entre as espécies. Isso destaca a importância das mudanças na disposição genômica do DNA satélite na evolução das espécies (GONZÁLEZ et al., 2020).

Tal resultado foi encontrado com base nas mesmas análises de bioinformática empregadas nesse trabalho para as espécies de *Piper* L. Os pesquisadores utilizaram os *softwares* apresentados nesse trabalho, gerando resultados que serviram de base para outras análises de citogenética, filogenética e biologia molecular, mostrando a relevância da caracterização das sequências satélite.

O uso da plataforma *Repeat Explorer* no estudo realizado com essas três espécies de *Piper* proporcionou caracterização das sequências de DNA satélites presentes nos genomas dessas espécies que até então eram desconhecidas. Tais informações, além de indicar o distanciamento entre as espécies, podem ser usadas para produção de sondas para localização nos cromossomos por meio de hibridização *in situ* fluorescente para produção de cariótipos com maior resolução para estudos citogenéticos comparativos.

## 5 CONCLUSÃO

As sequências de DNA satélite identificadas e caracterizadas nas espécies *Piper capense*, *Piper sarmentosum* e *Piper caninum* são recentes, surgindo após a diversificação entre as espécies.

## REFERÊNCIAS BIBLIOGRÁFICAS

- ARAÚJO, Marília; Família *Piperaceae*; **Infoescola "Navegando e Aprendendo"**. Disponível em: <https://www.infoescola.com/plantas/familia-piperaceae/>. Acesso em: segunda-feira, 27 de fevereiro de 2023.
- CRONQUIST, A; Um Sistema Integrado de Classificação das Plantas com Flor; **New York: Columbia University Press**; 1981.
- DAVIDSON, Alan; The Oxford Companion to Food; **Oxford University Press**. pp. 828. ISBN 978-0-19-211579-9; 1999.
- ENA. European Nucleotide Archive. **EMBL-EBI**. Disponível em: <https://www.ebi.ac.uk/ena/browser/home>. Acesso em: segunda-feira, 27 de fevereiro de 2023.
- GARRIDO-RAMOS, M.A.; DNA satélite: Um Tópico em Evolução. **Genes**, 8, 230; <https://doi.org/10.3390/genes8090230>; 2017.
- GARRIDO-RAMOS, M.A.; DNA satélites em Plantas: mais do que apenas lixo. **Cytogenet Genome Res.** 2015, 146, 153–170.
- GOECKS, J. et al.; Galaxy: uma abordagem abrangente para apoiar pesquisas computacionais acessíveis, reprodutíveis, e transparentes nas ciências da vida, **Genome Biol.**, vol. 11 pg. R86; 2010.
- GONZÁLEZ, María Laura, Jorge Chiapella, Juliana Topalian, Juan Domingo Urdampilleta, Diferenciação genômica de *Deschampsia antarctica* and *D. cespitosa* (Poaceae) baseada em DNA satélite, **Botanical Journal of the Linnean Society**, Volume 194, Issue 3, November 2020, Pages 326–341, <https://doi.org/10.1093/botlinnean/boaa045>.
- GUIMARÃES, E.F.; Carvalho-Silva, M.; Medeiros, E.V.S.S.; Monteiro, D.; Queiroz, G.A. *Piperaceae* in **Flora e Funga do Brasil**. Jardim Botânico do Rio de Janeiro. Disponível em: <<https://floradobrasil.jbrj.gov.br/FB190>>. Acesso em: 27 fev. 2023.
- GUIMARÃES et al. **Sistemática de angiospermas. 2009. (CD-room)**.
- HEGNAUER, R.; Umbelífera. In: Quimiotaxonomia de Plantas. Livros didáticos e monografias da área de Ciências Exatas, vol 31. **Birkhäuser, Basel**. [https://doi.org/10.1007/978-3-0348-9256-8\\_109](https://doi.org/10.1007/978-3-0348-9256-8_109); 1990.
- HUSSAIN, Khalid; Furqan; Hashmi, Kurshid; Latif, Abida; Ismail, Zhari; Sadikun, Amirin. "Uma Revisão da Literatura e os últimos avanços na Pesquisa de Piper sarmentosum". **Pharmaceutical Biology**. 50 (8): 1045–1052; 2012; doi:10.3109/13880209.2011.654229. PMID 22486533. S2CID 24679676.

JARAMILLO, M. Alejandra; Paul S. Manos; "Filogenia e padrões de diversidade floral no gênero *Piper* (Piperaceae)". **American Journal of Botany**. **88** (4): 706–716. (2001). doi:10.2307/2657072. JSTOR 2657072. PMID 11302858.

JUDD, W. S. et al. Sistemática vegetal: um enfoque filogenético. Porto Alegre: **Artmed**, 612p.; 2009.

KATO, K.; STANDLEY, D. M. MAFFT; Software de Alinhamento de Múltiplas Sequências Versão 7: Melhorias no Desempenho e Usabilidade. **Molecular Biology and Evolution**, v. 30, n. 4, p. 772–780, 1 abr. 2013.

LI, W.; COWLEY, A.; ULUDAG, M.; GUR, T.; MCWILLIAM, H.; SQUIZZATO, S.; PARK, Y. M.; BUSO, N.; LOPEZ, R. The EMBL-EBI; Web de bioinformática e framework de ferramentas programáticas. **Nucleic acids research**, v. 43, n. 1, p. 580-4, 1 jul. 2015.

LÓPEZ-FLORES, I.; GARRIDO-RAMOS, M.A. O conteúdo de DNA repetitivo de genomas eucarióticos. **Genome Dyn.** 2012, 7, 1–28.

MACAS, J, et al. Análise baseada em sequenciamento de próxima geração de DNA repetitivo na planta dióica modelo *Silene latifolia*; **PLoS One**, vol. 6 pg. e27335. 2011.

MELTERS, D.P.; Bradnam, K.R.; Young, H.A.; Telis, N.; May, M.R.; Ruby, J.G.; Sebra, R.; Peluso, P.; Eid, J.; Rank, D.; et al.; A análise comparativa de repetições em tandem de centenas de espécies revela percepções únicas sobre a evolução do centrômero. **Genome Biol.** 2013, 14, R10.

MENDES, Roselita Maria de Souza; CHAVES, Bruno Edson. **Livro Sistemática Vegetal: Noções básicas com enfoque em algumas famílias de angiospermas representativas no Brasil**. Edição de 2015, página 32. Disponível em: [https://educapes.capes.gov.br/bitstream/capes/176805/2/Livro\\_Ciencias%20Biologicas\\_Sistemica%20Vegetal.PDF](https://educapes.capes.gov.br/bitstream/capes/176805/2/Livro_Ciencias%20Biologicas_Sistemica%20Vegetal.PDF). Acesso em: segunda-feira, 27 de fevereiro de 2023.

NOVÁK, P, et al.; Agrupamento baseado em gráfico e caracterização de sequências repetitivas em dados de sequenciamento de próxima geração; **BMC Bioinformatics**, vol. 11 pg. 378. 2010.

NOVÁK, Petr, Laura Ávila Robledillo, Andrea Koblížková, Iva Vrbová, Pavel Neumann, Jiří Macas, TAREAN: Uma ferramenta computacional para a identificação e caracterização de DNA satélite a partir de leituras curtas desmontadas; **Nucleic Acids Research**, Volume 45, Issue 12, 7, Page e111, July 2017 <https://doi.org/10.1093/nar/gkx257>.

NOVÁK, Petr, Pavel Neumann, Jiří Pech, Jaroslav Steinhaisl, Jiří Macas, RepeatExplorer: um servidor da web baseado em Galaxy para a caracterização de todo o genoma de elementos eucarióticos repetitivos de leituras de sequência de próxima geração; **Bioinformatics**, Volume 29, Issue 6, March 2013, Pages 792–793, <https://doi.org/10.1093/bioinformatics/btt054>.

OKONECHNIKOV, K, Golosova O, Fursov M, the UGENE team. Unipro UGENE: um kit de ferramentas de bioinformática unificado. **Bioinformatics** 2012 28: 1166-1167. doi:10.1093/bioinformatics/bts091.

PAGAN, HJT, et al. O sequenciamento de pesquisa revela atividade elevada de transposon de DNA, novos elementos e variação em paisagens repetitivas entre morcegos; **Genome Biol. Evol.**, vol. 4 (pg. 575-585). 2012.

PECH, M.; Igo-Kemenes, T.; Zachau, H.G. Sequência de nucleotídeos de um componente altamente repetitivo do DNA de rato. **Nucleic Acids Res.** 1979, 7, 417-432.

RAPINI, A. **Sistemática vegetal: embriófitas** (apostila). Disponível em: Acesso em: 04 out. 2012.

RAVEN, P. H.; EVERT, R. F.; EICHHORN, S. E. **Biologia vegetal**. 7. ed. Rio de Janeiro: Guanabara Koogan, 2007. 830p.

RICHARD, G.-F.; KERREST, A.; DUJON, B. Genômica Comparativa e Dinâmica Molecular de Repetições de DNA em Eucariotos. **Microbiology and Molecular Biology Reviews**, Washington, v. 72, n. 4, p. 686-727, 2008. DOI: 10.1128/MMBR.00011-08.

RSTUDIO Team (2023). RStudio: Desenvolvimento Integrado para R. **RStudio**, PBC, Boston, MA URL <http://www.rstudio.com/>.

RUIZ-RUANO, F. J.; LÓPEZ-LEÓN, M. D.; CABRERO, J.; CAMACHO, J. P. M.; A análise de alto rendimento do satélite ilumina a evolução do DNA do satélite. **Scientific Reports**, v. 6, n. 1, p. 28333, 7 set 2016.

SAARELA, JM, Bull RD, Paradis MJ, Ebata SN, Peterson PM, Soreng RJ, Paszko B.; Filogenética molecular de gramíneas de estação fria nas subtribos Agrostidinae, Anthoxanthinae, Aveninae, Brizinae, Calothecinae, Koeleriinae e Phalaridinae (Poaceae, Pooideae, Poeae, Poeae grupo cloroplasto 1). **PhytoKeys** 87: 1-139.2017.

SALLEH, W.M.N.H.W.; Ahmad, F.; Yen, K.H.; Sirat, H.M. Composições Químicas, Atividades Antioxidantes e Antimicrobianas de Óleos Essenciais de *Piper caninum* Blume. **Int. J. Mol. Sci.** 12, 7720-7731. (2011) <https://doi.org/10.3390/ijms12117720>.

SALOMÃO, Karin. "Conheça a Rota das Especiarias". **Globo Rural**. Disponível em: <https://globorural.globo.com/Noticias/noticia/2014/02/conheca-rota-das-especiarias.html>. 2014.

SANTOS, D. Y. A. C.; CHOW, F.; FURLAN, C. M. A botânica no cotidiano. (apostila). Disponível em: [http://felix.ib.usp.br/Botanica\\_Cotidiano.pdf](http://felix.ib.usp.br/Botanica_Cotidiano.pdf). Acesso em: 05 out. 2012.

SANDEEP, Sen; Dayanandan, Selvadurai; Davis, Thomson; Ganesan, Rengaiyan; Jagadish, M.R.; Mathew, P.J.; Ravikanth, Gudalamani; ["Origem e evolução do](#)

[gênero Piper na Península da Índia](#)“(PDF). *Molecular Phylogenetics and Evolution*. 138: 102–113. doi:[10.1016/j.ympev.2019.05.033](https://doi.org/10.1016/j.ympev.2019.05.033). ISSN 1055-7903. PMID 31132521. S2CID 167218493. (September 2019).

SEIDEMANN, Johannes. Plantas de especiarias do mundo: uso econômico, botânica, taxonomia. **Springer**. pp. 292. ISBN 978-3-540-22279-8. (2005).

SOLOMON, Charmaine; Enciclopédia de Comida Asiática; **Asia Source**, 1998.

TEBBS, M.C. Piperaceae. In: Kubitzki, K., Rohwer, J.G., Bittrich, V. (eds); Plantas com Flores · Dicotiledôneas. As famílias e gêneros das plantas vasculares, vol 2. **Springer, Berlin, Heidelberg**. [https://doi.org/10.1007/978-3-662-02899-5\\_60](https://doi.org/10.1007/978-3-662-02899-5_60). (1993).

YUNCKER, T. G. As *Piperaceae* - um perfil de família. **Brittonia** 10: 1-7. 1958.

ZIMUDZI, C., Piper capense L.f. In: Schmelzer, G.H. & Gurib-Fakim, A.; Recursos Vegetais da África Tropical, **Prota**, Wageningen, Netherlands. (2008). Accessed 27 February 2023.

